October 16, 1989 LIDS-P-1920

On the Convergence of the Affine-Scaling Algorithm*

by

Paul Tseng† and Zhi-Quan Luo‡

Abstract

The affine-scaling algorithm, first proposed by Dikin, is presently enjoying great popularity as a potentially effective means of solving linear programs. An outstanding question about this algorithm is its convergence in the presence of degeneracy (which is important since "practical" problems tend to be degenerate). In this paper, we give new convergence results for this algorithm that do not require any non-degeneracy assumption on the problem. In particular, we show that if the stepsize choice of either Dikin or Barnes or Vanderbei, et. al. is used, then the algorithm generates iterates that converge at least linearly with a convergence ratio of $1-\beta/\sqrt{n}$, where n is the number of variables and $\beta \in (0,1]$ is a certain stepsize ratio. For one particular stepsize choice which is an extension of that of Barnes, the limit point is shown to have a cost which is within $O(\beta)$ of the optimal cost and, for β sufficiently small, is shown to be exactly optimal. We prove the latter result by using an unusual proof technique, that of analyzing the ergodic convergence of the corresponding dual vectors. For the special case of network flow problems, we show that it suffices to take $\beta = \frac{1}{6mC}$, where m is the number of constraints and C is the sum of the cost coefficients, to achieve exact optimality.

KEY WORDS: Linear program, affine-scaling, ergodic convergence.

* This research is partially supported by the U.S. Army Research Office, contract DAAL03-86-K-0171 (Center for Intelligent Control Systems), by the National Science Foundation, grant NSF-ECS-8519058, and by the Science and Engineering Research Board of McMaster University.

†Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139.

‡Communications Research Laboratory, Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada.

1. Introduction

Since the recent work of Karmarkar [Kar84], much interest has focussed on solving linear programming problems using interior point algorithms. These interior point algorithms can be classified roughly as either (i) projective-scaling (or potential reduction), or (ii) affine-scaling, or (iii) path-following. Both the projective-scaling algorithms, originated by Karmarkar, and the path-following algorithms, attributed to Frisch [Fri55], have very nice polynomial-time complexity (see for example [Gon89], [Kar84], [Ren88], [Vai87], [Ye88]) and the latter can be extended to solve convex (quadratic) programs and certain classes of linear complementarity problems (see for example [KMY87], [MoA87], [MeS88], [Tse89], [Ye89]). However it is the affine-scaling algorithm that has enjoyed most wide use in practice [AKRV89], [CaS85], [MSSP88], [MoM87], although its time complexity is suspected not to be polynomial. (Recently, it was shown that one primal dual version of this algorithm has a polynomial-time complexity, provided that it starts near the "centre" of the feasible set and the stepsizes are sufficiently small [MAR88].) The affine-scaling algorithm was proposed independently by a number of researchers [Bar86], [CaS85], [ChK86], [KoS87], [VMF86], and it was only recently discovered (in the West) that this algorithm was invented 20 years ago by the Russian mathematician I. I. Dikin [Dik67], [Dik74] (see discussions in [VaL88], [Dik88]). A key open question about this algorithm is its convergence in the absence of any non-degeneracy assumption on the problem. Presently it is only known that this algorithm is convergent under the assumption of either primal non-degeneracy [Dik74], [VaL88] or, if a certain stepsize ratio is small, dual non-degeneracy [Tsu89]. (Weaker results that require both primal and dual non-degeneracy are given in [Bar86], [MeS89], [VMF86].) Otherwise, no useful convergence result of any kind is known. (The continuous time version of this algorithm was shown by Adler and Monteiro [AdM88] to converge even when the problem is primal and/or dual degenerate, but the analysis therein do not readily extend to our discrete time case.) This situation is rather unfortunate since most problems that occur in practice are degenerate.

In this paper we give the first convergence results for the (discrete time) affine-scaling algorithm that do not require any non-degeneracy assumption on the problem. In particular, we consider versions of this algorithm proposed by, respectively, Dikin [Dik67], Barnes [Bar86], and Vanderbei, et. al. [VMF86], and we show that any sequence of iterates generated by either of these algorithms converge at least linearly with a convergence ratio of $1-\beta/\sqrt{n}$, where $\beta \in (0,1]$ is a certain stepsize ratio and n is the problem dimension. Moreover, for a particular version of the algorithm we show that the limit point has a cost that is within $O(\beta)$ of the optimal cost, where the constant inside the big O notation depends on the problem data only, and, for β sufficiently small, this limit point is exactly optimal. For single commodity network flow problems we estimate the size of β for which the latter holds to be $\frac{1}{6mC}$, where m is the number of constraints and C is the sum of the cost coefficients. Our convergence result for the small stepsize case significantly improves upon that obtained by Adler and Monteiro [AdM88] for the continuous time version of the affine-scaling algorithm (for which the stepsizes are infinitesimally small). Our proofs are also fundamentally different from those of the others. For example, in order to prove the $O(\beta)$-optimality result, instead of following closely the trajectory of the primal and/or dual iterates as is typicall done, we study the long term averages of the dual iterates, which exhibit a much more stable behaviour than the individual dual iterates. (Convergence in the average of iterates is known as ergodic convergence, e.g. [Pas79].) We show, by a very simple argument, that this long term average is bounded and, in the limit, satisfies $O(\beta)$-complementary slackness with the primal iterates.

2. Algorithm Description

Consider linear program in the following canonical form:

$$\begin{aligned} \text{Minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b, \\ & x \geq 0, \end{aligned} \tag{P}$$

where A is an $m \times n$ matrix, b is an m-vector, and c is an n-vector. In our notation, all vectors are column vectors and superscript T denotes transpose. We will denote, for any vector x, by x_j the j-th coordinate of x and by $\|x\|_1$ and $\|x\|_2$, respectively, the L_1-norm and the L_2-norm of x. We make the following standing assumption about (P), which is standard for interior point algorithms.

Assumption A. (P) has a finite optimal value and $\{x | Ax = b, x > 0\}$, the relative interior of its feasible set, is nonempty.

Consider the following version of the affine-scaling algorithm for solving (P): Given $x^k > 0$ satisfying $Ax^k = b$ (x^0 is assumed given), let w^k be the unique optimal solution of the following subproblem

$$\begin{aligned} \text{Minimize} \quad & c^T w \\ \text{subject to} \quad & Aw = 0, \\ & \|(X^k)^{-1} w\|_2^2 \leq n, \end{aligned} \tag{2.1}$$

where X^k is the $n \times n$ diagonal matrix whose j-th diagonal entry is x_j^k, and set

$$x^{k+1} = x^k + \lambda^k w^k, \tag{2.2}$$

where λ^k is a positive stepsize for which $x^k + \lambda^k w^k > 0$ (λ^k will be specified presently). Notice that $x^{k+1} > 0$ and (since $Aw^k = 0$) $Ax^{k+1} = Ax^k = b$. Also, since the zero vector is a feasible solution of (2.1), there holds $c^T w^k \leq 0$ (i.e., w^k is a descent direction at x^k) so that $c^T x^{k+1} \leq c^T x^k$. Hence, $\{c^T x^k\}$ is monotonically decreasing and x^{k+1} is a feasible solution of (P) for all k. Since the function $x \rightarrow c^T x$ is bounded from below on the feasible set for (P) (cf. Assumption A), this implies that $\{c^T x^k\}$ converges to a limit. [Also notice that the value used in the right hand side of the ellipsoid constraint in (2.1) is immaterial since w^k is scaled by λ^k in (2.2).]

All of the affine-scaling algorithms proposed for solving (P) differ only in their choices of the stepsize λ^k. We will consider primarily the following choice for λ^k:

$$\lambda^k = \frac{\beta}{\|(X^k)^{-1} w^k\|}, \tag{2.3}$$

where β is a fixed scalar in $(0,1)$ and $\|\cdot\|$ is any L_p-norm ($p \in [1,\infty]$). (The largest stepsize is obtained when $\|\cdot\|$ is the L_∞-norm.) When $\|\cdot\|$ is the L_2-norm, then the above choice of λ^k coincides with that proposed by Barnes [Bar86]. Alternatively, we can choose

$$\lambda^k = \frac{1}{\|(X^k)^{-1}w^k\|_2}, \tag{2.4}$$

which is the stepsize proposed in the the original algorithm of Dikin [Dik67], [Dik74]. Vanderbei, et. al. [VMF86] choose λ^k to be a fraction $\beta \in (0,1)$ of the largest stepsize that maintains feasibility of the new iterate, i.e. (compare with (2.3))

$$\lambda^k = \frac{\beta}{\max_j\{-w_j^k/x_j^k\}}. \tag{2.5}$$

It can be seen that all of the above stepsizes maintain $x^k + \lambda^k w^k > 0$. [For Dikin's stepsize (2.4), it can be shown that the positivity condition is not satisfied only if $x^k + \lambda^k w^k$ is an optimal solution of (P), in which case the algorithm can be terminated immediately.] In what follows, we will consider primarily the stepsize (2.3) and will allude to the other stepsizes only on occasions when our results apply to them as well. We remark that all of our results extend to a modified version of the stepsize of [VMF86], whereby an upper bound is placed on the positive components of the descent direction as well, i.e. λ^k is the minimum of $\frac{\beta}{\max_j\{-w_j^k/x_j^k\}}$ and $\frac{\eta}{\max_{w_j^k>0}\{w_j^k/x_j^k\}}$, for some fixed positive scalar η.

It is easily seen that the redundant rows of A can be removed without changing the iterates w^k and x^k given by (2.1)-(2.2) (since the feasible set for both (P) and (2.1) would remain unchanged). Hence, to simplify the presentation, we will without loss of generality make the following standing assumption:

Assumption B. The matrix A has full row rank.

Then, by attaching a Lagrange multiplier vector p^k to the constraints $Aw = 0$, we obtain from the Kuhn-Tucker conditions for (2.1) that w^k has the following closed form:

$$w^k = -\sqrt{n}\frac{(X^k)^2 r^k}{\|X^k r^k\|_2}, \tag{2.6}$$

where

$$r^k = c - A^T p^k, \tag{2.7}$$

and

$$p^k = (A(X^k)^2A^T)^{-1}A(X^k)^2c. \tag{2.8}$$

(The matrix inverse in (2.8) is well-defined since A has full row rank and X^k is a diagonal matrix with positive diagonal entries.) The m-vector p^k can be thought of as the dual vector corresponding to x^k, although it is not necessarily dual feasible.

This paper proceeds as follows: In Sections 3 and 4, we show that the iterates generated by (2.1)-(2.2), with the stepsizes given by either (2.3) or (2.4) or (2.5), converge at least linearly with a convergence ratio between $1-\beta/n$ and $1-\beta/\sqrt{n}$, depending on the choice of stepsizes used. In Section 5, we show that, for the stepsize choice (2.3), the limit point has a cost that is within $O(\beta)$ of the optimal cost and, for β sufficiently small, is exactly optimal. In Section 6, we show that, for the single commodity network flow problem, it suffices to take $\beta = \frac{1}{6m\|c\|_1}$ in order for exact optimality to be attained. In Section 7, we discuss various extensions.

3. Linear Convergence of the Costs

In this section, we analyze the rate of convergence of the costs $c^T x^k$ generated by the algorithm (2.1)-(2.2) [with stepsizes given by either (2.3) or (2.4) or (2.5)]. In particular, we show that, for all k sufficiently large, the costs $c^T x^k$ converge at least linearly with a convergence ratio between $1 - \beta/n$ and $1 - \beta/\sqrt{n}$, depending on the choice of the stepsize λ^k used. A similar result has been obtained earlier by Barnes [Bar86], but only for the stepsize (2.3) and under the additional assumption that (P) is both primal and dual non-degenerate.

First, we need the following result which says that the solution of a linear system is in some sense Lipschitz continuous in the right hand side (see for example [Hof52], [Rob73], [MaS87]):

Lemma 1. For any $k \times n$ matrix B, any $l \times n$ matrix C, any k-vector d and any l-vector e, if the linear system $Bx = d, Cx \geq e$ has a solution, then it has a solution whose $\|\cdot\|$ norm is at most $\mu(\|d\| + \|e\|)$, where μ is a constant that depends on B and C only.

Lemma 1 will be used in later analysis as well. Below we give the main result of this section.

Theorem 1. If $\{x^k\}$ is a sequence of iterates generated by (2.1)-(2.2), then

$$c^T x^{k+1} - v^\infty \leq (1 - \lambda^k)(c^T x^k - v^\infty)$$

for all k sufficiently large, where $v^\infty = \lim_{k\to\infty}\{c^T x^k\}$.

Proof: Let Ξ denote the set of feasible solutions for (P), i.e. $\Xi = \{x | Ax = b, x \geq 0\}$. First we claim that there exists a positive integer $\bar{k}$ such that

$$\min_{y\in\Xi, c^T y = v^\infty} \|(X^k)^{-1}(y - x^k)\|_2^2 \leq n, \forall k \geq \bar{k}. \tag{3.1}$$

To see this, suppose the contrary, so that there exists a subsequence K of $\{0, 1, ...\}$ such that

$$\min_{y\in\Xi, c^T y = v^\infty} \|(X^k)^{-1}(y - x^k)\|_2^2 > n, \forall k \in K. \tag{3.2}$$

By further passing into a subsequence if necessary, we will assume that, for each $j \in \{1, ..., n\}$, either $\{x_j^k\}_K$ converges to some limit, say x_j^∞, or $\{x_j^k\}_K \to \infty$. For each $k \in K$, consider the linear system $Ax = b, x \geq 0, c^T x = c^T x^k, x_j = x_j^k\ \forall j \in J$, where J is the set of indices j such that $\{x_j^k\}_K$ converges to some limit. This system is feasible since x^k is a solution, so that, by Lemma 1, there exists a solution ξ^k such that $\|\xi^k\| = O(\|b\| + |c^T x^k| + \sum_{j\in J} |x_j^k|)$. Then, the sequence $\{\xi^k\}_K$ is bounded and satisfies

$$\xi^k \in \Xi, \quad c^T \xi^k = c^T x^k, \quad \xi_j^k = x_j^k\ \forall j \in J,$$

for all $k \in K$. Since $\{\xi^k\}_K$ is bounded, by further passing into a subsequence if necessary, we will assume that it converges to some limit, say ξ^∞. Then, $\xi^\infty \in \Xi$, $c^T\xi^\infty = \upsilon^\infty$ and $\xi_j^\infty = x_j^\infty$ for all $j \in J$. For each $k \in K$, let $\Delta^k = x^k - \xi^k$ (so that $c^T\Delta^k = 0, A\Delta^k = 0, \Delta_j^k = 0$ for all $j \in J$, and $\Delta^k \geq 0$ if k is sufficiently large). Then, $y^k = \xi^\infty + \Delta^k$ has a cost of υ^∞, satisfies

$$\begin{aligned} \|(X^k)^{-1}(y^k - x^k)\|_2^2 &= \|(X^k)^{-1}(\xi^\infty - \xi^k)\|_2^2 \\ &= \sum_{j\in J, \xi_j^\infty = 0} \left(\frac{-\xi_j^k}{x_j^k}\right)^2 + \sum_{j \in J, \xi_j^\infty > 0 \text{ or } j \notin J} \left(\frac{\xi^\infty - \xi_j^k}{x_j^k}\right)^2, \end{aligned} \tag{3.3}$$

and, for all $k \in K$ sufficiently large, is in Ξ. Since $\xi_j^k = x_j^k$ for all $k \in K$ and all $j \in J$, each term in the second to the last sum of expression (3.3) is equal to 1. Also, since $\{\xi^k\}_K \to \xi^\infty$, $\{x_j^k\}_K \to \infty$ for all $j \notin J$, and $\{x_j^k\}_K \to \xi_j^\infty$ for all $j \in J$, then each term in the last sum of expression (3.3) is less than or equal to 1 for all $k \in K$ sufficiently large. Hence, for all $k \in K$ sufficiently large, y^k belongs to Ξ and satisfies $c^Ty^k = \upsilon^\infty$ and $\|(X^k)^{-1}(y^k - x^k)\|_2^2 \leq n$, a contradiction of (3.2). Hence (3.1) holds.

Now, consider any $k \geq \bar{k}$ and let y^k be any element of Ξ satisfying $c^Ty^k = \upsilon^\infty$, $\|(X^k)^{-1}(y^k - x^k)\|_2^2 \leq n$ [cf. (3.1)]. Then, $y^k - x^k$ is a feasible solution for the subproblem (2.1) and, since w^k is the optimal solution of (2.1), it must be that $c^Tw^k \leq c^Ty^k - c^Tx^k$. Since $c^Ty^k = \upsilon^\infty$, this together with (2.2) then yields

$$\begin{aligned} c^Tx^{k+1} &= c^Tx^k + \lambda^k c^T w^k \\ &\leq c^Tx^k + \lambda^k(\upsilon^\infty - c^Tx^k), \end{aligned}$$

Hence

$$c^Tx^{k+1} - \upsilon^\infty \leq (1 - \lambda^k)(c^Tx^k - \upsilon^\infty),$$

Q.E.D.

An open question is the estimation of $\bar{k}$. For example, if $\bar{k}$ is a polynomial in the size of the problem encoding, then, for linear network flow problems with polynomial-sized cost coefficients (e.g. maximum flow), we would obtain a polynomial-time algorithm (see Corollary 1 below and Theorem 4 in Section 6). Next, we bound the stepsize λ^k.

Lemma 2. The following hold:

(a) If λ^k is given by (2.3), then $\frac{\beta}{\sqrt{n}} \min_{\|x\|=1} \frac{\|x\|_2}{\|x\|} \leq \lambda^k \leq \beta$ for all k.

(b) If λ^k is given by (2.4), then $\lambda^k = 1/\sqrt{n}$ for all k.

(c) If λ^k is given by (2.5), then $\beta/\sqrt{n} \leq \lambda^k$ for all k.

Proof: Parts (a) and (b) follow from the observation that the ellipsoid constraint in (2.1) is tight for any optimal solution of (2.1), so that w^k satisfies $\|(X^k)^{-1}w^k\|_2 = \sqrt{n}$ for all k. To prove part (c), note from $\|(X^k)^{-1}w^k\|_2 = \sqrt{n}$ that $0 \leq \max_j\{\frac{-w_j^k}{x_j^k}\} \leq \sqrt{n}$ for all k. Q.E.D.

Notice that $\min_{\|x\|=1} \frac{\|x\|_2}{\|x\|}$ is lower bounded by $1/\sqrt{n}$ and is equal to 1 if $\|\cdot\|$ is an L_p-norm with $p \geq 2$. From Theorem 1 and Lemma 2, we immediately obtain the following corollary:

Corollary 1. If $\{x^k\}$ is a sequence of iterates generated by (2.1)-(2.2) with the stepsizes given by either (2.3) or (2.4) or (2.5), then $\{c^T x^k\}$ converges at least linearly with a convergence ratio of, respectively, $1 - \frac{\beta}{\sqrt{n}} \min_{\|x\|=1} \frac{\|x\|_2}{\|x\|}$, $1 - 1/\sqrt{n}$, and $1 - \beta/\sqrt{n}$.

4. Linear Convergence of the Primal Iterates

In this section, we establish that the sequence of iterates $\{x^k\}$ generated by $(2.1)-(2.2)$ [with stepsizes given by either (2.3) or (2.4) or (2.5)] in fact converges. Our proof is based on showing that the change in the iterate $x^{k+1}-x^k$ is $O(c^Tx^k - c^Tx^{k+1})$, so that, by the linear convergence result proven earlier (cf. Corollary 1), $\{x^k\}$ is a Cauchy sequence and therefore converges. Intuitively, $x^{k+1}-x^k$ should be $O(c^Tx^k - c^Tx^{k+1})$, for otherwise there would exist an n-vector in the space orthogonal to the cost vector c which can be subtracted from $x^{k+1}-x^k$ to obtain a "better" descent direction.

Theorem 2. If $\{x^k\}$ is a sequence of iterates generated by $(2.1)-(2.2)$ with the stepsizes given by either (2.3) or (2.4) or (2.5), then $\{x^k\}$ converges at least linearly with the same convergence ratio as that of $\{c^Tx^k\}$.

Proof: Let

$$z^k = x^{k+1} - x^k, \ \forall k.$$

From Theorem 1 we have that $\{c^Tz^k\}$ converges to zero at least linearly with a ratio of convergence given in Corollary 1. Below we show that $\|z^k\|$ is $O(-c^Tz^k)$, from which it immediately follows that $\{x^k\}$ converges at least linearly with the same convergence ratio as that of $\{c^Tx^k\}$.

First, we claim that each z^k can be decomposed as

$$z^k = \hat{z}^k + \tilde{z}^k, \tag{4.1a}$$

where $\hat{z}^k$ and $\tilde{z}^k$ are n-vectors satisfying

$$A\hat{z}^k = 0, \quad A\tilde{z}^k = 0, \quad c^T\hat{z}^k = 0, \quad c^T\tilde{z}^k = c^Tz^k, \tag{4.1b}$$

and $\|\tilde{z}^k\|$ is $O(-c^Tz^k)$. (To see this, for each $k \in K$, consider the linear system $Az = 0, c^Tz = c^Tz^k$. This system is feasible since z^k is a solution. By Lemma 1, there exists a solution $\tilde{z}^k$ such that $\|\tilde{z}^k\| = O(-c^Tz^k)$, where the constant in the big O notation depends on A and c only. Let $\hat{z}^k = z^k - \tilde{z}^k$.)

If $\|\hat{z}^k\|$ is also $O(-c^Tz^k)$, then clearly $\|z^k\|$ is $O(-c^Tz^k)$ [cf. (4.1a)] and we are done. Otherwise, suppose that there exists a subsequence K of $\{0,1,...\}$ such that $\{c^Tz^k/\|\hat{z}^k\|\}_K \to 0$. We will then establish a contradiction. First, by further passing into a subsequence if necessary, we will assume that the set of coordinates $\hat{z}_j^k$ that are of the same order of magnitude as $\|\hat{z}^k\|$ is fixed, i.e. there exists a nonempty $J \subseteq \{1,...,n\}$ such that

$$\{\tfrac{|\hat{z}_j^k|}{\|\hat{z}^k\|}\}_K \to 0 \ \forall j \notin J, \quad \lim_{k\to\infty, k\in K} \tfrac{|\hat{z}_j^k|}{\|\hat{z}^k\|} > 0 \ \forall j \in J, \tag{4.2}$$

Now, for each $k \in K$, consider the linear system $Az = 0, c^T z = 0, z_j = \hat{z}_j^k$ for all $j \notin J$. This system is feasible since $\hat{z}^k$ is a solution. By Lemma 1, there exists a solution ς^k such that $\|\varsigma^k\| = O(\sum_{j \notin J} |\hat{z}_j^k|)$, where the constant in the big O notation depends on A and c only. Then, by (4.2), we have

$$\{\tfrac{\|\varsigma^k\|}{\|\hat{z}^k\|}\}_K \to 0, \quad A\varsigma^k = 0, \quad c^T\varsigma^k = 0, \quad \varsigma_j^k = \hat{z}_j^k \; \forall j \notin J. \tag{4.3}$$

Let $\Delta^k = \hat{z}^k - \varsigma^k$ for all $k \in K$. Then, for every $k \in K$, there holds [cf. (4.1a), (4.1b) (4.3)] $c^T(z^k - \Delta^k) = c^T z^k, A(z^k - \Delta^k) = 0$ and

$$\begin{aligned} \|(X^k)^{-1}(z^k - \Delta^k)\|_2^2 &= \sum_{j \notin J} (\frac{z_j^k}{x_j^k})^2 + \sum_{j \in J} (\frac{z_j^k - \hat{z}_j^k + \varsigma_j^k}{x_j^k})^2 \\ &= \sum_{j \notin J} (\frac{z_j^k}{x_j^k})^2 + \sum_{j \in J} (\frac{\tilde{z}_j^k + \varsigma_j^k}{x_j^k})^2 \end{aligned} \quad . \tag{4.4}$$

Now, since $\|\tilde{z}^k\|$ is $O(-c^T z^k)$, our hypothesis $\{c^T z^k / \|\hat{z}^k\|\}_K \to 0$ implies $\{\tilde{z}^k / \|\hat{z}^k\|\}_K \to 0$, which together with (4.3) yields $\{(\tilde{z}^k + \varsigma^k)/\|\hat{z}^k\|\}_K \to 0$. Then, by (4.1a) and (4.2), we have $\{(\tilde{z}_j^k + \varsigma_j^k)/z_j^k\}_K \to 0$ for all $j \in J$, so that each j-th term in the last sum of (4.4) is strictly less than $(\frac{z_j^k}{x_j^k})^2$ for all $k \in K$ sufficiently large. Since J is nonempty, this together with (4.4) yields that, for all $k \in K$ sufficiently large,

$$\|(X^k)^{-1}(z^k - \Delta^k)\|_2^2 < \|(X^k)^{-1} z^k\|_2^2,$$

so that (also using $c^T z^k < 0$ and the observation that w^k is the unique positive multiple of z^k whose L_2-norm after pre-multiplication by $(X^k)^{-1}$ is $\sqrt{n}$) $\frac{c^T z^k \sqrt{n}}{\|(X^k)^{-1}(z^k - \Delta^k)\|_2} < \frac{c^T z^k \sqrt{n}}{\|(X^k)^{-1} z^k\|_2} = c^T w^k$. Since $c^T(z^k - \Delta^k) = c^T z^k$, this implies that the vector $(z^k - \Delta^k)\frac{\sqrt{n}}{\|(X^k)^{-1}(z^k - \Delta^k)\|_2}$ has a cost strictly lower than that of w^k. Also, since $A(z^k - \Delta^k) = 0$, this same vector can be seen to be a feasible solution of (2.1), contradicting the fact that w^k is the optimal solution of (2.1). Q.E.D.

5. Convergence to Near Optimality and Ergodic Convergence of the Dual Iterates

From Section 4 we have that $\{x^k\}$ converges at least linearly to some limit point, which is clearly a feasible solution of (P). Hence, it only remains to show that this limit point is an optimal (or approximately optimal) solution of (P). This, however, turns out to be a very difficult task because the trajectory of the dual vectors p^k near the relative boundary of the feasible set is quite unpredictable. We will resolve this difficulty by taking a long term weighted average of the dual vectors. By choosing the weights appropriately, we show that the sequence of "average" dual vectors is bounded and, in the limit, satisfies approximate complementary slackness with the primal limit point. This analysis, however, only works for the stepsize (2.3) (as well as the modified version of the stepsize (2.5) discussed in Section 2, which will not be treated here) and it remains open whether it can be extended to other stepsizes.

First, we give a characterization of approximate complementary slackness. For any $\epsilon > 0$, any x that is feasible for (P) (i.e. x satisfies $Ax = b$ and $x \geq 0$) and any m-vector p, we will say that x and p satisfy ϵ-complementary slackness (ϵ-CS for short) [TsB87a] if, for all $j \in \{1, ..., n\}$,

$$\begin{array}{llrl} x_j = 0 & \Rightarrow & & c_j - A_j^T p \geq -\epsilon, \\ x_j > 0 & \Rightarrow & \epsilon \geq & c_j - A_j^T p \geq -\epsilon, \end{array} \qquad (\epsilon - CS)$$

where A_j denotes the j-th column of A. From Proposition 8 in [TsB87b] we have the following lemma regarding primal dual pairs satisfying ϵ-CS:

Lemma 3. Any x that is feasible for (P) and satisfies ϵ-CS with some p is within $O(\epsilon)$ in cost of the optimal cost, where the constant in the big O notation depends on the problem data only.

Moreover, it can be seen that any x satisfying the hypothesis of Lemma 3 is an optimal solution of a perturbed problem whereby every cost component is perturbed by at most ϵ. Since we are dealing with linear programs, it is easily seen that if ϵ is sufficiently small, then every optimal solution of the perturbed problem is also an optimal solution of the original problem (P) (see discussions in [TsB87a, Section 5]). Although the size of ϵ for which this holds is in general very small, for certain special classes of problems it can be taken to be quite large. For example, for linear cost network flow problems with integer data, it has been shown that $\epsilon < 1/m$ suffices (see [BeE88], [BeT89; Chap. 5]).

The following lemma follows as an immediate consequence of our construction of the descent directions w^k.

Lemma 4. If $\{x^k\}$ is a sequence of iterates generated by (2.1)-(2.2) with stepsizes given by either (2.3) or (2.4) or (2.5), then $\{X^k r^k\} \to 0$.

Proof: By using (2.2), (2.6)-(2.8) and the idempotent property of the orthogonal projection operator $I - X^k A^T (A(X^k)^2 A^T)^{-1} A X^k$, it can be seen that $c^T x^k - c^T x^{k+1} = \lambda^k \sqrt{n} \|X^k r^k\|_2$ for all k. Since $\{c^T x^k\}$

converges so that $\{c^T x^k - c^T x^{k+1}\} \to 0$ and (cf. Lemma 2) λ^k is bounded from below for all k, this proves the claim. Q.E.D.

Below is our main result:

Theorem 3. If $\{x^k\}$ is a sequence of iterates generated by (2.1)-(2.2) with the stepsizes given by (2.3), then $\{x^k\}$ converges to a limit point which satisfies $O(\beta/(1-\beta))$-CS with some m-vector p.

Proof: For each k denote

$$\pi^k = \frac{p^k/\|X^k r^k\| + ... + p^0/\|X^0 r^0\|}{1/\|X^k r^k\| + ... + 1/\|X^0 r^0\|} \tag{5.1}$$

(so that π^k is a linear convex combination of p^k ,..., p^0). We will show that x^{k+1} and π^k satisfy $O(\beta/(1-\beta))$ -CS as $k \to \infty$.

Fix any $j \in \{1, ..., n\}$. From (2.2), (2.3) and (2.6) we have $x^{k+1} = x^k - \beta(X^k)^2 r^k/\|X^k r^k\|$ for all k, so that

$$\begin{aligned} x_j^{k+1} &= x_j^k - \beta(x_j^k)^2 r_j^k/\|X^k r^k\| \\ &= x_j^k(1 - \beta x_j^k r_j^k/\|X^k r^k\|) \\ &= x_j^k(1 + \beta x_j^k \delta_j^k), \end{aligned} \tag{5.2}$$

where we denote

$$\delta_j^k = -r_j^k/\|X^k r^k\|. \tag{5.3}$$

Thus, $x_j^{k+1}/x_j^k - 1 = \beta x_j^k \delta_j^k$ so that if we let

$$\theta_j^k = \frac{x_j^k \delta_j^k}{x_j^{k+1}} + \frac{x_j^{k-1}\delta_j^{k-1}}{x_j^k} + ... + \frac{x_j^0 \delta_j^0}{x_j^1},$$

we obtain that

$$\begin{aligned} \beta\theta_j^k &= \beta\left[\frac{x_j^k \delta_j^k}{x_j^{k+1}} + \frac{x_j^{k-1}\delta_j^{k-1}}{x_j^k} + ... + \frac{x_j^0 \delta_j^0}{x_j^1}\right] \\ &= \frac{1}{x_j^{k+1}}\left(\frac{x_j^{k+1}}{x_j^k} - 1\right) + \frac{1}{x_j^k}\left(\frac{x_j^k}{x_j^{k-1}} - 1\right) + ... + \frac{1}{x_j^1}\left(\frac{x_j^1}{x_j^0} - 1\right) \\ &= \frac{1}{x_j^k} - \frac{1}{x_j^{k+1}} + \frac{1}{x_j^{k-1}} - \frac{1}{x_j^k} + ... + \frac{1}{x_j^0} - \frac{1}{x_j^1} \\ &= \frac{1}{x_j^0} - \frac{1}{x_j^{k+1}}. \end{aligned} \tag{5.4}$$

We bound θ_j^k as follows: From (5.2) we have that, for every k,

$$(1+\beta)x_j^k \geq x_j^{k+1} > x_j^k \quad if \quad \delta_j^k > 0,$$
$$(1-\beta)x_j^k \leq x_j^{k+1} < x_j^k \quad if \quad \delta_j^k < 0,$$

so that

$$\frac{\delta_j^k}{1+\beta} \leq \frac{x_j^k \delta_j^k}{x_j^{k+1}} \leq \delta_j^k \quad if \quad \delta_j^k > 0,$$
$$\frac{\delta_j^k}{1-\beta} \leq \frac{x_j^k \delta_j^k}{x_j^{k+1}} \leq \delta_j^k \quad if \quad \delta_j^k < 0.$$

Since $1/(1+\beta) = 1 - \beta/(1+\beta)$ and $1/(1-\beta) = 1 + \beta/(1-\beta)$, this implies that

$$\delta_j^k + \ldots + \delta_j^0 - \frac{\beta}{1+\beta} \sum_{\delta_j^l > 0} \delta_j^l + \frac{\beta}{1-\beta} \sum_{\delta_j^l < 0} \delta_j^l \leq \theta_j^k \leq \delta_j^k + \ldots + \delta_j^0.$$

Hence (also using the fact $1/(1+\beta) < 1/(1-\beta)$),

$$\theta_j^k \leq \delta_j^k + \ldots + \delta_j^0 \leq \theta_j^k + \frac{\beta}{1-\beta}(|\delta_j^k| + \ldots + |\delta_j^0|).$$

Dividing all sides by $1/\|X^k r^k\| + \ldots + 1/\|X^0 r^0\|$ and using (5.3) gives

$$\begin{aligned} \frac{\theta_j^k}{1/\|X^k r^k\| + \ldots + 1/\|X^0 r^0\|} &\leq -\frac{r_j^k/\|X^k r^k\| + \ldots + r_j^0/\|X^0 r^0\|}{1/\|X^k r^k\| + \ldots + 1/\|X^0 r^0\|} \\ &\leq \frac{\theta_j^k}{1/\|X^k r^k\| + \ldots + 1/\|X^0 r^0\|} + \frac{\beta}{1-\beta} \frac{|r_j^k|/\|X^k r^k\| + \ldots + |r_j^0|/\|X^0 r^0\|}{1/\|X^k r^k\| + \ldots + 1/\|X^0 r^0\|}. \end{aligned} \tag{5.5}$$

Since the sequence $\{p^k\}$ [cf. (2.8)] is bounded by the following lemma given in [VaL88]:

<u>Lemma 5</u>. For any n-vector γ, the function $w^\gamma(x) = (A(X)^2A^T)^{-1}A(X)^2\gamma$, where X denotes the $n \times n$ diagonal matrix whose j-th diagonal entry is x_j, is bounded. Moreover, the bound depends on A and γ only.

then so is the sequence $\{r^k\}$ [cf. (2.7)]. Since the middle quantity in (5.5) is exactly $A_j^T\pi^k - c_j$ [cf. (2.7) and (5.1)] and the far right quantity in (5.5) is simply $\beta/(1-\beta)$ multiplied by a linear convex combination of $|r_j^k|, \ldots, |r_j^0|$, everyone of which according to the Lemma 5 is bounded by some scalar M depending on A and c only, it follows that

$$\frac{\theta_j^k}{1/\|X^k r^k\| + \ldots + 1/\|X^0 r^0\|} \leq A_j^T\pi^k - c_j \leq \frac{\theta_j^k}{1/\|X^k r^k\| + \ldots + 1/\|X^0 r^0\|} + \frac{\beta}{1-\beta}M. \tag{5.6}$$

Let x^∞ be the limit point of $\{x^k\}$ (so that x^∞ is feasible for (P)), which exists by Theorem 2. Consider any $j \in \{1, \ldots, n\}$. Suppose that $x_j^\infty = 0$. Then, $\{x_j^{k+1}\} \to 0$ so that, by (5.4), $\{\theta_j^k\} \to -\infty$. This together with (5.6) implies

$$\limsup_{k\to\infty}\{A_j^T\pi^k - c_j\} \le \frac{\beta}{1-\beta}M.$$

Now, suppose that $x_j^\infty > 0$. Then, since $\{x_j^k r_j^k\} \to 0$ (cf. Lemma 4) and $\{x_j^k\} \to x_j^\infty$, we have $\{r_j^k\} \to 0$, which together with the fact that $\{1/\|X^k r^k\|\}$ is bounded away from zero (cf. Lemma 4) implies that

$$\{\frac{r_j^k/\|X^k r^k\| + ... + r_j^0/\|X^0 r^0\|}{1/\|X^k r^k\| + ... + 1/\|X^0 r^0\|}\} \to 0.$$

Since the quantity on the left hand side of the above expression is exactly $c_j - A_j^T\pi^k$ [cf. (2.7) and (5.1)], it follows that $\{c_j - A_j^T\pi^k\} \to 0$. Now, since our choice of j was arbitrary, the above holds for all $j \in \{1, ..., n\}$, so that any limit point of $\{\pi^k\}$ (which exists by Lemma 5) satisfies $\frac{\beta}{1-\beta}M$-CS with x^∞. Q.E.D.

We remark that, by a more careful analysis, we can improve the bound for the j-th coordinate from $\frac{\beta}{1-\beta}M$ to $\frac{\beta}{1-\beta}M\omega_j$, where $\omega_1, ..., \omega_n$ are positive scalars such that $\omega_1 + ... + \omega_n \le \max_{\|x\|=1} \frac{\|x\|_1}{\|x\|}$. Also, from the proof of Theorem 1 we see that every limit point of the sequence of dual vectors $\{\pi^k\}$ is an $O(\beta/(1-\beta))$-optimal dual solution of (P). Unfortunately, there does not seem to be any practical way to evaluate the π^k's.

By Theorem 3 and the properties of the ϵ-CS mechanism (cf. Lemma 3), for β sufficiently small, the iterates x^k converge to an optimal solution of (P).

6. Estimating the Stepsize for Achieving Optimality

We had shown in Section 5 that, provided that the stepsize β is sufficiently small, the iterates x^k generated by (2.1)-(2.3) are guaranteed to converge to an optimal solution of (P). Hence, it is of interest to estimate the size of β for which this holds. Below we consider a special case of (P), namely, the single commodity network flow problem (i.e. A is the node-arc incidence matrix for a directed graph) [FoF62], [BeT89], [PaS82], [Roc84], and show that, for this problem, β need not be smaller than $\frac{1}{6m\|c\|_1}$.

Theorem 4. Suppose that (P) is a single commodity network flow problem and the components of c and b are all integers. Then, for any $\beta \le \frac{1}{6m\|c\|_1}$ and any sequence of iterates $\{x^k\}$ generated by (2.1)-(2.3), there holds $\{x^k\}$ converges at least linearly to an optimal solution of (P).

Proof: First we bound $\{p^k\}$. Fix any n-vector $x > 0$. Following [VaL88], we use Cramer's rule and the Cauchy-Binet theorem to write the i-th component of the corresponding dual vector
$p = (A(X)^2A^T)^{-1}A(X)^2c$ as

$$p_i = \frac{\sum_{1\le j_1<\cdots<j_m\le n}(x_{j_1}\cdots x_{j_m})^2 \det_{j_1,\ldots,j_m}(a_1,\ldots,a_m)\det_{j_1,\ldots,j_m}(a_1,\ldots,a_{i-1},c,a_{i+1},\ldots,a_m)}{\sum_{1\le j_1<\cdots<j_m\le n}(x_{j_1}\cdots x_{j_m})^2[\det_{j_1,\ldots,j_m}(a_1,\ldots,a_m)]^2},$$

where $\det_{j_1,\ldots,j_m}(\alpha_1,\ldots,\alpha_m)$ denotes the determinant of the $m\times m$ matrix obtained by selecting columns $j_1,\ldots,j_m$ from the $m\times n$ matrix whose rows are the n-vectors $\alpha_1,\ldots,\alpha_m$ and where a_i denotes the i-th row of A. By the total unimodularity property of node-arc incidence matrices, we have that each $\det_{j_1,\ldots,j_m}(a_1,\ldots,a_m)$ is either 0 or 1 or -1 (see for example [PaS82]). Hence,

$$|p_i| \le \frac{\sum_{1\le j_1<\cdots<j_m\le n}(x_{j_1}\cdots x_{j_m})^2|\det_{j_1,\ldots,j_m}(a_1,\ldots,a_{i-1},c,a_{i+1},\ldots,a_m)|}{\sum_{1\le j_1<\cdots<j_m\le n}(x_{j_1}\cdots x_{j_m})^2},$$

where the summation in both the numerator and the denominator are taken over only those indices j for which $\det_{j_1,\ldots,j_m}(a_1,\ldots,a_m)$ is nonzero. Then, the right hand side of the above expression is simply a linear convex combination of the $|\det_{j_1,\ldots,j_m}(a_1,\ldots,a_{i-1},c,a_{i+1},\ldots,a_m)|$'s, which in turn is upper bounded by the maximum of the $|\det_{j_1,\ldots,j_m}(a_1,\ldots,a_{i-1},c,a_{i+1},\ldots,a_m)|$'s. Now, by Cramer's rule,

$$\det_{j_1,\ldots,j_m}(a_1,\ldots,a_{i-1},c,a_{i+1},\ldots,a_m) = \sum_{l=1}^{m}(-1)^{l+1}c_{j_l}\det_{j_1,\ldots,j_{l-1},j_{l+1},\ldots,j_m}(a_1,\ldots,a_{i-1},a_{i+1},\ldots,a_m),$$

and, by the total unimodularity property of A, each determinant inside the above sum is either 0 or 1 or -1. Thus,

$$|\det_{j_1,\ldots,j_m}(a_1,\ldots,a_{i-1},c,a_{i+1},\ldots,a_m)| \le \sum_{l=1}^{m}|c_{j_l}|,$$

so that each component of p is at most $\|c\|_1$ in magnitude. Then, by using the fact that each column of a node-arc incidence matrix contains at most two nonzero entries and each nonzero entry is either a 1 or a -1, we obtain that each component of $A^T p$ is at most $2\|c\|_1$ in magnitude. This together with (2.8) and (2.7) implies $|r_j^k| \leq |c_j| + 2\|c\|_1$ for all k and all $j \in \{1,..,n\}$, so that the quantity M in the proof of Theorem 3 can be bounded by $3\|c\|_1$ and any limit point of $\{\pi^k\}$ given by (5.1), say π^∞, satisfies $\frac{3\beta\|c\|_1}{1-\beta}$-CS with the limit point of $\{x^k\}$, say x^∞ (which exists by Theorem 2). Since all problem data are integer, the results given in [BeE88], [BeT89, Chap. 5] can be applied to conclude that, for $\frac{3\beta\|c\|_1}{1-\beta} < 1/m$, x^∞ is an optimal solution of (P). By Theorem 2, $\{x^k\}$ converges at least linearly to x^∞. Q.E.D.

For general constraint matrices A (not necessarily a node-arc incidence matrix), we have by a similar argument as above that

$$|p_i| \leq \max_{1 \leq j_1 < \cdots < j_m \leq n} \frac{\det_{j_1,\ldots,j_m}(a_1,\ldots,a_{i-1},c,a_{i+1},\ldots,a_m)}{\det_{j_1,\ldots,j_m}(a_1,\ldots,a_m)},$$

where the maximum is taken over only those indices j for which the denominator is nonzero. This bound can be used to estimate the size of β for which exact optimality is achieved in a manner analogous to that described above for the network flow case. However, unless the matrix A has a certain special property such as total unimodularity, the resulting estimate would likely be very small.

7. Extensions

Consider the following dual of (P)

$$\begin{aligned} &\text{Minimize} \quad -b^T p \\ &\text{subject to} \quad c - A^T p \geq 0. \end{aligned} \tag{D}$$

Instead of Assumption A, we assume that (D) has a finite optimal value and the set $\{p|c > A^T p\}$ is nonempty. The following affine-scaling algorithm, sometimes called the dual affine-scaling algorithm, has been proposed to solve (D) (see for example [AKRV89], [Tsu89]): Given an m-vector p^k satisfying $c > A^T p^k$, compute

$$p^{k+1} = p^k + \frac{(A(S^k)^{-2}A^T)^{-1}b}{\sqrt{b^T(A(S^k)^{-2}A^T)^{-1}b}}, \tag{7.1}$$

where S^k denotes the $n \times n$ diagonal matrix whose j-th diagonal entry is the j-th coordinate of $s^k = c - A^T p^k$. We claim that we can conclude from the results derived in previous sections (cf. Theorems 1 to 4) that $\{p^k\}$ converges at least linearly. To see this, let $\bar{x}$ be any feasible solution of (P) (which exists by linear programming duality), so that $A\bar{x} = b$. Then, by plugging this into (D) and substituting in the slack n-vector $x = c - A^T p$, we can tranform (D) into the form

$$\begin{aligned} &\text{Minimize} \quad \bar{x}^T x \\ &\text{subject to} \quad x = c - A^T p \text{ for some } p, \\ &\qquad\qquad\quad x \geq 0. \end{aligned} \tag{D'}$$

The problem (D') is clearly of the same form as (P) (i.e., minimizing a linear function subject to linear equality and non-negativity constraints). Suppose that we apply (2.1)-(2.2), with stepsize given by (2.4), to (D'). Then, we obtain the iteration

$$x^{k+1} = x^k + \frac{w^k}{\|(X^k)^{-1}w^k\|_2}, \tag{7.2}$$

where w^k is the optimal solution of the subproblem

$$\begin{array}{lll} \text{Minimize} & \bar{x}^T w & \\ \text{subject to} & w = -A^T y & \text{for some } y, \\ & \|(X^k)^{-1}w\|_2^2 \leq n, & \end{array} \tag{7.3}$$

with X^k being the $n \times n$ diagonal matrix whose j-th diagonal entry is x_j^k. By writing down the Kuhn-Tucker optimality conditions for the above subproblem and using the identity $A\bar{x} = b$, we find that $w^k = -A^T(A(X^k)^{-2}A^T)^{-1}b$ and $\|(X^k)^{-1}w^k\|_2 = \sqrt{b^T(A(X^k)^{-2}A^T)^{-1}b}$, so that the iteration (7.2)-(7.3) can be written equivalently as

$$x^{k+1} = x^k - \frac{A^T(A(X^k)^{-2}A^T)^{-1}b}{\sqrt{b^T(A(X^k)^{-2}A^T)^{-1}b}}.$$

On the other hand, by multiplying both sides of (7.1) by $-A^T$ and then adding c to them, we obtain the following updating equation for s^k:

$$s^{k+1} = s^k - \frac{A^T(A(S^k)^{-2}A^T)^{-1}b}{\sqrt{b^T(A(S^k)^{-2}A^T)^{-1}b}},$$

which is clearly of the same form as the updating equation for x^k given above. Hence, Theorem 2 can be applied to conclude that the sequence $\{s^k\}$ converges at least linearly. Since A has full row rank so that p^k is uniquely determined by s^k, this implies that $\{p^k\}$ converges at least linearly. [We remark that analogous results hold for the iterations based on the other stepsize choices (2.3) and (2.5).]

Some, but not all, of our results extend to problems with upper bounds. Suppose that upper bound constraints of the form $x \leq u$ are added to the constraints of (P), where u is a non-negative n-vector some of whose components may have the extended value ∞. To solve this problem, we modify the subproblem (2.1) by replacing the $n \times n$ diagonal matrix X^k inside the ellipsoid constraint by the $n \times n$ diagonal matrix whose j-th diagonal entry is

$$q_j^k = \begin{cases} x_j^k, & \text{if } x_j^k \leq \alpha_j u_j; \\ u_j - x_j^k & \text{otherwise,} \end{cases}$$

where each α_j is a fixed scalar in $(0,1)$. By modifying the stepsize choices (2.3)-(2.5) accordingly so that the iterates remain inside the relative interior of the feasible set, it can be shown that Theorems 1 and 2 as well as Corollary 1 hold for the resulting algorithm.

An open question is the convergence of the iterates to exact optimality without assuming that the stepsize ratio β is sufficiently small. Worst case complexity is another direction for future research.

References

[AKRV89] Adler, I., Karmarkar, N., Resende, M. G. C., and Veiga, G., "Data structures and programming techniques for the implementation of Karmarkar's algorithm," ORSA Journal on Comput. 1 (1989), 84-106.

[AdM88] Adler, I. and Monteiro, R. D. C., "Limiting behavior of the affine scaling continuous trajectories for linear programming problems," Technical Report, Operations Research Center, University of California, Berkeley, CA (April, 1988).

[Bar86] Barnes, E. R., "A variation on Karmarkar's algorithm for solving linear programming problems," Math. Prog. 36 (1986), 174-182.

[BeT89] Bertsekas, D. P. and Tsitsiklis, J. N., Parallel and Distributed Computation: Numerical Methods, Prentice-Hall, Englewood Cliffs, NJ (1989).

[BeE88] Bertsekas, D. P. and Eckstein, J., "Dual coordinate step methods for linear network flow problems," Math. Prog. 42 (1988), 203-243.

[CaS85] Cavalier, T. M. and Soyster, A. L., "Some computational experience and a modification of Karmarkar's algorithm," ISME Working Paper 85-105, Pennsylvania State University, University Park, PA (1985).

[ChK85] Chandru, V. and Kochar, B. S., "A class of algorithms for linear programming," Research Memorandum No. 85-14, Purdue University, Lafayette, Indiana (1985).

[ChK86] Chandru, V. and Kochar, B. S., "Exploring special structures using a variant of Karmarkar's algorithm," Research Memorandum No. 86-10, Purdue University, Lafayette, Indiana (1986).

[Dik67] Dikin, I. I., "Iterative solution of problems of linear and quadratic programming," Soviet Math. Dokl. 8 (1967), 674-675.

[Dik74] Dikin, I. I., "On the speed of an iterative process," Upravlyaemye Sistemi 12 (1974), 54-60.

[Dik88] Dikin, I. I., Letter to the Editor, Math. Prog. 41 (1988), 393-394.

[FoF62] Ford, L. R. and Fulkerson, D. R., Flows in Networks, Princeton University Press, Princeton, NJ (1962).

[Fri55] Frisch, K. R., "The logarithmic potential method of convex programming," Technical Report, University Institute of Economics, Oslo, Norway (1955).

[Gon89] Gonzaga, C. C., "An algorithm for solving linear programming problems in $O(n^3L)$ operations," in Progress in Mathematical Programming: Interior-Point and Related Methods (N. Megiddo, ed.) Springer-Verlag, New York, NY (1989).

[Hof52] Hoffman, A. J., "On approximate solutions of systems of linear inequalities," J. Res. Natl. Bur. Standards 49 (1952), 263-265.

[Kar84] Karmarkar, N., "A new polynomial-time algorithm for linear programming," Combinatorica 4 (1984), 373-395.

[KMY87] Kojima, M., Mizuno, S. and Yoshise, A., "A polynomial- time algorithm for a class of linear complementarity problems," Report No. B-193, Department of Information Sciences, Tokyo Institute of Technology, Tokyo, Japan (1987).

[KoS87] Kortanek, K. O. and Shi, M., "Convergence results and numerical experiments on a linear programming hybrid algorithm," Eur. J. Oper. Res. 32 (1987), 47-61.

[MaS87] Mangasarian, O. L. and Shiau, T.-H., "Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems," SIAM J. Contr. & Optim. 25 (1987), 583-595.

[MSSP88] Marsten, R. E., Saltzman, M. J., Shanno, D. F., and Pierece, G. S., "Implementation of a dual affine interior point algorithm for linear programming," Working Paper, Center for the Management of Information, University of Arizona, Tucson, AZ (1988).

[MeS89] Megiddo, N. and Shub, M., "Boundary behavior of interior point algorithms in linear programming," Math. of Oper. Res. 14 (1989), 97-146.

[MeS88] Mehotra, S. and Sun, J., "A method of analytic centers for quadratically constrained convex quadratic programs," Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL (1988).

[MoM87] Monma, C. L. and Morton, A. J., "Computational experiments with a dual affine variant of Karmarkar's method for linear programming," Technical Report, Bell Communications Research, Morristown, NJ (1987).

[MoA87] Monteiro, R. C. and Adler, I., "Interior path following primal-dual algorithm - part II: convex quadratic programming," Technical Report, Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA (1987); to appear in Math. Prog.

[MAR88] Monteiro, R. D. C., Adler, I., and Resende, M. G. C., "A polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series expansion," Technical Report, Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA (March, 1988).

[PaS82] Papadimitriou, C. H. and Steiglitz, K., Combinatorial Optimization: Algorithms and Complexity, Prentice-Hall, Englewood Cliffs, NJ (1982).

[Pas79] Passty, G. B., "Ergodic convergence to a zero of the sum of monotone operators in Hilbert space," J. Math. Anal. Appl. 72 (1979), 383-390.

[Rob73] Robinson, S. M., "Bounds for errors in the solution set of a perturbed linear program," Lin. Algeb. & Appl. 6 (1973), 69-81.

[Roc84] Rockafellar, R. T., Network Flows and Monotropic Optimization, Wiley-Interscience, New York, NY (1984).

[Tse89] Tseng, P., "Complexity analysis of a linear complementarity algorithm based on a Lyapunov function," Report P-1883, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA (June, 1989).

[TsB87a] Tseng, P. and Bertsekas, D. P., "Relaxation methods for linear programs," Math. of Oper. 12 (1987), 569-596.

[TsB87b] Tseng, P. and Bertsekas, D. P., "Relaxation methods for monotropic programs," Report P-1697, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA (October, 1987); to appear in Math. Prog.

[Tsu89] Tsuchiya, T., "Global convergence property of the affine scaling methods for primal degenerate linear programming problems," Technical report, The Institute of Statistical Mathematics, Tokyo, Japan (August, 1989).

[Vai87] Vaidya, P. M., "An algorithm for linear programming which requires $O(((m+n)n^2+(m+n)^{1.5})L$ arithmetic operations," Technical Report, AT&T Bell Laboratories, Murray Hill, NJ (1987).

[Van89] Vanderbei, R. J., "Affine-scaling for linear programs with free variables," Math. Prog. 43 (1989), 31-44.

[VaL88] Vanderbei, R. J. and Lagarias, J. C., "I. I. Dikin's convergence result for the affine-scaling algorithm," Technical report, AT&T Bell Laboratories, Murray Hill, NJ (1988).

[VMF86] Vanderbei, R. J., Meketon, M. S., and Freeman, B. A., "A modification of Karmarkar's linear programming algorithm," Algorithmica 1 (1986), 395-407.

[Ye88] Ye, Y., "A class of potential functions for linear programming," Department of Management Sciences, The University of Iowa, Iowa City, Iowa (1988).

[Ye89] Ye, Y., "An extension of Karmarkar's algorithm and the trust region method for quadratic programming," in Progress in Mathematical Programming: Interior-Point and Related Methods (N. Megiddo, ed.) Springer-Verlag, New York, NY (1989).