A Simplified Global Convergence Proof of the Affine Scaling Algorithm*

R. D. C. Monteiro† T. Tsuchiya‡ Y. Wang§

October 6, 1992
Revised: March 10, 1993

Abstract

This paper presents a simplified and self-contained global convergence proof for the affine scaling algorithm applied to degenerate linear programming problems. Convergence of the sequence of dual estimates to the center of the optimal dual face is also proven. In addition, we give a sharp rate of convergence result for the sequence of objective function values. All these results are proved with respect to the long step version of the affine scaling algorithm in which we move a fraction λ, where $\lambda \in (0, 2/3]$, of the step to the boundary of the feasible region.

1 Introduction

The affine scaling algorithm, introduced by Dikin [6] in 1967, is one of the simplest and most efficient interior point algorithms for solving linear programming (LP) problems. Because of the theoretical and practical importance of the affine scaling algorithm, there are a number of papers which study its global and local convergence [3, 6, 7, 8, 11, 16, 19, 20, 21, 22, 23, 24] and the behavior of its associated continuous trajectories [2, 4, 16, 25]. Like in the simplex algorithm, the analysis of the affine scaling algorithm for (primal) degenerate LP problems is much harder than for (primal) nondegenerate LP problems.

Recently, Dikin [8] and Tsuchiya and Muramatsu [22] have succeeded in proving the global convergence for degenerate LP problems of the long step version of the affine scaling algorithm, that is the version in which the next iterate is determined by taking a fixed fraction $\lambda \in (0, 1)$ of the whole step to the boundary of the feasible region. Their study have definitely settled satisfactory answers to the case where $\lambda \in (0, 2/3]$.

Unfortunately, none of these two papers are self-contained. In order to fully understand Tsuchiya and Muramatsu's paper, it is necessary to read two of Tsuchiya's preceding papers

*This research was supported by the National Science Foundation (NSF) under Grant No. DDM-9109404 and the Overseas Research Scholars of the Ministry of Education, Science and Culture of Japan
†Department of Systems and Industrial Engineering, University of Arizona, Tucson, Arizona 85721.
‡The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-Ku, Tokyo, 106, Japan.
§Department of Systems and Industrial Engineering, University of Arizona, Tucson, Arizona 85721.

[20, 21] which contain nontrivial matrix decomposition results and long derivations. The same can be said about Dikin's paper which refers the reader to Tsuchiya's paper [20] in some important steps.

The main goal of this paper is to give a compact and self-contained global convergence proof of the long step version of the affine scaling algorithm. Two major simplifications are made in our analysis compared to the analysis of [22]. The first one is obtained when showing that the sequence of primal iterates converges. The way we prove this result is based on techniques contained in Tseng and Luo [19]. Another major simplification is obtained by showing that the the nontrivial and long matrix decomposition results derived in Tsuchiya [20, 21] are no longer needed. Instead, we use techniques which have been employed in Adler and Monteiro [2].

All the main results given in this paper have already been proved in the revised version of Tsuchiya and Muramatsu [22]. Namely, assuming that $\lambda \leq 2/3$, we show that the sequence of primal iterates converges to a point lying in the relative interior of the primal optimal face and that the sequence of dual estimates converge to the analytical center of the dual optimal face. Under the same assumptions, we also show that the sequence of objective function values converges Q-linearly with convergence rate equal to $1 - \lambda$. As a natural consequence of the results of this paper, we also derive the global convergence of the long step version of the affine scaling algorithm under primal nondegeneracy, for any fraction $\lambda \in (0, 1)$.

The history behind the affine scaling algorithm is as follows. The affine scaling algorithm was introduced by the Russian mathematician I. Dikin [6] in 1967, who published a convergence proof [7] in 1974. Dikin assumes primal nondegeneracy and takes a full step to the boundary of the inscribed ellipsoid. In 1985, the affine scaling algorithm was rediscovered in the west community by Barnes [3], Karmarkar and Ramakrishnan [15] and Vanderbei, Meketon and Freedman [24] following the introduction of the projective scaling algorithm by Karmarkar [14] in 1984. Both Vanderbei et al. [24] and Barnes [3] assume primal and dual nondegeneracy in the global convergence analysis, but, as opposed to the ellipsoid version of Barnes [3] and Dikin [7], Vanderbei et al. studies the long step version of the affine scaling algorithm for any $\lambda \in (0, 1)$. The long step version was also studied by Gonzaga [11] under primal nondegeneracy only. Results demonstrating the good performance of the long step version (with $\lambda = 0.95$ or $\lambda = 0.995$) of the affine scaling algorithm have been reported in several papers (eg., Adler, Karmarkar, Resende and Veiga [1] and Monma and Morton [17]).

The papers mentioned in the previous paragraph all assume primal nondegeneracy. Several papers gradually showed that this assumption could also be removed. Adler and Monteiro [2] and Witzgall et al. [25] independently investigated the convergence of the continuous trajectories of the affine scaling algorithm and their associated dual estimate trajectories without imposing any nondegeneracy condition. Assuming only dual nondegeneracy, Tsuchiya [21] shows the global convergence of the ellipsoid version of the affine scaling algorithm with fraction $\leq 1/8$. In a subsequent paper, Tsuchiya [20] shows that the dual nondegeneracy condition could be removed from the analysis of [21]. One of the main ideas used in these two papers is the use of a potential function to analyze the local behavior of the affine scaling algorithm near the boundary of the feasible region. This potential function was also used in the subsequent analysis of Dikin [8] and Tsuchiya and Muramatsu [22]. Slightly prior to [20], Tseng and Luo [19] proved the global convergence of a very short step version of the affine scaling algorithm under no nondegeneracy condition, where the step-size is taken to be $2^{-O(L)}$ and L is the input size of the problem.

Finally, the long step version was studied for degenerate LP problems by Dikin [8] and Tsuchiya and Muramatsu [22] with the fraction λ satisfying $\lambda \leq 1/2$ and $\lambda \leq 2/3$, respectively. Convergence of the sequence of dual estimates to the analytical center of the dual optimal face is also discussed in these two papers. One of the important ideas used in these two papers is given in the paper by Dikin [9] which analyzes the reduction of the potential function using long steps.

The global convergence of the affine scaling algorithm when the fraction λ is larger than 2/3 is still an open problem. However, as was conjectured by Tsuchiya, Hall and Vanderbei [12] shows that the sequence of dual estimates may no longer converge when $\lambda > 2/3$ (see also Section 6 of the revised version of [22]).

Our paper is organized as follows. In Section 2, we show that the sequence of primal iterates converge to a point lying in a dual degenerate face, for any fraction $\lambda \in (0,1)$. We end the section showing how this result can be used to derive the global convergence of the affine scaling algorithm when applied to primal nondegenerate LP problems. In Section 3, we show the global convergence of the affine scaling algorithm when applied to degenerate LP problems, for any $\lambda \leq 2/3$. In Section 4, also assuming that $\lambda \leq 2/3$, we show that the sequence of dual estimates converges to the analytical center of the dual optimal face and that the sequence of objective function values converge to the optimal value Q-linearly with convergence rate equal to $1 - \lambda$.

One may find that some "well-known results" about the affine scaling algorithm, including its global convergence for nondegenerate LP problems, the boundedness of the sequence of dual estimates and etc., are duplicated in this paper. Our aim was to make the paper as much self-contained as possible, and to present most of the important convergence results on the affine scaling algorithm in an organized way. Consequently, this article is also a survey or review on the convergence theory of the affine scaling algorithm. (While this paper was being revised for publication, we learned of another more recent survey article by Saigal [18].)

The following notation is used throughout our paper. We denote the vector of all ones by e. Its dimension is always clear from the context. $\mathbb{R}^n$, $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$ denote the n-dimensional Euclidean space, the nonnegative orthant of $\mathbb{R}^n$ and the positive orthant of $\mathbb{R}^n$, respectively. The set of all $m \times n$ matrices with real entries is denoted by $\mathbb{R}^{m\times n}$. Given an index set $J \subseteq \{1, \ldots, n\}$ and a vector $w \in \mathbb{R}^n$, we denote by w_J the subvector of w corresponding to J. Similarly, if E is an $m \times n$ matrix then E_J denotes the $m \times |J|$ submatrix of E corresponding to J. For a vector w, we let $\chi[w]$ denote the largest component of w. The Euclidean norm, the 1-norm and the ∞-norm are denoted by $\|\cdot\|$, $\|\cdot\|_1$ and $\|\cdot\|_\infty$, respectively. The diagonal matrix corresponding to a vector w is denoted by $diag(w)$. If J is a finite index set then $|J|$ denotes its cardinality, that is the number of elements of J. The superscript ^{T} denotes transpose.

2 Convergence of the Primal Sequence

In this section, we state the main terminology and assumptions used throughout our paper. We then briefly describe the affine scaling algorithm. The main result of this section (stated as Theorem 2.6 and Theorem 2.9) shows that the sequence of primal iterates generated by the affine scaling algorithm converges to a point lying in a dual degenerate face of the feasible region, that is a face on which the objective function is constant. Using this result, we give an alternative proof of the global convergence of the affine scaling algorithm applied to primal nondegenerate LP problems (see Theorem 2.11).

The most crucial result used in the proof of the convergence result mentioned above is Theorem 2.5. This theorem is based on Theorem 2 of Tseng and Luo [19].

Consider the following linear programming problem

$$\begin{array}{lll} \text{minimize}_{\;x} & c^T x & \\ \text{subject to} & Ax = b, & x \geq 0, \end{array} \tag{1}$$

and its associated dual problem

$$\begin{array}{lll} \text{maximize}_{\;(y,s)} & b^T y & \\ \text{subject to} & A^T y + s = c, & s \geq 0, \end{array} \tag{2}$$

where $A \in \mathbb{R}^{m \times n}$, $c, x, s \in \mathbb{R}^n$ and $b, y \in \mathbb{R}^m$.

We next introduce some notation which will be used throughout our paper. We frequently consider (possibly, empty) subsets of the affine space $\{x|Ax = b\}$ or of the feasible region $\{x|Ax = b, x \geq 0\}$ obtained by making some components of the vector x equal to 0. We adopt the convention of denoting by N the set of indices for which the associated variables are enforced to be equal to 0, and by B the complement of N with respect to $\{1, ..., n\}$. Formally, we consider the following notation:

$$\mathcal{P}_N \equiv \{x \in \mathbb{R}^n \mid A_B x_B = b, x_N = 0\}; \tag{3}$$

$$\mathcal{P}_N^+ \equiv \{x \in \mathcal{P}_N \mid x_B \geq 0\}; \tag{4}$$

$$\mathcal{P}_N^{++} \equiv \{x \in \mathcal{P}_N \mid x_B > 0\}; \tag{5}$$

$$\mathcal{D}_N \equiv \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n \mid A_B^T y = c_B, A_N^T y + s_N = c_N, s_B = 0\}; \tag{6}$$

$$\mathcal{D}_N^+ \equiv \{(y, s) \in \mathcal{D}_N \mid s_N \geq 0\}; \tag{7}$$

$$\mathcal{D}_N^{++} \equiv \{(y, s) \in \mathcal{D}_N \mid s_N > 0\}. \tag{8}$$

When $N = \emptyset$, we denote the sets $\mathcal{P}_N$, $\mathcal{P}_N^+$ and $\mathcal{P}_N^{++}$ by $\mathcal{P}$, $\mathcal{P}^+$ and $\mathcal{P}^{++}$, respectively. The sets $\mathcal{P}^+$ and $\mathcal{P}^{++}$ are the sets of feasible solutions and strictly feasible solutions of problem (1). Similarly, when $N = \{1, ..., n\}$, we denote the sets $\mathcal{D}_N$, $\mathcal{D}_N^+$ and $\mathcal{D}_N^{++}$ by $\mathcal{D}$, $\mathcal{D}^+$ and $\mathcal{D}^{++}$, respectively. $\mathcal{D}^+$ and $\mathcal{D}^{++}$ are the sets of feasible solutions and strictly feasible solutions of problem (2).

We impose the following assumptions throughout this paper.

Assumption 1: $Rank(A) = m$;

Assumption 2: The objective function $c^T x$ is not constant over the feasible region of (1);

Assumption 3: Problem (1) has an interior feasible solution, that is $\mathcal{P}^{++} \neq \emptyset$;

Assumption 4: Problem (1) has an optimal solution.

We now introduce important functions which are used in the description and in the analysis of the affine scaling algorithm. For every $x \in \mathbb{R}^n_{++}$, let

$$
\begin{aligned}
y(x) &\equiv (AX^2A^T)^{-1}AX^2c, && (9a)\\
s(x) &\equiv c - A^Ty(x), && (9b)\\
d(x) &\equiv X^2s(x) = X[I - XA^T(AX^2A^T)^{-1}AX]Xc, && (9c)
\end{aligned}
$$

where $X \equiv diag(x)$. We note that Assumption 1 implies that the inverse of AX^2A^T exists for every $x > 0$. The pair $(y(x), s(x))$ is called the *dual estimate* associated with the point x. $d(x)$ is referred to as the *affine scaling direction* associated with x.

The following result provides characterizations of the affine scaling direction $d(x)$ and the dual estimate $(y(x), s(x))$ as optimal solutions of certain quadratic programming problems.

Proposition 2.1 *The following statements hold.*

(a) *For every $x > 0$, $d(x)$ is the unique optimal solution of the following QP problem:*

$$
\begin{aligned}
&\text{maximize}_{\ p} && c^Tp - \tfrac{1}{2}\|X^{-1}p\|^2\\
&\text{subject to} && Ap = 0,
\end{aligned}
\qquad (10)
$$

where $X \equiv diag(x)$.

(b) *For every $x > 0$, $(y(x), s(x))$ is the unique solution of the following QP problem:*

$$
\begin{aligned}
&\text{minimize}_{\ (y,s)} && \tfrac{1}{2}\|Xs\|^2\\
&\text{subject to} && A^Ty + s = c.
\end{aligned}
\qquad (11)
$$

We leave the trivial proof of Proposition 2.1 to the reader.

We are now in a position to describe the affine scaling algorithm. For a good motivation of the method, we refer the reader to Dikin [6], Barnes [3], Vanderbei, Meketon and Freedman [24] and Vanderbei and Lagarias [23].

Affine Scaling Algorithm:

initially: Choose a fixed constant $\lambda \in (0, 1)$ and assume that $x^0 \in \mathcal{P}^{++}$ is available;

for $k = 0, 1, 2, \ldots$

$$
\begin{aligned}
d^k &= d(x^k);\\
X^k &= diag(x^k);\\
x^{k+1} &= x^k - \frac{\lambda}{\chi[(X^k)^{-1}d^k]}d^k && (12)
\end{aligned}
$$

end for

Recall that $\chi[\cdot]$ is used to denote the maximum component of a vector. We note that Assumptions 1–4 imply that $d(x)$ must have at least one positive component so that $\chi[X^{-1}d(x)] > 0$. Hence, the expression which determines x^{k+1} in the affine scaling algorithm is well-defined. Observe also that if λ were equal to 1, the next iterate would lie in the boundary of the feasible region. Thus, since we choose $\lambda \in (0,1)$, the next iterate is ensured to be an interior point.

Lemma 2.2 *For any $(\tilde{y}, \tilde{s}) \in \mathcal{D}$, there holds*

$$\tilde{s}^T d(x) = c^T d(x) = \|X^{-1}d(x)\|^2 = \|Xs(x)\|^2, \qquad \forall x > 0, \tag{13}$$

where $X \equiv diag(x)$.

Proof. By Proposition 2.1, we know that

$$Ad(x) = 0, \qquad A^T y(x) + s(x) = c. \tag{14}$$

Hence, we obtain

$$\tilde{s}^T d(x) = (c - A^T\tilde{y})^T d(x) = c^T d(x) - \tilde{y}^T A d(x) = c^T d(x). \tag{15}$$

From (14), we have that $(y(x), s(x)) \in \mathcal{D}$. Hence, using (15) we obtain

$$\begin{aligned} c^T d(x) &= s(x)^T d(x) \\ &= [Xs(x)]^T [X^{-1}d(x)] \\ &= \|X^{-1}d(x)\|^2 \\ &= \|Xs(x)\|^2, \end{aligned}$$

where the third equality follows from (9c). ∎

From now on, we denote the sequence of dual estimates $\{(y(x^k), s(x^k))\}$ simply by $\{(y^k, s^k)\}$. The following basic result has been proved in several papers (see for example Vanderbei and Lagarias [23]). For the sake of completeness, we give its proof here.

Proposition 2.3 *The following statements hold for the affine scaling algorithm.*

(a) $x^k \in \mathcal{P}^{++}$ *for all* $k \geq 0$;

(b) *The sequence of objective function values $\{c^T x^k\}$ strictly decreases and converges to a finite value;*

(c) $X^k s^k \longrightarrow 0$ *as* $k \longrightarrow \infty$, *where* $X^k \equiv diag(x^k)$.

Proof. Assume that $x^k \in \mathcal{P}^{++}$. Using (12), we obtain

$$(X^k)^{-1}x^{k+1} = e - \lambda \frac{(X^k)^{-1}d^k}{\chi[(X^k)^{-1}d^k]} \geq e - \lambda e = (1-\lambda)e > 0,$$

from which (a) follows. Using (12) and Lemma 2.2, we obtain

$$\begin{aligned} c^T x^{k+1} &= c^T x^k - \frac{\lambda}{\chi[(X^k)^{-1} d^k]} c^T d^k \\ &\le c^T x^k - \frac{\lambda}{\|(X^k)^{-1} d^k\|} \|(X^k)^{-1} d^k\|^2 \\ &= c^T x^k - \lambda \|X^k s^k\|, \end{aligned} \tag{16}$$

and hence

$$0 < \|X^k s^k\| \le \lambda^{-1}(c^T x^k - c^T x^{k+1}), \qquad \forall k \ge 0,$$

from which (b) and (c) obviously follows. ∎

Proposition 2.3 does not guarantee that the sequence $\{x^k\}$ converges and not even that $\{x^k\}$ is a bounded sequence. We now concentrate our efforts in showing that the sequence $\{x^k\}$ converges. The following result, which is due to Hoffman, plays an important role in proving this fact. For a proof of the result below, we refer the reader to Hoffman [13].

Lemma 2.4 *Let $F \in \mathbb{R}^{p\times q}$ be given. Then, there exists a constant $C = C(F)$ with the following property: for $f \in \mathbb{R}^p$ such that the system $Fw = f$ is feasible and $z \in \mathbb{R}^q$, there exists a solution $\bar{w}$ of $Fw = f$ such that*

$$\|\bar{w} - z\| \le C\|f - Fz\|.$$

The following crucial result is based on Theorem 2 of Tseng and Luo [19]. Its proof uses Hoffman's Lemma.

Theorem 2.5 *Let $v \in \mathbb{R}^q$ and $E \in \mathbb{R}^{l\times q}$ be given. Then, there exists a constant $L = L(v, E)$ with the property that for any diagonal matrix $D > 0$, the (unique) optimal solution $p^* = p^*(D)$ of*

$$\begin{aligned} &\text{maximize}_{\ p} && v^T p - \tfrac{1}{2}\|Dp\|^2 \\ &\text{subject to} && Ep = 0 \end{aligned} \tag{17}$$

satisfies

$$\|p^*\| \le L\, v^T p^*. \tag{18}$$

Proof. If $v \in Range(E^T)$ then $p^*(D) = 0$ for every diagonal matrix $D > 0$, and (18) obviously holds if we choose any $L \ge 0$. We may therefore assume that $v \notin Range(E^T)$. In this case, it is easy to verify that $v^T p^* > 0$ for every $p^* = p^*(D)$. To show that (18) also holds in this case, assume by contradiction that there exists a sequence of positive diagonal matrices $\{D^k\}$ such that

$$\lim_{k\to\infty} \frac{\|p^k\|}{v^T p^k} = \infty, \tag{19}$$

where p^k is the (unique) optimal solution of (17) with $D = D^k$. This implies that there exist a constant $M > 0$, an index set $J \subseteq \{1, \ldots, q\}$, $J \ne \emptyset$, and a subsequence $\{p^k\}_{k\in K}$ with the property that

$$\frac{|p_j^k|}{v^T p^k} \le M, \qquad \forall k \in K,\ \forall j \notin J; \tag{20}$$

$$\lim_{k\in K} \frac{|p_j^k|}{v^T p^k} = \infty, \qquad \forall j \in J. \tag{21}$$

Consider the following linear system

$$\begin{aligned} v^T p &= v^T p^k, & (22a)\\ Ep &= 0, & (22b)\\ p_j &= p_j^k, \qquad \forall j \notin J, & (22c) \end{aligned}$$

and note that p^k is a solution of this system. By Lemma 2.4, (22) has a solution $\hat{p}^k$ such that

$$\begin{aligned} \|\hat{p}^k\| &\le C(|v^T p^k| + \sum_{j\notin J} |p_j^k|) \\ &= C(v^T p^k + \sum_{j\notin J} |p_j^k|), \end{aligned} \tag{23}$$

where C is a positive constant depending only on v and E. Hence, we have by (20) that

$$\begin{aligned} \|\hat{p}^k\| &\le C[v^T p^k + (n-|J|)Mv^T p^k] \\ &= C[1+(n-|J|)M]v^T p^k \\ &= M_1 v^T p^k, \end{aligned} \tag{24}$$

for all $k \in K$, where $M_1 \equiv C[1+(n-|J|)M]$. Also, (21) implies that there exists $l > 0$ such that for all $k \ge l$, $k \in K$

$$|p_j^k| > M_1 v^T p^k, \qquad \forall j \in J. \tag{25}$$

From (24) and (25), we have

$$|p_j^k| > \|\hat{p}^k\|, \qquad \forall j \in J,\ \forall k \ge l,\ \forall k \in K. \tag{26}$$

Hence, it follows from (22c) and (26) that

$$\begin{aligned} \|D^k \hat{p}^k\|^2 &= \sum_{j\in J} (D_j^k \hat{p}_j^k)^2 + \sum_{j\notin J} (D_j^k \hat{p}_j^k)^2 \\ &< \sum_{j\in J} (D_j^k p_j^k)^2 + \sum_{j\notin J} (D_j^k p_j^k)^2 \\ &= \|D^k p^k\|^2, \qquad \forall k \ge l,\ \forall k \in K. \end{aligned} \tag{27}$$

Relations (22a) and (27) then imply

$$v^T \hat{p}^k - \frac{1}{2}\|D^k \hat{p}^k\|^2 > v^T p^k - \frac{1}{2}\|D^k p^k\|^2, \qquad \forall k \ge l,\ \forall k \in K, \tag{28}$$

which together with (22b) contradicts the fact that p^k is an optimal solution of (17) with $D = D^k$. ■

The following result was proved in Tseng and Luo [19] and Tsuchiya [20]. Our proof which is an immediate consequence of Theorem 2.5 is based on the presentation of Tseng and Luo [19]. (See also [5] and [18] for alternative proofs.)

Theorem 2.6 *For the affine scaling algorithm, the following statements hold.*

(a) *There exists a constant $M = M(c, A) > 0$ such that*

$$\|d^k\| \le M c^T d^k, \qquad \forall k \ge 0; \tag{29}$$

(b) *The sequence $\{x^k\}$ converges to a point $x^* \in \mathcal{P}^+$.*

(c) *For all $k \ge 0$, we have $\|x^k - x^*\| \le M(c^T x^k - v^*)$ where $v^* \equiv c^T x^* = \lim_{k\to\infty} c^T x^k$.*

Proof. By Proposition 2.1, we know that the affine scaling direction d^k is the optimal solution of

$$\begin{array}{ll} \text{maximize}_{\ p} & c^T p - \frac{1}{2}\|(X^k)^{-1} p\|^2 \\ \text{subject to} & Ap = 0, \end{array}$$

where $X^k \equiv diag(x^k)$. Then, (a) follows directly from Theorem 2.5. We now prove (b). Using (29), we obtain

$$\left\| \frac{\lambda}{\chi[(X^k)^{-1} d^k]} d^k \right\| \le M c^T \left(\frac{\lambda}{\chi[(X^k)^{-1} d^k]} d^k \right), \qquad \forall k \ge 0,$$

which, in view of (12), is equivalent to

$$\|x^{k+1} - x^k\| \le M(c^T x^k - c^T x^{k+1}), \qquad \forall k \ge 0. \tag{30}$$

Since $\{c^T x^k\}$ converges, (30) implies that

$$\sum_{k=0}^{\infty} \|x^{k+1} - x^k\| \le \lim_{k\to\infty} M(c^T x^0 - c^T x^k) < \infty.$$

This implies that $\{x^k\}$ is a Cauchy sequence, and therefore a convergent sequence. Clearly, $\lim_{k\to\infty} x^k \equiv x^* \in \mathcal{P}^+$. We now show (c). From (30), we also obtain

$$\|x^l - x^k\| \le M(c^T x^k - c^T x^l), \qquad \forall l > k \ge 0.$$

Letting $l \to \infty$, we obtain

$$\|x^k - x^*\| \le M(c^T x^k - v^*), \qquad \forall k \ge 0.$$

■

The constant $M(c, A)$ appearing in Theorem 2.6 will henceforth be denoted by M. We will now show that the smallest face containing the point x^* is a dual degenerate face. First, we formalize the notion of a dual degenerate face. A nonempty set of the form $\mathcal{P}_N^+$ where $N \subseteq \{1, \ldots, n\}$ is called a *face* of $\mathcal{P}^+$. The face is called *dual degenerate* if the objective function $c^T x$ is constant over the face. If the set $\mathcal{P}_N^+$ is a dual degenerate face and $\mathcal{P}_N^{++} \neq \emptyset$ then we call N a *dual degenerate index set.* The following result gives conditions under which N is a dual degenerate index set.

Proposition 2.7 *Let $N \subseteq \{1, \ldots, n\}$ be given. The index set N is dual degenerate if and only if $\mathcal{P}_N^{++} \neq \emptyset$ and $c_B \in Range(A_B^T)$.*

The equivalent condition stated in Proposition 2.7 can also be simply stated as $\mathcal{P}_N^{++} \neq \emptyset$ and $\mathcal{D}_N \neq \emptyset$. We leave the trivial proof of Proposition 2.7 to the reader.

We now define the following index sets associated with x^*:

$$
\begin{aligned}
B_* &= \{i \mid x_i^* > 0\}, &\quad (31a)\\
N_* &= \{i \mid x_i^* = 0\}. &\quad (31b)
\end{aligned}
$$

Then, the smallest face containing x^* is the set $\mathcal{P}_{N_*}^{+}$.

The following result is well-known and its proof can be found in Vanderbei and Lagarias [23]. For the sake of completeness, we give in the Appendix a new proof of this result based on Hoffman's lemma.

Proposition 2.8 *The set $\{(y(x), s(x)) \mid x > 0\}$ is bounded. In particular, the sequence of the dual estimates $\{(y^k, s^k)\}$ generated by the affine scaling algorithm is bounded.*

Using the fact that $\{(y^k, s^k)\}$ is a bounded sequence, we can now prove that N_* is a dual degenerate index set.

Proposition 2.9 *The set N_* is dual degenerate and the smallest face $\mathcal{P}_{N_*}^{+}$ which contains x^* is dual degenerate.*

Proof. The last assertion follows immediately from the first one. Since $\{(y^k, s^k)\}$ is a bounded sequence, it must have an accumulation point (y^*, s^*). Clearly, $A^T y^* + s^* = c$. Using Proposition 2.3(c), we obtain $X^* s^* = 0$, where $X^* \equiv diag(x^*)$. Since $x_{B_*}^* > 0$, this implies that $s_{B_*}^* = 0$. Hence, $A_{B_*}^T y^* = c_{B_*}$ which shows that $c_{B_*} \in Range(A_{B_*}^T)$. By Proposition 2.7, it follows that N_* is a dual degenerate index set. ∎

Before moving to the proof of the global convergence of the affine scaling algorithm under primal nondegeneracy, we give one more basic result on the local linear convergence property of this algorithm. This result was established under primal nondegeneracy assumption by Dikin and Zorkaltsev [10] and more recently by Barnes [3]. A proof of this result that does not require primal nondegeneracy was finally given by Tseng and Luo [19], Theorem 1 and Lemma 2(c). We use this result in Section 4 to show global convergence of the sequence of dual estimates and to obtain a sharper convergence rate for the case in which $\lambda \in (0, 2/3]$. For the sake of completeness, we give a proof of the result below in the Appendix.

Lemma 2.10 *The sequence $\{x^k\}$ generated by the affine scaling algorithm satisfies*

$$
\limsup_{k \to \infty} \frac{c^T x^{k+1} - v^*}{c^T x^k - v^*} \leq 1 - \frac{\lambda}{\sqrt{n}}. \qquad (32)
$$

Now, using Theorem 2.6(b), we give a different global convergence proof for the affine scaling algorithm applied to primal nondegenerate LP problems.

Theorem 2.11 *Assume that (1) is primal-nondegenerate, that is* $(AX^2A^T)^{-1}$ *exists for all* $x \in \mathcal{P}^+$, *where* $X \equiv diag(x)$. *Then, the following statements hold:*

(a) The limit x^* *of the sequence* $\{x^k\}$ *generated by the affine scaling algorithm is an optimal solution of (1) and the sequence of dual estimates* $\{(y(x^k),\ s(x^k))\}$ *converges to the unique optimal solution* $(y^*,\ s^*) \equiv (y(x^*), s(x^*))$ *of (2).*

(b) The pair of optimal solutions x^* *and* $(y^*,\ s^*)$ *satisfies strictly complementarity, or equivalently* x^* *lies in the relative interior of the optimal face of (1).*

Proof. We first prove (a). Under the primal-nondegeneracy assumption, the function $(y(x), s(x))$ defined in (9) is a continuous function of $x \in \mathcal{P}^+$. Since $\{x^k\}$ converges to $x^* \in \mathcal{P}^+$, the sequence $\{(y^k, s^k)\}$ also converges to $(y^*, s^*) \equiv (y(x^*), s(x^*)) \in \mathcal{D}$. By Lemma 2.3(c), we have $X^*s^* = 0$, where $X^* \equiv diag(x^*)$. Hence, it is sufficient to show that $s^* \geq 0$ in order to prove that x^* and (y^*, s^*) are optimal solutions for (1) and (2), respectively. Indeed, since $x^*_{B_*} > 0$ we have $s^*_{B_*} = 0$. Next, we show that $s^*_{N_*} \geq 0$. Assume by contradiction that there exists a $j \in N_*$ such that $s^*_j < 0$. Then we can find an integer $K \geq 0$ such that

$$s^k_j < 0, \qquad \forall k \geq K.$$

Hence, using (12) we obtain

$$x^{k+1}_j = x^k_j - \alpha_k (x^k_j)^2 s^k_j > x^k_j, \qquad \forall k \geq K,$$

where $\alpha_k \equiv \lambda/\chi[(X^k)^{-1}d^k] > 0$ and $X^k \equiv diag(x^k)$. But this contradicts the fact that x^k_j converges to $x^*_j = 0$. We next prove (b). Let $Z \equiv \{j \mid s^*_j > 0\}$. Using (9) and the fact that $s^* = c - A^Ty^*$ and $s^*_j = 0$ for all $j \notin Z$, we obtain

$$\begin{aligned} s_j(x) &= s^*_j - (A_j)^T(AX^2A^T)^{-1}AX^2s^* \\ &= -(A_j)^T(AX^2A^T)^{-1}A_ZX_Z^2s^*_Z, \qquad \forall j \notin Z, \end{aligned} \tag{33}$$

where A_j denotes the j-th column of A. It is easy to see that (33) implies that there exists $C_1 \geq 0$ such that

$$|s^k_j| \leq C_1 \|x^k_Z\|^2, \qquad \forall k \geq 0,\ \forall j \notin Z. \tag{34}$$

We will now show that

$$\sum_{k=0}^{\infty} \alpha_k \|x^k_Z\|^2 < \infty. \tag{35}$$

Indeed, since $s^*_Z > 0$ and $\lim_{k\to\infty} s^k_Z = s^*_Z > 0$, there exists an integer $k_0 \geq 0$ and a scalar $\mu > 0$ such that

$$s^k_Z > \mu e, \qquad \forall k \geq k_0. \tag{36}$$

Hence, if $j \in Z$ we have

$$\begin{aligned} \sum_{k=k_0}^{\infty} \alpha_k (x^k_j)^2 &\leq \mu^{-1} \sum_{k=k_0}^{\infty} \alpha_k (x^k_j)^2 s^k_j \\ &= \mu^{-1} \sum_{k=k_0}^{\infty} (x^k_j - x^{k+1}_j) \\ &= \mu^{-1}(x^{k_0}_j - x^*_j) \ = \ \mu^{-1}x^{k_0}_j \ < \ \infty, \end{aligned} \tag{37}$$

which obviously implies (35). We finally prove that $x_j^* > 0$ for all $j \notin Z$ and hence that statement (b) holds. Indeed, let $C_2 \geq 0$ be a constant such that

$$\|x^k\| \leq C_2, \qquad \forall k \geq 0. \tag{38}$$

Using (34), (35) and (38), we obtain

$$\sum_{k=0}^{\infty} \frac{|x_j^{k+1} - x_j^k|}{x_j^k} = \sum_{k=0}^{\infty} \alpha_k |x_j^k s_j^k| \leq C_1 C_2 \sum_{k=0}^{\infty} \alpha_k \|x_Z^k\|^2 < \infty, \qquad \forall j \notin Z. \tag{39}$$

It is easy to verify that this last relation implies

$$\sum_{k=0}^{\infty} |\log x_j^{k+1} - \log x_j^k| = \sum_{k=0}^{\infty} \left| \log \left(\frac{x_j^{k+1}}{x_j^k} \right) \right| < \infty, \qquad \forall j \notin Z,$$

which in turn implies that $\lim_{k\to\infty} x_j^k = x_j^* > 0$, for all $j \notin Z$. ∎

3 Global convergence of the affine scaling algorithm

This section is organized as follows. In Subsection 3.1, we show that the limit x^* of the primal sequence $\{x^k\}$ is an optimal solution of (1). In this subsection, we state Theorem 3.2 without proof. Subsection 3.2 contains important technical lemmas which are used to derive asymptotic estimates near dual degenerate faces for several quantities associated with the affine scaling algorithm. These estimates are then used in Subsection 3.3 to derive a proof of Theorem 3.2.

3.1 Main Theorems

We have shown in Section 2 that the primal sequence $\{x^k\}$ converges to some $x^* \in \mathcal{P}^+$. In this section, we show that x^* is an optimal solution of problem (1) if the fraction $\lambda \in (0, 2/3]$. An important ingredient used in the proof of these results is a potential function which in general can be defined with respect to any degenerate index set N. This potential function was used in the global convergence of the affine scaling algorithm presented in Tsuchiya [20, 21], Dikin [8] and Tsuchiya and Muramatsu [22].

We start by introducing the forementioned potential function.

Definition 1 *Let N be a dual degenerate index set and let v_N denote the the constant value of $c^T x$ over the face $\mathcal{P}_N^+$. The potential function with respect to N is defined as*

$$\psi_N(x) = |N| \log(c^T x - v_N) - \sum_{j \in N} \log x_j \tag{40}$$

for every $x \in \mathcal{P}^{++}$ such that $c^T x > v_N$.

Note that when $N = N_*$, we have $v_{N_*} = v^*$. Hence, the potential function with respect to the dual degenerate index set N_* can be written as

$$\psi_{N_*}(x) = |N_*| \log(c^T x - v^*) - \sum_{j \in N_*} \log x_j. \tag{41}$$

The fact that the limit point x^* is an optimal solution of problem (1) follows from two results stated below, namely Lemma 3.1 and Theorem 3.2. Lemma 3.1 follows as an immediate consequence of Theorem 2.6(c). The proof of Theorem 3.2 is the main goal of Subsection 3.3. Theorem 3.3 combines these two results to obtain the conclusion that x^* is an optimal solution of (1).

Lemma 3.1 *For the affine scaling algorithm, there exists a constant $\epsilon > 0$ such that*

$$\frac{c^T x^k - v^*}{\sum_{j \in N_*} x_j^k} \geq \epsilon, \qquad \forall k \geq 0. \tag{42}$$

Proof. By Theorem 2.6(c), we know that for some $M \geq 0$,

$$\|x^k - x^*\| \leq M(c^T x^k - v^*), \qquad \forall k \geq 0.$$

Since $x^*_{N_*} = 0$, this implies

$$\|x^k_{N_*}\| \leq M(c^T x^k - v^*), \qquad \forall k \geq 0.$$

Then, we obtain

$$\sum_{j \in N_*} x_j^k \leq \sqrt{|N_*|}\|x^k_{N_*}\| \leq \sqrt{|N_*|} M(c^T x^k - v^*), \qquad \forall k \geq 0.$$

Letting $\epsilon \equiv 1/(\sqrt{|N_*|}M)$, we obtain (42). ∎

The proof of the next result is one of the main goals of this section and is postponed to a later stage of this section after several preliminary results have been proved. Theorem 3.2 has been proved in Tsuchiya and Muramatsu [22]. However, our proof of this result differs in many ways from and is considerably simpler than the one given in [22] which uses complicated matrix decomposition results in [20].

Theorem 3.2 *For the affine scaling algorithm, assume that $\lambda \leq 2/3$. If x^* is not an optimal solution lying in the relative interior of the optimal face of problem (1) then there exist a constant $\zeta > 0$ and an integer K such that*

$$\psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k) < -\zeta, \qquad \forall k \geq K. \tag{43}$$

Combining Lemma 3.1 and Theorem 3.2, we can now show that x^* is an optimal solution of problem (1).

Theorem 3.3 *For the affine scaling algorithm, assume that $\lambda \leq 2/3$. Then the limit x^* of the primal sequence $\{x^k\}$ is an optimal solution lying in the relative interior of the optimal face of problem (1).*

Proof. Assume by contradiction that x^* is not an optimal solution lying in the relative interior of the optimal face of (1). By Theorem 3.2, it follows that

$$\lim_{k\to\infty} \psi_{N_*}(x^k) = -\infty. \tag{44}$$

Noting the fact that

$$|N_*| \log(\sum_{j\in N_*} x_j^k) - \sum_{j\in N_*} \log x_j^k \geq |N_*| \log |N_*|,$$

we obtain

$$\begin{aligned} \psi_{N_*}(x^k) &= |N_*| \log\left(\frac{c^T x^k - v^*}{\sum_{j\in N_*} x_j^k}\right) + |N_*| \log(\sum_{j\in N_*} x_j^k) - \sum_{j\in N_*} \log x_j^k \\ &\geq |N_*| \log\left(\frac{c^T x^k - v^*}{\sum_{j\in N_*} x_j^k}\right) + |N_*| \log |N_*|. \end{aligned} \tag{45}$$

Relations (44) and (45) then imply

$$\lim_{k\to\infty} \frac{c^T x^k - v^*}{\sum_{j\in N_*} x_j^k} = 0,$$

which contradicts Lemma 3.1. ∎

In the next two subsections, we concentrate our efforts in the proof of Theorem 3.2. Before going into the details, we provide below an intuitive justification for using the potential function (40) and for the result of Theorem 3.2. As is obviously seen from its form, the potential function (40) is based on Karmarkar's potential function [14]. As was observed in Bayer and Lagarias [4], Karmarkar's algorithm is equivalent to the affine scaling algorithm applied to a homogeneous linear programming problem. For the homogeneous linear programming problem, we can show that the affine scaling algorithm uniformly reduces the associated Karmarkar's potential function at every iteration.

We apply the above observations to our context as follows. Consider the LP problem

$$\begin{aligned} &\text{minimize}_{\ x} \quad c_{N_*}^T x_{N_*} + c_{B_*}^T (x_{B_*} - x_{B_*}^*) \\ &\text{subject to} \quad A_{N_*} x_{N_*} + A_{B_*}(x_{B_*} - x_{B_*}^*) = 0, \quad x_{N_*} \geq 0. \end{aligned} \tag{46}$$

which arises from (1) by discarding the constraints $x_{B_*} \geq 0$. By introducing a new variable $v_{B_*} \equiv x_{B_*} - x_{B_*}^*$, one can transform this problem to a homogeneous LP problem. Then the potential function (41) is the associated Karmarkar's potential function for this problem. As mentioned above, we can reduce (41) by taking long steps of the affine scaling algorithm applied to this homogeneous problem. On the other hand, we can show that the affine scaling direction for problem (46) at x^k asymptotically converges to the affine scaling direction for problem (1) at x^k, as long as $x_{B_*}^k$ is uniformly bounded away from zero. Hence, by moving along the affine scaling direction for (1), we should also be able to asymptotically reduce the potential function (40).

Although a proof along the above lines could be pursued in this paper, we use a slightly different approach. We note however that the above idea is used in the approach of Tsuchiya and Muramatsu [22].

3.2 Technical Lemmas

In this subsection, we prove some technical results to analyze the asymptotic behavior of the affine scaling direction and the dual estimate in a neighborhood of a dual degenerate face. Lemma 3.6 and Lemma 3.7 play an important role in the approach used in this paper. Although they can be implicitly obtained from the approach of Tsuchiya and Muramatsu [22], they are stated explicitly for the first time here. In addition, their proofs are new.

Lemma 3.4 *Let $N \subseteq \{1, \ldots, n\}$ be a dual degenerate index set and let $(\tilde{y}, \tilde{s}) \in \mathcal{D}_N$ be given. Then, the following statements hold:*

(a) $c^T x - v_N = \tilde{s}_N^T x_N$, $\forall x \in \mathcal{P}$;

(b) $\tilde{s}_N^T d_N(x) = \tilde{s}^T d(x) = c^T d(x) = \|X^{-1} d(x)\|^2 = \|X s(x)\|^2$, $\forall x > 0$ *and* $X \equiv diag(x)$;

(c) $\|X^{-1} d(x)\| \leq \|X_N \tilde{s}_N\|$, $\forall x > 0$ *and* $X \equiv diag(x)$.

Proof. We first prove (a). We know that $v_N = c^T \tilde{x}$ for any $\tilde{x} \in \mathcal{P}_N^+$. Using the fact that $A(x - \tilde{x}) = 0$, we obtain

$$\begin{aligned} c^T x - v_N &= c^T (x - \tilde{x}) \\ &= (A^T \tilde{y} + \tilde{s})^T (x - \tilde{x}) \\ &= \tilde{s}^T (x - \tilde{x}) \\ &= \tilde{s}_N^T x_N, \end{aligned} \tag{47}$$

where the last equality follows from the fact that $\tilde{x}_N = 0$ and $\tilde{s}_B = 0$. Hence, (a) is proved. Statement (b) is an immediate consequence of Lemma 2.2 and the fact that $\tilde{s}_B = 0$. Using Lemma 2.2 and the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned} \|X^{-1} d(x)\|^2 &= \tilde{s}^T d(x) \\ &= (X\tilde{s})^T (X^{-1} d(x)) \\ &\leq \|X\tilde{s}\| \|X^{-1} d(x)\| \\ &= \|X_N \tilde{s}_N\| \|X^{-1} d(x)\|, \end{aligned}$$

from which (c) follows. ∎

Lemma 3.5 *Let $N \subseteq \{1, \ldots, n\}$ be a dual degenerate index set and let $(\tilde{y}, \tilde{s}) \in \mathcal{D}_N$ be given. Then, the affine scaling direction $d(x)$ is the unique solution of the following QP problem:*

$$\begin{array}{ll} \text{maximize}_{\ p} & \tilde{s}_N^T p_N - \frac{1}{2}\|X^{-1} p\|^2 \\ \text{subject to} & Ap = 0, \end{array} \tag{48}$$

where $X \equiv diag(x)$.

Proof. The proof of this lemma follows as an immediate consequence of Proposition 2.1 and Lemma 3.4(a). ∎

The estimates obtained in the following two results play a crucial role in simplifying the proof of Theorem 3.2.

Lemma 3.6 *Let N be a dual degenerate index set. Then, there exists a constant $C \geq 0$ such that*

$$\|X_B^{-1} d_B(x)\| \leq C\|X_B^{-1}\|\|x_N\|\|X_N^{-1} d_N(x)\|, \qquad \forall x > 0.$$

Proof. By Lemma 3.5 we know that $d(x)$ solves problem (48) where $\tilde{s}_N$ is the $N-$component of a vector $(\tilde{y}, \tilde{s}) \in \mathcal{D}_N$. It follows that $d_B(x)$ solves the problem

$$\begin{array}{ll} \text{minimize }_{p_B} & \frac{1}{2}\|X_B^{-1} p_B\|^2 \\ \text{subject to} & A_B p_B = -A_N d_N(x). \end{array} \tag{49}$$

By Lemma 2.4, we know that the system $A_B p_B = -A_N d_N(x)$ has a solution $\bar{p}_B \in \mathbb{R}^{|B|}$ such that

$$\|\bar{p}_B\| \leq C\|d_N(x)\|$$

where C is a constant independent of x. Hence, we have

$$\begin{aligned} \|X_B^{-1} d_B(x)\| &\leq \|X_B^{-1} \bar{p}_B\| \\ &\leq \|X_B^{-1}\|\|\bar{p}_B\| \\ &\leq \|X_B^{-1}\|C\|d_N(x)\| \\ &\leq C\|X_B^{-1}\|\|X_N\|\|X_N^{-1} d_N(x)\| \\ &\leq C\|X_B^{-1}\|\|x_N\|\|X_N^{-1} d_N(x)\|. \end{aligned}$$

■

Lemma 3.7 *Let N be a dual degenerate index set. Then, there exists a constant $\bar{C} \geq 0$ such that for all $x > 0$,*

$$\min\{\|s(x) - s\| \mid (y, s) \in \mathcal{D}_N\} \leq \bar{C}\|X_B^{-1}\|^2\|x_N\|^2, \tag{50}$$

where $X_B \equiv diag(x_B)$.

Proof. Recall that any point $(y, s) \in \mathcal{D}_N$ is a solution of the linear system

$$\begin{aligned} A_B^T y &= c_B, \\ A_N^T y + s_N &= c_N, \\ s_B &= 0. \end{aligned}$$

By applying Lemma 2.4 with $z = (y(x), s(x))$, we know that there exists a constant $C_1 \geq 0$ with the following property: for every $x > 0$, there exists $(\hat{y}(x), \hat{s}(x)) \in \mathcal{D}_N$ such that

$$\|s(x) - \hat{s}(x)\| \leq C_1\|s_B(x)\|. \tag{51}$$

Squaring both sides of this expression and using the fact that $\hat{s}_B = 0$, we obtain

$$\|s_B(x)\|^2 + \|s_N(x) - \hat{s}_N(x)\|^2 \leq C_1^2\|s_B(x)\|^2,$$

or equivalently,

$$\|s_N(x) - \hat{s}_N(x)\|^2 \leq (C_1^2 - 1)\|s_B(x)\|^2. \tag{52}$$

By Proposition 2.8, we know that $\{s(x) \mid x > 0\}$ is bounded. Using (52), it also follows that $\{\hat{s}_N(x) \mid x > 0\}$ is bounded. Then, let $M \geq 0$ be such that for all $x > 0$,

$$\|s(x)\| \leq M, \qquad \|\hat{s}_N(x)\| \leq M. \tag{53}$$

On the other hand, using the fact that $(y(x), s(x))$ is the optimal solution of (11) and that $(\hat{y}(x), \hat{s}(x))$ is feasible to (11), we obtain

$$\|X_B s_B(x)\|^2 + \|X_N s_N(x)\|^2 \leq \|X_N \hat{s}_N(x)\|^2,$$

where $X_N = diag(x_N)$, and this implies

$$\begin{aligned}
\|s_B(x)\|^2 &\leq \|X_B^{-1}\|^2 \|X_B s_B(x)\|^2 \\
&\leq \|X_B^{-1}\|^2 \left\{ \|X_N \hat{s}_N(x)\|^2 - \|X_N s_N(x)\|^2 \right\} \\
&= \|X_B^{-1}\|^2 \, [X_N \hat{s}_N(x) + X_N s_N(x)]^T [X_N \hat{s}_N(x) - X_N s_N(x)] \\
&\leq \|X_B^{-1}\|^2 \, \|X_N \hat{s}_N(x) + X_N s_N(x)\| \, \|X_N \hat{s}_N(x) - X_N s_N(x)\| \\
&\leq \|X_B^{-1}\|^2 \|X_N\|^2 \|\hat{s}_N(x) + s_N(x)\| \, \|\hat{s}_N(x) - s_N(x)\| \\
&\leq 2M(C_1^2 - 1)^{1/2} \|X_B^{-1}\|^2 \|x_N\|^2 \|s_B(x)\|,
\end{aligned} \tag{54}$$

where the last inequality follows from (53) and (52). From (54), it immediately follows that

$$\|s_B(x)\| \leq 2M(C_1^2 - 1)^{1/2} \|X_B^{-1}\|^2 \|x_N\|^2.$$

In view of (51), this last relation gives

$$\|s(x) - \hat{s}(x)\| \leq 2MC_1(C_1^2 - 1)^{1/2} \|X_B^{-1}\|^2 \|x_N\|^2$$

which implies (50) if we let $\bar{C} \equiv 2MC_1(C_1^2 - 1)^{1/2}$. ∎

3.3 Proof of Theorem 3.2

The main goal of this subsection is to provide a proof of Theorem 3.2.

The following notation is used frequently in order to make the ideas clearer and the proofs more concise. Let $\{\tau^k\}$ and $\{\beta^k\}$ be two sequences of real numbers. We write $\tau^k = \mathcal{O}(\beta^k)$ to indicate that there exist an integer $k_0 \geq 0$ and a scalar $r \geq 0$ such that $|\tau_k| \leq r\beta_k$ for all $k \geq k_0$. Clearly, if $\beta_k > 0$ for all $k \geq 0$ then $|\tau_k| \leq r\beta_k$ holds for all $k \geq 0$ by taking a larger scalar r if necessary.

We next examine the change in the potential function (41) at two consecutive iterates generated by the affine scaling algorithm. Define the normalized scaled search direction as follows:

$$u^k \equiv \frac{(X^k)^{-1} d^k}{c^T x^k - v^*} = \frac{X^k s^k}{c^T x^k - v^*}, \qquad \forall k \geq 0, \tag{55}$$

where $X^k = diag(X^k)$. Using Lemma 2.2, we obtain that the ratio between $c^T x^{k+1} - v^*$ and $c^T x^k - v^*$ in terms of u^k is given as follows:

$$\begin{aligned} 0 &< \frac{c^T x^{k+1} - v^*}{c^T x^k - v^*} \\ &= 1 - \frac{\lambda}{\chi[(X^k)^{-1} d^k]} \frac{c^T d^k}{c^T x^k - v^*} \\ &= 1 - \frac{\lambda}{\chi[(X^k)^{-1} d^k]} \frac{\|(X^k)^{-1} d^k\|^2}{c^T x^k - v^*} \\ &= 1 - \frac{\lambda}{\chi[u^k]} \|u^k\|^2. \end{aligned} \tag{56}$$

In a similar manner, we have

$$0 < 1 - \lambda \le 1 - \frac{\lambda}{\chi[(X^k)^{-1} d^k]} [(X^k)^{-1} d^k]_i = \frac{x_i^{k+1}}{x_i^k} = 1 - \frac{\lambda}{\chi[u^k]} u_i^k. \tag{57}$$

Using these relations, we obtain

$$\begin{aligned} \psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k) &= |N_*| \log \left(\frac{c^T x^{k+1} - v^*}{c^T x^k - v^*} \right) - \sum_{i \in N_*} \log \left(\frac{x_i^{k+1}}{x_i^k} \right) \\ &= |N_*| \log \left(1 - \frac{\lambda}{\chi[u^k]} \|u^k\|^2 \right) - \sum_{j \in N_*} \log \left(1 - \frac{\lambda}{\chi[u^k]} u_j^k \right). \end{aligned} \tag{58}$$

The direction (55) plays an important role in the subsequent analysis.

Lemma 3.8 *The normalized scaled search direction u^k has the following properties:*

(a) $\|u_{B_*}^k\| / \|u_{N_*}^k\| = \mathcal{O}(\|x_{N_*}^k\|)$.

(b) *The sequence $\{u^k\}$ is bounded and $\|u_{B_*}^k\| = \mathcal{O}(\|x_{N_*}^k\|)$.*

(c) $|e^T u_{N_*}^k - 1| = \mathcal{O}(\|x_{N_*}^k\|^2)$.

(d) $\chi[u^k] = \chi[u_{N_*}^k]$ *for all k sufficiently large.*

Proof. Using Lemma 3.6 with $N = N_*$, we obtain

$$\begin{aligned} \frac{\|u_{B_*}^k\|}{\|u_{N_*}^k\|} &= \frac{\|(X_{B_*}^k)^{-1} d_{B_*}(x^k)\|}{\|(X_{N_*}^k)^{-1} d_{N_*}(x^k)\|} \\ &\le C \|(X_{B_*}^k)^{-1}\| \, \|x_{N_*}^k\|, \end{aligned}$$

where $X_{B_*}^k \equiv diag(x_{B_*}^k)$, which immediately implies (a). We now show (b). First observe that, by Theorem 2.6(c), there exists $M > 0$ such that

$$\|x_{N_*}^k\| \le M(c^T x^k - v^*), \qquad \forall k \ge 0. \tag{59}$$

Let $(\bar{y}, \bar{s}) \in \mathcal{D}_{N_*}$ be given. Using (59) and Lemma 3.4(c) with $N = N_*$, we obtain

$$\|u^k\| = \frac{\|(X^k)^{-1}d^k\|}{c^T x^k - v^*} \leq \frac{\|X^k_{N_*}\bar{s}_{N_*}\|}{M^{-1}\|x^k_{N_*}\|} \leq M\|\bar{s}_{N_*}\|, \qquad \forall k \geq 0.$$

Hence, $\{u^k\}$ is bounded and in view of (a), this obviously implies that $\|u^k_{B_*}\| = \mathcal{O}(\|x^k_{N_*}\|)$. We now show (c). Using Lemma 3.7 with $N = N_*$, we find that for all $k \geq 0$, there exists $(\hat{y}^k, \hat{s}^k) \in \mathcal{D}_{N_*}$ such that

$$\|\hat{s}^k_{N_*} - s^k_{N_*}\| = \mathcal{O}(\|x^k_{N_*}\|^2). \tag{60}$$

Since $(\hat{y}^k, \hat{s}^k) \in \mathcal{D}_{N_*}$, we know from Lemma 3.4(a) with $N = N_*$ that

$$c^T x^k - v^* = (\hat{s}^k_{N_*})^T x^k_{N_*}, \qquad \forall k \geq 0. \tag{61}$$

Hence, it follows from (55), (59), (60) and (61) that

$$\begin{aligned}
|e^T u^k_{N_*} - 1| &= \left| \frac{(s^k_{N_*})^T x^k_{N_*} - (\hat{s}^k_{N_*})^T x^k_{N_*}}{c^T x^k - v^*} \right| \\
&\leq \frac{\|s^k_{N_*} - \hat{s}^k_{N_*}\| \, \|x^k_{N_*}\|}{M^{-1}\|x^k_{N_*}\|} \\
&= \mathcal{O}(\|x^k_{N_*}\|^2).
\end{aligned}$$

Statement (d) follows easily from (a), (b) and (c). ∎

The following lemma was proved by Tsuchiya and Muramatsu [22], Lemma 5.1 (Lemma 3.1 in the revised version). It is an improvement of a result obtained in Dikin [9], and plays a substantial role in the proof of Theorem 3.2. For the sake of completeness, its proof is given in the Appendix.

Lemma 3.9 *Let $w \in \mathbb{R}^q$ and $\rho \in \mathbb{R}$. Define*

$$G_q(w, \rho) \equiv q \log(1 - \rho\|w\|^2) - \sum_{i=1}^{q} \log(1 - \rho w_i) \tag{62}$$

$$H_q(w, \rho) \equiv \frac{q\rho}{q - \rho} \left\| w - \frac{1}{q} e \right\|^2 \left(-q + \frac{\rho}{2\{1 - \rho\chi[w]\}} \right) \tag{63}$$

and consider the set

$$\Omega = \{(w, \rho) \in \mathbb{R}^q \times \mathbb{R} \mid e^T w = 1, \rho\|w\|^2 < 1 \text{ and } \rho w_i < 1, \forall i = 1, \ldots, q\}.$$

Then, the following inequality holds:

$$G_q(w, \rho) \leq H_q(w, \rho), \qquad \forall (w, \rho) \in \Omega.$$

To make use of Lemma 3.9 to find an upper bound for (58), we introduce

$$w^k_{N_*} \equiv \frac{u^k_{N_*}}{e^T u^k_{N_*}}, \qquad \rho^k \equiv \frac{\lambda}{\chi[w^k_{N_*}]}. \tag{64}$$

Note that $w^k_{N_*}$ is obtained from $u^k_{N_*}$ by enforcing the condition $e^T w^k_{N_*} = 1$ by a conical projection. Note also that $w^k_{N_*}$ is well-defined for all sufficiently large k due to Lemma 3.8(c).

Lemma 3.10 *Consider the sequences $\{w^k_{N_*}\}$ and $\{\rho^k\}$ as above and let $q \equiv |N_*|$. For sufficiently large k, we have*

$$(w^k_{N_*}, \rho^k) \in \Omega, \tag{65}$$

$$\psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k) \leq H_q(w^k_{N_*}, \rho^k) + \mathcal{O}(\|x^k_{N_*}\|^2), \tag{66}$$

where Ω and $H_q(\cdot,\cdot)$ are defined in Lemma 3.9.

To prove Lemma 3.10, we need two technical lemmas.

Lemma 3.11 *Consider the sequences $\{u^k\}$, $\{w^k_{N_*}\}$, and $\{\rho^k\}$ which are defined in relations (55) and (64). Then, the following statements hold:*

(a) $\|u^k_{N_*} - w^k_{N_*}\| = \mathcal{O}(\|x^k_{N_*}\|^2)$;

(b) $\liminf_{k\to\infty} \chi[u^k] = \liminf_{k\to\infty} \chi[u^k_{N_*}] = \liminf_{k\to\infty} \chi[w^k_{N_*}] \geq 1/|N_*|$;

(c) $\limsup_{k\to\infty} \chi[u^k] = \limsup_{k\to\infty} \chi[u^k_{N_*}] = \limsup_{k\to\infty} \chi[w^k_{N_*}] \leq 1/\lambda$;

(d) $\liminf_{k\to\infty} \rho^k \geq \lambda^2$;

(e) $\limsup_{k\to\infty} \rho^k \leq |N_*|\lambda$;

Proof. Using statements (b) and (c) of Lemma 3.8, we obtain

$$\|u^k_{N_*} - w^k_{N_*}\| = \left\| u^k_{N_*} - \frac{u^k_{N_*}}{e^T u^k_{N_*}} \right\| = \|u^k_{N_*}\| \frac{|e^T u^k_{N_*} - 1|}{e^T u^k_{N_*}} = \mathcal{O}(\|x^k_{N_*}\|^2)$$

and hence (a) follows. We now show (b) and (c) together. The equalities in (b) and (c) follow from (a) and Lemma 3.8(d). In view of (64), we have $e^T w^k_{N_*} = 1$ and this obviously implies that $\chi[w^k_{N_*}] \geq 1/|N_*|$, for all $k \geq 0$. Hence, $\liminf_{k\to\infty} \chi[w^k_{N_*}] \geq 1/|N_*|$ and (b) follows. To complete the proof of (c), we will show that $\chi[u^k] \leq 1/\lambda$ for all $k \geq 0$. Indeed, since $c^T x^k - v^* > 0$ for all $k \geq 0$, we obtain from relation (56) that,

$$0 < 1 - \lambda \frac{\|u^k\|^2}{\chi[u^k]} \leq 1 - \lambda \|u^k\|, \qquad \forall k \geq 0.$$

Hence,

$$\frac{1}{\lambda} \geq \|u^k\| \geq \chi[u^k], \qquad \forall k \geq 0,$$

as desired. Statements (d) and (e) now follow immediately from (64) and statements (b) and (c). ∎

Lemma 3.12 *For sufficiently large k, we have*

$$1-\lambda \le 1-\frac{\lambda}{\chi[u^k]}u_i^k = 1-\rho^k w_i^k, \qquad \forall i \in N_*. \tag{67}$$

Furthermore, there exists a constant $\delta > 0$ such that

$$\delta \le 1-\frac{\lambda}{\chi[u^k]}\|u^k\|^2 = 1-\rho^k\|w_{N_*}^k\|^2 + \mathcal{O}(\|x_{N_*}^k\|^2). \tag{68}$$

Proof. The inequality in (67) is obvious. Using relation (64) and Lemma 3.8(d), we can easily verify the equality in (67). We now show the equality in (68). Using (c) and (d) of Lemma 3.8 and relation (64), we can easily show that

$$\left\|\frac{\lambda}{\chi[u^k]}-\rho^k\right\| = \mathcal{O}(\|x_{N_*}^k\|^2). \tag{69}$$

Hence,

$$\begin{aligned}
&\left\|\frac{\lambda}{\chi[u^k]}\|u^k\|^2-\rho^k\|w_{N_*}^k\|^2\right\| \\
&= \left\|\left(\frac{\lambda}{\chi[u^k]}-\rho^k\right)\|u^k\|^2+\rho^k\left(\|u^k\|^2-\|w_{N_*}^k\|^2\right)\right\| \\
&\le \left|\frac{\lambda}{\chi[u^k]}-\rho^k\right|\|u^k\|^2+\rho^k\left(\|u_{N_*}^k-w_{N_*}^k\|\,\|u_{N_*}^k+w_{N_*}^k\|+\|u_{B_*}^k\|^2\right) \\
&= \mathcal{O}(\|x_{N_*}^k\|^2),
\end{aligned}$$

where the last equality follows from (a) and (b) of Lemma 3.8, (a) and (e) of Lemma 3.11 and relation (69). This last relation shows the equality in (68). Let $(\bar{y},\bar{s}) \in \mathcal{D}_{N_*}$ given. We will show that the inequality in (68) holds for $\delta \equiv (1-\lambda)/(M\|\bar{s}_{N_*}\|)$, where M is the constant introduced in Theorem 2.6(c). Indeed, using (56), Lemma 3.4(a), Theorem 2.6(c) and (57), we obtain

$$1-\frac{\lambda}{\chi[u^k]}\|u^k\|^2 = \frac{c^Tx^{k+1}-v^*}{c^Tx^k-v^*} \ge \frac{M^{-1}\|x_{N_*}^{k+1}\|}{\bar{s}_{N_*}^T x_{N_*}^k} \ge \frac{(1-\lambda)\|x_{N_*}^k\|}{M\|\bar{s}_{N_*}\|\|x_{N_*}^k\|} = \frac{1-\lambda}{M\|\bar{s}_{N_*}\|} = \delta. \tag{70}$$

■

Now we are ready to prove Lemma 3.10.

Proof of Lemma 3.10 From (67) and (68), it is easy to see that $(w_{N_*}^k,\rho^k) \in \Omega$, for all k sufficiently large, so that $H_q(w_{N_*}^k,\rho^k)$ and $G_q(w_{N_*}^k,\rho^k)$ are well-defined. From relations (58), (63), (67) and Lemma 3.9, we have for all k sufficiently large

$$\begin{aligned}
\psi_{N_*}(x^{k+1})-\psi_{N_*}(x^k) &= |N_*|\log\left(1-\frac{\lambda}{\chi[u^k]}\|u^k\|^2\right)-\sum_{j\in N_*}\log\left(1-\frac{\lambda}{\chi[u^k]}u_j^k\right) \\
&= G_q(w_{N_*}^k,\rho^k)+r^k \\
&\le H_q(w_{N_*}^k,\rho^k)+r^k,
\end{aligned}$$

where

$$r^k = |N_*| \log \left(\frac{1 - (\lambda/\chi[u^k])\|u^k\|^2}{1 - \rho^k \|w^k_{N_*}\|^2} \right)$$

From relation (68) of Lemma 3.12, it is easy to see $r^k = \mathcal{O}(\|x_N\|^2)$. Hence the lemma follows. ∎

The following result gives an upper bound for $\psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k)$ in terms of the distance between the vectors $w^k_{N_*}$ and $(1/|N_*|)e$.

Theorem 3.13 *For the affine scaling algorithm, assume that $\lambda \le 2/3$. For all $k \ge 0$, define $\eta_k \equiv \|w^k_{N_*} - (1/|N_*|)e\|$. Then, for all k sufficiently large, we have*

$$\psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k) \le -\frac{|N_*|\lambda^2 \eta_k^3}{1 + \eta_k} + \mathcal{O}(\|x^k_{N_*}\|^2). \tag{71}$$

Proof. Let $q \equiv |N_*|$. Using Lemma 3.11(d), we obtain

$$\liminf_{k \to \infty} \frac{q\rho^k}{q - \rho^k} \ge \frac{q\lambda^2}{q - \lambda^2} > \lambda^2. \tag{72}$$

Using the fact that $e^T w^k_{N_*} = 1$ and $\eta_k = \|w^k_{N_*} - (1/q)e\|$, we can easily show that

$$\chi[w^k_{N_*}] \ge \frac{1}{q} + \frac{\eta_k}{q} = \frac{1 + \eta_k}{q}, \qquad \forall k \ge 0. \tag{73}$$

Using relations (64) and (73) and the fact that $\lambda \le 2/3$, we obtain

$$\begin{aligned} -q + \frac{\rho^k}{2\{1 - \rho^k \chi[w^k_{N_*}]\}} &= -q + \frac{\lambda}{2(1-\lambda)} \frac{1}{\chi[w^k_{N_*}]} \\ &\le -q + \frac{q}{1 + \eta_k} \\ &= -\frac{q\eta_k}{1 + \eta_k}. \end{aligned} \tag{74}$$

Using Lemma 3.10 and relations (63), (72) and (74), we obtain

$$\begin{aligned} \psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k) &\le H_q(w^k_{N_*}, \rho^k) + \mathcal{O}(\|x^k_{N_*}\|^2) \\ &= \frac{q\rho^k}{q - \rho^k} \left\| w^k_{N_*} - \frac{1}{q} e \right\|^2 \left(-q + \frac{\rho^k}{2[1 - \rho^k \chi[w^k_{N_*}]]} \right) + \mathcal{O}(\|x^k_{N_*}\|^2) \\ &\le \lambda^2 \eta_k^2 \left(-\frac{q\eta_k}{1 + \eta_k} \right) + \mathcal{O}(\|x^k_{N_*}\|^2) \\ &= -\frac{q\lambda^2 \eta_k^3}{1 + \eta_k} + \mathcal{O}(\|x^k_{N_*}\|^2). \end{aligned}$$

for all k sufficiently large. ∎

We are now in a position to give the proof of Theorem 3.2.

Proof of Theorem 3.2 Assume that x^* is not a point lying in the relative interior of the optimal face of (1). Assume by contradiction that there exists a subsequence $\{x^k\}_{k\in K}$ for which $\lim_{k\in K} \psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k) \geq 0$. Due to relation (71) of Theorem 3.13, we have

$$\lim_{k\in K} \left\| w^k_{N_*} - \frac{1}{|N_*|} e \right\| = 0.$$

By Lemma 3.11(a), this implies $\lim_{k\in K} \|u^k_{N_*} - (1/|N_*|)e\| = 0$. By taking a subset of K if necessary, we may assume that

$$\frac{1}{2|N_*|} e \leq u^k_{N_*} = \frac{X^k_{N_*} s^k_{N_*}}{c^T x^k - v^*}, \qquad \forall k \in K.$$

Clearly, this relation implies that $s^k_{N_*} > 0$ for all $k \in K$. Hence, from Theorem 2.6(c) it follows that

$$\frac{e}{2|N_*|} \leq \frac{X^k_{N_*} s^k_{N_*}}{c^T x^k - v^*} \leq \frac{\|X^k_{N_*}\| s^k_{N_*}}{c^T x^k - v^*} \leq M s^k_{N_*}, \qquad \forall k \in K. \tag{75}$$

By Proposition 2.8, the sequence $\{(y^k, s^k)\}_{k\in K}$ is bounded, and hence it has an accumulation point (y^*, s^*) such that $s^*_{N_*} > 0$ due to (75). In view of Proposition 2.3(c) and the definition of B_*, we have $X^* s^* = 0$ and $x^*_{B_*} > 0$. Hence, x^* and (y^*, s^*) satisfies the strictly complementarity condition. This implies that x^* is a point lying in the relative interior of the primal optimal face contradicting our assumption. ∎

4 Convergence of the dual estimates

In this section we show that the sequence of dual estimates converge to the analytical center of the optimal dual face (Theorem 4.3). We also give a sharp rate of convergence result for the sequence $\{c^T x^k - v^*\}$ (Theorem 4.2).

The proof of the next lemma have almost been worked out in section 3. We only need to put the pieces together.

Lemma 4.1 *For the affine scaling algorithm, assume that $\lambda \leq 2/3$. Then,*

$$\lim_{k\to\infty} w^k_{N_*} = \frac{1}{|N_*|} e, \qquad \textit{and} \qquad \lim_{k\to\infty} u^k_{N_*} = \frac{1}{|N_*|} e. \tag{76}$$

Proof. In view of Lemma 3.11(a), it is sufficient to show that the first limit in relation (76) holds. Assume by contradiction that the first limit in (76) does not hold. Then there exists a constant $\delta > 0$ such that the set $\mathcal{B} = \mathcal{B}(\delta) \equiv \{k \,|\, \|w^k_{N_*} - e/|N_*|\| \geq \delta > 0\}$ is infinite. By Theorem 3.13, we have

$$\limsup_{k\in\mathcal{B}} (\psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k)) < 0. \tag{77}$$

As a consequence, we have

$$\sum_{k\in\mathcal{B}} \psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k) = -\infty. \tag{78}$$

Moreover, Theorem 3.13 also implies that

$$\psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k) \leq \mathcal{O}(\|x^k_{N_*}\|^2),$$

and hence there exists a constant $C \geq 0$ such that

$$\psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k) \leq C\|x^k_{N_*}\|^2, \qquad \forall k \geq 0.$$

Using this relation and Theorem 2.6(c) and Lemma 2.10, we obtain

$$\begin{aligned}
\sum_{k\notin\mathcal{B}} \psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k) &\leq C\sum_{k\notin\mathcal{B}} \|x^k_{N_*}\|^2 \\
&\leq C\sum_{k=1}^{\infty} \|x^k_{N_*}\|^2 \\
&\leq M^2 C\sum_{k=1}^{\infty} (c^T x^k - v^*)^2 < \infty.
\end{aligned} \tag{79}$$

Combining relations (78) and (79), we obtain

$$\sum_{k=0}^{\infty} \psi_{N_*}(x^{k+1}) - \psi_{N_*}(x^k) = -\infty,$$

and this implies that $\lim_{k\to\infty} \psi_{N_*}(x^k) = -\infty$. Using a similar argument as that used in the proof of Theorem 3.3, we can easily obtain a contradiction. Hence, (76) holds. ∎

We can now show that a sharper rate of convergence holds for the sequence $\{c^T x^k - v^*\}$ when $\lambda \leq 2/3$.

Theorem 4.2 *The following relation holds:*

$$\lim_{k\to\infty} \frac{c^T x^{k+1} - v^*}{c^T x^k - v^*} = 1 - \lambda. \tag{80}$$

Proof. Using relation (56), Lemma 3.8(b) and Lemma 4.1, we obtain

$$\lim_{k\to\infty} \frac{c^T x^{k+1} - v^*}{c^T x^k - v^*} = \lim_{k\to\infty} 1 - \lambda\frac{\|u^k\|^2}{\chi[u^k]} = 1 - \lambda.$$

∎

The next result shows that the sequence of dual estimates $\{(y^k, s^k)\}$ converges to the center of the dual optimal face. The center of the dual optimal face is the point $(y^*, s^*) \in \mathcal{D}_{N_*}$ such that

$$\begin{aligned}
(y^*, s^*_{N_*}) \;=\; &\operatorname{argmax} && \textstyle\sum_{j\in N_*} \log s_j \\
&\text{subject to} && A^T_{B_*} y = c_{B_*} \\
& && A^T_{N_*} y + s_{N_*} = c_{N_*}, \qquad (s_{N_*} > 0).
\end{aligned} \tag{81}$$

It can be easily verified that $(y^*, s^*_{N_*})$ is characterized as the unique point satisfying the following conditions.

$$s^*_{N_*} > 0; \tag{82a}$$

$$A_{N_*}(s^*_{N_*})^{-1} \in Range(A_{B_*}); \tag{82b}$$

$$[0, s^*_{N_*}]^T \in c + Range(A^T), \tag{82c}$$

where $(s^*_{N_*})^{-1} = (S^*_{N_*})^{-1}e$ and $S^*_{N_*} \equiv diag(s^*_{N_*})$.

We are now ready to state the main result of this section.

Theorem 4.3 *For the affine scaling algorithm, assume that $\lambda \le 2/3$. Then,*

$$\lim_{k\to\infty}(y^k, s^k) = (y^*, s^*). \tag{83}$$

Proof. By Lemma 4.1, we know that

$$\lim_{k\to\infty} u^k_{N_*} = \lim_{k\to\infty} \frac{X^k_{N_*}s^k_{N_*}}{c^Tx^k - v^*} = \frac{1}{|N_*|}e, \tag{84}$$

where $X^k \equiv diag(x^k)$. We next show that

$$\lim_{k\to\infty} \frac{|N_*|}{c^Tx^k - v^*}x^k_{N_*} = (s^*_{N_*})^{-1}. \tag{85}$$

Note first that by Lemma 3.1, the sequence $\{x^k_{N_*}/(c^Tx^k - v^*)\}$ is bounded. Let $a_{N_*} \in \mathbb{R}^{|N_*|}$ be an accumulation point of this sequence, that is

$$\lim_{k\in K} \frac{|N_*|}{c^Tx^k - v^*}x^k_{N_*} = a_{N_*}, \tag{86}$$

where $K \subseteq \{0, 1, 2, \ldots\}$ is an infinite set. Using (84) and the fact that $\{s^k\}$ is bounded, it is easy to see that $a_{N_*} > 0$. We will show that $(a_{N_*})^{-1}$ satisfies the conditions (82), where $(a_{N_*})^{-1}$ denotes the vector whose components are the inverses of the corresponding components of a_{N_*}. Clearly, $(a_{N_*})^{-1} > 0$. Also,

$$A_{N_*}x^k_{N_*} = b - A_{B_*}x^k_{B_*} = A_{B_*}x^*_{B_*} - A_{B_*}x^k_{B_*} \in Range(A_{B_*}), \qquad \forall k \ge 0,$$

which implies that $A_{N_*}a_{N_*} \in Range(A_{B_*})$ due to relation (86) and the fact that $Range(A_{B_*})$ is a closed set. Using (84) and (86), it follows that

$$\lim_{k\in K} s^k_{N_*} = (a_{N_*})^{-1}.$$

Since

$$[s^k_{B_*}, s^k_{N_*}]^T \in c + Range(A^T), \qquad \forall k \ge 0,$$

we have

$$[0, a^{-1}_{N_*}]^T = \lim_{k\in K}[s^k_{B_*}, s^k_{N_*}]^T \in c + Range(A^T).$$

Hence, $(a_{N_*})^{-1}$ satisfies the conditions (82), and therefore $a_{N_*} = (s^*_{N_*})^{-1}$. Since a_{N_*} is an arbitrary accumulation point of the sequence $\{x^k_{N_*}/(c^Tx^k - v^*)\}$, (85) follows. Clearly, (84) and (85) imply that $\lim_{k\to\infty} s^k = s^*$. Since $rank(A) = m$, it also follows that $\lim_{k\to\infty} y^k = y^*$. Hence, (83) follows. ∎

Appendix

In this appendix we give the proofs of Proposition 2.8, Lemma 2.10 and Lemma 3.9.

We start with the proof of Proposition 2.8. First, we need the following lemma whose proof uses Hoffman's Lemma.

Lemma A.1 Let $H \in \mathbb{R}^{l \times p}$ and $h \in \mathbb{R}^l$ be given such that $h \in Range(H)$. Then, there exists a constant $M > 0$ such that for every diagonal matrix $D > 0$, the (unique) optimal solution $\bar{w} = \bar{w}(D) \in \mathbb{R}^p$ of the problem

$$\begin{array}{ll} \text{minimize}_{\ w} & \|Dw\| \\ \text{subject to} & Hw = h \end{array} \tag{87}$$

satisfies $\|\bar{w}\| \leq M$.

Proof. Assume by contradiction that there exists a sequence of diagonal matrices $D^k > 0$ such that solution $w^k = w(D^k)$ of (87) satisfies

$$\lim_{k \to \infty} \|w^k\| = \infty.$$

This implies that there exist a constant $L > 0$, an index set $J \subseteq \{1, 2, \ldots, n\}$, $J \neq \emptyset$, and a subsequence $\{w^k\}_{k \in K}$ with the property that

$$|w_j^k| \leq L, \qquad \forall j \notin J, \ \forall k \in K; \tag{88}$$

$$\lim_{k \in K} |w_j^k| = \infty, \qquad \forall j \in J. \tag{89}$$

Consider the system

$$Hw = h \tag{90a}$$

$$w_j = w_j^k, \qquad \forall j \notin J, \tag{90b}$$

and observe that w^k is a solution of this system. By Lemma 2.4, (90) has a solution $\hat{w}^k$ such that

$$\|\hat{w}^k\| \leq L_1(\|h\|_1 + \sum_{j \notin J} |w_j^k|) \tag{91}$$

where L_1 is a constant independent of k. It follows from (88) and (91) that

$$\|\hat{w}^k\| \leq L_1(\|h\|_1 + (n - |J|)L) \stackrel{def}{=} L_2, \qquad \forall k \in K, \tag{92}$$

which shows that $\hat{w}^k$ is bounded. In view of (89), there exists an integer $k_0 \geq 0$ such that

$$|w_j^k| > L_2, \qquad \forall k \geq k_0, \ k \in K, \ \forall j \in J. \tag{93}$$

Then, it follows from (90b), (92) and (93) that

$$\begin{aligned} \|D\hat{w}^k\|^2 &= \sum_{j \in J} (D_{jj}^k \hat{w}_j^k)^2 + \sum_{j \notin J} (D_{jj}^k \hat{w}_j^k)^2 \\ &\leq \sum_{j \in J} (D_{jj}^k L_2)^2 + \sum_{j \notin J} (D_{jj}^k w_j^k)^2 \\ &< \sum_{j \in J} (D_{jj}^k w_j^k)^2 + \sum_{j \notin J} (D_{jj}^k w_j^k)^2 \\ &= \|D^k w^k\|^2, \qquad \forall k \geq k_0, \ k \in K, \end{aligned}$$

which together with (90a) contradicts the fact that w^k is an optimal solution of (87) with $D = D^k$. ∎

We are now in a position to prove Proposition 2.8.

Proof of Proposition 2.8 By Proposition 2.1, we know that $(y(x), s(x))$ is the unique solution of (11). Let $(\tilde{y}, \tilde{s}) \in \mathcal{D}$ and let $H \in \mathbb{R}^{(n-m)\times n}$ be a matrix whose rows form a basis of the null space of A. Then, for $(y, s) \in \mathbb{R}^m \times \mathbb{R}^n$, we have

$$A^T y + s = c \Longleftrightarrow Hs = H\tilde{s}.$$

Therefore, $s(x)$ is also the unique solution of the following problem

$$\begin{array}{ll} \text{maximize}_s & \frac{1}{2}\|Xs\|^2 \\ \text{subject to} & Hs = H\tilde{s}, \end{array} \tag{94}$$

where $X \equiv diag(x)$. Hence, by Lemma A.1, we know that the set $\{s(x) \mid x > 0\}$ is bounded. Since $rank(A) = m$, it also follows that $\{y(x) \mid x > 0\}$ is bounded. ∎

Below we give a proof of Lemma 2.10.

Proof of Lemma 2.10 It is easy to verify that $d^k/\|(X^k)^{-1}d^k\|$ is the optimal solution of

$$\begin{array}{ll} \text{maximize}_{\,p} & c^T p \\ \text{subject to} & Ap = 0 \\ & \|(X^k)^{-1}p\| \le 1. \end{array} \tag{95}$$

where $X^k \equiv diag(x^k)$. Hence, since $p = (x^k - x^*)/\|(X^k)^{-1}(x^k - x^*)\|$ is feasible to (95), we obtain

$$\frac{c^T(x^k - x^*)}{\|(X^k)^{-1}(x^k - x^*)\|} \le \frac{c^T d^k}{\|(X^k)^{-1}d^k\|} = \|(X^k)^{-1}d^k\| = \|X^k s^k\|, \tag{96}$$

where the two equalities follows from Lemma 2.2. Since $\{x^k\}$ converges to x^*, it is easy to verify that for some integer $k_0 \ge 0$,

$$\|(X^k)^{-1}(x^k - x^*)\| \le \sqrt{n}, \qquad \forall k \ge k_0. \tag{97}$$

Using relations (16), (96) and (97), we obtain

$$\begin{aligned} \frac{c^T x^{k+1} - v^*}{c^T x^k - v^*} &\le 1 - \lambda \frac{\|X^k s^k\|}{c^T x^k - v^*} \\ &= 1 - \lambda \frac{\|X^k s^k\|}{c^T(x^k - x^*)} \\ &\le 1 - \lambda \frac{1}{\|(X^k)^{-1}(x^k - x^*)\|} \\ &\le 1 - \frac{\lambda}{\sqrt{n}}. \end{aligned}$$

$\blacksquare$

Finally, we prove Lemma 3.9.

Proof of Lemma 3.9 Introducing a new variable $\tilde{v} = w - e/q$, we have

$$1 - \rho w_i = 1 - \frac{\rho}{q} - \rho \tilde{v}_i, \quad 1 - \rho \|w\|^2 = 1 - \frac{\rho}{q} - \rho \|\tilde{v}\|^2.$$

Since $w^T e = 1$, we obtain

$$0 < \rho < q, \tag{98}$$

$$\chi[\tilde{v}] = \chi[w] - \frac{1}{q} \geq 0, \tag{99}$$

$$\tilde{v}^T e = 0.$$

Using (98), we can write G_q in terms of $\tilde{v}$ as follows:

$$G_q(w, \rho) = q \log(1 - \theta \|\tilde{v}\|^2) - \sum_{i=1}^{q} \log(1 - \theta \tilde{v}_i), \tag{100}$$

where

$$\theta = \frac{q\rho}{q - \rho}.$$

We now make use of the following well-known inequalities:

$$\begin{aligned}
\log(1-\delta) &\leq -\delta \quad (\delta < 1), \\
\sum_{i=1}^{q} \log(1-\eta_i) &= \sum_{i:\eta_i > -\chi[\eta]} \left(-\eta_i - \frac{\eta_i^2}{2} - \frac{\eta_i^3}{3} - \cdots \right) + \sum_{i:\eta_i \leq -\chi[\eta]} \log(1-\eta_i) \\
&\geq \sum_{i:\eta_i > -\chi[\eta]} \left(-|\eta_i| - \frac{|\eta_i|^2}{2} - \frac{|\eta_i|^3}{2} - \cdots \right) + \sum_{i:\eta_i \leq -\chi[\eta]} \log(1-\eta_i) \\
&\geq \sum_{i:\eta_i > -\chi[\eta]} \left(-\eta_i - \frac{\eta_i^2}{2(1-|\eta_i|)} \right) + \sum_{i:\eta_i \leq -\chi[\eta]} \left(-\eta_i - \frac{\eta_i^2}{2(1-\chi[\eta])} \right) \\
&\geq -\eta^T e - \frac{\|\eta\|^2}{2(1-\chi[\eta])} \qquad (\eta \in R^q, \quad 0 \leq \chi[\eta] < 1).
\end{aligned}$$

Plugging these inequalities into (100) and using (99), we obtain

$$\begin{aligned}
G_q(w, \rho) &= q \log(1 - \theta \|\tilde{v}\|^2) - \sum_{i=1}^{q} \log(1 - \theta \tilde{v}_i) \\
&\leq -q\theta \|\tilde{v}\|^2 + \frac{\theta^2 \|\tilde{v}\|^2}{2(1 - \theta\chi[\tilde{v}])} \\
&= \theta \|\tilde{v}\|^2 \left(-q + \frac{\theta}{2(1 - \theta\chi[\tilde{v}])} \right),
\end{aligned} \tag{101}$$

provided that $\theta\chi[\tilde{v}] < 1$ and $\theta||\tilde{v}||^2 < 1$. It is easy to see that these conditions are satisfied under the assumptions of the lemma, since they are equivalent to $\rho\chi[w] < 1$ and $\rho||w||^2 < 1$.

Substituting the definition of $\tilde{v}$ and θ into the rightmost hand side of (101), we obtain

$$G_q(w,\rho) \leq \frac{q\rho}{q-\rho}\left\|w - \frac{1}{q}e\right\|^2\left(-q + \frac{\rho}{2(1-\rho\chi[w])}\right) = H_q(w,\rho).$$

∎

Acknowledgment

Part of this research was supported by the National Science Foundation (NSF) under Grant No. DDM-9109404. The first author thanks NSF for the financial support received during the completion of this work.

Part of this research was carried out while the second author was visiting the Center for Research on Parallel Computation and the Department of Computational and Applied Mathematics of Rice University, Houston, USA. He thanks his host Prof. J. E. Dennis and the colleagues there for providing the excellent research environment during his stay at Rice University.

The second author also thanks the first author for having arranged and supported a trip with his NSF grant to visit the University of Arizona where this research was initiated. He also thanks the University of Arizona for the congenial scientific atmosphere that it provided.

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*, 44:297–335, 1989. Errata in *Mathematical Programming*, 50:415, 1991.

[2] I. Adler and R. D. C. Monteiro. Limiting behavior of the affine scaling continuous trajectories for linear programming problems. *Mathematical Programming*, 50:29–51, 1991.

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

[4] D. A. Bayer and J. C. Lagarias. The nonlinear geometry of linear programming, Part I : Affine and projective scaling trajectories. *Transactions of the American Mathematical Society*, 314(2):499–526, 1989.

[5] J. Birge and C. Rosa. A simplified proof of the convergence of the affine scaling. Research Report 91-7, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan, 1991.

[6] I. I. Dikin. Iterative solution of problems of linear and quadratic programming. *Doklady Akademii Nauk SSSR*, 174:747–748, 1967. Translated in : *Soviet Mathematics Doklady*, 8:674–675, 1967.

[7] I. I. Dikin. On the convergence of an iterative process. *Upravlyaemye Sistemi*, 12:54–60, 1974. (In Russian).

[8] I. I. Dikin. The convergence of dual variables. Technical Report, Siberian Energy Institute, Irkutsk, Russia, December 1991.

[9] I. I. Dikin. Determination of the interior point of one system of linear inequalities. *Kibernetica and system analysis*, 1, 1992.

[10] I. I. Dikin and V. I. Zorkaltsev. *Iterative solutions of mathematical programming problems.* Nauka, Novosibirsk, 1980.

[11] C. C. Gonzaga. Convergence of the large step primal affine–scaling algorithm for primal nondegenerate linear programs. Technical Report ES–230/90, Dept. of Systems Engineering and Computer Science, COPPE Federal University of Rio de Janeiro, 21941 Rio de Janeiro, RJ, Brazil, September 1990.

[12] L. A. Hall and R. J. Vanderbei. Two-thirds is sharp for affine scaling. Technical Report SOR–92/9, Department of Civil Engineering and Operations Research, Princeton University, Princeton, New Jersey, 08544, September 1992.

[13] A. J. Hoffman. On approximate solutions of systems of linear inequalities. *Journal of Research of the National Bureau of Standards*, 49:263–265, 1952.

[14] N. K. Karmarkar. A new polynomial–time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[15] N. K. Karmarkar and R. G. Ramakrishnan. Further developments in the new polynomial-time algorithm for linear programming. Talk given at ORSA/TIMS National Meeting, Boston, MA, April, 1985.

[16] N. Megiddo and M. Shub. Boundary behavior of interior point algorithms in linear programming. *Mathematics of Operations Research*, 14:97–114, 1989.

[17] C. L. Monma and A. J. Morton. Computational experience with the dual affine variant of Karmarkar's method for linear programming. *Operations Research Letters*, 6:261–267, 1987.

[18] R. Saigal. A simplified proof of primal affine scaling method. Technical Report, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan, USA, 1993.

[19] P. Tseng and Z. Q. Luo. On the convergence of the affine–scaling algorithm. *Mathematical Programming*, 56:301–319, 1992.

[20] T. Tsuchiya. Global convergence of the affine–scaling methods for degenerate linear programming problems. *Mathematical Programming*, 52:377–404, 1991.

[21] T. Tsuchiya. Global convergence property of the affine scaling method for primal degenerate linear programming problems. *Mathematics of Operations Research*, 17(3):527–557, 1992.

[22] 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).

[23] R. J. Vanderbei and J. C. Lagarias. I. I. Dikin's convergence result for the affine–scaling algorithm. In J. C. Lagarias and M. J. Todd, editors, *Mathematical Developments Arising from Linear Programming : Proceedings of a Joint Summer Research Conference held at Bowdoin College, Brunswick, Maine, USA, June/July 1988*, volume 114 of *Contemporary Mathematics*, pages 109–119. American Mathematical Society, Providence, Rhode Island, USA, 1990.

[24] R. J. Vanderbei, M. S. Meketon, and B. A. Freedman. A modification of Karmarkar's linear programming algorithm. *Algorithmica*, 1(4):395–407, 1986.

[25] C. Witzgall, P. T. Boggs, and P. D. Domich. On the convergence behavior of trajectories for linear programming. In J. C. Lagarias and M. J. Todd, editors, *Mathematical Developments Arising from Linear Programming : Proceedings of a Joint Summer Research Conference held at Bowdoin College, Brunswick, Maine, USA, June/July 1988*, volume 114 of *Contemporary Mathematics*, pages 161–187. American Mathematical Society, Providence, Rhode Island, USA, 1990.