Complexity Analysis of an Interior Cutting Plane Method for Convex Feasibility Problems [1]

Jean-Louis GOFFIN
Faculty of Management, McGill University, 1001 Sherbrooke W, Montreal, Quebec, H3A 1G5, Canada
Zhi-Quan LUO
Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, L8S 4L7, Canada
and
Yinyu YE
Department of Management Sciences, The University of Iowa, Iowa City, Iowa 52242, USA.
June, 1994

Abstract

We further analyze the convergence and the complexity of a dual column generation algorithm for solving general convex feasibility problems defined by a separation oracle. The oracle is called at an *approximate* analytic center of the set given by the intersection of the linear inequalities which are the previous answers of the oracle. We show that the algorithm converges in finite time and is in fact a fully polynomial approximation algorithm, provided that the feasible region has an nonempty interior.

Keywords: Convex feasibility problem, potential reduction, column generation, cutting planes.

AMS subject classification: 90C25, 90C26, 90C60.

[1]The research of the first author is supported by the Natural Sciences and Engineering Research Council of Canada, grant number OPG0004152 and by the FCAR of Quebec; the research of the second author is supported by the Natural Sciences and Engineering Research Council of Canada, grant number OPG0090391; the research of the third author is supported by NSF grant DDM-9207347, an Iowa Business School Summer Grant and K. C. Wong Educational Foundation, Hong Kong, through Academia Sinica, Beijing, China.

1 Introduction

The problem studied is that of finding an interior point in a convex set, where we assume $\Gamma \subset \mathbb{R}^m$ has a nonempty interior and is contained in the cube $\Omega^0 = \{y \in \mathbb{R}^m : 0 \leq y \leq e\} = [0,1]^m$.

The set Γ is defined implicitly by a **separating oracle** which for every $\bar{y} \in \Omega^0$ either answers that $\bar{y}$ belongs to $\mathrm{int}(\Gamma)$ or generates a separating hyperplane $\{y \in \mathbb{R}^m : a^T y \leq a^T \bar{y}\} \supset \Gamma$. Without loss of generality, we assume that a is normalized so that $\|a\| = 1$.

The problem of finding an interior point to a convex set defined by a system (finite or infinite) of convex inequalities

$$\Gamma = \{y \in \mathbb{R}^m : f_i(y) \leq 0, \quad i = 1, 2, ...\},$$

where each $f_i : \mathbb{R}^m \mapsto \mathbb{R}$ is convex, can be cast in this manner. In particular, the separating oracle just need to select a to be $g_i/\|g_i\|$, where g_i is an arbitrary subgradient of any of the functions f_i satisfying $f_i(\bar{y}) \geq 0$, i.e. $g_i \in \partial f_i(\bar{y})$ (the subdifferential of f_i). Note that for any $g_i \in \partial f_i(\bar{y})$, $f_i(y) \leq f_i(\bar{y})$ implies $g_i^T(y - \bar{y}) \leq 0$. Thus, if $f_i(\bar{y}) \geq 0$, $f_i(y) \leq 0$ then $g_i^T(y - \bar{y}) \leq 0$. In fact, the requirement that $f_i(\bar{y})$ be computed exactly and that $g_i \in \partial f_i(\bar{y})$ can be significantly relaxed. All that is required from the oracle is that the following query be answered: find any i such that $f_i(\bar{y}) \geq 0$, and a such that $a^T(y - \bar{y}) \leq 0$ for all $y \in \{y \in \mathbb{R}^m : f_i(y) \leq 0\}$. This also works if the functions f_i are pseudoconvex.

The problem under investigation may be cast as that of finding the solution to an infinite system of linear inequalities, which is defined implicitly by the oracle. The classical centering methods that have been suggested for the above convex feasibility problem include the center of gravity method of Levin [12], the largest sphere method of Elzinga and Moore [5], the ellipsoid method of Yudin and Nemirovskii [16, 10, 18], the maximum volume ellipsoid method of Tarasov, Khacian and Erlich [20, 11] and the method of volumetric centers of Vaidya [21], among others.

Sonnevend [19], Goffin, Haurie and Vial [7] and Ye [22] proposed a column generation – or cutting plane – algorithm that computes $\bar{y}$ as the analytic center of the system of inequalities generated so far. In [22], Ye showed that this algorithm is a potential-reduction algorithm for solving the linear feasibility problem where $\Gamma = \{y \in \mathbb{R}^m : A^T y \leq c\}$ for some $A \in \mathbb{R}^{m \times d}$ and $c \in \mathbb{R}^d$. If d is finite, the algorithm finds a feasible solution in polynomial-time of n and d. When d is infinite, this algorithm was developed and tested for a variety of large scale problems, where they performed quite well (see, for example, Burke et al [2], Goffin, Haurie and Vial [7] and Mitchell [13][14]); but proofs of convergence and of complexity remained at large.

Recently, Atkinson and Vaidya [1] developed a modified version of the algorithm that uses a column elimination scheme and proved a polynomial complexity result [1]; Nesterov [17] proved a complexity result for the minimization of a convex function using a variant of the algorithm which uses a potential function consisted of the logarithm of a quadratic term representing the initial sphere localization.

In this paper, we show that for any given convex feasibility problem with nonempty interior, the algorithm of [7, 22] is a fully polynomial approximation scheme for the original version of the algorithm that uses only linear inequalities to approximate the solution set and does not use column elimination. A fully polynomial approximation scheme means that for every ϵ, the accuracy at termination, the method is polynomial in the dimension of the problem, m.

2 Preliminaries

Before we use approximate analytic centers, we need to develop some technical results to be used in the subsequent analysis. We start with a measure of proximity to the analytic center. Again, let Ω be a (bounded) polytope in $\mathbb{R}^m$ defined by n $(> m)$ linear inequalities, i.e.,

$$\Omega = \{y \in \mathbb{R}^m : c - A^T y \geq 0\}.$$

Recall that for a point y in the interior of Ω, the potential function

$$\phi(y) := \sum_{j=1}^{n} \ln s_j$$

where

$$s = c - A^T y,$$

and the max-potential of Ω is defined as the value of ϕ at the analytic center $\bar{y}$:

$$P(\Omega) := \max_{y \in \Omega} \phi(y).$$

Consider a normalized gradient vector of ϕ at y

$$\Delta s(y) = S^{-1}A^T(AS^{-2}A^T)^{-1}\nabla\phi(y) = -S^{-1}A^T(AS^{-2}A^T)^{-1}AS^{-1}e$$

and

$$\delta(y) = \|\Delta s(y)\| = \|S^{-1}A^T(AS^{-2}A^T)^{-1}AS^{-1}e\|,$$

or

$$\delta(y)^2 = e^T S^{-1}A^T(AS^{-2}A^T)^{-1}AS^{-1}e.$$

Note that $AS^{-2}A^T$ is symmetric positive definite, where S (or X) represents the diagonal matrix of vector s (or x). Thus, this quantity $\delta(y)$ is a norm of the gradient vector. Let

$$x(y) = S^{-1}(I - S^{-1}A^T(AS^{-2}A^T)^{-1}AS^{-1})e$$

we have

$$\Delta s(y) = Sx(y) - e \quad \text{and} \quad \delta(y) = \|Sx(y) - e\|.$$

Another measure would explicitly use the primal variable $x \geq 0$ and $Ax = 0$, the Lagrangean multiplier for Ω, where Ω is usually called the dual system. For a primal feasible x, a residual is defined as

$$\eta(x, y) = \|X(c - A^T y) - e\| = \|Xs - e\|.$$

We now establish some interesting inequalities relating $\delta(y)$ to $\eta(x, y)$. Some of these inequalities will be used in the subsequent analysis. Most results of this lemma were proved in Gonzaga [8] or den Hertog et al [4].

Lemma 2.1 *Let (y, s) be an interior point in Ω.*
i) If there is $x > 0$ with $Ax = 0$ such that $\eta(x, y) < 1$ then

$$\delta(y) \leq \eta(x, y);$$

Conversely, if $\delta(y) < 1$ then there is $x > 0$ with $Ax = 0$ such that

$$\eta(x, y) = \delta(y).$$

ii) Let $\Delta(y) = (AS^{-2}A^T)^{-1}\nabla\phi(y)$ and let $y^+ = y + \Delta(y)$. Then if $\delta(y) < 1$,

$$\delta(y^+) \leq \delta(y)^2.$$

iii) Let $\bar{y}$ be the analytic center of Ω. Then, if $\delta(y) < 1$,

$$\phi(\bar{y}) - \phi(y) \leq \frac{\delta(y)^2}{1 - \delta(y)^2}$$

or

$$P(\Omega) - \phi(y) \leq \frac{\delta(y)^2}{1 - \delta(y)^2}.$$

iv) Let $\bar{s} = c - A^T\bar{y}$ where $\bar{y}$ is the analytic center of Ω. Then, if $\delta(y) < 1$,

$$\|S^{-1}\bar{s} - e\| \leq \frac{\delta(y)}{1 - \delta(y)}.$$

Proof.
i) Given (y, s) in the interior of Ω, we can verify that $\delta(y)$ is the minimal value and $x(y)$ is the minimal solution of the least squares problem

$$\begin{array}{ll} \text{minimize} & \|Sx - e\| \\ \text{subject to} & Ax = 0. \end{array}$$

Since x in $\eta(x, y)$ is a feasible point for this problem, we must have

$$\delta(y) \leq \eta(x, y).$$

Conversely, let $x = x(y)$. Then, we have $Ax = 0$ and $\eta(x, y) = \delta(y)$. Moreover, $x > 0$ since $\delta(y) < 1$.

ii) Define $s^+ = s - S\Delta s(y)$ and note that from the proof of (i)

$$\delta(y^+) = \|S^+ x(y^+) - e\| \le \|S^+ x(y) - e\|.$$

But

$$\begin{aligned}
\|S^+ x(y) - e\|^2 &= \|(2S - S^2 X(y))x(y) - e\|^2 \\
&= \sum_{j=1}^{n} (s_j x(y)_j - 1)^4 \\
&\le (\sum_{j=1}^{n} (s_j x(y)_j - 1)^2)^2 \\
&= \|Sx(y) - e\|^4 \\
&= \delta(y)^4.
\end{aligned}$$

iii) Let y^+ be generated in (ii). Since ϕ is a concave function

$$\phi(y^+) - \phi(y) \le \nabla\phi(y)(y^+ - y) = \delta(y)^2.$$

Consider the sequence generated by this process with $y^0 = y$ and $y^{k+1} = y^k + \Delta(y^k)$, then from (ii), $y^k \to \bar{y}$ and from the above relation

$$\phi(y^{k+1}) - \phi(y^k) \le \delta(y^k)^2.$$

Thus, for any $k \ge 1$ we have

$$\phi(y^k) - \phi(y^0) \le \sum_{j=1}^{k} \delta(y)^{2^j} \le \sum_{j=1}^{k} (\delta(y)^2)^j \le \frac{\delta(y)^2}{1 - \delta(y)^2}.$$

iv) Let $y^0 = y$ and the sequence $\{y^k, s^k\}$ be generated by the process in (iii). Then, we show by induction that

$$\|(S^0)^{-1} s^k - e\| \le \sum_{j=1}^{2^k - 1} \delta(y)^j.$$

Obviously, this relation is true for $k = 1$. Now assuming it is true for k. Note that

$$s^{k+1} = s^k - S^k(S^k x(y^k) - e),$$

we have

$$\begin{aligned}
\|(S^0)^{-1}(s^{k+1} - s^k)\| &= \|(S^0)^{-1} S^k (S^k x(y^k) - e)\| \\
&\le \|(S^0)^{-1} S^k\| \|S^k x(y^k) - e\| \\
&\le \left(\sum_{j=0}^{2^k - 1} \delta(y)^j \right) \|S^k x(y^k) - e\| \\
&\le \left(\sum_{j=0}^{2^k - 1} \delta(y)^j \right) \delta(y)^{2^k}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\|(S^0)^{-1}s^{k+1} - e)\| &\leq \|(S^0)^{-1}s^k - e)\| + \|(S^0)^{-1}(s^{k+1} - s^k)\| \\
&\leq \sum_{j=1}^{2^k-1} \delta(y)^j + \left(\sum_{j=0}^{2^k-1} \delta(y)^j\right) \delta(y)^{2^k} \\
&= \sum_{j=1}^{2^{k+1}-1} \delta(y)^j.
\end{aligned}$$

This proves (iv). **Q.E.D.**

3 Newton Procedures for Computing Centers

We review two Newton's procedures to compute the analytic center. Either of them can be used in our cutting plane algorithm.

3.1 Dual Newton Procedure

The dual Newton iteration for finding the analytic center is as follows: given an interior point $y \in \mathbb{R}^m$, slacks $s = c - A^T y$ and a primal feasible $x = x(y)$ satisfying

$$\|Xs - e\| \leq \eta < 1, \tag{3.1}$$

we can apply the dual Newton method to generate a new iterate

$$y^+ = y + \Delta(y),$$

where $\Delta(y)$ is given by part (ii) of Lemma 2.1. The residual at (x^+, y^+, s^+), where $s^+ = c - A^T y^+$ and $x^+ = x(y^+)$ satisfies the following lemma.

Lemma 3.1 *Let (x^+, y^+, s^+) be generated as above. Then,*

$$\|X^+ s^+ - e\| \leq \eta^2.$$

3.2 Primal-dual Newton Procedure

Given an interior point $y \in \mathbb{R}^m$, $s = c - A^T y$, and a primal feasible $x \in \mathbb{R}^n$ satisfying

$$\|Xs - e\| \leq \eta < 1, \tag{3.2}$$

we can apply the primal-dual Newton method to generate a new iterate x^+ and (y^+, s^+) as follows.

Solve for d_x, d_y and d_s from the system of linear equations:

$$\begin{array}{rcl} Sd_x + Xd_s & = & \frac{x^T s}{n} e - Xs, \\ Ad_x & = & 0, \\ A^T d_y + d_s & = & 0. \end{array} \tag{3.3}$$

Then, let

$$x^+ = x + d_x, \quad y^+ = y + d_y, \quad s^+ = s + d_s.$$

We have a lemma for the residual at (x^+, y^+, s^+).

Lemma 3.2 *Let (x^+, y^+, s^+) be generated as above. Then,*

$$\|X^+ s^+ - e\| \leq \frac{\sqrt{2}\eta^2}{4(1-\eta)}.$$

Proof. See Mizuno [15].

4 Max-Potential Reduction

Now, we use an approximate center y^k to generate a cut, where (y^k, s^k) is an interior point in Ω and there is $x^k > 0$ with $Ax^k = 0$ such that

$$\|X^k s^k - e\| \leq \eta \tag{4.1}$$

for some $0 < \eta < 1$. Let us place a cut exactly at y^k, that is, we add a new inequality $a^T y \leq a^T y^k$ to Ω, and consider the new set

$$\Omega^+ = \{y : A^T y \leq c, \quad a^T y \leq a^T y^k\}.$$

We remark that our definition of Ω^+ is different from that of [22] which requires the new inequality to be $a^T y \leq a^T y^k + \beta r$, for some $0 < \beta < 1$, where

$$r = \sqrt{a^T (A(S^k)^{-2} A^T)^{-1} a}.$$

Thus, "more" is cut away from Ω in the current definition of Ω^+. This deep cut was first developed by Nesterov in his cutting plane method [17].

We now prove a lemma resembling Theorem 2 of Ye [22].

Lemma 4.1 *Let $(\bar{y}, \bar{s})$ be the analytic center of Ω and let*

$$\bar{r} = \sqrt{a^T (A\bar{S}^{-2} A^T)^{-1} a}.$$

Then

$$P(\Omega^+) \leq P(\Omega) + \ln(\bar{r}) - \alpha$$

for some constant α depending only on η. Moreover, if $0 < \eta < 1/100$, then we have $\alpha > 0$.

Proof. Let $\bar{y}$ and $\bar{y}^+$ be the analytic centers for Ω and Ω^+, respectively. Let $\bar{s}^+ = c - A^T\bar{y}^+ > 0$ and $\bar{s}^+_{n+1} = a^Ty^k - a^T\bar{y}^+$. Then we have

$$\begin{aligned}
\bar{s}^+_{n+1} &= a^T(y^k - \bar{y}^+) \\
&= a^T(A\bar{S}^{-2}A^T)^{-1}(A\bar{S}^{-2}A^T)(y^k - \bar{y}^+) \\
&= a^T(A\bar{S}^{-2}A^T)^{-1}A\bar{S}^{-2}(A^Ty^k - A^T\bar{y}^+) \\
&= a^T(A\bar{S}^{-2}A^T)^{-1}A\bar{S}^{-2}(-c + A^Ty^k + c - A^T\bar{y}^+) \\
&= a^T(A\bar{S}^{-2}A^T)^{-1}A\bar{S}^{-2}(\bar{s}^+ - s^k) \\
&= a^T(A\bar{S}^{-2}A^T)^{-1}A\bar{S}^{-1}(\bar{S}^{-1}\bar{s}^+ - \bar{S}^{-1}s^k) \\
&\le \|a^T(A\bar{S}^{-2}A^T)^{-1}A\bar{S}^{-1}\|\|\bar{S}^{-1}\bar{s}^+ - \bar{S}^{-1}s^k\| \\
&= \bar{r}\|\bar{S}^{-1}\bar{s}^+ - e + e - \bar{S}^{-1}s^k\| \\
&\le \bar{r}(\|\bar{S}^{-1}\bar{s}^+ - e\| + \|e - \bar{S}^{-1}s^k\|) \\
&\le \bar{r}(\|\bar{S}^{-1}\bar{s}^+ - e\| + \frac{\eta}{(1-\eta)^2}) \quad \text{(from Lemma 2.1).}
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\exp P(\Omega^+)}{\bar{r}\exp P(\Omega)} &= \frac{\bar{s}^+_{n+1}}{\bar{r}}\prod_{j=1}^n \frac{\bar{s}^+_j}{\bar{s}_j} \\
&\le (\|\bar{S}^{-1}\bar{s}^+ - e\| + \frac{\eta}{(1-\eta)^2})\prod_{j=1}^n \frac{\bar{s}^+_j}{\bar{s}_j}.
\end{aligned}$$

Note that we still have as in [22]

$$e^T\bar{S}^{-1}\bar{s}^+ = n.$$

This can be seen as follows:

$$\begin{aligned}
e^T\bar{S}^{-1}\bar{s}^+ &= e^T\bar{X}\bar{s}^+ \\
&= e^T\bar{X}(c - A^T\bar{y}^+) \\
&= e^T\bar{X}c \\
&= e^T\bar{X}(c - A^T\bar{y}) \\
&= e^T\bar{X}\bar{s} \\
&= e^Te \\
&= n,
\end{aligned}$$

where the first and sixth equalities are due to $\bar{X}\bar{s} = e$, and the second and fourth equalities follow from $A\bar{x} = 0$.

Thus, from the proof of Theorem 2 of [22]

$$(\|\bar{S}^{-1}\bar{s}^+ - e\| + \frac{\eta}{(1-\eta)^2})\prod_{j=1}^n \frac{\bar{s}^+_j}{\bar{s}_j} \le 4\exp(\frac{8\eta - 3}{2(1-\eta)^2}).$$

Thus,

$$P(\Omega^+) - P(\Omega) \le \ln(\bar{r}) + \ln(4) + \frac{8\eta - 3}{2(1-\eta)^2}.$$

Let

$$\alpha = -\ln(4) - \frac{8\eta - 3}{2(1-\eta)^2}.$$

Let $\eta = 1/100$. Then we have $\alpha > 0$ and the desired result. **Q.E.D.**

5 Update to a New Approximate Center

In this section, we show how to construct a pair (x, y, s) from (x^k, y^k, s^k) such that $(A, a)x = 0$ with $x > 0$, (y, s) is in the interior of Ω^+, and

$$\|Xs - e\| < \eta < 1.$$

Suppose a pair (x^k, y^k, s^k) is given which satisfies (4.1), we use the **dual** scaling for the construction of (x, y, s). Let

$$\begin{aligned}
r^k &= \sqrt{a^T(A(S^k)^{-2}A^T)^{-1}a}, && (5.1)\\
\Delta y &= -(\beta/r^k)(A(S^k)^{-2}A^T)^{-1}a, && (5.2)\\
\Delta s &= (\beta/r^k)A^T(A(S^k)^{-2}A^T)^{-1}a, && (5.3)\\
\Delta x &= -(\beta/r^k)(S^k)^{-2}A^T(A(S^k)^{-2}A^T)^{-1}a. && (5.4)
\end{aligned}$$

Then we set

$$y = y^k + \Delta y$$

and

$$x = \begin{bmatrix} x^k + \Delta x \\ \beta/r^k \end{bmatrix} = \begin{bmatrix} x^k - (\beta/r^k)(S^k)^{-2}A^T(A(S^k)^{-2}A^T)^{-1}a \\ \beta/r^k \end{bmatrix}.$$

It can be readily verified that

$$s = \begin{bmatrix} c - A^T(y^k + \Delta y) \\ a^T y^k - a^T(y^k + \Delta y) \end{bmatrix} = \begin{bmatrix} s^k + \Delta s \\ \beta r^k \end{bmatrix} = \begin{bmatrix} s^k + (\beta/r^k)A^T(A(S^k)^{-2}A^T)^{-1}a \\ \beta r^k \end{bmatrix}$$

First, we have

$$(A, a)x = Ax^k - (\beta/r^k)a + (\beta/r^k)a = 0.$$

Second, we have

$$s^k + (\beta/r^k)A^T(A(S^k)^{-2}A^T)^{-1}a = (S^k)(e + p^k)$$

and

$$x^k - (\beta/r^k)(S^k)^{-2}A^T(A(S^k)^{-2}A^T)^{-1}a = (S^k)^{-1}(X^kS^ke - p^k),$$

where

$$p^k = (\beta/r^k)(S^k)^{-1}A^T(A(S^k)^{-2}A^T)^{-1}a.$$

Note that we have

$$\|p^k\| = \beta. \tag{5.5}$$

On the other hand, we have

$$x_j^k s_j^k \geq 1 - \eta.$$

Thus, if we select η and β such that

$$1 - \eta - \beta > 0, \tag{5.6}$$

then both

$$s = \begin{bmatrix} (S^k)(e + p^k) \\ \beta r^k \end{bmatrix} > 0$$

and

$$x = \begin{bmatrix} (S^k)^{-1}(X^k S^k e - p^k) \\ \beta/r^k \end{bmatrix} > 0.$$

A simple calculation yields

$$Xs - e = \begin{bmatrix} X^k s^k - e \\ 0 \end{bmatrix} - \begin{bmatrix} (p^k)^2 \\ 1 - \beta^2 \end{bmatrix} - \begin{bmatrix} (X^k S^k - I)p^k \\ 0 \end{bmatrix}$$

where the vector

$$(p^k)^2 = ((p_1^k)^2, (p_2^k)^2, ...(p_n^k)^2)^T.$$

Therefore, we have

$$\begin{aligned} \|Xs - e\| &\leq \|X^k s^k - e\| + \sqrt{\|(p^k)^2\|^2 + (1 - \beta^2)^2} + \|X^k s^k - e\|\|p^k\| \\ &\leq \eta + \sqrt{\beta^4 + (1 - \beta^2)^2} + \eta\beta \end{aligned}$$

where the last step follows from (4.1) and (5.5). Let $\beta = 1/\sqrt{2}$ and $\eta = 0.15$, then

$$\|Xs - e\| \leq \gamma = 0.15 + 1/\sqrt{2} + 0.15/\sqrt{2} < 1.$$

Or, we can let

$$\eta = 1/100, \quad \text{and have} \quad \gamma := 1/100 + 1.01/\sqrt{2} < 1.$$

Furthermore, it can also be easily verified that (5.6) holds.

Hence, using this (y, s) as a starting pair, we can apply the **dual** Newton procedure of Section 3.1 to generate a pair (y^{k+1}, s^{k+1}) and $x^{k+1} = x(y^{k+1})$ such that

$$\begin{aligned} (A, a)x^{k+1} &= 0, \quad x^{k+1} > 0, \\ s^{k+1} &= (c^T, a^T y^k)^T - (A, a)^T y^{k+1} > 0 \end{aligned}$$

and

$$\|X^{k+1}s^{k+1} - e\| \leq \eta.$$

By Lemma 3.2 (part (ii)) and the above given values of γ and η, this can be accomplished in 4 **dual** Newton steps due to the fact $\gamma^{16} \leq \eta$. This column generation process can be repeated, and from Lemma 4.1 the nested sequence of polyhedral sets Ω^k generated by the algorithm satisfies

$$P(\Omega^{k+1}) \leq P(\Omega^k) + \ln(\bar{r}^k) - \alpha \tag{5.7}$$

where α is some constant,

$$\bar{r}^k = \sqrt{a_{k+1}^T(A(\bar{S}^k)^{-2}A^T)^{-1}a_{k+1}}$$

$(\bar{y}^k, \bar{s}^k)$ is the analytic center of Ω^k, and a_{k+1} is the cut generated at the kth iteration. Note that $(\bar{y}^k, \bar{s}^k)$ is solely used for analysis, and the algorithm does not need any knowledge of $(\bar{y}^k, \bar{s}^k)$.

We remark that, instead of using the **dual** Newton iterations of Section 3.1, one can use the **primal-dual** Newton iterations prescribed in Section 3.2 to obtain an approximate analytic center. Such updating procedure works with primal as well as dual variables.

Finally, the updating step (5.1)–(5.4) uses dual scaling. Alternatively, we could use the **primal** scaling or the **primal-dual** scaling. Specifically, in the **primal** scaling we use $A(X^k)^2A^T$ to replace $A(S^k)^{-2}A^T$ in (5.1)–(5.4), while in the **primal-dual** scaling we use $A(S^k)^{-1}X^kA^T$ to replace $A(S^k)^{-2}A^T$. The proofs and the conditions on η and β are similar; in particular, they both require 4 Newton iterations to recenter.

6 The Cutting Plane Method from Analytic Centers: Convergence and Complexity

The analysis of Goffin et al. [6] relies on two things: (i) a bound identical to (5.7); (ii) the assumption that the original feasibility problem has an interior point. Using the analysis above, we show that the technical results of [6] hold without much change for the algorithm using approximate centers. Thus, we derive the identical iteration complexity for the algorithm, and each iteration uses a constant number of primal-dual Newton steps. We extend here the proof of [6] for completeness.

Let the solution set Γ be contained in $\Omega^0 = \{y \in \mathbb{R}^m : 0 \leq y \leq e\}$, and $\text{int}(\Gamma)$ contain a full dimensional closed ball with $\epsilon < \frac{1}{2}$ radius. We also assume that there exists an oracle which for every $\bar{y} \in \Omega^0$ either returns that $\bar{y} \in \text{int}(\Gamma)$ or generates a separating hyperplane $\{y : a^Ty \leq a^T\bar{y}\} \supset \Gamma$, with $\|a\| = 1$ being assumed.

The interior-point column generation – or cutting plane – algorithm [19, 7, 22] from approximate analytic centers is as follows:

- **Step 0**

  Let

$$A^0 = (I, -I) \in \mathbb{R}^{m \times 2m},$$

$$c^0 = \begin{pmatrix} e \\ 0 \end{pmatrix} \in \mathbb{R}^{2m},$$

$$y^0 = \frac{1}{2} e \in \mathbb{R}^m,$$

$$s^0 = c^0 - (A^0)^T y^0 = \frac{1}{2} e \in \mathbb{R}^{2m},$$

$$x^0 = 2e \in \mathbb{R}^{2m}$$

  and

$$k = 1.$$

- **Step 1**

  Compute (y^k, s^k, x^k) with y^k an η-approximate analytic center of $\Omega^k = \{y \in \mathbb{R}^m : c^k - (A^k)^T y \geq 0\}$, using either the primal-dual procedure or the dual procedure of Section 3.2 with the updating scheme of section 5, with the appropriate values of η, starting at $(y^{k-1}, s^{k-1}, x^{k-1})$, an η-approximate "primal-dual" center of Ω^{k-1}.

  Query the oracle to see if $y^k \in \mathrm{int}(\Gamma)$ or not.

  If yes, stop; otherwise generate a hyperplane $\{y : a_{k+1}^T y \leq a_{k+1}^T y^k\} \supset \Gamma$ with $\|a_{k+1}\| = 1$, and let

$$\Omega^{k+1} = \{y \in \mathbb{R}^m : c^{k+1} - (A^{k+1})^T y \geq 0\}$$

  where

$$A^{k+1} = (A^k, a_{k+1})$$

  and

$$c^{k+1} = \begin{pmatrix} c^k \\ a_{k+1}^T y^k \end{pmatrix}$$

- **Step 2** Set $k := k + 1$ and return to Step 1.

Let the potential function computed at the **exact** analytic center $\bar{y}^k$ be:

$$P(\Omega^k) = \sum_{j=1}^{2m+k} \ln(c^k - (A^k)^T \bar{y}^k)_j.$$

Clearly the following relations hold, provided that termination has not occurred:

1.
$$\Gamma \subset \Omega^k \quad \forall k \tag{6.1}$$

2.
$$P(\Omega^{k+1}) \leq P(\Omega^k) + \frac{1}{2}\ln(\bar{r}^k)^2 - \alpha \quad \text{(by Lemma 4.1)} \tag{6.2}$$

where
$$(\bar{r}^k)^2 = a_{k+1}^T (A^k (\bar{S}^k)^{-2} (A^k)^T)^{-1} a_{k+1}$$
and
$$\bar{s}^k = c^k - (A^k)^T \bar{y}^k, \; \bar{S}^k = \text{ diag}(\bar{s}^k).$$

Lemma 6.1 *For all $k \geq 0$,*
$$P(\Omega^k) \geq (2m+k)\ln \epsilon.$$

Proof.

From (6.1), $\Gamma \subset \Omega^k$. Thus Ω^k contains a full dimensional ball with ϵ radius. Let the center of this ball be $\bar{y}$.

Then $c^k - (A^k)^T \bar{y} \geq \epsilon e$; thus
$$P(\Omega^k) = \sum_{j=1}^{2m+k} \ln(c^k - (A^k)^T \bar{y}^k)_j \geq \sum_{j=1}^{2m+k} \ln(c^k - (A^k)^T \bar{y})_j \geq \sum_{j=1}^{2m+k} \ln \epsilon$$
where $\bar{y}^k$ denotes the analytic center of Ω^k. **Q.E.D.**

Lemma 6.2 *Let $s = c^k - (A^k)^T y$ for any $y \in \Omega^k$. Then*

1. $0 \leq s_j \leq 1, \quad j = 1, \ldots, 2m$

2. $0 \leq s_j \leq \sqrt{m}, \quad j = 2m+1, \ldots, 2m+k.$

Proof. The proof given here is the same as [6]; we include it here for completeness.

For $j = 1, \ldots, m$, $s_j = 1 - y_j$; since $0 \leq y_j \leq 1$, $0 \leq s_j \leq 1$.
For $j = m+1, \ldots, 2m$, $s_j = y_{j-m}$; since $0 \leq y_{j-m} \leq 1$, $0 \leq s_j \leq 1$.
For $j = 2m+1, \ldots, 2m+k$,
$$s_j = a_{j-2m}^T y^{j-2m} - a_{j-2m}^T y \leq \|a_{j-2m}\| \, \|y^{j-2m} - y\| = \|y^{j-2m} - y\| \leq \sqrt{m}$$
The last inequality is due to the fact that $0 \leq y^{j-2m} \leq e$ and $0 \leq y \leq e$ or $y^{j-2m} \in \Omega^0$ and $y \in \Omega^0$. **Q.E.D.**

Lemma 6.1 indicates that in order to prove finite convergence, one needs to show that $P(\Omega^k)$ grows more slowly than $2m+k$. By Lemma 4.1, this means finding upper bounds on $\bar{r}^k$. In the following Lemma this is achieved by using a construction due to Nesterov [17] which bounds

$$A^k(\bar{S}^k)^{-2}(A^k)^T,$$

from below by using a certain matrix B^k which is simple enough to handle.

Lemma 6.3 *Let* $s = c^k - (A^k)^T y$ *for any* $y \in \Omega^k$ *and* $B^0 = 8I$, $B^{k+1} = B^k + \frac{1}{m} a_{k+1} a_{k+1}^T$. *Then*

$$A^k(S^k)^{-2}(A^k)^T \succeq B^k;$$

that is,

$$A^k(S^k)^{-2}(A^k)^T - B^k$$

is positive definite.

Proof. Let $Y^k = \mathrm{diag}(y^k)$. Then

$$\begin{aligned} A^k(S^k)^{-2}(A^k)^T &= (Y^k)^{-2} + (I - Y^k)^{-2} + \sum_{j=1}^{k} \frac{a_j a_j^T}{(s_{2m+j}^k)^2} \\ &\succeq (Y^k)^{-2} + (I - Y^k)^{-2} + \frac{1}{m} \sum_{j=1}^{k} a_j a_j^T \quad \text{(by Lemma 6.2)} \\ &\succeq 8I + \frac{1}{m} \sum_{j=1}^{k} a_j a_j^T \quad (\text{as } 0 \le y^k \le e) \\ &= B^k \end{aligned}$$

Q.E.D.

Lemma 6.4 *Let* $\bar{s}^k = c^k - (A^k)^T \bar{y}^k$ *be the slacks at the analytic center* $\bar{y}^k$ *of* Ω^k *and* $(\omega^k)^2 = a_{k+1}^T (B^k)^{-1} a_{k+1}$, *then*

$$(\omega^k)^2 \ge a_{k+1}^T (A^k(\bar{S}^k)^{-2}(A^k)^T)^{-1} a_{k+1} = (\bar{r}^k)^2. \tag{6.3}$$

Proof. Using Lemma 6.3

$$(\omega^k)^2 = a_{k+1}^T (B^k)^{-1} a_{k+1} \ge a_{k+1}^T (A^k(\bar{S}^k)^{-2}(A^k)^T)^{-1} a_{k+1} = (\bar{r}^k)^2. \tag{6.4}$$

Q.E.D.

This implies that upper bounds on the series of $(\omega^k)^2$ will lead to upper bounds on the series $(\bar{r}^k)^2$.

The following lemma adapts a result of Nesterov [17] to the technical hypotheses of this paper.

Lemma 6.5

$$\sum_{j=0}^{k}(\omega^j)^2 \leq 2m^2 \ln(1+\frac{k+1}{8m^2}).$$

Proof.

Note that

$$\det B^{k+1} = \det(B^k + \frac{1}{m}a_{k+1}a_{k+1}^T) = (1+\frac{(\omega^k)^2}{m})\det B^k.$$

Thus

$$\ln\det B^{k+1} = \ln\det B^k + \ln(1+\frac{(\omega^k)^2}{m}).$$

But

$$\frac{(\omega^k)^2}{m} \leq \frac{1}{8}a_{k+1}^T a_{k+1} = \frac{1}{8};$$

hence

$$\begin{aligned} \ln(1+\frac{(\omega^k)^2}{m}) &\geq \frac{(\omega^k)^2}{m} - \frac{(\frac{(\omega^k)^2}{m})^2}{2(1-\frac{(\omega^k)^2}{m})} \\ &= \frac{(\omega^k)^2}{m}(1-\frac{\frac{(\omega^k)^2}{m}}{2(1-\frac{(\omega^k)^2}{m})}) \\ &\geq \frac{(\omega^k)^2}{2m}. \end{aligned}$$

Thus we have

$$\ln\det B^{k+1} \geq \ln\det B^0 + \sum_{j=0}^{k}\frac{(\omega^j)^2}{2m} = m\ln 8 + \sum_{j=0}^{k}\frac{(\omega^j)^2}{2m}.$$

But

$$\frac{1}{m}\ln\det B^{k+1} \leq \ln\frac{\text{trace } B^{k+1}}{m} = \ln(8+\frac{k+1}{m^2}).$$

Thus

$$\sum_{j=0}^{k}\frac{(\omega^j)^2}{2m} \leq m\ln(8+\frac{k+1}{m^2}) - m\ln 8$$

or

$$\sum_{j=0}^{k}(\omega^j)^2 \leq 2m^2\ln(1+\frac{k+1}{8m^2}).$$

Q.E.D.

Theorem 6.1 *The column generation scheme stops with a feasible solution as soon as k satisfies:*

$$\frac{\epsilon^2}{m} \geq \frac{\frac{1}{2} + 2m\ln(1+\frac{k+1}{8m^2})}{2m+k+1} e^{-2\alpha\frac{k+1}{k+1+2m}}.$$

Proof. From relation (6.2) and Lemma 6.1,

$$\begin{aligned}(2m+k+1)\ln\epsilon &\leq P(\Omega^{k+1}) \\ &\leq P(\Omega^0) + \frac{1}{2}\sum_{j=0}^{k}\ln(\bar{r}^j)^2 - (k+1)\alpha \\ &= 2m\ln\frac{1}{2} + \frac{1}{2}\sum_{j=0}^{k}\ln(\bar{r}^j)^2 - (k+1)\alpha.\end{aligned}$$

Thus

$$\begin{aligned}\ln\epsilon + \frac{k+1}{2m+k+1}\alpha &\leq \frac{1}{2(2m+k+1)}\left[2m\ln\frac{1}{4} + \sum_{j=0}^{k}\ln(\bar{r}^j)^2\right] \\ &\leq \frac{1}{2}\ln\frac{2m\frac{1}{4} + \sum_{j=0}^{k}(\bar{r}^j)^2}{2m+k+1} \quad \text{(from the concavity of } \ln\text{)} \\ &\leq \frac{1}{2}\ln\frac{\frac{m}{2} + \sum_{j=0}^{k}(\omega^j)^2}{2m+k+1} \quad \text{(from Lemma 6.4)} \\ &\leq \frac{1}{2}\ln\frac{\frac{m}{2} + 2m^2\ln(1+\frac{k+1}{8m^2})}{2m+k+1} \quad \text{(from Lemma 6.5)}\end{aligned}$$

or

$$\frac{\epsilon^2}{m} \leq \frac{\frac{1}{2} + 2m\ln(1+\frac{k+1}{8m^2})}{2m+k+1} e^{-2\alpha\frac{k+1}{k+1+2m}}.$$

Q.E.D.

Theorem 6.1 implies that the complexity of the column generation scheme, counted by the calls to the oracle is $O^*(\frac{m^2}{\epsilon^2})$; the notation O^* means that lower order terms are ignored. Notice that whether α is positive or not does not alter the estimate of k significantly.

Theorem 6.2 *The approximate analytic center algorithm, which uses the updating scheme of Section 5 and either the dual method of 3.1 or the primal-dual method of Section 3.2, measured by the number of Newton steps or by the number of factorizations, is, for appropriate values of η and β which depend on the exact mix of recentering and updating steps, a fully polynomial approximation algorithm.*

Proof. For the prescribed values of η and β, the number of Newton steps is at most

4. One additional factorization may be required for the updating procedure. **Q.E.D.**

The largest value of η that guarantees $\gamma^{16} \leq \eta$ (so that 4 dual Newton steps are enough to recenter) is about $\eta = .09$ with $\beta = .691$. In this case α may be negative, nonetheless the algorithm will still terminate after $O^*(\frac{m^2}{\epsilon^2})$ of iterations.

References

[1] D. S. Atkinson and P. M. Vaidya, "A cutting plane algorithm that uses analytic centers", University of Illinois at Urbana–Champaign, Urbana, IL 61820, USA, June 1992.

[2] J. V. Burke, A. A. Goldstein, P. Tseng, and Y. Ye, "Translation cuts for minimization," Complexity in Numerical Optimization (Editor: P.M. Pardalos), World Scientific (1993), pp. 57-73.

[3] G. B. Dantzig and P. Wolfe, "The decomposition algorithm for linear programming", *Econometrica* **29** (1961) 767–778.

[4] D. den Hertog, C.Roos and J. Ph. Vial, "A complexity reduction for the long-step path-following algorithm for linear programming", *SIAM Journal on Optimization* **2** (1992) 71–87.

[5] J. Elzinga and T. Moore, "A central cutting plane algorithm for convex programming", *Mathematical Programming* **8** (1975) 134-145.

[6] J.-L. Goffin, Z. Luo, and Y. Ye, "On the complexity of a column generation algorithm for convex or quasiconvex feasibility problems", *Large Scale Optimization: State of the Art*, W.W. Hager, D.W. Hearn and P.M. Pardalos, editors, Kluwer Academic Publishers, 1993, 187–196.

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

[8] C. C. Gonzaga, "Path-following methods for linear programming", *SIAM Review* **34** (1992) 167-224.

[9] J. E. Kelley, "The cutting plane method for solving convex programs", *Journal of the SIAM* **8** (1960) 703–712.

[10] L.G. Khacian, "Polynomial algorithms in linear programming," *Zh. vychisl. Mat. mat. Fiz,* **20**, No. 1 (1980) 51-68; translated in *USSR Computational Mathematics and Mathematical Physics* **20**, No. 1 (1980) 53-72.

[11] L. G. Khacian and M. J. Todd, "On the complexity of approximating the maximal inscribed ellipsoid for a polytope", *Mathematical Programming* **61** (1993) 137-160.

[12] A. Levin, "An algorithm of minimization of convex functions", *Soviet Math. Doklady* **160** (1965), 6, 1244-1247.

[13] J. E. Mitchell, "Karmarkar's algorithm and combinatorial optimization problems," Ph.D. Thesis, Department of ORIE, Cornell University (Ithaca, NY, 1988).

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

[15] S. Mizuno, "A new polynomial time method for a linear complementarity problem", Technical Report 16, Department of IE and Management, Tokyo Institute of Technology, Tokyo, Japan, Feb. 1989.

[16] A. Nemirovsky and D. Yudin, "Problem complexity and method efficiency in Optimization", Wiley-Interscience, NY 1983.

[17] Y. Nesterov, "Cutting plane algorithms from analytic centers: efficiency estimates", University of Geneva, Geneva, Switzerland, December 1992.

[18] N. Z. Shor, "Minimization Methods for Non-Differentiable Functions", Springer-Verlag, Berlin, Heidelberg (1985).

[19] G. Sonnevend, "New algorithms in convex programming based on a notion of "centre" (for systems of analytic inequalities) and on rational extrapolation", in K. H. Hoffmann, J. B. Hiriat-Urruty, C. Lemarechal, and J. Zowe, editors, *Trends in Mathematical Optimization: Proceedings of the 4th French–German Conference on Optimization in Irsee, West–Germany, April 1986*, volume 84 of *International Series of Numerical Mathematics*, pages 311–327. Birkhäuser Verlag, Basel, Switzerland, 1988.

[20] S. Tarasov, L.G. Khachiyan, I. Erlich, "The method of inscribed ellipsoids", *Soviet Math. Doklady* **37** (1988).

[21] P. Vaidya, " A new algorithm for minimizing convex functions over convex sets" 1989, to appear in *Mathematical Programming.*

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