MATHEMATICS OF OPERATIONS RESEARCH
Vol 15, No 3, August 1990
Printed in U S A

AN EXTENSION OF KARMARKAR TYPE ALGORITHM TO A CLASS OF CONVEX SEPARABLE PROGRAMMING PROBLEMS WITH GLOBAL LINEAR RATE OF CONVERGENCE*†

RENATO D. C. MONTEIRO‡ AND ILAN ADLER§

We describe a primal-dual interior point algorithm for a class of convex separable programming problems subject to linear constraints. Each iteration updates a penalty parameter and finds a Newton step associated with the Karush-Kuhn-Tucker system of equations which characterizes a solution of the logarithmic barrier function problem for that parameter. It is shown that the duality gap is reduced at each iteration by a factor of $(1 - \delta/\sqrt{n})$, where δ is positive and depends on some parameters associated with the objective function.

1. Introduction. In Monteiro and Adler [11], an interior point primal-dual algorithm to solve convex quadratic programming problems has been presented which converges in $O(\sqrt{n}L)$ iterations with an average number of $O(n^{2.5})$ arithmetic operations per iteration. The present work discusses a variation of this algorithm which solves a class of separable convex programming problems subject to linear constraints.

Recently, with the advent of the new interior point algorithm by Karmarkar [6] for solving linear programming problems, some attention has been devoted to study classes of problems that can be solved by interior point algorithms in polynomial time.

The algorithm discussed in this paper is based on the logarithmic barrier function method and on the idea of following the path of minimizers for the logarithmic barrier family of problems, that is, the so called "central path". This path has been extensively studied in Bayer and Lagarias [1] and Meggido [9]. The logarithmic barrier function approach is usually attributed to Frisch [3] and is formally studied in Fiacco and McCormick [2] in the context of nonlinear optimization. Algorithms for linear programming problems based on following the central path have been presented in [12], [13], [5], [7] and [10]. All of these algorithms were shown to converge in polynomial number of arithmetic operations. Subsequently, Monteiro and Adler [11] developed an extension of the techniques in their previous paper [10] and presented an algorithm to solve convex quadratic programming problems which achieves a complexity of $O(\sqrt{n}L)$ iterations, in the worst case, with an average computational effort per iteration of $O(n^{2.5})$ arithmetic operations, similar to the linear programming case presented in [10]. A similar result was obtained by Kojima, Mizuno and Yoshise [8]. The present work shows that an approach similar to the one used in [10] and [11] can also be applied to solve some class of separable convex programming

*Received November 27, 1987; revised December 4, 1988.
AMS 1980 subject classification. Primary: 90C25.
IAOR 1973 subject classification. Main: Programming: Convex.
OR/MS Index 1978 subject classification. Primary: 654 Programming/Nonlinear/algorithms.
Key words. Interior-point methods, convex separable programming, Karmarkar's algorithm, global rate of convergence, barrier function, path following.
†Research partially funded by the Brazilian Post-Graduate Education Federal Agency (CAPES) and by the United States Navy Office of Naval Research under contract N00014-87-K-0202.
‡AT & T Bell Laboratories.
§University of California, Berkeley.

0364-765X/90/1503/0408/$01.25

problems subject to linear constraints. An interesting feature of the algorithm presented in this paper is that it has global linear rate of convergence, for this particular class of problems. We show that the algorithm improves the value of the objective function at every iteration by at least a factor of $(1 - \delta/\sqrt{n})$, where δ is positive and depends on some parameters associated with the objective function.

Our paper is organized as follows. In §2, the problem is introduced and some theoretical background is discussed. In §3, we present the algorithm. In §4, we prove results related to the convergence properties of the algorithm. In §5, we discuss how to initialize the algorithm. In §6, we conclude the paper with some remarks.

2. Problem description and some theoretical background. In this section, we introduce the class of problems which this paper is concerned with. We also review some results pertinent to the present work. A detailed discussion of these results can be found in [2] and [9]. We consider the following separable programming problem.

$$(P) \qquad \min \sum_{j=1}^{n} \psi_j(x_j) \quad \text{s.t.} \quad Ax = b, \quad x \geqslant 0,$$

where $x = (x_1, \ldots, x_n)^T$ is an n-vector, b is an m-vector, A is an $m \times n$ matrix. The functions ψ_j, $j = 1, \ldots, n$ are assumed to belong to the collection of functions Y defined as follows.

DEFINITION 2.1. Let I be a nonempty open subset of $\mathbf{R}$ which contains the interval $(0, \infty)$. We say that a function $\psi: I \to \mathbf{R}$ belongs to the collection Y if ψ satisfies the following:

(a) $\psi \in C^3$.

(b) If 0 is not in the set I then $\lim_{x \to o^+} \psi(x) = \infty$.

(c) There exist real numbers $M \geqslant 0$ and $p \geqslant 0$ such that for all reals $x > 0$ and $y > 0$, we have

$$(2.1) \qquad y|\psi'''(y)| \leqslant M \max\left\{\left(\frac{x}{y}\right)^p, \left(\frac{y}{x}\right)^p\right\}\psi''(x).$$

Observe that if condition (c) holds with M and p, it holds with any $M_0 \geqslant M$ and $p_0 \geqslant p$. From the definition above, it follows that, if $\psi \in$ Y then $\psi''(x) \geqslant 0$ for all real number $x > 0$. Also, if $\psi''(x) = 0$ for some real $x > 0$ then ψ is of the form $\psi(x) = dx + f$ for some constants d and f. Therefore, excepting this last case, $\psi \in$ Y is always a strictly convex function over the positive real numbers. In any case, the objective function of problem (P) is convex over the positive orthant. Thus, problem (P) enjoys all the properties of a convex programming type problem (see Proposition 2.2 below).

We now give some examples of functions which belong to the collection Y.

(1) $\psi(x) = dx + f$, for any real numbers d and f, and with $M = 0$ and $p = 0$.

(2) $\psi(x) = -\log x$, for $x > 0$, with $M = 2$ and $p = 2$.

(3) $\psi(x) = x^d$ for some real d, such that $d < 0$ or $d > 1$, and with $M = |d - 2|$ and $p = |d - 2|$. Note that, when $d < 0$, this function is only defined for $x > 0$.

(4) $\psi(x) = -x^d$, $x \geqslant 0$, for some real $0 < d < 1$, and with $M = |d - 2|$ and $p = |d - 2|$.

(5) Any positive linear combination of functions in the collection Y belongs to Y.

Since each function ψ_j in the objective function of problem (P) is in Y, there exists, for each $j = 1, \ldots, n$, a pair of constants $M_j \geqslant 0$ and $p_j \geqslant 0$ such that ψ_j, M_j and p_j

satisfy relation (2.1). If we let

$$M = \max_{j=1,\ldots,n} M_j, \quad p = \max_{j=1,\ldots,n} p_j, \tag{2.2}$$

then each function ψ_j, together with M and p, still satisfies relation (2.1). This observation will be used later in §4.

In order to simplify the notation in what follows, we let $\Psi(x) = \sum_{j=1}^{n} \psi_j(x_j)$. So if I_j denotes the domain of definition of the function ψ_j, $j = 1, \ldots, n$, the domain of the function Ψ is just the Cartesian product $\prod_{j=1}^{n} I_j$. The gradient of the function Ψ at the point $x = (x_1, \ldots, x_n)^T$ is given by the vector $\nabla\Psi(x) \equiv (\psi_1'(x_1), \ldots, \psi_n'(x_n))^T$ and the Hessian of Ψ at x is given by the diagonal matrix $\nabla^2\Psi(x) \equiv \operatorname{diag}(\psi_1''(x_1), \ldots, \psi_n''(x_n))$.

The Lagrangian dual problem corresponding to problem (P) is defined by

$$(D) \qquad \begin{aligned} \max \quad & \Psi(x) - \nabla\Psi(x)^T x + b^T y \\ \text{s.t.} \quad & -\nabla\Psi(x) + A^T y + z = 0, \\ & z \geqslant 0, \end{aligned}$$

where x and z are n-vectors and y is an m-vector. The relationship between problems (P) and (D) is provided by the following result known as the duality theorem for convex programming problems.

PROPOSITION 2.2. (a) *If problem (P) is unbounded then problem (D) is infeasible. If problem (D) is unbounded then problem (P) is infeasible.*

(b) *If problem (P) has an optimal solution x^0 then there exist y^0 and z^0 such that the point $(x, y, z) = (x^0, y^0, z^0)$ is an optimal solution of problem (D). Moreover, the optimal values of both problems are identical.*

We define the set of interior feasible solutions of problems (P) and (D) as

$$S \equiv \{x \in \mathbf{R}^n;\ Ax = b,\ x > 0\},$$

$$T \equiv \{(x, y, z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n;\ -\nabla\Psi(x) + A^T y + z = 0,\ z > 0\},$$

respectively, and let

$$W \equiv \{(x, y, z);\ (x, y, z) \in T,\ x \in S\}.$$

For a point $w = (x, y, z) \in W$, we define its duality gap $g(w)$ to be $g(w) = \nabla\Psi(x)^T x - b^T y$ which is simply the value of the objective function of problem (P) at x minus the value of the objective function of problem (D) at (x, y, z). From the definition of the set W, one can easily verify that if $w \in W$ then $g(w) = x^T z$.

We impose the following assumptions:

ASSUMPTION 2.3. (a) The set S is nonempty and bounded.

(b) $\operatorname{rank}(A) = m$.

Condition (b) of Definition 2.1 and the fact that the set S is bounded imply that problem (P) has an optimal solution. Hence, by Proposition 2.2, problem (D) also has an optimal solution.

The algorithm we consider in this paper is motivated by the application of the logarithmic barrier function technique to problem (P). The logarithmic barrier function method consists of examining the family of problems

$$(P_\mu) \qquad \min \Psi(x) - \mu \sum_{j=1}^{n} \ln x_j$$

$$\text{s.t.} \quad Ax = b,$$

$$x > 0,$$

where $\mu > 0$ is the barrier penalty parameter. This technique is well known in the context of general constraint optimization problems. One solves the problem penalized by the logarithmic barrier function term for several values of the parameter μ, with μ decreasing to zero, and the result is a sequence of feasible points converging to a "solution" of the original problem. The interested reader can refer to [2] for a detailed discussion of this technique in the context of nonlinear constrained optimization. Recently this method was first reconsidered in Gill, Murray, Saunders, Tomlin and Wright [4] where a similarity with Karmarkar's algorithm is discussed. A comprehensive analysis of the logarithmic barrier function approach as applied to linear programming and linear complementary problems with positive semi-definite matrices is given in [9].

Throughout this paper, we use the following notation. If $x = (x_1, \ldots, x_n)^T$ is an n-vector, then the corresponding capital letter X denotes the diagonal matrix $\mathrm{diag}(x_1, \ldots, x_n)$. Observe that the objective function of problem (P_μ) is a strictly convex function. This implies that problem (P_μ) has at most one global minimum, and that this global minimum, if it exists, is completely characterized by the Karush-Kuhn-Tucker stationary condition:

$$\nabla\Psi(x) - \mu X^{-1}e - A^T y = 0,$$

$$Ax = b, \quad x > 0$$

where e denotes the n-vector of ones and y is the vector of Lagrangian multiplier associated with the equality constraints of problem (P_μ). By introducing the n-vector z, this system can be rewritten in an equivalent way as

$$(2.3.a) \qquad ZXe - \mu e = 0,$$

$$(2.3.b) \qquad Ax = b, \quad x > 0,$$

$$(2.3.c) \qquad -\nabla\Psi(x) + A^T y + z = 0.$$

It turns out that, under Assumption 2.3, problem (P_μ) (and consequently system (2.3)) has a solution for all $\mu > 0$ (c.f. [2] and [9]). The Karush-Kuhn-Tucker system (2.3) provides important information which we now point out. Assume that $\mu > 0$ is fixed in the system (2.3). Since $x > 0$, equation (2.3.a) implies that $z > 0$. Equation (2.3.c) then implies that the triple (x, y, z) is an interior feasible solution for the dual problem (D). From (c) of Assumption 2.3, it follows that there is a unique y satisfying (2.3). We denote the unique triple that satisfies (2.3) by $w(\mu) = (x(\mu), y(\mu), z(\mu))$.

At this point, we observe that (a) of Assumption 2.3 implies that the set W, and therefore the set T, is nonempty. Indeed, $w(\mu) \in W$, for all $\mu > 0$. From (2.3.a), it follows that the duality gap at $w(\mu)$ is given by $g(w(\mu)) = n\mu$ for all $\mu > 0$ and therefore $g(w(\mu))$ converges to zero as μ approaches zero. This implies that the objective function value of problem (P) at $x(\mu)$ and the objective function value of problem (D) at $(x(\mu), y(\mu), z(\mu))$ converge to the common optimal value of problems (P) and (D). In fact, we have the following stronger result (cf. [2] and [9]).

PROPOSITION 2.4. *Under Assumption* 2.3, *as* $\mu \to 0$, $x(\mu)$ *and* $(x(\mu), y(\mu), z(\mu))$ *converge to optimal solutions of problems* (P) *and* (D) *respectively.*

The following notation will be useful later. If $w = (x, y, z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n$ then $f(w) = (f_1(w), \dots, f_n(w))^T \in \mathbf{R}^n$ denotes the n-vector XZe, that is, $f_j(w) = x_j z_j$, $j = 1, \dots, n$. We denote by Γ the set (or path) of solutions $w(\mu)$, $\mu > 0$, for the system (2.3), i.e.,

$$\Gamma = \{w(\mu) \equiv (x(\mu), y(\mu), z(\mu)); \mu > 0\}.$$

The algorithm which will be presented in the next section is based on the idea of following this path Γ closely. The path Γ will serve as a criterion to guide the points generated by the algorithm.

3. The algorithm. The algorithm presented in this section extends the one presented in [10] for linear programming problems and in [11] for convex quadratic programming problems. We also refer the reader to [7] and [10] for a motivation of the directions generated by the algorithm that we now describe.

In order to motivate the idea behind the algorithm, we need to introduce a definition.

DEFINITION 3.1. Let θ with $0 \leqslant \theta < 1$ be given. We say that a point $w \in W$ is θ-centered with respect to $\mu > 0$ if $\|f(w) - \mu e\| \leqslant \theta\mu$ where $\|.\|$ denotes the Euclidean norm.

Obviously, the solution $w(\mu)$ of system (2.3) is θ-centered with respect to μ for any $\theta \geqslant 0$. We can view θ-centered points $w \in W$ as points that are close to the central path Γ, where the criterion of closeness is given according to Definition 3.1. The following trivial observation is important for our purposes.

Observation 3.2. If $w \equiv (x, y, z) \in W$ is θ-centered with respect to the parameter $\mu > 0$ then $g(w) \equiv x^T z \leqslant (1 + \theta)n\mu$.

The algorithm described in this paper generates a sequence of points (w^k) in the set W and a strictly decreasing sequence of positive parameters (μ_k) converging to 0 such that w^k is θ-centered with respect to μ_k for all $k \geqslant 1$. In view of Observation 3.2, we would like the sequence of parameters (μ_k) to approach 0 as fast as possible. However, as we will see later, the speed of convergence of the sequence of parameters (μ_k) is constrained by the condition that the sequence (w^k) be θ-centered.

We now describe how the algorithm iterates. Given a point $w \in W$ θ-centered with respect to $\mu > 0$, we want to construct a point $\hat{w} \in W$ which is θ-centered with respect to a smaller parameter $\hat{\mu} > 0$ (the values of θ and of the ratio $\hat{\mu}/\mu$ will be specified later in this section). To determine a point $\hat{w} \in W$ with this property, consider the direction $\Delta w \equiv (\Delta x, \Delta y, \Delta z) \in \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n$ determined by the following system of linear equations

$$Z\,\Delta x + X\,\Delta z = XZe - \hat{\mu}e, \tag{3.1.a}$$

$$A\,\Delta x = 0, \tag{3.1.b}$$

$$-\nabla^2\Psi(x)\,\Delta x + A^T\,\Delta y + \Delta z = 0. \tag{3.1.c}$$

The direction Δw is just the Newton direction for the system (2.3) with $\mu = \hat{\mu}$. After some algebra, one obtains the following expressions for Δx, Δy and Δz:

$$\Delta x = (Z + XQ)^{-1}\left[I - XA^T\left(A(Z+XQ)^{-1}XA^T\right)^{-1}A(Z+XQ)^{-1}\right](XZe - \hat{\mu}e),$$

$$\Delta y = -\left[\left(A(Z+XQ)^{-1}XA^T\right)^{-1}A(Z+XQ)^{-1}\right](XZe - \hat{\mu}e),$$

$$\Delta z = Q\,\Delta x - A^T\Delta y$$

where $Q \equiv \nabla^2\Psi(x)$ denotes the diagonal Hessian matrix of the function Ψ at x. After computing the direction Δw according to system (3.1), the next iterate $\hat{w} \equiv (\hat{x}, \hat{y}, \hat{z})$ is determined as follows:

(3.2.a) $$\hat{x} = x - \Delta x,$$

(3.2.b) $$\hat{y} = y - \Delta y,$$

(3.2.c) $$\hat{z} = \nabla\Psi(\hat{x}) - A^T\hat{y}.$$

Observe that z is defined to be the slack vector for the dual problem (D) corresponding to $\hat{x}$ and $\hat{y}$. The following relation expresses the vector $\hat{z}$ in terms of the vector z and the direction Δz:

(3.3) $$\hat{z} = z - \Delta z + \nabla\Psi(\hat{x}) - \nabla\Psi(x) + \nabla^2\Psi(x)\,\Delta x.$$

This expression is verified by using (3.1.c) and (3.2.c). We denote the direction $\Delta w = (\Delta x, \Delta y, \Delta z)$ determined by the system (3.1) and the point $\hat{w} = (\hat{x}, \hat{y}, \hat{z})$ given by (3.2) as $\Delta w(w, \hat{\mu})$ and $\hat{w}(w, \hat{\mu})$ respectively, in order to indicate their dependence on the point $w = (x, y, z)$, and on the penalty parameter $\hat{\mu}$.

We are now ready to give a detailed description of the algorithm. Let $\theta > 0$ be a constant defined as

(3.4) $$\theta = \min\left(\frac{1}{2(1+M)}, \frac{1}{4(p+1)}, \frac{1}{8}\right)\cdot$$

where M and p are defined by (2.2). At the beginning of the algorithm, we assume that an initial point $w^0 \equiv (x^0, y^0, z^0) \in W$ θ-centered with respect to a positive parameter μ_0 is given, that is,

(3.5) $$\left\|f(w^0) - \mu_0 e\right\| \leqslant \theta\mu_0.$$

Let $\delta > 0$ be defined as

(3.6) $$\delta = \theta - 2\theta^2.$$

Note that δ is a positive real number. Indeed,

(3.7) $$\tfrac{3}{4}\theta \leqslant \delta \leqslant \theta$$

by (3.4) and (3.6). For later reference, we also note that θ and δ satisfy

(3.8) $$\frac{\theta + \delta}{1 - \theta} = 2\theta.$$

We now state the algorithm.

ALGORITHM 3.3.

Step 0. *Let θ and δ be as in (3.4) and (3.6). Let $w^0 \in W$ and $\mu_0 > 0$ satisfy (3.5). Let ϵ be a given tolerance for the duality gap. Set $k := 0$.*

Step 1. *If $(x^k)^T z^k \leqslant \epsilon$, stop.*

Step 2. *Set $\mu_{k+1} := \mu_k(1 - \delta/\sqrt{n})$.*
Calculate $\Delta w^k \equiv \Delta w(w^k, \mu_{k+1})$.
Set $w^{k+1} := \hat{w}(w^k, \mu_{k+1})$.
Set $k := k + 1$ and go to Step 1.

end of algorithm.

In the following sections, we prove that all points generated by Algorithm 3.3 are in the set W and that the kth iterate w^k is θ-centered with respect to the parameter μ_k for all $k \geqslant 1$. We also show that Algorithm 3.3 terminates in at most $O(\sqrt{n}\ \max(\log \epsilon^{-1}, \log n, \log \mu_0))$ iterations.

4. Convergence results. In this section, we present convergence results for the algorithm described in §3. The convergence results presented in this section are similar to the ones presented in [10] and [11] for linear programming problems and quadratic programming problems respectively.

Let $w = (x, y, z) \in W$ and $\hat{\mu} > 0$. Let $\Delta w = (\Delta x, \Delta y, \Delta z)$ denote the direction $\Delta w(w, \hat{\mu})$ and let $\hat{w} = (\hat{x}, \hat{y}, \hat{z})$ denote the point $\hat{w}(w, \hat{\mu})$. The next result provides expressions for the products of complementary variables $f_j(\hat{w}) = \hat{x}_j \hat{z}_j$, $j = 1, \dots, n$.

PROPOSITION 4.1. *Let $w \in W$ and $\hat{\mu} > 0$ be given and consider Δw and $\hat{w}$ as above. Then the following expressions hold*:

$$f_j(\hat{w}) = \hat{\mu} + \Delta x_j \Delta z_j + (x_j - \Delta x_j) \times \left[\psi_j'(x_j - \Delta x_j) - \psi_j'(x_j) + \psi_j''(x_j)\,\Delta x_j\right], \tag{4.1}$$

$$(\Delta x)^T(\Delta z) = \sum_{j=1}^{n} \psi_j''(x_j)(\Delta x_j)^2 \geqslant 0. \tag{4.2}$$

PROOF. By definition of $f_j(\hat{w})$ and expression (3.3), we have for all $j = 1, \dots, n$

$$\begin{aligned} f_j(\hat{w}) &= \hat{x}_j \hat{z}_j \\ &= (x_j - \Delta x_j)\left[z_j - \Delta z_j + \psi_j'(x_j - \Delta x_j) - \psi_j'(x_j) + \psi_j''(x_j)\,\Delta x_j\right] \\ &= x_j z_j - (x_j \Delta z_j + z_j \Delta x_j) + \Delta x_j \Delta z_j \\ &\quad + (x_j - \Delta x_j)\left[\psi_j'(x_j - \Delta x_j) - \psi_j'(x_j) + \psi_j''(x_j)\,\Delta x_j\right] \\ &= \hat{\mu} + \Delta x_j \Delta z_j \\ &\quad + (x_j - \Delta x_j)\left[\psi_j'(x_j - \Delta x_j) - \psi_j'(x_j) + \psi_j''(x_j)\,\Delta x_j\right] \end{aligned}$$

where the last inequality follows from (3.1.a). This shows (4.1). Multiplying expressions (3.1.b) and (3.1.c) on the left by $(\Delta y)^T$ and $(\Delta x)^T$ respectively, and combining, we obtain the equality part in (4.2). Noting that $\psi_j''(x_j) \geqslant 0$ for all $x_j > 0$ and all $j = 1, \dots, n$, we obtain the inequality part in (4.2). This completes the proof of the proposition. ∎

We now state and prove a result that provides bounds necessary to show that the points generated by Algorithm 3.3 are feasible and remain close to the path Γ.

Let $w = (x, y, z) \in W$ and $\hat{\mu} > 0$. Let $\Delta w = (\Delta x, \Delta y, \Delta z)$ be the direction $\Delta w(w, \hat{\mu})$. Let $\Delta f \equiv (\Delta X)(\Delta Z)e$, where ΔX and ΔZ are the diagonal matrices corresponding to the vectors Δx and Δz respectively. The next result provides an upper bound on the Euclidean norm of the vector Δf.

LEMMA 4.2. *Let Δf be defined as above. Then, we have*

$$\|\Delta f\| \leqslant \frac{\|f(w) - \hat{\mu}e\|^2}{2f_{\min}} \quad \textit{where} \tag{4.3}$$

$$f_{\min} \equiv \min\{x_j z_j;\ j = 1, \ldots, n\}.$$

Furthermore, we have

$$\sum_{j=1}^{n} \psi_j''(x_j)(\Delta x_j)^2 \leqslant \frac{\|f(w) - \hat{\mu}e\|^2}{2f_{\min}}, \tag{4.4}$$

$$\|D^{-1}\Delta x\|^2 + \|D\Delta z\|^2 \leqslant \frac{\|f(w) - \hat{\mu}e\|^2}{f_{\min}} \tag{4.5}$$

where D is the diagonal matrix defined by

$$D = (Z^{-1}X)^{1/2}. \tag{4.6}$$

PROOF. By equation (3.1.a), we have

$$D^{-1}\Delta x + D\Delta z = (XZ)^{-1/2}(XZ - \hat{\mu}e).$$

It then follows that

$$\|D^{-1}\Delta x + D\Delta z\|^2 = \|(XZ)^{-1/2}(XZe - \hat{\mu}e)\|^2 \tag{4.7}$$

$$= \sum_{j=1}^{n} \frac{(f_j(w) - \hat{\mu})^2}{x_j z_j} \leqslant \frac{\|f(w) - \hat{\mu}e\|^2}{f_{\min}}.$$

On the other hand, we have

$$\|D^{-1}\Delta x + D\Delta z\|^2 = \|D^{-1}\Delta x\|^2 + 2(\Delta x)^T(\Delta z) + \|D\Delta z\|^2. \tag{4.8}$$

Inequality (4.5) follows immediately from (4.7) and (4.8) and the inequality in (4.2). Similarly

$$\sum_{j=1}^{n} \psi_j''(x_j)(\Delta x_j)^2 = \Delta x^T \Delta z \leqslant \frac{\|f(w) - \hat{\mu}e\|^2}{2f_{\min}} \tag{4.9}$$

which proves (4.4). Using the Cauchy-Schwarz inequality, we obtain

$$\|\Delta f\| \leqslant \sum_{j=1}^{n} |\Delta x_j \Delta z_j| \tag{4.10}$$

$$= \sum_{j=1}^{n} |D_{jj}^{-1} \Delta x_j| \, |D_{jj} \Delta z_j|$$

$$\leqslant \|D^{-1} \Delta x\| \, \|D \Delta z\|.$$

Relations (4.10) and (4.5) then imply (4.3). This completes the proof of the lemma. ■

The next lemma provides important relations that will be useful in the proof of the main theorem.

LEMMA 4.3. *Suppose $w = (x, y, z) \in W$ and $\mu > 0$ satisfy*

$$\|f(w) - \mu e\| \leqslant \theta\mu \tag{4.11}$$

where θ is a positive constant such that $\theta < 1$. Let $\Delta w \equiv (\Delta x, \Delta y, \Delta z)$ be the direction $\Delta w(w, \hat{\mu})$ where $\hat{\mu} = \mu(1 - \delta/\sqrt{n}) > 0$ with $\delta > 0$. Then the following relations hold:

$$\|\Delta f\| \leqslant \frac{(\theta + \delta)^2 \mu}{2(1 - \theta)}, \tag{4.12}$$

$$\frac{|\Delta x_j|}{x_j} \leqslant \frac{(\theta + \delta)}{(1 - \theta)}, \qquad j = 1, \ldots, n, \tag{4.13}$$

$$\sum_{j=1}^{n} \psi_j''(x_j)(\Delta x_j)^2 \leqslant \frac{(\theta + \delta)^2 \mu}{2(1 - \theta)}. \tag{4.14}$$

PROOF. Relation (4.11) implies that

$$f_{\min} = \min\{x_j z_j\} \geqslant (1 - \theta)\mu. \tag{4.15}$$

Since $\|e\| = \sqrt{n}$ and $\mu - \hat{\mu} = (\delta/\sqrt{n})\mu$, it follows from relation (4.11) that

$$\|f(w) - \hat{\mu} e\|^2 \leqslant (\|f(w) - \mu e\| + \|\mu e - \hat{\mu} e\|)^2 \tag{4.16}$$

$$\leqslant (\theta\mu + (\mu - \hat{\mu})\|e\|)^2$$

$$\leqslant (\theta\mu + \delta\mu)^2$$

$$\leqslant (\theta + \delta)^2 \mu^2.$$

Using Lemma 4.2 and the two previous relations, we immediately obtain (4.12), (4.14) and the following inequality.

$$\|D^{-1} \delta x\|^2 \leqslant \frac{(\theta + \delta)^2 \mu}{(1 - \theta)} \tag{4.17}$$

where the matrix D is given by (4.6). To show (4.13), observe that for all $j = 1, \ldots, n$, we have

$$\left(\frac{\Delta x_j}{x_j}\right)^2 \leqslant \sum_{k=1}^{n} \left(\frac{\Delta x_k}{x_k}\right)^2 = \sum_{k=1}^{n} \frac{\left(D_{kk}^{-1}\,\Delta x_k\right)^2}{x_k z_k}$$

$$\leqslant \frac{\|D^{-1}\,\Delta x\|^2}{(1-\theta)\mu} \leqslant \left(\frac{\theta+\delta}{1-\theta}\right)^2$$

where in the last two inequalities, we used relations (4.15) and (4.17) respectively. This completes the proof of the lemma. ∎

The following lemma provides an inequality that will be useful in the proof of the main theorem. We leave the reader to verify its validity.

LEMMA 4.4. *Let p be a real number such that $p \geqslant 1$. Then for all real numbers u, with $0 \leqslant u \leqslant 1$, the inequality $(1-u)^p \geqslant 1 - pu$ holds.*

We now state the key result that will enable us to show the convergence of Algorithm 3.3. The following theorem shows the behavior of one iteration of Algorithm 3.3.

THEOREM 4.5. *Let $\theta > 0$ and $\delta > 0$ satisfy relations* (3.4) *and* (3.6) *respectively. Let $w = (x, y, z) \in W$ and $\mu > 0$ satisfy $\|f(w) - \mu e\| \leqslant \theta\mu$. Let $\hat{w} = (\hat{x}, \hat{y}, \hat{z})$ denote the point $\hat{w}(w, \hat{\mu})$ where $\hat{\mu} > 0$ is defined as*

$$\hat{\mu} = \mu\left(1 - \delta/\sqrt{n}\,\right). \tag{4.18}$$

Then the following statements hold:

(a) *$\hat{w} \in W$ and*

$$\|f(\hat{w}) - \hat{\mu} e\| \leqslant \theta\hat{\mu}. \tag{4.19}$$

(b) *$g(\hat{w}) \equiv \hat{x}^T\hat{z} \leqslant (1+\theta)n\hat{\mu}$.*

PROOF. (a) We first show that $\hat{w} = (\hat{x}, \hat{y}, \hat{y})$ is in the set W (cf. §2). Since $w \in W$, $Ax = b$. From relation (3.1.b) and the definition of $\hat{x}$ given by (3.2.a), it follows that $A\hat{x} = b$. From the way $\hat{z}$ was defined in relation (3.2.c), it follows that $-\nabla\Psi(\hat{x}) + A^T\hat{y} + \hat{z} = 0$. Therefore, we only need to show that the vectors $\hat{x}$ and $\hat{z}$ are strictly positive to conclude that $\hat{w} \in W$. Observe that by expressions (3.8) and (4.13)

$$|\Delta x_j| \leqslant 2\theta x_j \tag{4.20}$$

for all $j = 1, \ldots, n$. Therefore,

$$\hat{x}_j = x_j - \Delta x_j \geqslant x_j - |\Delta x_j| = (1 - 2\theta)x_j$$

for all $j = 1, \ldots, n$. Since $\theta \leqslant 1/8$ by (3.4) and $x_j > 0$, the last relation implies that $\hat{x}_j > 0$, for all $j = 1, \ldots, n$. Assume inequality (4.19), $\hat{x}_j\hat{z}_j \geqslant (1-\theta)\hat{\mu} > 0$. Therefore, we must have $\hat{z}_j > 0$, for all $j = 1, \ldots, n$.

We observe that the proof of (4.19) does not depend on the condition that $\hat{z} > 0$. We now show (4.19). If let $r = (r_1, \ldots, r_n)^T \in \mathbf{R}^n$ denote the vector whose jth component is given by

$$r_j = (x_j - \Delta x_j)\left[\psi_j'(x_j - \Delta x_j) - \psi_j'(x_j) + \psi_j''(x_j)\,\Delta x_j\right] \tag{4.21}$$

then expression (4.1) and the triangle inequality for norms imply

$$\|f(\hat{w}) - \hat{\mu}e\| \leqslant \|\Delta f\| + \|r\|. \tag{4.22}$$

Our objective is to bound the right-hand side of the above inequality. By the mean value theorem, there exists a real t_j, $0 < t_j < 1$, such that

$$\xi_j \equiv (1 - t_j)x_j + t_j(x_j - \Delta x_j) = x_j - t_j \Delta x_j > 0$$

satisfies

$$\psi_j'(x_j - \Delta x_j) - \psi_j'(x_j) + \psi_j''(x_j)\Delta x_j = \tfrac{1}{2}\psi_j'''(\xi_j)(\Delta x_j)^2.$$

From the definition of r_j in (4.21), the previous expression and the fact that each function ψ_j is in the collection Y, it follows that

$$\begin{aligned} |r_j| &= \tfrac{1}{2}(\Delta x_j)^2(x_j - \Delta x_j)|\psi_j'''(\xi_j)| \\ &= \frac{1}{2}(\Delta x_j)^2\frac{(x_j - \Delta x_j)}{\xi_j}|\xi_j\psi_j'''(\xi_j)| \\ &\leqslant \frac{1}{2}(\Delta x_j)^2\frac{(x_j - \Delta x_j)}{\xi_j}M\psi_j''(x_j)\max\left\{\left(\frac{x_j}{\xi_j}\right)^p, \left(\frac{\xi_j}{x_j}\right)^p\right\} \end{aligned} \tag{4.23}$$

where M and p are as given in (2.2). Now, using (4.20), one can easily show that

$$\frac{(x_j - \Delta x_j)}{\xi_j} \leqslant 1 + 2\theta \leqslant \frac{1}{1 - 2\theta}, \quad \frac{x_j}{\xi_j} \leqslant \frac{1}{1 - 2\theta} \quad \text{and}$$

$$\frac{\xi_j}{x_j} \leqslant 1 + 2\theta \leqslant \frac{1}{1 - 2\theta}.$$

Using these three last estimates in relation (4.23), we obtain

$$|r_j| \leqslant \frac{\psi_j''(x_j)(\Delta x_j)^2 M}{2(1 - 2\theta)^{p+1}}. \tag{4.24}$$

Using expression (3.4) and Lemma 4.4, we obtain

$$(1 - 2\theta)^{p+1} \geqslant 1 - 2(p + 1)\theta \geqslant \tfrac{1}{2}$$

since $\theta \leqslant 1/4(p + 1)$. Using this last estimate in (4.24), we obtain, for all $j = 1, \ldots, n$, $|r_j| \leqslant M\psi_j''(x_j)(\Delta x_j)^2$. Summing up the above inequality over all $j = 1, \ldots, n$ and using relation (4.14), we obtain

$$\sum_{j=1}^{n} |r_j| \leqslant M\frac{(\theta + \delta)^2}{2(1 - \theta)}\mu.$$

Relations (4.12) and (4.22) together with the previous inequality then imply that

$$\begin{aligned}\|f(\hat{w}) - \hat{\mu}e\| &\leqslant \|\Delta f\| + \|r\| \\ &\leqslant \|\Delta f\| + \sum_{j=1}^{n} |r_j| \\ &\leqslant \frac{(\theta+\delta)^2}{2(1-\theta)}(1+M)\mu \\ &= 2\theta^2(1-\theta)(1+M)\mu\end{aligned}$$

where the equality is due to (3.8). Since $\theta \leqslant [2(1+M)]^{-1}$ by (3.4) and $\hat{\mu} = (1 - \delta/\sqrt{n})\mu \geqslant (1-\delta)\mu \geqslant (1-\theta)\mu$ by (4.18) and (3.7), the last relation implies that $\|f(\hat{w}) - \hat{\mu}e\| \leqslant \theta\hat{\mu}$ and this completes the proof of (a).

(b) Statement (b) follows from statement (a) and Observation 3.2. This completes the proof of the theorem. □

As a consequence of Theorem 4.5, we have the following corollary.

COROLLARY 4.6. *All points w^k generated by Algorithm* 3.3 *satisfy*

(a) *For all* $k = 1, 2, \ldots$, $w^k \in W$ *and* $\|f(w^k) - \mu_k e\| \leqslant \theta\mu_k$.

(b) *For all* $k = 1, 2, \ldots$, $g(w^k) \equiv (x^k)^T z^k \leqslant (1+\theta)n\mu_k$, *where* $\mu_k = \mu_0(1 - \delta/\sqrt{n})^k$.

PROOF. This result follows trivially by arguing inductively and using Theorem 4.5. ■

We now derive an upper bound on the total number of iterations performed by Algorithm 3.3. The following result follows easily from Corollary 4.6 and is proved in §4 of [10].

PROPOSITION 4.7. *The total number of iterations performed by Algorithm* 3.3 *is no greater than* $k^* \equiv \lceil \log((1+\theta)n\epsilon^{-1}\mu_0)\sqrt{n}/\delta \rceil$ *where* $\epsilon > 0$ *denotes the tolerance for the duality gap and* μ_0 *is the initial penalty parameter.*

If the constants M and p are both $O(1)$, which is usually the case, then by (3.4) and (3.7), δ^{-1} is $O(1)$. In this case, Proposition 4.7 asserts that Algorithm 3.3 terminates in at most $O(\sqrt{n}\max(\log \epsilon^{-1}, \log n, \log \mu_0))$ iterations.

5. Initialization of the algorithm. In order to initialize the algorithm, the approach is to use a transformed problem equivalent to original one that satisfies the initial condition (3.5). Therefore, by solving the transformed problem, we are able to obtain a solution for the original problem.

Consider the convex programming problem

$$(\tilde{P}) \qquad \begin{aligned} &\min \tilde{\Psi}(\tilde{x}) \\ \text{s.t. } &\tilde{A}\tilde{x} = \tilde{b}, \\ &\tilde{x} \geqslant 0, \end{aligned}$$

where $\tilde{A}$ is an $\tilde{m} \times \tilde{n}$ matrix which has full row rank and $\tilde{b}$ is a vector of length $\tilde{m}$. Assume that $\tilde{\Psi}(\tilde{x}) = \sum_{j=1}^{\tilde{n}} \tilde{\psi}(\tilde{x}_j)$ where each $\tilde{\psi}_j$ belongs to the collection Y (see Definition 2.1). We also assume that the set of feasible solutions of problem $(\tilde{P})$ is bounded. Therefore, $(\tilde{P})$ has an optimal solution.

Let $n = \tilde{n} + 2$ and $m = \tilde{m} + 1$. Let λ be a large constant satisfying $e^T\tilde{x} < n\lambda$ for all feasible solution $\tilde{x}$ of problem $(\tilde{P})$. The constant λ exists since the set of feasible solutions of $(\tilde{P})$ is bounded by assumption. Also, let $K > 0$ be a sufficiently large constant. Consider the transformed problem as follows.

$$(P) \qquad \min \tilde{\Psi}(\tilde{x}) + K\tilde{x}_n$$

$$\text{s.t.} \quad \tilde{A}\tilde{x} + (\lambda^{-1}\tilde{b} - \tilde{A}e)\tilde{x}_n = \tilde{b},$$

$$e^T\tilde{x} + \tilde{x}_{n-1} + \tilde{x}_n = n\lambda,$$

$$\tilde{x} \geqslant 0, \quad \tilde{x}_{n-1} \geqslant 0, \quad \tilde{x}_n \geqslant 0,$$

where $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_{n-2})^T$ is an $(n-2)$-vector and $\tilde{x}_{n-1}$ and $\tilde{x}_n$ are scalars. This problem can be cast in the notation of problem (P) of § 2 as follows. Let $x = (\tilde{x}^T, \tilde{x}_{n-1}, \tilde{x}_n)^T \in \mathbf{R}^n$ and define $b \in \mathbf{R}^m$ and $A \in \mathbf{R}^{m\times n}$ as follows.

$$b = \begin{pmatrix} \tilde{b} \\ n\lambda \end{pmatrix}, \quad A = \begin{bmatrix} \tilde{A} & 0 & \lambda^{-1}\tilde{b} - \tilde{A}e \\ e^T & 1 & 1 \end{bmatrix}.$$

Define $\Psi(x) = \tilde{\Psi}(\tilde{x}) + K\tilde{x}_n$. With this notation, we can then rewrite problem (P) as in §2.

Let $(\tilde{x}_*, \tilde{y}_*, \tilde{z}_*)$ denote an optimal solution of the dual problem corresponding to problem $(\tilde{P})$ such that $\tilde{x}_*$ is an optimal solution of $(\tilde{P})$. The following result provides a theoretical lower bound on the constant K in such a way that if K is larger than this bound then the optimal solutions of problem (P) immediately give optimal solutions for problem $(\tilde{P})$.

Proposition 5.1. *Assume that the cost coefficient K in the objective function of problem (P) satisfies $K > (\lambda^{-1}\tilde{b} - \tilde{A}e)^T\tilde{y}_*$. Then*

(1) *The optimal value of problem (P) is equal to the optimal value of problem $(\tilde{P})$.*

(2) *If $x = (x_1, \dots, x_n)^T$ is an optimal solution of problem (P) then $x_n = 0$. Moreover, $\tilde{x} = (x_1, \dots, x_{n-2})^T$ is an optimal solution of $(\tilde{P})$.*

The proof of Proposition 5.1 is straightforward and is left to the reader (see §5 of [11] for a proof of this proposition for the case the objective function of $(\tilde{P})$ is a convex quadratic function). We now verify that problem (P) satisfies Assumption 2.3 of §2. Condition (c) of Assumption 2.3 is obviously satisfied since $\tilde{A}$ was assumed to have full row rank. We verify conditions (a) and (b) of Assumption 2.3 jointly by exhibiting a point $w^0 = (x^0, y^0, z^0)$ which is in the set W defined in §2 and satisfying the criterion of closeness (3.5). Let $x^0 \equiv \lambda e \in \mathbf{R}^n$. Observe that $Ax^0 = b$. Let $y^0 = (0, \dots, 0, -\mu_0/\lambda)^T \in \mathbf{R}^m$ where μ_0 satisfies

$$\mu_0 \geqslant \frac{\lambda\|\nabla\Psi(\lambda e)\|}{\theta}. \tag{5.1}$$

Let $z^0 \in \mathbf{R}^n$ denote the slack vector $\nabla\Psi(x^0) - A^Ty^0$ for the dual (D) corresponding to the pair $x = x^0$ and $y = y^0$. Since $A^Ty^0 = -\lambda^{-1}\mu_0 e$, it follows that $z^0 = \nabla\Psi(\lambda e) + \lambda^{-1}\mu_0 e$. Then

$$\sum_{j=1}^{n} \left(x_j^0 z_j^0 - \mu_0\right)^2 = \lambda^2\|\nabla\Psi(\lambda e)\|^2$$

and therefore the criterion of closeness (3.5) is satisfied due to expression (5.1).

Observe that Algorithm 3.3, applied to problem (P), will obtain a feasible solution $x = (x_1, \ldots, x_n)$ to (P) whose value is within the optimal value by ϵ where ϵ is the specified tolerance for the duality gap. If we let $\tilde{x} = (x_1, \ldots, x_{n-2})$ then $\tilde{x}$ will not be a feasible solution to problem $(\tilde{P})$, since x, being an interior point, has last component different from zero. However, by choosing the tolerance ϵ sufficiently small, $\tilde{x}$ will approximate feasibility to any desired degree of accuracy.

6. Remarks. The purpose of this paper is to present a theoretical result. Thus in order to simplify the presentation, we constructed $\hat{\mu} = \mu(1 - \delta/\sqrt{n})$. Obviously, one can use $\hat{\mu}$ which is less than or equal to the above value, but still satisfies (a) of Theorem 4.5. In this way, one can accelerate the convergence of the algorithm.

If the objective function of problem (P) is of the form $\Psi(x) = \sum_{j=1}^{n} \psi_j(x_j) + (1/2)x^T Q x$ where Q is a $n \times n$ symmetric positive semidefinite matrix and each function ψ_j belongs to the collection Y, then all the results of this paper are still valid. The only change to be noted is in relation (4.2), which in this case would be

$$\begin{aligned}(\Delta x)^T(\Delta z) &= \Delta x^T \, \nabla^2 \Psi(x) \, \Delta x \\ &\geqslant \sum_{j=1}^{n} \psi_j''(x_j)(\Delta x_j)^2 \geqslant 0.\end{aligned}$$

The class of functions introduced in Definition 2.1 is dictated by the proof of Theorem 4.5. Although this class of functions is not intuitive, it contains the familiar class of functions given after Definition 2.1. We should mention that Algorithm 3.3 can be slightly modified to handle problems minimizing a general convex function subject to linear (or convex nonlinear) constraints. However, in this case, we may not guarantee the strong convergence property achieved in this paper for the special class of functions described in §2. We conjecture about the possibility of extending this strong convergence property to a larger class of functions.

Acknowledgement. The authors wish to thank two anonymous referees for many useful comments. This research was partially funded by the Brazilian Post-Graduate Education Federal Agency (CAPES) and by the United States Navy Office of Naval Research under contract N00014-87-K-0202. Their financial support is gratefully acknowledged.

References

[1] Bayer, D. A. and Lagarias, J. C. (1989). The Nonlinear Geometry of Linear Programming. Part I: Affine and Projective Scaling Trajectories. Part II: Legendre Transform Coordinates and Central Trajectories. *Trans. Amer. Math. Soc.* **314** 499–581.

[2] Fiacco, A. and McCormick, G. (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley and Sons, New York.

[3] Frisch, K. R. (1955). The Logarithm Potential Method of Convex Programming. Unpublished Manuscript, University Institute of Economics, Oslo, Norway.

[4] Gill, P. E., Murray, W., Saunders, M. A., Tomlin, J. A. and Wright, M. H. (1986). On Projected Newton Barrier Methods for Linear Programming and an Equivalence to Karmarkar's Projective Method. *Math. Programming* **36** 183–209.

[5] Gonzaga, C. C. (1989). An Algorithm for Solving Linear Programming Problems in $O(n^3 L)$ Operations, in *Progress in Mathematical Programming, Interior-Point Algorithms and Related Methods.* N. Megiddo (Ed.), Springer-Verlag, 1–28.

[6] Karmarkar, N. (1984). A New Polynomial Time Algorithm for Linear Programming. *Combinatorica* **4** 373–395.

[7] Kojima, M., Mizuno, S and Yoshise, A. (1989). A Primal-Dual Interior Point Algorithm for Linear Programming, in *Progress in Mathematical Programming, Interior-Point Algorithms and Related Meth-*

ods. N. Megiddo (Ed.), Springer-Verlag, 29–47.

[8] Kojima, M., Mizuno, S. and Yoshise, A. (1989). A Polynomial-Time Algorithm For a Class of Linear Complementary Problems. *Math. Programming* **44** 1–26.

[9] Megiddo, N. (1989). Pathways to the Optimal Set in Linear Programming, in *Progress in Mathematical Programming, Interior-Point Algorithms and Related Methods*. N. Megiddo (Ed.), Springer-Verlag, 131–158.

[10] Monteiro, R. D. C. and Adler, I. (1989). Interior Path Following Primal-Dual Algorithms. Part I: Linear Programming. *Math. Programming* **44** 27–41.

[11] ______ and ______ (1987). Interior Path Following Primal-Dual Algorithms. Part II: Convex Quadratic Programming. *Math Programming* **44** 43–66.

[12] Renegar, J. (1986). A Polynomial-Time Algorithm Based on Newton's Method for Linear Programming. *Math. Programming* **40** 59–93.

[13] Vaidya, P. M. (1987). An Algorithm for Linear Programming Which Requires $O(((m + n)n^2 + (m + n)^{1.5}n)L)$ Arithmetic Operations. Manuscript, AT & T Bell Laboratories, Murray Hill, NJ (to appear in *Math. Programming*).

MONTEIRO: AT & T BELL LABORATORIES, ROOM 1F-437, CRAWFORDS CORNER ROAD, HOLMDEL, NEW JERSEY 07733

ADLER: DEPARTMENT OF INDUSTRIAL ENGINEERING AND OPERATIONS RESEARCH, UNIVERSITY OF CALIFORNIA, BERKELEY, CALIFORNIA 94720