Mathematical Programming 75 (1996) 77-110

# Superlinear convergence of the affine scaling algorithm

T. Tsuchiya [a,*], R.D.C. Monteiro [b,1]

[a] *The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-Ku, Tokyo 106, Japan*
[b] *School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA*

Received 4 March 1993; revised manuscript received 27 February 1996

**Abstract**

In this paper we show that a variant of the long-step affine scaling algorithm (with variable stepsizes) is two-step superlinearly convergent when applied to general linear programming (LP) problems. Superlinear convergence of the sequence of dual estimates is also established. For homogeneous LP problems having the origin as the unique optimal solution, we also show that $\frac{2}{3}$ is a sharp upper bound on the (fixed) stepsize that provably guarantees that the sequence of primal iterates converge to the optimal solution along a unique direction of approach. Since the point to which the sequence of dual estimates converge depend on the direction of approach of the sequence of primal iterates, this result gives a plausible (but not accurate) theoretical explanation for why $\frac{2}{3}$ is a sharp upper bound on the (fixed) stepsize that guarantees the convergence of the dual estimates.

*Keywords:* Interior point algorithms; Affine scaling algorithm; Linear programming; Superlinear convergence; Global convergence

## 1. Introduction

The affine scaling (AS) algorithm, introduced by Dikin [6] in 1967, is one of the simplest and most efficient interior point algorithms for solving linear programming

* Corresponding author. E-mail: tsuchiya@sun312.ism.ac.jp. The work of this author was based on research supported by the Overseas Research Scholars of the Ministry of Education, Science and Culture of Japan, 1992.

[1] The work of this author was based on research supported by the National Science Foundation (NSF) under grant DDM-9109404 and the Office of Naval Research (ONR) under grant N00014-93-1-0234. This work was done while the second author was a faculty member of the Systems and Industrial Engineering Department at the University of Arizona.


Published by Elsevier Science B.V.
*PII* S0025-5610(96)00025-1

(LP) problems. Because of its theoretical and practical importance, there are a number of papers which study its global and local convergence [4,6–8,12,16,25,26,28–31] as well as its continuous trajectories [3,16,32]. For computational experiments and implementation issues related to the AS algorithm, we refer the reader to [1,2,5,18,21,22].

Recently, Dikin [8] and Tsuchiya and Muramatsu [29] proved global convergence of the long-step version of the AS algorithm [31] for degenerate LP problems. This long-step version is the one 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. Assuming that $\lambda = \frac{1}{2}$, Dikin [8] showed the sequence of primal iterates converges to a point lying in the relative interior of the optimal face and that the sequence of dual estimates converges to the analytic center of the dual optimal face. Independently, Tsuchiya and Muramatsu [29] obtained an analogous result under the less restrictive condition that $\lambda \leqslant \frac{2}{3}$. They also demonstrated that the asymptotic reduction rate of the objective function value is exactly $1-\lambda$, under the same assumption that $\lambda \leqslant \frac{2}{3}$. A simplified and self-contained proof of these results can be found in the recent survey by Monteiro, Tsuchiya and Wang [19].

In this paper we focus our attention on the asymptotic convergence properties of the long-step AS algorithm with variable stepsizes $\lambda_k$. Specifically, we develop a variant which is two-step superlinearly convergent by properly choosing the sequence of stepsizes $\{\lambda_k\}$. The algorithm is based on a centrality measure in the space of the "small" variables. When this measure is small, we show that, asymptotically, it is possible to take stepsizes sufficiently close to 1 to force the reduction rate of the objective function value as close to 0 as desired without loosing too much centrality. At the next step, if necessary, we select the stepsize $\lambda_k = \frac{1}{2}$ to recover the centrality of the iterate.

This paper is organized as follows. In Section 2, we introduce basic assumptions, terminology and notation. The long-step AS algorithm and some of its basic properties are also reviewed.

The main content of the paper is given in Sections 3, 4 and 5. The main result obtained in Section 3 is somewhat independent of (though related to) the results of Sections 4 and 5. It deals with the case of the AS algorithm applied to a homogeneous LP problem with the origin as the unique optimal solution. In this case, we show that, when the sequence of stepsizes $\{\lambda_k\}$ satisfies $\liminf_{k\to\infty} \lambda_k > \frac{2}{3}$, the direction of approach of the primal iterates towards the (unique) optimal solution always oscillates. This result contrasts with the case where $\lambda_k = \lambda \leqslant \frac{2}{3}$ for all $k \geqslant 0$, for which it is shown that the direction of approach is unique. Since the point to which the sequence of dual estimates converges depends on the direction of approach of the sequence of primal iterates, this result gives a plausible (but not accurate) theoretical explanation for why $\frac{2}{3}$ is a sharp upper bound on the (fixed) stepsize $\lambda$ that provably guarantees the convergence of the dual estimates. Specific examples illustrating that $\frac{2}{3}$ is indeed sharp in the above sense were given by Tsuchiya and Muramatsu [29] and Hall and Vanderbei [13].

The above result is obtained by observing that the sequence of points obtained by conically projecting the sequence of the AS iterates for the homogeneous problem onto a constant-cost hyperplane (that is, a hyperplane where the objective function is constant) is exactly the sequence obtained by applying Newton's method (with variable stepsizes) to the optimization problem defining the analytic center of the polyhedron determined by the intersection of the constant-cost hyperplane with the feasible (conical) region of the homogeneous problem. In conjunction with this, we also show that the projected sequence converges quadratically to the analytic center when $\lambda_k = \lambda = \frac{1}{2}$ for all $k \geqslant 0$. This result suggests that the AS iteration with $\lambda_k = \frac{1}{2}$ can be used as a kind of centering step to keep the iterate "well-centered".

In Section 4, we show that the relation established in Section 3 between the AS algorithm for the homogeneous problem and Newton's method for the analytic center problem can be used to analyze the sequence of AS iterates for general LP problems. Close to a constant-cost face, it is possible to approximate the original problem by a homogeneous problem in the sense that the AS directions at a point $x$ for the two problems asymptotically approach each other as $x$ approaches the face. Hence, near a constant-cost face, the iterates generated by the AS algorithm applied to a general problem behave very much like the ones generated by the AS algorithm applied to a homogeneous problem. The analysis of Section 4 forms the basis for the development of the superlinear AS algorithm presented in Section 5.

We show in Section 5 that the new variant of the AS algorithm, whose sequence of stepsizes asymptotically alternate between $\lambda_k = \frac{1}{2}$ and $\lambda_k \sim 1$, is two-step superlinearly convergent with $Q$-order $1+p$ with respect to the sequence of objective function values, where $p$ is any a priori chosen constant in the interval $(0, \frac{1}{3})$. Superlinear convergence of the sequences of primal iterates and dual estimates to a point in the relative interior of the optimal face and to the analytic center of the dual optimal face, respectively, with $R$-order $1+p$ is also shown. Finally, we give some remarks in Section 6.

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. The symbols $\mathbb{R}^n$, $\mathbb{R}^n_+$ 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 $\mathbb{R}^{m\times n}$. If $J$ is a finite index set then $|J|$ denotes its cardinality, that is the number of elements of $J$. For $J \subseteq \{1,\dots,n\}$ and $w \in \mathbb{R}^n$, we let $w_J$ denote the subvector $[w_i]_{i\in J}$; moreover, 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 $w \in \mathbb{R}^n$, we let $\max(w)$ denote the largest component of $w$, $\operatorname{diag}(w)$ denote the diagonal matrix whose $i$-th diagonal element is $w_i$ for $i = 1,\dots,n$ and $w^{-1}$ denote the vector $[\operatorname{diag}(w)]^{-1}e$ whenever it is well-defined. The Euclidean norm, the 1-norm and the $\infty$-norm are denoted by $\|\cdot\|$, $\|\cdot\|_1$ and $\|\cdot\|_\infty$, respectively. The superscript $^{\mathrm{T}}$ denotes transpose.

To avoid introducing several constants throughout the paper, we use the following notation. Given functions $g_1(x)$ and $g_2(x)$ which are defined for points on a set $E$, we say that $g_1(x) = \mathcal{O}(g_2(x))$ for every $x \in E$ if there exists some constant $M$ such that $\|g_1(x)\| \leqslant M\|g_2(x)\|$ for every $x \in E$. When the conditions $g_1(x) = \mathcal{O}(g_2(x))$ for every $x \in E$ and $g_2(x) = \mathcal{O}(g_1(x))$ for every $x \in E$ hold then we simply write $g_1(x) \sim g_2(x)$ for every $x \in E$.

## 2. Affine scaling algorithm

In this section, we state the main terminology and assumptions used throughout our paper and describe the AS algorithm. We also review some basic properties of the AS algorithm that are needed in the subsequent sections.

Consider the following LP problem

$$\begin{aligned} &\text{minimize}_x \quad c^T x \\ &\text{subject to} \quad Ax = b, \quad x \geqslant 0, \end{aligned} \tag{1}$$

and its associated dual problem

$$\begin{aligned} &\text{maximize}_{(y,s)} \quad b^T y \\ &\text{subject to} \quad A^T y + s = c, \quad s \geqslant 0, \end{aligned} \tag{2}$$

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 and definitions which will be used throughout our paper. Given a point $x \in \mathbb{R}^n$, let $B(x) \equiv \{i : x_i \neq 0\}$ and $N(x) \equiv \{i : x_i = 0\}$. Clearly, $(N(x), B(x))$ determines a partition of $\{1, \ldots, n\}$. Associated with any specific partition $(N, B)$ of $\{1, \ldots, n\}$, we let

$$\mathcal{P}_N \equiv \{x \in \mathbb{R}^n : Ax = b, x_N = 0\}, \tag{3}$$

$$\mathcal{P}_N^+ \equiv \{x \in \mathcal{P}_N : x_B \geqslant 0\}, \tag{4}$$

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

$$\mathcal{D}_B \equiv \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n : A^T y + s = c, s_B = 0\}, \tag{6}$$

$$\mathcal{D}_B^+ \equiv \{(y, s) \in \mathcal{D}_B : s_N \geqslant 0\}, \tag{7}$$

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

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 $B = \emptyset$, we denote the sets $\mathcal{D}_B$, $\mathcal{D}_B^+$ and $\mathcal{D}_B^{++}$ 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).

A *constant-cost face* of an LP problem is a nonempty face of the feasible polyhedron over which the objective function is constant. Every nonempty face $\mathcal{F}$ of $\mathcal{P}^+$ is uniquely determined by a partition $(N, B)$ in the sense that $\mathcal{P}_N^+ = \mathcal{F}$ and $\mathcal{P}_N^{++} \neq \emptyset$. Every partition $(N, B)$ which is uniquely associated with a constant-cost face of (1) is called a *constant-cost partition*. If $(N, B)$ is a constant-cost partition then the constant value of the objective function $c^T x$ over $\mathcal{P}_N^+$ is denoted by $\nu_N$. The partition associated with the optimal face of (1) is referred to as the *optimal partition*.

The following result can be easily shown.

**Lemma 2.1.** *The following statements hold:*
(a) *$(N, B)$ is a constant-cost partition if and only if $\mathcal{P}_N^{++} \neq \emptyset$ and $\mathcal{D}_B \neq \emptyset$;*
(b) *if $(N, B)$ is a constant-cost partition then $c^{\mathrm{T}}x - \nu_N = \bar{s}_N^{\mathrm{T}} x_N$ for any $x \in \mathcal{P}$ and $(\bar{y}, \bar{s}) \in \mathcal{D}_B$.*

We impose the following assumptions throughout this paper.

**Assumption 1.** $\mathrm{Rank}(A) = m$.

**Assumption 2.** The objective function $c^{\mathrm{T}}x$ is not constant over the feasible region of problem (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 analysis of the AS algorithm. For every $x \in \mathbb{R}^n_{++}$, let

$$y(x) \equiv (AX^2A^{\mathrm{T}})^{-1}AX^2c, \tag{9a}$$

$$s(x) \equiv c - A^{\mathrm{T}}y(x), \tag{9b}$$

$$d(x) \equiv X^2 s(x) = X[I - XA^{\mathrm{T}}(AX^2A^{\mathrm{T}})^{-1}AX]Xc, \tag{9c}$$

where $X \equiv \mathrm{diag}(x)$. Note that Assumption 1 implies that the inverse of $AX^2A^{\mathrm{T}}$ exists for every $x > 0$. The quantities $(y(x), s(x))$ and $d(x)$ are the *dual estimate* and the *AS direction* associated with the point $x$, respectively. For the purpose of future reference, we note that (9c) implies

$$X^{-1}d(x) = Xs(x). \tag{10}$$

**Lemma 2.2.** *The following statements hold:*
(a) *for any $(\bar{y}, \bar{s}) \in \mathcal{D}$, $d(x)$ is the (unique) optimal solution of*

$$\begin{array}{ll} \text{maximize}_d & \bar{s}^{\mathrm{T}}d - \frac{1}{2}\|X^{-1}d\|^2 \\ \text{subject to} & Ad = 0; \end{array} \tag{11}$$

(b) *if $(N, B)$ is a constant-cost partition then there exists a constant $C_0 > 0$ such that*

$$\|X_B^{-1}d_B(x)\| \leqslant C_0 \|X_B^{-1}\| \|X_N\| \|X_N^{-1}d_N(x)\| \quad \forall x > 0.$$

**Proof.** The proof of (a) is straightforward. The proof of (b) is given in Monteiro et al. [19, Lemma 3.6]. □

We are ready to describe the AS algorithm. For a good motivation of the method, we refer the reader to Dikin [6], Barnes [4], Vanderbei, Meketon and Freedman [31] and Vanderbei and Lagarias [30].

**Algorithm 1** (Affine Scaling Algorithm)
*Step* 0. Assume $x^0 \in \mathcal{P}^{++}$ is available. Set $k := 0$.
*Step* 1. Choose $\lambda_k \in (0,1)$, and let

$$d^k = d(x^k), \tag{12a}$$

$$X^k = \mathrm{diag}(x^k), \tag{12b}$$

$$x^{k+1} = x^k - \frac{\lambda_k}{\max((X^k)^{-1}d^k)} d^k. \tag{12c}$$

*Step* 2. $k := k+1$ and return to Step 1.

We note that Assumptions 1–4 imply that, for every $x \in \mathcal{P}^{++}$, the direction $d(x)$ must have at least one positive component so that $\max(X^{-1}d(x)) > 0$. Hence, the expression which determines $x^{k+1}$ in the AS algorithm is well-defined. Observe also that if $\lambda_k$ were equal to 1, the iterate $x^{k+1}$ would lie in the boundary of the feasible region. Thus, since we choose $\lambda_k \in (0,1)$, $x^{k+1}$ is ensured to be a point in $\mathcal{P}^{++}$.

The following basic result whose proof can be found in Vanderbei and Lagarias [30, p. 118] or in Monteiro et al. [19, Proposition 2.8] will be needed later on.

**Proposition 2.3.** *For any full row rank matrix $A \in \mathbb{R}^{m\times n}$ and any vector $c \in \mathbb{R}^n$, the set $\{(y(x), s(x)) : x > 0\}$ is bounded, where $(y(x), s(x))$ is defined in* (9).

We next summarize the main results that have been proved for the AS algorithm. Proofs of these results can be found in Tsuchiya and Muramatsu [29] and in the survey paper by Monteiro et al. [19]. The results below are stated in more general terms than they have been stated originally to accommodate the needs of the current paper. But their proofs follow along the same lines pursued in the above two references. Let $\{x^k\}$ denote the sequence of iterates generated by Algorithm 1 and let $\{(y^k, s^k)\}$ denote the sequence of dual estimates defined as $(y^k, s^k) = (y(x^k), s(x^k))$ for all $k \geqslant 0$.

**Proposition 2.4.** *The following statements hold*:
(a) *the sequence $\{x^k\}$ converges to some point $x^* \in \mathcal{P}^+$;*
(b) *there exists $M > 0$ such that $\|x^k - x^*\| \leqslant M(c^{\mathrm{T}}x^k - c^{\mathrm{T}}x^*)$ for all $k \geqslant 0$;*
(c) *the sequence $\{(y^k, s^k)\}$ is bounded.*

*If, in addition, we have $\liminf_{k\to\infty} \lambda_k > 0$ then*:
(d) *$(N_*, B_*) \equiv (N(x^*), B(x^*))$ is a constant-cost partition, or equivalently, the smallest face containing $x^*$, namely $\mathcal{P}^+_{N_*}$, is a constant-cost face;*
(e) *every accumulation point $(y^*, s^*)$ of $\{(y^k, s^k)\}$ is in $\mathcal{D}_{B_*}$ (hence, $X^*s^* = 0$).*

The *analytic center* of the optimal face of problem (2) is the (unique) point defined as

$$(\bar{y}^a, \bar{s}^a) = \operatorname{argmax}\Big\{ \sum_{j\in N_{\rm opt}} \log s_j : (y,s) \in \mathcal{D}^{++}_{B_{\rm opt}} \Big\}, \tag{13}$$

where $(N_{\rm opt}, B_{\rm opt})$ denotes the optimal partition of (1).

**Proposition 2.5.** *If the sequence $\{\lambda_k\}$ satisfies $\lambda_k \leqslant \frac{2}{3}$ for all $k \geqslant 0$ and $\liminf_{k\to\infty} \lambda_k > 0$, then the following statements hold:*

(a) *$\{x^k\}$ converges to a point $x^*$ lying in the relative interior of the optimal face of (1) (hence, $(N_*, B_*) \equiv (N(x^*), B(x^*))$ is the optimal partition of (1));*
(b) *$\{(y^k, s^k)\}$ converges to $(\bar{y}^a, \bar{s}^a)$;*
(c) $\lim_{k\to\infty} X^k_{N_*} s^k_{N_*}/(c^{\rm T}x^k - c^{\rm T}x^*) = e/|N_*|$;
(d) *for any $(\bar{y}, \bar{s}) \in \mathcal{D}_{B_*}$, we have*

$$\lim_{k\to\infty} \frac{x^k_{N_*}}{c^{\rm T}x^k - c^{\rm T}x^*} = \frac{(\bar{s}^a)^{-1}}{|N_*|} = \operatorname{argmax}\left\{ \sum_{j\in N_*} \log x_j : \begin{array}{l} A_{N_*}x_{N_*} \in \operatorname{Range}(A_{B_*}) \\ \bar{s}^{\rm T}_{N_*}x_{N_*} = 1,\ x_{N_*} > 0 \end{array} \right\}. \tag{14}$$

**Remark.** Statement (d) is not explicitly stated in [19] and [29]; however, the first equality in (d) follows immediately from (b) and (c), while the second one follows by verifying that $(\bar{s}^a)^{-1}/|N_*|$ satisfies the optimality condition for the optimization problem in (14) (see Lemma 4.7).

## 3. Asymptotic behavior of the AS algorithm for a homogeneous problem

It was shown in the original version of Tsuchiya and Muramatsu [29] that the sequence of dual estimates $\{(y^k, s^k)\}$ converges to the analytic center $(\bar{y}^a, \bar{s}^a)$ of the dual optimal face whenever $\lambda_k = \lambda \in (0, \frac{2}{3})$ for all $k \geqslant 0$ (see Proposition 2.5(b)). Later, Tsuchiya and Muramatsu [29] pointed out that their result holds even for $\lambda_k = \frac{2}{3}$. Furthermore, they [29] and Hall and Vanderbei [13] gave specific examples showing that the bound $\frac{2}{3}$ on the (fixed) stepsize is tight with respect to the property that $\lim_{k\to\infty}(y^k, s^k) = (\bar{y}^a, \bar{s}^a)$. In this section, we give a plausible explanation for the tightness of the bound $\frac{2}{3}$.

Specifically, we show for any homogeneous LP problem that $\frac{2}{3}$ is a sharp upper bound on the fixed stepsize that provably guarantees that the sequence $\{x^k\}$ converges to the optimal solution along a unique direction of approach. For an arbitrary LP problem and for $\lambda \leqslant \frac{2}{3}$, the uniqueness of the direction of approach of $\{x^k\}$ follows as a consequence of Proposition 2.5(c) and Lemma 4.9. The main result of this section shows that the direction of approach of $\{x^k\}$ towards the optimal solution is not unique, whenever the sequence of stepsizes $\{\lambda_k\}$ satisfies $\liminf_{k\to\infty} \lambda_k > \frac{2}{3}$ (e.g., $\lambda_k = \lambda > \frac{2}{3}$ for all $k \geqslant 0$), the LP problem is homogeneous and 0 is its unique optimal solution.

Since the accumulation points of the sequence $\{(y^k, s^k)\}$ are determined by the set of directions of approach of $\{x^k\}$ (this fact can be proved by using similar arguments as in Adler and Monteiro [3, Theorem 4.1]), the above result gives a plausible theoretical explanation for why $\frac{2}{3}$ is a tight bound on the (fixed) stepsize that guarantees the convergence of $\{(y^k, s^k)\}$ to the analytic center $(\bar{y}^a, \bar{s}^a)$. This explanation is not accurate though since existence of two or more directions of approach of $\{x^k\}$ does not imply (but is likely to result in) nonconvergence of the sequence $\{(y^k, s^k)\}$.

The main observation used in this section is that the sequence of points $\{r^k\}$ obtained by conically projecting $\{x^k\}$ onto a constant-cost hyperplane (that is, a hyperplane where the objective function is constant) is exactly the sequence obtained by applying Newton's method with a sequence of variable stepsizes $\{\tau_k\}$ to the optimization problem defining the analytic center, say $r^*$, of the polyhedron determined by the intersection of the constant-cost hyperplane with the feasible (conical) region of the homogeneous problem. One important consequence of this observation is that the sequence $\{r^k\}$ converges quadratically to $r^*$ when $\lambda_k = \lambda = \frac{1}{2}$ for all $k \geqslant 0$. This result suggests that the AS iteration with $\lambda_k = \frac{1}{2}$ can be used as a kind of centering step to keep the iterates "well-centered". Another important consequence is that when $\lambda_k = \lambda > \frac{2}{3}$ for all $k \geqslant 0$, the corresponding sequence of Newton stepsizes $\{\tau_k\}$ satisfies $\liminf_{k\to\infty} \tau_k > 2$ from which nonconvergence of the sequence $\{r^k\}$ easily follows.

The following homogeneous problem is considered in this section. Given a vector $\tilde{c} \in \mathbb{R}^p$ and a subspace $H \subseteq \mathbb{R}^p$, the problem is to

$$\text{minimize } \{\tilde{c}^T\tilde{x} : \tilde{x} \in H,\ \tilde{x} \geqslant 0\}. \tag{15}$$

Define

$$H^{++} \equiv \{\tilde{x} \in H : \tilde{x} > 0\},$$
$$H^{++}_{>} \equiv \{\tilde{x} \in H^{++} : \tilde{c}^T\tilde{x} > 0\},$$
$$H^{++}_{1} \equiv \{\tilde{x} \in H^{++} : \tilde{c}^T\tilde{x} = 1\}.$$

Throughout this section we assume that $H^{++}_{>} \neq \emptyset$, or equivalently, $H^{++}_{1} \neq \emptyset$.

The AS direction at a point $\tilde{x} \in H^{++}$ is the (unique) solution $\tilde{d}(\tilde{x})$ of the problem

$$\text{maximize } \{\tilde{c}^T\tilde{d} - \tfrac{1}{2}\|\tilde{X}^{-1}\tilde{d}\|^2 : \tilde{x} \in H\}, \tag{16}$$

where $\tilde{X} \equiv \operatorname{diag}(\tilde{x})$.

One of our goals in this section is to give the relationship between the direction $\tilde{d}(\tilde{x})$ and the Newton direction at the point $r = \tilde{x}/\tilde{c}^T\tilde{x} \in H^{++}_1$ with respect to the following maximization problem:

$$\text{maximize } \left\{\sum_{i=1}^{p} \log r_i : r \in H^{++}_1\right\}. \tag{17}$$

Given $r \in H^{++}_1$, the Newton direction of (17) at $r$ is the (unique) solution $\eta(r)$ of the problem

$$\text{maximize } \{(r^{-1})^{\mathrm{T}}(-\eta) - \tfrac{1}{2}\eta^{\mathrm{T}}R^{-2}\eta : \eta \in H,\ \tilde{c}^{\mathrm{T}}\eta = 0\}, \tag{18}$$

where $R = \mathrm{diag}(r)$ and the variable $\eta$ belongs to $\mathbb{R}^p$. With this $\eta(r)$, one iteration of the Newton method with a unit stepsize at the point $r^k$ is written as $r^{k+1} = r^k - \eta(r^k)$.

The proof of the following result is straightforward.

**Lemma 3.1.** *Assume that $H^{++} \neq \emptyset$. Then, the following statements are equivalent:*
(a) *$\tilde{x} = 0$ is the unique optimal solution of* (15);
(b) *$H_1^{++}$ is nonempty and bounded;*
(c) *problem* (17) *has a (unique) optimal solution.*

The optimal solution of (17), when it exists, is denoted by $r^*$. The following result plays an important role in several parts of the paper.

**Lemma 3.2.** *The following statements hold:*
(a) *the function $r \mapsto \eta(r)$ is continuous on $H_1^{++}$;*
(b) *for $r \in H_1^{++}$, $\eta(r) \neq 0$ if and only if $r$ is not the optimal solution of* (17);
(c) *if the optimal solution $r^*$ of* (17) *exists then*

$$\limsup_{r\to r^*,\ r\in H_1^{++}} \frac{\|r - r^* - \eta(r)\|}{\|r - r^*\|^2} < \infty, \tag{19}$$

$$\limsup_{r\to r^*,\ r\in H_1^{++}} \frac{\|\eta(r)\|}{\|r - r^*\|} = 1. \tag{20}$$

**Proof.** The proof of (a) and (b) are straightforward. Relation (19) is a standard property of Newton methods and it holds whenever some reduced Hessian (see Fletcher [11, p. 260]) of the objective function of (17) is nonsingular at $r^*$. This last property follows due to the fact that the (full) Hessian of $\sum_{i=1}^{p} \log r_i$ is negative definite at $r^*$. Relation (20) follows as an immediate consequence of (19). □

**Lemma 3.3.** *Assume that $\{r^k\} \subset H_1^{++}$ is a sequence determined by the recurrence relation $r^{k+1} = r^k - \tau_k\eta(r^k)$, where $\{\tau_k\}$ is a sequence of scalars such that $\liminf_{k\to\infty} \tau_k > 2$. Then*
(i) *$r^k = r^*$ holds for all $k$ sufficiently large (this can happen only if $r^*$ exists), or*
(ii) *$\{r^k\}$ can not converge to a point in $H_1^{++}$.*

**Proof.** Assume that (i) does not hold. We will show that (ii) must hold. Indeed, in view of Lemma 3.2(b), we know that if $r^{k_0} = r^*$ for some $k_0$ then $r^k = r^*$ for all $k > k_0$. Since we are assuming that (i) does not hold, we conclude that $r^k \neq r^*$ for all $k \geqslant 0$. To show that (ii) holds, assume for contradiction that $\{r^k\}$ converges to a point $r^\infty \in H_1^{++}$. The relation $r^{k+1} = r^k - \tau_k\eta(r^k)$ and the fact that $\liminf_{k\to\infty} \tau_k > 2$ imply that $\lim_{k\to\infty} \eta(r^k) = 0$. By (a) and (b) of Lemma 3.2, we conclude that $r^\infty = r^*$, and hence that $\lim_{k\to\infty} r^k = r^*$. Using the relation $r^{k+1} = r^k - \tau_k\eta(r^k)$ and Lemma 3.2(c), we obtain

$$\begin{aligned}\liminf_{k\to\infty}\frac{\|r^{k+1}-r^*\|}{\|r^k-r^*\|} &= \liminf_{k\to\infty}\frac{\|-(\tau_k-1)\eta(r^k)+(r^k-r^*-\eta(r^k))\|}{\|r^k-r^*\|}\\ &\geqslant \liminf_{k\to\infty}\frac{(\tau_k-1)\|\eta(r^k)\|-\|(r^k-r^*-\eta(r^k))\|}{\|r^k-r^*\|}\\ &= \liminf_{k\to\infty}(\tau_k-1)>1.\end{aligned}$$

Hence, we conclude that $\|r^{k+1}-r^*\| > \|r^k-r^*\|$ for every $k$ sufficiently large, which contradicts the fact that $\{r^k\}$ converges to $r^*$. □

Given a point $\tilde{x}\in H_>^{++}$, we define

$$\tilde{u}(\tilde{x}) = \frac{\tilde{X}^{-1}\tilde{d}(\tilde{x})}{\tilde{c}^{\mathrm{T}}\tilde{x}}; \tag{21}$$

$$r(\tilde{x}) = \frac{\tilde{x}}{\tilde{c}^{\mathrm{T}}\tilde{x}}, \tag{22}$$

where $\tilde{X}=\mathrm{diag}(\tilde{x})$.

**Lemma 3.4.** *The following relations hold for every $\tilde{x}\in H_>^{++}$:*

$$\tilde{c}^{\mathrm{T}}\tilde{d}(\tilde{x}) = \|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2, \tag{23}$$

$$e^{\mathrm{T}}\tilde{u}(\tilde{x}) = 1, \tag{24}$$

$$\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\| \leqslant \|\tilde{X}\tilde{c}\|, \tag{25}$$

$$\eta(r(\tilde{x})) = -r(\tilde{x}) + \frac{\tilde{d}(\tilde{x})}{\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2} = R(\tilde{x})\left(-e+\frac{\tilde{u}(\tilde{x})}{\|\tilde{u}(\tilde{x})\|^2}\right), \tag{26}$$

*where $R(\tilde{x})=\mathrm{diag}(r(\tilde{x}))$.*

**Proof.** Let $\tilde{x}\in H_>^{++}$ be given. Since $\tilde{d}(\tilde{x})$ is a solution of (16), we have

$$\tilde{c}-\tilde{X}^{-2}\tilde{d}(\tilde{x})\in H^{\perp}, \qquad \tilde{d}(\tilde{x})\in H, \tag{27}$$

where $H^{\perp}$ denotes the orthogonal complement of $H$. Multiplying the first relation in (27) on the left by $\tilde{d}(\tilde{x})^{\mathrm{T}}$ and using the second relation, we obtain (23). Multiplying the first relation in (27) on the left by $\tilde{x}^{\mathrm{T}}$ and using the fact that $\tilde{x}\in H$, we obtain $\tilde{c}^{\mathrm{T}}\tilde{x} = e^{\mathrm{T}}(\tilde{X}^{-1}\tilde{d}(\tilde{x}))$, which is equivalent to (24), due to (21). Using (23) and the Cauchy-Schwarz inequality, we obtain

$$\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2 = (\tilde{X}\tilde{c})^{\mathrm{T}}(\tilde{X}^{-1}\tilde{d}(\tilde{x})) \leqslant \|\tilde{X}\tilde{c}\|\,\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|,$$

which clearly implies (25). It remains to show (26). To simplify notation, let $r\equiv r(\tilde{x})$. Since $\eta(r)$ is the unique optimal solution of (18), the first equality in relation (26) follows once we show that $-r+\tilde{d}(\tilde{x})/\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2$ satisfies the optimality condition for (18), that is, $\eta=-r+\tilde{d}(\tilde{x})/\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2$ satisfies

$$-r^{-1}-R^{-2}\eta\in H^{\perp}+\mathbb{R}c, \quad \eta\in H, \quad \tilde{c}^{\mathrm{T}}\eta=0, \tag{28}$$

where $\mathbb{R}c = \{\lambda c : \lambda \in \mathbb{R}\}$ and $R = \mathrm{diag}(r)$. Indeed, relations (22) and (27) imply

$$-r^{-1} - R^{-2}\left(-r + \frac{\tilde{d}(\tilde{x})}{\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2}\right) = -\frac{R^{-2}\tilde{d}(\tilde{x})}{\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2}$$
$$= -\frac{(\tilde{c}^{\mathrm{T}}\tilde{x})^2\tilde{X}^{-2}\tilde{d}(\tilde{x})}{\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2} \in H^{\perp} + \mathbb{R}c.$$

Since $H$ is a subspace, $\tilde{x} \in H$ and $\tilde{d}(\tilde{x}) \in H$, we conclude that

$$-r + \frac{\tilde{d}(\tilde{x})}{\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2} = -\frac{\tilde{x}}{\tilde{c}^{\mathrm{T}}\tilde{x}} + \frac{\tilde{d}(\tilde{x})}{\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2} \in H.$$

Using (22) and (23), we obtain

$$\tilde{c}^{\mathrm{T}}\left(-r + \frac{\tilde{d}(\tilde{x})}{\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2}\right) = -\left(\tilde{c}^{\mathrm{T}}r - \frac{\tilde{c}^{\mathrm{T}}\tilde{d}(\tilde{x})}{\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2}\right)$$
$$= -\left(1 - \frac{\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2}{\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2}\right) = 0.$$

Hence, the first equality in (26) follows. The second equality in (26) follows from (21) and (22). □

Given a point $\tilde{x} \in H^{++}_{>}$ and a scalar $\lambda \in (0,1)$, let

$$\tilde{\theta}(\tilde{x}) = \frac{\|\tilde{u}(\tilde{x})\|^2}{\max(\tilde{u}(\tilde{x}))}; \tag{29}$$

$$\tilde{x}^{+}(\lambda) = \tilde{x} - \frac{\lambda}{\max(\tilde{X}^{-1}\tilde{d}(\tilde{x}))}\tilde{d}(\tilde{x}). \tag{30}$$

**Lemma 3.5.** *Let $\tilde{x} \in H^{++}_{>}$ and $\lambda > 0$ be such that $\tilde{x}^{+}(\lambda) \in H^{++}_{>}$. Then, the following relations hold:*

$$\tilde{c}^{\mathrm{T}}\tilde{x}^{+}(\lambda) = \tilde{c}^{\mathrm{T}}\tilde{x}(1 - \lambda\tilde{\theta}(\tilde{x})), \tag{31}$$

$$\frac{1}{\sqrt{p}} \leqslant \|\tilde{u}(\tilde{x})\| \leqslant \tilde{\theta}(\tilde{x}) \leqslant \frac{1}{\lambda}, \tag{32}$$

$$r(\tilde{x}^{+}(\lambda)) = r(\tilde{x}) - \frac{\lambda\tilde{\theta}(\tilde{x})}{1 - \lambda\tilde{\theta}(\tilde{x})}\eta(r(\tilde{x})). \tag{33}$$

**Proof.** Let $\tilde{x}^{+} = \tilde{x}^{+}(\lambda)$. Using (21), (23), (29) and (30), we obtain

$$\tilde{c}^{\mathrm{T}}\tilde{x}^{+} = \tilde{c}^{\mathrm{T}}\left(\tilde{x} - \frac{\lambda}{\max(\tilde{X}^{-1}\tilde{d}(\tilde{x}))}\tilde{d}(\tilde{x})\right) = \tilde{c}^{\mathrm{T}}\tilde{x}\left(1 - \frac{\lambda\|\tilde{X}^{-1}\tilde{d}(\tilde{x})\|^2}{(\tilde{c}^{\mathrm{T}}\tilde{x})\max(\tilde{X}^{-1}\tilde{d}(\tilde{x}))}\right)$$
$$= \tilde{c}^{\mathrm{T}}\tilde{x}\left(1 - \frac{\lambda\|\tilde{u}(\tilde{x})\|^2}{\max(\tilde{u}(\tilde{x}))}\right) = \tilde{c}^{\mathrm{T}}\tilde{x}(1 - \lambda\tilde{\theta}(\tilde{x})),$$

and hence (31) follows. The first and second inequalities in (32) follow from (24) and (29). Since $\tilde{x}$ and $\tilde{x}^{+}$ are in $H^{++}$, relation (31) imply that $1 - \lambda\tilde{\theta}(\tilde{x}) > 0$, from which the third inequality in (32) follows. We next show (33). Using (30), (21) and (29), we obtain

$$\begin{aligned}\tilde{x}^+ &= \tilde{X}\left(e - \frac{\lambda}{\max(\tilde{X}^{-1}\tilde{d}(\tilde{x}))}\tilde{X}^{-1}\tilde{d}(\tilde{x})\right) = \tilde{X}\left(e - \frac{\lambda}{\max(\tilde{u}(\tilde{x}))}\tilde{u}(\tilde{x})\right)\\ &= \tilde{X}\left(e - \lambda\tilde{\theta}(\tilde{x})\frac{\tilde{u}(\tilde{x})}{\|\tilde{u}(\tilde{x})\|^2}\right).\end{aligned}$$

This relation together with (31) yield

$$\begin{aligned}r(\tilde{x}^+) = \frac{\tilde{x}^+}{\tilde{c}^{\mathrm{T}}\tilde{x}^+} &= \frac{\tilde{X}\left[e - (\lambda\tilde{\theta}(\tilde{x})/\|\tilde{u}(\tilde{x})\|^2)\,\tilde{u}(\tilde{x})\right]}{\tilde{c}^{\mathrm{T}}\tilde{x}(1-\lambda\tilde{\theta}(\tilde{x}))}\\ &= \frac{R(\tilde{x})\left[e - (\lambda\tilde{\theta}(\tilde{x})/\|\tilde{u}(\tilde{x})\|^2)\,\tilde{u}(\tilde{x})\right]}{1-\lambda\tilde{\theta}(\tilde{x})}\\ &= r(\tilde{x}) - \frac{\lambda\tilde{\theta}(\tilde{x})}{1-\lambda\tilde{\theta}(\tilde{x})}R(\tilde{x})\left(-e + \frac{\tilde{u}(\tilde{x})}{\|\tilde{u}(\tilde{x})\|^2}\right).\end{aligned}$$

where $R(\tilde{x}) = \mathrm{diag}(r(\tilde{x}))$. Combining the last relation with (26), we obtain (33). □

In the remaining part of this section, we let $\{\tilde{x}^k\}$ and $\{\lambda_k\}$ denote the sequence of iterates and stepsizes for the AS algorithm applied to problem (15) and define $\tilde{r}^k \equiv r(\tilde{x}^k) = \tilde{x}^k/\tilde{c}^{\mathrm{T}}\tilde{x}^k$ for all $k \geqslant 0$. When $\lambda_k = \lambda \in (\frac{2}{3}, 1)$ for all $k \geqslant 0$, the following result shows that $\{\tilde{x}^k\}$ can not converge to $\tilde{x} = 0$ along a unique direction of approach.

**Theorem 3.6.** *Assume that $\{\lambda_k\} \subseteq (\frac{2}{3}, 1)$ satisfies $\liminf_{k\to\infty} \lambda_k > \frac{2}{3}$ and that $\{\tilde{x}^k\} \subseteq H_{\geq}^{++}$. Then, the following statements hold*:

(a) $\lim_{k\to\infty} \tilde{c}^{\mathrm{T}}\tilde{x}^k = 0$ *(hence, if 0 is the unique optimal solution of (15) then* $\lim_{k\to\infty} \tilde{x}^k = 0$);

(b) $\{\tilde{r}^k\}$ *does not converge (i.e., $\{\tilde{r}^k\}$ has at least two accumulation points), or $\tilde{r}^k = r^*$ for all $k$ sufficiently large.*

**Proof.** Since, by assumption, $\{\tilde{x}^k\} \subseteq H_{\geq}^{++}$, we have $\tilde{c}^{\mathrm{T}}\tilde{x}^k > 0$ for all $k \geqslant 0$. Moreover, (32) and the assumption that $\lambda_k > \frac{2}{3}$ imply that $\liminf_{k\to\infty} \lambda_k\tilde{\theta}(\tilde{x}^k) > 0$. Since, by (31), we have $\tilde{c}^{\mathrm{T}}\tilde{x}^{k+1} = \tilde{c}^{\mathrm{T}}\tilde{x}^k(1 - \lambda_k\tilde{\theta}(x^k))$, we conclude that $\lim_{k\to\infty} \tilde{c}^{\mathrm{T}}\tilde{x}^k = 0$. We next show (b). If $\tilde{r}^k = r^*$ holds for some $k = k_0$, then, we see, in view of (33) and the fact that $\tilde{r}^k = r^*$ implies $\tilde{u}^k = e/|N|$ due to (26), that $\tilde{r}^k = r^*$ holds for all $k \geqslant k_0$. Now we deal with the case where $\tilde{r}^k \neq r^*$ for all $k$. Assume for contradiction that $\{\tilde{r}^k\}$ converges to some point $\tilde{r}^\infty$. It follows from relation (33) that

$$\tilde{r}^{k+1} = \tilde{r}^k - \frac{\lambda_k\tilde{\theta}(\tilde{x}^k)}{1-\lambda_k\tilde{\theta}(\tilde{x}^k)}\eta(\tilde{r}^k) \quad \forall k \geqslant 0. \tag{34}$$

This relation together with the fact that $\{\tilde{r}^k\}$ converges and $\liminf_{k\to\infty} \lambda_k\tilde{\theta}(\tilde{x}^k) > 0$ imply that

$$\lim_{k\to\infty} \eta(\tilde{r}^k) = 0. \tag{35}$$

By (26), we have

$$(\tilde{R}^k)^{-1}\eta(\tilde{r}^k) = -e + \frac{\tilde{u}^k}{\|\tilde{u}^k\|^2} \quad \forall k \geqslant 0, \tag{36}$$

where $\tilde{u}^k = \tilde{u}(x^k)$ and $\tilde{R}^k = \operatorname{diag}(\tilde{r}^k)$. We now consider two cases: 1. $\tilde{r}^\infty > 0$ and 2. $\tilde{r}^\infty \not> 0$, and show that both of them are not possible. Consider first case 1. In this case, it follows from (35) and (36) that

$$\lim_{k\to\infty} \frac{\tilde{u}^k}{\|\tilde{u}^k\|^2} = e, \tag{37}$$

and hence, in view of (29), we obtain

$$\lim_{k\to\infty} \tilde{\theta}(\tilde{x}^k)^{-1} = \lim_{k\to\infty} \max\left(\frac{\tilde{u}^k}{\|\tilde{u}^k\|^2}\right) = 1.$$

This relation together with the assumption that $\liminf_{k\to\infty} \lambda_k > \frac{2}{3}$ yield

$$\lim_{k\to\infty} \frac{\lambda_k \tilde{\theta}(\tilde{x}^k)}{1-\lambda_k\tilde{\theta}(\tilde{x}^k)} = \lim_{k\to\infty} \frac{\lambda_k}{1-\lambda_k} > 2.$$

Hence, in view of Lemma 3.3, we conclude that $\tilde{r}^k$ can not converge to a point $\tilde{r}^\infty > 0$. This shows that case 1 can not occur. We now consider case 2 in which $\tilde{r}^\infty \not> 0$. Let $Z = \{i : \tilde{r}_i^\infty = 0\}$. It is shown in Lemma 3.7 below that $\lim_{k\to\infty}[(\tilde{R}^k)^{-1}\eta(\tilde{r}^k)]_Z = -e$, and hence that $\eta_Z(\tilde{r}^k) > 0$ for every $k$ sufficiently large. This observation together with (34) clearly imply that $\tilde{r}_Z^{k+1} > \tilde{r}_Z^k$ for every $k$ sufficiently large, a conclusion that contradicts the fact that $0 = \tilde{r}_Z^\infty = \lim_{k\to\infty} \tilde{r}_Z^k$. □

Note that the assumption $\{\tilde{x}^k\} \subseteq H_>^{++}$ in Theorem 3.6 is automatically satisfied if $\tilde{x} = 0$ is the unique optimal solution of problem (15).

We next state and prove the result that was needed in proof of Theorem 3.6.

**Lemma 3.7.** *Let a sequence $\{r^k\} \subseteq H_1^{++}$ and an index set $Z \subsetneq \{1,\ldots,p\}$ be given such that $\lim_{k\to\infty} r_Z^k = 0$. Then, $\lim_{k\to\infty}[(R^k)^{-1}\eta(r^k)]_Z = -e$, where $R^k \equiv \operatorname{diag}(r^k)$.*

**Proof.** Let $\tilde{x}$ be an arbitrary point in $H_>^{++}$ and $\tilde{A}$ be a full row rank matrix such that $\operatorname{Null}(\tilde{A}) = H$. Then, it is easy to see that $\tilde{d}(\tilde{x}) = \tilde{X}^2\tilde{s}(\tilde{x})$, where $\tilde{s}(\tilde{x}) \equiv \tilde{c} - \tilde{A}^{\mathrm{T}}(\tilde{A}\tilde{X}^2\tilde{A}^{\mathrm{T}})^{-1}\tilde{A}\tilde{X}^2\tilde{c}$. Hence, we obtain

$$\tilde{u}(\tilde{x}) = \frac{\tilde{X}^{-1}\tilde{d}(\tilde{x})}{\tilde{c}^{\mathrm{T}}\tilde{x}} = \frac{\tilde{X}\tilde{s}(\tilde{x})}{\tilde{c}^{\mathrm{T}}\tilde{x}} = R(\tilde{x})\tilde{s}(\tilde{x}), \tag{38}$$

where $R(\tilde{x}) \equiv \operatorname{diag}(r(\tilde{x}))$. By Proposition 2.3, there exists a constant $L_0 > 0$ such that $\|\tilde{s}(\tilde{x})\| \leqslant L_0$ for all $\tilde{x} > 0$. This observation together with (38) imply

$$\|\tilde{u}_Z(\tilde{x})\| \leqslant L_0\|r_Z(\tilde{x})\| \quad \forall \tilde{x} \in H_>^{++}. \tag{39}$$

Let $\{\tilde{x}^k\}$ be a sequence such that $r(\tilde{x}^k) = r^k$ (for example, $\tilde{x}^k = r^k$ for every $k$). From (39) and the assumption that $\lim_{k\to\infty} r^k_Z = 0$, we conclude that $\lim_{k\to\infty} \tilde{u}_Z(\tilde{x}^k) = 0$. Using this observation together with (24) and (26), we obtain

$$\lim_{k\to\infty} [(R^k)^{-1}\eta(r^k)]_Z = \lim_{k\to\infty}\left(-e + \frac{\tilde{u}_Z(\tilde{x}^k)}{\|\tilde{u}(\tilde{x}^k)\|^2}\right) = -e. \qquad \square$$

We close this section with a result showing that the sequence $\{\tilde{r}^k\}$ converges quadratically to $r^*$ whenever $r^*$ exists and $\lambda_k = \lambda = \frac{1}{2}$ for all $k \geqslant 0$.

**Theorem 3.8.** *If $r^*$ exists and $\lambda_k = \frac{1}{2}$ for all $k \geqslant 0$ then $\{\tilde{r}^k\}$ converges quadratically to $r^*$.*

**Proof.** Assume that $r^*$ exists and that $\lambda_k = \frac{1}{2}$ for all $k \geqslant 0$. We first show that $\{\tilde{r}^k\}$ converges to $r^*$. Indeed, by Lemma 3.1 and the fact that $r^*$ exists, we conclude that $\tilde{x} = 0$ is the unique optimal solution of (15). By Proposition 2.5(a), it follows that $x^* \equiv \lim_{k\to\infty} \tilde{x}^k = 0$ and, and hence that $N_* \equiv N(x^*) = \{1,\dots,p\}$. Hence, if $\tilde{A}$ is a full row rank matrix such that $H = \mathrm{Null}(\tilde{A})$ then, by Proposition 2.5(d) with $(\bar{y},\bar{s}) = (0,\tilde{c})$, we conclude that

$$\lim_{k\to\infty} \tilde{r}^k = \lim_{k\to\infty} \frac{\tilde{x}^k}{\tilde{c}^{\mathrm{T}}\tilde{x}^k} = \operatorname{argmax}\left\{\sum_{j=1}^{p} \log \tilde{x}_j : \tilde{A}\tilde{x} = 0,\ \tilde{c}^{\mathrm{T}}\tilde{x} = 1,\ \tilde{x} > 0\right\} = r^*,$$

where the last equality follows from the definition of $r^*$. It remains to show that $\{\tilde{r}^k\}$ converges quadratically to $r^*$. In view of (34), quadratic convergence follows once we show that

$$\frac{\lambda_k\tilde{\theta}(\tilde{x}^k)}{1-\lambda_k\tilde{\theta}(\tilde{x}^k)} = 1 + \mathcal{O}(\|\eta(\tilde{r}^k)\|). \tag{40}$$

First observe that (29) and (36) imply that

$$\tilde{\theta}(\tilde{x}^k)^{-1} - 1 = \max\left(\frac{\tilde{u}^k}{\|\tilde{u}^k\|^2} - 1\right) = \max\left((\tilde{R}^k)^{-1}\eta(\tilde{r}^k)\right) = \mathcal{O}(\eta(\tilde{r}^k)).$$

Using the fact that $\lambda_k = \frac{1}{2}$ for all $k \geqslant 0$, it is now easy to see that the above relation implies (40). $\square$

It is easily seen that any other fixed stepsize $\lambda \in (0, \frac{2}{3})$ such that $\lambda \neq \frac{1}{2}$ would yield only linear convergence of the sequence $\{\tilde{r}^k\}$ to $r^*$. Hence, the above result shows that $\lambda = \frac{1}{2}$ is the best stepsize as far as the speed of convergence of the sequence $\{\tilde{r}^k\}$ to $r^*$ is concerned.

## 4. Technical results

In this section, we show that the relation between the conical projection of the AS sequence $\{\bar{x}^k\}$ for the homogeneous problem (15) and the Newton iterates for the analytic center problem (17) carries over to the context of general LP problems. The main idea is to approximate the original LP problem by a homogeneous LP problem near a constant-cost face in the sense that the AS directions at a feasible point $x$ for the two problems approach each other as $x$ approaches the constant-cost face. We can then apply the techniques developed in the previous section to the approximate homogeneous problem and thereby obtain conclusions about the AS sequence $\{x^k\}$ for the original LP problem. The results of this section are rather technical but they form the basis for the development of the superlinearly convergent algorithm of Section 5.

Associated with a given constant-cost partition $(N, B)$, there is a homogeneous LP problem defined in the $x_N$-space. Near the face $\mathcal{P}_N^+$, the AS direction associated with this homogeneous problem provides a good approximation of $d_N(x)$ as we will see below in Lemma 4.3. To motivate and introduce the homogeneous LP problem associated with $(N, B)$, consider the following problem

$$\begin{array}{ll} \text{minimize}_x & c_N^{\mathrm{T}} x_N + c_B^{\mathrm{T}} x_B \\ \text{subject to} & A_N x_N + A_B x_B = b, \quad x_N \geqslant 0, \end{array} \tag{41}$$

obtained by removing the constraint $x_B \geqslant 0$ from (1). The homogeneous LP problem is obtained by eliminating $x_B$ from the above problem as follows. Let $(\bar{y}, \bar{s}) \in \mathcal{D}_B$ be given and note that $b \in \text{Range}(A_B)$ since $\mathcal{P}_N^+ \neq \emptyset$. Due to Lemma 2.1(b), problem (41) can be written as

$$\begin{array}{ll} \text{minimize}_x & \bar{s}_N^{\mathrm{T}} x_N \\ \text{subject to} & A_N x_N \in \text{Range}(A_B), \quad x_N \geqslant 0, \end{array} \tag{42}$$

which is the homogeneous problem associated with $(N, B)$. This problem can be identified as a problem of the form (15) by letting $\tilde{c} = \bar{s}_N$, $\tilde{x} = x_N$ and $H = \mathcal{H}_N \equiv \{\tilde{x} \in \mathbb{R}^{|N|} : A_N \tilde{x} \in \text{Range}(A_B)\}$. The corresponding problems (16), (17) and (18) in this case become:

$$\begin{array}{ll} \text{maximize}_{\tilde{d}_N} & \bar{s}_N^{\mathrm{T}} \tilde{d}_N - \frac{1}{2} \|X_N^{-1} \tilde{d}_N\|^2 \\ \text{subject to} & A_N \tilde{d}_N \in \text{Range}(A_B), \end{array} \tag{43}$$

$$\begin{array}{ll} \text{maximize}_{r_N} & \sum_{i \in N} \log r_i \\ \text{subject to} & A_N r_N \in \text{Range}(A_B) \\ & \bar{s}_N^{\mathrm{T}} r_N = 1, \quad r_N > 0, \end{array} \tag{44}$$

and

$$\begin{array}{ll} \text{maximize}_{\eta_N} & (r_N^{-1})^{\mathrm{T}}(-\eta_N) - \frac{1}{2}\eta_N R_N^{-2}\eta_N \\ \text{subject to} & A_N\eta_N \in \operatorname{Range}(A_B) \\ & \bar{s}_N^{\mathrm{T}}\eta_N = 0, \quad \eta_N > 0, \end{array} \tag{45}$$

respectively, where $R_N \equiv \operatorname{diag}(r_N)$.

We now introduce the notation needed for the development in this and the next section. Unless otherwise specified, $(N, B)$ denotes a constant-cost partition of (1). Let

$$\mathcal{Q}_N^{++} \equiv \{x \mid x \in \mathcal{P}^{++},\ c^{\mathrm{T}}x - \nu_N > 0\}, \tag{46}$$

and, given $x \in \mathcal{Q}_N^{++}$, define

$$u(x) \equiv \frac{X^{-1}d(x)}{c^{\mathrm{T}}x - \nu_N} = \frac{X^{-1}d(x)}{\bar{s}_N^{\mathrm{T}}x_N}, \tag{47}$$

$$r_N(x) \equiv \frac{x_N}{c^{\mathrm{T}}x - \nu_N} = \frac{x_N}{\bar{s}_N^{\mathrm{T}}x_N}, \tag{48}$$

$$\tilde{u}_N(x) \equiv \frac{X_N^{-1}\tilde{d}_N(x)}{c^{\mathrm{T}}x - \nu_N} = \frac{X_N^{-1}\tilde{d}_N(x)}{\bar{s}_N^{\mathrm{T}}x_N}, \tag{49}$$

where $\tilde{d}_N(x)$ denotes the (homogeneous) AS direction of problem (42) at $x_N$, that is, the optimal solution of (43). (Note that $\tilde{d}_N(x)$ and $\tilde{u}_N(x)$ are really functions of $x_N$ but for simplicity of notation we view them as a function of $x$. Note also that the vector $u(x)$ depends on $(N, B)$ but this dependence is ignored for simplicity of notation.) Given $x \in \mathcal{Q}_N^{++}$, we let $\eta_N(x)$ denote the Newton direction associated with (44) at the point $r_N = r_N(x)$, that is, the optimal solution of (45). (It would be more accurate to view $\eta_N(\cdot)$ as a function of $r_N$ but for simplicity of notation, we view it as a function of $x$.) Clearly, it follows that $\eta(x) = \eta(r_N(x))$ for every $x \in \mathcal{Q}_N^{++}$. Finally, given $x \in \mathcal{P}^{++}$ and $\lambda > 0$, we let

$$x^+(\lambda) \equiv x - \frac{\lambda}{\max(X^{-1}d(x))}d(x). \tag{50}$$

The following result provides a preliminary relation between the AS direction $d(x)$ and the homogeneous AS direction $\tilde{d}_N(x)$.

**Lemma 4.1.** *For every $x \in \mathcal{P}^{++}$, the vector $(\delta_N, \delta_B) \equiv (d_N(x) - \tilde{d}_N(x), d_B(x))$ is the (unique) optimal solution of the following QP problem:*

$$\begin{array}{ll} \text{minimize}_{(\delta_N,\delta_B)} & \frac{1}{2}\|X_N^{-1}\delta_N\|^2 + \frac{1}{2}\|X_B^{-1}\delta_B\|^2 \\ \text{subject to} & A_N\delta_N + A_B\delta_B = -A_N\tilde{d}_N(x). \end{array} \tag{51}$$

**Proof.** The vector $(d_N(x) - \tilde{d}_N(x), d_B(x))$ is clearly feasible for problem (51). To prove that $(d_N(x) - \tilde{d}_N(x), d_B(x))$ is optimal for (51), it is sufficient to show that

$$\begin{pmatrix} X_N^{-2}(d_N(x) - \tilde{d}_N(x)) \\ X_B^{-2}d_B(x) \end{pmatrix} \in \operatorname{Range}(A^{\mathrm{T}}). \tag{52}$$

Fix some $(\bar{y},\bar{s}) \in \mathcal{D}_B \neq \emptyset$. Since $\bar{s}_B = 0$, $\tilde{d}_N(x)$ solves (43) and, by Lemma 2.2(a), $d(x)$ solves (11), we have

$$\begin{pmatrix} \bar{s}_N - X_N^{-2}\tilde{d}_N(x) \\ 0 \end{pmatrix} \in \mathrm{Range}(A^{\mathrm{T}}), \tag{53}$$

$$\begin{pmatrix} \bar{s}_N - X_N^{-2}d_N(x) \\ -X_B^{-2}d_B(x) \end{pmatrix} \in \mathrm{Range}(A^{\mathrm{T}}). \tag{54}$$

Combining (53) and (54), we obtain (52). □

The following technical lemma is well-known and is used in the proof of next result.

**Lemma 4.2.** *Let $F \in \mathbb{R}^{p\times q}$ be given. Then, there exists a constant $C_1 = C_1(F)$ with the following property: for any $f \in \mathbb{R}^p$ such that the system $Fw = f$ is feasible and any $z \in \mathbb{R}^q$, there exists a solution $\bar{w}$ of $Fw = f$ such that*

$$\|\bar{w} - z\| \leqslant C_1\|f - Fz\|.$$

**Lemma 4.3.** *The following statements hold:*

(a) $\|u_B(x)\| = \mathcal{O}(\|X_B^{-1}\|\,\|X_N\|\,\|u_N(x)\|)$ *for all* $x \in \mathcal{Q}_N^{++}$;

(b) $\|\tilde{u}_N(x) - u_N(x)\| = O(\|X_B^{-1}\|^2\|X_N\|^2\|u_N(x)\|)$ *for all* $x \in \mathcal{Q}_N^{++}$ *such that* $\|X_B^{-1}\|\,\|X_N\|$ *is sufficiently small.*

**Proof.** The proof of (a) follows immediately from (47) and Lemma 2.2(b). We next show (b). Fix $x > 0$. Since $A_N\tilde{d}_N(x) \in \mathrm{Range}(A_B)$ and $A_Bd_B(x) = -A_Nd_N(x)$, it follows from Lemma 4.2 that there exists $\tilde{d}_B(x)$ such that

$$A_B\tilde{d}_B(x) = -A_N\tilde{d}_N(x), \quad \|\tilde{d}_B(x) - d_B(x)\| \leqslant C_2\|\tilde{d}_N(x) - d_N(x)\|, \tag{55}$$

where $C_2$ is a constant independent of $x$. Using the second relation in (55), we obtain

$$\begin{aligned}\|X_B^{-1}(d_B(x) - \tilde{d}_B(x))\| &\leqslant \|X_B^{-1}\|\,\|d_B(x) - \tilde{d}_B(x)\| \\ &\leqslant C_2\|X_B^{-1}\|\,\|d_N(x) - \tilde{d}_N(x)\| \\ &\leqslant C_2\|X_B^{-1}\|\,\|X_N\|\,\|X_N^{-1}d_N(x) - X_N^{-1}\tilde{d}_N(x)\|. \end{aligned} \tag{56}$$

The first relation in (55) implies that $(\delta_N,\delta_B) = (0,\tilde{d}_B(x))$ is feasible to problem (51), and hence, by Lemma 4.1, we obtain

$$\|X_B^{-1}d_B(x)\|^2 + \|X_N^{-1}(d_N(x) - \tilde{d}_N(x))\|^2 \leqslant \|X_B^{-1}\tilde{d}_B(x)\|^2. \tag{57}$$

Thus, we obtain

$$\begin{aligned}\|X_N^{-1}(d_N(x) - \tilde{d}_N(x))\|^2 &\leqslant \|X_B^{-1}\tilde{d}_B(x)\|^2 - \|X_B^{-1}d_B(x)\|^2 \\ &= [X_B^{-1}(\tilde{d}_B(x) + d_B(x))]^{\mathrm{T}}[X_B^{-1}(\tilde{d}_B(x) - d_B(x))] \\ &\leqslant \|X_B^{-1}(d_B(x) + \tilde{d}_B(x))\|\,\|X_B^{-1}(\tilde{d}_B(x) - d_B(x))\|. \end{aligned} \tag{58}$$

Combining this last relation with (56), we obtain

$$\begin{aligned}\|X_N^{-1}(d_N(x) - \tilde{d}_N(x))\| &\leqslant C_2\|X_B^{-1}\|\,\|X_N\|\,\|X_B^{-1}(d_B(x) + \tilde{d}_B(x))\| \\ &\leqslant C_2\|X_B^{-1}\|\,\|X_N\|\,\left\{2\|X_B^{-1}d_B(x)\| + \|X_B^{-1}(\tilde{d}_B(x) - d_B(x))\|\right\} \\ &\leqslant 2C_2\|X_B^{-1}\|\,\|X_N\|\,\|X_B^{-1}d_B(x)\| \\ &\quad + C_2^2\|X_B^{-1}\|^2\,\|X_N\|^2\|X_N^{-1}d_N(x) - X_N^{-1}\tilde{d}_N(x)\|,\end{aligned}$$

from which we conclude that

$$\|X_N^{-1}(d_N(x) - \tilde{d}_N(x))\| \leqslant 4C_2\|X_B^{-1}\|\,\|X_N\|\,\|X_B^{-1}d_B(x)\|,$$

whenever $C_2^2\|X_B^{-1}\|^2\,\|X_N\|^2 \leqslant \frac{1}{2}$. The last relation together with (a) imply

$$\|X_N^{-1}(d_N(x) - \tilde{d}_N(x))\| = \mathcal{O}\Big(\|X_B^{-1}\|^2\,\|X_N\|^2\,\|X_N^{-1}d_N(x)\|\Big).$$

After dividing both sides of this relation by $c^{\mathrm{T}}x - \nu_N$, we obtain the desired result. □

**Lemma 4.4.** *The following relations hold:*

$$e^{\mathrm{T}}\tilde{u}_N(x) = 1 \quad \forall x \in \mathcal{Q}_N^{++}, \tag{59}$$

*and*

$$|e^{\mathrm{T}}u_N(x) - 1| = \mathcal{O}(\|X_B^{-1}\|^2\|X_N\|^2\|u_N(x)\|), \tag{60}$$

*for every $x \in \mathcal{Q}_N^{++}$ with $\|X_B^{-1}\|\,\|X_N\|$ sufficiently small.*

**Proof.** Relation (59) is an immediate consequence of (24). Using (59) and Lemma 4.3, we obtain

$$\begin{aligned}|e^{\mathrm{T}}u_N(x) - 1| &= |e^{\mathrm{T}}(u_N(x) - \tilde{u}_N(x))| \\ &\leqslant \|e\|\,\|u_N(x) - \tilde{u}_N(x)\| \\ &= \mathcal{O}(\|X_B^{-1}\|^2\|X_N\|^2\|u_N(x)\|),\end{aligned}$$

for every $x \in \mathcal{Q}_N^{++}$ with $\|X_B^{-1}\|\,\|X_N\|$ sufficiently small. □

**Lemma 4.5.** *The following relations hold:*

$$c^{\mathrm{T}}x^+(\lambda) - \nu_N = (1 - \lambda\theta(x))(c^{\mathrm{T}}x - \nu_N) \quad \forall x \in \mathcal{P}^{++}, \tag{61}$$

*and*

$$r_N(x^+(\lambda)) = r_N(x) - \frac{\lambda\theta(x)}{1 - \lambda\theta(x)}R_N(x)\left(-e + \frac{u_N(x)}{\|u(x)\|^2}\right), \tag{62}$$

*for every $x \in \mathcal{Q}_N^{++}$ and $\lambda > 0$ such that $x^+(\lambda) \in \mathcal{Q}_N^{++}$, where $R_N(x) = \mathrm{diag}(r_N(x))$ and*

$$\theta(x) \equiv \frac{\|u(x)\|^2}{\max(u(x))}. \tag{63}$$

**Proof.** The proof of (61) is similar to the proof of (31) and uses the fact that $c^{\mathrm{T}}d(x) = \|X^{-1}d(x)\|^2$. Also, the proof of (62) follows along the same line as the proof of (33). We omit the details. □

The next lemma is the main result of this section. It generalizes relation (33) of Lemma 3.5 to the context of general LP problems.

**Lemma 4.6.** *We have*:

$$r_N(x^+(\lambda)) = r_N(x) - \frac{\lambda\theta(x)}{1-\lambda\theta(x)}(\eta_N(x) + R_N(x)h_N(x)), \tag{64}$$

*where*

$$h_N(x) \equiv \left(-e + \frac{u_N(x)}{\|u(x)\|^2} - R_N^{-1}(x)\eta_N(x)\right) = \mathcal{O}(\|X_B^{-1}\|^2\|X_N\|^2), \tag{65}$$

*for every* $x \in \mathcal{Q}_N^{++}$ *such that* $\|X_B^{-1}\|\,\|X_N\|$ *is sufficiently small.*

**Proof.** First note that (59) and Lemma 4.3 imply that

$$\|\tilde{u}_N(x)\| \geqslant \frac{1}{\sqrt{|N|}}, \qquad \|u_N(x)\| \geqslant \frac{\|\tilde{u}_N(x)\|}{2} \geqslant \frac{1}{2\sqrt{|N|}}, \tag{66}$$

for every $x \in \mathcal{Q}_N^{++}$ with $\|X_B^{-1}\|\,\|X_N\|$ sufficiently small. Due to relation (26), we have

$$R_N^{-1}(x)\eta_N(x) = R_N^{-1}(x)\eta_N(r_N(x)) = -e + \frac{\tilde{u}_N(x)}{\|\tilde{u}_N(x)\|^2}, \tag{67}$$

and hence,

$$\begin{aligned} h_N(x) &= \frac{u_N(x)}{\|u(x)\|^2} - \frac{\tilde{u}_N(x)}{\|\tilde{u}_N(x)\|^2} \\ &= \frac{\tilde{u}_N(x)}{\|\tilde{u}_N(x)\|^2}\left(\frac{\|\tilde{u}_N(x)\|^2}{\|u(x)\|^2} - 1\right) + \frac{u_N(x) - \tilde{u}_N(x)}{\|u(x)\|^2}. \end{aligned}$$

Thus, we obtain

$$\begin{aligned} \|h_N(x)\| &\leqslant \frac{|\,\|\tilde{u}_N(x)\|^2 - \|u(x)\|^2|}{\|\tilde{u}_N(x)\|\,\|u(x)\|^2} + \frac{\|u_N(x) - \tilde{u}_N(x)\|}{\|u(x)\|^2} \\ &\leqslant \frac{\|u_B(x)\|^2 + \|\tilde{u}_N(x) - u_N(x)\|\,[\,\|\tilde{u}_N(x)\| + \|u_N(x)\|\,]}{\|\tilde{u}_N(x)\|\,\|u_N(x)\|^2} \\ &\quad + \frac{\|u_N(x) - \tilde{u}_N(x)\|}{\|u_N(x)\|^2} \\ &= \frac{\|u_B(x)\|^2}{\|\tilde{u}_N(x)\|\,\|u_N(x)\|^2} + \frac{\|\tilde{u}_N(x) - u_N(x)\|}{\|\tilde{u}_N(x)\|\,\|u_N(x)\|^2} + 2\frac{\|u_N(x) - \tilde{u}_N(x)\|}{\|u_N(x)\|^2}. \end{aligned} \tag{68}$$

Now, using (66) and Lemma 4.3, it is easy to see that each one of the terms in the right hand side of the above expression is $\mathcal{O}(\|X_B^{-1}\|^2\|X_N\|^2)$ for every $x \in \mathcal{Q}_N^{++}$ with

$\|X_B^{-1}\|\,\|X_N\|$ sufficiently small. Hence, (65) follows. Relation (64) is an immediate consequence of (62) and (65). □

A natural question to be asked is: for which constant-cost partitions $(N,B)$ does problem (44) have an optimal solution? The following result shows that the optimal partition is the only one. Recall that $(\bar{y}^a, \bar{s}^a)$ denotes the analytic center of the dual optimal face, that is, the point defined in (13).

**Lemma 4.7.** *Let $(N,B)$ be a constant-cost partition. Then, problem (44) has an optimal solution if and only if $(N,B)$ is the optimal partition of (1), in which case $(\bar{s}_N^a)^{-1}/|N|$ is the (unique) optimal solution of (44).*

**Proof.** If $r_N^*$ is the optimal solution of (44) then by considering the optimality conditions of (44), we can easily show that

$$\begin{pmatrix} (r_N^*)^{-1}/|N| \\ 0 \end{pmatrix} \in \mathrm{Range}(A^{\mathrm{T}}) + \mathbb{R}c. \tag{69}$$

Hence, $\mathcal{D}_B^{++} \neq \emptyset$. Due to the assumption that $(N,B)$ is a constant-cost partition, we have $\mathcal{P}_N^{++} \neq \emptyset$. Hence, we conclude that $(N,B)$ is the optimal partition of (1). Conversely, by considering the optimality conditions of (13), we can easily verify that $(\bar{s}_N^a)^{-1}/|N|$, where $(\bar{s}_N^a)^{-1} \equiv [\mathrm{diag}(\bar{s}_N^a)]^{-1}e$, satisfies the optimality conditions of (44) with $\bar{s}_N = \bar{s}_N^a$. We omit the details of the proof. □

**Lemma 4.8.** *Let $(N,B)$ be a constant-cost partition and assume that $\{\alpha_k\} \subseteq \mathbb{R}^{++}$, $\{\tilde{x}^k\} \subseteq \mathcal{P}^{++}$ and $\{(\tilde{y}^k, \tilde{s}^k)\} \subseteq \mathcal{D}$ are sequences satisfying the following conditions:*

(a) *$\{\tilde{x}_N^k/\alpha_k\}$ is bounded;*

(b) *$\lim_{k\to\infty}(\tilde{X}_N^k \tilde{s}_N^k)/\alpha_k = a_N > 0$, where $\tilde{X}_N^k = \mathrm{diag}(\tilde{x}_N^k)$ and $a_N$ is some $|N|$-dimensional vector;*

(c) *$\lim_{k\to\infty} \tilde{s}_B^k = 0$ and $\{\tilde{s}_N^k\}$ is bounded.*

*Then, $(N,B)$ is the optimal partition of (1) and we have:*

$$\lim_{k\to\infty} (\tilde{y}^k, \tilde{s}^k) = (\bar{y}, \bar{s}), \tag{70}$$

$$\lim_{k\to\infty} x_N^k/\alpha_k = (\bar{S}_N)^{-1} a_N, \tag{71}$$

*where $\bar{S}_N = \mathrm{diag}(\bar{s}_N)$ and*

$$(\bar{y}, \bar{s}) \equiv \mathrm{argmax}\Big\{ \sum_{j\in N} a_j \log s_j : (y,s) \in \mathcal{D}_B^{++} \Big\}. \tag{72}$$

*In particular, if $a_N$ is a positive multiple of the vector of all ones then $(\bar{y}, \bar{s})$ is equal to the analytic center $(\bar{y}^a, \bar{s}^a)$ defined in (13).*

**Proof.** Since $(N,B)$ defines a constant-cost face, we have $\mathcal{P}_N^{++} \neq \emptyset$. Hence, to show that $(N,B)$ is the optimal partition of (1) and that (70) holds, it is sufficient to show that any accumulation point $(\bar{y}, \bar{s})$ of $(\tilde{y}^k, \tilde{s}^k)$ satisfies the optimality condition for (72), namely

$$(\bar{y}, \bar{s}) \in \mathcal{D}_B^{++}, \tag{73}$$

$$A_N \bar{S}_N^{-1} a_N \in \text{Range}(A_B), \tag{74}$$

where $\bar{S}_N = \text{diag}(\bar{s}_N)$. Indeed, let $\mathcal{K}$ be an infinite index set such that $\lim_{k\in\mathcal{K}}(\tilde{y}^k, \tilde{s}^k) = (\bar{y}, \bar{s})$. Using (c) and the assumption that $\{(\tilde{y}^k, \tilde{s}^k)\} \subset \mathcal{D}$, we conclude $(\bar{y}, \bar{s}) \in \mathcal{D}_B$. Moreover, (a) and (b) imply that $\bar{s}_N > 0$ and that

$$\lim_{k\in\mathcal{K}} \frac{\tilde{x}_N^k}{\alpha_k} = \bar{S}_N^{-1} a_N. \tag{75}$$

Thus, we conclude that (73) holds. Since $\mathcal{P}_N^{++} \neq \emptyset$, we have $b \in \text{Range}(A_B)$. This observation together with the fact that $\{\tilde{x}^k\} \subseteq \mathcal{P}^{++}$ imply

$$A_N \left( \frac{\tilde{x}_N^k}{\alpha_k} \right) \in \text{Range}(A_B). \tag{76}$$

Relation (74) now follows immediately from (75) and (76). The limit (71) is an immediate consequence of (70) and (b). □

**Lemma 4.9.** *Let $(N, B)$ be a constant-cost partition and let $\{\tilde{x}^k\} \subseteq \mathcal{Q}_N^{++}$ be a sequence such that $\{r_N(\tilde{x}^k)\}$ is bounded and $\lim_{k\to\infty} \|(\tilde{X}_B^k)^{-1}\| \, \|\tilde{X}_N^k\| = 0$. Then the following statements are equivalent:*

(a) $\lim_{k\to\infty} u_N(\tilde{x}^k) = e/|N|$;

(b) $\lim_{k\to\infty} \|(R_N^k)^{-1}\eta_N(\tilde{x}^k)\| = 0$, *where* $R_N^k \equiv \text{diag}(r_N(\tilde{x}^k))$;

(c) $(N, B)$ *is the optimal partition of* (1) *and* $\lim_{k\to\infty} r_N(\tilde{x}^k) = (\bar{s}_N^a)^{-1}/|N|$,

*in which case,* $\lim_{k\to\infty} (y(\tilde{x}^k), s(\tilde{x}^k)) = (\bar{y}^a, \bar{s}^a)$.

**Proof.** From Lemma 4.3 and the assumption that $\lim_{k\to\infty} \|(\tilde{X}_B^k)^{-1}\| \, \|\tilde{X}_N^k\| = 0$, it follows that

$$\lim_{k\to\infty} u_N(\tilde{x}^k) = e/|N| \iff \lim_{k\to\infty} \tilde{u}_N(\tilde{x}^k) = e/|N|. \tag{77}$$

Due to relations (26) and (59), we have

$$\begin{aligned} \|(R_N^k)^{-1}\eta_N(\tilde{x}^k)\|^2 &= \left\| -e + \frac{\tilde{u}(\tilde{x}^k)}{\|\tilde{u}_N(\tilde{x}^k)\|^2} \right\|^2 = |N| - 2\frac{e^{\mathrm{T}}\tilde{u}_N(\tilde{x}^k)}{\|\tilde{u}_N(\tilde{x}^k)\|^2} + \frac{1}{\|\tilde{u}_N(\tilde{x}^k)\|^2} \\ &= |N| - \frac{1}{\|\tilde{u}_N(\tilde{x}^k)\|^2}. \end{aligned} \tag{78}$$

Using the fact that $e^{\mathrm{T}}\tilde{u}_N(\tilde{x}^k) = 1$, it is easy to see that $\lim_{k\to\infty} \|\tilde{u}_N(\tilde{x}^k)\| = 1/\sqrt{|N|}$ if and only if $\lim_{k\to\infty} \tilde{u}_N(\tilde{x}^k) = e/|N|$. This observation together with (77) and (78) immediately imply the equivalence of (a) and (b). The proof of the implication (c) ⇒ (b) follows immediately from (a) and (b) of Lemma 3.2 and Lemma 4.7. It remains to show the implication (a) ⇒ (c). Indeed, assume that (a) holds and let $\alpha_k \equiv c^{\mathrm{T}}\tilde{x}^k - \nu_N$

and $(\tilde{y}^k, \tilde{s}^k) \equiv (y(\tilde{x}^k), s(\tilde{x}^k))$ for all $k \geqslant 0$. We will show that the sequences $\{\tilde{x}^k\}$, $\{(\tilde{y}^k, \tilde{s}^k)\}$ and $\{\alpha_k\}$ satisfy conditions (a), (b) and (c) of Lemma 4.8 with $a_N = e/|N|$. Since by assumption $\{r_N(\tilde{x}^k)\}$ is bounded and $r_N(\tilde{x}^k) = \tilde{x}_N^k/\alpha_k$ for every $k$, condition (a) of Lemma 4.8 holds. Condition (a) implies that (b) of Lemma 4.8 with $a_N = e/|N|$ is satisfied since

$$u(\tilde{x}^k) = \frac{\tilde{X}^k s(\tilde{x}^k)}{c^{\mathrm{T}}\tilde{x}^k - \nu_N} = \frac{\tilde{X}^k \tilde{s}^k}{\alpha_k}. \tag{79}$$

Clearly, $\{(\tilde{y}^k, \tilde{s}^k)\}$ is bounded in view of Proposition 2.3. Hence, to show that condition (c) of Lemma 4.8 holds, it is sufficient to prove that $\lim_{k\to\infty} \tilde{s}_B^k = 0$. Indeed, first observe that (a), the assumption that $\lim_{k\to\infty} \|(\tilde{X}_B^k)^{-1}\| \, \|\tilde{X}_N^k\| = 0$ and Lemma 4.3(a) imply that $\{u(\tilde{x}^k)\}$ is bounded. Using this observation, the fact that $\alpha_k = \mathcal{O}(\|\tilde{X}_N^k\|)$ and (79), we obtain

$$\|\tilde{s}_B^k\| \leqslant \|(\tilde{X}_B^k)^{-1}\| \, \|\tilde{X}^k \tilde{s}^k\| = \alpha_k \|(\tilde{X}_B^k)^{-1}\| \, \|u(\tilde{x}^k)\| \leqslant \mathcal{O}(\|(\tilde{X}_B^k)^{-1}\| \, \|\tilde{X}_N^k\|),$$

which, together with the assumption that $\lim_{k\to\infty} \|(\tilde{X}_B^k)^{-1}\| \, \|\tilde{X}_N^k\| = 0$, clearly imply that $\lim_{k\to\infty} \tilde{s}_B^k = 0$. Using Lemma 4.8, we conclude that (c) holds and $\lim_{k\to\infty}(y(\tilde{x}^k), s(\tilde{x}^k)) = (\bar{y}^a, \bar{s}^a)$. □

We observe that it is possible to give a direct proof of the implication (b) ⇒ (c) by using Lemma 3.7, Lemma 3.2(b) and Lemma 4.7. The proof given above shows instead the implication (a) ⇒ (c) via Lemma 4.8, which is simpler in the sense that it does not need the machinery introduced in the Section 2 and in the first part of this section. It also illustrates a basic principle that has been used in the convergence analysis of the AS algorithm (see Tsuchiya and Muramatsu [29] or Monteiro et al. [19, Theorem 4.3]).

## 5. A superlinearly convergent affine scaling algorithm

In this section we present a variant of Algorithm 1 which is globally and two-step superlinearly convergent. After we state the algorithm, its global convergence and superlinear convergence are proved.

To describe the variant of Algorithm 1 that will be studied in this section, we assume that two constants $p$ and $q$ are given such that

$$p, q \in (0,1), \qquad p < \frac{q}{q+2}. \tag{80}$$

Examples of constants satisfying these conditions are: $p = 0.3$ and $q = 0.95$. Observe that $p$ can be chosen as close as to $\frac{1}{3}$ as it is desired. For the purpose of future reference, we note that (80) implies that

$$\frac{2(q-p)}{1+p} > q. \tag{81}$$

The following variant of Algorithm 1 will be shown later to converge two-step superlinearly with order at least $1+p<\frac{4}{3}$.

**Algorithm SLA**

*Step* 0. Assume that constants $p$ and $q$ satisfying (80) and a point $x^0 \in \mathcal{P}^{++}$ are given. Set $k:=0$.

*Step* 1. Compute $d^k \equiv d(x^k)$ according to (9c) and let

$$N_k = \{i : x_i^k \leqslant |\, e^{\mathrm{T}}[(X^k)^{-1}d^k]\,|^{1/2}\}, \tag{82a}$$

$$\sigma_k = e^{\mathrm{T}}\big((X_{N_k}^k)^{-1}d_{N_k}^k\big), \tag{82b}$$

$$\epsilon_k = \sqrt{|N_k| - \frac{\sigma_k^2}{\|(X^k)^{-1}d^k\|^2}} = \sqrt{|N_k| - \frac{(e^{\mathrm{T}}u_{N_k}(x^k))^2}{\|u(x^k)\|^2}}. \tag{82c}$$

*Step* 2. If

$$\epsilon_k < \sigma_k^q, \tag{83}$$

then (Predictor step)

$$\lambda_k = \max(0.5, 1-\sigma_k^p) \tag{84}$$

else (Corrector step)

$$\lambda_k = 0.5.$$

*Step* 3. $$x^{k+1} = x^k - \lambda_k \frac{d^k}{\max((X^k)^{-1}d^k)}. \tag{85}$$

*Step* 4. $k := k+1$ and return to Step 1.

The first expression for $\epsilon_k$ is the one that should be used to compute it. The second one is used during the analysis of the algorithm and is a consequence of (47). It is easy to see that term within the square root of the first or second expression for $\epsilon_k$ is nonnegative so that $\epsilon_k$ is well-defined.

The basic procedure is to alternate the choice of the stepsize between $\lambda_k = 0.5$ and $\lambda_k \sim 1$. Since Algorithm SLA is a variant of Algorithm 1 in which $\lambda_k \geqslant \frac{1}{2}$ for all $k \geqslant 0$, we conclude that it satisfies all the statements (a)–(e) of Proposition 2.4. As in Section 2, we denote the limit point of the sequence $\{x^k\}$ by $x^*$ and let $(N_*, B_*) \equiv (N(x^*), B(x^*))$. By Proposition 2.4(d), $(N_*, B_*)$ is a constant-cost partition. Recall that the constant value of $c^{\mathrm{T}}x$ over the face $\mathcal{P}_{N_*}^+$ is denoted by $\nu_{N_*}$. Clearly, $\nu_{N_*} = c^{\mathrm{T}}x^*$. Throughout this section, the function $u(\cdot)$ refers to the one associated with the partition $(N_*, B_*)$ and the following notation is used: $u^k = u(x^k)$, $\tilde{u}_{N_*}^k = \tilde{u}_{N_*}(x^k)$, $r_{N_*}^k = r_{N_*}(x^k)$, $R_{N_*}^k = \mathrm{diag}(r_{N_*}^k)$, $\eta_{N_*}^k = \eta_{N_*}(x^k)$, $(y^k, s^k) = (y(x^k), s(x^k))$, for all $k \geqslant 0$.

The global convergence analysis of Algorithm SLA is much simpler than its superlinear convergence analysis and is obtained in Theorem 5.3. So we next explain the underlying idea behind the superlinear convergence analysis of Algorithm SLA. It is shown in Proposition 5.1 that $\sigma_k \sim c^{\mathrm{T}}x^k - \nu_{N_*}$ and $\|x^k_{N_*}\| = \mathcal{O}(c^{\mathrm{T}}x^k - \nu_{N_*})$ from which it is easy to conclude that $N_k = N_{\mathrm{opt}}$ for all $k$ sufficiently large, where $(N_{\mathrm{opt}}, B_{\mathrm{opt}})$ denotes the optimal partition of (1). Moreover, Lemma 5.2 shows that $\epsilon_k$ is a measure of centrality for the "small" variables $x^k_{N_*}$, the ones that dictate the speed of convergence of the (or, any interior point) algorithm. When the measure of centrality $\epsilon_k$ is small, a predictor step with stepsize $\lambda_k$ asymptotically approaching 1 is taken. The behavior of the predictor steps is analyzed in Lemma 5.5; the main conclusion is that the measure of progress $c^{\mathrm{T}}x - \nu_{N_*}$ is reduced at a superlinear rate while the centrality measure "slowly" deteriorates. At the next step, if the the small variables are not well-centered (i.e., the test (83) fails), then a corrector step is taken with stepsize $\lambda_k = \frac{1}{2}$. The effect of this step is analyzed in Lemma 5.6; the main conclusion is that $c^{\mathrm{T}}x - \nu_{N_*}$ is reduced at a linear rate while the centrality measure is improved at a quadratic rate. Lemma 5.7 shows that, asymptotically, one corrector step suffices to recover the centrality of the small variables and hence that a predictor step is taken in every two steps of Algorithm SLA. Using these conclusions, it is now easy to prove the superlinear convergence of Algorithm SLA (see Theorem 5.10).

Some basic properties of Algorithm SLA which follows almost immediately from the analysis of Section 4 are given in the following result.

**Lemma 5.1.** *The following statements hold*:

(a) *the sequences* $\{u^k\}$, $\{\tilde{u}^k_{N_*}\}$ *and* $\{r^k_{N_*}\}$ *are bounded*;

(b) $\{\|(X^k_{B_*})^{-1}\|\,\|X^k_{N_*}\|\}$, $\{\|u^k_{N_*} - \tilde{u}^k_{N_*}\|\}$, $\{\|u^k_{B_*}\|\}$, $\{e^{\mathrm{T}}u^k_{N_*} - 1\}$ *and* $\{\|s^k_{B_*}\|\}$ *converge to* 0 *according to*:

$$\|(X^k_{B_*})^{-1}\|\,\|X^k_{N_*}\| = \mathcal{O}\big(c^{\mathrm{T}}x^k - \nu_{N_*}\big), \tag{86}$$

$$\|u^k_{N_*} - \tilde{u}^k_{N_*}\| = \mathcal{O}\big((c^{\mathrm{T}}x^k - \nu_{N_*})^2\big), \tag{87}$$

$$|e^{\mathrm{T}}u^k_{N_*} - 1| = \mathcal{O}\big((c^{\mathrm{T}}x^k - \nu_{N_*})^2\big), \tag{88}$$

$$\|u^k_{B_*}\| = \mathcal{O}\big(c^{\mathrm{T}}x^k - \nu_{N_*}\big), \tag{89}$$

$$\|s^k_{B_*}\| = \mathcal{O}\big((c^{\mathrm{T}}x^k - \nu_{N_*})^2\big). \tag{90}$$

(c) $N_k = N_*$ *for all* $k$ *sufficiently large and the following relations hold*:

$$\lim_{k\to\infty} \frac{\sigma_k}{(c^{\mathrm{T}}x^k - \nu_{N_*})} = 1, \tag{91}$$

$$\lim_{k\to\infty} \sigma_k = 0. \tag{92}$$

**Proof.** By Proposition 2.4(b) and the fact that $x^*_{N_*} = 0$, we have

$$\|x^k_{N_*}\| \leqslant \|x^k - x^*\| = \mathcal{O}(c^{\mathrm{T}}x^k - \nu_{N_*}). \tag{93}$$

Clearly, this implies that $\{r^k_{N_*}\}$ is bounded and that (86) holds, since $\lim_{k\to\infty} x^k_{B_*} = x^*_{B_*} > 0$. By (47) and (10), we have

$$u^k = \frac{X^k s^k}{c^{\mathrm{T}} x^k - \nu_{N_*}}, \tag{94}$$

from which we conclude that $u^k_{N_*} = R^k_{N_*} s^k_{N_*}$. This fact together with Proposition 2.4(c) and the fact that $\{r^k_{N_*}\}$ is bounded imply that $\{u^k_{N_*}\}$ is also bounded. Due to Lemma 4.3, Lemma 4.4, the fact that $\{u^k_{N_*}\}$ is bounded and (86), we conclude that (87), (88) and (89) hold. Clearly, (87) and (89) imply that $\{u^k\}$ and $\{\tilde{u}^k_{N_*}\}$ are bounded. Relation (90) follows immediately from (89), (94) and the fact that $\lim_{k\to\infty} x^k_{B_*} > 0$. It remains to show (c). Let $\tau_k \equiv e^{\mathrm{T}}((X^k)^{-1} d^k) = (x^k)^{\mathrm{T}} s^k$. Using (88), (89) and (94), we obtain

$$\lim_{k\to\infty} \frac{\tau_k}{c^{\mathrm{T}} x^k - \nu_{N_*}} = \lim_{k\to\infty} e^{\mathrm{T}} u^k = 1,$$

and hence, that

$$\lim_{k\to\infty} \tau_k = 0. \tag{95}$$

These two relations together with (93) imply that

$$x^k_i = \mathcal{O}(c^{\mathrm{T}} x^k - \upsilon_{N_*}) \leqslant \mathcal{O}(\tau_k) \leqslant \sqrt{\tau_k},$$

for all $i \in N_*$ and $k$ sufficiently large. Moreover, (95) and the fact that $\lim_{k\to\infty} x^k_{B_*} = x^*_{B_*} > 0$ imply that $x^k_i > \sqrt{\tau_k}$ for all $i \in B_*$ and $k$ sufficiently large. From these two observations and (82$a$), we conclude that $N_k = N_*$ for all $k$ sufficiently large. Relation (92) follows immediately from (91), which in turn is an immediate consequence of (88) and the fact that $\sigma_k/(c^{\mathrm{T}} x^k - \nu_{N_*}) = e^{\mathrm{T}} u^k_{N_k} = e^{\mathrm{T}} u^k_{N_*}$ for all $k$ sufficiently large. □

**Remark.** From the previous lemma, it immediately follows that $\sigma_k \sim c^{\mathrm{T}} x^k - \upsilon_{N_*}$. This means that any quantity appearing in the analysis below whose order is $\mathcal{O}(\sigma_k)$ is also $\mathcal{O}(c^{\mathrm{T}} x^k - \nu_{N_*})$ and vice versa.

**Lemma 5.2.**

$$\big|\|(R^k_{N_*})^{-1} \eta^k_{N_*}\| - \epsilon_k\big| = \mathcal{O}(\sigma_k). \tag{96}$$

**Proof.** Due to relations (26) and (59), we have (see (78))

$$\|(R^k_{N_*})^{-1} \eta^k_{N_*}\|^2 = |N_*| - \frac{1}{\|\tilde{u}^k_{N_*}\|^2}. \tag{97}$$

Define

$$\phi_k \equiv \frac{1}{\|\tilde{u}^k_{N_*}\|^2} - \frac{(e^{\mathrm{T}} u^k_{N_*})^2}{\|u^k\|^2} \quad \forall k \geqslant 0. \tag{98}$$

Using the inequality $(\alpha - \gamma)^2 \leqslant |\alpha^2 - \gamma^2|$ for $\alpha > 0$ and $\gamma > 0$, relation (97) and Lemma 5.1(c), we obtain for every $k$ sufficiently large that

$$\begin{aligned}\left(\epsilon_k - \|(R^k_{N_*})^{-1}\eta^k_{N_*}\|\right)^2 &\leqslant \left|\epsilon_k^2 - \|(R^k_{N_*})^{-1}\eta^k_{N_*}\|^2\right| \\ &= \left| |N_*| - \frac{(e^{\mathrm{T}}u^k_{N_*})^2}{\|u^k\|^2} - \|(R^k_{N_*})^{-1}\eta^k_{N_*}\|^2 \right| \\ &\leqslant |\phi_k|.\end{aligned}$$

It remains to show that $|\phi_k| = \mathcal{O}(\sigma_k^2)$. Indeed, using a bounding scheme similar to the one in (68) together with (66), Lemma 4.3 and Lemma 4.4, it is easy to see that

$$|\phi_k| = \mathcal{O}(\|(X^k_{B_*})^{-1}\|^2\, \|X^k_{N_*}\|^2).$$

This relation together with (86) and (91) imply $|\phi_k| = \mathcal{O}(\sigma_k^2)$. □

The next result establishes global convergence of Algorithm SLA.

**Theorem 5.3.** *$(N_*, B_*)$ is the optimal partition of (1), or equivalently, $x^*$ lies in the relative interior of the optimal face of (1).*

**Proof.** If the condition $\epsilon_k < \sigma_k^q$ is satisfied for finitely many indices $k$, the result follows from Proposition 2.5 since, in this case, $\sigma_k = \frac{1}{2}$ for all $k$ sufficiently large. Assume now that the set $\mathcal{K}$ of all indices $k$ satisfying $\epsilon_k < \sigma_k^q$ is infinite. By (92) and the definition of $\mathcal{K}$, we have $\lim_{k\in\mathcal{K}} \epsilon_k = 0$, and hence, in view of Lemma 5.2, we conclude that $\lim_{k\in\mathcal{K}} \|(R^k_{N_*})^{-1}\eta^k_{N_*}\| = 0$. Using the equivalence between (b) and (c) of Lemma 4.9, we conclude that $(N_*, B_*)$ is the optimal partition of (1). □

We now focus our attention on the superlinear convergence analysis of Algorithm SLA. We start with the following technical result.

**Lemma 5.4.** *Consider the function $\theta(x)$ defined in (63) with $N = N_*$ and let $\theta_k \equiv \theta(x^k)$ for all $k \geqslant 0$. For all $k$ sufficiently large, we have:*

$$|\theta_k^{-1} - 1| \leqslant \|(R^k_{N_*})^{-1}\eta^k_{N_*}\| + \mathcal{O}(\sigma_k^2), \tag{99}$$

*where $\theta_k \equiv \theta(x^k)$ for all $k$.*

**Proof.** By (89) and (88), we have $\max(u^k) = \max(u^k_{N_*})$ for every $k$ sufficiently large. This observation together with (63), (65), (86) and (91) imply

$$\theta_k^{-1} = \frac{\max(u^k_{N_*})}{\|u^k\|^2} = \max\left(e + (R^k_{N_*})^{-1}\eta^k_{N_*} + h_{N_*}(x^k)\right),$$

where $h_{N_*}(x^k) = \mathcal{O}(\|(X^k_{B_*})^{-1}\|^2\|X^k_{N_*}\|^2) \leqslant \mathcal{O}(\sigma_k^2)$. Hence, we obtain

$$|\theta_k^{-1} - 1| \leqslant \|(R^k_{N_*})^{-1}\eta^k_{N_*}\| + \|h_{N_*}(x^k)\| \leqslant \|(R^k_{N_*})^{-1}\eta^k_{N_*}\| + \mathcal{O}(\sigma_k^2). \qquad \square$$

To simplify our presentation, we introduce the following set of indices:

$$\mathcal{K}_P = \{k : \text{a predictor step is taken at the } k\text{th iteration}\}.$$

In the remaining part of this section, we use $r^*_{N_*}$ to denote the point $(\bar{s}^q_{N_*})^{-1}/|N_*|$. By Lemma 4.7 and Theorem 5.3, we know that $r^*_{N_*}$ is the (unique) optimal solution of problem (44) with $N = N_*$. The main result about the predictor steps is given next.

**Lemma 5.5.** *For every $k \in \mathcal{K}_P$, we have:*
(a) $\|r^k_{N_*} - r^*_{N_*}\| = \mathcal{O}(\sigma^q_k)$;
(b) $\sigma_{k+1} \sim (c^{\mathrm{T}}x^{k+1} - \nu_{N_*}) \sim (c^{\mathrm{T}}x^k - \nu_{N_*})^{1+p} \sim \sigma_k^{1+p}$;
(c) $\|r^{k+1}_{N_*} - r^*_{N_*}\| = \mathcal{O}(\sigma_{k+1}^{(q-p)/(1+p)})$ *and hence,* $\lim_{k\in\mathcal{K}_P} \|r^{k+1}_{N_*} - r^*_{N_*}\| = 0$.

**Proof.** Observe that (83) implies that $\epsilon_k = \mathcal{O}(\sigma^q_k)$ for all $k \in \mathcal{K}_P$. Using this observation, Lemma 5.2, (92) and the fact that $q < 1$, we obtain that

$$\|(R^k_{N_*})^{-1}\eta^k_{N_*}\| = \mathcal{O}(\sigma^q_k) \quad \forall k \in \mathcal{K}_P \text{ sufficiently large,} \tag{100}$$

and hence, $\lim_{k\to\infty} \|(R^k_{N_*})^{-1}\eta^k_{N_*}\| = 0$. It then follows from Lemma 4.9 that $\lim_{k\to\infty} r^k_{N_*} = r^*_{N_*}$. This observation together with Lemma 3.2(c) imply

$$\|r^k_{N_*} - r^*_{N_*} - \eta^k_{N_*}\| = \mathcal{O}(\|r^k_{N_*} - r^*_{N_*}\|^2), \tag{101}$$

$$\|r^k_{N_*} - r^*_{N_*}\| \sim \|\eta^k_{N_*}\| \sim \|(R^k_{N_*})^{-1}\eta^k_{N_*}\|. \tag{102}$$

for all $k \in \mathcal{K}_P$ sufficiently large. Statement (a) now follows from relations (100) and (102). We next show (b). By (61), we have $c^{\mathrm{T}}x^{k+1} - \nu_{N_*} = (1 - \lambda_k\theta_k)(c^{\mathrm{T}}x^k - \nu_{N_*})$ for all $k \geqslant 0$. In view of (91), (b) follows once we show that

$$1 - \lambda_k\theta_k \sim \sigma^p_k \quad \forall k \in \mathcal{K}_P \text{ sufficiently large.} \tag{103}$$

Using (100), Lemma 5.4, (92) and the fact that $q < 2$, we conclude that

$$|\theta_k - 1| = \mathcal{O}(\sigma^q_k) \quad \forall k \in \mathcal{K}_P \text{ sufficiently large.} \tag{104}$$

Using this observation and the fact that, by (84) and (92), we have $\lambda_k = 1 - \sigma^p_k$ for all $k \in \mathcal{K}_P$ sufficiently large, we obtain

$$|(1 - \lambda_k\theta_k) - \sigma^p_k| = |(1 - \lambda_k\theta_k) - (1 - \lambda_k)| = \lambda_k|\theta_k - 1| = \mathcal{O}(\sigma^q_k).$$

Using (92), the fact that, by (80), $q > p$, and the above relation, we conclude that $1 - \lambda_k\theta_k \sim \sigma^p_k$, and hence (b) follows. It remains to show (c). By (64), we have

$$r^{k+1}_{N_*} = r^k_{N_*} - \frac{\lambda_k\theta_k}{1 - \lambda_k\theta_k}\left(\eta^k_{N_*} + R^k_{N_*}h_{N_*}(x^k)\right). \tag{105}$$

Moreover, by (65), (86) and (91), we have

$$h_{N_*}(x^k) = \mathcal{O}(\|(X^k_{B_*})^{-1}\|^2\,\|X^k_{N_*}\|^2) \leqslant \mathcal{O}((c^{\mathrm{T}}x^k - \nu_{N_*})^2) \leqslant \mathcal{O}(\sigma^2_k). \tag{106}$$

Using relations (101), (102), (103), (105) and (106), Lemma 5.1(a) and statements (a) and (b), we obtain that, for all $k \in \mathcal{K}_P$ sufficiently large,

$$
\begin{aligned}
\|r^{k+1}_{N_*} - r^*_{N_*}\| &= \left\| (r^k_{N_*} - r^*_{N_*} - \eta^k_{N_*}) - \left(\frac{\lambda_k\theta_k}{1-\lambda_k\theta_k} - 1\right)\eta^k_{N_*} - \frac{\lambda_k\theta_k}{1-\lambda_k\theta_k} R^k_{N_*} h_{N_*}(x^k) \right\| \\
&\leqslant \|r^k_{N_*} - r^*_{N_*} - \eta^k_{N_*}\| + \frac{|2\lambda_k\theta_k - 1|}{1-\lambda_k\theta_k}\|\eta^k_{N_*}\| + \frac{\lambda_k\theta_k}{1-\lambda_k\theta_k}\|R^k_{N_*}\|\,\|h_{N_*}(x^k)\| \\
&\leqslant \mathcal{O}(\|r^k_{N_*} - r^*_{N_*}\|^2) + \mathcal{O}\left(\frac{\|r^k_{N_*} - r^*_{N_*}\|}{\sigma_k^p}\right) + \mathcal{O}(\sigma_k^{2-p}) \\
&\leqslant \mathcal{O}(\sigma_k^{q-p}) \leqslant \mathcal{O}(\sigma_{k+1}^{(q-p)/(1+p)}). \qquad \square
\end{aligned} \tag{107}
$$

The main result about the corrector steps is as follows.

**Lemma 5.6.** *Assume that $\mathcal{K}$ is an infinite index set such that $\lim_{k\in\mathcal{K}} r^k_{N_*} = r^*_{N_*}$ and $\lambda_k = \frac{1}{2}$ for all $k \in \mathcal{K}$. Then, the following statements hold:*

(a) $\lim_{k\in\mathcal{K}} (c^{\mathrm{T}}x^{k+1} - \nu_{N_*})/(c^{\mathrm{T}}x^k - \nu_{N_*}) = \frac{1}{2}$;

(b) $\lim_{k\in\mathcal{K}} \sigma_{k+1}/\sigma_k = \frac{1}{2}$;

(c) $\|r^{k+1}_{N_*} - r^*_{N_*}\| = \mathcal{O}(\|r^k_{N_*} - r^*_{N_*}\|^2 + \sigma_k^2)$ *for every* $k \in \mathcal{K}$ (*hence,* $\lim_{k\in\mathcal{K}} r^{k+1}_{N_*} = r^*_{N_*}$).

**Proof.** The assumption $\lim_{k\in\mathcal{K}} r^k_{N_*} = r^*_{N_*}$ implies that $\lim_{k\in\mathcal{K}} \|(R^k_{N_*})^{-1}\eta^k_{N_*}\| = 0$, which together with (92) and Lemma 5.4 imply $\lim_{k\in\mathcal{K}} \theta_k = 1$. This relation together with the assumption that $\lambda_k = \frac{1}{2}$ for all $k \in \mathcal{K}$ imply

$$\lim_{k\in\mathcal{K}} 1 - \lambda_k\theta_k = \tfrac{1}{2}. \tag{108}$$

This in turn implies (a), due to (61). Statement (b) follows immediately from (a) and (91). We next prove (c). Since $\lim_{k\in\mathcal{K}} r^k_{N_*} = r^*_{N_*}$, Lemma 3.2(c) implies that

$$\|r^k_{N_*} - r^*_{N_*} - \eta^k_{N_*}\| = \mathcal{O}(\|r^k_{N_*} - r^*_{N_*}\|^2), \tag{109}$$

$$\|r^k_{N_*} - r^*_{N_*}\| \sim \|\eta^k_{N_*}\| \sim \|(R^k_{N_*})^{-1}\eta^k_{N_*}\|. \tag{110}$$

hold for every $k \in \mathcal{K}$ sufficiently large. Hence, by Lemma 5.4, we have

$$|\theta_k - 1| = \mathcal{O}(\|(R^k_{N_*})^{-1}\eta^k_{N_*}\| + \sigma_k^2) = \mathcal{O}(\|r^k_{N_*} - r^*_{N_*}\| + \sigma_k^2), \tag{111}$$

for every $k \in \mathcal{K}$ sufficiently large. Using relations (108), (109), (110) and (111), Lemmas 4.6 and 5.1(b) and the assumption that $\lambda_k = \frac{1}{2}$ for all $k \in \mathcal{K}$, we obtain by using an argument similar to (105), (106) and (107) that

$$
\begin{aligned}
\|r^{k+1}_{N_*} - r^*_{N_*}\| &\leqslant \|r^k_{N_*} - r^*_{N_*} - \eta^k_{N_*}\| + \frac{|2\lambda_k\theta_k - 1|}{1-\lambda_k\theta_k}\|\eta^k_{N_*}\| + \frac{\lambda_k\theta_k}{1-\lambda_k\theta_k}\|R^k_{N_*}\|\,\|h_{N_*}(x^k)\| \\
&\leqslant \mathcal{O}(\|r^k_{N_*} - r^*_{N_*}\|^2) + \mathcal{O}(|\theta_k - 1|\,\|r^k_{N_*} - r^*_{N_*}\|) + \mathcal{O}(\sigma_k^2) \\
&\leqslant \mathcal{O}(\|r^k_{N_*} - r^*_{N_*}\|^2 + \sigma_k^2),
\end{aligned}
$$

for every $k \in \mathcal{K}$ sufficiently large. $\square$

The next result shows that, asymptotically, a predictor step must occur at every two steps as long as the set $\mathcal{K}_P$ is infinite.

**Lemma 5.7.** *For every k sufficiently large, the following implication holds:*

$$k \in \mathcal{K}_P, \quad k+1 \notin \mathcal{K}_P \implies k+2 \in \mathcal{K}_P. \tag{112}$$

**Proof.** Let $\mathcal{K}'_P \equiv \{k : k \in \mathcal{K}_P,\ k+1 \notin \mathcal{K}_P\}$. In view of Lemma 5.5(c), it follows that the set $\mathcal{K} \equiv \{k+1 : k \in \mathcal{K}'_P\}$ satisfies the assumption of Lemma 5.6. Hence, it follows from (b) and (c) of Lemma 5.6 that

$$\|r_{N_*}^{k+2} - r_{N_*}^*\| = \mathcal{O}(\|r_{N_*}^{k+1} - r_{N_*}^*\|^2 + \sigma_{k+1}^2) \quad \forall k \in \mathcal{K}'_P, \tag{113}$$

$$\sigma_{k+2} \sim \sigma_{k+1} \quad \forall k \in \mathcal{K}'_P, \tag{114}$$

and $\lim_{k\in\mathcal{K}'_P} \|r_{N_*}^{k+2} - r_{N_*}^*\| = 0$. Hence, in view of Lemma 3.2(c), we have

$$\|r_{N_*}^{k+2} - r_{N_*}^*\| \sim \|\eta_{N_*}^{k+2}\| \sim \|(R_{N_*}^{k+2})^{-1}\eta_{N_*}^{k+2}\| \quad \forall k \in \mathcal{K}'_P. \tag{115}$$

Using Lemma 5.2, (115), (113), (114) and Lemma 5.5(c), we obtain

$$\begin{aligned}
\epsilon_{k+2} &\leqslant \|(R_{N_*}^{k+2})^{-1}\eta_{N_*}^{k+2}\| + \mathcal{O}(\sigma_{k+2}) \\
&\leqslant \mathcal{O}(\|r_{N_*}^{k+2} - r_{N_*}^*\| + \sigma_{k+2}) \\
&\leqslant \mathcal{O}(\|r_{N_*}^{k+1} - r_{N_*}^*\|^2 + \sigma_{k+1}^2 + \sigma_{k+2}) \\
&\leqslant \mathcal{O}(\sigma_{k+1}^{2(q-p)/(1+p)} + \sigma_{k+1}^2 + \sigma_{k+2}) \\
&\leqslant \mathcal{O}(\sigma_{k+2}^s),
\end{aligned}$$

for every $k \in \mathcal{K}'_P$, where $s \equiv \min\{2(q-p)/(1+p), 1\}$. By (81), we have $s > q$. This observation together with the above relation imply that $\epsilon_{k+2} < \sigma_{k+2}^q$, or equivalently, $k+2 \in \mathcal{K}_P$, for every $k \in \mathcal{K}'_P$ sufficiently large. □

The next result shows that the set $\mathcal{K}_P$ is infinite. In view of the Lemma 5.7, this clearly implies that, asymptotically, a predictor step occurs at every two steps of Algorithm SLA.

**Lemma 5.8.** *The set $\mathcal{K}_P$ is infinite.*

**Proof.** Assume for contradiction that there exists an integer $k_0$ such that $k \notin \mathcal{K}_P$ for every $k \geqslant k_0$. By Proposition 2.5(c), we have

$$\lim_{k\to\infty} u_{N_*}^k = \lim_{k\to\infty} X_{N_*}^k s_{N_*}^k/(c^{\mathrm{T}}x^k - \nu_{N_*}) = e/|N_*|.$$

This relation together with Lemma 5.1(a) and Lemma 4.9 imply that $\lim_{k\to\infty} r_{N_*}^k = r_{N_*}^*$. Hence, it follows that the assumptions of Lemma 5.6 are satisfied with $\mathcal{K} = \{k : k \geqslant k_0\}$. By using statements (b) and (c) of this lemma, we conclude that $\lim_{k\to\infty} \sigma_k/\sigma_{k+1} = 2$ and that, for some constant $L_0 > 0$,

$$\|r_{N_*}^{k+1} - r_{N_*}^*\| \leqslant L_0(\|r_{N_*}^k - r_{N_*}^*\|^2 + \sigma_k^2) \quad \forall k \geqslant k_0. \tag{116}$$

Hence, by taking a larger $k_0$ if necessary, we may assume that

$$\sigma_k \leqslant 3\sigma_{k+1} \quad \forall k \geqslant k_0, \tag{117}$$

$$\|r^k_{N_*} - r^*_{N_*}\| \leqslant \frac{1}{18L_0} \quad \forall k \geqslant k_0, \tag{118}$$

where (118) is due to the fact that $\lim_{k\to\infty} r^k_{N_*} = r^*_{N_*}$. We next show by induction that

$$\|r^k_{N_*} - r^*_{N_*}\| \leqslant L_1\sigma_k^2 \quad \forall k \geqslant k_0, \tag{119}$$

where $L_1 \equiv \max\{18L_0, \|r^{k_0}_{N_*} - r^*_{N_*}\|/\sigma^2_{k_0}\}$. Indeed, (119) obviously hold for $k = k_0$ in view of the definition of $L_1$. If (119) holds for $k = l \geqslant k_0$ then (116), (117) and (118) imply

$$\|r^{l+1}_{N_*} - r^*_{N_*}\| \leqslant \frac{1}{18}\|r^l_{N_*} - r^*_{N_*}\| + L_0\sigma_l^2 \leqslant \left(\frac{L_1}{18} + L_0\right)\sigma_l^2 \leqslant \frac{L_1}{9}\sigma_l^2 \leqslant L_1\sigma^2_{l+1}.$$

where the third inequality follows from the definition of $L_1$. We have thus proved that (119) holds. Using (119) and Lemma 3.2(c), we conclude that $\|(R^k_{N_*})^{-1}\eta^k_{N_*}\| = \mathcal{O}(\sigma_k^2)$ for all $k$. This observation together with Lemma 5.2 then imply that $\epsilon_k = \mathcal{O}(\sigma_k)$ for all $k$. Using (92) and the fact that $q < 1$, we conclude that $\epsilon_k < \sigma_k^q$, or equivalently, $k \in \mathcal{K}_P$, for every $k$ sufficiently large. Since this conclusion contradicts our initial assumption, the result follows. □

The next result is needed in the proof that the sequence $\{(y^k, s^k)\}$ converges two-step superlinearly to the analytic center of the dual optimal face.

**Lemma 5.9.** *If* $\lim_{k\to\infty} r^k_{N_*} = r^*_{N_*}$ *then*

$$\left\|u^k_{N_*} - \frac{e}{|N_*|}\right\| = \mathcal{O}(\|r^k_{N_*} - r^*_{N_*}\| + (c^{\mathrm{T}}x^k - \nu_{N_*})^2) \quad \forall k \geqslant 0. \tag{120}$$

**Proof.** Lemma 3.2(c) together with $\lim_{k\to\infty} r^k_{N_*} = r^*_{N_*}$ imply $\|(R^k_{N_*})^{-1}\eta^k_{N_*}\| \sim \|r^k_{N_*} - r^*_{N_*}\|$. Using this observation, Lemma 5.1(a) and relations (67) and (78) with $N = N_*$, we obtain

$$\begin{aligned}\left\|\tilde{u}^k_{N_*} - \frac{e}{|N_*|}\right\| &\leqslant \left\|\tilde{u}^k_{N_*}\left(1 - \frac{1}{|N_*|\,\|\tilde{u}^k_{N_*}\|^2}\right)\right\| + \left\|\frac{\tilde{u}^k_{N_*}}{|N_*|\,\|\tilde{u}^k_{N_*}\|^2} - \frac{e}{|N_*|}\right\| \\ &\leqslant \frac{\|\tilde{u}^k_{N_*}\|\,\|(R^k_{N_*})^{-1}\eta^k_{N_*}\|^2}{|N_*|} + \frac{\|(R^k_{N_*})^{-1}\eta^k_{N_*}\|}{|N_*|} \\ &\leqslant \mathcal{O}(\|r^k_{N_*} - r^*_{N_*}\|) \quad \forall k \geqslant 0.\end{aligned}$$

This relation together with (87) then imply

$$\begin{aligned}\left\|u^k_{N_*} - \frac{e}{|N_*|}\right\| &\leqslant \|u^k_{N_*} - \tilde{u}^k_{N_*}\| + \left\|\tilde{u}^k_{N_*} - \frac{e}{|N_*|}\right\| \\ &= \mathcal{O}(\|r^k_{N_*} - r^*_{N_*}\| + (c^{\mathrm{T}}x^k - \nu_{N_*})^2) \quad \forall k \geqslant 0. \qquad \square\end{aligned}$$

The next result establishes the two-step superlinear convergence of Algorithm SLA.

**Theorem 5.10.** *Algorithm SLA has the following properties:*

(a) *the sequence* $\{c^{\mathrm{T}}x^k\}$ *converges 2-step superlinearly to the optimal value* $\nu_{N_*} = c^{\mathrm{T}}x^*$ *with Q-order at least* $1+p$, *namely*

$$\limsup_{k\to\infty} \frac{c^{\mathrm{T}}x^{k+2} - \nu_{N_*}}{(c^{\mathrm{T}}x^k - \nu_{N_*})^{1+p}} < \infty; \tag{121}$$

(b) *the sequence* $\{x^k\}$ *converges 2-step superlinearly with R-order at least* $1+p$ *to a point lying in the relative interior of the optimal face of problem* (1);

(c) *the sequence* $\{(y^k, s^k)\}$ *converges 2-step superlinearly with R-order at least* $1+p$ *to the analytic center of the optimal face of the dual problem* (2), *that is, the point* $(\bar{y}^a, \bar{s}^a)$ *defined in* (13).

**Proof.** It follows from Lemma 5.7 that if $k$ is sufficiently large then either $k \in \mathcal{K}_P$ or $k+1 \in \mathcal{K}_P$. This fact together with Lemma 5.5(b) clearly imply (a). By Proposition 2.4(b), we know that $\|x^k - x^*\| = \mathcal{O}(c^{\mathrm{T}}x^k - \nu_{N_*})$ and hence, in view of (a), it follows that $\{x^k\}$ converges 2-step superlinearly to $x^*$ with $R$-order at least $1+p$. We next show (c). From (a) and (90), it follows that that $\{s^k_{B_*}\}$ converges to $\bar{s}^a_{B_*} = 0$ two-step superlinearly with $R$-order at least $1+p$. We next analyze the convergence of $\{s^k_{N_*}\}$. From (a) and (c) of Lemma 5.5 and Lemma 5.7, it is easy to see that $\|r^k_{N_*} - r^*_{N_*}\| = \mathcal{O}((c^{\mathrm{T}}x^k - \nu_{N_*})^t)$, where $t \equiv (q-p)/(1+p)$. Using this observation, the relation $r^*_{N_*} = (\bar{s}^a_{N_*})^{-1} (>0)$ and Lemma 5.9, we obtain

$$\begin{aligned}
\|s^k_{N_*} - \bar{s}^a_{N_*}\| &= \left\| (R^k_{N_*})^{-1} u^k_{N_*} - \frac{(R^*)^{-1}e}{|N_*|} \right\| && (122)\\
&\leqslant \|(R^k_{N_*})^{-1}\| \left\| u^k_{N_*} - \frac{e}{|N_*|} \right\| + \frac{\|(R^k_{N_*})^{-1}e - (R^*)^{-1}e\|}{|N_*|} && (123)\\
&= \mathcal{O}\left( \left\| u^k_{N_*} - \frac{e}{|N_*|} \right\| + \|r^k_{N_*} - r^*_{N_*}\| \right) && (124)\\
&= \mathcal{O}(\|r^k_{N_*} - r^*_{N_*}\| + (c^{\mathrm{T}}x^k - \nu_{N_*})^2) && (125)\\
&= \mathcal{O}((c^{\mathrm{T}}x^k - \nu_{N_*})^t). && (126)
\end{aligned}$$

This clearly implies that $\{s^k_{N_*}\}$ converges to $\bar{s}^a_{N_*}$ two-step superlinearly with $R$-order at least $1+p$. □

## 6. Concluding remarks

In this paper we have demonstrated that a variant of the long-step AS algorithm is two-step superlinearly convergent with $Q(R)$-order as close to $\frac{4}{3}$ as desired. Practical efficiency of this algorithm is not known at this moment, but the results of this paper may suggest possible ways to implement the AS algorithm more reliably and efficiently. We believe that the analysis of this paper is important from the theoretical point of view since it shows that the AS algorithm with certain stepsizes is also able to keep the sequence of iterates well-centered, at least asymptotically. This is in some sense an unexpected result in view of the (pure) steepest descent nature of the AS algorithm.

One interesting research problem is to improve the order of convergence of the algorithm of Section 5. It would also be interesting to develop a variant of the AS algorithm with convergence order equal to any number less than or equal to two, a property which many primal-dual algorithms (e.g. [33] and [17]) and the Iri and Imai's algorithm [27] have been shown to have.

We believe that our analysis can be directly applied to the long-step variant of Karmarkar's algorithm [15] presented in [20]. It seems possible to show that this variant of Karmarkar's algorithm enjoys superlinear convergence without sacrificing its polynomial complexity by properly choosing the sequence of stepsizes according to the ideas suggested in this paper.

Another algorithm whose analysis could benefit from the techniques in this paper is Todd's low complexity algorithm [23]. During the predictor steps, his algorithm moves along the AS direction with stepsize less than $\frac{1}{2}$ (namely $\frac{1}{5}$ of the step to the boundary of the largest inscribed ellipsoid). Since the AS step with $\lambda_k \leqslant \frac{1}{2}$ works as a kind of corrector step, it seems possible to show that Todd's algorithm may not need any corrector step asymptotically (cf. [24]). Moreover, it seems possible to apply our analysis to show that a variant of Todd's algorithm is superlinearly convergent without sacrificing its polynomial complexity.

## Acknowledgment

This research was carried out while the first author was visiting the Center for Research on Parallel Computation and the Department of Computational and Applied Mathematics of Rice University, Houston, USA from March of 1992 to February of 1993. He thanks his host Prof. J.E. Dennis and the colleagues there for providing the excellent research environment during his stay at Rice University.

Part of this research was supported by the National Science Foundation (NSF) under grant DDM-9109404 and the Office of Naval Research (ONR) under grant N00014-93-1-0234. The second author thanks NSF and ONR for the financial support received during the completion of this work.

The first author also thanks the second author for having arranged and supported a trip with his NSF grant to visit the University of Arizona in the summer of 1992 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, "Data structures and programming techniques for the implementation of Karmarkar's algorithm," *ORSA Journal on Computing* 1 (1989) 84–106.

[2] 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. (Errata in *Mathematical Programming* 50 (1991) 415.)

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

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

[5] Y.-C. Cheng, D.J. Houck, J.-M. Liu, M.S. Meketon, L. Slutsman, R.J. Vanderbei and P. Wang, "The AT&T KORBX System", *AT&T Technical Journal* 68 (3) (1989) 7–19.

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

[7] I.I. Dikin, "On the convergence of an iterative process," *Upravlyaemye Sistemi* 12 (1974) 54–60 (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, USSR, 1980).

[11] R. Fletcher, *Practical Methods of Optimization*, 2nd edition (Wiley, New York, 1987).

[12] 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 (Rio de Janeiro, Brazil, Sept. 1990).

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

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

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

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

[17] S. Mehrotra, "Quadratic convergence in a primal-dual method," *Mathematics of Operations Research* 18 (1993) 741–751.

[18] 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 (1987) 261–267.

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

[20] M. Muramatsu and T. Tsuchiya, "Convergence analysis of the projective scaling algorithm based on a long-step homogeneous affine scaling algorithm," *Mathematical Programming* 72 (1996) 291–305.

[21] M.G.C. Resende and G. Veiga, "An efficient implementation of a network interior point method," in: D.S. Johnson and C.C. McGeoch, eds., *Network Flows and Matching: First DIMACS Implementation Challenge*, DIMACS Series on Discrete Mathematics and Theoretical Computer Science, Vol. 12 (American Mathematical Society, 1993) pp. 299–348.

[22] L. Sinha and B. Freedman and N. Karmarkar and A. Putcha and K. Ramakrishnan, "Overseas network planning," *Proc. Third Internat. Network Planning Sysmposium – Networks'86* (IEEE Communications Society, held on June 1–6, 1986, Tarpon Springs, FL, 1986) 121–124.

[23] M. Todd, "A low complexity interior-point algorithm for linear programming," *SIAM Journal on Optimization* 2 (1992) 198–209.

[24] M. Todd and J.-P. Vial, "Todd's complexity interior-point algorithm is a predictor-corrector path-following algorithm," *Operations Research Letters* 11 (1992) 199–207.

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

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

[27] T. Tsuchiya, "Quadratic convergence of Iri and Imai's method for degenerate linear programming problems," *Jounral of Optimization Theory and Applications* 87 (3) (1995) 703–726.

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

[29] T. Tsuchiya and M. Muramatsu, "Global convergence of a long-step affine scaling algorithm for degenerate linear programming problems," *SIAM Journal on Optimization* 5 (3) (1995) 525–551.

[30] R.J. Vanderbei and J.C. Lagarias, "I.I. Dikin's convergence result for the affine-scaling algorithm," *Contemporary Mathematics* 114 (1990) 109–119.

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

[32] C. Witzgall, P.T. Boggs and P.D. Domich, "On the convergence behavior of trajectories for linear programming," *Contemporary Mathematics* 114 (1990) 109–119.

[33] Y. Ye, O. Güler, R.A. Tapia and Y. Zhang, "A quadratically convergent $O(\sqrt{n}L)$-iteration algorithm for linear programming," *Mathematical Programming* 59 (1993) 151–162.