

CONVEX NONDIFFERENTIABLE OPTIMIZATION: A SURVEY FOCUSSED ON THE ANALYTIC CENTER CUTTING PLANE METHOD.

JEAN-LOUIS GOFFIN* AND JEAN-PHILIPPE VIAL**

HEC/Logilab Technical Report 99.02
February 1999

Abstract. We present a survey of nondifferentiable optimization problems and methods with special focus on the analytic center cutting plane method. We propose a self-contained convergence analysis, that uses the formalism of the theory of self-concordant functions, but for the main results, we give direct proofs based on the properties of the logarithmic function. We also provide an in depth analysis of two extensions that are very relevant to practical problems: the case of multiple cuts and the case of deep cuts.

We further examine extensions to problems including feasible sets partially described by an explicit barrier function, and to the case of nonlinear cuts. Finally, we review several implementation issues and discuss some applications.

Keywords Nondifferentiable optimization, Analytic center, Cutting Plane Methods.

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.

The paper has been presented at the 2nd School on Optimization (December 10-18,1998), Universidad de Chile, Santiago, Chile.

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

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

1. Introduction. In a famous example, Wolfe [87] showed that standard optimization methods fail to solve a simple convex nondifferentiable optimization problem. Since nondifferentiable problems often arise as the result of some mathematical transformation –e.g., Lagrangian relaxation or Benders decomposition– of an originally large, and perhaps intractable, problem, there is a definite need for efficient methods. This inspired and gave rise to a significant and sophisticated literature. Our goal is to review this field, with a special focus on a cutting plane method based on the concept of analytical centers.

Convex nondifferentiable optimization can be characterized by the nature of the information, which can be collected and used in the design of algorithms. Essentially, one has to replace differential by subdifferential calculus. At a test point, one can compute subgradients of functions, i.e., supports to the epigraphs, and linear separators between the point and the feasible set. One often calls *oracle* the mechanism that computes the cutting planes. Given a sequence of trial points, the oracle produces cutting planes that provide a polyhedral outer approximation of the problem. The main issue in the design of algorithms is to make a clever use of this information to construct trial points.

Many methods exist. Some use part of the information (e.g., the most recently generated cutting planes), some use all of it. We shall provide a brief survey of the main approaches. A linear approximation often turns out to be a very poor representation of underlying nonlinearities. Cutting plane algorithms face the risk of choosing trial points, which look attractive with respect to the polyhedral approximation, but turn out to be almost irrelevant for the original problem. Therefore, most methods propose a regularizing mechanism. Methods of centers exploit this idea: centers are less sensitive to the introduction of cutting planes than are the extreme points of the polyhedral approximation.

Among many possible centers, the analytic center of a polytope has a number of advantages. The analytic center is an analytic function of the data describing the polytope, and thus its sensitivity to changes in this data is highly predictable. Furthermore, interior point methods offer highly efficient and robust methods to compute these centers. The analytic center cutting plane method (ACCPM) has been built on that principle. The theory underlying the method has been studied in depth by several authors, who provided complexity estimates for the basic method and for several of its enhancements. The method has been implemented; it has been also applied with success to a wide variety of problems. Our intent here is to collect many results and put them in a unified format to give the reader a general perspective on the theoretical and practical properties of the method.

The paper is organized as follows. In Section 2, we review the best-known sources of nondifferentiable optimization problems. In Section 3, we present the most prominent methods for solving them. In Section 4, we analyze the basic version of ACCPM, in the context of a convex feasibility problem, endowed with an oracle producing one central cut at a time. The analytic center is associated with the logarithmic barrier, and the analysis relies very much on the properties of this function. The more general class of self-concordant functions shares many properties of the logarithmic function: it can be used to define more general analytic center cutting plane methods. Our presentation of the method does not apply this larger class. Although we focus on the logarithmic barrier, we used formulations that can be extended to self-concordant functions, making it easy to establish links with more recent developments. In Section 5, we review two extensions that are very relevant in practice: the case of multiple

cuts and deep cuts. We further examine extensions to problems involving feasible sets partially described by an explicit barrier function, and to the case of nonlinear cuts. Finally, in Section 6, we discuss implementation issues, and briefly mention some applications.

Throughout the paper we use notation that is becoming more or less standard in the field of interior point methods.

The vector of all ones is denoted $e = (1, \dots, 1)^T$. To a given a vector x , we associate the diagonal matrix $X = \text{diag}(x)$. It is often necessary to take the component-wise product of two vectors. This is the so-called Hadamard product, which we denote

$$s \bullet x = Sx = SXe = Xs = x \bullet s.$$

The scalar product between vectors a and b is denoted $\langle a, b \rangle$. We may thus write

$$\langle a, b \rangle = \langle a \bullet b, e \rangle.$$

Given a positive definite matrix H , we define the norm associated with it by

$$\|a\|_H = \langle Ha, a \rangle.$$

If $F : R^n \mapsto R$ is a twice differentiable function, we denote its gradient as F' and its Hessian as F'' .

In presenting complexity estimates, we use the traditional notation $O(\cdot)$, “order of”. In some results, the bound on the number of iterations is the solution of a complicated equation, with no closed form solution. We use the notation $O^*(\cdot)$ to denote the dependance of the dominant terms in the solution.

2. Sources of nondifferentiable problems.

2.1. Lagrangian relaxation. The problem under investigation is

$$(2.1) \quad \min \{f(x) \mid h(x) \leq 0, x \in X\},$$

where f is convex and h is a vector-valued convex function, while the set X is arbitrary, and may, for instance, include integrality restrictions.

The standard *Lagrangian* is

$$L(x, u) = f(x) + \langle u, h(x) \rangle,$$

where u is the vector of Lagrange multipliers or dual variables.

The weak duality relationship

$$\min_{x \in X} \max_{u \geq 0} L(x, u) \geq \max_{u \geq 0} \min_{x \in X} L(x, u),$$

holds under no assumptions (the convexity of f and g is not even necessary). The left-hand side of the equation is the optimal value of the original problem, that is:

$$\min_{x \in X} \max_{u \geq 0} L(x, u) = \min \{f(x) \mid h(x) \leq 0, x \in X\}.$$

The dual function $L(u) = \min_{x \in X} L(x, u)$ is a concave nondifferentiable function, taking values in the extended reals. The dual problem $\max_{u \geq 0} L(u)$ always provides a lower bound on the original problem. If X is convex and classical regularity conditions are satisfied, the optimal value of the dual problem equals the optimal value of the

original problem. If X is finite then $L(u)$ is a piecewise linear function, but the number of pieces is likely to be exponential.

The key observation is that if $\bar{x} \in X$ is such that $L(\bar{u}) = L(\bar{x}, \bar{u})$, then $L(u)$ satisfies the subgradient inequality

$$L(u) \leq L(\bar{u}) + \langle h(\bar{x}), u - \bar{u} \rangle, \forall u \geq 0.$$

The Lagrangian relaxation is attractive if the problem $\min_{x \in X} L(x, u)$ is easy to solve and u is of moderate size. A typical situation where this property holds is when X is a Cartesian product and $L(x, u)$ is separable on this product space. The case where X is not convex, as, for instance, including integrality constraints, has led to a host of very successful applications. See, for instance, Held and Karp [43, 44] in their solution to the travelling salesman problem, Graves [42] in hierarchical planning, Geoffrion [29], Shapiro [78] and Fisher [24].

2.2. Dantzig-Wolfe column generation. Let $f(x) = \langle c, x \rangle$ and $h(x) = Ax - b$, and assume that $X = \{x : Dx \leq d\}$. X can be represented by the convex hull of its extreme points $\{x^k : k \in I\}$ plus the conic hull of its extreme rays $\{x^k : k \in J\}$.

This allows us to express problem (2.1) as the following program:

$$(2.2) \quad \begin{aligned} \min_{\lambda} \quad & \sum_{k \in I \cup J} \lambda_k \langle c, x^k \rangle \\ \text{s.t.} \quad & \sum_{k \in I \cup J} \lambda_k Ax^k \leq b, \\ & \sum_{k \in I} \lambda_k = 1, \\ & \lambda_k \geq 0 \quad \text{for all } k \in I \cup J. \end{aligned}$$

The dual of (2.2) is:

$$(2.3) \quad \begin{aligned} \max_{z, u} \quad & z - b^T u \\ \text{s.t.} \quad & z \leq \langle c, x^k \rangle + \langle u, Ax^k \rangle \text{ for all } k \in I, \end{aligned}$$

$$(2.4) \quad \begin{aligned} & 0 \leq \langle c, x^k \rangle + \langle u, Ax^k \rangle \text{ for all } k \in J, \\ & u \geq 0. \end{aligned}$$

The cardinality of I and J is likely to be exponential.

Dantzig-Wolfe decomposition [13, 14, 25] is directly related to Lagrangian relaxation. Let $L(u)$ be the dual function:

$$L(u) = \min_x \{ \langle c, x \rangle + \langle u, Ax - b \rangle \mid Dx \leq d \},$$

where $u \geq 0$ is a vector of Lagrange multipliers. Clearly, the so-called optimality constraints (2.3) define a set of extreme supports for $L(u)$. (Any support of L at $u \in \text{dom } L$ is a convex combination of the extreme supports.) Similarly, the so-called feasibility constraints (2.4) define an envelope of $\text{dom } L$.

Dantzig-Wolfe decomposition is thus equivalent to Lagrangian relaxation. It can also be extended to the case of X not convex, for instance by including integrality restrictions, as in first outlined in Gilmore and Gomory [30, 31].

2.3. Benders decomposition. Benders decomposition [7] deals with the problem

$$\begin{aligned} \min \quad & f(x) + g(y) \\ & h(x) + k(y) \leq 0, \\ & x \in X, \quad y \in Y, \end{aligned}$$

where $X \subset R^n$ is an arbitrary set, $Y \subset R^p$ is convex, $f : X \mapsto R$ and $g : Y \mapsto R$ are convex, $h : R^n \mapsto R^m$ is convex and $k : Y \mapsto R^m$ is convex. Finally, for the sake of simplicity, we assume that g, h and k are continuously differentiable. Let us introduce the function

$$(2.5) \quad Q(x) = \min \{ g(y) \mid k(y) \leq -h(x), y \in Y \}.$$

Q may take infinite values, is convex and usually nondifferentiable.

If we assume that the problem defining Q has a finite optimum and satisfies the usual regularity conditions, then, if \bar{u} is an optimal multiplier in the definition of Q ,

$$Q(x) \geq Q(\bar{x}) + \langle Dh(\bar{x})^T \bar{u}, x - \bar{x} \rangle, \forall x \in X,$$

where, $Dh(\bar{x})$ is the Jacobian of h at \bar{x} .

The equivalent nondifferentiable problem, under the assumption that Q is finite, is $\min_{x \in X} \{ f(x) + Q(x) \}$, where the value of Q and of one subgradient is computed by the solution of (2.5). Feasibility cuts are introduced to account for the x values for which the Problem (2.5) is infeasible. A complete description is given in [85] in the case of stochastic linear programming.

Benders decomposition can also be used in the case where X is not convex, in particular the set of integers. But this does not fall within the context of this paper as the master problem formulation is a nondifferentiable optimization problem with integer constraints.

2.4. L_p optimization. The L_p -norm is defined as $\|x\|_p = (\sum_{j=1}^n |x_j|^p)^{1/p}$, with $1 \leq p < \infty$. The case $p = \infty$ is defined by $\|x\|_\infty = \max_{j=1, \dots, n} |x_j|$. The case $p = 2$ corresponds to the standard Euclidean norm.

The L_p approximation problem is defined as

$$\min_{x \in R^n} f(x),$$

where

$$f(x) = \sum_{i=1}^m w_i \|A_i x - b_i\|_p,$$

and the w_i 's are positive weights.

A specific example would be the weighted single facility location problem, also known as the Fermat–Weber problem [86], with

$$f(x) = \sum_{i=1}^m w_i \|x - b_i\|_p,$$

where x is the location of a new facility and b_i represent the existing facilities.

Extensions to multiple facilities would be described as

$$f(x) = \sum_{k=1, \dots, q; l=1, \dots, q} w_{k,l} \|x^k - x^l\|_p + \sum_{i=1, \dots, m; k=1, \dots, q} w_{k,i} \|x^k - b^i\|_p,$$

where the n -dimensional vectors $(x^k)_{k=1, \dots, q}$ represent the new facilities. Convex constraints on the facility locations could be added.

In each of these cases the function $f(x)$ is convex and nondifferentiable, and the computation of $f(x)$ and a subgradient $g(x)$ is quite easy.

Some of the L_p problems can be solved directly by interior point methods. It is shown in [70] that there exist self-concordant barriers for inequalities of the type $t_i \geq |x_i|^p$, with $p \geq 1$. Whenever the formulation allows replacing the L_p norm by their p -th power, the problem can be solved by interior point methods. This remark does not apply to functions that are convex combinations of L_p norms, with an exception for the case $p = 2$. Then the so-called ice-cream cone inequality $t \geq \|x\|$ is endowed with the self-concordant barrier $\log(t^2 - \|x\|^2)$.

2.5. Variational inequalities. Let $H(x)$ be a multi-valued monotone operator defined on a compact convex set $Q \subset R^n$. The strong formulation of the variational inequality problem is

$$\text{find } x^* \in Q \text{ and } h_* \in H(x^*) : \langle h_*, x^* - x \rangle \leq 0, \forall x \in Q.$$

The strong formulation raises issues about existence of solutions and computational schemes. For these reasons, one may prefer the weak formulation

$$\text{find } x^* \in Q : \langle h_x, x^* - x \rangle \leq 0, \forall x \in Q, h_x \in H(x).$$

Since H is monotone, any strong solution is a weak solution. The converse is not true in general. However, a sufficient condition for it is that H be either single-valued and continuous, or multi-valued and maximal monotone [57, 70].

Let X_w^* be the set of solutions for the weak formulation of the VI problem. This set can be conveniently represented by means of the so-called ‘‘gap’’ function [4],

$$\varphi(x) = \max_u \{ \langle h_u, x - u \rangle \mid u \in Q, h_u \in H(u) \}.$$

The gap function is convex and continuous, but nondifferentiable. Clearly, $\varphi(x) \geq 0$, and $\varphi(x) = 0$ if and only if $x \in X_w^*$. However, $\varphi(x)$ is defined as the optimal value of a nonconvex problem: explicit values cannot be computed, and of course, no subgradient is available. However, one can easily construct a separating plane in the horizontal space x . Indeed, given $x = \bar{x} \in Q$ and $h_{\bar{x}} \in H(\bar{x})$, the cut

$$\langle h_{\bar{x}}, x - \bar{x} \rangle \leq 0, x \in Q$$

defines a subset of Q containing the solution set X_w^* .

2.6. Semi-infinite programming. Semi-infinite programs pertain to the class of problems

$$\min \{ f(x) \mid g(x, t) \leq 0, t \in T \},$$

where f and g convex in x , and T is a set of infinite or exponential cardinality.

The oracle at the point \bar{x} finds a point $\bar{t} \in T$ such that $g(\bar{x}, \bar{t}) > 0$, and generates the cut $g(\bar{x}, \bar{t}) + \langle g'_x(\bar{x}, \bar{t}), x - \bar{x} \rangle \leq 0$. The cut might be computed by solving $\max_{t \in T} g(\bar{x}, t)$. This is often achieved by dynamic programming (for example in the cutting stock problem [30, 31]), or by global optimization.

2.7. Linear matrix inequalities and eigenvalue optimization. The linear matrix inequality problem (LMI) [9, 10] attempts to find a solution to the problem $A(x) = A_0 + \sum_{i=1}^n x_i A_i \succeq 0$, where $(A_i)_{i=0,\dots,n}$ are symmetric $K \times K$ real matrices, $(x_i)_{i=1,\dots,n}$ are scalars and the symbol “ \succeq ” represents an ordering on the cone of symmetric real matrices, i.e., $A \succeq B$ means that the symmetric matrix $A - B$ is positive semi-definite. The solution set $S = \{x \mid A(x) \succeq 0\}$ of our problem of interest can be stated differently as $S = \{x \mid \langle A(x)h, h \rangle \geq 0, \forall h \in R^K\}$.

The condition $A(x) \succeq 0$ can conveniently be described by means of the eigenvalues. Assume, for simplicity, that at a given x , an eigenvalue decomposition of $A(x)$ has been computed and let $(\lambda_i)_{i=1,\dots,K}$ be the set of eigenvalues. The matrix $A(x)$ can be written

$$A(x) = V \Lambda V^T = \sum_{k=1}^K \lambda_k v_k v_k^T,$$

where $\Lambda = \text{Diag}(\lambda_k)_{k=1,\dots,K}$ is the diagonal matrix of eigenvalues and $(v_k)_{k=1,\dots,K}$ is a set of unit orthogonal eigenvectors of $A(x)$. It is well-known that $A(x) \succeq 0$ if and only if $\lambda_{\min} = \min_{i \leq K} \{\lambda_i\} \geq 0$. The constraint $A(x) \succeq 0$ is equivalent to $\lambda_{\min}(A(x)) \geq 0$.

Given an arbitrary x , define the vector

$$h(x) = \sum_{k=1}^K \delta_k v_k, \quad \text{with } \delta_k = \begin{cases} 1 & \text{if } \lambda_k < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, $h = 0$ if $x \in S$; else, $\langle A(x)h(x), h(x) \rangle < 0$. If $h \neq 0$, we can construct the valid cut

$$\langle A(z)h(x), h(x) \rangle = \langle A_0 h(x), h(x) \rangle + \sum_{i=1}^n z_i \langle A_i h(x), h(x) \rangle \geq 0, \quad \forall z \in S.$$

Alternative methods using the interior point framework, and the theory of self-concordant functions or barriers on the cone of positive semi-definite matrices, have been used with significant success [10]. The comparative efficiency of the nondifferentiable optimization schemes reviewed in this paper is still to be tested.

3. Review of existing methods.

3.1. Canonical formulations.

3.1.1. Optimization. The generic optimization problem amenable to a cutting plane approach is given by:

$$\min \{f(x) \mid x \in C\},$$

where C is a simple bounded set, such as a cube, or possibly a sphere. The case of a cube would lead to the use of linear programming, while a sphere requires the use of quadratic programming.

The convex function f is described by a supporting *oracle*, that is: for every $\bar{x} \in C$ a subroutine returns the value $f(\bar{x})$ and an arbitrary element $g(\bar{x})$ of the subdifferential $\partial f(\bar{x})$ of f at \bar{x} .

The subgradient inequality implies by convexity that the set

$$\{(z, x) \mid f(\bar{x}) + \langle g(\bar{x}), x - \bar{x} \rangle \leq z \leq f(\bar{x})\}$$

contains the optimal set in the space of the epigraph, while the set

$$\{x \mid \langle g(\bar{x}), x - \bar{x} \rangle \leq 0\}$$

contains the optimal points in the horizontal space of the variable x .

3.1.2. Feasibility. The generic feasibility formulation attempts to compute a point in a set S , where S is contained in a bounded set C , say a cube, and is assumed to contain an interior. The set S is defined by a separation oracle which for every $\bar{x} \in C$ returns either a statement that \bar{x} is feasible, or a vector $g(\bar{x})$ which separates the point \bar{x} from the set S , i.e.:

$$S \subset \{x \mid \langle g(\bar{x}), x - \bar{x} \rangle \leq 0\}.$$

We shall refer to this as a *pure separation* feasibility problem. To discuss complexity, we will assume that a sphere of radius ϵ is contained in S .

A strict separation occurs if there is a negative scalar γ such that

$$S \subset \{x \mid \langle g(\bar{x}), x - \bar{x} \rangle \leq \gamma\}.$$

In many practical cases the set S is defined by the functional constraint, $S = \{x \mid h(x) \leq 0\}$, where h is a convex function. A separator at $\bar{x} \notin S$, i.e., $h(\bar{x}) > 0$, obtains by considering taking any element $g(\bar{x})$ of the subdifferential $\partial h(\bar{x})$. The inequality

$$S \subset \{x \mid \langle g(\bar{x}), x - \bar{x} \rangle \leq -h(\bar{x})\},$$

provides a strict separation. Cuts such as these have been called *deep cuts*. The case of multiple functional constraints $h_i(x) \leq 0$, $i = 1, \dots, m$, can be transformed into the single functional constraint $h(x) \leq 0$ with $h(x) = \min_{i \leq m} h_i(x)$.

3.1.3. Optimization and feasibility. The generic optimization problem amenable to a cutting plane approach is given by:

$$\min \{f(x) \mid x \in S\},$$

where $S \subset C$, and C is a simple bounded set, such as a cube. The convex function f is described by a supporting *oracle*, as in section 3.1.1 that is: for every $\bar{x} \in C$ a subroutine returns the value $f(\bar{x})$ and an arbitrary element $g(\bar{x})$ of the subdifferential $\partial f(\bar{x})$ of f at \bar{x} .

The set S could be described by either one of:

- a projection operator,
- a function $h(x)$,
- a pure separation oracle.

We note that the classical description of Dantzig-Wolfe decomposition leads by duality to a supporting cut for the objective and a separation for the feasibility set.

3.2. Solution methods. The solution methods that use the cutting plane approach have been developed for different versions of the oracles. We will try, where it makes sense, to describe them in the case of the pure separation feasibility oracle.

An algorithm will use some or all of the information generated up to now to choose a new point where to call the oracle.

3.2.1. Kelley-Cheney-Goldstein cutting plane [51, 12]. To deal with the pure feasibility case, we define the $(k+1)$ -st iterate as a solution of the linear program,

$$\min\{z \mid z \geq \langle g_i, x - x_i \rangle, \forall i \leq k ; x \in C\};$$

the oracle either answers that $x_{k+1} \in S$, i.e., x_{k+1} solves the feasibility problem, or returns g_{k+1} with the property that $\{x \mid \langle g_{k+1}, x - x_{k+1} \rangle \leq 0\}$ contains the set S . In the latter case the outer approximation to S is updated to

$$\{x \mid 0 \geq \langle g_i, x - x_i \rangle, \forall i \leq k + 1 ; x \in C\}.$$

This method may fail to converge in the case of a pure separation, if the $\|g_i\|$'s are not bounded. To illustrate the point we consider the simple problem with $S = \{1/8\}$ and $C = [0, 1]$. The oracle returns at $x \neq 1/8$ the separating vector

$$g(x) = \begin{cases} \frac{2x}{x-1/4} & \text{if } \frac{1}{4} < x \leq 1, \\ 1 & \text{if } \frac{1}{8} < x \leq \frac{1}{4}, \\ -1 & \text{if } 0 \leq x < \frac{1}{8}. \end{cases}$$

If the first cutting plane is generated at $x = 0$ and the second one at \bar{x} , with $1/4 < \bar{x} \leq 1$, the algorithm will compute the next point as the minimizer of:

$$\min\{z \mid z \geq -x, z \geq \langle \frac{2\bar{x}}{\bar{x}-1/4}, x - \bar{x} \rangle, x \in C\};$$

The minimizer of this is

$$x = \frac{\bar{x} + 1/4}{2},$$

showing that the sequence generated by the algorithm converges to $x = 1/4$.

For convergence to be guaranteed one must make the assumption that the feasible set is described by a Lipschitzian convex constraint. Nevertheless, the complexity of this method is very poor [67].

If the problem to be solved is that of optimization

$$\min \{f(x) \mid x \in C\},$$

then the next iterate x_{k+1} is defined as the solution of the LP:

$$\min\{z \mid z \geq f(x_i) + \langle g_i, x - x_i \rangle, \forall i \leq k ; x \in C\}$$

where $g_i \in \partial f(x_i)$. The oracle will then answer: $f(x_{k+1})$ and $g_{k+1} \in \partial f(x_{k+1})$, and this new supporting plane will be added to the current approximation to f .

The piecewise linear function $f_k(x) = \max_{i \leq k} f(x_i) + \langle g_i, x - x_i \rangle$ is a lower approximation to f , and clearly $f_k \leq f_{k+1} \leq f$. The method converges, but the estimate in $O(1/\epsilon^n)$ is very poor [9], and can be as bad as this [67].

3.2.2. Chebyshev centers or largest sphere. In the pure feasibility case, the next point where the separation oracle is called, x_{k+1} , solves the linear program

$$\min\{z \mid z \geq \langle \frac{g_i}{\|g_i\|}, x - x_i \rangle, \forall i \leq k\}.$$

This is a linear program, and x_{k+1} is the center of the largest sphere contained in:

$$\{x \mid 0 \geq \langle g_i, x - x_i \rangle, \forall i \leq k\}.$$

This simply amounts to a variant of Kelley's cutting plane method, where the separators g_i 's have been scaled to have unit Euclidean norm. It has significantly better complexity properties than Kelley's method. In fact the example of the previous section will be solved by this algorithm, indicating power of the idea of centering.

It has been described in the optimization case by Elzinga-Moore ([23]).

3.2.3. Subgradient optimization. This is usually described in the case of the unconstrained problem $\min f(x)$, and the iteration is given by:

$$\min\{\langle -g_k, x - x_k \rangle + \frac{1}{2t_k} \|x - x_k\|^2\}.$$

This leads to $x_{k+1} = x_k - t_k g_k = x_1 - \sum_{i=1}^k t_i g_i$. By recursive substitution, one gets an equivalent formulation for the subgradient step

$$\min\{\sum_1^k t_i \langle -g_i, x - x_i \rangle + \frac{1}{2} \|x - x_1\|^2\}.$$

Many proposals, both heuristic and theoretical, for the choice of stepsizes t_k have been made; see [80, 43, 75].

If the problem is constrained, i.e., $\min\{f(x) \mid x \in S\}$, where $S \subset C$, and if a projection map Π_S is available, then the most used version of the subgradient algorithm is: $x_{k+1} = \Pi_S(x_k - t_k g_k)$. The convergence estimate is $O(1/\epsilon^2)$, for an optimal choice of the stepsize [67].

3.2.4. Bundle methods and level set bundle methods. The usual problem solved by bundle methods is: $\min\{f(x) \mid x \in C\}$.

After k iterations the piecewise linear function $f_k(x) = \max_{i=1, \dots, k} f(x_i) + \langle g_i, x - x_i \rangle$ has been computed. If x_{k+1} was selected to be the minimizer of f_k over C , then this would be Kelley's cutting plane method. The fundamental idea of bundle methods is to add a quadratic regularization to f_k and define the next point where to call the oracle \bar{x}_{k+1} as the minimizer of $f_k(x) + \frac{1}{2t_k} \|x - x_k\|^2$ over C . In most implementations of bundle methods, the next iterate may not be moved to \bar{x}_{k+1} from x_k unless a descent of f is thus achieved; see the works by Lemaréchal [58, 46] and Kiwiel [54]. A very nice feature of bundle methods is their ability to limit the number of cutting planes involved in the definition of f_k .

Note that if only the last cutting plane is used in the definition of f_k , i.e. $f(x_k) + \langle g_k, x - x_k \rangle$ then this reduces to subgradient optimization.

The specific rules for choosing t_k are beyond the scope of these notes.

To describe the level bundle methods, for which a full complexity is available, one has to define:

- the best recorded function value $f_k^* = \min_{i \leq k} f(x_i)$;
- the minimum value of $f_k(x)$ over C , denoted as \underline{f}_k ; note that this involves the computation of the point used by Kelley's method.

The optimal value of the original problem f^* lies between \underline{f}_k and f_k^* .

The next point where to call the oracle, and next iterate x_{k+1} , is defined by the solution of the quadratic programming problem:

$$\min_{x \in C} \{\|x - x_k\|^2 \mid f_k(x) \leq \alpha f_k^* + (1 - \alpha) \underline{f}_k\},$$

where $0 < \alpha < 1$. The next iterate is thus chosen as the projection of the current point x_k on a level set of the piecewise linear function $f_k(x)$ restricted to C ; the level is chosen as a combination between the best recorded function value and “Kelley’s value”. A complexity analysis was given in [57] as $O(1/\epsilon^2)$.

If the functional f_k is reduced to the last supporting plane $f(x_k) + \langle g_k, x - x_k \rangle$ then this method is again a version of subgradient optimization, with x_{k+1} defined as:

$$x_{k+1} = \Pi_C \left(x_k - \frac{g_k}{\|g_k\|^2} (f_k^* - \underline{f}_k) (1 - \alpha) \right).$$

3.2.5. The ellipsoid method [80, 67]. Here it is assumed that a sphere E_0 contains the set S . At a current iteration k an ellipsoid E_k with center x_k is available. The ellipsoid E_k contains the set S and is defined as $E_k = \{x \mid (x - x_k)^T H_k^{-1} (x - x_k) \leq 1\}$. The oracle is called at x_k , and if x_k does not belong to S , the separator g_k is returned.

The next iterate x_{k+1} is the center of the ellipsoid E_{k+1} which is the minimum volume ellipsoid containing $E_k \cap \{x \mid \langle g_k, x - x_k \rangle \leq 0\}$. Clearly, by induction, $E_{k+1} \supset S$. The set E_{k+1} is defined by

$$x_{k+1} = x_k - \frac{1}{n+1} \frac{H_k g_k}{\sqrt{g_k^T H_k g_k}},$$

$$H_{k+1} = \frac{n^2}{n^2 - 1} \left(H_k - \frac{2}{n+1} \frac{H_k g_k g_k^T H_k}{g_k^T H_k g_k} \right),$$

where we let n to be the dimension of x .

It can be shown that the volume of E_{k+1} equals the volume of E_k reduced by the factor $(1 - 1/(n+1)^2)$. This immediately leads, if we assume that S contains a sphere of radius ϵ to a complexity of $O(n^2 \log(1/\epsilon))$.

3.2.6. Centering methods. A variety of centering methods have been proposed such as

1. the center of gravity method [59]
2. the center of the maximum volume ellipsoid [83, 52]
3. the volumetric center [84, 2].

These methods all have polynomial complexity $O(n \log(1/\epsilon))$ as measured by the number of calls to the separation oracle; but there is no evidence available today that they are effective in practice, as the computation of a new center appears to be rather time consuming.

4. The analytic center cutting plane. From now on, the paper is devoted to an in-depth analysis of the analytic center cutting plane method. The convergence of the method was first studied in [3]. The scheme involves cut elimination and the proof is rather involved. However, the authors obtain a convergence estimate in $O(n(\log 1/\epsilon)^2)$ calls to the oracle. In the same journal issue, the paper [68] analyzes a version of the method with a proximal term for the minimization of a convex function with uniformly bounded derivatives. The bound in the number of iterations is then $O(L^2 R^2 / \epsilon^2)$, where L is the Lipschitz constant and R the diameter of a ball

containing the optimal set. Finally, the method with linear cuts and box constraints, but no proximal term, is studied in [34] in the framework of feasibility problems. The convergence estimate is $O^*(n^2/\epsilon^2)$. Additional convergence results may be found in [1, 55]. We study here a similar method; see [18] for a slightly different viewpoint.

The problem under study is

$$\text{Find } y \in C \cap [0, 1]^n,$$

where C is a closed convex set in R^n . The set C is described by an oracle which returns the cut $\langle a, \bar{y} - y \rangle \geq 0$ at $\bar{y} \notin C$. We assume throughout $\|a\| = 1$.

As shown in Section 3, the cutting plane method calls the oracle at a point lying in a polytope. This polytope varies as the iterations proceed. We shall use the general notation

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

to represent the polytope. In this definition A is an $n \times m$ matrix, $c \in R^m$ and the variable y in which the minimization is done lies in R^n . We assume that A has full row rank and that \mathcal{F}_D is bounded and has a nonempty interior. We conveniently associate the vector of slacks variables $s = c - A^T y \in R^m$ with $y \in \mathcal{F}_D$.

The analytic center of \mathcal{F}_D is the unique solution of the minimization problem:

$$(4.1) \quad \min\{\varphi_D(s) = -\sum_{i=1}^n \log s_i \mid A^T y + s = c, s > 0\}.$$

Let us introduce the notation

$$F(y) = \varphi_D(c - A^T y).$$

The function F is the so-called logarithmic barrier function. It belongs to the class of self-concordant logarithmically homogeneous barriers for the set \mathcal{F}_D which are studied in [70]. Clearly, $\text{dom } F = \text{int } \mathcal{F}_D$. The analytic center of \mathcal{F}_D is the minimizer of F

$$y^* = \arg \min F(y).$$

The basic step of the analytic center cutting plane method with approximate centers¹ is defined as follows:

¹The concept of approximate minimizer will be made precise in the next section.

Initialization

$0 < \eta < 1$ is the centering parameter.

$$\mathcal{F}_D^1 = \{y \mid 0 \leq y \leq e\}.$$

$$F_1(y) = -\sum_{i=1}^n \log y_i(1 - y_i).$$

Basic Step

Let y^k be an approximate minimizer of F_k

$$(\text{For } k = 1, \text{ take } y^1 = \frac{1}{2}e = \arg \min F_1(y).)$$

The oracle returns a^k .

$$\text{Update: } \mathcal{F}_D^{k+1} = \mathcal{F}_D^k \cap \{y \mid \langle a^k, y - y^k \rangle \leq 0\}.$$

$$F_{k+1}(y) = F_k(y) - \log \langle a^k, y^k - y \rangle.$$

The method, and its convergence analysis, rely on the theory of interior point methods. We need thus present some useful results from this theory.

4.1. Useful results from interior point methods. Analytic centers can be found via a damped Newton's method. To this end, we recall the derivatives of F

$$F'(y) = As^{-1} \quad \text{and} \quad F''(y) = AS^{-2}A^T.$$

The Newton step with respect to F is

$$p(y) = -[F''(y)]^{-1}F'(y) = -(AS^{-2}A^T)^{-1}As^{-1}.$$

The analytic center is uniquely defined by the condition $F'(y) = 0$, i.e., $As^{-1} = 0$.

The Hessian $F''(y)$ and its inverse define norms that are extensively used to measure proximity. For instance, given y and $\bar{y} \in \mathcal{F}_D$, we introduce the notation

$$\|\bar{y} - y\|_y = \|\bar{y} - y\|_{F''(y)}.$$

Note that

$$\begin{aligned} \|\bar{y} - y\|_y &= \langle (AS^{-2}A^T)(\bar{y} - y), \bar{y} - y \rangle^{\frac{1}{2}}, \\ &= \langle S^{-1}A^T(\bar{y} - y), S^{-1}A^T(\bar{y} - y) \rangle^{\frac{1}{2}}, \\ &= \|s^{-1} \bullet (\bar{s} - s)\|. \end{aligned}$$

The condition $\|s^{-1} \bullet (\bar{s} - s)\| < 1$ (or $\|\bar{y} - y\|_y < 1$) on all $\bar{s} = c - A^T\bar{y}$, $\bar{y} \in \mathcal{F}_D$, defines the so-called Dikin ellipsoid around s (or y). For this reason, we name $\|\bar{y} - y\|_y$ the Dikin distance between \bar{y} and y at y .

The following lemmas are due to Nesterov [69]. They hold for the more general case of self-concordant functions [70]. Proofs are also given in [72].

LEMMA 4.1. *Let $y, \bar{y} \in \text{int } \mathcal{F}_D$. Then*

$$(4.2) \quad \frac{\|\bar{y} - y\|_y}{1 + \|\bar{y} - y\|_y} \leq \|\bar{y} - y\|_{\bar{y}}.$$

Moreover, if $\|\bar{y} - y\|_y < 1$, then

$$(4.3) \quad \|\bar{y} - y\|_{\bar{y}} \leq \frac{\|\bar{y} - y\|_y}{1 - \|\bar{y} - y\|_y}.$$

Lemma 4.1 relates the two Dikin distances between y and \bar{y} taken at y and \bar{y} respectively.

LEMMA 4.2. *For any $s, \bar{s} \in \mathcal{F}_D$ and corresponding y and \bar{y} , we have*

$$(4.4) \quad \langle F'(\bar{y}) - F'(y), \bar{y} - y \rangle \geq \frac{\|\bar{y} - y\|_y^2}{1 + \|\bar{y} - y\|_y}.$$

$$(4.5) \quad F(\bar{y}) \geq F(y) + \langle F'(y), \bar{y} - y \rangle + \omega(\|\bar{y} - y\|_y),$$

where $\omega(t) = t - \log(1 + t)$.

Inequalities in the opposite direction also hold, under the provision that \bar{y} belongs to Dikin's ellipsoid around y .

LEMMA 4.3. *For any $y, \bar{y} \in \mathcal{F}_D$ such that $\|\bar{y} - y\|_y < 1$ we have*

$$(4.6) \quad \langle F'(\bar{y}) - F'(y), \bar{y} - y \rangle \leq \frac{\|\bar{y} - y\|_y^2}{1 - \|\bar{y} - y\|_y}.$$

$$(4.7) \quad F(\bar{y}) \leq F(y) + \langle F'(y), \bar{y} - y \rangle + \omega_*(\|\bar{y} - y\|_y),$$

where $\omega_*(t) = -t - \log(1 - t)$.

Lemmas 4.1 to 4.3 can be proved directly without resorting to the theory of self-concordant functions. For the sake of completeness we give in the Appendix a direct proof of Lemma 4.1 and repeat the proofs of [69] for Lemmas 4.2 and 4.3.

A damped Newton method taken at $y \in \text{int}\mathcal{F}_D$ is defined by $y^+ = y + \alpha p(y)$, where $p(y) = -[F''(y)]^{-1}F'(y)$ and $\alpha > 0$ is the stepsize. In the convergence analysis of Newton's method, the norm

$$\|p(y)\|_{F''(y)} = \langle [F''(y)]^{-1}F'(y), F'(y) \rangle^{\frac{1}{2}}$$

plays a critical role. The following lemma, sometimes called potential reduction lemma, is a direct consequence of Lemma 4.3.

LEMMA 4.4 (Linear convergence). *Assume $y \in \text{int}\mathcal{F}_D$ and $\|p(y)\|_{F''(y)} \geq \eta > 0$. Let $\alpha = (1 + \|p(y)\|_{F''(y)})^{-1}$ and $y^+ = y + \alpha p(y)$. Then, $y^+ \in \text{int}\mathcal{F}_D$ and*

$$F(y^+) \leq F(y) - \eta + \log(1 + \eta).$$

Proof. Rewrite (4.7) using $\langle F'(y), y^+ - y \rangle = -\alpha^2 \|p(y)\|_{F''(y)}^2$ and $\alpha = 1/(1 + \|p(y)\|_{F''(y)})$. ■

COROLLARY 4.5. *The number of damped Newton steps to reach a point y^+ with $\|p(y^+)\|_{F''(y^+)} < \eta$ from a point $y \in \text{int}\mathcal{F}_D$ is bounded by*

$$K = \left\lceil \frac{F(y) - F^*}{\eta - \log(1 + \eta)} \right\rceil,$$

where $F^* = \min F(y)$.

The above convergence result can be complemented with a finer² local analysis.

LEMMA 4.6 (Quadratic convergence). *Assume $\|p(y)\|_{F''(y)} \leq \eta < 1$ and let $y^+ = y + p(y)$. Then $y^+ \in \text{int}\mathcal{F}_D$ and*

$$\|p(y^+)\|_{F''(y^+)} \leq \|p(y)\|_{F''(y)}^2.$$

COROLLARY 4.7. *Assume $\|p(y)\|_{F''(y)} \leq \eta < 1$. Let y^* be the analytic center of \mathcal{F}_D . Then,*

$$\|y^* - y\|_y \leq \frac{\eta}{1-\eta}$$

$$F(y^*) \leq F(y) \leq F(y^*) + \frac{\eta^2}{1-\eta^2}.$$

The proofs of Lemma 4.6 and Corollary 4.7 can be found in [76] or [90].

4.2. Convergence of the cutting plane scheme. In the description of the analytic center cutting plane method, we used the statement: “ y^k is an approximate minimizer of F_k ”. We need first to make this statement precise.

Let $0 < \eta < 1$. An η -center of \mathcal{F}_D , equivalently an η -minimizer of F , is any point $y \in \mathcal{F}_D$ such that $\|p(y)\|_{F''(y)} \leq \eta$.

To prove convergence, we shall proceed with three steps. In a first step, we consider the optimal value of F on \mathcal{F}_D and study how it varies when a new cut is added. In a second step, we show that the cumulated variations of these optimal values yield to a bound on the number of calls to the oracle. In the third step, we propose a damped Newton scheme to compute a new approximate center after a cut has been added; we show that this last operation can be done in a number of steps bounded above by an absolute constant.

4.2.1. Behavior of the barrier function. Let us introduce the two quantities

$$\sigma_1 = \frac{-(1+\eta) + \sqrt{(1-\eta)^2 + 4}}{2(1-\eta)}, \quad \text{and} \quad \sigma_2 = \frac{1+\eta + \sqrt{(1-\eta)^2 + 4}}{2(1-\eta)}.$$

One checks that σ_2 and $-\sigma_1$ are the two solutions of the quadratic equation

$$(1-\eta)t^2 - (1+\eta)t - 1 = 0.$$

The functions σ_2 and σ_1 are respectively increasing and decreasing on $0 \leq \eta < 1$. The following bounds are easily computed

$$\frac{1}{2} < \sigma_1 \leq \frac{\sqrt{5}-1}{2} < \frac{\sqrt{5}+1}{2} \leq \sigma_2.$$

For $\eta = 0$, one has $\sigma_2 = \frac{\sqrt{5}+1}{2}$ and $\sigma_1 = \frac{\sqrt{5}-1}{2}$; while $\lim_{\eta \rightarrow 1} \sigma_2 = +\infty$ and $\lim_{\eta \rightarrow 1} \sigma_1 = \frac{1}{2}$.

In the sequel we shall denote F_k^* the minimum value of F_k on \mathcal{F}_D^k .

LEMMA 4.8. *Assume y^k is an η -center of \mathcal{F}_D^k and y_*^{k+1} the exact center of \mathcal{F}_D^{k+1} . Then*

$$\sigma_1 \leq \|y_*^{k+1} - y^k\|_{y^k} \leq \sigma_2,$$

²The results are slightly stronger in the case of the logarithmic barrier than for general self-concordant functions.

where $\|y_*^{k+1} - y^k\|_{y^k} = \|y_*^{k+1} - y^k\|_{F_k''(y^k)}$.

Proof. The first optimality condition on $F_{k+1}(y) = F_k(y) - \log\langle a^k, y^k - y \rangle$ is

$$F_{k+1}'(y_*^{k+1}) = F_k'(y_*^{k+1}) + \frac{a^k}{\langle a^k, y^k - y_*^{k+1} \rangle} = 0.$$

Multiplying by $y^k - y_*^{k+1}$, and adding $\langle -F_k'(y^k), y_*^{k+1} - y^k \rangle$ on both sides of the equation yields

$$(4.8) \quad \langle F_k'(y_*^{k+1}) - F_k'(y^k), y_*^{k+1} - y^k \rangle = 1 - \langle F_k'(y^k), y_*^{k+1} - y^k \rangle.$$

Note that

$$\langle F_k'(y^k), y_*^{k+1} - y^k \rangle = \langle [F_k''(y^k)]^{-\frac{1}{2}} F_k'(y^k), [F_k''(y^k)]^{\frac{1}{2}} (y_*^{k+1} - y^k) \rangle.$$

Since $\|F_k'(y^k)\|_{[F_k''(y^k)]^{-1}} = \|p(y^k)\|_{F_k''(y^k)} \leq \eta$, by Cauchy-Schwarz inequality:

$$(4.9) \quad -\eta \|y_*^{k+1} - y^k\|_{y^k} \leq \langle F_k'(y^k), y_*^{k+1} - y^k \rangle \leq \eta \|y_*^{k+1} - y^k\|_{y^k}.$$

Denote $r = \|y_*^{k+1} - y^k\|_{y^k}$. By Lemma 4.2 and (4.9),

$$1 + \eta r \geq \langle F_k'(y_*^{k+1}) - F_k'(y^k), y_*^{k+1} - y^k \rangle \geq \frac{\|y_*^{k+1} - y^k\|_{y^k}^2}{1 + \|y_*^{k+1} - y^k\|_{y^k}}.$$

Thus $(1 - \eta)r^2 - (1 + \eta)r - 1 \leq 0$, implying $\|y_*^{k+1} - y^k\|_{y^k} \leq \sigma_2$.

Similarly, by Lemma 4.3 and (4.9),

$$1 - \eta r \leq \langle F_k'(y_*^{k+1}) - F_k'(y^k), y_*^{k+1} - y^k \rangle \leq \frac{\|y_*^{k+1} - y^k\|_{y^k}^2}{1 - \|y_*^{k+1} - y^k\|_{y^k}}.$$

Thus $(1 - \eta)r^2 + (1 + \eta)r - 1 \geq 0$, implying $\|y_*^{k+1} - y^k\|_{y^k} \geq \sigma_1$. This concludes the proof of the lemma. \blacksquare

LEMMA 4.9. *Let y^k be the η -center of \mathcal{F}_D^k at which the oracle is called. Let y_*^{k+1} be the exact center of \mathcal{F}_D^{k+1} . Then,*

$$F_{k+1}^* \geq F_k(y^k) + \theta - \log \|a^k\|_{[F_k''(y^k)]^{-1}},$$

with $F_{k+1}^* = F_{k+1}(y_{k+1}^*)$ and $\theta = (1 - \eta)\sigma_2 - \log(1 + \sigma_2) - \log \sigma_2$.

Proof. Let us define $r = \|y_*^{k+1} - y^k\|_{y^k}$. As shown in the proof of Lemma 4.8, the following inequality holds

$$\langle F_k'(y^k), y_*^{k+1} - y^k \rangle \geq -\|p(y^k)\|_{F_k''(y^k)} \|y_*^{k+1} - y^k\|_{y^k} \geq -\eta r.$$

By Lemma 4.2 and the above inequality, we have

$$\begin{aligned} F_k(y_*^{k+1}) &\geq F_k(y^k) + \langle F_k'(y^k), y_*^{k+1} - y^k \rangle + \omega(\|y_*^{k+1} - y^k\|_{y^k}), \\ &\geq F_k(y^k) - \eta r + r - \log(1 + r). \end{aligned}$$

On the other hand,

$$\begin{aligned} F_{k+1}(y_*^{k+1}) &= F_k(y_*^{k+1}) - \log \langle a^k, y^k - y_*^{k+1} \rangle, \\ &\geq F_k(y_*^{k+1}) - \log \|a^k\|_{[F_k''(y^k)]^{-1}} - \log r. \end{aligned}$$

Thus,

$$F_{k+1}^* \geq F_k(y^k) + (1 - \eta)r - \log(1 + r) - \log r - \log \|a^k\|_{[F_k''(y^k)]^{-1}}.$$

One easily checks that the function $(1 - \eta)r - \log(1 + r) - \log r$ achieves its minimum value at $s = \sigma_2$. Hence, the result. \blacksquare

4.2.2. Bound on the calls to the oracle. Let us denote $\tau_j = \|a^j\|_{[F_j''(y^j)]^{-1}}$ and $\theta = (1 - \eta)\sigma_2 - \log(1 + \sigma_2) - \log \sigma_2$.

LEMMA 4.10. *Assume C contains a ball of radius $\epsilon > 0$. If the algorithm has not stopped at the k -th iteration, then*

$$k\theta - \sum_{j=1}^k \log \tau_j + 2n \log 2 \leq F_{k+1}^* \leq -(2n + k) \log \epsilon.$$

Proof. Since we assumed $\|a^k\| = 1$, the slack $s_{2n+i} = \langle a^i, y^i - y \rangle$, $1 \leq i \leq k$, is just the distance from y to the i -th cutting plane. Since the oracle returns a cutting plane at the k -th iteration (and denies $y^k \in \mathcal{F}_D^k$), then $\mathcal{F}_D^{k+1} \supset C$. Since C contains a ball of radius ϵ , then $(s_*)_{2n+i}^k \geq \epsilon$, $i = 1, \dots, k$, where $(s_*)^k$ are the slacks at the analytic center y_*^k . Moreover, since $\mathcal{F}_D^1 \supset C$, then $\epsilon < \frac{1}{2}$, and

$$\sum_{j=1}^n \log(1 - (y_*)_j)(y_*)_j \geq 2n \log \epsilon.$$

This proves the right-hand side inequality. The inequality on the left follows from Lemma 4.9 and $F_1(e/2) = 2n \log 2$. \blacksquare

We need to bound the sum of the logarithms of the τ 's from above. First we note that the inequality between the geometric and the arithmetic means yields

$$\begin{aligned} \sum_{i=1}^k \log \tau_i - 2n \log 2 &= \frac{1}{2} \left(\sum_{i=1}^k \log \tau_i^2 - 2n \log 4 \right), \\ (4.10) \qquad \qquad \qquad &\leq \frac{2n + k}{2} \log \frac{\frac{n}{2} + \sum_{i=1}^k \tau_i^2}{2n + k}. \end{aligned}$$

We shall now bound the right-hand side of (4.10).

LEMMA 4.11. *For all k , the following inequality holds:*

$$\tau_k^2 \leq \omega_k^2 = \langle B_k^{-1} a^k, a^k \rangle,$$

with $B_k = 8I + \frac{1}{n} \sum_{i=1}^{k-1} a^i (a^i)^T$.

Proof. Since $0 < y^k < 1$ for all k , then $F_1''(y^k) = Y_k^{-2} + (I - Y_k)^{-2} \succeq 8I$, and $s_{2n+i}^k = \langle \tilde{a}^i, y^i - y^k \rangle \leq \|a^i\| \|y^i - y^k\| \leq \sqrt{n}$. Hence,

$$F_k''(y^k) = F_1''(y^k) + \sum_{i=1}^{k-1} s_{2n+i}^{-2} a^i (a^i)^\top \succeq 8I + \frac{1}{n} \sum_{i=1}^{k-1} a^i (a^i)^\top = B_k.$$

From $B_k^{-1} \succeq [F_k''(y^k)]^{-1}$, we conclude

$$(\tau_k)^2 = \langle [F_k''(y^k)]^{-1} a^k, a^k \rangle \leq \langle B_k^{-1} a^k, a^k \rangle = \omega_k^2.$$

■

LEMMA 4.12.

$$\sum_{j=1}^k \omega_j^2 \leq 2n^2 \log \left(1 + \frac{k}{8n^2} \right).$$

Proof. Let $\tilde{a}^k = B_k^{-1/2} a^k$. Then $B_{k+1} = B_k^{1/2} (I + \frac{1}{n} \tilde{a}^k (\tilde{a}^k)^\top) B_k^{1/2}$ and

$$\det B_{k+1} = \det \left(I + \frac{1}{n} \tilde{a}^k (\tilde{a}^k)^\top \right) \det B_k.$$

The eigenvalues of the matrix $(I + \frac{1}{n} \tilde{a}^k (\tilde{a}^k)^\top)$ are 1 with order $(n-1)$ and $(1 + \frac{1}{n} (\tilde{a}^k)^\top \tilde{a}^k) = (1 + \frac{1}{n} \omega_k^2)$ with order 1. Hence,

$$\log \det B_{k+1} = \log \left(1 + \frac{\omega_k^2}{n} \right) + \log \det B_k.$$

Using $B_k \succeq 8I$, we have

$$\omega_k^2 = \langle B_k^{-1} a^k, a^k \rangle \leq \frac{1}{8} (a^k)^\top a^k = \frac{1}{8}.$$

Since $\log(1+u) \geq u/2$ for $0 \leq u \leq 1$, one has $\log(1 + \frac{\omega_k^2}{n}) \geq \frac{\omega_k^2}{2n}$. Therefore,

$$\log \det B_{k+1} \geq \log \det B_1 + \sum_{j=1}^k \frac{\omega_j^2}{2n} = n \log 8 + \sum_{j=1}^k \frac{\omega_j^2}{2n}.$$

In view of the inequality between the geometric and the arithmetic means and of the relation $\text{tr}(a^j (a^j)^\top) = \|a^j\|^2 = 1$, we may write

$$\begin{aligned} \log \det B_{k+1} &\leq n \log \text{tr} \left(\frac{B_{k+1}}{n} \right), \\ &= n \log \left(\text{tr} \frac{8}{n} I + \sum_{j=1}^k \text{tr} \frac{a^j (a^j)^\top}{n^2} \right) = n \log \left(8 + \frac{k}{n^2} \right). \end{aligned}$$

Subtracting $n \log 8$ on both sides, we get

$$\sum_{j=1}^k \omega_j^2 \leq 2n^2 \log \left(1 + \frac{k}{8n^2} \right).$$

THEOREM 4.13. *The algorithm stops with a solution as soon as k satisfies:*

$$\frac{\varepsilon^2}{n} \geq \frac{\frac{1}{2} + 2n \log(1 + \frac{k}{8n^2})}{2n + k} \exp\left(-2\theta \frac{k}{2n + k}\right).$$

Proof. The proof is a direct application of Lemmas 4.10, 4.11 and 4.12. ■

For any value of θ , positive or negative, the exponential term on the right-hand side of the inequality tends to a constant as k increases. Neglecting lower order terms, we conclude that, asymptotically, the complexity of the number of calls to the oracle is of the same order as n^2/ε^2 , independent of θ . Following [34] we introduce the notation $O^*(n^2/\varepsilon^2)$ to describe this situation.

4.2.3. Complexity of the computation of the approximate centers. Our goal now is to show that after adding a new cut, we can easily retrieve a point in the vicinity of the analytic center of the updated set \mathcal{F}_D^{k+1} . To this end, we use a damped Newton process starting at some interior point of \mathcal{F}_D^{k+1} . To exhibit such a point, we define (similarly to [65]) a search direction d by

$$(4.11) \quad d = -\frac{[F_k''(y^k)]^{-1} a^k}{\|a^k\|_{[F_k''(y^k)]^{-1}}} = \arg \min \{ \langle a^k, \delta \rangle \mid \langle F_k''(y^k) \delta, \delta \rangle \leq 1 \},$$

and choose the re-entry point as $y(\alpha) = y^k + \alpha d$, for some $0 < \alpha < 1$. The motivation behind (4.11) is to balance the terms $\log \langle a^k, y^k - y \rangle$ and $F_k(y)$ in

$$F_{k+1}(y) = F_k(y) - \log \langle a^k, y^k - y \rangle.$$

Remembering that y^k is nearly centered, we neglect the first order approximation of F_k and use the level set defined by its second-order approximation, that is the Dikin ellipsoid $\{y \mid \langle F_k''(y^k)(y^k - y), (y^k - y) \rangle \leq 1\}$. Note that

$$\langle a^k, -d \rangle = \|a^k\|_{[F_k''(y^k)]^{-1}} = \tau_k, \quad \text{and} \quad \|d\|_{F_k''(y^k)} = 1.$$

THEOREM 4.14. *For any $k \geq 1$, the number of damped Newton steps to generate a η -approximate analytic center is bounded by*

$$\left\lceil \frac{g(\eta)}{\eta - \log(1 + \eta)} \right\rceil,$$

with

$$g(\eta) = -(1 - \eta)(\sigma_2 + \sigma_1) + \log \frac{(1 + \sigma_2)\sigma_2}{(1 - \sigma_1)\sigma_1}.$$

Proof. Let us evaluate F_{k+1} at $y = y^k + \alpha d$, with $\alpha \leq 1$, where d is defined by (4.11). Using $\langle a^k, -d \rangle = \tau_k$, we have

$$(4.12) \quad F_{k+1}(y^k + \alpha d) = F_k(y^k + \alpha d) - \log \langle a^k, -\alpha d \rangle = -\log \alpha - \log \tau_k.$$

By Lemma 4.3, we have

$$(4.13) \quad F_k(y^k + \alpha d) \leq F_k(y^k) + \langle F'_k(y^k), \alpha d \rangle - \alpha - \log(1 - \alpha).$$

Since

$$\langle F'_k(y^k), \alpha d \rangle \leq \alpha \|F'_k(y^k)\|_{[F''_k(y^k)]^{-1}} \cdot \|d\|_{F''_k(y^k)} \leq \alpha \|p(y_k)\|_{F''_k(y^k)} \leq \alpha \eta,$$

we conclude that

$$(4.14) \quad F_k(y^k + \alpha d) \leq F_k(y^k) - (1 - \eta)\alpha - \log(1 - \alpha).$$

Putting (4.12) and (4.14) together, we obtain

$$F_{k+1}(y^k + \alpha d) \leq F_k(y^k) - (1 - \eta)\alpha - \log(1 - \alpha) - \log \alpha - \log \tau_k.$$

By differentiating the right-hand side with respect to α , we show that the right hand side is minimized when $\alpha = \sigma_1$.

By Lemma 4.9,

$$F_{k+1}^* \geq F_k(y^k) + (1 - \eta)\sigma_2 - \log(1 + \sigma_2) - \log \sigma_2 - \log \tau_k.$$

Therefore,

$$F_{k+1}(y^k + \alpha d) - F_{k+1}^* \leq g(\eta),$$

with

$$g(\eta) = -(1 - \eta)(\sigma_2 + \sigma_1) + \log \frac{(1 + \sigma_2)\sigma_2}{(1 - \sigma_1)\sigma_1}.$$

■

The best value for the bound on the number of Newton steps is

$$\min_{0 < \eta < 1} \left\{ \frac{g(\eta)}{\eta - \log(1 + \eta)} \right\}.$$

For instance, one gets the integral bound 19 with $\eta = .74$.

The search direction does not take into account that $F'_k(y) \neq 0$ at an approximate center. Better results may be achieved in practice if one replaces the direction by

$$d = \arg \min \{ \langle a^k + F'_k(y^k), \delta \rangle \mid \langle F''_k(y^k)\delta, \delta \rangle \leq 1 \}.$$

The explicit formula for this direction is

$$\hat{d} = - \frac{[F''_k(y^k)]^{-1} \hat{a}^k}{\|\hat{a}^k\|_{[F''_k(y^k)]^{-1}}}$$

with

$$\hat{a}^k = a^k + F'_k(y^k).$$

This modification is likely to be beneficial in practice, though it does not lead to any improvement in the complexity estimates.

5. Extensions towards a practical implementation.

5.1. Multiple cuts. In practice, some oracles generate multiple cuts. This is typical of decomposition approaches in which the subproblem breaks down into many independent ones. See [21]. The problem has been dealt with by several authors [89, 37, 38]. Here we give a proof that is adapted from [74].

We suppose that the oracle generates p central cuts at a time. At iteration k , the cuts are denoted

$$-\gamma_j^k = \langle a^{kj}, y - y^k \rangle \leq 0, \quad j = 1, \dots, p.$$

Let us denote $a^{kj} = a^{(k-1)p+j}$ the j -th cut, $j = 1, \dots, p$, generated at the k -th call of the oracle, and $A^k = (a^{k1} \dots a^{kp})$ the matrix whose column vectors are associated with the new cuts.

After adding the p cuts the potential function becomes

$$F_{k+1}(y) = F_k(y) - \sum_{j=1}^p \log \gamma_j^k.$$

The function is not defined at $y = y^k$. As in the case of a single cut, we must construct a search direction d along which to look for a good starting point for the computation of the next approximate analytic center y^{k+1} .

Since we assumed that the solution set has a nonempty interior, we can assert that there exists a y such that $\gamma^k > 0$. However, this does not imply that finding such a y has a closed form solution. Thus, we must propose a method for computing a suitable restoration direction and discuss its complexity.

5.1.1. Optimal restoration direction. Using the same motivation as in the single cut case, we define the restoration direction as the optimal solution to the problem:

$$(5.1) \quad \min \left\{ - \sum_{j=1}^p \log \gamma_j^k \mid \gamma + (A^k)^T d = 0, \|d\|_{F_k''(y^k)} \leq 1 \right\}.$$

We shall now show that the optimal solution of (5.1) can be computed via the simple unconstrained optimization problem

$$(5.2) \quad \min \{ G(\beta) = \frac{p}{2} \langle \beta, V\beta \rangle - \sum_{j=1}^p \log \beta_j \}$$

where

$$V = (A^k)^T (F_k''(y^k))^{-1} (A^k).$$

Note that (5.2) is a problem in R^p .

The norm of a^{kj} with respect to $[F_k''(y^k)]^{-1}$ is

$$\tau_j^k = \langle a^{kj}, [F_k''(y^k)]^{-1} a^{kj} \rangle^{\frac{1}{2}}.$$

We define the vector $\tau^k = \{\tau_j^k\}$.

It can be shown [38] that the existence of an optimal solution to Problem 5.2 follows from the initial assumption that there exists an interior solution to Problem 5.1.

For the sake of simpler notation we shall drop the index k in the next theorem.

THEOREM 5.1. *Let (d^*, γ^*) and β^* be the optimal solutions of (5.1) and (5.2) respectively. Then*

- i) $(\beta^*)^{-1} = pV\beta^*$.
- ii) $d^* = -[F_k''(y^k)]^{-1}A\beta^*$ and $\gamma^* = -A^T d^* = V\beta^*$.
- iii) $\|d^*\|_{F_k''(y^k)} = \langle \beta^*, V\beta^* \rangle^{1/2} = 1$.
- iv) $\sum_{j=1}^p \log \beta_j^* + \sum_{j=1}^p \log \gamma_j^* + p \log p = 0$.

The proof of the theorem is given in [38]: it is a straightforward application of the necessary and sufficient first order optimality conditions. We leave it to the reader.

5.1.2. Computation of the restoration direction. The function $G(\beta)$ is the logarithmic barrier studied in the previous section plus a quadratic term. It can be shown directly [38], or using the theory of self-concordant functions [70], that Lemmas 4.2, 4.3, 4.4 and³ 4.6 hold. Starting from the point

$$\beta^0 = \frac{\tau^{-1}}{\langle \tau^{-1}, V\tau^{-1} \rangle^{1/2}}$$

we can apply a damped Newton scheme that reduces G by at least $\delta - \log(1 + \delta)$, where δ is any number $0 < \delta \leq \|G'(\beta)\|_{[G''(\beta)]^{-1}}$. Let us fix $\delta < 1$. We propose the following computational scheme. Apply damped Newton steps as long as $\|G'(\beta)\|_{[G''(\beta)]^{-1}} > \sqrt{\delta}$. Terminate with one additional pure Newton step. Let β^+ be the point thus obtained.

THEOREM 5.2. *The point β^+ satisfies*

- i) $\|p\beta^+ \cdot V\beta^+ - e\| \leq \delta$.
- ii) *Letting $d^+ = -[F_k''(y^k)]^{-1}A\beta^+$ and $\gamma^+ = -A^T \beta^+ = V\beta^+$, we have $\gamma^+ > 0$ and $\sqrt{1 - \delta} \leq \|d^+\|_{F_k''(y^k)} \leq \sqrt{1 + \delta}$.*

To bound the number of necessary Newton steps to reach a near optimal direction \hat{d} , we just need to bound the difference $G(\beta^0) - G(\beta^*)$. Using the expression of β^0 one easily verifies that

$$G(\beta^0) \leq \sum_{i=1}^p \log \tau_i + \frac{p}{2} + p \log p \leq p\left(\frac{1}{2} + \log p - 3 \log 2\right).$$

thus $\tau_j \leq 1/8 \|a^j\|^2 = 1/8$. The following lower bound on $G(\beta^*)$ is obtained [38] under the assumption that the feasible set contains a ball of radius ϵ :

$$G(\beta^*) \geq \frac{p}{2} + p \log p + p \log \left(\frac{\epsilon(1 - \eta)}{(m + 1)(1 + \eta)} \right),$$

where $m = 2n + kp$ is the total number of inequalities (cuts and the box sides) present in the definition of F_k .

³In the case of general self-concordant functions, a slightly weaker version of Lemma 4.6 holds. In the case of Problem (5.2) the lemma holds exactly as stated in the previous section.

After $O(p \log m/\epsilon)$ Newton steps, we obtain a point β^+ that is near optimal. To construct a feasible restoration direction we can take $\gamma^+ = V\beta^+$, but we must make sure that $\gamma^+ > 0$ and $y + d^+ \in \text{dom}F_k$. The first property follows from statement i) of Theorem 5.2. To obtain a direction satisfying the second property we define

$$(5.3) \quad \hat{d} = \frac{1}{\sqrt{1+\delta}}d^+ \quad \text{and} \quad \hat{\gamma} = \frac{1}{\sqrt{1+\delta}}\gamma^+.$$

By Theorem 5.2, $y + \hat{d} \in \text{dom}F_k$.

The next theorem relates the objective value of Problem (5.1) at $\hat{\gamma}$ with the optimal value.

THEOREM 5.3. *The approximate solution $\hat{\gamma}$ of Problem (5.1) satisfies*

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

Proof. The inequality on the left is trivial. From the definition of $\hat{\gamma}$,

$$\sum_{j=1}^p \log \hat{\gamma}_j = \sum_{j=1}^p \log \gamma_j^+ - p \log \sqrt{1+\delta}.$$

From $p\beta^+ \cdot \gamma^+ \geq (1-\delta)e$ (Theorem 5.2), and the optimality of β^* , and $p\beta^* \cdot \gamma^* = e$:

$$\begin{aligned} \sum_{j=1}^p \log \gamma_j^+ &\geq -\sum_{j=1}^p \log \beta_j^+ - p \log p + p \log(1-\delta), \\ &\geq -\sum_{j=1}^p \log \beta_j^* - p \log p + p \log(1-\delta), \\ &= \sum_{j=1}^p \log \gamma_j^* + p \log(1-\delta). \end{aligned}$$

The inequality on the right follows immediately. ■

5.1.3. Computing the next analytic center. The next theorem extends the result obtained in the single cut case. Define

$$\sigma_1(p) = \frac{-(p+\delta) + \sqrt{(p-\delta)^2 + 4p}}{2(1-\delta)}, \quad \text{and} \quad \sigma_2(p) = \frac{p+\delta + \sqrt{(p-\delta)^2 + 4p}}{2(1-\delta)}.$$

Note that in the single cut case, we retrieve $\sigma_1(1) = \sigma_1$ and $\sigma_2(1) = \sigma_2$.

THEOREM 5.4. *The following inequality holds*

$$\sigma_1(p) \leq \|y_*^{k+1} - y^k\|_{y^k} \leq \sigma_2(p).$$

To compute a new approximate analytic center y^{k+1} , we use a damped Newton method with the initial point

$$y(\alpha) = y^k + \alpha \hat{d}^k, \quad 0 < \alpha \leq 1,$$

where \hat{d}^k is computed as in (5.3).

THEOREM 5.5. *The computation of the analytic center is performed in $O(p \log p)$ damped Newton iterations.*

Proof. The bound on the number of iterations is proportional to the difference $F_{k+1}(y(\alpha)) - F_{k+1}^*$. By convexity,

$$\begin{aligned} F_{k+1}^* &= F_k(y_*^{k+1}) - \sum_{j=1}^p \log \langle a^{kj}, y_*^{k+1} - y^k \rangle, \\ &\geq F_k(y^k) + \langle F'_k(y^k), y_*^{k+1} - y^k \rangle - \sum_{j=1}^p \log \langle a^{kj}, y_*^{k+1} - y^k \rangle. \end{aligned}$$

Since $\|y_*^{k+1} - y^k\|_{y^k} \leq \sigma_2(p)$, the direction $d = \frac{1}{\sigma_2(p)}(y_*^{k+1} - y^k)$ is feasible for Problem (5.1). Thus,

$$p \log \sigma_2(p) - \sum_{j=1}^p \log \langle a^{kj}, y_*^{k+1} - y^k \rangle \geq - \sum_{j=1}^p \log \gamma_j^*,$$

where γ^* is the optimal solution of Problem (5.1). Therefore,

$$(5.4) \quad F_{k+1}^* \geq F_k(y^k) - \eta \sigma_2(p) - \sum_{j=1}^p \log \gamma_j^* - p \log \sigma_2(p).$$

On the other hand, for $\alpha < 1$, by Lemma 4.3

$$\begin{aligned} F_{k+1}(y(\alpha)) &= F_k(y(\alpha)) - \sum_{j=1}^p \log \langle a^{kj}, \alpha \hat{d} \rangle - \sum_{j=1}^p \log(\alpha \hat{\gamma}_j), \\ &\leq F_k(y^k) + \alpha \langle F'_k(y^k), \hat{d} \rangle - \alpha \|\hat{d}\|_{F''_k(y^k)} - \log(1 - \alpha \|\hat{d}\|_{F''_k(y^k)}) - p \log \alpha - \sum_{j=1}^p \log \hat{\gamma}_j, \\ &\leq F_k(y^k) + \alpha(\eta - 1) - p \log \alpha - \log(1 - \alpha) - \sum_{j=1}^p \log \gamma_j^* + p \log \frac{\sqrt{1 + \delta}}{1 - \delta}. \end{aligned}$$

The last inequality follows from $\|\hat{d}\|_{F''_k(y^k)} \leq 1$ and from Theorem 5.3.

$$\begin{aligned} F_{k+1}(y(\alpha)) - F_{k+1}^* &\leq \alpha(\eta - 1) - p \log \alpha - \log(1 - \alpha) + \\ &\quad + p \log \frac{\sqrt{1 + \delta}}{1 - \delta} + \eta \sigma_2(p) + p \log \sigma_2(p). \end{aligned}$$

Since $\sigma_2(p)$ is of order p , the result follows. ■

5.1.4. Complexity of the multiple cut process. Let us revisit the basic convexity inequality that gave rise to inequality (5.4). Using

$$\langle a^{kj}, y_*^{k+1} - y^k \rangle \leq \|a^{kj}\|_{[F''_k(y^k)]^{-1}} \|y_*^{k+1} - y^k\|_{F''_k(y^k)} \leq \tau_j^k \sigma_2(p),$$

we may write the alternative inequality

$$\begin{aligned}
F_{k+1}^* &= F_k(y_*^{k+1}) - \sum_{j=1}^p \log \langle a^{kj}, y_*^{k+1} - y^k \rangle, \\
(5.5) \quad &\geq F_k(y^k) - \eta \sigma_2(p) - \sum_{j=1}^p \log \tau_j^k - p \log \sigma_2(p).
\end{aligned}$$

We thus obtain the counterpart of Lemma 4.10

$$(2n + kp) \ln \epsilon \geq F_{k+1}^* \geq 2n \log 2 - \sum_{i=1}^k \sum_{j=1}^p \log \tau_j^k + k\theta,$$

with $\theta = (\eta \sigma_2(p) + p \log \sigma_2(p))$. The single cut analysis needs further adjustment. Rather straightforwardly, inequality (4.10) becomes

$$(5.6) \quad \sum_{j=1}^k \sum_{i=1}^p \log \tau_i^j - 2n \log 2 \leq \frac{2n + kp}{2} \log \frac{\frac{n}{2} + \sum_{j=1}^k \sum_{i=1}^p (\tau_i^j)^2}{2n + kp}.$$

Modifying the definitions of B_k and ω_i^k into $B_k = 8I + \frac{1}{n} \sum_{j=1}^{k-1} \sum_{i=1}^p a^{ji} (a^{ji})^\top$ and $\omega_i^k = \langle B_k^{-1} a^{ki}, a^{ki} \rangle^{\frac{1}{2}}$, we state the modified Lemma 4.12

LEMMA 5.6.

$$\sum_{j=1}^k \sum_{i=p}^k (\omega_i^j)^2 \leq 2n^2 p \log \left(1 + \frac{kp}{8n^2} \right).$$

Proof. The proof is a slight modification of the proof of Lemma 4.12. Let $\tilde{a}^{ki} = B_k^{-1/2} a^{ki}$. Then,

$$\det B_{k+1} = \det \left(I + \sum_{i=1}^p \tilde{a}^{ki} (\tilde{a}^{ki})^\top \right) \det B_k.$$

Clearly, $I + \sum_{i=1}^p \tilde{a}^{ki} (\tilde{a}^{ki})^\top \succeq I + a^{ki} (a^{ki})^\top$ for all $i \leq k$; moreover, the eigenvalues of the matrix on the left-hand side are all larger or equal to 1. We conclude that

$$\log \det B_{k+1} \geq \log \det \left(8I + \frac{1}{n} \tilde{a}^{ki} (\tilde{a}^{ki})^\top \right) + \log \det B_k, \quad \forall i \leq p.$$

Summing those inequalities over i and using $\log(1 + (\omega_i^j)^2/n) \geq (\omega_i^j)^2/(2n)$ yields

$$p \log \det B_{k+1} \geq \sum_{j=1}^k \sum_{i=1}^p \frac{(\omega_i^j)^2}{2n} + p \log \det B_1.$$

Finally, noting that

$$\text{tr} B_{k+1} = 8n + \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^p \|a^{ji}\|^2 = 8n + \frac{kp}{n},$$

and $\log \det B_1 = n \log 8$, we easily prove the result. ■

We can now state the main convergence result for the multiple cut process.

THEOREM 5.7. *The algorithm stops with a solution as soon as k satisfies:*

$$\frac{\varepsilon^2}{n} \geq \frac{\frac{1}{2} + 2np \log(1 + \frac{kp}{8n^2})}{2n + kp} \exp\left(-2\theta \frac{k}{2n + kp}\right).$$

The above theorem reveals that the multiple cut process worsens the complexity estimate by a factor proportional to p . This negative result is not all too surprising. Suppose for instance that the oracle generates a single cut that is further replicated p times. The impact on the geometry of the set of localization \mathcal{F}_D is the same, whether the cut is repeated or not. On the other hand, repetition leads to an artificial shift of the analytic center that may turn out to be detrimental to convergence.

However, a close look at the arguments used in the proofs suggests that the multiple cut process is likely to enhance convergence in practice. Indeed, the inequality that bounds the variation of F_*^k could be much sharper if we used $-\sum_{j=1}^p \log \gamma_j^*$ instead of $-\sum_{j=1}^p \log \tau_j^k$, i.e., if we substituted (5.4) to (5.5). Clearly, if the p cuts span a very narrow cone, then the term $-\sum_{j=1}^p \log \gamma_j^*$ may be very large, while $-\sum_{j=1}^p \log \tau_j^k$ is almost as bad as if the same cut was repeated p times.

5.2. Deep cuts. So far we studied the case of an oracle returning “central cuts”, that is, cuts passing through the query point. In practice, cuts are often deep, that is, they cut off the query point. A deep cut generated at the approximate center \bar{y} takes the form

$$\langle a^{m+1}, y \rangle \leq c_{m+1} \leq \langle a^{m+1}, \bar{y} \rangle,$$

with strict inequality if the cut is deep. Here, $m = 2n + k - 1$ is the total number of cuts, including the box sides, at the start of iteration k . Recall that under our general assumptions $\|y\| \leq \sqrt{n}$ since $0 \leq y \leq e$ and $\|a^{m+1}\| = 1$; hence $c_{m+1} \leq \sqrt{n}$. We also find it convenient to denote the depth of the cut by

$$d_{m+1} = \langle a^{m+1}, \bar{y} \rangle - c_{m+1} \geq 0.$$

Convergence analysis breaks down in at least two places. First, a deep cut may push the new analytic center far away: the bound $\|y_*^{k+1} - y^k\|_{y^k} \leq \sigma_2$ does not necessarily hold. The second difficulty is of the same nature: even though the set \mathcal{F}_D^{k+1} has a nonempty interior, one cannot directly find a point in this set. One needs to resort to special techniques to compute the new analytic center. If the cut lies far away from the query point, the computation will be all the more difficult. However, those negative arguments are balanced by a positive one: the deeper the cut, the more the localization set \mathcal{F}_D^{k+1} shrinks, and the faster the cutting plane scheme converges.

There are at least two ways of retrieving the new analytic center. The first one, consists in artificially setting the new constraint at the analytic center, and then pushing the constraint to its correct position using a path-following scheme. An alternative approach consists of exploiting duality, as in [36].

5.2.1. Further results from interior point methodology. In the linear case, one easily defines the dual of Problem (4.1), for instance using Wolfe’s duality, we introduce

$$(5.7) \quad \min\{\varphi_P(x) = \langle c, x \rangle - \sum_{i=1}^n \ln x_i \mid Ax = 0, x > 0\}.$$

We denote

$$\mathcal{F}_P = \{x \geq 0 \mid Ax = 0\}$$

the feasible set of (5.7). This set is reminiscent of Karmarkar's canonical problem of linear programming. For this reason we name it "primal".

There is a simple duality relation between functions φ_P and φ_D : it follows from the inequality $\log t \leq t - 1$ that holds for all $t \geq 1$, with equality if and only if $t = 1$. Let $s \in \text{int}\mathcal{F}_D$ and $x \in \text{int}\mathcal{F}_P$. Then, setting $t_i = x_i s_i$, we get

$$\varphi_P(x) + \varphi_D(s) = \langle c, x \rangle - \sum_{i=1}^m \log x_i s_i \geq \langle c, x \rangle - \sum_{i=1}^m x_i s_i + m = m,$$

with equality if and only if $x \bullet s = e$, that is, if and only if x and s are the analytic centers of \mathcal{F}_P and \mathcal{F}_D , respectively.

The function $\varphi_P(x)$ enjoys the same properties as the function $\varphi_D(s)$. Namely, Lemmas 4.2 and 4.3 hold. This suggests that a damped Newton method is well fit to compute an approximate center.

Taking the second order approximation of φ_P , one may write the projected Newton direction as the solution of the problem

$$\min \left\{ \frac{1}{2} \langle X^{-2} p(x), p(x) \rangle + \langle c - x^{-1}, p(x) \rangle \mid Ap(x) = 0 \right\}.$$

The first order optimality conditions are

$$\begin{aligned} x^{-1} \bullet p(x) - X A^T y &= e - x \bullet c, \\ AX(x^{-1} \bullet p(x)) &= 0. \end{aligned}$$

Thus $x^{-1} \bullet p(x) = P_{AX}(e - c \bullet x)$, where P_{AX} denotes the projection operator onto the null space of AX . An alternative expression for $p(x)$ is

$$(5.8) \quad p(x) = x \bullet (e - s \bullet x),$$

with

$$(5.9) \quad s(x) = c - A^T y(x) \quad \text{and} \quad y(x) = (AX^2 A^T)^{-1} AX^2 c.$$

One easily checks that the variable s which appears in the above equation can alternatively be defined as

$$s(x) = \text{argmin} \{ \|e - s \bullet x\| \mid s = c - A^T y \}.$$

The proximity measure takes the form

$$\|p(x)\|_{\varphi_P''(x)} = \|e - s(x) \bullet x\|.$$

Thus

$$\langle \varphi_P'(x), p(x) \rangle = \langle x \bullet \varphi_P'(x), x^{-1} \bullet p(x) \rangle = \langle c \bullet x - e, P_{AX}(e - c \bullet x) \rangle = - \|p(x)\|_{\varphi_P''(x)}^2.$$

Using the above expression, one easily shows that Lemma 4.4 holds; thus Corollary 4.5 holds. It gives a bound for the number of damped Newton steps to generate an η -center. It can also be shown [76, 90] that the quadratic convergence Lemma 4.6 and Corollary 4.7 also hold.

5.2.2. Restoration step. To compute the new analytic center after introducing a deep cut, we proceed with the primal problem. Let $\tilde{x} = \begin{pmatrix} x \\ x_{m+1} \end{pmatrix}$, $\tilde{c} = \begin{pmatrix} c \\ c_{m+1} \end{pmatrix}$ and

$$\tilde{\varphi}_P(\tilde{x}) = \langle \tilde{c}, \tilde{x} \rangle - \sum_{j=1}^{m+1} \log \tilde{x}_j = \varphi_P(x) + c_{m+1}x_{m+1} - \log x_{m+1}.$$

To initiate a damped Newton process, one must define an initial feasible point in the \tilde{x} variables. To this end, we consider the problem

$$\max\{x_{m+1} \mid A\Delta x + a^{m+1}x_{m+1} = 0, \|X^{-1}\Delta x\| \leq 1\}.$$

Its solution is

$$\Delta x = -\frac{1}{\tau_{m+1}} X^2 A^\top (AX^2 A^\top)^{-1} a^{m+1},$$

with

$$\tau_{m+1}^2 = \langle a^{m+1}, (AX^2 A^\top)^{-1} a^{m+1} \rangle.$$

Let $\tilde{x}(\alpha) = \begin{pmatrix} x + \alpha \Delta x \\ \alpha/\tau \end{pmatrix}$. By construction $\tilde{x}(\alpha) \in \text{int}\tilde{\mathcal{F}}_P$ for $0 < \alpha < 1$, where $\tilde{\mathcal{F}}_P$ is the obvious extension of \mathcal{F}_P to account for the new variable x_{m+1} .

Starting at $x(\alpha)$ a primal Newton method with damped steps eventually converges to a point x satisfying $\|e - s(x) \bullet x\| < \eta$, where $s(x) = c - A^\top y(x)$ is defined by (5.9). The point $y(x)$ which appears as a by-product of the computation of the Newton direction is the query point \bar{y} for the next call to the oracle.

5.2.3. Complexity analysis. Contrary to the central cut case, one cannot a priori bound the number of damped Newton steps which are necessary to compute a new approximate analytic center after introducing a new cut. Let ν be that number, when the process is started at $\tilde{x}(\alpha)$, for some $0 < \alpha < 1$. Each step decreases $\tilde{\varphi}_P$ by at least $\rho = \eta - \log(1 + \eta)$. Thus,

$$(5.10) \quad \tilde{\varphi}_P^* + \nu\rho \leq \tilde{\varphi}_P(\tilde{x}(\alpha)).$$

To get an inequality on the opposite side, we proceed as follows. First, note that

$$\tilde{\varphi}_P(\tilde{x}(\alpha)) = \varphi_P(x(\alpha)) + \alpha \frac{c_{m+1}}{\tau} - \log \alpha + \log \tau.$$

Next, use Lemma 4.3 to write

$$(5.11) \quad \varphi_P(x(\alpha)) \leq \varphi_P(x) + \alpha \langle \varphi'_P(x), \Delta x \rangle - \alpha \|\Delta x\|_x - \log(1 - \alpha \|\Delta x\|_x).$$

Let us now gather the two terms

$$\langle \varphi'_P(x), \Delta x \rangle + \frac{c_{m+1}}{\tau} = \frac{1}{\tau} (\langle c - x^{-1}, -X^2 A^\top (AX^2 A^\top)^{-1} a^{m+1} \rangle + \langle a^{m+1}, \bar{y} \rangle - d_{m+1}),$$

where y is the point at which the oracle has been called. Since the query point y is given by (5.9), we have $\bar{y} = (AX^2 A^\top)^{-1} AX^2 c$. Replacing y by its value and using $AX^2 x^{-1} = Ax = 0$, we obtain

$$\langle \varphi'_P(x), \Delta x \rangle + \frac{c_{m+1}}{\tau} = -\frac{d_{m+1}}{\tau} \leq 0.$$

Since $\|\Delta x\|_x = 1$, we get

$$(5.12) \quad \varphi_P(x(\alpha)) \leq \varphi_P(x) - \alpha\left(1 + \frac{d_{m+1}}{\tau}\right) - \log(1 - \alpha).$$

Finally, using $\varphi_P^* + \varphi_D^* = m$ and $\tilde{\varphi}_P^* + \tilde{\varphi}_D^* = m + 1$, we get from (5.10), (5.11) and (5.12)

$$(5.13) \quad \tilde{\varphi}_D^* \geq \varphi_D^* + \nu\rho + 1 - \log\tau + \alpha\left(1 + \frac{d_{m+1}}{\tau}\right) + \log\alpha(1 - \alpha).$$

Inequality (5.13) shows that the deeper the cut, i.e., the larger d_{m+1} , then the bigger is the increase of $\tilde{\varphi}_P$. Additionally, the larger the number of intermediary iterations to retrieve the approximate analytic center, the larger again is the increase of $\tilde{\varphi}_P$. Those larger gains on the potential are likely to reduce the number of calls to the oracle, as it will be displayed in the next theorem. Unfortunately, this analysis cannot be used to reduce the overall complexity.

THEOREM 5.8. *The algorithm stops with a solution as soon as k satisfies:*

$$\frac{\varepsilon^2}{n} \geq \frac{\frac{1}{2} + 2n \log\left(1 + \frac{k+1}{8n^2}\right)}{2n+k} \exp\left(-2\frac{\rho \sum_{j=1}^k \nu_j + k}{2n+k}\right) \exp\left(-2\theta \frac{k}{2n+k}\right),$$

where $\rho = \eta - \log(1 + \eta)$, $\theta = \alpha + \log\alpha(1 - \alpha)$ and $\alpha = \frac{\sqrt{5}+1}{2}$. Here, ν_j is the number of Newton iterations to compute the analytic center after introducing the j -th cut.

Since the first exponential term is less than 1, we can assert that the number of calls to the oracle is asymptotically of order $O^*(n^2/\varepsilon^2)$. It follows that the total number of Newton iterations $\sum_{j=1}^k \nu_j$ is asymptotically bounded by an estimate of order $O^*\left(\frac{n}{\varepsilon^2} \log \frac{n^2}{\varepsilon^2}\right)$.

5.3. Nonlinear cutting plane schemes. The cutting plane scheme is based on the assumption that the only accessible information about the optimization problem is given by linear cuts generated by the oracle. In many instances, there may be additional sources of information such as functional constraints given in explicit form, or a known expression for a nonlinear objective. Of course, this explicit nonlinear information can be translated into linear approximations, i.e., cutting planes, but sometimes with a considerable loss of information. This situation calls for an enhancement of the basic cutting plane scheme.

In some other situations, the oracle may itself generate nonlinear cuts. For instance, the underlying functions defining the feasible set may be strongly convex. Conceivably, the linear cuts may be strengthened with a quadratic term. We shall review these two extensions. In both cases, the analysis must resort to the theory of self-concordant functions.

5.3.1. The homogeneous analytic center cutting plane method. The paper [72] considers three convex problems. In each of them Q is a closed bounded convex set in R^n .

- 1. Feasibility** Find $y \in Y^* \cap Q$, where Y^* is a closed convex set described by an oracle and $Q \cap Y^*$ has a non-empty interior.
- 2. Variational inequality** find $y^* \in Q$ such that $\langle a, y - y^* \rangle \geq 0$, for all $y \in Q$ and $a \in A(y)$, where A is a multivalued monotone mapping.
- 3. Constrained minimization** $\min\{f(y) \mid y \in Q\}$, where $f(y)$ is convex and sub-differentiable on some open convex set containing Q .

It is assumed that the set Q is described by a ν -normal self-concordant barrier. Let us briefly recall that self-concordant functions extend the logarithmic function studied in the previous section. Formally a function f is self-concordant if for all $y \in \text{dom} f$ and $u \in R^n$

$$|f'''(y)[u, u, u]| \leq M_f \|u\|_{f''(y)}^{3/2},$$

where M_f is some positive constant. (For the definition and the properties of this class of functions, see [70].)

In [72], the original problem is embedded into a homogeneous space. To describe the embedding, we shall concentrate on the case of variational inequalities only. The other cases receive a similar treatment. Let

$$Y^* = \{y^* \in Q \mid \langle a, y - y^* \rangle \geq 0, \text{ for all } y \in Q, \text{ and } a \in A(y)\}.$$

A point x in the embedding has the form (y, t) , where $t > 0$ is the scalar variable associated with the extra dimension. The set Q generates a cone K defined by

$$K = \{x = (y, t) \mid t > 0, \frac{y}{t} \in Q\}.$$

Similarly, the solution set Y^* generates a cone X^* .

Let us now describe the oracle and barrier function in the embedding space. Assume $G(y)$ is a ν -self-concordant barrier for Q . $G(y)$ can be transformed into a self-concordant $F(x)$ barrier for K (see [69])

$$F(x) = c_1 H\left(\frac{y}{t}\right) - c_2 \nu \ln t,$$

where the c_i 's are two absolute constants.

Note first that for any $a \in A(\bar{y})$, the inequality $\langle a, \bar{y} - y \rangle \geq 0$ holds for all $y \in Y^*$; it is thus a separation oracle. Let $x = (y, t) \in \text{int} K$ and $a \in A(y)$. Define $y(x) = y/t \in Q$ and

$$g(x) = \frac{\hat{g}(x)}{\|\hat{g}(x)\|}, \quad \text{with } \hat{g}(x) = (a, -\langle a, y(x) \rangle).$$

Clearly g defines a homogeneous separating hyperplane for X^* .

The canonical problem takes the form

$$\text{Find } x \in K \cap X^*, \quad x \neq 0.$$

The homogeneous cutting plane algorithm can be summarized by

Initialization

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

Basic Step

Let $x^k = \arg \min F_k(x)$.

The oracle returns a^k .

$$F_{k+1}(x) = F_k(x) - \log \langle a^k, x^k - x \rangle.$$

Note that the method involves the presence of a proximal term.

The convergence analysis is more direct than in the standard case. It is based on the study of the quantity

$$\mu_k(x) = \frac{1}{S_k} \sum_{i=0}^{k-1} \lambda_{ik} \langle g(x_i), x_i - x \rangle,$$

where $\lambda_{ik} = \langle g(x_i), x_i - x_k \rangle^{-1} > 0$, for $0 \leq i < k-1$, and $S_k = \sum_{i=0}^{k-1} \lambda_{ik}$. In [72] it is proved that μ_k decreases as $1/\sqrt{k}$. Finally, the method uses the candidate solution

$$\bar{y}_k = \frac{1}{P_k} \sum_{i=0}^{k-1} \pi_{ik} y(x_i),$$

where $\pi_{ik} = \lambda_{ik} / \|\hat{g}(x_i)\|$ and $P_k = \sum_{i=0}^{k-1} \pi_{ik}$.

THEOREM 5.9. *The proximal analytic center cutting plane method yields an ϵ -approximate solution \bar{y}*

$$\phi(\bar{y}) = \max_{u \in Q} \{ \langle a, y - u \rangle \mid a \in A(u) \} \leq \epsilon$$

after k iterations, with k satisfying

$$\frac{k}{\sqrt{k+\nu}} \leq \frac{L(1+R^2)}{\epsilon\theta_3} e^{\theta_2\sqrt{\nu}},$$

where R is the radius of a ball containing Q , L is a Lipschitz constant for A over Q , and θ_2 and θ_3 are absolute constants (independent of the problem).

It must be pointed out that the last generated point y^k may **not** converge towards the solution set, as shown in a simple counter-example in [72].

For the minimization problem, a similar approach leads to a convergence theorem, but, here, the last generated point is now the candidate point.

THEOREM 5.10. *For any $k \geq 1$ we have*

$$\min_{0 \leq i \leq k-1} f(y_i) - f^* \leq \frac{L}{\sqrt{k+\nu}} \left[\sqrt{\nu} + \frac{e}{\theta_3} \left(1 + \frac{\nu}{k} \right) \right] [1+R^2]^{1+\frac{\nu}{k}},$$

where L is a Lipschitz constant for f and R the radius of a ball containing Q .

It is shown in [72] that an appropriate scaling of the problem makes the complexity estimate in the above theorem proportional to the scale radius and not to the square of R .

An implementable version of the homogeneous cutting plane algorithm involves the use of approximate minimizers. The convergence analysis has been extended to this case by Nesterov, Péton and Vial [71].

5.3.2. Nonlinear cuts. One of the exciting properties of interior point methods is their ability to deal with quadratic functions as easily as with linear functions. To a limited extent, this statement extends to the case of self-concordant convex functions.

This has led to the development of extensions of ACCPM that use quadratic or nonlinear cuts. These extensions seem very promising in practice, but the theory still needs some improvements. In [61] and [62] it is shown how to introduce shallow quadratic and nonlinear cuts; however the cuts are restricted to be generated from a

finite family. The paper [79] shows convergence and complexity in the case of linear cuts but with a sliding nonlinear objective, which creates a shallow nonlinear cut. The paper [17] addresses the issue of introducing a nonlinear Jacobian cut in the context of variational inequalities; the experimental results are impressive, but no convergence analysis is given, as the algorithm needs to drop cuts, which seems out of the range of currently available convergence proofs. Finally, in the case of strongly monotone variational inequalities, [63] give a complexity analysis.

6. Implementation and applications.

6.1. Some implementation issues. The analytic center cutting plane method described in Section 4, and its extensions of Section 5 to multiple cuts and to deep cuts, have been coded and made accessible to researchers. It is available from Logilab for academic research⁴ under the acronym ACCPM [39]. We discuss some of the implementation issues, with main reference to ACCPM code.

6.1.1. Underlying interior point method. ACCPM is implemented in a purely linear framework. Three main methods can be used to compute the analytic center: primal, dual and primal-dual, with some necessary adaptation to handle infeasible starts. In Section 4, the analysis relies on a dual method (free variables and inequality constraints). In Section 5, it is argued that a primal method has a clear advantage in recovering feasibility after adding a deep cut. On the other hand, common wisdom in linear programming grants the primal-dual method with superior efficiency, especially in the case of infeasible starts.

Computing analytic centers has received far less attention than solving linear programming. The experience is thus too limited to credit one method of a superior advantage over the others. ACCPM is based on the primal projective algorithm described in [15]. In contrast, the papers [16, 66] report implementation with the primal-dual method. The remarkable fact in all cases is that a “good” restoration direction allows very fast computation of a new analytic center after introduction of a new cut. Experience shows that one or two Newton iterations suffice after introducing a single cut.

6.1.2. Linear algebra. Just as in plain interior point methods for linear programming, the main issues are: the computation of the symmetric matrix AD^2A^T , where D is a diagonal scaling matrix; and the computation of the Cholesky factors of AD^2A^T . The general structure of A is $A = (G^T, E^T)^T$, where G is the matrix of subgradients generated by the oracle and E is a matrix with a structure of generalized upper bound GUB constraints. The GUB structure appears when the oracle deals with an additive objective function and returns separate subgradients, one for each piece of the additive function. The GUB matrix may be absent, as in the case of pure feasibility problems, or partially present.

The matrix E introduces a sparsity structure that is easily exploitable. Indeed the matrix AD^2A^T can be partitioned into four blocks: the lower diagonal one is the diagonal matrix ED^2E^T . By block pivoting on this matrix [5], the factorization of AD^2A^T boils down to the factorization of $GD^2G^T + (GD^2E^T)(ED^2E^T)^{-1}(ED^2G^T)$.

In many applications GD^2G^T is dense, either because G itself is dense, or because the rapidly increasing number of generated columns (in particular in the multiple cut case) induces fill-in into the product matrix, even though G may remain sparse. However, it was found necessary, for instance on multicommodity flow problems [32],

⁴Web site <http://ecolu-info.unige.ch/logilab>

to resort to sparse linear algebra, at least for an efficient computation the matrix GD^2G^T . ACCPM also offers an option for sparse Cholesky factorization, but the dynamically changing size of the matrix G makes it necessary to perform a new symbolic factorization after each cut addition.

Finally, let us mention that ACCPM is rather sensitive, at least on some very large problems, to the accuracy in the computation of the Newton direction. Insufficient precision may lead to a stall. The current implementation of ACCPM uses an iterative refinement process that considerably improves performances.

6.1.3. Additive cuts. As pointed out in [48, 21], a proper formulation in a decomposition approach may have a tremendous impact on the overall computational effort. In particular, when the objective is additive and when the subgradients are computed for each term of the sum, it is far better to keep the disaggregated information generated by each individual term rather than aggregating it in a single subgradient (or cut). The latter approach is usually known as a multiple cuts (or multiple columns) scheme, and is often a must in practical implementations.

Multiple cuts pose a challenge when it comes to the point of computing the new analytic center by the interior point method. For example, in instances of the nonlinear multicommodity flow problem with up to 10000 commodities and 2000 arcs that were successfully handled by ACCPM, the matrix G has 2000 rows (hence, A has 12000 rows,) and each call to the oracle increases the number of columns by several thousands. The experience shows that the primal projective algorithm faces this situation successfully: the number of Newton iterations to compute a new analytic center is generally of the order 10 to 20; rarely does it exceed these figures.

6.1.4. Initial box constraints. Theory and practice require that the iterates remain within a box. Some sides of the box may be naturally specified by the oracle, e.g., nonnegativity constraints $y \geq 0$. Most of the time, the sides are artificial bounds that should be inactive at the optimum. ACCPM proposes default values. If the iterate gets “too” close to a side, ACCPM automatically pushes this side of the box. This mechanism turns out to be highly efficient in practice. (See e.g., [41].)

The user may also enter guess values for the box sides; good guesses may significantly enhance performance.

6.1.5. Weights on some constraints and column deletion. In the context of optimization problem, there is always a cut associated with the objective. When the oracle provides a better bound (in the context of a minimization problem, an improved upper bound,) at the query point, a new objective cut should be added. In practice, it is easier to translate the objective cut. However, this clearly implies a loss of information. To compensate for it, the objective cut is repeated a number of times. This strategy was found much more efficient [21]. The best repetition factor p seems to be the number of generated columns. A p -times repetition of the constraint [35] is equivalent to a weight p on the associated logarithmic barrier term.

In some examples, the rapidly increasing size of the matrix G becomes a real challenge for the method. Clearly, some of the early generated columns become irrelevant and should be discarded. Otherwise, the Newton iterations to compute analytic centers become more and more costly as the number of generated columns increases. Though highly desirable, column deletion is perilous. There is still a lack of theoretical study⁵, at least in the framework described in this paper. Tests based on

⁵With the exception of [3]; however, the method in this paper has not been implemented; it relies on shallow cuts, a mechanism which could considerably slow down the method.

inscribed and enclosing ellipsoids have been developed to decide on elimination [19]. Unfortunately, those tests are computationally costly. Besides, updating the analytic center after a cut deletion seems to be much more difficult than in the case of cut addition.

The best strategy in problems with large number of multiple cuts is to perform a drastic deletion only once, say after one third of the estimated total number of iterations. See [19].

6.1.6. Constructing oracles. A cutting plane method can be viewed as a dialog between the main program, say ACCPM, and the oracle. The latter is problem dependent; it is left to the responsibility of the user.

Recent development, tend to automatize this process in the case of large scale structured problems, e.g., block angular structures. SET is the acronym of Structure Exploiting Tool [26], a device that is appended to standard algebraic modeling languages and that allows the user to pass the information relative to the structure. The use of a decomposition scheme can be then fully automatized, leaving the user free to use either the standard Kelley-Goldstein-Cheney-Dantzig-Wolfe-Benders scheme or ACCPM.

6.1.7. General assessment on accpm behavior. The general behavior of ACCPM can be paraphrased in few sentences. The strong points of this algorithm are: robustness and reliability (ACCPM works with similar speed and predictability on very different problems), simplicity (no tuning), stability (insensitive to degeneracy). The stability issue is very important, in particular when the algorithm is used to solve repeatedly a subproblem within a larger optimization problem, e.g., a Lagrangian relaxation within a branch and bound approach. ACCPM shares with some other methods, but not all, the nice feature of delivering dual variables.

On the negative side, two factors may severely reduce performance:

1. The Newton iterations are costly, especially if the cuts lie in a large dimensional space.
2. The algorithm may be slow at identifying the optimum, even when all necessary pieces of an original piecewise linear problem have been generated.

The last item puts ACCPM in a weak position with respect to Kelley’s approach for problems such as the linear multicommodity flow⁶, where the cuts necessary to describe the optimum are few per commodity and quickly generated. Kelley’s strategy stops at once when all necessary cuts are present⁷. In contrast, ACCPM keeps taking cautious steps to guard against a possible bad behavior of the oracle.

6.2. Some applications. To conclude, we briefly review a few interesting applications. Nonlinear multicommodity flow problems very naturally lend to a highly disaggregated formulation. Since the cuts (columns) are made up of indicator vectors of paths on a network, the problem is sparse. By exploiting sparsity, ACCPM could solve [32, 40] extremely large instances, with up to 5000 arcs and/or 10000 commodities. The survey paper [73] investigates several methods such as the *flow deviation* [28], a *primal-dual proximal* method [64] and the *projection method* [8] on large nonlinear multicommodity flow problems. The most striking result is that ACCPM is not always the fastest, but it is constantly “good” and is by far the most stable one.

⁶At least on problems of medium to large size, but perhaps not on huge instances.

⁷Kelley’s algorithm, just like the simplex, has the finite termination criterion. However, just as the simplex—and contrary to ACCPM and IPM’s, it may converge poorly on some problems.

The capacity planning for survivable telecommunications networks is formulated in [60, 77] as a very large structured LP that is solved by a two-level decomposition scheme. Those papers provide instances of problems on which the optimal (Kelley's) strategy fails. Stochastic programming and multi-sectorial planning are traditional applications of a decomposition approach. ACCPM is used in [5] and [6] to solve these problems. The paper [27] considers a stochastic programming approach to portfolio management, with the aim of handling a large number of scenarios. By distributing subproblems computations on a system with 10 parallel PC's, the analytic center method could solve a version with one million scenarios.

Computation of economic equilibria is a very promising area of application for ACCPM. A recent thesis [11] and the paper [16] give ample evidence of the solving-power of the method on these reputedly difficult problems.

Finally, we would like to mention applications to integer programming. In the first application [20], ACCPM is used to solve a Lagrangian relaxation of the capacitate multi-item lot sizing problem with set-up times. A full integration of ACCPM in a column generation, or Lagrangian relaxation, framework, for structured integer programming problems (clustering, Weber problem with multiple sites [22]), shows that the reliability and robustness of ACCPM in applications where a nondifferentiable problem must be solved repeatedly (i.e., at every node of the branch and bound tree) makes it a very powerful alternative to both Kelley's cutting plane method and to subgradient optimization.

REFERENCES

- [1] A. Altman and K.C. Kiwiel, "A note on some analytic center cutting plane methods for convex feasibility and minimization problems," *Computational Optimization and Applications* 5 (1996) 175–180.
- [2] K.M. Anstreicher, "Towards a practical volumetric cutting plane method for convex programming," Working Paper, Department of Management Sciences, The University of Iowa, Iowa city, IA, January 1997.
- [3] D. S. Atkinson and P. M. Vaidya (1995), "A cutting plane algorithm that uses analytic centers", *Mathematical Programming*, series B, **69**, 1–43.
- [4] A. Auslender (1976), *Optimisation. Méthodes Numériques*, Masson, Paris.
- [5] 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.
- [6] O. Bahn, A. Haurie, S. Kypreos, and J.P. Vial, Advanced mathematical programming modeling to assess the benefits from international CO₂ abatement cooperation, Technical Report 97.10, HEC/Management Studies, University of Geneva, Switzerland, 1997.
- [7] J. F. Benders, "Partitioning procedures for solving mixed-variables programming problems", *Numerische Mathematik* 4 (1962), 238-252.
- [8] D.P. Bertsekas and R.G. Gallager, *Data Networks*, Prentice-Hall, 1987.
- [9] S. Boyd and C. H. Barratt, *Linear Controller Design: Limits of Performance*, Prentice Hall, 1991.
- [10] S. Boyd, L. El Ghaoui, E. Feron and V. Balakrishnan, "Linear Matrix Inequalities in System and Control Theory", SIAM Studies in Applied Mathematics, 1994.
- [11] B. Büeler, Computing Economic Equilibria and its Application to International Trade of Permits: an Agent-Based Approach, PhD Thesis (forthcoming, December 1997), Swiss Federal Institute of Technology, Zurich, Switzerland, 1997.
- [12] E.W. Cheney and A.A. Goldstein, "Newton's method for convex programming and Tchebycheff approximation", *Numer. Math.* **1** (1959) 253–268.
- [13] G.B. Dantzig and P. Wolfe, The decomposition algorithm for linear programming, *Econometrica* 29, 4 (1961) 767-778.
- [14] G. B. Dantzig and P. Wolfe, "Decomposition principle for linear programs", *Operations Research*, **8**, (1960) 101-111.

- [15] G. de Ghellinck and J.-Ph. Vial, "A polynomial Newton method for linear programming", *Algorithmica* 1 (1986), 425-453.
- [16] M. Denault and J.-L. Goffin, "On a Primal-Dual Analytic Center Cutting Plane Method for Variational Inequalities", *Computational Optimization and Applications*, vol. 12, pp 127-156, 1999.
- [17] M. Denault and J.-L. Goffin, "Variational Inequalities with Quadratic Cuts", Gerad Technical report G-98-69, December 1998, 41 pp.
- [18] D. den Hertog, J. Kaliski, C. Roos and T. Terlaky, "A logarithmic barrier cutting plane method for convex programming", *Annals of Operations Research* 58 (1995) 69-98.
- [19] O. du Merle, Interior points and cutting planes: development and implementation of methods for convex optimization and large scale structured linear programming, Ph.D Thesis, Department of Management Studies, University of Geneva, Geneva, Switzerland, 1995 (in French).
- [20] O. du Merle, J.-L. Goffin, C. Trouiller and J.-P. Vial, "A Lagrangian Relaxation of the Capacitated Multi-Item Lot Sizing Problem Solved with an Interior Point Cutting Plane Algorithm", Logilab Technical Report 97.5, Department of Management Studies, University of Geneva, Switzerland, April 1997.
- [21] O. du Merle, J.-L. Goffin and J.-P. Vial, "On Improvements to the Analytic Center Cutting Plane Method", *Computational Optimization and Applications*, 11 (1998), 37-52.
- [22] O. du Merle, P. Hansen, B. Jaumard, N. Mladenovic, "An Interior Point Algorithm for Minimum Sum of Squares Clustering", GERAD Tech. Report G-97-53, August 1997, 28 pages.
- [23] J. Elzinga and T.G. Moore, "A central cutting plane algorithm for the convex programming problem", *Mathematical Programming* 8 (1975) 134-145.
- [24] M. L. Fisher, "The Lagrangian relaxation method for solving integer programming problems", *Management Science* 27 (1981) 1-18.
- [25] L.R. Ford Jr and D.R. Fulkerson "A suggested computation for multicommodity flows", *Management Science*, 5 pp 97-101, 1958.
- [26] E. Fragnière, J. Gondzio, R. Sarkissian and J.-Ph. Vial, Structure exploiting tool in algebraic modeling languages, Logilab Technical Report 97.2, Department of Management Studies, University of Geneva, Switzerland, June 1997.
- [27] E. Fragnière, J. Gondzio and J.-P. Vial, "A Planning Model with one Million Scenarios Solved on an Affordable Parallel Machine", Logilab Technical Report 98.11, Department of Management Studies, University of Geneva, Switzerland, June, 1998.
- [28] L. Fratta, M. Gerla and L. Kleinrock, "The flow deviation method : an approach to store-and-forward communication network design", *Networks*, 3 (1973), pp. 97-133.
- [29] A. M. Geoffrion, "Lagrangian relaxation for integer programming", *Mathematical Programming Study*, 2 (1974) 82-114.
- [30] P. C. Gilmore and R. E. Gomory, "A linear programming approach to the cutting stock problem", *Operations Research*, 9, pp 849-859, 1960.
- [31] P. C. Gilmore and R. E. Gomory, "A linear programming approach to the cutting stock problem- part II", *Operations Research*, 11, pp 863-888, 1963.
- [32] 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.
- [33] J.-L. Goffin, A. Haurie and J.-P. Vial (1992), "Decomposition and nondifferentiable optimization with the projective algorithm", *Management Science* 38, 284-302.
- [34] 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.
- [35] J.-L. Goffin and J.-P. Vial (1993), "On the Computation of Weighted Analytic Centers and Dual Ellipsoids with the Projective Algorithm", *Mathematical Programming* 60, 81-92.
- [36] J.-L. Goffin and J.-P. Vial (1996), "Shallow, deep and very deep cuts in the analytic center cutting plane method", *Mathematical Programming*, 84 1999 pp 89-103
- [37] J.-L. Goffin and J.-P. Vial (1997), "A Two-Cut Approach in the Analytic Center Cutting Plane Method", Logilab Technical Report 97.20, Department of Management Studies, University of Geneva, Switzerland, August 1997. To appear in *Mathematical Methods in Operations Research*, Volume 49, Issue 1, 1999.
- [38] J.-L. Goffin and J.-Ph. Vial (1998), "Multiple Cuts in the Analytic Center Cutting Plane Method", Logilab, HEC Working Paper 98.10, Department of Management Studies, University of Geneva.
- [39] 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.

- [40] J. Gondzio, R. Sarkissian and J.-P. Vial, Using an Interior Point Method for the Master Problem in a Decomposition Approach, Technical Report *European Journal of Operational Research*, volume 101, 1997,577–587.
- [41] J. Gondzio and J.-P. Vial, Warm Start and epsilon-subgradients in the Cutting Plane Scheme for Block-angular Linear Programs, Logilab Technical Report 97.1, Department of Management Studies, University of Geneva, Switzerland, June 1997. To appear in *Computational Optimization and Applications*.
- [42] S. C. Graves, “Using Lagrangean techniques to solve hierarchical production planning problems”, *Management Science* **28** (1982) 260-274.
- [43] M. Held and R. M. Karp, “The traveling-salesman problem and minimum spanning trees”, *Operations Research* **18** (1970) 1138-1162.
- [44] M. Held and R. M. Karp, “The traveling-salesman problem and minimum spanning trees: Part II”, *Mathematical Programming* **1** (1971) 6-25.
- [45] M. Held, P. Wolfe and H. P. Crowder, “Validation of subgradient optimization,”, *Mathematical Programming* **6** (1974) 62-88.
- [46] J.B. Hiriart-Urruty and c. Lemaréchal, “Convex Analysis and Minimization Algorithms I and II”, Springer-Verlag, Berlin Heidelberg 1993.
- [47] P. Huard, “Resolution of Mathematical Programming with Nonlinear Constraints by the Method of Centers”, in J. Abadie, editor, *Nonlinear Programming*, pages 207–219. North Holland, Amsterdam, The Netherlands, 1967 .
- [48] K.L. Jones, I.J. Lustig, J.M. Farvolden and W.B. Powell, ”Multicommodity network flows: the impact of formulation on decomposition,” *Mathematical Programming*, volume 62, 1993, 95–117.
- [49] J. A. Kaliski and Y. Ye, “A decomposition variant of the potential reduction algorithm for linear programming”, *Management Science* 39 (1993) 757–776.
- [50] N.K. Karmarkar (1984), “A New Polynomial-Time Algorithm for Linear Programming”, *Combinatorica* **4**, 373–395.
- [51] J. E. Kelley, “The cutting plane method for solving convex programs”, *Journal of the SIAM* **8** (1960), 703-712.
- [52] L. G. Khachiyan and M. J. Todd, “On the complexity of approximating the maximal inscribed ellipsoid for a polytope,” *Mathematical Programming* 61 (1993) 137–160.
- [53] K. Kim and J. L. Nazareth. ”The decomposition principle and algorithms for linear programming”, *Linear Algebra and Its Applications*, 152:119–133, 1991.
- [54] K.C. Kiwiel , *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics, vol 1133, Springer Verlag, Berlin, 1985.
- [55] K.C. Kiwiel, “Efficiency of the analytic center cutting plane method for convex minimization,” *SIAM Journal on Optimization* Vol. 7, No. 2 (1997) 336–346.
- [56] L. Lasdon, *Optimization Theory for Large Scale Systems* , (MacMillan, 1970).
- [57] C. Lemaréchal, A. Nemirovskii, Y. Nesterov, “New Variants of Bundle Methods”, *Rapports de Recherche INRIA* No 1508 (1991), INRIA-Rocquencourt, also *Mathematical Programming* 69 (1995), 111–147.
- [58] C. Lemaréchal, “Nondifferentiable Optimization” in G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, editors, *Optimization*, volume 1 of *Handbooks in Operations Research and Management Science*, pages 529–572. North Holland, Amsterdam, The Netherlands, 1989.
- [59] A. Levin, “An algorithm of minimization of convex functions,” *Soviet Mathematics Doklady* 160 (1965) 1244–1247.
- [60] A. Lissier, R. Sarkissian and J.-P. Vial, Optimal joint synthesis of base and reserve telecommunication networks, CNET Technical Note NT/PAA/ATR/ORI/4491, France Telecom, Issy-les-Moulineaux,France, November 1995.
- [61] Z.Q. Luo and J. Sun, “An analytic center based column generation algorithm for convex quadratic feasibility problems,” Preprint, Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada, November 1995.
- [62] Z.Q. Luo and J. Sun, “Cutting surfaces and analytic center: A polynomial algorithm for a convex feasibility problem defined by self-concordant inequalities,” Preprint, Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada, June 1996.
- [63] H.-J. Lüthi and B. Büeler, ”The Analytic Center Quadratic Cut Method (ACQCM) for Strongly Monotone Variational Inequality Problems”, manuscript, ETH Zürich, 1998.
- [64] P. Mahey, A. Ouorou, L. LeBlanc, J. Chifflet, ”A new proximal decomposition algorithm for routing in telecommunication networks”, to appear *Networks*.
- [65] J. E. Mitchell and M.J. Todd (1992), “Solving combinatorial optimization problems using

- Karmarkar's algorithm," *Mathematical Programming* **56**, 245–284.
- [66] J. E. Mitchell (1997), "Computational experience with an interior point cutting plane method", Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY, USA.
- [67] A.S. Nemirovskii and D.B. Yudin, "Problem Complexity and Method Efficiency in Optimization", John Wiley, Chichester (1983).
- [68] Y. Nesterov (1995), "Cutting plane algorithms from analytic centers: efficiency estimates", *Mathematical Programming*, series B, **69**, 149–176.
- [69] Yu. Nesterov. *Introductory lectures on Convex Optimization*. Unpublished manuscript, Louvain, Belgium, 1996.
- [70] Yu. Nesterov and A. Nemirovsky. *Interior Point Polynomial Algorithms in Convex Programming: Theory and Applications*, SIAM, Philadelphia, 1994.
- [71] Yu. Nesterov, O. Péton and J.-P. Vial, "Homogeneous Analytic Center Cutting Plane Methods with Approximate Centers", HEC Technical Report 98.3, Department of Management Studies, University of Geneva, Switzerland, revised version, June 1998. To appear in *Computational Optimization and Applications*.
- [72] Yu. Nesterov and J.-Ph. Vial. "Homogeneous Analytic Center Cutting Plane Methods for Convex Problems and Variational Inequalities", Logilab Technical Report 1997.4, 1997. To appear in *SIAM Journal on Optimization*.
- [73] A. Ouorou, P. Mahey and J.-Ph. Vial, "A survey of algorithms for convex multicommodity flow problems", HEC Working Paper 97.13, Department of Management Studies, University of Geneva, Switzerland, June 1997.
- [74] O. Péton and J.-Ph. Vial. "Multiple Cuts with the Homogeneous Analytic Center Cutting Plane Method". Forthcoming.
- [75] B. T. Polyak, "Minimization of unsmooth functionals (in Russian)", *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* **9** (1969) 509-521 [English translation: *U.S.S.R. Computational Mathematics and Mathematical Physics* **9** (3) (1969) 14-29.]
- [76] C. Roos, T. Terlaky and J.-P. Vial (1997), *Theory and Algorithms for Linear Optimization: An Interior point Approach*. John Wiley & Sons, Chichester.
- [77] R. Sarkissian, *Telecommunications Networks: Routing and Survivability Optimization Using a Central Cutting Plane Method*, PhD thesis, Ecole Polytechnique Fédérale, Lausanne, Switzerland, 1997.
- [78] J. F. Shapiro, "A survey of Lagrangean techniques for discrete optimization", [in *Discrete Optimization II* P. L. Hammer, E. L. Johnson and B. H. Korte, eds.], *Annals of Discrete Mathematics* **5** (1979b) 113-138.
- [79] F. Sharifi-Mokhtarian and J.L. Goffin, "A Nonlinear Analytic Center Cutting Plane Method for a Class of Convex Programming Problems," *SIAM Journal on Optimization*, vol. 8 No. 4, November 1998, pp 1108–1131.
- [80] N. Z. Shor, *Minimization Methods for Non-differentiable Functions*, Springer, Berlin, 1985.
- [81] N. Z. Shor, *Nondifferentiable Optimization and Polynomial Problems*, Kluwer Academic Publishers, Boston Dordrecht Londonm, 1998.
- [82] G. Sonnevend, "New algorithms in convex programming based on a notion of "center" (for systems of analytic inequalities) and on rational extrapolation," in: K. H. Hoffmann, J. B. Hiriart-Urruty, C. Lemaréchal, and J. Zowe, eds., *Trends in Mathematical Optimization: Proceedings of the 4th French-German Conference on Optimization* (Irsee, West-Germany, April 1986); also: *International Series of Numerical Mathematics* 84 (Birkhäuser Verlag, Basel, Switzerland, 1988) 311–327.
- [83] S. Tarasov, L.G. Khachiyan, I. Erlich, "The method of inscribed ellipsoids," *Soviet Mathematics Doklady* **37** (1988).
- [84] P.M. Vaidya, "A new algorithm for minimizing convex functions over convex sets," *Mathematical Programming* **73** (1996) 291–341.
- [85] R. Van Slyke and R. J.-B. Wets, "L-shaped linear programs with applications to optimal control and stochastic programming", *SIAM J. Appl. Math.* **17** (1969) 638–663.
- [86] E. Weiszfeld, "Sur le point par lequel le somme des distances de n points donnés est minimum", *Tohoku Mathematical Journal*, Vol. 43(1937), pp. 355–386.
- [87] P. Wolfe (1975), "A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions", *Mathematical Programming Study* **3**, pp 145–173.
- [88] Y. Ye (1992), "A potential reduction algorithm allowing column generation", *SIAM Journal on Optimization* **2**, 7–20.
- [89] Y. Ye (1997), "Complexity Analysis of the Analytic Center Cutting Plane Method That Uses Multiple Cuts", *Mathematical Programming* **78**, 85–104.
- [90] Y. Ye (1997), *Interior Point Algorithms: Theory and Analysis*. John Wiley & Sons, New York.

Appendix.

LEMMA 6.1. *Let $u \in R^n$ be a strictly positive vector. Then,*

$$(6.1) \quad \frac{\|e - u\|}{1 + \|e - u\|} \leq \|e - u^{-1}\|.$$

Moreover, if $\|e - u\| < 1$, then.

$$(6.2) \quad \|e - u^{-1}\| \leq \frac{\|e - u\|}{1 - \|e - u\|}.$$

Proof. lemma trivially holds if $u = e$. Assume $u \neq e$. Since $u > 0$, we may write

$$\|e - u\| = \|u \bullet (u^{-1} - e)\| \leq \max\{u_i\} \|e - u^{-1}\|.$$

Since $\|e - u\| \geq \max |u_i - 1| \geq \max\{u_i\} - 1$, one has

$$\|e - u\| + 1 \geq \max\{u_i\} \geq \frac{\|e - u\|}{\|e - u^{-1}\|}.$$

This proves the first part of the lemma.

To prove the second part of the lemma we substitute the vectors u and u^{-1} for one another in the above proof. Hence,

$$\frac{\|e - u^{-1}\|}{1 + \|e - u^{-1}\|} \leq \|e - u\|$$

also holds. Assuming $\|e - u\| < 1$ and solving for $\|e - u^{-1}\|$, we get

$$\|e - u^{-1}\| \leq \frac{\|e - u\|}{1 - \|e - u\|}.$$

■

Lemma 6.1 can be interpreted as follows. Let $y, \bar{y} \in \mathcal{F}_D$ and let $s = c - A^T y$ and $\bar{s} = c - A^T \bar{y}$. Define $u = s^{-1} \bullet \bar{s}$. Then

$$\|e - u\| = \|s^{-1} \bullet (\bar{s} - s)\| = \|S^{-1} A^T (\bar{y} - y)\| = \|\bar{y} - y\|_y.$$

Thus, $\|e - u\|$ is just the distance from \bar{y} to y in the Dikin metric at y . It appears that Lemma 6.1 provides a new direct proof of Lemma 4.1.

Proof of Lemma 4.2. (This is essentially the proof in [69]. See also [72].)

Denote $y_\tau = y + \tau(\bar{y} - y)$, $\tau \in [0, 1]$, and $r = \|\bar{y} - y\|_y$. From the mean value theorem

$$\begin{aligned} \langle F'(\bar{y}) - F'(y), \bar{y} - y \rangle &= \int_0^1 \langle F''(y_\tau) (\bar{y} - y), \bar{y} - y \rangle d\tau, \\ &= \int_0^1 \frac{1}{\tau^2} \langle AS_\tau^{-2} A^T (y_\tau - y), y_\tau - y \rangle d\tau, \\ &= \int_0^1 \frac{1}{\tau^2} \|y_\tau - y\|_{y_\tau}^2 d\tau. \end{aligned}$$

In view of (6.1),

$$\begin{aligned} \int_0^1 \frac{1}{\tau^2} \|y_\tau - y\|_{y_\tau}^2 d\tau &\geq \int_0^1 \frac{1}{\tau^2} \frac{\|y_\tau - y\|_y^2}{(1 + \|y_\tau - y\|_y)^2} d\tau, \\ &\geq \int_0^1 \frac{r^2 d\tau}{(1 + \tau r)^2} = r \int_0^r \frac{dt}{(1+t)^2} = \frac{r^2}{1+r}. \end{aligned}$$

Using (4.4) and the mean value theorem, we have

$$\begin{aligned} F(\bar{y}) - F(y) - \langle F'(y), \bar{y} - y \rangle &= \int_0^1 \langle F'(y_\tau) - F'(y), \bar{y} - y \rangle d\tau, \\ &= \int_0^1 \frac{1}{\tau} \langle F'(y_\tau) - F'(y), y_\tau - y \rangle d\tau, \\ &\geq \int_0^1 \frac{r^2 \tau d\tau}{1 + \tau r} = \int_0^r \frac{t dt}{1+t} = \omega(r). \end{aligned}$$

■

Proof of Lemma 4.3.

The proof is analogous to Lemma 4.2, but uses (6.2) of Lemma 4.1 instead of (6.1). ■