USING THE CENTRAL PATH FOR COMPUTING THE ANALYTIC CENTER OF A POLYTOPE*

CLOVIS C. GONZAGA† AND MARLI CARDIA‡

Abstract. In this work we develop an algorithm for computing the analytic center of a polytope after the addition of a deep cut. This problem constitutes the main procedure in central cutting plane methods for convex optimization. When deep cuts are considered, the greatest difficulty lies in retrieving feasibility to initiate a centering algorithm. Our strategy consists in starting at a nearly centered point for the old polytope, and then following the primal-dual central path for an auxiliary linear programming problem. When this path following algorithm stops, the resulting point is a nearly central point for the new polytope. We show that this method performs well for shallow or deep cuts, and prove that its application to the convex feasibility problem matches the best existing complexity bounds.

1. Introduction. Consider the polytope

$$\Omega = \{y \in I\!R^m \mid A^T y \leq c\},$$

where $c \in I\!R^n$ and A is an $m \times n$ matrix with maximum rank m. Given a vector $a \in I\!R^m$ and a number K, the new polytope resulting from the addition of a cut $a^T y \leq K$ is

$$\tilde{\Omega} = \{y \in I\!R^m \mid a^T y \leq K,\ y \in \Omega\}.$$

Assume that a point $y^0 \in \Omega$ is known, and that y^0 is an approximation to the analytic center of Ω (this will be formally stated ahead). With respect to this point, the cut is said to be *deep* if $K << a^T y^0$, or equivalently, if the cut is given by $a^T y \leq a^T y^0 + \gamma$, where $\gamma << 0$. This means that the available point y^0 does not belong to $\tilde{\Omega}$ and is far from it. The cut is said to be shallow if $y^0 \in \tilde{\Omega}$. There is an intermediate situation in which $y^0 \notin \tilde{\Omega}$ but y^0 is near that set in a sense that will become clear below: in this situation the cut is also called shallow.

We shall assume that $\Omega^0 = \{y \mid A^T y < c\}$ and $\tilde{\Omega}^0 = \{y \mid a^T y < K,\ y \in \Omega^0\}$ are non-empty.

We want to design a method for computing an approximate analytic center of $\tilde{\Omega}$, starting at y^0. Note that for deep cuts, y^0 does not belong to $\tilde{\Omega}^0$, and then no interior centralization algorithm can be initiated at y^0.

This computation is central in cutting plane methods based on analytic centers. Cutting plane methods were developed for solving convex optimization problems. They generate a sequence of linear programming problems whose feasible sets provide increasingly refined polyhedral approximations of the original problem. The linear inequalities defining these approximations are generated by an oracle as hyperplanes which separate the present test point from the solution set.

The crucial question in the development of cutting plane algorithms is the choice of the test point in each polyhedral approximation. Several different choices can be found in the literature. One of the first is the classical Kelley cutting plane method [15], in which the test point is an optimal solution of the linear relaxation, i.e., each iteration

*MANUSCRIPT COMPLETED IN AUGUST, 2002.
†Department of Mathematics, Federal University of Santa Catarina, Florianópolis, Brazil
‡Department of Mathematics, Federal University of Paraná, Curitiba, Brazil

solves a linear programming problem. This method may be very inefficient, both theoretically and practically, which motivated much research to improve complexity bounds and practical performance. An important class of methods is based in the computation of a "center" of the polytope, which is then separated from the solution set by a "central cut".

The first such method was the center of gravity method bu Levin [18], followed by the maximal sphere method by Elzinga and Moore [4], the ellipsoid method by Yudin and Nemirovski [34], Shor [28] and Khachiyan [16], the maximum volume ellipsoid of Tarasov, Khachiyan and Erlich [30] and Khachiyan and Todd [17], the volumetric center method by Vaidya [31] and the cutting plane methods based on analytic centers. These last methods have been very successful, due to the existence of fast algorithms for computing analytic centers.

The concept of analytic center was introduced by Sonnevend [29], who also suggested its usage in cutting plane methods. Since then, several authors have been working on such algorithms, including Goffin and Vial [5] and Ye [33]. These methods have been used and tested in a variety of large scale problems with encouraging results: Goffin, Haurie, and Vial [8], report results related to non-smooth convex optimization problems, Bahn *et al.* [3] applied the approach to geometric programming problems, Bahn *et al.* [2] in stocastic linear programming problems, Goffin *et al.* [7] in multicommodity flow problems.

The complexity analysis of central cutting plane methods is relatively easy in problems in which only a given number of cuts will be generated, but it becomes involved in the case of general convex feasibility problems. In this case the complexity bounds will depend on the dimension n of the space and on a precision parameter $\varepsilon > 0$, and not directly on the number of cuts generated.

Atkinson and Vaidya [1] developed such a result for a central cutting plane method which discards constraints which have become irrelevant. Other methods do not discard old constraints and achieve similar polynomiality resuts: Nesterov [24] uses a hybrid method involving an additional quadratic term; Goffin, Luo and Ye [9], based on Nesterov's arguments together with results from Ye [33], and without the quadratic term, found a complexity bound of $O^*\left(n^2/\varepsilon^2\right)$. Here O^* means that lower order terms were ignored (see [9]). Mokhtarian and Goffin [23] and Goffin and Vial [6] relaxed the need for central cuts made in [9], obtaining the same complexity bounds for shallow and deep cuts. Finally, Nesterov and Vial [26] and Nesterov, Péton and Vial [25], study a central cutting plane in a projective space and show a complexity bound dependent only on the parameter of a self-concordant barrier for the feasible set.

Most of the cutting plane methods mentioned above do not accept the cuts as they are generated by the oracle. They move the cuts so as to make them shallow or central.

In this work we develop an algorithm for computing an approximate center of the new polytope after the addition of a cut as it given by the oracle, possibly very deep. Goffin and Vial [6] have considered this case: they obtain an initial feasible point by means of the application of the Mitchell and Todd direction [20] to the dual of the problem (which is in primal format). The iterations are then Newton steps for a problem in a conical formulation like the one in Karmarkar [14].

Our work follows a different approach. Starting from an approximate analytic center of Ω, we define the following auxiliary linear programming problem in dual format:

$$
(D) \qquad \begin{array}{ll} \text{minimize} & a^T y \\ \text{subject to} & A^T y + s = c \\ & s \geq 0, \end{array}
$$

or equivalently,

$$\underset{y \in \Omega}{\text{minimize}}\, a^T y.$$

Define also the sets

$$\Omega_K = \{y \in I\!R^m \mid A^T y \leq c,\ a^T y \leq K\} = \{y \in \Omega \mid a^T y \leq K\}.$$

It is well known (and we shall discuss this ahead) that the central path associated with (D) coincides with the set of analytic centers of Ω_K for $K \in (\min_{y \in \Omega} a^T y,\ +\infty)$. This is studied in [11], where K provides one of the possible parameterizations of the central path for D, corresponding to Renegar's approach. This means that for a given value of K, the analytic center of $\tilde{\Omega} = \Omega_K$ can be found by following the central path for (D).

Our main tasks will be the following:

1. Describe a procedure for initializing the path following method. If K is large (the case of a shallow cut), this step may already solve the problem.
2. Describe a procedure for terminating the path following method so as to obtain an approximate center of $\tilde{\Omega} = \Omega_K$.

The path following method itself can be one of several found in the literature, like Mizuno-Todd-Ye [22] or Gonzaga [12]. The most efficient primal-dual method implemented nowadays seems to be Mehrotra's predictor-corrector algorithm [19].

We shall also discuss the usage of a short step path following algorithm, for complexity computations. We must describe the effect of the procedure on the potential function used in the complexity analysis of Goffin and Vial [6], to show that the same complexity bounds are attained by our procedure when applied to a convex feasibility problem.

The work is organized as follows: in Sec. 2 we discuss properties of the central path, analytic centers and barrier functions. Sec. 3 studies the auxiliary problem. Section 4 describes the algorithm for the introduction of a deep cut, and Section 5 solves the inverse problem, finding the maximum cut depth attainable with a single Newton step. Section 6 shows that the complexity bounds for the convex feasibility obtained by Goffin, Luo and Ye [9] are also achieved by our method.

Our notation is quite standard: $I\!R^n$ is the Euclidean space, $I\!R^n_+$ is the non negative orthant and $I\!R^n_{++}$ is the positive orthant. The letter e denotes the vector of ones, with dimension given by the context. Given a vector x, the upper case X denotes the matrix $\text{diag}(x_1, x_2, \ldots, x_n)$. Given vectors $x, s \in I\!R^n$ x^{-1}, xs, $\frac{x}{s}$ denote the vectors with components x_i^{-1}, $x_i s_i$ and $\frac{x_i}{s_i}$, $i = 1, ..., n$. $\|\cdot\|$ denotes the Euclidean norm, and $x^T s$ is the internal product of the vectors x and s. The null space and the range space of a matrix A are respectively $\mathcal{N}(A)$ and $\mathcal{R}(A)$.

The next section lists known properties of the central points, in the format which will be used in our treatment.

2. Barrier function, central path and analytic center. This is a section in which we list several properties of central points in linear programming. The results are quoted directly from references [11] and [13].

We shall study the primal and dual linear programming problems and their optimality conditions:

$$
\begin{array}{ccc}
(P) & (D) & (PD) \\
\begin{array}{lrcl}
\min & c^T x & & \\
\text{s. t.} & Ax & = & b \\
& x & \geq & 0
\end{array}
&
\begin{array}{lrcl}
\max & b^T y & & \\
\text{s. t.} & A^T y + s & = & c \\
& s & \geq & 0
\end{array}
&
\begin{array}{rcl}
xs & = & 0 \\
Ax & = & b \\
A^T y + s & = & c \\
x, s & \geq & 0
\end{array}
\end{array}
$$

where $c \in I\!\!R^n$, $b \in I\!\!R^m$ and $A \in I\!\!R^{m \times n}$ $(m < n)$ has full row rank.

The primal feasible set and the set of primal interior points are given by

$$\mathcal{X} = \{x \in I\!\!R^n \mid Ax = b, x \geq 0\} \quad , \quad \mathcal{X}^0 = \{x \in \mathcal{X} \mid x > 0\}.$$

The dual feasible set and the set of dual interior points are given by

$$\mathcal{D} = \{(y, s) \in I\!\!R^{n+m} \mid A^T y + s = c, \ s \geq 0\} \quad , \quad \mathcal{D}^0 = \{(y, s) \in \mathcal{D} \mid s > 0\}.$$

We shall assume that $\mathcal{X}^0 \neq \emptyset$ and $\mathcal{D}^0 \neq \emptyset$. It is known that the primal and dual feasible sets cannot be simultaneously bounded (and this will be clear in a moment, when we define central points). We shall assume that the dual feasible set is bounded.

The dual problem can be stated in terms of either y or s alone, because A has full row rank. The feasible sets and sets of interior solutions will then be denoted by

$$\Omega = \{y \in I\!\!R^m \mid (y, s) \in \mathcal{D}\} \quad , \quad \Omega^0 = \{y \in I\!\!R^m \mid (y, s) \in \mathcal{D}^0\}$$

and

$$\mathcal{S} = \{s \in I\!\!R^n \mid (y, s) \in \mathcal{D}\} \quad , \quad \mathcal{S}^0 = \{s \in I\!\!R^n \mid (y, s) \in \mathcal{D}^0\}.$$

From now on, we shall speak with no warning of dual solutions $y \in \Omega$ or $s \in \mathcal{S}$, whichever is more convenient. Whenever we define one of the sets, say, $\Omega, \Omega_j, \ldots$, the associated set $\mathcal{S}, \mathcal{S}_j, \ldots$ is also considered as defined with coherent notation.

In terms of y, (D) is

$$\underset{y \in \Omega}{\text{maximize}} \, b^T y.$$

Let us state (D) in terms of s. To obtain the constraint, let B be a matrix whose rows span $\mathcal{N}(A)$, for instance $B = P_A$, the projection matrix into $\mathcal{N}(A)$. The equality constraint can be written as $Bs = Bc$. To obtain the cost, multiply the equality constraint by $(AA^T)^{-1}A$ (which exists because A has full row rank), and get $y = -(AA^T)^{-1}A(s - c)$. Substituting into the objective of D and discarding the constant term, we get

$$
(D') \qquad
\begin{array}{lrl}
\text{maximize} & -\rho^T s & \\
\text{subject to} & Bs & = Bc \\
& s & \geq 0,
\end{array}
$$

with $\rho = A^T (AA^T)^{-1} b$.

This problem is equivalent to (D) in the sense that its feasible set is $\mathcal{S}$, and its optimal set coincides with the set of optimal slacks for (D). We can also write the problem as minimize$\{\rho^T s \mid Bs = Bc, s \geq 0\}$, and $-\rho$ will be the steepest descent direction in $\mathcal{N}(B)$.

The cost vector satisfies $\rho \in \mathcal{R}(A^T) = \mathcal{N}(B)$. In interior point methods it is common to scale the problem about an interior solution s^0 by setting $\bar{s} = S_0^{-1}s$, where $S_0 = \operatorname{diag}(s^0)$. Doing this in (D) by multiplying the constraint by S_0^{-1}, the scaled cost will be $\bar{\rho} = \bar{A}^T(\bar{A}\bar{A}^T)^{-1}b$, with $\bar{A} = S_0^{-1}A$. Note that now $\bar{\rho} \in \mathcal{R}(\bar{A}^T)$. In primal problems this is usually called *projected scaled cost*, and will be important ahead.

Most of the standard references in interior point methods study the primal problem. The symmetry between (P) and (D') allows us to adapt their results for the dual problem. This is done in our paper [13]. Here we shall quote the results directly for the dual problem, because this is what we need for this paper.

Central points. The logarithmic barrier function is defined by

$$s \in I\!R^n_{++} \mapsto p(s) = -\sum_{i=1}^{n} \ln s_i.$$

The central paths associated with (P) and (D) are respectively

$$\mu \in I\!R_{++} \mapsto x(\mu) = \operatorname*{argmin}_{x \in \mathcal{X}^0} \left\{ \frac{c^T x}{\mu} + p(x) \right\}.$$

$$\mu \in I\!R_{++} \cup \{+\infty\} \mapsto (y(\mu), s(\mu)) = \operatorname*{argmax}_{(y,s) \in \mathcal{D}^0} \left\{ \frac{b^T y}{\mu} - p(s) \right\},$$

with the convention that $1/\infty = 0$. Only $\mathcal{D}$ is bounded, by hypothesis. Then the dual central path starts at $\mu = +\infty$, and the starting point is the analytic center of $\mathcal{D}$ (or of $\mathcal{S}$, or of Ω).

The analytic center and the central points of $\mathcal{D}$ are well defined, since $p(\cdot)$ is strictly convex and unbounded near the boundaries of $\mathcal{X}$ and $\mathcal{S}$.

The following lemma proved for instance in [11] relates the primal and dual central paths to the so called primal-dual central path:

LEMMA 2.1. *Given $\mu \in (0, +\infty), (x(\mu), s(\mu))$ is the unique solution of the relaxed primal-dual equations given by*

$$xs = \mu\, e, \quad x \in \mathcal{X},\ s \in \mathcal{S}.$$

From this lemma we see that for μ growing, $s(\mu)$ approaches the analytic center of $\mathcal{S}$, and $x(\mu)$ will diverge.

Proximity measures and centering. Let $s \in \mathcal{S}^0$ be a given dual solution, and consider a penalty parameter value $\mu \geq 0$. The Newton step from s for the problem of finding the central point $s(\mu)$ is well known, and described in detail in [13]. It is computed as follows:

$$h(s,\mu) = S\bar{h}(s,\mu), \tag{2.1}$$

where $\bar{h}(s,\mu)$ is the scaled centering step, which corresponds to the Newton centering step from $\bar{s} = e$ after a change of coordinates (scaling) which maps s to e. It is given by

$$\bar{h}(s,\mu) = -\bar{A}^T(\bar{A}\bar{A}^T)^{-1}\left(\frac{b}{\mu} - \bar{A}e\right) = -\frac{\bar{\rho}}{\mu} + \bar{e}_p, \tag{2.2}$$

with $\bar{A} = AS^{-1}$. In this expression, $\bar{e}_p = \bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{A}e$ is the projection of e into $\mathcal{R}(\bar{A}^T)$.

In terms of y, the Newton centering direction satisfies

$$\begin{array}{lrcl} (2.3) & h(y,\mu) & = & (\bar{A}\bar{A}^T)^{-1}\left(\dfrac{b}{\mu} - \bar{A}e\right) = -(\bar{A}\bar{A}^T)^{-1}\bar{A}\bar{h}(s,\mu), \\ & \bar{h}(s,\mu) & = & -\bar{A}^T h(y,\mu). \end{array}$$

Proximity. Let $\mu \geq 0$, $x \in \mathcal{X}^0$ and $s \in \mathcal{S}^0$ be given. Proximity measures are estimates of the distance from s to $s(\mu)$, or from (x,s) to $(x(\mu),s(\mu))$. The following measures will be relevant:

The primal-dual proximity:

$$\delta(x,s,\mu) = \left\| \frac{xs}{\mu} - e \right\|.$$

The dual proximity:

$$\delta(y,\mu) \equiv \delta(s,\mu) = \left\| \bar{h}(s,\mu) \right\|.$$

The dual proximity to the analytic center: for $\mu = +\infty$, $y(+\infty)$ or $s(+\infty)$ characterize the analytic center of $\mathcal{D}$, and the following result is deduced from (2.2) in [13]:

$$(2.4)\ \delta(s) = \left\|\bar{h}(s,+\infty)\right\| = \|\bar{e}_p\| = \left\|\bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{A}e\right\| = \min\{\|xs - e\| \mid Ax = 0\}.$$

As above, $\delta(y) \equiv \delta(s)$. The notation for the proximities differs in the arguments, which should not cause any confusion.

The dual and primal-dual proximities are related by the following fact: for $\mu > 0$ and $s \in \mathcal{S}^0$,

$$(2.5) \qquad \delta(s,\mu) = \min\{\delta(x,s,\mu) \mid x \in \mathcal{X}\}.$$

For given $\mu > 0$ and $s \in \mathcal{S}^0$, the primal solution for which this minimum is assumed is given by the following important theorem:

THEOREM 2.2. *Let $\mu > 0$ and $s \in \mathcal{S}^0$ be such that $\delta(s,\mu) < 1$. Then*

$$(2.6) \qquad x = \mu S^{-1}(e - \bar{h}(s,\mu)) \in \mathcal{X}^0$$

and $\delta(s,\mu) = \delta(x,s,\mu)$. Besides this,

$$(2.7) \qquad \delta(s + h(s,\mu),\mu) \leq \delta(x, s + h(s,\mu),\mu) \leq (\delta(s,\mu))^2.$$

The last statement of the theorem is possibly the most important fact in interior point methods, ensuring the quadratic convergence of the Newton centering algorithm, and proved in many references. The first statement, proved in [11], is not so well known: it shows how to associate a nearly central primal solution to a given nearly central dual solution, obtaining a nearly central primal-dual solution. This will allow us to initialize a primal-dual path following algorithm from a given dual solution.

Primal-dual centering. Primal-dual path following methods are described in many references, like the books [27] and [32]. We shall use properties of these methods ahead, and here we quote the main result on primal-dual centering (without describing the procedure): given an interior primal-dual solution (x, s) such that $\delta(x, s, \mu) = \delta < 1$ for some $\mu > 0$, a primal-dual Newton centering step results in a solution (x^+, s^+) such that

$$\delta(x^+, s^+, \mu) \leq \frac{1}{\sqrt{8}} \frac{\delta^2}{1-\delta}. \tag{2.8}$$

This establishes the quadratic convergence of the centering procedure for the primal-dual case. It is proved by Mizuno [21] and in the books cited above.

3. The auxiliary problem.

In this section we present the main constructions needed for adding a cut to a polytope. Consider the problem stated in the Introduction, in which a a cut is added to the polytope $\Omega \subset I\!R^m$ to form the polytope $\tilde{\Omega}$. We assume that both polytopes have non-empty interiors.

The problem. Assume that an approximate analytic center y^0 for Ω is known, and we want to find an approximate analytic center for $\tilde{\Omega}$. Formally, let $0 < \alpha \leq \beta < 1$ be given (typically $\beta \leq 0.5$ and either $\alpha = \beta$ or $\alpha = \beta^2$). Assume that $\delta(y^0) \leq \alpha$. Find $y \in \tilde{\Omega}$ such that $\tilde{\delta}(y) \leq \beta$, where $\delta(\cdot)$ and $\tilde{\delta}(\cdot)$ denote respectively the proximities to the analytic centers of Ω and $\tilde{\Omega}$.

These sets can also be written in terms of the slack variables, defining

$$\begin{array}{rcl} \mathcal{S} & = & \{s \in I\!R^n_+ \mid A^T y + s = c,\ y \in I\!R^m\} \\ \tilde{\mathcal{S}} & = & \{\tilde{s} = (s, \tilde{s}_{n+1}) \in I\!R^{n+1}_+ \mid A^T y + s = c,\ a^T y + s_{n+1} = K,\ y \in I\!R^m\}. \end{array} \tag{3.1}$$

With the data above, let us construct the following *auxiliary* linear programming problem in dual format

$$(D) \qquad \begin{array}{lll} \text{maximize} & -a^T y & \\ \text{subject to} & A^T y + s & = c \\ & s & \geq 0, \end{array}$$

which has as associated *auxiliary* primal problem

$$(P) \qquad \begin{array}{lll} \text{minimize} & c^T x & \\ \text{subject to} & Ax & = -a \\ & x & \geq 0. \end{array}$$

This pair of problems has an associated primal-dual central path

$$\mu > 0 \quad \longmapsto \quad (x(\mu), s(\mu)).$$

As we did in Problem (D'), the dual projected cost scaled about an interior solution $s \in \mathcal{S}^0$ will be

$$\bar{\rho}_s = -\bar{A}^T(\bar{A}\bar{A}^T)^{-1} a, \tag{3.2}$$

with $\bar{A} = S^{-1}A$. We introduce the following notation for the norm of this projected cost:

$$r_s = \|\bar{\rho}_s\| = \left\|\bar{A}^T(\bar{A}\bar{A}^T)^{-1} a\right\|.$$

Remark on notation. We are using the notation $\bar{A}$ and $\bar{\rho}_s$ to indicate vectors or matrices scaled about a given solution s. This solution will always be fixed and well known, and hence we shall drop the subscripts in $\bar{\rho}_s$ and r_s. There is a special situation: when the scaling is about the analytic center of $\mathcal{S}$, denoted σ, the scaled projected cost and its norm will be denoted $\hat{r} = \|\hat{\rho}\|$.

For computing the proximity to a central point, (2.2) gives

$$\bar{h}(s,\mu) = \bar{A}^T(\bar{A}\bar{A}^T)^{-1}\left(\frac{a}{\mu} + \bar{A}e\right) = -\frac{\bar{\rho}}{\mu} + \bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{A}e = -\frac{\bar{\rho}}{\mu} + \bar{e}_p. \tag{3.3}$$

The cost variation along $h(y,\mu)$ is

$$a^T h(y,\mu) = \bar{\rho}^T\bar{h}(s,\mu), \tag{3.4}$$

because $-\bar{A}\bar{\rho} = a$ by (3.2) and $\bar{h}(s,\mu) = -\bar{A}^T h(y,\mu)$ by (2.3).

From the analytic center to a central point. Our first lemma will give us the tools for solving the following problem: given an approximate center y^0 for Ω, with $\delta(y^0) \le \alpha \le \beta < 1$, choose $\mu > 0$ as small as possible and a primal-dual solution (x,y,s) such that $\delta(x,s,\mu) < \beta$.

We shall use the results of Sec. 2, with the substitution $b = -a$. Define $\bar{A} = A(S^0)^{-1}$, $\bar{\rho} = -\bar{A}^T(\bar{A}\bar{A}^T)^{-1}a$, $r \equiv r_{s^0} = \|\bar{\rho}\|$ and $\bar{h}(s^0,\mu)$ as in (3.3) for $\mu > 0$.

Remark. Examining (3.3) we become aware of a complicating factor: the role of μ as the parameter of the central path depends on the size r of the projected cost. Adding to a a vector orthogonal to $\mathcal{N}(B)$, i.e., a vector in $\mathcal{R}(A)$, has no effect on r, but changing the size of the projected cost affects it. In this section the results on the parameter will be associated with μ/r, with r fixed.

LEMMA 3.1. *Consider an interior dual solution $(y^0,s^0) \in \mathcal{D}$ such that $\delta(s^0) \le \alpha$. Given $\mu > 0$ with $\delta(s^0,\mu) < 1$, consider the pair (x_μ, s_μ) defined by*

$$\begin{aligned} s_\mu &= s^0 + h(s^0,\mu) = S^0(e + \bar{h}(s^0,\mu)) \\ x_\mu &= \mu(S^0)^{-1}(e - \bar{h}(s^0,\mu)). \end{aligned} \tag{3.5}$$

Then (x_μ, s_μ) is feasible for (P) and (D) and satisties

$$\delta(x_\mu, s_\mu, \mu) \le (\alpha + r/\mu)^2.$$

Proof. Consider the vectors x_μ and x_μ as above. Here $\bar{h}(s^0,\mu)$ is the scaled Newton step (2.2), $s_\mu > 0$ because $\|\bar{h}(s^0,\mu)\| < 1$, and x_μ is the primal solution in (2.6). Using Theorem 2.2, (x_μ, s_μ) is primal-dual feasible and

$$\delta(s^0, x_\mu, \mu) = \delta(s^0, \mu).$$

By (2.7),

$$\delta(x_\mu, s_\mu, \mu) \le (\delta(s^0,\mu))^2.$$

From (3.3), with $r = \|\bar{\rho}\|$ and $\delta(s^0) = \|\bar{e}_p\| \le \alpha$,

$$\delta(s^0,\mu) = \left\|\bar{h}(s^0,\mu)\right\| \le r/\mu + \alpha,$$

completing the proof □

This lemma gives immediate solutions to the problem stated above. Let us examine some possible computations.

Example 1. Solutions for the problem using a single linear system decomposition.
First solution: using the construction indicated in Lemma 3.1, we obtain for each $\mu > 0$, $\delta(x_\mu, s_\mu, \mu) \leq (\alpha + r/\mu)^2$. Setting $(\alpha + r/\mu)^2 = \beta$, we get the result

$$\frac{\mu}{r} = \frac{1}{\sqrt{\beta} - \alpha}.$$

The case in which $\beta = \alpha$ is interesting. In this case, Fig. 3.1 shows the values of μ/r associated with each α. We can also see that in the extreme situation in which $\alpha = 0$ and $\beta = 1$, this method results in $\mu = r$.
Second solution: A smaller value for μ can be found if we do not use the inequality above, computing directly the value of $\delta(x_\mu, s_\mu, \mu)$ for each $\mu > 0$. Using (2.3), we define the following mappings, needed in Lemma 3.1:

$$\mu > 0 \longmapsto \left\{ \begin{array}{rl} h(y^0, \mu) & = -(\bar{A}\bar{A}^T)^{-1}(a/\mu + \bar{A}e) \\ \bar{h}(s^0, \mu) & = -\bar{A}^T h(y^0, \mu) \\ s_\mu & = s^0 + S^0 \bar{h}(s^0, \mu) \\ x_\mu & = \mu(S^0)^{-1}(e - \bar{h}(s^0, \mu)) \\ y_\mu & = y^0 + h(y^0, \mu) \\ \delta(x_\mu, s_\mu, \mu) & = \|x_\mu s_\mu/\mu - e\|. \end{array} \right. \tag{3.6}$$

The computation of these functions needs only one system decomposition and two back substitutions, which appear in the computation of $(\bar{A}\bar{A}^T)^{-1}a$ and $(\bar{A}\bar{A}^T)^{-1}\bar{A}e$.

Now we can find a small μ that satisfies $\delta(x_\mu, s_\mu, \mu) \leq \beta$ either by means of a one-dimensional search on μ, or by solving a quartic equation which can be easily derived from the expressions above. Remember that the search procedure should be initialized with the value of μ given in the first solution above.

Example 2. Solution using two system decompositions.
Here we propose to use one of the solution methods above to find (x_μ, s_μ, μ), followed by a primal-dual centralization which produces (x_μ^+, s_μ^+, μ) with $\delta(x_\mu^+, s_\mu^+, \mu) \leq \beta$. All that we need is to compute $\bar{\delta} > 0$ such that

$$\beta = \frac{1}{\sqrt{8}} \frac{\bar{\delta}^2}{1 - \bar{\delta}},$$

and then use either of the methods above to find μ such that $\delta(x_\mu, s_\mu, \mu) \leq \bar{\delta}$. From x_μ, s_μ, compute x^+, s^+ by primal-dual centralization, with the desired result ensured by (2.8).

The interesting case in which the first solution method is used with $\beta = \alpha$ is depicted in Fig. 3.1. We see that the primal-dual centralization leads to quite better results. With one step, the minimum value of μ/r attainable is $\mu/r = 4$, for $\alpha = 0.25$. With the primal-dual centralization, $\mu/r = 1.85$ is achieved for $\alpha = 0.12$. We also see that the values of μ/r are not very sensitive to variations in $\alpha \in (0.1, 0.5)$. In particular, $\mu/r = 2.4$ can be reached for $\alpha = 0.1$. This will be useful in Section 6.

From a central point to an analytic center.

Let us now examine the problem in the opposite direction. The next lemma shows under what conditions a point y near the central path for the auxiliary problem is also near the analytic center of the set $\tilde{\Omega}$.

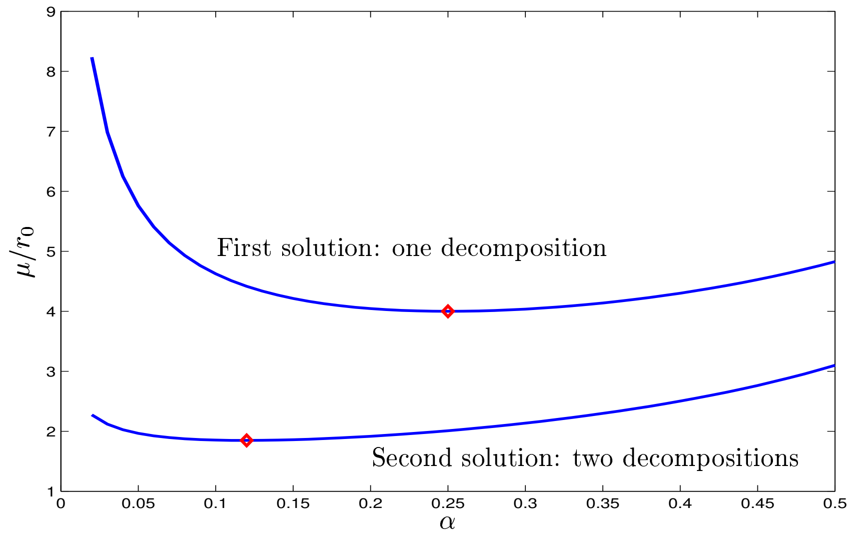

FIG. 3.1. *From an α-center to an α-central point in Example 1.*

LEMMA 3.2. *Let (x, y, s) be a primal-dual feasible solution for the problems (P) and (D) such that $a^T y \leq K$, and consider $\mu > 0$. Then*

$$\tilde{\delta}(y)^2 \leq \delta(x, s, \mu)^2 + \left(\frac{K - a^T y}{\mu} - 1\right)^2.$$

In particular, for $\mu = K - a^T y$, $\tilde{\delta}(y) \leq \delta(x, s, \mu)$.

Proof. After the introduction of the cut, the feasible dual solutions $y \in I\!R^m$ and $\tilde{s} \in I\!R^{n+1}$ defined in (3.1) satisfy

$$\begin{bmatrix} A^T \\ a^T \end{bmatrix} y + \tilde{s} = \begin{bmatrix} c \\ K \end{bmatrix}, \quad \tilde{s} \geq 0.$$

The primal solutions $\tilde{x} \in I\!R^{n+1}$ associated with the feasibility problem $\tilde{s} \in \tilde{\mathcal{S}}$ satisfy

$$[A\ a]\tilde{x} = 0, \quad \tilde{x} \geq 0.$$

Let (x, y, s) be feasible for (P) and (D) such that $a^T y \leq K$ and consider $\mu > 0$. We shall construct the following vectors in $I\!R^{n+1}$:

$$\tilde{x} = \begin{bmatrix} x/\mu \\ 1/\mu \end{bmatrix}, \quad \tilde{s} = \begin{bmatrix} s \\ K - a^T y \end{bmatrix}.$$

We now prove that these vectors provide feasible solutions for the augmented problem. By direct substitution,

$$\begin{bmatrix} A^T \\ a^T \end{bmatrix} y + \begin{bmatrix} s \\ K - a^T y \end{bmatrix} = \begin{bmatrix} A^T y + s \\ a^T y + K - a^T y \end{bmatrix} = \begin{bmatrix} c \\ K \end{bmatrix}.$$

From the hypotheses we know that $s \geq 0$ and $s_{n+1} = K - a^T y \geq 0$, and hence $\tilde{s} \in \tilde{\mathcal{S}}$. Besides this, using the primal feasibility,

$$[A\ a] \begin{bmatrix} x/\mu \\ 1/\mu \end{bmatrix} = (Ax + a)/\mu = 0.$$

Thus the vectors $(\tilde{x}, y, \tilde{s})$ are a feasible primal-dual solution for the problem of finding the analytic center of $\tilde{\Omega}$. By (2.4), the proximity to the analytic center satisfies

$$\begin{aligned}\tilde{\delta}(y)^2 &\leq \|\tilde{x}\tilde{s} - e\|^2 \\ &= \left\| \begin{bmatrix} xs/\mu - e \\ (K - a^T y)/\mu - 1 \end{bmatrix} \right\|^2 \\ &= \|xs/\mu - e\|^2 + ((K - a^T y)/\mu - 1)^2.\end{aligned}$$

completing the proof. □

COROLLARY 3.3. *Let (x, y, s) be a primal-dual feasible solution for (P) and (D) such that $\delta(x, s, \mu) < 1$ for a given $\mu > 0$. Then for $K = a^T y + \mu$, $\tilde{\delta}(y) \leq \delta(x, s, \mu)$.*

Proof. Direct application of Lemma 3.2 □

4. The algorithm for introducing a cut. Last section gave us two results: Lemma 3.1 shows the way to go from an approximate center of Ω to a nearly central primal-dual solution for the auxiliary problem; Lemma 3.2 and its corollary show the way to go from a nearly central primal-dual solution to an approximate center of the new polytope $\tilde{\Omega}$. The former result provides the initialization for a search procedure; the latter, a termination rule.

Set us state again the problem to be solved, keeping the same notation:

The direct problem. Consider the sets Ω and $\tilde{\Omega} = \{y \in \Omega |\ a^T y \leq K\}$, for a given value of K and constants $0 < \alpha \leq \beta < 1$. Given $y^0 \in \Omega$ such that $\delta(y^0) \leq \alpha$, find $y \in \tilde{\Omega}$ such that $\tilde{\delta}(y) \leq \beta$.

Using the results of Lemma 3.2, it is enough to solve the following problem:

Direct problem (restated). In the conditions above, find (x, u, s) feasible for (P) and (D) and $\mu > 0$ such that

$$\delta(x, s, \mu)^2 + ((K - a^T y)/\mu - 1)^2 \leq \beta^2, \tag{4.1}$$

or in a still simpler formulation, such that

$$\delta(x, s, \mu) \leq \beta \quad , \quad \mu = K - a^T y. \tag{4.2}$$

In fact, if (4.1) or (4.2) is satisfied, then Lemma 3.2 implies $\tilde{\delta}(y) \leq \beta$.

The solution of the direct problem is the main purpose of this paper. Before stating the main algorithm, let us show some examples of direct problem solutions.

Example 3. A shallow cut: consider the direct problem. Compute $\mu > 0$ and a primal-dual solution (x_μ, y_μ, s_μ) as in Example 1. Now $\delta(x'_\mu, s'_\mu, \mu') \leq \beta$ for any $\mu' \in [\mu, +\infty]$. Check the condition $K - a^T y_\mu \geq \mu$. If it holds, then the cut is shallow and for some $\mu' \in [\mu, +\infty)$, we must have $K - a^T y_{\mu'} = \mu'$. Computing this value μ' by a line search on μ using the maps (3.6), the problem is solved.

Example 4. A deeper shallow cut: as above, compute $\mu > 0$ and a primal-dual solution (x_μ, y_μ, s_μ) as in Example 1 and check the condition $K - a^T y_\mu \geq \mu$. If it holds, then Example 3 provides a solution; otherwise we may use (4.1) and solve the following one-dimensional problem:

$$\begin{array}{ll} \text{minimize} & \delta(x_\mu, s_\mu, \mu)^2 + ((K - a^T y_\mu)/\mu - 1)^2 \\ \text{subject to} & K - a^T y_\mu \ > \ 0 \\ & \mu \ > \ 0. \end{array}$$

If this problem results in a value lower than β^2, then the cut can be introduced.

Example 5. Interpolation between two nearly central points.
Consider two feasible primal-dual solutions as follows:

$$(x^+, y^+, s^+) = (x, y, s) + (dx, dy, ds) \quad , \quad \mu^+ = \mu - d\mu.$$

Assume that they satisfy the following condition, very common in interior path following algorithms: for any $\lambda \in [0, 1]$, $\delta(x + \lambda dx, s + \lambda ds, \mu - \lambda d\mu) \leq \beta$.

Assume that for the given $K > 0$, $K - a^T y \leq \mu$ and $K - a^T y^+ \geq \mu^+$.

By interpolation we can immediately find $\lambda \in [0, 1]$ such that $\mu' = \mu - \lambda d\mu$ together with $(x', y', s') = (x, y, s) + \lambda(dx, dy, ds)$ satisfy $K - a^T y' = \mu'$ and $\delta(x', s', \mu') \leq \beta$. In fact it is trivial to check this for $\lambda = (K - a^T y - \mu)/(a^T dy - d\mu)$. Using Corollary 3.3, $\tilde{\delta}(y') \leq \beta$.

Now we are ready to state the main algorithm for solving the direct problem. The algorithm has a first step in which the maps (3.6) are used, as in Examples 1 or 2 in the last section. If the cut is shallow, the algorithm terminates as in Examples 3 or 4.

If the cut is not shallow, then we start a primal-dual path following algorithm from this point. We shall not specify the path-following method in detail. We assume that the path-following algorithm works in a $\beta-$neighborhood of the central path, generating a sequence (x^k, y^k, s^k, μ^k), such that for $k = 0, 1, \ldots$, $\delta(x^k, s^k, \mu^k) \leq \beta$. Each iteration of the path following algorithm computes a step

$$(x^{k+1}, y^{k+1}, s^{k+1}, \mu^{k+1}) = (x^k + dx^k, y^k + dy^k, s^k + ds^k, \mu^k - d\mu^k)$$

such that $d\mu^k > 0$ and for $\lambda \in [0, 1]$, $\delta(x^k + \lambda dx^k, sk + \lambda ds^k, \mu^k - \lambda d\mu^k) \leq \beta$.

These characteristics are shared by most path following algorithms which stay in the $\beta-$neighborhood of the central path, like the Mizuno-Todd-Ye algorithm [22] or the largest step algorithm [12].

We present the algorithm in a simple formulation, and comment on possible variations afterwards.

ALGORITHM 4.1. *Introduction of a deep cut.*
Data: A precision $\epsilon > 0$, $y^0 \in \Omega$ such that $\delta(y^0) \leq \alpha \leq \beta < 1$, $a \in I\!R^n$, $K \in I\!R$.
k=0
First step (see Example 3)
  Construct the maps in (3.6) and compute μ^1 such that
  $\delta(x^1, s^1) = \beta$, where $(x^1, y^1, s^1) = (x_{\mu^1}, y_{\mu^1}, s_{\mu^1})$.
  If $K - a^T y^1 \geq \mu^1$ then
    (The cut is shallow)
    Compute $\mu \in [\mu^1, +\infty]$ such that $K - a^T y_\mu = \mu$.
    Set $y^+ = y_\mu$ and STOP
Path following:
  $k = 1$.
  REPEAT
    Compute $(dx, dy, ds, d\mu)$ such that
      $\delta(x^k + \lambda dx, s^k + \lambda ds, \mu^k - \lambda d\mu) \leq \beta$ for $\lambda \in [0, 1]$.
    $(x^{k+1}, y^{k+1}, s^{k+1}, \mu^{k+1}) = (x^k + dx, y^k + dy, s^k + ds, \mu^k - d\mu)$.
    $k = k + 1$.
  UNTIL $\mu^k \leq \epsilon$ or $K - a^T y^k \geq \mu^k$.
Interpolation (see Example 5)
  IF $\mu^k < \epsilon$ then STOP with no solution.

ELSE

Set $\lambda = \dfrac{\mu^{k-1} - K + a^T y^{k-1}}{d\mu - a^T dy}$.

Set $y^+ = y^{k-1} + \lambda dy$.

If the algorithm stops with a solution y^*, then clearly $\tilde{\delta}(y^*) \leq \beta$, as we proved in Example 5. Otherwise, the duality gap for the auxiliary problem is smaller than $n\epsilon$, where n is the number of constraints. In this case the cut is excessively deep or infeasible.

Remarks.
Using a larger initial step like in Example 2 does not seem to make sense, because then if the cut is shallow, the interpolation between y^0 and y^+ cannot be done. Anyway, the second Newton step in the example should already be a step in the main path following method.

In practical implementations, one should use large neighborhoods of the central path to generate deep cuts, and spend more effort in centralizations only when the stopping rule is met, if necessary.

The termination condition is simple, based on Corollary 3.3. A more elaborate condition, resulting in a better centered resulting point may be devised using the full power of Lemma 3.2, as we did in Example 4. This is a simple exercise, possibly with little interest.

5. Shallow cuts: the inverse problem. We shall loosely say that a cut is shallow if we can find the new approximate center using a single Newton iteration (of course, we could also use a fixed number of iterations). The study of shallow cuts includes most of the cases studied in the literature, and for this reason we show what can be obtained using our approach.

We have already shown in Examples 3 and 4 how to compute shallow cuts using Lemmas 3.1 and 3.2 for a given value of K (the direct problem), whenever this is possible. Now we tackle the problem of deciding for which values of K we can ensure that the cut is shallow.

The inverse problem. Consider the polytope Ω, the vector $a \in I\!R^m$, the numbers $0 < \alpha \leq \beta < 1$ as before. Given $y^0 \in \Omega$ such that $\delta(y^0) \leq \alpha$, find $K \in I\!R$ as small as possible and $y \in \tilde{\Omega} = \{z \in I\!R^m;\ a^T z \leq K,\ z \in \Omega\}$ such that $\tilde{\delta}(y) \leq \beta$, using a single Newton iteration (a single system decomposition).

Remark. Remember that by definition (see (3.2)), $r = r_{s^0} = \|\bar{\rho}\| = \left\|\bar{A}^T(\bar{A}\bar{A}^T)^{-1}a\right\|$. This is the norm of the projected cost, after a scaling that transfers s^0 to the vector e. We could normalize the problem by setting $r = 1$, but we chose not to do it: all cost-related entities will then depend on r, and appear as K/r, μ/r, etc. The reader may simplify the geometry by assuming without loss of generality that $s^0 = e$ and $r = 1$. In this scaled space the unit ball centered in e is mapped onto the *Dikin ellipsoid* in Ω centered in y^0 (see the geometrical interpretation ahead).

We start from the maps defined in (3.6), and define the new map

$$\mu > 0 \mapsto K_\mu = a^T y_\mu + \mu.$$

LEMMA 5.1. *Consider the inverse problem. For any $\mu > 0$,*

(i) $\dfrac{K_\mu}{r} \leq \dfrac{a^T y^0}{r} - \dfrac{r}{\mu} + \dfrac{\mu}{r} + \alpha.$
(ii) If $K = K_\mu$ *then* $\tilde{\delta}(y_\mu) \leq (\alpha + r/\mu)^2$.

Proof. By definition,

$$K_\mu = a^T y_\mu + \mu = a^T y^0 + a^T h(y^0, \mu) + \mu. \tag{5.1}$$

Let us develop the term $a^T h(y^0, \mu)$ in this expression: from (3.3) and (3.4),

$$a^T h(y^0, \mu) = \bar{\rho}^T \bar{h}(s^0, \mu) = \bar{\rho}^T(-\bar{\rho}/\mu + \bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{A}e).$$

Using the definition of $\delta(y^0)$ in (2.4), $r = \|\bar{\rho}\|$, and the Cauchy-Schwarz inequality,

$$a^T h(y^0, \mu) \leq -\frac{r^2}{\mu} + r\delta(y^0) \leq -\frac{r^2}{\mu} + r\alpha,$$

because $\delta(y^0) \leq \alpha$. Introducing this into (5.1), the proof of this item is complete.
(ii) If $K = K_\mu$, then the result is a consequence of Lemma 3.2 followed by Lemma 3.1:

$$\tilde{\delta}(y_\mu) \leq \delta(x_\mu, s_\mu, \mu)^2) \leq (\alpha + r/\mu)^2,$$

completing the proof. ☐

Let us define the depth of the cut $a^T y \leq K$ as

$$\Delta = \frac{a^T y^0 - K}{r}.$$

A cut through y^0 has $\Delta = 0$. If the cut excludes y^0, then $\Delta > 0$. The depth Δ is the Euclidean distance from s^0 to the new polytope in terms of s, for the scaled problem.

Choosing $K = K_\mu$ like we did in Example 3, (i) gives us the following: we can only ensure $\Delta \geq 0$ for the extreme case in which $\alpha = 0$, $\beta = 1$ e $r/\mu = 1$. For any other choice of $0 \leq \alpha \leq \beta \leq 1$, $\Delta < 0$, and we cannot cut the point y^0.

Deeper cuts will have $K < K_\mu$. Let us study this using the full power of Lemma 3.2.

LEMMA 5.2. *Consider the inverse problem. If* $\mu > 0$ *and* $K \in I\!R$ *are such that* $0 < K - a^T y_\mu < \mu$, *then*

$$\begin{aligned}\tilde{\delta}(y_\mu)^2 &\leq \left(\frac{r}{\mu} + \alpha\right)^4 + \left(\frac{K - K_\mu}{\mu}\right)^2 \\ &\leq \left(\frac{r}{\mu} + \alpha\right)^4 + \left(\frac{r}{\mu}\left(\Delta + \frac{r}{\mu} - \alpha\right) - 1\right)^2.\end{aligned} \tag{5.2}$$

Proof. From Lemma 3.2, we have for $\tilde{\Omega}$ (adding and subtracting K_μ from the numerator)

$$\tilde{\delta}(y_\mu)^2 \leq \delta(x_\mu, s_\mu, \mu)^2 + \left(\frac{K - K_\mu}{\mu} + \frac{K_\mu - a^T y_\mu}{\mu} - 1\right)^2.$$

By Lemma 3.1, $\delta(x_\mu, s_\mu, \mu) \leq (\alpha + r/\mu)^2$ and by definition, $K_\mu - a^T y_\mu = \mu$, and hence the first inequality holds.

From Lemma 5.1(i),

$$0 < \frac{K_\mu - K}{r} \le \frac{a^T y^0 - K}{r} - \frac{r}{\mu} + \frac{\mu}{r} + \alpha = \Delta - \frac{r}{\mu} + \frac{\mu}{r} + \alpha.$$

Multiplying by r/μ we obtain the second inequality, completing the proof. □

The deepest possible cut for given α and β ensured by this lemma corresponds to the maximum value of Δ such that the last term of (5.2) is bounded by β^2. We shall not try to derive an analytic solution for this, but it is easy to obtain numerical solutions.

Remark. The following reasoning is a nice exercise: setting

$$\left(\alpha + \frac{r}{\mu}\right)^2 = \beta_1 \quad , \quad \frac{r}{\mu}\left(\Delta - \frac{r}{\mu} + \alpha\right) + 1 = \beta_2,$$

we must maximize Δ subject to $\beta_1^2 + \beta_2^2 = \beta^2$ and $\beta_1 > \alpha$. With a little manipulation, we isolate

$$\Delta = \sqrt{\beta_1} - \frac{1 - \beta_2}{\sqrt{\beta_1} - \alpha} - 2\alpha.$$

Now, setting $\beta_1 = \beta \cos\phi$ and $\beta_2 = \beta \sin\phi$, and hence $\beta_1^2 + \beta_2^2 = \beta^2$, we maximize Δ by a search in $\phi \in [0, \pi/2]$, with the constraint $\beta_1 > \alpha$.

Solving this problem numerically (using for instance the method described above), we plot in Fig. 5.1 the values of the maximum depth Δ as a function of β for several values of α.

The figure in the bottom shows in detail the parameter values for which $\Delta > 0$, i.e., for which it is possible to cut the point y^0 and obtain a β-proximal center using a single Newton step.

The table lists some of these values for the parameters. The extreme situation, for $\alpha = 0$ and $\beta = 1$, allows a cut with depth $\Delta = 0.52$, which is thus the largest depth that can be guaranteed using a single Newton step.

α	β	r/μ	Δ
0	1	0.8	0.52
0.01	0.99	0.9	0.48
0.01	0.9	0.77	0.33
0.1	0.9	0.6	0.1
0.2	0.8	0.56	-0.47

These results reproduce the ones previously found by Goffin and Vial [6]. In this paper there is also a geometrical interpretation of the results, which we now reproduce in our language.

Geometrical interpretation.

In the problem studied above, consider the point y^0 and another point y such that $a^T y = K$. Assume that the cut defined by $a^T y = K$ is such that $\Delta > 0$. Then we have by definition of Δ, $a^T(y^0 - y) = r\Delta$.

Let now s^0 and s be the slacks associated with these points. We have $\rho^T(s^0 - s) = r\Delta$. Scaling the problem with $\bar{s} \equiv (S^0)^{-1}s$, we get $\bar{\rho}^T(e - \bar{s}) = r\Delta$. Using the Cauchy-Schwarz inequality and noting that by definition $r = \|\bar{\rho}\|$, we conclude that $\|e - s\| \ge \Delta$.

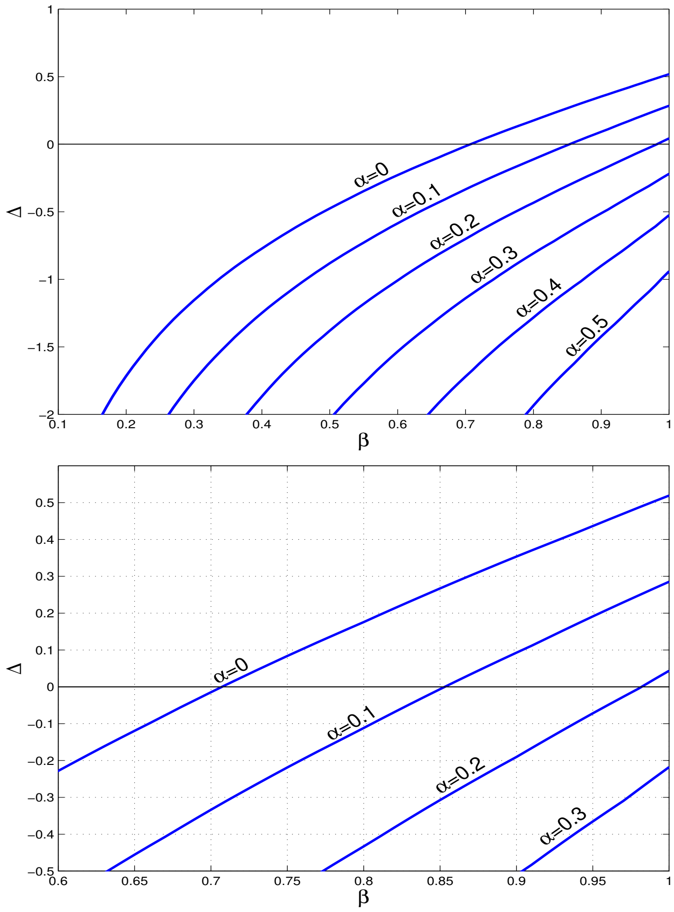

FIG. 5.1. *Maximum depths ensured by a single Newton step as a function of β for several values of α.*

This means that in the problem scaled in terms of s, the new cut excludes the ball with radius Δ centered in e. For the original problem, this ball corresponds to the Dikin ellipsoid with radius Δ, which can be represented either in terms of s or in terms of y. In Fig. 5.2 we represent a polytope Ω, a point y^0 near the analytic center of Ω and a cut with $\Delta = 0.4$. We also represent the Dikin ellipsoids with radius equal to 1 and Δ. We can see that the cut excludes the ellipsoid with radius Δ. The figure also shows the central path associated with the cost a, on which the analytic center of the new polytope will be located.

6. The convex feasibility problem. The algorithm developed in this paper has its most interesting application in the convex feasibility problem. We shall state this problem following Goffin, Luo and Ye [9], and show that the same complexity bounds obtained by them is achieved by our method.

Consider a unit box $\Omega_{2m} \subset I\!R^m$ defined by the $2m$ inequalities $0 \leq y_i \leq 1$, $i = 1, \ldots, m$, and a convex set $\Omega^* \subset \Omega_{2m}$ such that Ω^* contains a ball of radius $\varepsilon \in (0, 1)$. Define $y^{2m} = 0.5e$, the analytic center of Ω_{2m}. We want to find a point $y \in \Omega^*$.

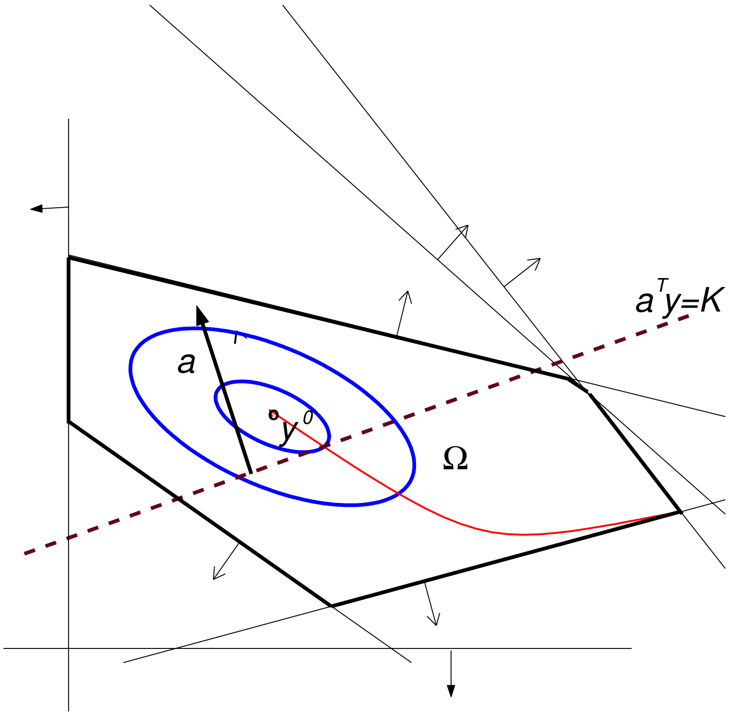

FIG. 5.2. *A cut with $\Delta > 0$ excludes the Dikin ellipsoid with radius Δ. In the example, $\Delta = 0.4$.*

A central cutting plane method generates for $j = 2m+1, \ldots$ sequences of polytopes $\Omega_j \subset I\!R^m$, points $y^j \in \Omega_j$, vectors $a_j \in I\!R^m$ and constants $c_j \in I\!R$ with the following characteristics:

- The point y^j is an approximate analytic center of Ω_j, i.e., $\delta_j(y^j) \le 0.1$, where $\delta_j(\cdot)$ stands for the proximity to the analytic center of Ω_j.
- $\Omega_j = \{y \in \Omega_{j-1} \mid a_j^T y \le c_j\}$.
- $\Omega^* \subset \Omega_j$.
- y^{j-1} is separated from Ω_j by the hyperplane defining the constraint $a_j^T y \le c_j$.
- $\|a_j\| = 1$.

Each iteration $j = 2m+1, \ldots$ of the algorithm starts with Ω_{j-1} and computes an approximate analytic center y^j. If $y^j \in \Omega^*$, then the algorithm terminates; otherwise it calls an oracle which generates a constraint $a_j^T y \le c_j$ which cuts the point y^j, i.e., $a_j^T y^j \ge c_j$.

Each polytope Ω_j, $j = 2m, \ldots$, is characterized by the linear system $A_j^T y \le c$, where the first $2m$ constraints define the unit box, and the other are the hyperplanes generated by the oracle. Although the problem is stated using the polytopes Ω_j, the algorithm and the analysis will be mostly based on the associated set of slacks $\mathcal{S}_j = \{s \in I\!R^m_+ \mid A_j^T y + s = c\}$. Each polytope has an associated barrier function $s > 0 \mapsto p(s) = -\sum_{i=1}^j \log s_i$. The *potential* of the polytope Ω_j is defined as $P(\Omega_j) = p(\sigma_j)$, where σ_j is the analytic center of $\mathcal{S}_j$. We also define for each $j = 2m+1, \ldots$ the value

$$\hat{r}_j = \|\hat{\rho}_j\| = \left\| \hat{A}_{j-1}^T (\hat{A}_{j-1} \hat{A}_{j-1}^T)^{-1} a_j \right\|, \tag{6.1}$$

which comes from (3.2) with $\hat{A}_j = \Sigma_j^{-1} A_j$. The values $\hat{r}_j$ cannot be computed, because the analytic centers are not known, but we shall see below that we have very good approximations to it.

A very refined analysis made in the cited references [24, 9] leads to the following result: assume that each iteration $j = 2m+1, \ldots$ uses $k_j \geq 1$ Newton steps to find y^j, and achieves the following result:

$$P(\Omega_j) \geq P(\Omega_{j-1}) + \zeta + k_j \xi - \log(\hat{r}_j), \tag{6.2}$$

where $\xi > 0$ and ζ (possibly negative) are constants. Then the algorithm terminates with $j = O^*(m^2/\varepsilon^2)$, and the total number of Newton iterations is bounded by

$$\sum_{j > 2m} j_k = O^* \left(\frac{m^2}{\varepsilon^2} \log \left(\frac{1}{\varepsilon} \right) \right).$$

We shall now show that our path-following algorithm has these properties, using technical results proved in Gonzaga and Cardia [13].

For proving complexity results, we shall use a primal short steps path following algorithm. Efficient primal-dual path following algorithms like the one by Mizuno, Todd and Ye [22] perform at least as well as the short steps method, and should lead to the same bounds. In most of this section we use the same notation as in the previous sections, dropping the indices j. We start from a nearly central point for the polytope Ω and add the cut $a^T y \leq K$.

We state the algorithm and then show that it is well defined.

ALGORITHM 6.1. *Short steps algorithm for the introduction of a deep cut.*

Data: A primal-dual pair (y^0, s^0) such that $\delta(s^0) \leq 0.1$, $a \in I\!R^m$, $K \in I\!R$ such that $K \leq a^T y^0$.

Initialization: set $\bar{A} = (S^0)^{-1} A$, $r = \left\| \bar{A}^T (\bar{A} \bar{A}^T)^{-1} a \right\|$, $\mu^1 = 2.4\, r$.

First step (see Example 3)

Using three Newton steps, find (x^1, y^1, s^1) such that $\delta(x^1, s^1, \mu^1) \leq 0.01$

Path following:

$k = 1$.

WHILE $K - a^T y^k < \mu^k$

Compute μ^{k+1} such that $\delta(s^k, \mu^{k+1}) = 0.1$.

$s^{k+1} = s^k + h(s^k, \mu^{k+1})$.

$x^{k+1} = \mu^{k+1} (S^k)^{-1} (e - \bar{h}(s^k, \mu^{k+1}))$.

k=k+1.

END WHILE

Find the result by interpolation, as in Example 5.

Now we describe the properties of this algorithm.

Initialization: we saw in Example 2 that it is easy to obtain (x, y, s) such that $\delta(x, s, \mu^1) \leq 0.1$ for $\mu^1 = 2.4\, r$ in two Newton iterations. One extra centralization step gives (x^1, y^1, s^1) such that $\delta(s^1, \mu^1) \leq \delta(x^1, s^1, \mu^1) \leq 0.01$.

For $k > 1$, Each iteration finds μ^{k+1} such that $\delta(s^k, \mu^{k+1}) = 0.1$, which is done using (2.2), and the Newton step reduces the proximity to $\delta(s^{k+1}, \mu^{k+1}) \leq \delta(x^{k+1}, s^{k+1}, \mu^{k+1}) \leq 0.01$.

So, the algorithm is well defined. Note that it generates a sequence of primal-dual pairs (s^k, x^k), but in reality it is a purely dual algorithm, in the sense that x^k is not used in the computation of the step. So, the computation of x^k is not needed at all, unless for the interpolation in the last iteration.

In the proofs that follow we use several properties of proximities proved in the reference [13]. We shall only summarize these properties here, referencing directly the lemmas in that paper. We shall use the scaled norm of a vector, defined for given $s, h \in I\!R^n$, $s > 0$, as $\|h\|_s = \left\|S^{-1}h\right\|$. The proximity measure is a very good approximation for the scaled distance between a nearly central and the associated central point. The following results are taken from [13]:
(i) Relationship between scalings: If $\|u - v\|_u \leq 0.5$ for $u, v \in I\!R^m_{++}$, then for $h \in I\!R^m$ $\|h\|_v \leq 2\|h\|_u$ (from Lemma 2.2).
(ii) Relationship between proximity and the scaled norm: If $\delta(s,\mu) \leq 0.45$, then $\|s - s(\mu)\|_{s(\mu)} \leq 1.05\delta(s,\mu)$ (from Lemma 3.1).
(iii) Effect of a change of scale on the scaled projected cost: If $\delta(s) \leq 0.01$, then $0.98\,\hat{r} \leq r_s \leq 1.02\,\hat{r}$ (from Lemma 3.5).
(iv) The central path does not approach the analytic center $\sigma = s(+\infty)$: If $\mu \leq 2.5\hat{r}$ and $\delta(s,\mu) \leq 0.1$, then $\delta(s) \geq 0.2$ (from Lemma 3.7) and $p(s) - p(\sigma) \geq 0.018$ (from Lemma 3.9).

We start by proving that the algorithm does not stop in the first iteration.

LEMMA 6.2. *The first iteration of Algorithm 6.1 produces* (x^1, y^1, s^1) *such that* $K - a^T y^1 < \mu^1$.

Proof. Let σ be the (unknown) analytic center of the polytope $\mathcal{S}$. The projected cost $\hat{\rho}$ scaled about σ and its norm $\hat{r}$ are given by (6.1) (ommiting the indices). For a nearly central point s^0 with $\delta(s^0) \leq 0.01$, it follows from (iii) that

$$2.3 \leq \frac{\mu^1}{\hat{r}} \leq 2.5. \tag{6.3}$$

Now consider the four points $s^0, \sigma, s(\mu^1), s^1$. We show that

$$\left\|s^0 - s^1\right\|_\sigma < 1: \tag{6.4}$$

Since $\delta(s^0) \leq 0.01$ (s^0 is very near σ), (ii) gives $\left\|s^0 - \sigma\right\|_\sigma < 0.0105$.
Since $\delta(s^1,\mu^1) \leq 0.01$ (s^1 is very near $s(\mu^1)$), (ii) gives $\left\|s^1 - s(\mu^1)\right\|_{s(\mu^1)} < 0.0105$.
Since by definition $\delta(\sigma,\mu^1) = \hat{r}/\mu_1 \leq 1/2.3 < 0.45$ (σ is near $s(\mu^1)$), (ii) gives $\left\|\sigma - s(\mu^1)\right\|_{s(\mu^1)} < 1.05/2.3$.
Let us combine these three inequalities: Using (i) and the triangle inequality on the last two inequalities

$$\left\|s^1 - \sigma\right\|_\sigma \leq 2\left\|s^1 - \sigma\right\|_{s(\mu^1)} < 0.98.$$

Using the triangle inequality on this and the first inequality we complete the proof of (6.4). The cost variations are the same in the original and scaled spaces. Hence, using (6.4),

$$a^T(y^0 - y^1) = \rho^T(s^0 - s^1) = \hat{\rho}^T\Sigma^{-1}(s^0 - s^1) \leq \hat{r}\,\left\|s^0 - s^1\right\|_\sigma \leq \hat{r}.$$

Finally with $K \leq a^T y^0$,

$$K - a^T y^1 \leq a^T(y^0 - y^1) \leq \hat{r} = \frac{\hat{r}}{\mu^1}\mu^1 \leq \frac{\mu^1}{2.3} < \mu^1,$$

completing the proof. ∎

Finally, we can use results on the variation of the barrier function along the central path derived in [13] to prove the main complexity result. Consider an iteration j of the cutting plane algorithm, and, as before, drop the indices j, assuming that Algorithm 6.1 starts from a polytope $\mathcal{S}$. Let $\bar{k}$ be the index of the last iteration. This last iteration may correspond to a very short step, due to the interpolation, and we shall ignore it below.

This is what we know about the variation in the barrier function:
First iteration: from the analytic center σ of $\mathcal{S}$ to s^1 such that $\hat{r}/\mu^1 \geq 0.4$ and $\delta(s^1, \mu^1) \leq 0.1$, (iv) ensures that

$$p(s^1) \geq p(\sigma) + 0.018.$$

Intermediate iterations: The variation of $p(\cdot)$ is given by the next lemma.

LEMMA 6.3. *Consider an iteration $1 < k < \bar{k}$ of Algorithm 6.1. Then $p(s^{k+1}) \geq p(s^k) + 0.07\delta(s^k) \geq p(s^k) + 0.014$.*

Proof. Let us simplify the notation by setting $s = s^k$, $\mu = \mu^k$, $\mu^+ = \mu^{k+1}$. We know by construction that $\delta(s, \mu) \leq 0.01$ and $\delta(s, \mu^+) = 0.1$. We shall use a very common trick in this sort of proofs: we assume without loss of generality that $s = e$. This can always be achieved by a change of variables (scaling) which transfers s to e and does not affect the barrier function variations.

From $s = e$, the Newton centering step $h(e, \mu^+)$ satisfies $\|h(e, \mu^+)\| = 0.1$, and

$$h(e, \mu^+) = -\frac{\rho}{\mu^+} + e_p, \tag{6.5}$$

where ρ is the projected cost and e_p is the projection of e into $\mathcal{R}(A^T)$. By (iv), $\delta(e) = \|e_p\| \geq 0.2$, because by (6.3) $\mu \leq 2.5\,\hat{r}$. This already proves the last inequality in the lemma. We must prove that $p(e + h(e, \mu^+)) - p(e) \geq 0.07\,\|e_p\|$.

Denote $\mu^+ = \mu - d\mu$, and define $\tau = h(e, \mu) = \rho/\mu + e_p$. From $\|\tau\| \leq 0.01$ and $\|e_p\| \geq 0.2$, we deduce

$$0.95\,\|e_p\| \leq \|\tau + e_p\| \leq 1.05\,\|e_p\|\,. \tag{6.6}$$

From (6.5),

$$\begin{aligned} \mu^+ h(e, \mu^+) &= -\rho + \mu\, e_p - d\mu\, e_p \\ &= \mu\tau - d\mu\, e_p \\ &= \mu^+\tau - d\mu(e_p + \tau). \end{aligned}$$

Thus,

$$h(e, \mu^+) = \tau - \frac{d\mu}{\mu^+}(e_p + \tau).$$

Taking norms and using (6.6), $0.1 \leq 0.01 + 1.05\,\|e_p\|\,d\mu/\mu^+$, and thus

$$\frac{d\mu}{\mu^+} \geq \frac{0.09}{1.05\,\|e_p\|}.$$

Using the left inequality in (6.6) and noting that $e_p^T(e_p + \tau) > 0$ because $\|e_p\| > \|\tau\|$,

$$\begin{aligned} -e_p^T h(e, \mu^+) &= -e_p^T(\tau - (e_p + \tau)d\mu/\mu^+) \\ &\geq \|e_p\| \left(\frac{0.09}{1.05} \frac{\|e_p + \tau\|}{e_p} - 0,01 \right) \\ &\geq 0.07\,\|e_p\|\,. \end{aligned}$$

Using the convexity of $p(\cdot)$, the variation of $p(\cdot)$ along $h(e, \mu^+)$ is given by

$$p(e + h(e, \mu^+)) - p(e) \geq \nabla p(e)^T h(e, \mu^+) = -e^T h(e, \mu^+).$$

But e_p is the projection of e into $\mathcal{R}(A^T)$ and $h(e, \mu^+) \in \mathcal{R}(A^T)$. Hence $-e^T h(e, \mu^+) = -e_p^T h(e, \mu^+) \geq 0.07 \, \|e_p\|$, completing the proof. □

We can finally prove the main result.

THEOREM 6.4. *The cutting plane method applied to the convex feasibility problem using Algorithm 6.1 terminates in $k = O^*(m^2/\varepsilon^2)$ iterations, and the total number of Newton steps is bounded by $O^*\left(\dfrac{m^2}{\varepsilon^2} \log\left(\dfrac{1}{\varepsilon}\right)\right)$.*

Proof. Consider an iteration j of the cutting plane algorithm, which starts with a polytope Ω_{j-1} and uses Algorithm 6.1 to find an approximate center of $\Omega_j = \{y \in \Omega_{j-1} \mid a_j^T y \leq c_j\}$. We must prove the inequality (6.2).

Assume that the Algorithm 6.1 stopped in $\bar{k} > 1$ iterations, and define $k = \bar{k} - 1$. We shall study the solution (y^k, s^k), ignoring the last iteration, because it may be a very short step due to the interpolation. We have by construction $a_j^T y^k + \mu^k \geq c_j$. Define $K = a_j^T y^k + \mu^k$ and $\tilde{\Omega} = \{y \in \Omega_{j-1} \mid a_j^T y \leq K\}$. Then $\Omega_j \subset \tilde{\Omega}$, and it is easy to see that $P(\Omega_j) \geq P(\tilde{\Omega})$. We shall prove that the inequality (6.2) is satisfied by $P(\tilde{\Omega})$.

Using the results on the variation of the barrier function derived above,

$$p(s^k) \geq p(\sigma) + 0.018 + 0.014(k-1),$$

where σ is the analytic center of $\mathcal{S}_{j-1}$, and hence $P(\Omega_{j-1}) = p(\sigma)$. The number of Newton steps is $k_j = k + 3$ (counting three for the first and one for the last iteration), and hence

$$p(s^k) \geq P(\Omega_{j-1}) - 0.038 + 0.014 \, k_j. \tag{6.7}$$

Now consider the polytope $\tilde{\Omega}$, and compute the barrier function $\tilde{p}(\tilde{s})$ for this polytope at the point $\tilde{s} = (s^k, K - a_j^T s^k)$ (see (3.1) for the notation). We have

$$\tilde{p}(\tilde{s}) = -\log(K - a_j^T s^k) + p(s^k) = -\log(\mu^k) + p(s^k),$$

by definition of K. Since $\mu^k \leq 2.5 \, \hat{r}_j$ by (6.3), it follows that

$$\tilde{p}(\tilde{s}) \geq P(\Omega_{j-1}) - \log(\hat{r}_j) - 0.038 - \log(2.5) + 0.014 \, k_j.$$

By Lemma 3.2, $\tilde{\delta}(\tilde{s}) \leq \delta(s^k, \mu^k) \leq 0.01$. Let $\tilde{\sigma}$ be the analytic center of $\tilde{\mathcal{S}}$, the polytope of slacks associated with $\tilde{\Omega}$. Then in $\tilde{\mathcal{S}}$, the scaled distance from $\tilde{s}$ to $\tilde{\sigma}$ satisfies $\xi = \|\tilde{s} - \tilde{\sigma}\|_{\tilde{\sigma}} \leq 0.0105$ by (ii), and by a well known property of the logarithmic barrier function near its minimizer (see for instance [11]),

$$\tilde{p}(\tilde{s}) \leq \tilde{p}(\tilde{\sigma}) + \frac{\xi^2}{2(1-\xi)} < \tilde{p}(\tilde{\sigma}) + 10^{-4}.$$

It follows that $P(\tilde{\Omega}) = \tilde{p}(\tilde{\sigma}) \geq \tilde{p}(\tilde{s}) - 10^{-4}$ and finally, with $\zeta = -\log(2.5) - 0.0381$,

$$P(\Omega_j) \geq P(\tilde{\Omega}) \geq P(\Omega_{j-1}) + \zeta - \log(\hat{r}_j) + 0.014 \, k_j.$$

This corresponds to (6.2), completing the proof. □

REFERENCES

[1] Atkinson, D.S., Vaidya, P.M. "A Cutting Plane Algorithm that uses Analytic Centers", *Mathematical Programming*, n.69, pp.1-43, 1995.

[2] Bahn, O., Goffin, J.-L., Vial, J.-P. *et al.* "Experimental Behaviour of an Interior Point Cutting Plane Algorithm for Convex Programming: An Application to Geometric Programming", *Discrete Applied Mathematics*, n.49, pp.3-23, 1994.

[3] Bahn, O., Merle, O. du, Goffin, J.-L. *et al.* "A Cutting Plane Method from Analytic Centers for Stochastic Programming", *Mathematical Programming*, n.69, pp.45-73, 1995.

[4] Elzinga, J., Moore, T. "A Central Cutting Plane Algorithm for Convex Programming", *Mathematical Programming*, n.8, pp.134-145, 1975.

[5] Goffin, J.-L., Vial, J.P. "Cutting Planes and Column Generation Techniques with the Projective Algorithm", *Journal of Optimization Theory and its Applications*, n.65, pp.409-429, 1989.

[6] Goffin, J.-L., Vial, J.P. *Shallow, Deep and very Deep Cuts in the Analytic Center Cutting Plane Method.* Logilab Technical Report 96-3, Department of Management Studies, University of Geneva, Switzerland, 1996.

[7] Goffin, J.L., Gondizio, J., Sarkisian, R. *et al.* "Solving Nonlinear Multicommodity Flows Problems by the Analytic Centers Cutting Plane Method", *Mathematical Programming*, v.76, n.1, pp.131-154, 1997.

[8] Goffin, J.L., Haurie, A., Vial, J.P. "Decomposition and Nondifferentiable Optimization with the Projective Algorithm", *Management Science*, n.38, pp.284-302, 1992.

[9] Goffin, J.L., Luo, Z.Q., Ye, Y. "Complexity Analysis of an Interior Cutting Plane for Convex Feasibility Problems", *SIAM Journal of Optimization*, n.6, pp.638-652, 1996.

[10] Gonzaga, C.C. "Interior Point Algorithms for Linear Programming with Inequality Constraints", *Mathematical Programming*, n.52, pp.209-255, 1991.

[11] Gonzaga, C.C. "Path-Following Methods for Linear Programming", *SIAM Review*, v.34, n.2, pp.167-224, 1992.

[12] Gonzaga,C.C.,"The Largest Step Path Following Algorithm for Monotone Linear Complementarity Problems", *Mathematical Programming*, n. 76, pp. 309-332, 1997.

[13] Gonzaga,C.C., Cardia,M. "Properties of the central points in linear programming problems", Technical Report, Department of Mathematics, Federal University of Santa Catarina, Brazil, 2002.

[14] Karmarkar, N. "A New Polynomial Time Algorithm for Linear Programming", *Combinatoria*, n.4, pp.373-395, 1984.

[15] Kelley, J.E. "The Cutting-Plane Method for Solving Convex Programs". *Journal of the SIAM*, n.8, 1960.

[16] Khachiyan, L.G. "A Polynomial Algorithm for Linear Programming", *Soviet. Math. Doklady*, n.20, pp.191-194, 1979.

[17] Khachiyan, L.G., Todd, M.J. "On the Complexity of Approximating the Maximal Inscribed Ellipsoid for a Polytope", *Mathematical Programming*, n.61, pp.137-160, 1993.

[18] Levin, A. "An Algorithm of Minimization of Convex Functions". *Soviet. Math. Doklady*, v.160, n.6, pp.1244-1247, 1965.

[19] Mehrotra, S.,"On the Implementation of a Primal-Dual Interior Point Method", SIAM *Journal on Optimization*, n. 2, 1992.

[20] Mitchell, J.E., Todd, M.J. "Solving Combinatorial Optimization Problems using Karmarkar's Algorithms", *Mathematical Programming*, n.56, pp.245-284, 1992.

[21] Mizuno, S. "A New Polynomial Time Method for a Linear Complementarity Problem". *Mathematical Programming*, n.56, 1992.

[22] Mizuno, S., Todd, M.J., Ye, Y., "On Adaptive-Step Primal-Dual Interior-Point Algorithms for Linear Programming",*Mathematics of Operations Research*, vol. 18, n.4, 1993.

[23] Mokhtarian, F.S., Goffin, J.L. *Using the Primal-Dual Infeasible Newton Method in the Analytic Centers Method for Problems Defined by Deep Cutting Planes.* Cahier du GERAD G 94-41, ISSN: 0711-2440, 1994.

[24] Nesterov, Y. "Cutting Plane Algorithms from Analytic Centers: Efficiency Estimates", *Mathematical Programming*, n.69, pp.149-176, 1995.

[25] Nesterov, Y., Péton, O., Vial, J.P.,*Homogeneous Analytic Center Cutting Plane Methods with Approximate Centers.* Logilab Technical Report 98, Department of Management Studies, University of Geneva, Switzerland,1998.

[26] Nesterov, Y., Vial, J.P. *Homogeneous Analytic Center Cutting Plane Methods for Convex Problems and Variational Inequalities.* Logilab Technical Report 97-4, Department of Management Studies, University of Geneva, Switzerland, 1997.

[27] Roos,C., Terlaky, T., Vial, J.P. *Theory and Algorithms for Linear Optimization:An Interior*

Point Approach,Chichester, John Willey and Sons, 1997.

[28] Shor, N. "A Cutting Plane Method for Solving Convex Programming Problems" *Cybernetics*, n.1, pp.42-50, 1977.

[29] Sonnevend, G. "An Analytical Centre for Polyhedron and New Class of Global Algorithms for Linear (smooth convex) Programming". In: *Lectures Notes in Control and Information Sciences*, n.84, New York, Springer Verlag, 1985.

[30] Tarasov, S., Khachiyan, L.G., Erlich, I. "The Method of Inscribed Ellipsoids", *Soviet. Math. Doklady*, n.37, 1988.

[31] Vaidya, P.M. "A New Algorithm for Minimizing a Convex Function over Convex Sets". In: *Proceedings of the 30th Annual Symposium of Foundations if Computer Science*, Research Triangle Park, NC, USA, pp.338-343, IEEE Computer Society Press, Los Alamitos, CA, 1990.

[32] S. Wright, *Primal-Dual Interior-Point Methods*, Philadelphia, SIAM Publications, 1997.

[33] Ye, Y. "A Potential Reduction Algorithm allowing Column Generation", *SIAM Journal of Optimization*, n.2, pp.7-20, 1992.

[34] Yudin, D., Nemirovsky, A. "Informational Complexity and Efficient Methods for Solving Convex Extremal Problems", *Ekonomika i matem. metody*, v.12, n.2, pp.357-369, 1976.