A large-step analytic center method for a class of smooth convex programming problems

den Hertog, D.; Roos, C.; Terlaky, T.

Published in:
SIAM Journal on Optimization

Publication date:
1992

SIAM J. OPTIMIZATION
Vol. 2, No. 1, pp. 55–70, February 1992

A LARGE-STEP ANALYTIC CENTER METHOD FOR A CLASS OF SMOOTH CONVEX PROGRAMMING PROBLEMS*

D. DEN HERTOG†, C. ROOS†, AND T. TERLAKY‡

Abstract. In this paper, a large-step analytic center method for smooth convex programming is proposed. The method is a natural implementation of the classical method of centers. It is assumed that the objective and constraint functions fulfil the so-called Relative Lipschitz Condition, with Lipschitz constant $M > 0$. A great advantage of the method, over the existing path-following methods, is that the steps can be made long by performing linesearches.

In this method linesearches are performed along the Newton direction with respect to a strictly convex potential function when located far away from the central path. When sufficiently close to this path a lower bound for the optimal value is updated. It is proven that the number of iterations required by the algorithm to converge to an ϵ-optimal solution is $O((1 + M^2)\sqrt{n}|\ln \epsilon|)$ or $O((1 + M^2)n|\ln \epsilon|)$, depending on the updating scheme for the lower bound.

Key words. convex programming, analytic center method, Newton method

AMS(MOS) subject classification. 90C25

1. Introduction. Since Karmarkar [6] presented his projective method for the solution of the linear programming problem in 1984, many other variants have been developed by researchers. Among them are the large-step path-following methods such as those proposed by Roos and Vial [9]; Gonzaga [3]; and Den Hertog, Roos, and Terlaky [1]; and the potential reduction methods such as those proposed by Ye [12], Freund [2], and Gonzaga [4]. The advantages of these methods are that they do not use projective transformations as the projective methods do, and that they do not need to follow the so-called central path closely, contrary to the small-step path-following methods.

In Jarre [5] and Mehrotra and Sun [8] small-step path-following algorithms are proposed for smooth convex programming problems. Again, the great disadvantage of these methods is that they are based on very small stepsizes to remain in the vicinity of the central trajectory. This characteristic makes these methods unattractive for practical use. To accelerate his method, Jarre proposed a (higher-order) extrapolation scheme.

In this paper we propose a large-step path-following method for smooth convex programming problems, which fulfil the so-called Relative Lipschitz Condition. Jarre [5] also uses this condition. Our method is a generalization of the method for linear programming presented in [1] and is also based on Jarre's paper.

In our method we do a linesearch along the Newton direction with respect to a certain strictly convex potential function. If we are close to the current analytic center we update the lower bound somehow, whereafter we do linesearches aiming at getting close to the analytic center associated with the new lower bound. We prove that after a linesearch the potential value reduces with at least a certain constant. Using this result, we prove that the number of iterations required by the algorithm to converge

* Received by the editors April 11, 1990; accepted for publication (in revised form) January 29, 1991.

† Faculty of Technical Mathematics and Computer Science, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, the Netherlands.

‡ Eötvös University, Department of Operations Research, 1088 Budapest, Muzeum Krt. 6-8, Hungary. This work was performed while the author was visiting Delft University of Technology, P.O. Box 5031, 2600 GA Delft, the Netherlands.

to an ϵ-optimal solution is bounded by a polynomial in $|\ln \epsilon|$, the dimension of the problem, and the Lipschitz constant.

We note that Kojima, Mizuno, and Yoshise [7] already proposed a primal–dual potential reduction algorithm for linear complementarity problems. To our knowledge our algorithm is the first large-step algorithm for (a class of) smooth convex programming. Our algorithm can also be viewed as a natural implementation of the classical method of centers. In a coming report we will deal with a natural implementation of the logarithmic barrier function method.

This paper is organized as follows. In §2 we will do some preliminary work. In §3 we describe our algorithm. Then in §4 we prove some lemmas needed for the convergence analysis in §5.

2. Preliminaries. We consider the primal formulation of the smooth convex programming problem,

$$\text{(CP)} \qquad \max \{f_0(y) \; : \; y \in \mathcal{F}\},$$

where $\mathcal{F}$ denotes the feasible region, which is given by

$$\mathcal{F} := \{y \in \mathbb{R}^m \; : f_i(y) \le 0, \;\; 1 \le i \le n\};$$

the functions $-f_0(y)$ and $f_i(y)$, $1 \le i \le n$, are convex functions with continuous first- and second-order derivatives in $\mathcal{F}$. We assume an additional smoothness condition, namely, that the Hessian matrix of $f_i(y)$, $0 \le i \le n$, fulfils the so-called Relative Lipschitz Condition, which will be specified below. Moreover, we suppose that the interior of the feasible region $\mathcal{F}$, denoted as $\mathcal{F}'$, is nonempty and bounded. This assumption is not essential.

Wolfe's [11] formulation of the dual problem associated with this primal problem is

$$\text{(D)} \qquad \begin{array}{l} \min \; f_0(y) - \sum_{i=1}^n u_i f_i(y), \\ \sum_{i=1}^n u_i \nabla f_i(y) = \nabla f_0(y), \\ u_i \ge 0. \end{array}$$

Note that there is no symmetry between the primal and dual problem, as in linear programming, because the dual problem (D) contains both y and u variables. Moreover, the dual problem is not necessarily convex!

However, it is a well-known result that if y is a feasible solution of (CP) and $(\overline{y}, u)$ is a feasible solution of the dual problem (D), then

$$f_0(y) \le f_0(\overline{y}) - \sum_{i=1}^n u_i f_i(\overline{y}).$$

Due to the assumption that $\mathcal{F}'$ is nonempty, the Slater condition is satisfied, and hence (D) has a minimum solution and the extremum values are equal.

We associate the following potential function with (CP):

$$\phi(y, z) = -q \ln(f_0(y) - z) - \sum_{i=1}^n \ln(-f_i(y)),$$

where z is a lower bound for the optimal value z^*, and q is a positive integer value, which will be discussed below. For $q = n$ this potential function is exactly the same as the one used by Jarre [5].

It can be proved that $\phi(y,z)$ is strictly convex on its domain $\mathcal{F}$ (see Jarre [5, p. 8]). It also takes infinite values on the boundary of the feasible set. Hence this potential function achieves the minimal value in its domain (for fixed z) at a unique point, which is denoted by $y(z)$. The necessary and sufficient Karush–Kuhn–Tucker conditions for these minima are:

$$\begin{array}{lll} & f_i(y) \le 0, & 1 \le i \le n, \\ (1) & \sum_{i=1}^n u_i \nabla f_i(y) = \nabla f_0(y), & u \ge 0, \\ & -f_i(y)u_i = \frac{f_0(y)-z}{q}, & 1 \le i \le n. \end{array}$$

Using this it can easily be verified that $y(z)$ lies on the so-called central path of the problem, which is the set of analytic centers for $\mathcal{F} \cap \{y : f_0(y) \ge \mu\}$, where μ varies from $-\infty$ to z^*.

We can rewrite $\phi(y,z)$ as

$$\phi(y,z) = -\sum_{i=1}^{n+q} \ln(-f_i(y)),$$

where $-f_i(y) = f_0(y) - z$ for $n+1 \le i \le n+q$. The first- and second-order derivatives of $\phi(y,z)$ are given by

$$g(y,z) := \nabla\phi(y,z) = \sum_{i=1}^{n+q} \frac{\nabla f_i(y)}{-f_i(y)}$$

and

$$H(y,z) := \nabla^2\phi(y,z) = \sum_{i=1}^{n+q} \left[\frac{\nabla^2 f_i(y)}{-f_i(y)} + \frac{\nabla f_i(y) \nabla f_i(y)^T}{f_i(y)^2} \right].$$

If no confusion is possible we will write, for shortness' sake, g and H instead of $g(y,z)$ and $H(y,z)$.

In the sequel to this paper we will also use the quadratic approximation $q_y(x,z)$ for $\phi(y,z)$ when x is near the point y, defined as

$$q_y(x,z) := \phi(y,z) + g^T(x-y) + \tfrac{1}{2}(x-y)^T H(x-y).$$

We will use the H-norm $\|.\|_H$ to measure closeness of points, and especially closeness to the central trajectory. The definition of this norm is as follows:

$$\|x\|_H = \sqrt{x^T H x}.$$

Because H is positive definite, $\|.\|_H$ defines a norm.

Having introduced this notation we are able to formulate the Relative Lipschitz Condition:

$$\exists M > 0 : \ \forall v \in \mathbb{R}^m \quad \forall y, y+h \in \mathcal{F}' :$$

$$|v^T(\nabla^2 f_i(y+h) - \nabla^2 f_i(y))v| \le M\|h\|_H v^T \nabla^2 f_i(y) v, \tag{2}$$

for all $1 \le i \le n+1$.

This condition is also used by Jarre [5]. In general, the condition might be hard to check for a given problem.

3. The algorithm. In our algorithm we do not need to stay close to the central path, as in Jarre [5]. If we are far away from the central path we do a linesearch along the Newton direction with respect to $\phi(y,z)$. The Newton direction $p(y,z)$ associated with $\phi(y,z)$ at y is given by

$$p(y,z) = -H(y,z)^{-1}g(y,z) = -H^{-1}g.$$

If no confusion is possible we will write, for shortness' sake, p instead of $p(y,z)$. This process is repeated until we are sufficiently close to the central path. More precisely, we stop doing linesearches if $\|p\|_H \le \tau$, where τ is a certain tolerance. This proximity criterion is also used by Jarre [5]. In the algorithm we will use $\tau = \frac{1}{8(1+2M)}$, which will appear to be appropriate later on. (Note that $\|p\|_H = 0$ if and only if $y = y(z)$.) If the proximity criterion is satisfied, we update the lower bound z as follows: $\overline{z} := z + \theta(f_0(y) - z)$, for some $0 < \theta < 1$, and the whole process is repeated again and again until some stopping criterion is satisfied. Note that $\overline{z}$ is really a lower bound, because $\overline{z} < f_0(y) \le z^*$.

We can now describe the algorithm.

Algorithm
Input:
θ is the updating factor, $0 < \theta < 1$;
$\tau = \frac{1}{8(1+2M)}$ is the proximity tolerance;
t is an accuracy parameter, $t \in \mathbb{N}$;
y^0 is a given interior feasible point and $z^0 < f_0(y^0)$ is a lower bound for the optimal value, such that $\|p(y^0,z^0)\|_{H(y^0,z^0)} \le \tau$ and $z^* - z^0 \le \frac{1}{\epsilon}$.

begin
  $y := y^0$; $z := z^0$;
  while $f_0(y) - z > e^{-t}$ **do**
  begin (outer step)
    while $\|p\|_H > \tau$ **do**
    begin (inner step)
      $\bar{\alpha} := \arg\min_{\alpha>0} \{\phi(y+\alpha p, z) : y + \alpha p \in \mathcal{F}'\}$
      $y := y + \bar{\alpha} p$
    end (end inner step)
    $z := z + \theta(f_0(y) - z)$;
  end (end outer step)
end.

For finding the initial point that satisfies the input assumptions of the algorithm we refer the reader to Jarre [5] and Mehrotra and Sun [8]. Later on, the "centering assumption" will be alleviated.

4. Preliminary lemmas. In §5 we will prove the complexity result on the Algorithm. In this section we deal with some lemmas that will be needed to obtain an upper bound for the total number of outer and inner iterations. The lemmas are built up as follows:

- Lemma 4.1 gives an upper bound for the error in the quadratic approximation if the functions $-f_0(y)$ and $f_i(y)$, $1 \le i \le n$, are linear or quadratic;
- Lemma 4.2 states the same as Lemma 4.1, but now $-f_0(y)$ and $f_i(y)$, $1 \le i \le n$, are general convex functions;

- Lemma 4.3 states that if the proximity criterion holds, then y lies close to the exact center $y(z)$ (with respect to the H-norm);
- Lemma 4.4 states that if we do a linesearch along the Newton direction, then a sufficient decrease in the potential value can be guaranteed;
- Lemma 4.5 gives an upper bound for the difference in potential value of the current iterate and the exact center;
- Lemma 4.6 states that if the lower bound is updated, then the potential value increases with a constant;
- Lemma 4.7 gives a relation between the objective value in the exact center and the current point;
- Lemma 4.8 gives an upper bound for the gap between the optimal value and the lower bound z;
- Lemma 4.9 states that the gap $f_0(y(z)) - z$ decreases monotonically if z increases.

The following lemma improves Lemma 2.1 of Jarre [5].

LEMMA 4.1. *If $-f_0(y)$ and $f_i(y)$, $1 \leq i \leq n$, are linear or quadratic functions with positive semidefinite Hessian matrix, and if $y \in \mathcal{F}'$ and $\|d\|_H < 1$, then $y+d \in \mathcal{F}'$ and*

$$|\phi(y+d,z) - q_y(y+d,z)| < \frac{\|d\|_H^3}{3(1-\|d\|_H)}.$$

Proof. We expand $\phi(y+d,z)$ in a Taylor series about y:

$$\phi(y+d,z) = q_y(y+d,z) + \sum_{i=3}^{\infty} t_i, \tag{3}$$

where t_i is the ith-order Taylor term in the expansion. Note that ϕ only takes finite values in $\mathcal{F}'$. Hence, if $\sum_{i=3}^{\infty} t_i$ can be shown to converge for d such that $\|d\|_H < 1$, then it follows that $y + d \in \mathcal{F}'$. It can be proved that

$$|t_i| \leq \frac{1}{i}\|d\|_H^i. \tag{4}$$

The proof of this inequality is quite technical. Therefore, it is omitted here (see Appendix). From (4), and using the fact that $\|d\|_H < 1$, we derive that

$$\sum_{i=3}^{\infty} |t_i| \leq \sum_{i=3}^{\infty} \frac{\|d\|_H^i}{i} \leq \frac{\|d\|_H^3}{3(1-\|d\|_H)}.$$

Substituting this into (3) yields

$$|\phi(y+d,z) - q_y(y+d,z)| \leq \sum_{i=3}^{\infty} |t_i| \leq \frac{\|d\|_H^3}{3(1-\|d\|_H)}.$$

Thus the lemma has been proved. □

LEMMA 4.2. *If the functions $f_i(y)$ satisfy the Relative Lipschitz Condition with Lipschitz constant $M > 0$, and if*

$$y \in \mathcal{F}' \quad \text{and} \quad \|d\|_H < \min\left\{\frac{1}{2}, \frac{1}{2M^{1/3}}\right\},$$

then $y + d \in \mathcal{F}'$ *and*

$$|\phi(y+d,z) - q_y(y+d,z)| < \frac{\|d\|_H^3}{3(1-\|d\|_H)}(1+2M).$$

Proof. Using Lemma 4.1, one can use the same reasoning as in the proof of Lemma 2.10 of Jarre [5] to obtain the result of the lemma. □

The next lemma simplifies Lemma 2.16 of Jarre [5].

LEMMA 4.3. *If* $\|p\|_H \le \frac{1}{8(1+2M)}$, *then*

$$\|y - y(z)\|_H \le \frac{5}{2}\|p\|_H.$$

Proof. Let h be arbitrary, such that $\|h\|_H = \frac{3}{2}\|p\|_H$. We consider the values on the ellipsoid $\{y+p+h \ : \ \|h\|_H = \frac{3}{2}\|p\|_H\}$. We have

$$\|h+p\|_H \le \|h\|_H + \|p\|_H = \frac{5}{2}\|p\|_H < \frac{1}{3(1+2M)}. \tag{5}$$

With the help of Lemma 4.2, and using the fact that $y + p = \arg\min_x q_y(x,z)$, we obtain

$$\begin{aligned}
\phi(y+p+h,z) &> q_y(y+p+h,z) - \frac{1}{3(1-\frac{1}{3})}\|p+h\|_H^3(1+2M) \\
&\ge q_y(y+p,z) + \frac{1}{2}\|h\|_H^2 - \frac{1}{2}\|p+h\|_H^3(1+2M) \\
&\ge q_y(y+p,z) + \frac{9}{8}\|p\|_H^2 - \frac{125}{16}\|p\|_H^3(1+2M) \\
&\ge q_y(y+p,z) + 9\|p\|_H^3(1+2M) - \frac{125}{16}\|p\|_H^3(1+2M) \\
&> q_y(y+p,z) + \|p\|_H^3(1+2M).
\end{aligned}$$

Using Lemma 4.2 once more, we also obtain that

$$\phi(y+p,z) < q_y(y+p,z) + \frac{1}{2}\|p\|_H^3(1+2M).$$

Hence $\phi(y+p+h,z) > \phi(y+p,z)$. Thus in the center, $y+p$, of the ellipsoid the potential value is less than the value on its boundary. Therefore, by the strict convexity of ϕ, the minimum of ϕ is in the interior of the ellipsoid, which means that $\|y - y(z)\|_H \le \|p+h\|_H$. Now using (5), the lemma follows. □

LEMMA 4.4. *If* $\|p\|_H \ge \frac{1}{8(1+2M)}$, *then the decrease* $\Delta\phi$ *in the potential function after a linesearch along the Newton direction* p *satisfies*

$$\Delta\phi \ge \frac{1}{140(1+2M)^2}.$$

Proof. Let λ be a steplength such that

$$\|\lambda p\|_H \le \min\left\{\frac{1}{2}, \frac{1}{2M^{1/3}}\right\}. \tag{6}$$

Then, as a consequence of Lemma 4.2, we have

$$\phi(y+\lambda p, z) \leq q_y(y+\lambda p, z) + \frac{\lambda^3\|p\|_H^3}{3(1-\lambda\|p\|_H)}(1+2M).$$

Now using the definition of q_y, we obtain

$$\begin{aligned}\phi(y,z) - \phi(y+\lambda p, z) &\geq -\lambda g^T p - \frac{1}{2}\lambda^2 p^T H p - \frac{\lambda^3\|p\|_H^3}{3(1-\lambda\|p\|_H)}(1+2M) \\ &= \lambda\|p\|_H^2 - \frac{1}{2}\lambda^2\|p\|_H^2 - \frac{\lambda^3\|p\|_H^3}{3(1-\lambda\|p\|_H)}(1+2M).\end{aligned}$$

Replacing λ by the value

$$\frac{1}{9(1+2M)\|p\|_H},$$

which satisfies (6), yields the lemma. □

LEMMA 4.5. *If* $\|p\|_H \leq \frac{1}{8(1+2M)}$, *then*

$$\phi(y,z) - \phi(y(z),z) \leq 4\|p\|_H^2. \tag{7}$$

Proof. Let d be defined as $y(z) - y$. Using Lemmas 4.2 and 4.3 we get

$$\begin{aligned}\phi(y(z),z) &\geq q_y(y+d,z) - \frac{\|d\|_H^3}{3(1-\|d\|_H)}(1+2M) \\ &= \phi(y,z) - p^T H d + \frac{1}{2}d^T H d - \frac{\|d\|_H^3}{3(1-\|d\|_H)}(1+2M).\end{aligned}$$

Using the Cauchy–Schwarz inequality we may write

$$-p^T H d \geq -\|p\|_H\|d\|_H.$$

Also using Lemma 4.3, we obtain

$$\begin{aligned}\phi(y,z) - \phi(y(z),z) &\leq \|p\|_H\|d\|_H - \frac{1}{2}\|d\|_H^2 + \frac{\|d\|_H^3}{3(1-\|d\|_H)}(1+2M) \\ &\leq \|p\|_H\|d\|_H + \frac{\|d\|_H^3}{24(1-\|d\|_H)\|p\|_H} \\ &\leq \frac{5}{2}\|p\|_H^2 + \frac{\frac{125}{8}}{24(1-\frac{5}{16})}\|p\|_H^2 \\ &\leq 4\|p\|_H^2.\end{aligned}$$

□

LEMMA 4.6. *Let* $\bar{z}$ *be the new lower bound, i.e.,* $\bar{z} = z + \theta(f_0(y) - z)$, *where* $0 < \theta < 1$, *then*

$$\phi(y,\bar{z}) - \phi(y,z) = -q\ln(1-\theta).$$

Proof. The proof is simple and straightforward. We have

$$f_0(y) - \bar{z} = f_0(y) - z - \theta(f_0(y) - z) = (1-\theta)(f_0(y) - z).$$

Hence

$$\phi(y, \bar{z}) - \phi(y, z) = -q \ln \frac{f_0(y) - \bar{z}}{f_0(y) - z} = -q \ln(1-\theta). \qquad \square$$

LEMMA 4.7. *If* $\|y - y(z)\|_H \le \beta$, *then*

$$f_0(y(z)) - z \le \left(1 + \frac{\beta}{\sqrt{q}}\right)(f_0(y) - z).$$

Proof. The lemma is trivial if $f_0(y(z)) \le f_0(y)$. So let us assume that $f_0(y(z)) > f_0(y)$. By definition we have

$$\begin{aligned}
\beta^2 &\ge \|y - y(z)\|_H^2 \\
&= (y - y(z))^T \left[\sum_{i=1}^{n+q} \left(\frac{\nabla f_i(y) \nabla f_i(y)^T}{f_i(y)^2} + \frac{\nabla^2 f_i(y)}{-f_i(y)}\right)\right](y - y(z)) \\
&\ge (y - y(z))^T q \frac{\nabla f_0(y) \nabla f_0(y)^T}{(f_0(y) - z)^2}(y - y(z)) \\
&\ge q \frac{(f_0(y) - f_0(y(z)))^2}{(f_0(y) - z)^2},
\end{aligned}$$

where the last inequality follows from the convexity of $-f_0(y)$ and the assumption that $f_0(y(z)) > f_0(y)$. Consequently,

$$f_0(y(z)) - f_0(y) \le \frac{\beta}{\sqrt{q}}(f_0(y) - z).$$

This means that

$$f_0(y(z)) - z \le \left(1 + \frac{\beta}{\sqrt{q}}\right)(f_0(y) - z). \qquad \square$$

LEMMA 4.8. *If* $\|y - y(z)\|_H \le \beta$, *then*

$$z^* - z \le \left(1 + \frac{n}{q}\right)\left(1 + \frac{\beta}{\sqrt{q}}\right)(f_0(y) - z).$$

Proof. The exact center $y(z)$ minimizes the potential function for z. The necessary and sufficient conditions for these minima are (1). From these conditions we derive that $(u(z), y(z))$ is dual-feasible. Moreover, using $z^* \le f_0(y(z)) - \sum_{i=1}^n u_i(z) f_i(y(z))$, it follows that

$$z^* - f_0(y(z)) \le -\sum_{i=1}^n u_i(z) f_i(y(z)) = \frac{n}{q}(f_0(y(z)) - z).$$

Consequently,

$$(z^* - z) - (f_0(y(z)) - z) \le \frac{n}{q}(f_0(y(z)) - z).$$

This means that

$$z^* - z \leq \left(1 + \frac{n}{q}\right)(f_0(y(z)) - z) \leq \left(1 + \frac{n}{q}\right)\left(1 + \frac{\beta}{\sqrt{q}}\right)(f_0(y) - z),$$

where the last inequality follows from Lemma 4.7. This proves the lemma. □

The next lemma generalizes an inequality of Vaidya [10] for the LP-case to the present convex case.

LEMMA 4.9. *The gap $f_0(y(z)) - z$ decreases monotonically if $z < z^*$ increases.*

Proof. We have that $u(z)$ and $y(z)$ satisfy the Karush–Kuhn–Tucker conditions (1). Taking the derivative with respect to z of the last two equations in (1), we obtain

$$\sum_{i=1}^{n} u_i' \nabla f_i(y) + \sum_{i=1}^{n} u_i H_i y' = H_0 y', \tag{8}$$

$$-u_i' f_i(y) - u_i \nabla f_i(y)^T y' = \frac{\nabla f_0(y)^T y' - 1}{q}, \qquad i = 1, \cdots, n, \tag{9}$$

where the prime denotes the derivative with respect to z and H_i denotes the Hessian matrix of $f_i(y)$. The Jacobian of this system of equations is clearly nonsingular for $z < z^*$, and hence, as a consequence of the implicit function theorem, we may conclude that u' and y' exist for $z < z^*$. Multiplying (9) with u_i and using (1), we get

$$\frac{f_0(y) - z}{q} u_i' - u_i^2 \nabla f_i(y)^T y' = \frac{\nabla f_0(y)^T y' - 1}{q} u_i.$$

Multiplying this equation with $\nabla f_i(y)$, summing over i, and using (8) and (1) results in

$$-\frac{f_0(y) - z}{q}\left(\sum_{i=1}^{n} u_i H_i y' - H_0 y'\right) - \sum_{i=1}^{n} u_i^2 \nabla f_i(y)^T y' \nabla f_i(y) = \frac{\nabla f_0(y)^T y' - 1}{q} \nabla f_0(y).$$

Now, taking the inner product with y', we obtain

$$\frac{\nabla f_0(y)^T y' - 1}{q} \nabla f_0(y)^T y' = -\frac{f_0(y) - z}{q}(y')^T \left(\sum_{i=1}^{n} u_i H_i - H_0\right) y'$$
$$- \sum_{i=1}^{n} u_i^2 (\nabla f_i(y)^T y')^2 \leq 0.$$

We conclude that $0 \leq \nabla f_0(y)^T y' \leq 1$, which means that the derivative of $f_0(y(z)) - z$, which is equal to $\nabla f_0(y)^T y' - 1$, is not positive. This proves the lemma. □

5. Convergence analysis. Based on the lemmas in the previous section, we will give upper bounds for the total number of outer iterations and inner iterations.

THEOREM 5.1. *Let $\beta \leq \frac{5}{16(1+2M)}$; then after at most*

$$K = \frac{(1 + \frac{n}{q})(1 + \frac{\beta}{\sqrt{q}})}{\theta} O(|\ln \epsilon|)$$

outer iterations, the algorithm finds an ϵ-optimal solution for (CP).

Proof. Let z^k be the lower bound in the kth outer iteration. Then we have

$$\begin{aligned}\frac{z^* - z^k}{z^* - z^{k-1}} &= \frac{z^* - (z^{k-1} + \theta(f_0(y^{k-1}) - z^{k-1}))}{z^* - z^{k-1}} \\ &= 1 - \theta\frac{f_0(y^{k-1}) - z^{k-1}}{z^* - z^{k-1}} \\ &\leq 1 - \frac{\theta}{(1+\frac{n}{q})(1+\frac{\beta}{\sqrt{q}})},\end{aligned}$$

where y^{k-1} is the iterate at the end of the $(k-1)$th outer iteration. The last inequality follows from Lemma 4.8. Hence after K outer iterations we have

$$\begin{aligned} z^* - f_0(y^K) &\leq z^* - z^{K+1} \\ &\leq \left(1 - \frac{\theta}{(1+\frac{n}{q})(1+\frac{\beta}{\sqrt{q}})}\right)(z^* - z^K) \\ &\leq \left(1 - \frac{\theta}{(1+\frac{n}{q})(1+\frac{\beta}{\sqrt{q}})}\right)^K (z^* - z^0).\end{aligned}$$

This means that $z^* - f_0(y^K) \leq \epsilon$ certainly holds if

$$\left(1 - \frac{\theta}{(1+\frac{n}{q})(1+\frac{\beta}{\sqrt{q}})}\right)^K (z^* - z^0) \leq \epsilon.$$

Taking logarithms, this inequality reduces to

$$-K \ln\left(1 - \frac{\theta}{(1+\frac{n}{q})(1+\frac{\beta}{\sqrt{q}})}\right) \geq |\ln \epsilon| + \ln(z^* - z^0).$$

Since $-\ln(1-v) > v$, this will certainly hold if

$$K > \frac{(1+\frac{n}{q})(1+\frac{\beta}{\sqrt{q}})}{\theta}(|\ln \epsilon| + \ln(z^* - z^0)).$$

Now using the assumption on z^0, i.e. $z^* - z^0 \leq \frac{1}{\epsilon}$, the theorem follows. □

From Lemma 4.8 it follows that it suffices to take

$$t = \ln \frac{(1+\frac{n}{q})(1+\frac{\beta}{\sqrt{q}}) - 1}{\epsilon},$$

i.e., for such t the algorithm ends up with a solution y such that $z^* - f_0(y) \leq \epsilon$.

Now we give an upper bound for the total number of inner iterations during an arbitrary outer iteration.

THEOREM 5.2. *The total number P of inner iterations during an arbitrary outer iteration satisfies*

$$P\delta \leq 1 + \frac{\theta\beta q}{\sqrt{q} + \beta} + \frac{\theta^2 q}{1-\theta},$$

where δ is the guaranteed decrease in each inner iteration, and $\beta \leq \frac{5}{16(1+2M)}$.

Proof. We denote the used lower bound in an arbitrary outer iteration by $\tilde{z}$, while the lower bound in the previous outer iteration is denoted by $\overline{z}$. The iterate at the beginning of the outer iteration is denoted by y. Hence y is centered with respect to $y(\overline{z})$ and $\tilde{z} = \overline{z} + \theta(f_0(y) - \overline{z})$. Because of Lemma 4.4 and definition of $y(\tilde{z})$ we have

$$P\delta \leq \phi(y, \tilde{z}) - \phi(y(\tilde{z}), \tilde{z}). \tag{10}$$

Let us call the right-hand side of (10) $\Phi(y, \tilde{z})$. According to the mean value theorem there is a $\hat{z} \in (\overline{z}, \tilde{z})$ such that

$$\Phi(y, \tilde{z}) = \Phi(y, \overline{z}) + \left.\frac{d\,\Phi(y, z)}{d\,z}\right|_{z=\hat{z}} (\tilde{z} - \overline{z}). \tag{11}$$

Let us now look at $\frac{d\,\Phi(y,z)}{d\,z}$. We have

$$\frac{d\,\phi(y, z)}{d\,z} = \frac{q}{f_0(y) - z}$$

and

$$\begin{aligned}\frac{d\,\phi(y(z), z)}{d\,z} &= -q\frac{\nabla f_0(y(z))^T y' - 1}{f_0(y(z)) - z} + \sum_{i=1}^{n} \frac{\nabla f_i(y(z))^T y'}{-f_i(y(z))} \\ &= -q\frac{\nabla f_0(y(z))^T y' - 1}{f_0(y(z)) - z} + \frac{q}{f_0(y(z)) - z}\sum_{i=1}^{n} u_i(z)\nabla f_i(y(z))^T y' \\ &= \frac{q}{f_0(y(z)) - z},\end{aligned}$$

where the two last equations follow from (1). So

$$\left.\frac{d\,\Phi(y, z)}{d\,z}\right|_{z=\hat{z}} = q\left.\left(\frac{1}{f_0(y) - z} - \frac{1}{f_0(y(z)) - z}\right)\right|_{z=\hat{z}} \leq q\left(\frac{1}{f_0(y) - \tilde{z}} - \frac{1}{f_0(y(\overline{z})) - \overline{z}}\right),$$

where the last inequality follows from the fact that $\tilde{z} > \hat{z}$ and from Lemma 4.9. Substituting this into (11) gives

$$\begin{aligned}\Phi(y, \tilde{z}) &\leq \Phi(y, \overline{z}) + q\left(\frac{1}{f_0(y) - \tilde{z}} - \frac{1}{f_0(y(\overline{z})) - \overline{z}}\right)(\tilde{z} - \overline{z}) \\ &= \Phi(y, \overline{z}) + q\theta\left(\frac{1}{1-\theta} - \frac{f_0(y) - \overline{z}}{f_0(y(\overline{z})) - \overline{z}}\right) \\ &\leq 1 + q\theta\left(\frac{1}{1-\theta} - \frac{1}{1 + \frac{\beta}{\sqrt{q}}}\right) \qquad (12) \\ &= 1 + q\theta\left(\frac{\theta}{1-\theta} + \frac{\beta}{\sqrt{q} + \beta}\right),\end{aligned}$$

where inequality (12) follows because $\Phi(y, \overline{z}) \leq 1$ according to Lemma 4.5, and $f_0(y(\overline{z})) - \overline{z} \leq (1 + (\beta/\sqrt{q}))(f_0(y) - \overline{z})$ according to Lemma 4.7. □

From Theorem 5.1 we know that the total number of outer iterations is at most

$$\frac{(1 + \frac{n}{q})(1 + \frac{\beta}{\sqrt{q}})}{\theta} O(|\ln \epsilon|).$$

Hence the total number of inner iterations during the whole process is given by

$$\frac{1}{\delta}\left(1+\frac{n}{q}\right)\left(1+\frac{\beta}{\sqrt{q}}\right)\left(\frac{1}{\theta}+\frac{\beta q}{\sqrt{q}+\beta}+\frac{\theta q}{1-\theta}\right)O(|\ln \epsilon|). \tag{13}$$

This makes clear that if we take $q = n$, then

- if we take $\theta = \nu/\sqrt{n}$, $\nu > 0$ and independent of n, M, and ϵ, then the algorithm has an $O((1+M)^2\sqrt{n}|\ln\epsilon|)$ iteration bound.
- if we take $0 < \theta < 1$, independent of n, M, and ϵ, then the algorithm has an $O((1+M)^2 n|\ln\epsilon|)$ iteration bound.

The first case corresponds to a small reduction factor θ. In this case we can return to the vicinity of the central trajectory in $O((1+M)^2)$ steps, while the lower bound must be updated $O(\sqrt{n}|\ln\epsilon|)$ times. In the path-following algorithm of Jarre [5], the same iteration bound is obtained for

$$\theta = \frac{1}{200\sqrt{n}(1+M^2)}.$$

The second case corresponds to a large reduction factor θ. In this case we can return to the vicinity of the central trajectory in $O((1+M)^2 n)$ linesearches, while the lower bound must be updated $O(|\ln\epsilon|)$ times.

Remark 1. We note that the upper bound for the number of iterations is not better than Jarre's. However, while Jarre's bound is more or less exact, our bound can be very pessimistic, because of the linesearches involved in the inner iterations. This can also be one of the reasons for the fact that a large reduction factor gives a worse bound than a small reduction factor, while one would expect the contrary.

Remark 2. The "centering assumption" $\|p(y^0,z^0)\|_{H(y^0,z^0)} \le \tau$ can be alleviated to

$$\phi(y^0,z^0) - \phi(y(z^0),z^0) \le O(\sqrt{n}|\ln\epsilon|)$$

for the first case, and to

$$\phi(y^0,z^0) - \phi(y(z^0),z^0) \le O(n|\ln\epsilon|)$$

for the second case. This follows easily from Lemma 4.4.

Remark 3. Note that the updating factor θ is independent from M, contrary to Jarre's [5] method.

Remark 4. For linear programming problems, i.e., $M = 0$, we can find an exact solution if we take $\epsilon = 2^{-L}$, where L denotes the input length of the problem. In this case our results reduce to an $O(\sqrt{n}L)$ iteration bound if we take $\theta = \nu/\sqrt{n}$, $\nu > 0$ and independent of n, M, and ϵ; and to an $O(nL)$ iteration bound if we take $0 < \theta < 1$, independent of n, M, and ϵ. These results are also obtained by Den Hertog, Roos, and Terlaky [1]; Gonzaga [3]; and Roos and Vial [9].

Appendix: Proof of the inequalities (4). Since each function f_i is assumed to be linear or quadratic, the kth-order term in its Taylor expansion has the following form:

$$t_k = \frac{1}{k!}\sum_{j=1}^{n+q}\sum_{i=0}^{\lfloor k/2\rfloor} a_{k,i}\frac{(\nabla f_j(y)^T d)^{k-2i}(d^T\nabla^2 f_j(y)d)^i}{(-f_j(y))^{k-i}}, \qquad k \ge 1, \tag{14}$$

where $a_{k,i}$ has to be determined yet. For shortness' sake we use the following notations:

$$\chi_j := \frac{\nabla f_j(y)^T d}{-f_j(y)},$$
$$\psi_j := \frac{d^T \nabla^2 f_j(y) d}{-f_j(y)},$$
$$D^2 := \|d\|_H^2.$$

Using these notations (14) becomes

$$t_k = \frac{1}{k!} \sum_{j=1}^{n+q} \sum_{i=0}^{\lfloor k/2 \rfloor} a_{k,i} \chi_j^{k-2i} \psi_j^i. \tag{15}$$

We also have

$$D^2 = \sum_{j=1}^{n+q} (\chi_j^2 + \psi_j).$$

Now we derive a recursive formula for $a_{k,i}$. Using the chain rule for taking derivatives, we obtain from (14) an expression for t_{k+1}:

$$t_{k+1} = \frac{1}{(k+1)!} \sum_{j=1}^{n+q} \left[\sum_{i=0}^{\lfloor k/2 \rfloor} (k-i) a_{k,i} \chi_j^{k-2i+1} \psi_j^i + \sum_{i=0}^{\lfloor k/2 \rfloor} (k-2i) a_{k,i} \chi_j^{k-2i-1} \psi_j^{i+1} \right].$$

This can be rewritten as

$$t_{k+1} = \frac{1}{(k+1)!} \sum_{j=1}^{n+q} \left[\sum_{i=0}^{\lfloor k/2 \rfloor} (k-i) a_{k,i} \chi_j^{k-2i+1} \psi_j^i + \sum_{i=1}^{\lfloor k/2 \rfloor + 1} (k-2i+2) a_{k,i-1} \chi_j^{k-2i+1} \psi_j^i \right].$$

From this the following recursive formula can be derived:

$$\begin{aligned} &a_{1,0} = 1, \quad a_{1,i} = 0, \quad i \neq 0, \\ &a_{k+1,i} = (k-i) a_{k,i} + (k-2i+2) a_{k,i-1}, \qquad k \geq 1. \end{aligned} \tag{16}$$

From this recursive scheme we derive an explicit formula for $a_{k,i}$:

$$a_{k,i} = \begin{cases} \frac{k!(k-i-1)!}{i!(k-2i)!2^i} & \text{if } 0 \leq i \leq \lfloor \frac{k}{2} \rfloor, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

We prove this formula by using induction. For $k = 1, 2$ our formula is certainly correct, as follows by inspection. From (16) it readily follows that $a_{k,0} = (k-1)!$, $k \geq 1$. This is in accordance with (17), since

$$(k-1)! = \frac{k!(k-1)!}{0!k!2^0}.$$

Now suppose that the formula is correct for some value of k, $k \geq 1$. Then, using (16) and (17) one has, for $i \geq 1$,

$$\begin{aligned} a_{k+1,i} &= (k-i)a_{k,i} + (k-2i+2)a_{k,i-1} \\ &= (k-i)\frac{k!(k-i-1)!}{i!(k-2i)!2^i} + (k-2i+2)\frac{k!(k-i)!}{(i-1)!(k-2i+2)!2^{i-1}} \\ &= \frac{(k+1)!(k-i)!}{i!(k-2i+1)!2^i}. \end{aligned}$$

This proves that formula (17) is correct indeed.

We proceed by deriving an upper bound for t_k. To this end we consider the following optimization problem:

$$\max \left(\frac{1}{k!} \sum_{j=1}^{n+q} \sum_{i=0}^{\lfloor k/2 \rfloor} a_{k,i} \chi_j^{k-2i} \psi_j^i \;:\; \sum_{j=1}^{n+q} (\chi_j^2 + \psi_j) = D^2,\ \chi_j \geq 0,\ \psi_j \geq 0,\ 1 \leq j \leq n \right),$$

where the maximization is done over χ_j and ψ_j. The nonnegativity of ψ_j is an obvious consequence of its definition; the nonnegativity of χ_j can be assumed, since the constraints are sign independent as far as χ_j is concerned, whereas positive values will give a larger objective than negative values. The Kuhn–Tucker optimality conditions for this problem are given by

$$\sum_{i=0}^{\lfloor k/2 \rfloor -1} (k-2i) a_{k,i} \chi_j^{k-2i-1} \psi_j^i + \left(\left\lfloor \frac{k+1}{2} \right\rfloor - \left\lfloor \frac{k}{2} \right\rfloor \right) a_{k,\lfloor k/2 \rfloor} \psi^{\lfloor k/2 \rfloor} \leq 2\alpha \chi_j, \tag{18}$$

$$\sum_{i=1}^{\lfloor k/2 \rfloor} i a_{k,i} \chi_j^{k-2i} \psi_j^{i-1} \leq \alpha, \tag{19}$$

$$\chi_j \left(\sum_{i=0}^{\lfloor k/2 \rfloor -1} (k-2i) a_{k,i} \chi_j^{k-2i-1} \psi_j^i + \left(\left\lfloor \frac{k+1}{2} \right\rfloor - \left\lfloor \frac{k}{2} \right\rfloor \right) a_{k,\lfloor k/2 \rfloor} \psi^{\lfloor k/2 \rfloor} - 2\alpha \chi_j \right) = 0, \tag{20}$$

$$\psi_j \left(\sum_{i=1}^{\lfloor k/2 \rfloor} i a_{k,i} \chi_j^{k-2i} \psi_j^{i-1} - \alpha \right) = 0, \tag{21}$$

where α is the multiplier.

From these conditions we will derive that either ψ_j or χ_j must be zero for each j. Assume that neither ψ_j nor χ_j is zero. From this we shall derive a contradiction. By multiplying (20) by ψ_j and (21) by $2\chi_j^2$ and subtracting, we derive

$$\sum_{i=0}^{\lfloor k/2 \rfloor -1} \left[(k-2i)a_{k,i} - 2(i+1)a_{k,i+1} \right] \chi_j^{k-2i} \psi_j^{i+1} + \left(\left\lfloor \frac{k+1}{2} \right\rfloor - \left\lfloor \frac{k}{2} \right\rfloor \right) a_{k,\lfloor k/2 \rfloor} \chi_j \psi^{\lfloor k/2 \rfloor +1} = 0.$$

It is easy to see that $\lfloor \frac{k+1}{2} \rfloor - \lfloor \frac{k}{2} \rfloor$ is 1 if k is odd, and 0 if k is even. Furthermore, let

$$\alpha_{k,i} := (k-2i)a_{k,i} - 2(i+1)a_{k,i+1}.$$

We easily obtain that $\alpha_{k,0} = 0$ and $\alpha_{k,i} > 0$, for $i \geq 1$, by using the general formula (17) for $a_{k,i}$:

$$\begin{aligned}\alpha_{k,i} &= (k-2i)\frac{k!(k-i-1)!}{i!(k-2i)!2^i} - 2(i+1)\frac{k!(k-i-2)!}{(i+1)!(k-2i-2)!2^{i+1}} \\ &= \frac{k!(k-i-2)!}{i!(k-2i-2)!2^i}\left[\frac{k-i-1}{k-2i-1} - 1\right] \\ &> 0.\end{aligned}$$

Hence, we have obtained a contradiction. This means that either χ_j or ψ_j is zero in the maximum. Consequently, the objective function either consists of "pure" χ_j terms or "pure" ψ_j terms.

Now if k is even, it easily follows that the largest objective value is obtained if $\psi_j = 0$, since, using the general formula (17), we have

$$a_{k,0} = 2^{(k/2)-1}a_{k,k/2},$$

which means that the coefficient of the pure χ_j term is greater than the coefficient of the pure ψ_j term. Now assume that $\psi_j > 0$ if k is odd. Then (19) holds with equality. By multiplying (19) by $2\chi_j$ and subtracting from (18), we derive

$$\sum_{i=0}^{\lfloor k/2 \rfloor - 1} \left[(k-2i)a_{k,i} - 2(i+1)a_{k,i+1}\right]\chi_j^{k-2i-1}\psi_j^i + a_{k,\lfloor k/2 \rfloor}\psi^{\lfloor k/2 \rfloor} = 0. \tag{22}$$

Again, we have obtained a contradiction, since we assumed that $\psi_j > 0$. Consequently, both for k even and k odd, the maximum is reached if $\psi_j = 0$. So an upper bound for the maximum of the function is

$$\frac{1}{k!}\sum_{j=1}^{n+q}(k-1)!\chi_j^k = \frac{1}{k}\sum_{j=1}^{n+q}\chi_j^k \leq \frac{1}{k}\left(\sum_{j=1}^{n+q}\chi_j^2\right)^{k/2} = \frac{1}{k}(D^2)^{k/2} = \frac{1}{k}D^k.$$

Hence the proof of (4) is complete.

REFERENCES

[1] D. DEN HERTOG, C. ROOS, AND T. TERLAKY (1991), *A potential reduction variant of Renegar's short-step path-following method for linear programming*, Linear Algebra Appl., 152, pp. 43–68.

[2] R. M. FREUND (1988), *Polynomial-time algorithms for linear programming based only on primal scaling and projected gradients of a potential function*, Working Paper OR 182–88, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA.

[3] C. C. GONZAGA (1989), *Large-steps path-following methods for linear programming, Part* I: *Barrier function method*, Report ES–210/89, Department of Systems Engineering and Computer Sciences, COPPE–Federal University of Rio de Janeiro, Rio de Janeiro, Brasil; also in SIAM J. Optimization, 1 (1991), pp. 268–279.

[4] ——— (1989), *Large-step path-following methods for linear programming, Part* II: *Potential reduction method*, Report ES–211/89, Department of Systems Engineering and Computer Sciences, COPPE–Federal University of Rio de Janeiro, Rio de Janeiro, Brasil; also in SIAM J. Optimization, 1 (1991), pp. 280–292.

[5] F. JARRE (1989), *The method of analytic centers for smooth convex programs*, Ph.D. thesis, Institut für Angewandte Mathematik und Statistik, Universität Würzburg, Würzburg, Germany.

[6] N. KARMARKAR (1984), *A new polynomial-time algorithm for linear programming*, Combinatorica, 4, pp. 373–395.

[7] M. KOJIMA, S. MIZUNO, AND A. YOSHISE (1991), *An $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems*, Math. Programming, 50, pp. 331–342.

[8] S. MEHROTRA AND J. SUN (1988), *An interior point algorithm for solving smooth convex programs based on Newton's method*, Tech. Report 88-08, Department of Industrial Engineering and Management Science, Northwestern University, Evaston, IL.

[9] C. ROOS AND J.-PH. VIAL (1990), *Long steps with the logarithmic penalty barrier function in linear programming*, in Economic Decision-Making: Games, Economics and Optimization, dedicated to Jacques H. Drèze; J. Gabszevwicz, J.-F. Richard, and L. Wolsey, eds., Elsevier Science Publisher B.V., Amsterdam, pp. 433–441.

[10] P. M. VAIDYA (1990), *An algorithm for linear programming which requires $O(((m+n)n^2 + (m+n)^{1.5}n)L)$ arithmetic operations*, Math. Programming, 47, pp. 175–201.

[11] PH. WOLFE (1961), *A duality theorem for nonlinear programming*, Quart. Appl. Math., 19, pp. 239–244.

[12] Y. YE (1991), *An $O(n^3L)$ potential reduction algorithm for linear programming*, Math. Programming, 50, pp. 239–258.