A simple proof of a primal affine scaling method

Romesh Saigal*

*Department of Industrial and Operations Engineering,
The University of Michigan, Ann Arbor, Michigan 48109-2117, USA*

E-mail: rsaigal@engin.umich.edu

In this paper, we present a simpler proof of the result of Tsuchiya and Muramatsu on the convergence of the primal affine scaling method. We show that the primal sequence generated by the method converges to the interior of the optimum face and the dual sequence to the analytic center of the optimal dual face, when the step size implemented in the procedure is bounded by 2/3. We also prove the optimality of the limit of the primal sequence for a slightly larger step size of $2q/(3q-1)$, where q is the number of zero variables in the limit. We show this by proving the dual feasibility of a cluster point of the dual sequence.

Keywords: Linear programming, affine scaling methods, interior point methods.

1. Introduction

We consider here the linear programming problem:

$$\begin{aligned}
\text{minimize} \quad & c^T x \\
& Ax = b, \\
& x \geq 0,
\end{aligned} \qquad \text{LP}$$

where A is an $m \times n$ matrix and b and c are appropriate vectors. We also assume that

ASSUMPTION 1

The linear program has an interior solution.

ASSUMPTION 2

The objective function is not constant on the feasible region.

ASSUMPTION 3

The matrix A has rank m.

There have been some significant developments in the convergence theory of the affine scaling method. In the series of papers (Tsuchiya [14,15], Dikin [5,6] and Tsuchiya and Muramatsu [16]), convergence is shown when the step size is restricted to two-thirds of the distance to the boundary. Prior to these results, convergence of the primal sequence was known, but its optimality was only known under non-degeneracy, see for example Vanderbei and Lagarias [18].

On the basis of the global convergence analysis by the use of a local potential function, developed by Tsuchiya [14,15], and the analysis of reduction of potential function developed by Dikin [6], Dikin [5] proved the global convergence of the long step affine scaling method when the step size is restricted to 1/2. Stimulated by [6], but independently, Tsuchiya and Muramatsu [16] proved that the method converges to the optimum solution if the step size is restricted to 2/3, and that in this case, the primal sequence converges to the relative interior of the optimal face of the primal and the dual sequence to the analytic center of the optimal face of the dual. Hall and Vanderbei [8] present an example that shows the 2/3 is tight in the sense that with a larger step size, the dual sequence may not converge. In this paper, we give a simple proof of the two-third result. Another simplification of the proof can be found in Monteiro et al. [10]. Our proof parallels their proof in many places, but is simpler. Wherever possible, we will point out the similarity of our results to theirs. In addition, we also prove the convergence of the primal sequence to optimality for a slightly larger step size of $2q/(3q-1)$, where q is the number of zero variables in the limit of the primal sequence. In this case, we show that the dual sequence has a subsequence converging to a dual feasible solution. For the step size strictly less than $2q/(3q-1)$, we show that the primal converges to the interior of the primal optimal face. Independently, this result has also been obtained by Gonzaga [7].

In addition to the introduction, this paper has nine other sections. In section 2, we present the affine scaling method we will deal with in this paper. In section 3, we study some important properties of the sequences generated by this method, and in section 4 we present some preliminary results. In section 5, we prove the convergence of the primal sequence. In section 6, we develop additional properties of the sequences, and in section 7 we introduce the concept of local potential function. In section 8, we prove the convergence of the dual sequence when the step size is restricted to 2/3, while in section 9 we prove the optimality of the limit of the primal sequence for a step size restricted by $2q/(3q-1)$. The final section is the appendix.

2.   Primal affine scaling method

We refer the reader to Barnes [2] and Vanderbei et al. [17] for more on this method. The iterative scheme of the primal affine scaling method is the following, and is known as the large step version. The method presented in [2] and [17] is known as the short step version.

Step 0. Let x^0 be an interior point solution, $0 < \alpha < 1$, and let $k = 0$.

Step 1. Tentative solution to the dual:

$$y^k = (AX_k^2 A^T)^{-1} AX_k^2 c.$$

Step 2. Tentative dual slack:

$$s^k = c - A^T y^k.$$

If $s^k \leq 0$, then STOP. The solution is unbounded.

Step 3. Min-ratio test:

$$\theta_k = \min \left\{ \frac{\| X_k s^k \|}{x_j^k s_j^k} : s_j^k > 0 \right\}$$

$$= \frac{\| X_k s^k \|}{\phi(X_k s^k)},$$

where $\phi(x) = \max_j x_j$. If $\theta_k = 1$, set $\alpha = 1$.

Step 4. Next interior point:

$$x^{k+1} = x^k - \alpha \theta_k \frac{X_k^2 s^k}{\| X_k s^k \|}.$$

Step 5. Iterative step: If $x_j^{k+1} = 0$ for some j, then STOP. x^{k+1} is an optimal solution. Otherwise, set $k = k + 1$ and go to step 1.

It is shown in Vanderbei and Lagarius [18] that if the algorithm stops in step 5, then an optimal solution has been found.

2.1. ELLIPSOIDAL APPROXIMATING PROBLEM

As is now well known (see, for example, Barnes [2]), the direction $X_k^2 s^k / \| X_k s^k \|$ generated by this algorithm is obtained by solving the following ellipsoidal approximating problem:

At the kth iterate, given an interior point $x^k > 0$, the direction $v^k = X_k^2 s^k$ to the next iterate x^{k+1}, generated in step 4, is obtained by scaling the solution to the following ellipsoidal approximating problem:

$$\begin{aligned} \text{maximize} \quad & c^T v \\ & Av = 0, \qquad\qquad \text{EAP} \\ & \| X_k^{-1} v \|^2 \leq 1. \end{aligned}$$

2.2.   LEAST-SQUARES PROBLEM

Another important observation about this method is that the dual estimate y^k, calculated in step 1, is readily seen as generated by the following "least-squares" problem:

$$y^k = \arg\min{}_y \|X_k(c - A^T y)\|$$

and we note that this estimate, given a primal feasible solution x^k, is chosen to minimize the complementary slackness violation without the nonnegativity constraint on the dual variable $s = c - A^T y$.

We now state some important properties of the sequences generated by this method.

3.   Properties of sequences

Let $\{x^k\}$, $\{y^k\}$, $\{s^k\}$ and $\{v^k\} = \{X_k^2 s^k\}$ be the sequences generated by the primal affine scaling algorithm. If these sequences are finite, then the last point of the first sequence solves the linear program, or the problem has an unbounded solution. Thus, we henceforth assume that they are not finite. We now establish some important properties of these sequences.

THEOREM 1

$x^k > 0$ and $\theta_k \geq 1$ for all k. Also,

$$c^T v^k = c^T X_k^2 s^k = \| X_k s^k \|^2 = \| X_k^{-1} v^k \|^2$$

for all k.

Proof

The first part readily follows. To see the second part first observe that, from the definition $c^T v^k = c^T X_k^2 s^k$, and

$$c^T X_k^2 s^k = c^T X_k (I - X_k A^T (AX_k^2 A^T)^{-1} AX_k) X_k c$$
$$= \| P_k X_k c \|^2,$$

where $P_k = I - X_k A^T (AX_k^2 A^T)^{-1} AX_k$ is the projection matrix into the null space $\mathcal{N}(AX_k)$ of the matrix AX_k, with $P_k = P_k^2 = P_k^T$. Now

$$P_k X_k c = X_k(c - A^T(AX_k^2 A^T)^{-1} AX_k^2 c) = X_k s^k$$

and we are done.   □

THEOREM 2

$\{c^T x^k\}$ is a strictly decreasing sequence with

$$c^T x^{k+1} = c^T x^k - \alpha \theta_k \| X_k s^k \|.$$

Either this sequence diverges to $-\infty$, or it is bounded, and thus converges with $\| X_k s^k \| \to 0$ as $k \to 0$, and $\infty > \sum_{k=1}^{\infty} c^T (x^k - x^{k+1})$.

Proof

It is readily seen, from step 4 and theorem 1, that

$$c^T x^{k+1} = c^T x^k - \alpha \theta_k \frac{c^T X_k^2 s^k}{\| X_k s^k \|}$$

$$= c^T x^k - \alpha \theta_k \| X_k s^k \|.$$

From assumption 2, $\| X_k s^k \| \neq 0$, and so $\{c^T x^k\}$ is strictly decreasing. Hence, it either diverges to $-\infty$ or is bounded. In the latter case, since every bounded monotone sequence converges and $\alpha \theta_k \geq \alpha >$, $X_k s^k \to 0$. Now

$$\sum_{k=1}^{\infty} c^T (x^k - x^{k+1}) = c^T x^1 - c^*,$$

where $\lim_{k \to \infty} c^T x^k = c^*$ and we are done. $\qquad \square$

We now prove some preliminary results, and then establish the convergence of sequences $\{x^k\}$, $\{y^k\}$ and $\{s^k\}$.

4. Preliminary results

In this section, we prove some preliminary results and two theorems related to the ellipsoidal approximating problem discussed in section 2.1.

LEMMA 3

Let r_j and $s_j > 0$ for each $j = 1, \ldots, l$ be arbitrary real numbers. Then

$$\frac{\left| \sum_{j=1}^{l} r_j \right|}{\sum_{j=1}^{l} s_j} \leq \max_j \frac{|r_j|}{s_j}.$$

Proof

Let

$$\frac{|r_k|}{s_k} = \max_j \frac{|r_j|}{s_j}.$$

Then for each $j = 1,\ldots,l$, $s_j|r_k| \geq s_k|r_j|$. Adding all these l inequalities, we have

$$|r_k| \sum_{j=1}^{l} s_j \geq s_k \sum_{j=1}^{l} |r_j| \geq s_k \left| \sum_{j=1}^{l} r_j \right|$$

and we are done.   □

The next theorem was proved by Dikin [4]. Our proof is from Vanderbei and Lagarias [18]. Other proofs can be found in Todd [12] and Vavasis [19].

THEOREM 4

For every $x > 0$, the function $\|(AX^2A^T)^{-1}AX^2p\| \leq q(A)\|p\|$, where $q(A) > 0$ is a constant.

Proof

By the Cauchy-Binet Theorem and Cramer's rule we can write the ith component

$$((AX^2A^T)^{-1}AX^2p)_i = \frac{\sum_J (x_{j_1} \ldots x_{j_m})^2 \det(A_J) \det(A_J^{(i)})}{\sum_{J,} (x_{j_1} \ldots x_{j_m})^2 (\det(A_J))^2}$$

where $A^{(i)}$ is the matrix obtained by replacing the ith row of A by p^T, and the sum is over all possible J which are ordered sets of m elements $1 \leq j_1 \leq \ldots \leq j_m \leq n$. Since a term in the denominator is zero only when the term in the numerator is zero, from lemma 3 we can use

$$q(A, p) = \sqrt{n} \max_J \frac{|\det(A_J^{(i)})|}{|\det(A_J)|}$$

as an upper bound on $\|(AX^2A^T)^{-1}AX^2p\|$. Since the determinant of a matrix is a linear function of a row of the matrix, we can assume that $q(A, p) \leq q(A)\|p\|$ for some constant $q(A) > 0$ which is only a function of A.   □

Given a $k \times l$ matrix Q of rank k, a k vector $c \neq 0$ and an $l \times l$ diagonal matrix D with positive diagonal entries, consider the problem:

$$\text{maximize} \quad c^T x$$
$$x^T QDQ^T x \leq 1. \tag{1}$$

The following result establishes a relation between the objective function value and the solution vector of this problem.

THEOREM 5

Let $\hat{x}$ solve problem (1). There exists a constant $p(Q, c) > 0$ such that

$$c^T \hat{x} \geq p(Q, c) \| \hat{x} \|.$$

Proof

Let $c, B_1, \ldots, B_{k-1}$ be a basis for $\mathbb{R}^k$ with $B = (B_1 \ldots, B_{k-1})$ an orthonormal basis for the orthogonal complement of the one-dimensional subspace spanned by c. Thus, $c^T B = 0$ and $B^T B = I$. Expressing the columns of Q in this basis, we obtain

$$Q = ca^T + BR,$$

where a is an l vector and R is a $(k - 1) \times l$ matrix. Since Q has full rank k, so does (a, R^T). Also, let

$$x = uc + Bv,$$

where u is a scalar. Rewriting problem (1) in the variables u and v, we get

$$c^T x = uc^T c,$$

$$x^T QDQ^T x = (uc^T c)^2 a^T Da + 2(uc^T c)a^T DR^T v + v^T RDR^T v,$$

and if $w = c^T cu$, problem (1) is equivalent to

$$\text{maximize } w$$

$$w^2 a^T Da + 2wa^T DR^T v + v^T RDR^T v \leq 1.$$

Let λ be the multiplier on the constraint. Then the optimality conditions for this problem can be stated as follows:

$$1 + \lambda(2wa^T Da + 2a^T DR^T v) = 0, \tag{2}$$

$$wRDa + RDR^T v = 0, \tag{3}$$

and we note that equation (3) has the solution

$$\hat{v} = -w(RDR^T)^{-1} RDa$$

and let $\hat{w}$ be the solution obtained by substituting $\hat{v}$ into equation (2). Then, after computing $\hat{x}$, we note that

$$c^T \hat{x} = \hat{w},$$

$$\hat{x} = \hat{u}c + B\hat{v}$$

$$= \frac{\hat{w}}{c^T c} c - \hat{w}B(RDR^T)^{-1} RDa.$$

Thus,

$$\frac{\hat{x}}{c^T\hat{x}} = \frac{c}{c^Tc} - B(RDR^T)^{-1}RDa,$$

and, noting that $\|B\| \le 1$, we get

$$\frac{\|\hat{x}\|}{|c^T\hat{x}|} \le \frac{1}{\|c\|} + \|(RDR^T)^{-1}RDa\|.$$

Since $x = 0$ is an interior feasible solution, $c^T\hat{x} > 0$. If, from theorem 4, $q(R, a)$ bounds the second term on the right-hand side above, since both R and a are only functions of Q and c, we get our result with $p(Q, c) = 1/(1/\|c\| + q(R, a))$. $\quad\square$

We now establish an important corollary to the preceding theorem. It was proved by Tseng and Luo [13] and Tsuchiya [14]. See also theorem 2.6 of Monteiro et al. [10]. Given an $x > 0$, consider the ellipsoidal approximating problem EAP of section 2.2.

COROLLARY 6

There exists a constant $p(A, c) > 0$ such that for every $x > 0$, if $\hat{v}$ solves EAP,

$$\|\hat{v}\| \le p(A, c)c^T\hat{v}.$$

Proof

Let the set of vectors $\{Q_1, \dots, Q_{n-m}\}$ be an orthonormal basis for the null space $\mathcal{N}(A)$ of A; i.e., for $Q = (Q_1, \dots, Q_{n-m})$, $Q^TQ = I$. Since, in the above problem, $v \in \mathcal{N}(A)$, we note that $c^Tv = c_n^Tv$, where c_n is the projection of c into the null space of A. Define $\bar{c}$ and u such that

$$c_n = Q\bar{c},$$

$$v = Qu.$$

Expressing the problem in variables u, we obtain an equivalent problem:

$$\text{maximize} \quad \bar{c}^Tu$$

$$u^TQ^TX_k^{-2}Qu \le 1.$$

Let u^k solve this problem. From theorem 5, there is a constant $p(Q, \bar{c}) > 0$ such that

$$c^Tv^k = \bar{c}^Tu^k$$

$$\ge p(Q, \bar{c})\|u^k\|$$

$$= p(Q, \bar{c})\|v^k\|,$$

since $\|v^k\|^2 = (u^k)^TQ^TQu^k = \|u^k\|^2$. Since $Q, \bar{c}$ are only functions of A, c, we are done. $\quad\square$

Let R be an arbitrary $r \times l$ matrix, with rank $\leq r$ and let $b \in \mathcal{R}(R)$. Also, let $z \in \mathbb{R}^n$ be any vector. Consider the following problem:

$$\text{minimize} \quad \|z - x\|^2$$

$$Rx = b.$$

We can then prove:

THEOREM 7

There is a constant $q(R)$ such that for every z, for the solution $x(z)$ of the above problem we have

$$\|z - x(z)\| \leq q(R)\|Rz - b\|.$$

Proof

Let $\bar{R}$ be an $\bar{r} \times l$ submatrix of R which has full row rank. Since $b \in \mathcal{R}(R)$,

$$\{x : Rx = b\} = \{x : \bar{R}x = \bar{b}\}$$

for some $\bar{b}$. Then, it is readily seen that $x(z)$ solves the problem if and only if for some y,

$$2(z - x(z)) - \bar{R}^T y = 0,$$

$$\bar{R}x(z) = \bar{b}$$

and the result is readily seen by setting $q(R) = \|\bar{R}^T(\bar{R}\bar{R}^T)^{-1}\|$. $\qquad \square$

The following simple lemma on the natural logarithmic function is important. This result was first proved by Tsuchiya [14].

LEMMA 8

Let $w \in \mathbb{R}^q$ and $0 < \lambda < 1$ be such that $w_j \leq \lambda$. Then

$$\sum_{j=1}^{q} \log(1 - w_j) \geq -e^T w - \frac{\|w\|^2}{2(1 - \lambda)}.$$

Proof

The result follows as a consequence of the following facts: When $a < 0$, by considering the function $g(x) = \log(1 + x) - x + x^2/(2(1 - \lambda))$, it is readily confirmed that its derivative is positive when $x > 0$, and thus $g(|a|)$ is an increasing function of $|a|$ and thus

$$\log(1 - a) \geq -a - \frac{a^2}{2(1 - \lambda)}.$$

To see that second fact, let $0 \le a \le \lambda$. Then

$$\log(1 - a) = -a - \frac{a^2}{2} - \frac{a^3}{3} -$$

$$\ge -a - \frac{a^2}{2(1 - a)}$$

$$\ge -a - \frac{a^2}{2(1 - \lambda)}.$$

The result now readily follows from the above two facts.   □

5.   Convergence of primal sequence

We assume here that the sequence $\{c^T x^k\}$ is bounded. We now prove that the primal sequence $\{x^k\}$ converges. We prove this in the next theorem.

THEOREM 9

Let the assumptions 1 through 3 hold, $\{c^T x^k\}$ be bounded, and the sequence $\{x^k\}$ be generated by the primal affine scaling algorithm. Then x^k converges, say to x^*.

Proof

Consider the approximating problem of section 2.1. Note that

$$\alpha \theta_k \hat{v}^k = x^k - x^{k+1},$$

where $\hat{v}^k = X_k^2 s^k / \| X_k s^k \|$ is the solution to the ellipsoidal approximating problem EAP of section 2.1. We will now show that $\sum_{k=1}^{\infty} \alpha \theta_k \| \hat{v}^k \| < \infty$ and thus the sequence $\{x^k\}$ converges. But from theorem 2 and corollary 6,

$$\infty > c^T x^1 - c^* = \sum_{k=1}^{\infty} c^T (x^k - x^{k+1}) = \sum_{k=1}^{\infty} \alpha \theta_k c^T \hat{v}^k \ge p(A, c) \sum_{k=1}^{\infty} \alpha \theta_k \| \hat{v}^k \|,$$

and thus $\{x^k\}$ converges, say to x^*, and we are done.   □

6.   More on the sequences

We have seen in the previous section that the primal sequence converges. Let x^* be such that

$$\lim_{k \to \infty} x^k = x^*.$$

Define

$$N = \{j : x_j^* = 0\},$$

$$B = \{j : x_j^* > 0\},$$

and

$$q = |N|,$$

where $|N|$ is the number of elements in N. Let F be the face of $P = \{x : Ax = b, x \geq 0\}$ such that x^* lies in the relative interior of the face F, i.e., it is the smallest face of P containing x^*. We now prove an important property of the sequence $\{x^k\}$. Such a result was first proved by Tsuchiya [14]. Also see lemma 3.1 of [10].

THEOREM 10

There is a $\delta > 0$ and an $M > 0$ such that for each $k = 1, 2, \ldots$

$$\frac{c^T x^k - c^*}{\| x^k - x^* \|} \geq \frac{1}{M}, \quad \frac{c^T x^k - c^*}{\sum_{j \in N} x_j^k} \geq \delta, \quad \text{and} \quad \frac{c^T x^k - c^*}{\sum_{j \in B} | x_j^k - x_j^* |} \geq \delta.$$

Proof

Note that using the results of theorem 2 and corollary 6 with $M = q(A, c)$,

$$c^T x^k - c^* = \sum_{j=0}^{\infty} c^T (x^{k+j} - x^{k+j+1})$$

$$\geq \frac{1}{M} \sum_{j=0}^{\infty} \| x^{k+j} - x^{k+j+1} \|$$

$$\geq \frac{1}{M} \left\| \sum_{j=0}^{\infty} (x^{k+j} - x^{k+j+1}) \right\|$$

$$= \frac{1}{M} \| x^k - x^* \|$$

and we have our first result.

Also,

$$c^T x^k - c^* \geq \frac{\| x_N^k \|}{M} \geq \frac{\sum_{j \in N} x_j^k}{\sqrt{q} M},$$

$$c^T x^k - c^* \geq \frac{\| x_B^k - x_B^* \|}{M} \geq \frac{\sum_{j \in B} | x_j^k - x_j^* |}{\sqrt{n - q} M}$$

and we have the remaining parts with

$$\delta = \min \left\{ \frac{1}{\sqrt{q} M}, \frac{1}{\sqrt{n - q} M} \right\}.$$

$\square$

The following is a well-known result on the convergence rate of the sequence $\{c^T x^k - c^*\}$. Also, see lemma 2.10 of [10].

THEOREM 11

There exists an $L \geq 1$ such that for all $k \geq L$,

$$c^T x^{k+1} - c^* \leq \left(1 - \frac{\alpha}{\sqrt{n}}\right)(c^T x^k - c^*).$$

Proof

Consider the problem EAP of section 2.1. We note that $X_k^2 s^k / \| X_k s^k \|$ solves this problem and that $(x^k - x^*)/\| X_k^{-1}(x^k - x^*) \|$ is feasible for it. Thus, from the above fact and theorem 1,

$$\frac{c^T x^k - c^*}{\| X_k^{-1}(x^k - x^*) \|} \leq \frac{c^T X_k^2 s^k}{\| X_k s^k \|} = \| X_k s^k \|. \tag{4}$$

Now $\| X_k^{-1}(x^k - x^*) \|^2 = \| e_B - X_{k,B}^{-1} x_B^* \|^2 + \| e_N \|^2$, and since $X_{k,B}^{-1} x_B^* \to e_B$, there is an $L \geq 1$ such that for all $k \geq L$, $\| e_B - X_{k,B}^{-1} x_B^* \|^2 \leq n - q$. Thus, for each $k \geq L$, $\| X_k^{-1}(x^k - x^*) \| \leq \sqrt{n}$. Now, as

$$c^T x^{k+1} - c^* = c^T x^k - c^* - \frac{\alpha}{\phi(X_k s^k)} c^T X_k^2 s^k,$$

we have

$$\frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} \leq 1 - \alpha \frac{\| X_k s^k \|}{c^T x^k - c^*}$$

and the result follows from the equation (4).   □

Now define

$$D = \{(y, s) : A_B^T y = c_B, A_N^T y + s_N = c_N, s_B = 0\}$$

to be the set of all tentative dual solutions that are complementary to each solution in F. We can readily establish the following:

LEMMA 12

$D \neq \varnothing$.

Proof

From step 1 and theorem 4, the sequences $\{y^k\}$ and $\{s^k\}$ are bounded, and thus have cluster points y^* and s^*, respectively. From theorem 2, since $X_k s^k \to 0$, $s_B^* = 0$. The result now follows since $A^T y^* + s^* = c$.   □

Consider the sequence $\{u^k\}$ defined by

$$u^k = \frac{X_k s^k}{c^T x^k - c^*}.$$

We now state an important result about this sequence. Its proof is given in the appendix.

THEOREM 13

The sequence $\{u^k\}$ has the following properties:

1. It is bounded.
2. There is an $L \geq 1$ such that for each $k \geq L$,
    (a) $\|u^k\|^2 = \|u_N^k\|^2 + \gamma_k$ and $\sum_{k=L}^{\infty} |\gamma_k| < \infty$.
    (b) $e^T u_N^k = 1 + \delta_k$ and $\sum_{k=L}^{\infty} |\delta_k| < \infty$.
    (c) $1/\alpha \geq \phi(u_N^k) \geq 1/2q$.
    (d) $\phi(u^k) = \phi(u_N^k)$.

7. Local potential function

To show the convergence of the dual sequence, it is necessary to control the step size. In particular, the step size α is required to be less than or equal to $2/3$. This result is achieved by the use of a local version of the potential function, introduced by Tsuchiya [14], which we now introduce.

For any $x > 0$ with $c^T x - c^* > 0$ and $N \subset \{1,\dots,n\}$ with $q = |N|$, define

$$F_N(x) = q \log(c^T x - c^*) - \sum_{j \in N} \log(x_j).$$

For the sequence $\{x^k\}$, we can prove

THEOREM 14

There exists an $L \geq 1$ such that for all $k \geq L$,

$$F_N(x^{k+1}) - F_N(x^k) = q \log\left(1 - \theta\|w_N^k\|^2 - \theta\left(\frac{2\delta_k}{q} + \gamma_k\right)\right) - \sum_{j \in N} \log(1 - \theta w_j^k),$$

where

$$w_N^k = u_N^k - \frac{1}{q} e, \quad \tilde{\alpha} = \frac{\alpha}{\phi(u_N^k)}, \quad \theta = \frac{q\tilde{\alpha}}{q - \tilde{\alpha}}$$

and

$$\sum_{k=L}^{\infty} |\delta_k| < \infty, \quad \sum_{k=L}^{\infty} |\gamma_k| < \infty.$$

Proof

   Since
$$c^T x^{k+1} - c^* = c^T x^k - c^* - \frac{\alpha}{\phi(X_k s^k)} c^T X_k^2 s^k,$$

from theorem 1 and simple algebra we see that

$$\frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} = 1 - \frac{\alpha}{\phi(u^k)} \|u^k\|^2 \tag{5}$$

and, for each $j \in N$,

$$\frac{x_j^{k+1}}{x_j^k} = 1 - \frac{\alpha}{\phi(u^k)} u_j^k.$$

From theorem 13 part 2(a), there is an $L \geq 1$ such that for all $k \geq L$,

$$\frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} = 1 - \frac{\alpha}{\phi(u_N^k)} \|u_N^k\|^2 - \frac{\alpha \gamma_k}{\phi(u_N^k)}.$$

Now let $k \geq L$, and note that from theorem 13 part 2(b),

$$\|w_N^k\|^2 = \|u_N^k\|^2 - \frac{2e^T u_N^k}{q} + \frac{1}{q} = \|u_N^k\|^2 - \frac{1}{q} - \frac{2\delta_k}{q},$$

and we have our result by observing that

$$\frac{q}{q - \tilde{\alpha}} (1 - \tilde{\alpha}\|u_N^k\|^2 - \tilde{\alpha}\gamma_k) = 1 - \theta\|w_N^k\|^2 - \theta\left(\frac{2\delta_k}{q} + \gamma_k\right)$$

and

$$\frac{q}{q - \tilde{\alpha}} (1 - \tilde{\alpha}u_j^k) = 1 - \theta w_j^k$$

and the constant $q/(q - \tilde{\alpha})$ cancels.   □

8.   Convergence of dual sequence

   We are now ready to prove that with the step size α less than or equal to 2/3 the dual sequence converges to the analytic center of the optimal face of the dual polytope. Throughout this section, we will assume that the sequence $\{c^T x^k\}$ is bounded and we now define the analytic center of the optimal dual face as the solution (y^*, s^*) to the following problem:

$$\text{maximize} \quad \sum_{j \in N} \log s_j$$

$$(y, s) \in D, \qquad\qquad \text{ACP}$$

$$s_N > 0.$$

The solution to ACP must satisfy the following K.K.T. conditions:

$$A_B x_B + A_N x_N = 0,$$
$$s_N^{-1} + x_N = 0,$$
$$A_B^T y = c_B,$$
$$A_N^T y + s_N = c_N,$$
$$s_N > 0,$$
$$s_B = 0.$$

We now establish a lemma.

LEMMA 15

If the analytic center defined by ACP exists, it is unique.

Proof

The uniqueness follows from the strict concavity of the log function. $\square$

We are now ready to prove our main theorem.

THEOREM 16

Let the sequence generated by the affine scaling algorithm be infinite, and the step size α be less than or equal to $2/3$. Then, there exist vectors x^*, y^* and s^* such that

1. $x^k \to x^*$,

2. $y^k \to y^*$,

3. $s^k \to s^*$, and $s^* \geq 0$.

The limits are optimal solutions for their respective problems, and satisfy the strict complementarity condition. In addition, the dual solution (y^*, s^*) converges to the analytic center of the optimal dual face, and the primal solution to the relative interior of the optimal primal face.

Proof

From lemma 8 and theorems 13 and 14 and the fact that for any $a < 1$, $\log(1 - a) < -a$, we obtain an L such that for each $k \geq L$ with $0 < \phi(w_N^k)$, we have

$$F_N(x^{k+1}) - F_N(x^k) \leq -q\theta\left(\|w_N^k\|^2 + \frac{2\delta_k}{q} + \gamma_k\right) + \theta\left(e^T w_N^k + \theta\frac{\|w_N^k\|^2}{2(1 - \theta\phi(w_N^k))}\right)$$

$$= \theta \| w_N^k \|^2 \left(-q + \frac{\theta}{2(1 - \theta \phi(w_N^k))} \right) - \theta(q\gamma_k + \delta_k)$$

$$= \theta \| w_N^k \|^2 \left(-q + \frac{\alpha}{2(1 - \alpha)} \frac{1}{\phi(u_N^k)} \right) - \theta(q\gamma_k + \delta_k). \tag{6}$$

From theorem 13 part 2(c), it can be shown that $\theta \geq \alpha^2$. Now, as $w_N^k = u_N^k - (1/q)e$, we note that, when $\phi(w_N^k) > 0$, $\phi(u_N^k) \geq (1 + \phi(w_N^k))/q$.

Thus, we have

$$F_N(x^{k+1}) - F_N(x^k) \leq \theta \| w_N^k \|^2 \left(-q + \frac{\alpha}{2(1 - \alpha)} \frac{q}{1 - \phi(w_N^k)} \right) - \theta(q\gamma_k + \delta_k).$$

For each $\alpha \leq 2/3$, $\alpha/(2(1 - \alpha)) \leq 1$. Hence, when $\phi(w_N^k) > 0$,

$$F_N(x^{k+1}) - F_N(x^k) \leq - \frac{q\theta \phi(w_N^k) \| w_N^k \|^2}{1 + \phi(w_N^k)} - \theta(q\gamma_k + \delta_k).$$

In the case $\phi(w_N^k) \leq 0$, then $w_N^k \leq 0$. So, $\log(1 - \theta w_j^k) \geq 0$. From theorem 14 and the fact that $\log(1 - a) \leq -a$, we have

$$F_N(x^{k+1}) - F_N(x^k) \leq - q\theta \| w_N^k \|^2 - \theta(q\gamma_k + 2\delta_k).$$

From theorem 10, $\sum_{k=L}^{\infty} (F_N(x^{k+1}) - F_N(x^k)) > -\infty$ and, from theorem 13 parts 2(a) and (b), $0 \leq \sum_{k=L}^{\infty} (|\delta_k| + \gamma_k) < \infty$. In the case $\{\phi(w_N^k)\}$ has a strictly positive or a strictly negative cluster point, we must have $\| w_N^k \| \to 0$. In the case $\phi(w_N^k) \to 0$, since $e^T w_N^k = \delta_k$ and $\delta_k \to 0$, $w_N^k \to 0$. Thus, we get

$$\lim_{k \to \infty} u_N^k = \frac{1}{q} e. \tag{7}$$

Now, consider the sequences $\{y^k\}$, $\{s^k\}$, $\{x_j^k/(c^T x^k - c^T x^*)\}$ for each $j \in N$ and $\{x_j^k - x_j^*/(c^T x^k - c^T x^*)\}$ for each $j \in B$. Let $s^k \to s^*$ on some subsequence $K \subset \{1, 2, \ldots\}$. From theorem 4, steps 1 and 2 and theorem 10, these are bounded for each $1 \leq j \leq n$. Thus, on some common subsequence $K \subset K'$,

$$y^k \to y^*, \quad s^k \to s^*$$

$$\frac{q x_j^k}{c^T x^k - c^T x^*} \to a_j \text{ for each } j \in N$$

and

$$\frac{q(x_j^k - x_j^*)}{c^T x^k - c^T x^*} \to b_j \quad \text{for each } j \in B.$$

In view of (7), $a_j > 0$ for each $j \in N$. Also, for each $j \in N$, since

$$s_j^k = \frac{c^T x^k - c^T x^*}{x_j^k}\left(w_j^k - \frac{1}{q}\right), \quad \text{as } w_j^k \to 0,$$

we note that $\lim_{k \in K} s_j^k = a_j^{-1}$.

Now, $A_N x_N^k + A_B x_B^k = A_B x_B^*$; we see that

$$A_N a + A_B b = 0$$

and thus $s_j = a_j^{-1}$ for each $j \in N$ and $x_N = -a$, $x_B = -b$ and $s_B = 0$, $s_N = s_N^*$, $y = y^*$ solve the K.K.T. conditions for the analytic center problem ACP. Thus, s_B^k converges to a^{-1}, the analytic center, on each convergent subsequence, and hence parts (2) and (3) of the theorem are established by lemma 15. We finish the proof by observing that x^* is primal feasible, (y^*, s^*) is dual feasible and the pair satisfies the complementary slackness theorem (by theorem 2) and therefore are optimal solutions for the respective problems. The strict complementarity holds as $s_N^* > 0$.   □

The following is a sharp bound on the convergence rate of $\{c^T x^k - c^*\}$. See also theorem 4.2 of [10].

THEOREM 17

Let $\alpha \le 2/3$. Then

$$\lim_{k \to \infty} \frac{c^T x^{k+1} - c^*}{c^T x^k - c^*} = 1 - \alpha.$$

Proof

Follows readily from theorem 13 parts 2 (a), (c) and (d), equation (5) and equation (7).   □

9.   Convergence to optimality for larger step size

In this section, we will show convergence to optimality for a slightly larger step size than 2/3. This result has also, independently, been obtained by Gonzaga [7]. We now establish a lemma.

LEMMA 18

Assume that on some subsequence K, $\|w_N^k\| \not\to 0$. Then there is an $L \ge 1$ such that for every $k \ge L$ and $k \in K$,

$$\phi(u_N^k) \ge \frac{1}{q} + \frac{\|w_N^k\|}{\sqrt{q(q-1)}} - \frac{2|\delta_k|}{q}.$$

Proof

Consider the optimization problem:

$$\text{minimize} \quad \rho$$

$$\sum_{i=1}^{q-1} x_i + \rho = 1 + \delta_k,$$

$$x_i - \rho \leq 0,$$

$$\sum_{i=1}^{q-1} \left(x_i - \frac{1}{q} \right)^2 + \left(\rho - \frac{1}{q} \right)^2 = \| w_N^k \|^2 .$$

It is readily confirmed that the solution of this optimization problem is:

$$\rho - \frac{1}{q} = \sqrt{\frac{\| w_N^k \|^2 - \frac{\delta^2}{q}}{q(q-1)}} + \frac{\delta_k}{q}$$

and thus we obtain our result by observing that $\phi(u_N^k) \geq \rho$ for every u_N^k with $e^T u_N^k = 1 + \delta_k$ and $u_N^k - (1/q)e = w_N^k$; and for every $a > b > 0$, $\sqrt{a^2 - b^2} \geq a - b$. $\square$

We are now ready to prove the main theorem:

THEOREM 19

Let $2/3 < \alpha \leq 2q/(3q - 1)$. Then, there exists a subsequence $K \subset \{1, 2, \ldots\}$ and vectors x^*, y^* and s^* such that

1. $\lim_{k \in K} x^k \to x^*$,

2. $\lim_{k \in K} y^k \to y^*$,

3. $\lim_{k \in K} s^k \to s^*$, and $s^* \geq 0$.

Thus, x^* and (y^*, s^*) are optimum solutions of the dual pair. Also, when $\alpha < 2q/(3q - 1)$, they satisfy the strict complementarity condition.

Proof

As a result of theorem 10, either on some subsequence K', $F(x^{k+1}) - F(x^k) \geq 0$ for each $k \in K'$ or $\lim(F(x^{k+1}) - F(x^k)) = 0$. Thus, on some subsequence, from equation (6)

$$\liminf_{k \in K'} \left(-q + \frac{\alpha}{2(1 - \alpha)} \frac{1}{\phi(u_N^k)} \right) \geq 0.$$

Let $K \subset K'$ be such that as $k \in K$ goes to infinity, $x^k \to x^*$, $y^k \to y^*$, $s^k \to s^*$, $u_N^k \to u_N^*$ and $\|w_N^k\| \to \theta \geq 0$. If $\theta = 0$, then we are done since $u_N^* = (1/q)e > 0$. Otherwise, from lemma 18

$$-q + \frac{\alpha}{2(1-\alpha)} \frac{q}{1 + \theta\sqrt{\frac{q}{q-1}}} \geq 0.$$

Letting $\alpha \leq 2q/(3q-1)$, it is readily confirmed that

$$\theta \leq \frac{1}{\sqrt{q(q-1)}}.$$

Since $\|u_N^* - (1/q)e\| = \theta$ and $e^T u_N^* = 1$, we can conclude that $u_N^* \geq 0$ since it lies in the largest inscribed hypersphere in the simplex $\{x : e^T x = 1, x \geq 0\}$. The strict complementarity follows since under the hypothesis, u_N^* lies strictly inside the hypersphere, and is thus strictly positive, and we are done. $\qquad\square$

Appendix

We now prove theorem 13. These proofs parallel proofs of similar results in [10].

Proof

To see (1), note that from theorem 1, for any $(\tilde{y}, \tilde{s}) \in D$, $\|X_k s^k\|^2 = c^T v^k = (A^T \tilde{y} + \tilde{s})^T v^k = \tilde{s}_N^T v_N^k = (X_{k,N} \tilde{s}_N)^T (X_{k,N}^{-1} v_N^k) \leq \phi(x_N^k)\|\tilde{s}_N\|\|X_k s^k\|$. From theorem 10, we obtain that $\|u^k\| \leq M\|\tilde{s}_N\|$ for $M = p(A, c)$.

To see 2(a), from corollary 6 and for any $(\tilde{y}, \tilde{s}) \in D$, $\|v_B^k\| \leq \|v^k\| \leq Mc^T v^k = M\tilde{s}_N v_N^k \leq M\|\tilde{s}_N\|\|v_N^k\| \leq M\phi(x_N^k)\|\tilde{s}_N\|\|X_{k,N} s_N^k\|$. Also,

$$\|u_B^k\| \leq \frac{\|X_{k,B}^{-1}\|\|v_B^k\|}{c^T x^k - c^*} \leq \|X_{k,B}^{-1}\| M\phi(x_N^k)\|\tilde{s}_N\|\|u_N^k\|.$$

From theorem 9, $\phi(x_N^k) \to 0$ and $x_B^k \to x_B^* > 0$. Thus there exist an $\hat{M}$ and an $\bar{L} \geq 1$ such that for every $k \geq \bar{L}$,

$$\gamma_k = \|u_B^k\|^2 \leq (M\|X_{k,B}^{-1}\|\phi(x_N^k)\|\tilde{s}_N\|\|u_N^k\|)^2 < 1, \; M\|X_{k,B}^{-1}\|\|\tilde{s}_N\|\|u_N^k\| < \hat{M}.$$

Since $\|u^k\|^2 = \|u_N^k\|^2 + \|u_B^k\|^2$, we are done if $\sum_{k=\bar{L}}^{\infty} \phi(x_N^k) < \infty$. But from theorem 10 and theorem 11, there is an $\tilde{L} \geq 1$ such that for all $k \geq \tilde{L}$,

$$\phi(x_N^k) \leq \|x^k - x^*\|$$
$$\leq M(c^T x^k - c^*)$$
$$\leq M\left(1 - \frac{\alpha}{\sqrt{n}}\right)^{k-\tilde{L}} (c^T x^{\tilde{L}} - c^*)$$

and thus the sum is finite.

To see part 2(b), let $(\tilde{y}^k, \tilde{s}^k)$ solve the following problem:

$$\text{minimize} \quad \|s^k - s\|$$
$$(y, s) \in D.$$

From theorem 7, there is a constant $\tilde{N} > 0$ such that $\|\tilde{s}^k - s^k\| \leq \tilde{N} \|s_B^k\|$. Since $\{s^k\}$ is bounded, so is $\{\tilde{s}^k\}$, and let $\|s^k\| \leq \hat{N}$ and $\|\tilde{s}^k\| \leq \hat{N}$ for each k. Also, since s^k solves the least-squares problem (LSQ), we get $\|X_{k,B} s_B^k\|^2 + \|X_{k,N} s_N^k\|^2 \leq \|X_{k,N} \tilde{s}_N^k\|^2$.

Now,

$$\|s_B^k\|^2 \leq \|X_{k,B}^{-1}\|^2 \|X_{k,B} s_B^k\|^2$$

$$\leq \|X_{k,B}^{-1}\|^2 (\|X_{k,N} \tilde{s}_N^k\|^2 - \|X_{k,N} s_N^k\|^2)$$

$$= \|X_{k,B}^{-1}\|^2 (X_{k,N} \tilde{s}_N^k - X_{k,N} s_N^k)^T X_{k,N} \tilde{s}_N^k + X_{k,N} s_N^k)$$

$$\leq \|X_{k,B}^{-1}\|^2 \phi(x_N^k)^2 \|\tilde{s}_N^k - s_N^k\| \|\tilde{s}_N^k + s_N^k\|$$

$$\leq \|X_{k,B}^{-1}\|^2 \phi(x_N^k)^2 \, 2\hat{N}\tilde{N} \|s_B^k\|.$$

Thus, $\|\tilde{s}_N^k - s_N^k\| \leq 2\hat{N}\tilde{N}^2 \|X_{k,B}^{-1}\|^2 \phi(x_N^k)^2$. Now, $c^T x^k - c^* = c^T(x^k - x^*) = (\tilde{s}^k + A^T \tilde{y}^k)^T (x^k - x^*) = (\tilde{s}_N^k)^T x_N^k$. Thus, from theorem 10,

$$|e^T u_N^k - 1| = \frac{|(s_N^k)^T x_N^k - (\tilde{s}_N^k)^T x_N^k|}{c^T x^k - c^*}$$

$$\leq \frac{\|s_N^k - \tilde{s}_N^k\| \|x_N^k\|}{M^{-1} \|x_N^k\|}$$

$$\leq 2M\hat{N}\tilde{N}^2 \|X_{k,B}^{-1}\|^2 \phi(x_N^k)^2.$$

Thus, there exists an $\hat{L} \geq 1$ such that for all $k \geq \hat{L}$,

$$|\delta_k| = |e^T u_N^k - 1| \leq 2M\hat{N}\tilde{N}^2 \|X_{k,B}^{-1}\|^2 \phi(x_N^k)^2 < 1.$$

The required sum can be shown to be finite by the same argument as for the sum in 2(a).

To see part 2(c), using the result of part 2(b), there exists an $L^* \geq \tilde{L}$ such that for all $k \geq L^*$,

$$\phi(u_N^k) \geq \frac{1 + \delta_k}{q} \geq \frac{1}{2q}.$$

Also, since $1 - (\alpha/\phi(u_N^k)\|u_N^k\|^2 > 0$, we have

$$\phi(u_N^k) \leq \frac{\|u_N^k\|^2}{\phi(u_N^k)} \leq \frac{1}{\alpha}.$$

To see part 2(d), using the results of 2(a) and 2(c), we note that there is an $L' \geq L^*$ such that for all $k \geq L'$,

$$\phi(u_B^k) \leq \|u_B^k\| \leq \hat{M}\phi(x_N^k) < \frac{1}{2q} \leq \phi(u_N^k)$$

and we are done with $L = \max\{\bar{L}, \hat{L}, L'\}$, since $\phi(u^k\} = \max\{\phi(u_B^k), \phi(u_N^k)\}$. $\qquad\square$

References

[1] I. Adler, N.K. Karmarkar, M.G.C. Resende and G. Veiga, An implementation of Karmarkar's algorithm for linear programming, Mathematical Programming 44(1989)297–335.

[2] E.R. Barnes, A variation on Karmarkar's algorithm for solving linear programming problems, Mathematical Programming 36(1986)174–182.

[3] I.I. Dikin, Iterative solution of problems of linear and quadratic programming, Soviet Mathematics Doklady 8(1967)674–675.

[4] I.I. Dikin, On the convergence of an iterative process, Upravlyaemye Sistemi 12(1974)54–60 (in Russian).

[5] I.I. Dikin, The convergence of dual variables, Technical Report, Siberian Energy Institute, Irkutsk, Russia (1991).

[6] I.I. Dikin, Determination of interior point of one system of linear inequalities, Kibernetica and System Analysis 1 (1992).

[7] C. Gonzaga, unpublished note (1992).

[8] L.A. Hall and R.J. Vanderbei, Two-thirds is sharp for affine scaling, OR Letters 13(1993) 197–201.

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

[10] R.D.C. Monteiro, T. Tsuchiya and Y. Wang, A simplified global convergence proof of the affine scaling algorithm, Annals of Operations Research 47(1993)443–482.

[11] M. Muramatsu and T. Tsuchiya, A convergence analysis of a long-step variant of the projective scaling algorithm, Research Memorandum 454, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan (November 1992).

[12] M.J. Todd, A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm, Operations Research 38(1990)1006–1018.

[13] P. Tseng and Z.Q. Luo, On the convergence of the affine-scaling algorithm, Mathematical Programming 56(1992)301–319.

[14] T. Tsuchiya, Global convergence of the affine-scaling methods for degenerate linear programming problems, Mathematical Programming 52(1991)377–404.

[15] T. Tsuchiya, Global convergence property of the affine scaling method for primal degenerate linear programming problems, Mathematics of Operations Research 17(1992)527–557.

[16] T. Tsuchiya and M. Muramatsu, Global convergence of a long-step affine scaling algorithm for degenerate linear programming problems, Research Memorandum 423, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan (January 1992; revised September 1992), to appear in SIAM Journal on Optimization.

[17] R.J. Vanderbei, M.J. Meketon and B.A. Freedman, A Modification of Karmarkar's linear programming algorithm, Algorithmica 1(1986)395–407.

[18] R.J. Vanderbei and J.C. Lagarias, I.I. Dikin's convergence result for the affine-scaling algorithm, in: *Mathematical Developments Arising from Linear Programming: Proceedings of a Joint Summer Research Conference*, ed. J.C. Lagarias and M.J. Todd, Bowdoin College, Brunswick, Maine, USA (1988), vol. 114 of Contemporary Mathematics (American Mathematical Society, Providence, RI, USA, 1990) pp. 109–119.

[19] S.A. Vavasis, Stable numerical algorithms for equilibrium systems, TR92-1280, Department of Computer Science, Cornell University, Ithaca, New York (1992).