Global Convergence of the Affine Scaling Algorithm for Convex Quadratic Programming

Renato D.C. Monteiro* T. Tsuchiya†

March 24, 1995

Abstract

In this paper we give a global convergence proof of the second order affine scaling algorithm for convex quadratic programming problems, where the next iterate is the point which minimizes the objective function over the intersection of the feasible region with the ellipsoid centered at the current point and whose radius is a fixed fraction $\beta \in (0,1]$ of the radius of the largest "scaled" ellipsoid inscribed in the nonnegative orthant. The analysis is based on the local Karmarkar potential function introduced by Tsuchiya. For any $\beta \in (0,2/3]$ and without assuming any nondegeneracy assumption on the problem, it is shown that the sequences of primal iterates and dual estimates converge to optimal solutions of the quadratic program and its dual, respectively.

1 Introduction

The affine scaling (AS) algorithm for linear programming (LP) was introduced by Dikin [6] in 1967 but remained unknown to the west community until the late 80's. It was later rediscovered by Barnes [2] and Vanderbei et al. [26] as a natural simplification of Karmarkar's algorithm [12]. Because of the theoretical and practical importance of the AS algorithm, there are a number of papers which study its global and local convergence [2, 6, 7, 8, 10, 18, 20, 21, 23, 24, 26, 25] and the behavior of its associated continuous trajectories [1, 3, 13, 14, 27]. Like in the simplex algorithm, the analysis of the affine scaling algorithm for primal degenerate LP problems is much harder than for primal nondegenerate LP problems. By introducing a local Karmarkar potential function, Tsuchiya [20, 21] shows the global convergence of the short-step version of the AS algorithm where the next iterate minimizes the (linear) objective function over the intersection of the feasible region with the ellipsoid centered at the current point and whose radius is a fixed fraction β, $\beta \leq 1/8$, of the radius of the largest "scaled" ellipsoid inscribed in the nonnegative orthant. The global convergence of the long step version of the affine scaling algorithm for degenerate LP problems, that is the version where 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, was established by Dikin [8] for any $\lambda \in (0,1/2]$ and by Tsuchiya and Muramatsu [24] for any $\lambda \in (0,2/3]$. We refer the reader to Monteiro et al.

*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, 30332. The work of this author was based on research supported by the Office of Naval Research under grants N00014-93-1-0234 and N00014-94-1-0340.

†The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-Ku, Tokyo, 106, Japan.

[15] and Saigal [16] for a simplified and self-contained analysis of the AS algorithm in the context of LP problems.

In this paper, we focus our attention on the second order affine scaling algorithm for convex quadratic programming (QP), and prove its global convergence without any nondegeneracy assumption by using a local potential function similar to the one used in the LP context. This version of the AS algorithm was originally proposed by Dikin (see the references in his book [9]), and was later rediscovered by Ye [28] and Ye and Tse [30] after the introduction of Karmarkar's algorithm [12]. The algorithm analyzed in this paper is similar to the short-step version of the AS algorithm for LP problems in that the next iterate is the point which minimizes the (quadratic) objective function over the intersection of the feasible region with the ellipsoid centered at the current point and whose radius is a fixed fraction β of the radius of the largest "scaled" ellipsoid inscribed in the nonnegative orthant. The papers [28, 30] assume that the QP problem is primal nondegenerate and also impose an extra dual nondegeneracy assumption involving the objective function. These assumptions considerably simplify the convergence analysis of the AS algorithm.

Global convergence of the second order AS algorithm which drops the assumption of primal nondegeneracy but still keeps some sort of dual nondegeneracy is given in Tsuchiya [22] for strictly convex QP problems and any $\beta \in (0, 1/8]$. Sun [17] proves global convergence for this algorithm without imposing any nondegeneracy assumption. His analysis, which generalizes the one given in Tseng and Luo [18] for LP problems, is still restrictive in the sense that it allows only very small radius β, i.e. $\beta = 2^{-\mathcal{O}(L)}$, where L is the input size of the problem. In an unpublished manuscript, reported in the presentation [19], Tsuchiya proves global convergence of the second order AS algorithm for strictly convex QP with $\beta \in (0, 1/8]$. Part of this paper is based on the ideas developed in that manuscript.

Convergence of the first order AS algorithm for linearly constrained convex programming problems is proved in Gonzaga and Carlos [11] under the assumption of primal nondegeneracy. In this algorithm, the search direction is obtained by minimizing a first order linear approximation of the objective function at the current point over the same ellipsoid mentioned above and the next point is determined by performing a line search along this direction.

Convergence analysis of the second order AS algorithm for non-convex QP problems under the assumption of primal and dual nondegeneracy can be found in the papers Ye [29] and Bonnans and Bouhtou [4].

This paper is organized as follows. In Section 2 we describe the second order AS algorithm for QP problems and review some basic results about this method. After introducing a local assumption which requires that a partition of the variables satisfy certain conditions, we prove in Section 3 that the sequence of primal iterates and the sequence of dual estimates generated by the second order AS algorithm with $\beta \in (0, 2/3]$ converge to optimal solutions of the given QP problem and its dual, respectively. Two important tools play a crucial role in the convergence analysis: the local potential function and an asymptotic decomposition formula of the scaled projection matrix near the boundary of the feasible region. Finally, we show in Section 4 that the local assumption of Section 3 is satisfied by some (in fact, unique) partition of variables. Appendix A presents a general framework for analyzing the asymptotic reduction of the local potential function. The asymptotic decomposition formula for the scaled projection matrix is presented in Tsuchiya [21] but its derivation is quite involved. For this reason we give a simplified proof of this formula in Appendix B.

The following notation is used throughout our paper. We denote the vector of all ones by e. Its

dimension is always clear from the context. $\mathbb{R}^p$, $\mathbb{R}^p_+$ and $\mathbb{R}^p_{++}$ denote the p-dimensional Euclidean space, the nonnegative orthant of $\mathbb{R}^p$ and the positive orthant of $\mathbb{R}^p$, respectively. The set of all $p \times q$ matrices with real entries is denoted by $\mathbb{R}^{p\times q}$. Given a vector $x \in \mathbb{R}^p$ and an index set $\alpha \subseteq \{1,\ldots,p\}$, we denote the subvector $[x_i]_{i\in\alpha}$ by x_α. Given x and y in $\mathbb{R}^p$, $x \leq y$ means $x_i \leq y_i$ for every $i = 1,\ldots,n$. If $\alpha \subseteq \{1,\ldots,p\}$, $\beta \subseteq \{1,\ldots,q\}$ and $Q \in \mathbb{R}^{p\times q}$, we let $Q_{\alpha\beta}$ denote the submatrix $[Q_{ij}]_{i\in\alpha, j\in\beta}$. If $\alpha = \{1,\ldots,p\}$ we denote $Q_{\alpha\beta}$ simply by Q_β. For a vector w, we let $\chi[w]$ denote the largest component of w. The Euclidean norm, the 1-norm and the ∞-norm are denoted by $\|\cdot\|$, $\|\cdot\|_1$ and $\|\cdot\|_\infty$, respectively. The diagonal matrix corresponding to a vector w is denoted by $diag(w)$ and the vector whose i-th component is $1/w_i$ is denoted by w^{-1}. If J is a finite index set then $|J|$ denotes its cardinality, that is the number of elements of J. We say that (B,N) is a partition of $\{1,\ldots,p\}$ if $B \cup N = \{1,\ldots,p\}$ and $B \cap N = \emptyset$. The superscript ^{T} denotes transpose.

2 Notation and preliminary results

In this section we describe the second order AS algorithm for QP problems and review some basic results about this method.

We consider the following convex quadratic program (QP)

$$\begin{array}{ll} \text{minimize} & f(x) = \frac{1}{2}x^TQx + c^Tx \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array} \tag{1}$$

and its associated dual QP

$$\begin{array}{ll} \text{maximize} & b^Ty - \frac{1}{2}x^TQx \\ \text{subject to} & A^Ty + s = \nabla f(x) \\ & s \geq 0, \end{array} \tag{2}$$

where c, x, $s \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $Q \in R^{n\times n}$ is a symmetric positive semidefinite matrix.

The following notation is used throughout the paper. The quantity $(x^TQx)^{1/2}$ is simply denoted by $\|x\|_Q$. Clearly, $\|\cdot\|_Q$ is a norm only when Q is positive definite; in general, it fails to be a norm due to the fact that $\|x\|_Q = 0$ does not imply $x = 0$. If $\{\tau_k\}$ and $\{\beta_k\}$ are two sequences of real numbers and $\mathcal{K}$ is an infinite index set of nonnegative integers then the notation $\tau_k = \mathcal{O}(\beta_k)$ for $k \in \mathcal{K}$ means that there exists a scalar $r \geq 0$ such that $|\tau_k| \leq r\beta_k$ for every $k \in \mathcal{K}$ sufficiently large; clearly, if $\beta_k > 0$ for all $k \in \mathcal{K}$ then $|\tau_k| \leq r\beta_k$ holds for all $k \in \mathcal{K}$ by taking a larger scalar r if necessary. When $\mathcal{K}$ is equal to the set of all nonnegative integers, we simply write $\tau_k = \mathcal{O}(\beta_k)$. The notation $\tau_k \sim \beta_k$ for $k \in \mathcal{K}$ means that $\tau_k = \mathcal{O}(\beta_k)$ for $k \in \mathcal{K}$ and $\beta_k = \mathcal{O}(\tau_k)$ for $k \in \mathcal{K}$.

Given a partition (B,N) of $\{1,\ldots,n\}$, we let

$$\begin{array}{rcl} \mathcal{P}_N & \equiv & \{x \in \mathbb{R}^n \mid A_Bx_B = b,\ x_N = 0\}; \\ \mathcal{P}_N^+ & \equiv & \{x \in \mathcal{P}_N \mid x_B \geq 0\}; \\ \mathcal{P}_N^{++} & \equiv & \{x \in \mathcal{P}_N \mid x_B > 0\}; \\ \mathcal{D}_B(x) & \equiv & \{(y,s) \in \mathbb{R}^m \times \mathbb{R}^n \mid A^Ty + s = \nabla f(x),\ s_B = 0\}; \\ \mathcal{D}_B^+(x) & \equiv & \{(y,s) \in \mathcal{D}_B \mid s_N \geq 0\}; \\ \mathcal{D}_B^{++}(x) & \equiv & \{(y,s) \in \mathcal{D}_B \mid s_N > 0\}. \end{array} \tag{3}$$

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 $\mathcal{D}_B(x)$, $\mathcal{D}_B^+(x)$ and $\mathcal{D}_B^{++}(x)$ by $\mathcal{D}(x)$, $\mathcal{D}^+(x)$ and $\mathcal{D}^{++}(x)$, respectively.

Throughout this paper, we make the following assumption.

Assumption 2.1 *The following assumptions are made with respect to problem (1):*

(A) $\mathrm{rank}\,(A) = m$*;*

(B) $\mathcal{P}^{++} \neq \emptyset$*;*

(C) the set of optimal solutions of problem (1) is nonempty and bounded.

We next describe the second order affine scaling algorithm which is analyzed in this paper.

Algorithm AS: Let $x^0 \in \mathcal{P}^{++}$ and $\beta \in (0, 2/3]$ be given. Set $k = 0$.

step 1: compute the solution x^{k+1} of the problem

$$\min\{f(x) \mid Ax = b,\ \|(X^k)^{-1}(x - x^k)\|^2 \leq \beta^2\}; \tag{4}$$

step 2: if $\|(X^k)^{-1}(x^{k+1} - x^k)\| < \beta$ then stop; otherwise, increment k by 1 and go to step 1.

end of Algorithm AS.

Using the fact that $\|(X^k)^{-1}(x^{k+1} - x^k)\| \leq \beta$, it follows that

$$(1 - \beta)x^k \leq x^{k+1} \leq (1 + \beta)x^k, \qquad \forall k \geq 0. \tag{5}$$

As a consequence, it follows that all iterates generated by Algorithm AS lie in $\mathcal{P}^{++}$.

Proposition 2.2 *The following implications hold:*

(a) if $\|(X^k)^{-1}(x^{k+1} - x^k)\| < \beta$ *for some* k *then* x^{k+1} *is an optimal solution of problem (1);*

(b) if $\|(X^k)^{-1}(x^{k+1} - x^k)\| = \beta$ *for all* k *then every accumulation point of the sequence* $\{x^k\}$ *lies on the relative boundary of* $\mathcal{P}^+$*, that is, the set* $\mathcal{P}^+ \backslash \mathcal{P}^{++}$*.*

When case (a) of Proposition 2.2 occurs, Algorithm AS finds an optimal solution of problem (1) in a finite number of iterations. Throughout the paper we assume that case (b) of Proposition 2.2 holds. One of our goals will be to show that the sequence of iterates $\{x^k\}$ generated by Algorithm AS converges to an optimal solution of (1).

Throughout this paper, we let $X^k \equiv diag(x^k)$ and $g^k = \nabla f(x^k)$ for every k. The following quantities associated with the sequence $\{x^k\}$ generated by Algorithm AS play a fundamental role in the sequel:

$$y^k \equiv \left[A(X^k)^2A^T\right]^{-1} A(X^k)^2 g^{k+1}, \tag{6}$$

$$s^k \equiv g^{k+1} - A^T y^k = g^{k+1} - A^T \left[A(X^k)^2A^T\right]^{-1} A(X^k)^2 g^{k+1}, \tag{7}$$

$$P(x^k) \equiv I - X^k A^T \left[A(X^k)^2A^T\right]^{-1} AX^k, \tag{8}$$

$$d^k \equiv (X^k)^2 s^k = X^k P(x^k) X^k g^{k+1}. \tag{9}$$

Using the fact that $\|(X^k)^{-1}(x^{k+1} - x^k)\| = \beta$ and x^{k+1} satisfies the optimality condition associated with (1), we can easily prove the following result.

Lemma 2.3 *The following relations hold for every integer $k \geq 0$:*

$$x^{k+1} = x^k - \beta \frac{d^k}{\|(X^k)^{-1}d^k\|} = x^k - \beta \frac{(X^k)^2 s^k}{\|X^k s^k\|}; \tag{10}$$

$$(g^{k+1})^T d^k = (s^k)^T d^k = \|(X^k)^{-1}d^k\|^2 = \|X^k s^k\|^2; \tag{11}$$

$$\beta \|X^k s^k\| = (g^{k+1})^T (x^k - x^{k+1}) \leq f(x^k) - f(x^{k+1}). \tag{12}$$

The following result is well-known (see for example Ye [28] (or Ye and Tse [30]) and Sun [17]).

Lemma 2.4 *The following statements hold:*

(a) the sequence $\{x^k\}$ converges to some point $x^* \in \mathcal{P}^+$;

(b) $f(x^{k+1}) - f(x^*) \leq (1 - \beta/\sqrt{n})(f(x^k) - f(x^*))$ for every k sufficiently large;

(c) there exists a constant $C_0 > 0$ such that $\|x^k - x^*\|^2 \leq C_0[f(x^k) - f^*]$ for every $k \geq 0$;

(d) the sequence $\{(y^k, s^k)\}$ is bounded;

(e) $\sum_{k=0}^{\infty} \|x^k s^k\| < \infty$ (in particular, every accumulation point (y^*, s^*) of $\{(y^k, s^k)\}$ satisfies $x^* s^* = 0$).

Proof. The proof of (a), (b) and (c) are given in Theorem 1 and Theorem 2 of Sun [17] under the conditions stated in Assumption 2.1. It has been proved in Vanderbei and Lagarias [25] that the matrix valued mapping $U(x) \equiv (AX^2A^T)^{-1}AX^2$ is bounded on $\mathcal{P}^{++}$ whenever Assumption 2.1(A) holds. Since $y^k = U(x^k)\nabla f(x^{k+1})$, $s^k = \nabla f(x^{k+1}) - A^T y^k$ and $\{\nabla f(x^{k+1})\}$ is bounded, it follows that $\{(y^k, s^k)\}$ is bounded. Statement (e) is an immediate consequence of (12). ∎

The following trivial result is frequently used in our presentation.

Lemma 2.5 *If (y^*, s^*) is an accumulation point $\{(y^k, s^k)\}$ and (B, N) is a partition of $\{1, \ldots, n\}$ such that $\{i \mid s_i^* \neq 0\} \subseteq N$ then $\nabla f(x^*)^T(x - x^*) = (s_N^*)^T x_N$ for any $x \in \mathcal{P}$.*

Proof. First note that $A^T y^k + s^k = \nabla f(x^{k+1})$ implies $A^T y^* + s^* = \nabla f(x^*)$. This relation together with the fact that $A(x - x^*) = 0$, $x^* s^* = 0$ and $s_B^* = 0$ imply that

$$\nabla f(x^*)^T(x - x^*) = (s^*)^T(x - x^*) = (s_N^*)^T x_N.$$

∎

3 Convergence to an optimal solution

In this section we review an important asymptotic decomposition result for the projection matrix $P(x^k)$. This result is used in several places of our presentation and it is important in its own right. We also introduce a potential function that generalizes the one used in the context of linear programs and plays an important role in the convergence analysis of this section. After introducing a local assumption which guarantees the existence of a partition (B, N) of $\{1, \ldots, n\}$ satisfying certain properties, we prove that the sequence of primal iterates and the sequence of dual estimates generated by the second order AS algorithm with $\beta \in (0, 2/3]$ converge to optimal solutions of the given QP problem and its dual, respectively. The main goal of the next section is to show the existence of the partition (B, N) required for the analysis of this section.

We start this section by stating the following result on a decomposition of the projection matrix $P(x^k)$ whose proof can be found in Tsuchiya [21]. For completeness, a simpler proof of this result is given in the appendix. It is convenient to introduce the following notation. Let $P(x)$ denote the projection matrix onto the null space of the matrix AX, $\tilde{P}_B(x_B)$ denote the projection matrix onto the null space of the matrix $A_B X_B$ and $\tilde{P}_N(x_N)$ denote the projection matrix onto the subspace $\{p_N \mid AXp = 0\} = \{p_N \mid A_N X_N p_N \in \text{Range}(A_B)\}$. Hence, for every $h \in \mathbb{R}^n$,

$$\begin{aligned} P(x)h &= \text{argmin}_p\{\|p - h\|^2 \mid AXp = 0\}, && (13)\\ \tilde{P}_N(x_N)h_N &= \text{argmin}_{p_N}\{\|p_N - h_N\|^2 \mid AXp = 0\}. && (14)\\ \tilde{P}_B(x_B)h_B &= \text{argmin}_{p_B}\{\|p_B - h_B\|^2 \mid A_B X_B p_B = 0\}, && (15) \end{aligned}$$

Lemma 3.1 *For every $x \in \mathcal{P}^{++}$, let*

$$R(x) \equiv P(x) - \begin{pmatrix} \tilde{P}_N(x_N) & 0 \\ 0 & \tilde{P}_B(x_B) \end{pmatrix}. \tag{16}$$

Then the following statements hold:

(a) for every $x \in \mathcal{P}^{++}$, $R(x)$ is a symmetric matrix such that

$$R_{NN}(x) = \tilde{P}_N(x_N) R_{NN}(x) \tilde{P}_N(x_N), \quad R_{NB}(x) = \tilde{P}_N(x_N) R_{NB}(x); \tag{17}$$

(b) there exists a constant $C_1 > 0$ such that for every $x \in \mathcal{P}^{++}$ with $\Delta_N(x) \equiv \|x_N\| \, \|x_B^{-1}\|$ sufficiently small, there hold

$$\begin{aligned} \|R_{NN}(x)\| &\leq C_1 \Delta_N(x)^2, \\ \|R_{BB}(x)\| &\leq C_1 \Delta_N(x)^2, \\ \|R_{BN}(x)\| &= \|R_{NB}^T(x)\| \leq C_1 \Delta_N(x). \end{aligned}$$

Throughout this section, we make a local assumption that will be used to show that Algorithm AS converges to an optimal solution of problem (1) (see Theorem 3.6). This assumption imposes conditions on the way the sequences $\{x^k\}$ and $\{y^k, s^k\}$ generated by Algorithm AS behave. In Section 4, we show that it also holds without assuming any nondegeneracy assumption on problem (1).

Before stating the local assumption, we introduce the following notation that will be used throughout our presentation.

$$\begin{array}{rcl} g^* & = & \nabla f(x^*), \\ \tilde{s}^k & = & g^* - A^T[A(X^k)^2A^T]^{-1}A(X^k)^2g^*, \\ \tilde{d}^k & = & X^kP(x^k)X^kg^* = (X^k)^2\tilde{s}^k. \\ N_* & = & \{i \mid x_i^* = 0\}, \qquad B_* \;=\; \{i \mid x_i^* > 0\}, \end{array} \tag{18}$$

Note that $\tilde{d}^k$ and $\tilde{s}^k$ are the search direction and the dual estimate corresponding to the affine scaling algorithm applied to the LP problem whose objective function is $(g^*)^Tx = (\lim_{k\to\infty} \nabla f(x^k))^Tx$.

Local Assumption 3.2 *There exists a partition (B, N) of $\{1, \ldots, n\}$ such that $N \subseteq N_*$ and the following properties are satisfied:*

(A) the sequence $\{\Delta_N^k\}$ defined as $\Delta_N^k \equiv \|x_N^k\| \, \|(x_B^k)^{-1}\|$ for all $k \geq 0$ satisfies $\sum_{k=0}^{\infty} \Delta_N^k < \infty$;

(B) there exists $\tau_0 > 0$ such that $(g^*)^T(x^k - x^*) \geq \tau_0 \|x_N^k\|$ for every k sufficiently large;

(C) the sequences $\{u^k\}$ and $\{\tilde{u}^k\}$ defined as

$$u^k \equiv \frac{(X^k)^{-1}d^k}{(g^*)^T(x^k - x^*)}, \qquad \tilde{u}^k \equiv \frac{(X^k)^{-1}\tilde{d}^k}{(g^*)^T(x^k - x^*)}, \qquad \forall k \geq 0, \tag{19}$$

satisfy the following properties: $\{u^k - \tilde{u}^k\}$ is bounded and $u_N^k - \tilde{u}_N^k = \mathcal{O}(\Delta_N^k)$;

(D) $\lim_{k\to\infty} s_i^k = 0$ for every $i \in N_* \backslash N$.

Our goal in the remaining of this section is to show that Local Assumption 3.2 is a sufficient condition for $\{x^k\}, \{(y^k, s^k)\}$ to converge to optimal solutions of (1) and (2). We start with the following lemma.

Lemma 3.3 *The following properties hold:*

(a) $\{\tilde{u}^k\}$ is bounded;

(b) $|e^T\tilde{u}_N^k - 1| = \mathcal{O}((\Delta_N^k)^2)$ (and hence $\|\tilde{u}_N^k\| \geq 1/|N| - \mathcal{O}((\Delta_N^k)^2)$;

(c) $\|\tilde{u}_B^k\| = \mathcal{O}(\Delta_N^k)$.

Proof. First note that

$$(X^k)^{-1}\tilde{d}^k = P(x^k)X^kg^* = P(x^k)X^k\bar{s}, \tag{20}$$

for any $(\bar{y}, \bar{s}) \in \mathcal{D}(x^*)$. Let $(\bar{y}, \bar{s})$ be an accumulation point of the sequence $\{(y^k, s^k)\}$. Clearly, $(\bar{y}, \bar{s}) \in \mathcal{D}(x^*)$ and, by Local Assumption 3.2(D) and Lemma 2.4(e), we have $\bar{s}_B = 0$. Then, by (19), (20) and Local Assumption 3.2(B), we obtain

$$\|\tilde{u}^k\| = \frac{\|P(x^k)X^k\bar{s}\|}{(g^*)^T(x^k - x^*)} \leq \frac{\|X^k\bar{s}\|}{(g^*)^T(x^k - x^*)} \leq \frac{\|\bar{s}_N\|\|x_N^k\|}{(g^*)^T(x^k - x^*)} \leq \mathcal{O}(1),$$

and hence (a) follows. By (20), Lemma 3.1 and the fact that $\bar{s}_B = 0$, we obtain

$$(X^k)^{-1}\tilde{d}^k = P(x^k)X^k\bar{s} = \begin{pmatrix} \left\{\tilde{P}_N(x_N^k) + R_{NN}(x^k)\right\} X_N^k \bar{s}_N \\ R_{BN}(x^k)X_N^k\bar{s}_N \end{pmatrix}, \tag{21}$$

where $\|R_{BN}(x^k)\| = \mathcal{O}(\Delta_N^k)$ and $\|R_{NN}(x^k)\| = \mathcal{O}((\Delta_N^k)^2)$. Hence,

$$\|\tilde{u}_B^k\| = \frac{\|R_{BN}(x^k)X_N^k\bar{s}_N\|}{(g^*)^T(x^k - x^*)} \leq \frac{\mathcal{O}(\|x_N^k\|\Delta_N^k)}{(g^*)^T(x^k - x^*)} \leq \mathcal{O}(\Delta_N^k),$$

where the last inequality follows from Local Assumption 3.2(B). This shows (c). To show (b), we claim that

$$e^T\tilde{P}_N(x_N^k)X_N^k\bar{s}_N = (g^*)^T(x^k - x^*), \qquad \forall k \geq 0. \tag{22}$$

Before showing (22), we show how (22) implies (b). Indeed, by (19), (21) and (22), we obtain

$$|e^T\tilde{u}_N^k - 1| = \left|\frac{e^T\left[(X_N^k)^{-1}\tilde{d}_N^k - \tilde{P}_N(x_N^k)X_N^k\bar{s}_N\right]}{(g^*)^T(x^k - x^*)}\right| = \left|\frac{e^T R_{NN}(x^k)X_N^k\bar{s}_N}{(g^*)^T(x^k - x^*)}\right| \leq \mathcal{O}((\Delta_N^k)^2),$$

where the last inequality follows from Local Assumption 3.2(B) and the fact that $\|R_{NN}(x^k)\| = \mathcal{O}((\Delta_N^k)^2)$. We have thus shown (b). It remains to verify (22). Let $p_N^k \equiv P_N(x_N^k)X_N^k\bar{s}_N$. Since p_N^k solves problem (14) with $x_N = x_N^k$ and $h_N = X_N^k\bar{s}_N$, the optimality condition for this problem yields

$$(A_B)^T\hat{y}^k = 0, \quad X_N^k(A_N)^T\hat{y}^k = p_N^k - X_N^k\bar{s}_N, \tag{23}$$

for some vector $\hat{y}^k \in \mathbb{R}^m$. Since $b = A_{B_*}x_{B_*}^* \in \text{Range}(A_B)$, it follows that $A_N x_N^k = b - A_B x_B^k \in \text{Range}(A_B)$. Using this observation and the first equality in (23), we obtain $e^T[X_N^k(A_N)^T\hat{y}^k] = (A_N x_N^k)^T\hat{y}^k = 0$. Hence, the second equality in (23) yields

$$e^T p_N^k = \bar{s}_N^T x_N^k = (g^*)^T(x^k - x^*),$$

where the last equality follows from Lemma 2.5. This shows the claim. ∎

The following potential function, usually referred to as th local Karmarkar potential function, plays an important role in our analysis. It was introduced by Tsuchiya to analyze the affine scaling algorithms for LP in [19] and for QP in [22]. For every $x \in \mathcal{P}_{++}$ satisfying $(g^*)^T(x - x^*) > 0$, let

$$\psi_N(x) = |N| \log[(g^*)^T(x - x^*)] - \sum_{i \in N} \log x_i. \tag{24}$$

The following result provides an upper bound for the difference $\psi_N(x^{k+1}) - \psi_N(x^k)$.

Lemma 3.4 *Suppose that Local Assumption 3.2 hold and that $|N| \geq 2$. Then, there exists $\tau > 0$ such that for every k sufficiently large,*

$$\psi_N(x^{k+1}) - \psi_N(x^k) \leq -\tau\lambda_k\eta_k^2\left[1 - \frac{\lambda_k}{2(1-\lambda_k)}(1+\eta_k)^{-1}\right] + \mathcal{O}(\Delta_N^k) \leq \mathcal{O}(\Delta_N^k), \tag{25}$$

where

$$\lambda_k \equiv \frac{\beta\,\chi(\tilde{u}^k)}{\|u^k\|}, \text{ and} \qquad \eta_k \equiv \left\|\frac{\tilde{u}_N^k}{e^T\tilde{u}_N^k} - \frac{1}{|N|}e\right\|. \tag{26}$$

Proof. Let (y^*, s^*) be an accumulation point of $\{(y^k, s^k)\}$. Observe that Local Assumption 3.2(D) implies that $\{i \mid s_i^* \neq 0\} \subseteq N$. Hence, it follows from Lemma 2.5 that $(g^*)^T(x^k - x^*) = (s_N^*)^T x_N^k$ for every $k \geq 0$. Using Local Assumption 3.2(B), the above observation and relation (5), we obtain

$$\frac{(g^*)^T(x^{k+1} - x^*)}{(g^*)^T(x^k - x^*)} \geq \frac{\tau_0 \|x_N^{k+1}\|}{(s_N^*)^T x_N^k} \geq \frac{\tau_0}{\|s_N^*\|} \frac{\|x_N^{k+1}\|}{\|x_N^k\|} \geq \frac{\tau_0 (1-\beta)}{\|s_N^*\|}. \tag{27}$$

Using relation (10), the fact that $(g^*)^T \tilde{d}^k = \|(X^k)^{-1}\tilde{d}^k\|^2$ for all k, the definition of λ_k, u^k and $\tilde{u}^k$, Lemma 3.3(b) and conditions (B) and (C) of Local Assumption 3.2, we obtain

$$\begin{aligned}
\frac{(g^*)^T(x^{k+1} - x^*)}{(g^*)^T(x^k - x^*)} &= 1 - \frac{\beta}{(g^*)^T(x^k - x^*)} \frac{(g^*)^T d^k}{\|(X^k)^{-1} d^k\|} \\
&= 1 - \frac{\beta}{(g^*)^T(x^k - x^*)} \frac{(g^*)^T \tilde{d}^k}{\|(X^k)^{-1} d^k\|} + \frac{\beta}{(g^*)^T(x^k - x^*)} \frac{(g^*)^T(\tilde{d}^k - d^k)}{\|(X^k)^{-1} d^k\|} \\
&= 1 - \frac{\beta}{(g^*)^T(x^k - x^*)} \frac{\|(X^k)^{-1}\tilde{d}^k\|^2}{\|(X^k)^{-1} d^k\|} + \frac{\beta}{(g^*)^T(x^k - x^*)} \frac{(s_N^*)^T(\tilde{d}_N^k - d_N^k)}{\|(X^k)^{-1} d^k\|} \\
&= 1 - \beta \frac{\|\tilde{u}^k\|^2}{\|u^k\|} + \frac{\beta}{(g^*)^T(x^k - x^*)} \frac{(X_N^k s_N^*)^T(\tilde{u}_N^k - u_N^k)}{\|u^k\|} \\
&= 1 - \lambda_k \frac{\|\tilde{u}^k\|^2}{\chi(\tilde{u}^k)} + \mathcal{O}\left(\frac{\|x_N^k\|}{(g^*)^T(x^k - x^*)} \frac{\|\tilde{u}_N^k - u_N^k\|}{\|u_N^k\|} \right) \\
&= 1 - \lambda_k \frac{\|\tilde{u}^k\|^2}{\chi(\tilde{u}^k)} + \mathcal{O}(\Delta_N^k).
\end{aligned} \tag{28}$$

Using relations (5) and (10), the definition of λ_k, u^k and $\tilde{u}^k$ and Local Assumption 3.2(C), we obtain for every $i \in N$ that

$$\begin{aligned}
1 - \beta \leq \frac{x_i^{k+1}}{x_i^k} &= 1 - \beta \frac{(x_i^k)^{-1} d_i^k}{\|(X^k)^{-1} d^k\|} \\
&= 1 - \beta \frac{(x_i^k)^{-1} \tilde{d}_i^k}{\|(X^k)^{-1} d^k\|} + \beta \frac{(x_i^k)^{-1}(\tilde{d}_i^k - d_i^k)}{\|(X^k)^{-1} d^k\|} \\
&= 1 - \beta \frac{\tilde{u}_i^k}{\|u^k\|} + \beta \frac{\tilde{u}_i^k - u_i^k}{\|u^k\|} \\
&= 1 - \lambda_k \frac{\tilde{u}_i^k}{\chi(\tilde{u}^k)} + \mathcal{O}\left(\frac{\|\tilde{u}_N^k - u_N^k\|}{\|u^k\|} \right) \\
&= 1 - \lambda_k \frac{\tilde{u}_i^k}{\chi(\tilde{u}^k)} + \mathcal{O}(\Delta_N^k).
\end{aligned} \tag{29}$$

Relations (27), (28) and (29) and the expression

$$\psi_N(x^{k+1}) - \psi_N(x^k) = |N| \log \left(\frac{(g^*)^T(x^{k+1} - x^*)}{(g^*)^T(x^k - x^*)} \right) - \sum_{i \in N} \log \frac{x_i^{k+1}}{x_i^k}$$

then imply that

$$\psi_N(x^{k+1}) - \psi_N(x^k) \leq |N| \log \left(1 - \lambda_k \frac{\|\tilde{u}^k\|^2}{\chi(\tilde{u}^k)} \right) - \sum_{i \in N} \log \left(1 - \lambda_k \frac{\tilde{u}_i^k}{\chi(\tilde{u}^k)} \right) + \mathcal{O}(\Delta_N^k). \tag{30}$$

We will now show that $\{\tilde{u}^k\}$ and $\{\lambda_k\}$ satisfy all the conditions (a)-(e) of Proposition A.4. Indeed, conditions (b), (c) and (d) follows immediately from Lemma 3.3. Condition (a) follows from (27), (28) and the fact that $\lim_{k\to\infty} \Delta_N^k = 0$, in view of Local Assumption 3.2(A). It remains to verify that (e) holds. Assume for contradiction that (e) does not hold. Then there exists an infinite index set $\mathcal{K}$ such that

$$\frac{\lambda_k}{2(1-\lambda_k)} > 1 + \eta_k, \qquad \forall k \in \mathcal{K}. \tag{31}$$

Clearly, this implies that $\liminf_{k\in\mathcal{K}} \lambda_k \geq 2/3$. Noting that Lemma 3.3 and conditions (A) and (C) of Local Assumption 3.2 imply that $\lim_{k\to\infty} \|\tilde{u}_B^k\| = 0$, $\lim_{k\to\infty} e^T u_N^k = 1$, and $\lim_{k\to\infty} \|u_N^k - \tilde{u}_N^k\| = 0$, it is easy to see that $\limsup_{k\to\infty} \|\tilde{u}^k\|/\|u^k\| \leq 1$. This fact and the inequalities $\beta \leq 2/3$ and $\chi(u) \leq \|u\|$, for any vector u, yield

$$\limsup_{k\to\infty} \lambda_k = \beta \limsup_{k\to\infty} \frac{\chi(\tilde{u}^k)}{\|u^k\|} \leq \frac{2}{3} \limsup_{k\to\infty} \frac{\chi(\tilde{u}^k)}{\|\tilde{u}^k\|} \leq \frac{2}{3}. \tag{32}$$

This expression together with $\liminf_{k\in\mathcal{K}} \lambda_k \geq 2/3$ imply

$$\lim_{k\in\mathcal{K}} \lambda_k = 2/3, \text{ and } \lim_{k\in\mathcal{K}} \frac{\chi(\tilde{u}^k)}{\|\tilde{u}^k\|} = 1.$$

The first relation in the last expression together with (31) then imply that $\lim_{k\in\mathcal{K}} \eta_k = 0$, that is,

$$\lim_{k\in\mathcal{K}} \frac{\tilde{u}_N^k}{e^T \tilde{u}_N^k} = \frac{1}{|N|} e.$$

It is now easy to see that this relation and the fact that $\lim_{k\to\infty} \|\tilde{u}_B^k\| = 0$ and $\lim_{k\to\infty} e^T \tilde{u}_N^k = 1$ contradicts the fact that $\lim_{k\in\mathcal{K}} \chi(\tilde{u}^k)/\|\tilde{u}^k\| = 1$ when $|N| \geq 2$. Hence, we conclude that $\{\tilde{u}^k\}$ and $\{\lambda_k\}$ satisfy all the conditions (a)-(e) of Proposition A.4. It follows from Proposition A.4, relation (30), Lemma 3.3 and Local Assumption 3.2(A) that for all k sufficiently large,

$$\begin{aligned}
&\psi_N(x^{k+1}) - \psi_N(x^k) \\
&\quad \leq \quad -\tau\lambda_k\eta_k^2 \left[1 - \frac{\lambda_k}{2(1-\lambda_k)}(1+\eta_k)^{-1}\right] + \mathcal{O}(\Delta_N^k) + \mathcal{O}\left(|e^T\tilde{u}_N^k - 1| + \|\tilde{u}_B^k\|^2\right) \\
&\quad = \quad -\tau\lambda_k\eta_k^2 \left[1 - \frac{\lambda_k}{2(1-\lambda_k)}(1+\eta_k)^{-1}\right] + \mathcal{O}(\Delta_N^k),
\end{aligned}$$

This shows the first inequality in (25). The second inequality in (25) follows from the fact that $\lambda_k/[2(1-\lambda_k)] \leq 1 + \eta_k$ for every k sufficiently large. ∎

The next result shows that $\lim_{k\to\infty} u_N^k = e/|N|$.

Lemma 3.5 *Suppose that Local Assumption 3.2 hold. Then,*

$$\lim_{k\to\infty} u_N^k = \frac{1}{|N|} e. \tag{33}$$

Proof. If $|N| = 1$ then (33) follows immediately from Lemma 3.3(b) and the fact that, by Local Assumption 3.2(C), $\lim_{k\to\infty} u_N^k - \tilde{u}_N^k = 0$. We now consider the case in which $|N| \geq 2$. First observe that the sequence $\{\psi_N(x^k)\}$ is bounded below since

$$\begin{aligned}\psi_N(x^k) &= |N| \log\left(\frac{(g^*)^T(x^k - x^*)}{\sum_{i\in N} x_i^k}\right) + |N| \log(\sum_{i\in N} x_i^k) - \sum_{i\in N} \log x_i^k \\ &\geq |N| \log\left(\frac{\tau_0}{|N|^{1/2}}\right) + |N| \log|N|, \end{aligned} \tag{34}$$

where the last inequality is due to the inequalities $\|x_N^k\| \geq (\sum_{i\in N} x_i^k)/\sqrt{|N|}$ and $(\sum_{i\in N} x_i^k)/|N| \geq (\prod_{i\in N} x_i^k)^{1/|N|}$ and Local Assumption 3.2(B). We next show that the sequence $\{\eta_k\}$ defined in (26) converges to 0. Assume for contradiction that, for some $\epsilon > 0$ and some infinite index set $\mathcal{K}$, we have $\eta_k \geq \epsilon$ for all $k \in \mathcal{K}$. We claim that $\limsup_{k\in\mathcal{K}} \psi_N(x^{k+1}) - \psi(x_N^k) < 0$. Indeed, due to relation (32), we have $\limsup_{k\to\infty} \lambda_k \leq 2/3$, and hence

$$\liminf_{k\in\mathcal{K}} \left(1 - \frac{\lambda_k}{2(1-\lambda_k)}(1+\eta_k)^{-1}\right) \geq 1 - (1+\epsilon)^{-1} > 0.$$

Moreover, using (a) and (b) of Lemma 3.3 and (26), it is easy to see that $\liminf_{k\to\infty} \lambda_k > 0$. These two last observations, the fact that $\lim_{k\to\infty} \Delta_N^k = 0$ and Lemma 3.4 imply that the claim holds. By reducing $\mathcal{K}$ if necessary, we may assume that, for some constant $\gamma > 0$, $\psi_N(x^{k+1}) - \psi(x_N^k) \leq -\gamma$ for all $k \in \mathcal{K}$. Hence, it follows that

$$\sum_{k\in\mathcal{K}} \psi_N(x^{k+1}) - \psi(x_N^k) = -\infty. \tag{35}$$

Moreover, Lemma 3.4 implies that, for some constant $L_0 > 0$, $\psi_N(x^{k+1}) - \psi(x_N^k) \leq L_0\Delta_N^k$ for all $k \geq 0$. This implies that

$$\sum_{k\notin\mathcal{K}} \psi_N(x^{k+1}) - \psi_N(x_N^k) \leq L_0 \sum_{k\notin\mathcal{K}} \Delta_N^k < \infty,$$

where the last inequality is due to Local Assumption 3.2(A). Combining this last relation with (35), we conclude that

$$\sum_{k=0}^{\infty} \psi_N(x^{k+1}) - \psi_N(x_N^k) = -\infty,$$

and hence, $\lim_{k\to\infty} \psi_N(x_N^k) = -\infty$. But this contradicts (34) and hence we conclude that $\{\eta_k\}$ converges to 0. Using (26), the fact that, by Lemma 3.3(b), $\lim_{k\to\infty} e^T\tilde{u}_N^k = 1$ and, by Local Assumption 3.2(C), $\lim_{k\to\infty} u_N^k - \tilde{u}_N^k = 0$, we can now easily see that (33) holds. ∎

The next result establishes that the sequence $\{x^k\}$ generated by Algorithm AS converges to an optimal solution of problem (1) and that the sequence of dual estimates $\{(y^k, s^k)\}$ converges to an optimal solution of (2).

Theorem 3.6 *Suppose that the partition (B, N) satisfies Local Assumption 3.2. Then the limit point x^* of the sequence $\{x^k\}$ generated by Algorithm AS is an optimal solution of problem (1) and the associated sequence of dual estimates $\{(y^k, s^k)\}$ converges to an optimal solution (y^*, s^*)*

of the dual problem (2) having the following property: (y^*, s^*) is the unique optimal solution of the problem

$$\begin{array}{ll} \text{maximize} & \sum_{i\in N} \log s_i \\ \text{subject to} & A^T y + s = g^*, \\ & s_B = 0,\ s_N > 0. \end{array} \tag{36}$$

Proof. Let (y^*, s^*) denote the optimal solution (y^*, s^*) of (36). Since A has full row rank, it follows that if $\lim_{k\to\infty} s^k = s^*$ then $\lim_{k\to\infty} y^k = y^*$. Hence, it is sufficient to show that $\lim_{k\to\infty} s^k = s^*$. It can be easily verified that s^* is the unique point satisfying the following relations:

$$s_N > 0,\ s_B = 0; \tag{37}$$

$$s \in g^* + \text{Range}(A^T); \tag{38}$$

$$A_N(s_N)^{-1} \in \text{Range}(A_B), \tag{39}$$

where $(s_N)^{-1}$ denotes the subvector $[(s_i)^{-1}]_{i\in N}$. We will show that every accumulation point of the (bounded) sequence $\{s^k\}$ satisfies relations (37), (38) and (39), thereby showing that $\lim_{k\to\infty} s^k = s^*$. Indeed, let s^∞ denote an accumulation point of $\{s^k\}$, that is, $s^\infty = \lim_{k\in\mathcal{K}} s^k$ for some infinite index set $\mathcal{K}$. By Lemma 2.4(e) and Local Assumption 3.2(D), we know that $s^\infty_B = 0$. Letting k tend to ∞ in the relation $s^k \in g^{k+1} + \text{Range}(A^T)$, we conclude that $s^\infty \in g^* + \text{Range}(A^T)$. By Lemma 3.5, we have

$$\lim_{k\to\infty} u^k_N = \lim_{k\to\infty} \frac{X^k_N s^k_N}{(g^*)^T(x^k - x^*)} = \frac{1}{|N|} e,$$

where $X^k_N = \text{diag}(x^k_N)$, and by Local Assumption 3.2(B), we know that $\|x^k_N\|/[(g^*)^T(x^k - x^*)]$ is bounded. From these two observations, we can easily see that $s^\infty_N > 0$ and that

$$\lim_{k\in\mathcal{K}} \frac{|N|}{(g^*)^T(x^k - x^*)} x^k_N = (s^\infty_N)^{-1}. \tag{40}$$

Since $b \in \text{Range}(A_{B_*}) \subseteq \text{Range}(A_B)$ and $Ax^k = b$, we conclude that

$$A_N \left(\frac{|N|}{(g^*)^T(x^k - x^*)} x^k_N \right) \in \text{Range}(A_B), \qquad \forall k \geq 0.$$

Letting $k \in \mathcal{K}$ tend to ∞ in this relation and using (40), we obtain that

$$A_N(s^\infty_N)^{-1} \in \text{Range}(A_B).$$

We have thus shown that s^∞ satisfies (37), (38) and (39), and hence it follows that $s^\infty = s^*$. ∎

4 Global convergence proof

The main goal of this section is to show the existence of the partition (B, N) satisfying the conditions (A)-(D) of Assumption 3.2.

The following technical lemma is easy to verify and is frequently used in our presentation.

Lemma 4.1 *Let $F \in \mathbf{R}^{p\times q}$ be given. Then, there exists a constant $C_2 = C_2(F)$ with the following property: for any $f \in \mathbb{R}^p$ such that the system $Fw = f$ is feasible and any $z \in \mathbf{R}^q$, there exists a solution $\bar{w}$ of $Fw = f$ such that*

$$\|\bar{w} - z\| \leq C_2\|f - Fz\|.$$

Our next goal is to show that for index set N such that $N \subseteq N_*$, we have $f(x^k) - f(x^*) = \mathcal{O}(\|x_N^k\|)$ (see Theorem 4.4). This result plays a fundamental role in the analysis of this section.

We use the following notation. Given a point $x \in \mathcal{P}^{++}$, let

$$\mathcal{E}(x) = \mathcal{E}(x;\beta) \equiv \{w \in \mathcal{P} \mid \|X^{-1}(w - x)\| \leq \beta\}. \tag{41}$$

Lemma 4.2 $f(x^{k+1}) - f(x^*) = \mathcal{O}(\|X^k s^k\|)$.

Proof. Consider the point $\tilde{x}^{k+1}$ defined as

$$\tilde{x}^{k+1} \equiv x^k + \frac{\beta(x^* - x^k)}{\|(X^k)^{-1}(x^* - x^k)\|}.$$

Clearly, $\tilde{x}^{k+1} \in \mathcal{E}(x^k)$ (see (41)). Since x^{k+1} is the optimal solution of (4), we know that $(g^{k+1})^T(x - x^{k+1}) \geq 0$ for all $x \in \mathcal{E}(x^k)$. In particular, $x = \tilde{x}^{k+1}$ yields

$$(g^{k+1})^T\left(x^k - x^{k+1} + \frac{\beta\,(x^* - x^k)}{\|(X^k)^{-1}(x^* - x^k)\|}\right) \geq 0,$$

or equivalently,

$$\|(X^k)^{-1}(x^* - x^k)\| - \beta \geq \frac{\beta\,(g^{k+1})^T(x^{k+1} - x^*)}{(g^{k+1})^T(x^k - x^{k+1})}. \tag{42}$$

It is easy to see that $\lim_{k\to\infty} \|(X^k)^{-1}(x^* - x^k)\| = |N_*|^{1/2}$. This relation together with (12), (42) and the inequality $f(x^{k+1}) - f(x^*) \leq (g^{k+1})^T(x^{k+1} - x^*)$ imply

$$\begin{aligned}\limsup_{k\to\infty} \frac{f(x^{k+1}) - f(x^*)}{\|X^k s^k\|} &\leq \beta \limsup_{k\to\infty} \frac{(g^{k+1})^T(x^{k+1} - x^*)}{(g^{k+1})^T(x^k - x^{k+1})} \\ &\leq \limsup_{k\to\infty} \left(\|(X^k)^{-1}(x^* - x^k)\| - \beta\right) = |N_*|^{1/2} - \beta,\end{aligned}$$

and the result follows. ∎

Lemma 4.3 *Assume that $\{a_k\}$ is an unbounded sequence of positive scalars such that $a_{k+1} = \mathcal{O}(a_k)$. Then there exists a subsequence $\{a_k\}_{k\in\mathcal{K}}$ such that $\lim_{k\in\mathcal{K}} a_k = \infty$ and $a_k \leq a_{k+1}$ for all $k \in \mathcal{K}$.*

Proof. Let $\mathcal{K}_0 \equiv \{k \mid a_{k-1} \leq a_k\}$. We first show that the subsequence $\{a_k\}_{k\in\mathcal{K}_0}$ is unbounded. Indeed, assume for contradiction that, for some $L_0 > 0$, $a_k \leq L_0$ for every $k \in \mathcal{K}_0$. Let k_0 denote the smallest index in $\mathcal{K}_0$. We claim that $a_k \leq L_0$ for every $k \geq k_0$, a fact that contradicts the assumption that $\{a_k\}$ is unbounded. To prove the claim, let $k \geq k_0$ be given. Clearly, if $k \in \mathcal{K}_0$ then $a_k \leq L_0$. If $k \notin \mathcal{K}_0$ then let $j(k)$ denote the largest index $j \in \mathcal{K}_0$ such that $j < k$. Since $j(k) \in \mathcal{K}_0$ and $j(k) + 1, \ldots, k \notin \mathcal{K}_0$, we have $L_0 \geq a_{j(k)} > a_{j(k)+1} > \ldots > a_k$. This shows that

$a_k \leq L_0$ for every $k \geq k_0$ and, hence, that $\{a_k\}_{k\in\mathcal{K}_0}$ is unbounded. Hence, there exists an infinite index set $\mathcal{K}$ such that $a_k \leq a_{k+1}$ for every $k \in \mathcal{K}$ and $\lim_{k\in\mathcal{K}} a_{k+1} = \infty$. Since $a_{k+1} = \mathcal{O}(a_k)$, it also follows that $\lim_{k\in\mathcal{K}} a_k = \infty$. Hence, the result follows. ∎

Now we are ready to see the theorem.

Theorem 4.4 *Let $N \subseteq N_*$ be given. Then the following statements hold:*

(a) $f(x^k) - f(x^*) = \mathcal{O}(\|x_N^k\|)$;

(b) $\|x^k - x^*\|^2 = \mathcal{O}(\|x_N^k\|)$.

Proof. Observe that (b) follows as a consequence of (a) and Lemma 2.4(c). We next show (a). Assume for contradiction that $a_k \equiv [f(x^k) - f(x^*)]/\|x_N^k\|$ is unbounded. Clearly, we have $a_{k+1} \leq (1-\beta)^{-1} a_k = \mathcal{O}(a_k)$. Hence, it follows from Lemma 4.3 that there exists an infinite index set $\mathcal{K}$ such that $\lim_{k\in\mathcal{K}} a_k = \infty$ and $a_k \leq a_{k+1}$ for all $k \in \mathcal{K}$. Using the definition of a_k, the last inequality and Lemma 2.4(b), we obtain

$$\frac{\|x_N^{k+1}\|}{\|x_N^k\|} \leq \frac{f(x^{k+1}) - f(x^*)}{f(x^k) - f(x^*)} \leq \gamma_0 < 1, \qquad \forall k \in \mathcal{K}, \tag{43}$$

where $\gamma_0 = 1 - \beta/\sqrt{n} < 1$. It is easy to see that this relation implies that

$$\liminf_{k\in\mathcal{K}} \|(X_N^k)^{-1}(x_N^{k+1} - x_N^k)\| > 0. \tag{44}$$

We can easily verify the existence of an index set $\hat{N} \subseteq \{1, \ldots, n\}$ and an infinite index set $\mathcal{K}_0 \subseteq \mathcal{K}$ such that

$$\|x_{\hat{N}}^k\| = \mathcal{O}(\|x_N^k\|), \qquad \forall k \in \mathcal{K}_0, \tag{45}$$

$$\lim_{k\in\mathcal{K}_0} \frac{x_i^k}{\|x_N^k\|} = \infty, \qquad \forall i \notin \hat{N}. \tag{46}$$

These relations imply that $\lim_{k\in\mathcal{K}} \Delta_{\hat{N}}^k = 0$. Clearly, $N \subseteq \hat{N} \subseteq N_*$. In view of (45) and the fact that $\lim_{k\in\mathcal{K}} a_k = \infty$, we obtain

$$\lim_{k\in\mathcal{K}_0} \frac{f(x^k) - f(x^*)}{\|x_{\hat{N}}^k\|} = \infty. \tag{47}$$

Using Lemma 4.1 with $z = x^{k+1}$, it follows that for every $k \geq 0$ the feasible system (e.g., $x = x^k$ is a feasible solution)

$$Ax = b, \qquad x_{\hat{N}} = x_{\hat{N}}^k \tag{48}$$

has a solution $\hat{x}^{k+1}$ such that

$$\|\hat{x}^{k+1} - x^{k+1}\| = \mathcal{O}\left(\|x_{\hat{N}}^k - x_{\hat{N}}^{k+1}\|\right) = \mathcal{O}\left(\|x_{\hat{N}}^k\|\right). \tag{49}$$

We will now show that the half line starting at x^k and passing through the point $\hat{x}^{k+1}$ contains a point in $\mathcal{E}(x^k)$ whose f–function value is smaller than $f(x^{k+1})$ whenever $k \in \mathcal{K}_0$ is sufficiently large, a fact that that contradicts the definition of x^{k+1}. We first show that

$$\limsup_{k\in\mathcal{K}_0} \|(X^k)^{-1}(\hat{x}^{k+1} - x^k)\| < \beta. \tag{50}$$

Indeed, the fact that $\hat{x}^{k+1}_{\hat{N}} = x^k_{\hat{N}}$ implies

$$\begin{aligned}
\|(X^k)^{-1}(\hat{x}^{k+1} - x^k)\|^2 &= \|(X^k_{\hat{B}})^{-1}(\hat{x}^{k+1}_{\hat{B}} - x^k_{\hat{B}})\|^2 \\
&\leq \|(X^k_{\hat{B}})^{-1}(x^{k+1}_{\hat{B}} - x^k_{\hat{B}})\|^2 + \|(X^k_{\hat{B}})^{-1}(\hat{x}^{k+1}_{\hat{B}} - x^{k+1}_{\hat{B}})\|^2 \\
&\quad + 2\,\|(X^k_{\hat{B}})^{-1}(\hat{x}^{k+1}_{\hat{B}} - x^{k+1}_{\hat{B}})\|\,\|(X^k_{\hat{B}})^{-1}(x^{k+1}_{\hat{B}} - x^k_{\hat{B}})\| \\
&\leq \beta^2 - \|(X^k_{\hat{N}})^{-1}(x^{k+1}_{\hat{N}} - x^k_{\hat{N}})\|^2 + \|(X^k_{\hat{B}})^{-1}(\hat{x}^{k+1}_{\hat{B}} - x^{k+1}_{\hat{B}})\|^2 \\
&\quad + 2\beta\,\|(X^k_{\hat{B}})^{-1}(\hat{x}^{k+1}_{\hat{B}} - x^{k+1}_{\hat{B}})\|. \qquad (51)
\end{aligned}$$

By (49) and the fact that $\lim_{k\in\mathcal{K}} \Delta^k_{\hat{N}} = 0$, we have

$$\lim_{k\in\mathcal{K}_0} \|(X^k_{\hat{B}})^{-1}(\hat{x}^{k+1}_{\hat{B}} - x^{k+1}_{\hat{B}})\| \leq \lim_{k\in\mathcal{K}_0} \|(X^k_{\hat{B}})^{-1}\|\,\|\hat{x}^{k+1}_{\hat{B}} - x^{k+1}_{\hat{B}}\| = \lim_{k\in\mathcal{K}_0} \mathcal{O}(\|(X^k_{\hat{B}})^{-1}\|\|x^k_{\hat{N}}\|) = 0, \quad (52)$$

where the last equality follows from (45) and (46). Combining (44), (51) and (52), we obtain (50).

For $\lambda \geq 0$, consider the point $x^{k+1}(\lambda) \equiv x^k + \lambda(\hat{x}^{k+1} - x^k)$. It follows from (50) that there exists $\epsilon > 0$ such that $x^{k+1}(\lambda) \in \mathcal{E}(x^k)$ for every $\lambda \in [0, 1+\epsilon]$ and $k \in \mathcal{K}_0$ sufficiently large. We now estimate the difference $f(x^{k+1}) - f(x^{k+1}(\lambda))$. It is easy to see that

$$\begin{aligned}
f(x^{k+1}) - f(x^{k+1}(\lambda)) &= f(x^{k+1}) - f(x^k) + f(x^k) - f(x^{k+1}(\lambda)) \\
&= \nabla f(x^k)^T(x^{k+1} - x^{k+1}(\lambda)) + \frac{1}{2}\|x^{k+1} - x^k\|^2_Q - \frac{1}{2}\|x^{k+1}(\lambda) - x^k\|^2_Q
\end{aligned} \qquad (53)$$

Using (49) and the definition of $x^{k+1}(\lambda)$, we obtain

$$\begin{aligned}
\nabla f(x^k)^T(x^{k+1} - x^{k+1}(\lambda)) &= (1-\lambda)\nabla f(x^k)^T(x^{k+1} - x^k) + \lambda \nabla f(x^k)^T(x^{k+1} - \hat{x}^{k+1}) \\
&\geq (\lambda - 1)\nabla f(x^k)^T(x^k - x^{k+1}) - \mathcal{O}(\|x^k_{\hat{N}}\|) \\
&= (\lambda - 1)\left[\|x^k - x^{k+1}\|^2_Q + \beta\|X^k s^k\|\right] - \mathcal{O}(\|x^k_{\hat{N}}\|), \qquad (54)
\end{aligned}$$

where the last equality follows from (12). Moreover,

$$\begin{aligned}
\|x^{k+1}(\lambda) - x^k\|^2_Q &= \lambda^2\|\hat{x}^{k+1} - x^k\|^2_Q \\
&\leq \lambda^2\left[\|\hat{x}^{k+1} - x^{k+1}\|^2_Q + \|x^{k+1} - x^k\|^2_Q + 2\|\hat{x}^{k+1} - x^{k+1}\|_Q\,\|x^{k+1} - x^k\|_Q\right] \\
&\leq \lambda^2\|x^{k+1} - x^k\|^2_Q + \mathcal{O}(\|x^k_{\hat{N}}\|), \qquad \forall \lambda \in [0, 1+\epsilon].
\end{aligned}$$

Combining (53), (54) and (55) and rearranging, we obtain that for every $\lambda \in (1, 1+\epsilon]$,

$$\frac{f(x^{k+1}) - f(x^{k+1}(\lambda))}{(\lambda-1)\|X^k s^k\|} \geq \left\{\beta - \frac{1}{2}(\lambda-1)\frac{\|x^{k+1} - x^k\|^2_Q}{\|X^k s^k\|} - \mathcal{O}\left(\frac{\|x^k_{\hat{N}}\|}{(\lambda-1)\|X^k s^k\|}\right)\right\}. \qquad (55)$$

Now, by (5), (43) and Lemma 4.2, it follows that

$$f(x^k) - f(x^*) \leq (1-\beta)^{-1}\left[f(x^{k+1}) - f(x^*)\right] = \mathcal{O}(\|X^k s^k\|), \qquad \forall k \in \mathcal{K} \supseteq \mathcal{K}_0. \qquad (56)$$

This relation, relation (47) and the identity $f(x^k) - f(x^{k+1}) = \beta\|X^k s^k\| + \|x^{k+1} - x^k\|_Q^2/2$ imply that

$$\lim_{k\in\mathcal{K}_0} \frac{\|x_{\hat{N}}^k\|}{\|X^k s^k\|} = 0, \qquad \frac{\|x^{k+1} - x^k\|_Q^2}{\|X^k s^k\|} \le L_0, \qquad \forall k \in \mathcal{K}_0, \tag{57}$$

where L_0 is some positive constant. Let $\bar{\lambda} \equiv 1 + \min\{\epsilon, \beta/L_0\}$. Then, it follows from (55) with $\lambda = \bar{\lambda}$ that

$$\frac{f(x^{k+1}) - f(x^{k+1}(\bar{\lambda}))}{(\bar{\lambda} - 1)\,\|X^k s^k\|} \ge \frac{\beta}{2} - \mathcal{O}\left(\frac{\|x_{\hat{N}}^k\|}{\|X^k s^k\|}\right) > 0, \qquad \forall k \in \mathcal{K}_0 \text{ sufficiently large.}$$

Hence, $f(x^{k+1}(\bar{\lambda})) < f(x^{k+1})$ for every $k \in \mathcal{K}_0$ sufficiently large. This conclusion together with the fact that $x^{k+1}(\bar{\lambda}) \in \mathcal{E}(x^k)$ contradicts the definition of x^{k+1}. ∎

The following set is frequently used in our presentation. Let

$$\Gamma \equiv \{N \mid N \subseteq N_* \subseteq \{1, \ldots, n\},\ \liminf_{k\to\infty} \Delta_N^k = 0\}. \tag{58}$$

We now give an outline for the remaining part of this section. The main result of this section is Theorem 4.15 which establishes global convergence of Algorithm AS. Its statement is very similar to the one of Theorem 3.6 except that it guarantees the existence of a unique partition $(\bar{B}, \bar{N})$ satisfying Local Assumption 3.2. Let $\bar{N}$ denote a minimal element of the set Γ with respect to inclusion and let $\bar{B} \equiv \{1, \ldots, n\}\backslash\bar{N}$. The main goal from now on is to show that $(\bar{B}, \bar{N})$ satisfies all the conditions of Local Assumption 3.2. In Theorem 4.15 we show that there can be only one partition satisfying Local Assumption 3.2 from which it also follows that the set Γ has only one minimal element, in other words, Γ has a smallest element with respect to inclusion. Theorem 4.13 shows that $(\bar{B}, \bar{N})$ satisfies condition (A) of Local Assumption 3.2. In fact, once this condition is verified, it is not difficult to obtain the other ones. We separate the proof of Theorem 4.13 into several lemmas. The major part of the proof are given in Lemma 4.10, Lemma 4.11 and Lemma 4.12. All the results between (and including) Lemma 4.5 and 4.9 are mainly technical ones.

Lemma 4.5 *Assume that an infinite index set $\mathcal{K}$ and an index set $N \subseteq \{1, \ldots, n\}$ are such that $N \subseteq N_*$ and $\lim_{k\in\mathcal{K}} \Delta_N^k = 0$. Then, $\lim_{k\in\mathcal{K}} s_B^k = 0$, where $B \equiv \{1, \ldots, n\}\backslash N$.*

Proof. By Theorem 4.4(a), it follows that there exists $L_0 > 0$ such that $f(x^k) - f(x^*) \le L_0\|x_N^k\|$ for every integer $k \ge 0$. Using this fact and inequality (12), we obtain

$$\begin{aligned}
\|s_B^k\| &\le \|(x_B^k)^{-1}\|\,\|x_B^k s_B^k\| \le \|(x_B^k)^{-1}\|\,\|x^k s^k\| \\
&\le \|(x_B^k)^{-1}\|\frac{f(x^k) - f(x^{k+1})}{\beta} \le \|(x_B^k)^{-1}\|\frac{f(x^k) - f(x^*)}{\beta} \\
&\le \frac{L_0}{\beta}\|(x_B^k)^{-1}\|\,\|x_N^k\| = \mathcal{O}(\Delta_N^k).
\end{aligned}$$

Since by assumption $\lim_{k\in\mathcal{K}} \Delta_N^k = 0$, the result immediately follows. ∎

Lemma 4.6 *Suppose that $N \in \Gamma$ and $B \equiv \{1, \ldots, n\}\backslash N$. Then the following statements hold:*

(a) $g_B^* \in Range(A_B^T)$;

(b) $x^* \in \mathcal{P}_N^+$ *and* $x \mapsto (g^*)^T x$ *is constant over the face* $\mathcal{P}_N^+$;

(c) $\mathcal{P}_N^{++} \neq \emptyset$.

Proof. Since $N \in \Gamma$, we have $\liminf_{k\to\infty} \Delta_N^k = 0$. By Lemma 4.5, the sequence $\{(y^k, s^k)\}$ has an accumulation point $(\bar{y}, \bar{s})$ such that $\bar{s}_B = 0$. Since $A^T\bar{y} + \bar{s} = g^*$ and $\bar{s}_B = 0$, (a) follows. Statement (b) follows from (a) and the fact that $N \subseteq N_*$. We next show (c). Let $\mathcal{K}$ be an infinite index set such that $\lim_{k\in\mathcal{K}} \Delta_N^k = 0$. Consider the feasible system (e.g., $x = x^*$ is a feasible solution)

$$Ax = b, \qquad x_N = 0. \tag{59}$$

Using Lemma 4.1 with $z = x^k$, there exist a constant $L_0 > 0$ and a sequence $\{w^k\}$ of solutions of (59) such that

$$\|x^k - w^k\| \leq L_0 \|x_N^k\|, \qquad \forall k \geq 0.$$

Hence, we obtain

$$\|(X_B^k)^{-1}(w_B^k - x_B^k)\| \leq \|(x_B^k)^{-1}\| \, \|w_B^k - x_B^k\| \leq L_0 \|(x_B^k)^{-1}\| \, \|x_N^k\| = L_0 \Delta_B^k,$$

from which we conclude that $\lim_{k\in\mathcal{K}} \|(X_B^k)^{-1}(w_B^k - x_B^k)\| = 0$. Clearly, this implies that $w_B^k > 0$ for sufficiently large $k \in \mathcal{K}$, and hence that $w^k \in \mathcal{P}_N^{++}$, for all $k \in \mathcal{K}$ sufficiently large. ∎

Lemma 4.7 *Assume that* $\bar{N}$ *is a minimal element of* Γ *with respect to inclusion and let* $\bar{B} = \{1, \ldots, n\} \backslash \bar{N}$. *Then there exist scalars* $\delta_0 > 0$ *and* $\eta_0 > 0$ *such the following implication holds for every* k *sufficiently large:*

$$\Delta_{\bar{N}}^k \leq \delta_0 \Longrightarrow \frac{\|X_{\bar{N}}^k s_{\bar{N}}^k\|}{\|x_{\bar{N}}^k\|} \geq \eta_0 \tag{60}$$

Proof. Assume for contradiction that the conclusion of the lemma does not hold. Then there exists an infinite index set $\mathcal{K}$ such that

$$\lim_{k\in\mathcal{K}} \Delta_{\bar{N}}^k = 0, \qquad \lim_{k\in\mathcal{K}} \frac{\|X_{\bar{N}}^k s_{\bar{N}}^k\|}{\|x_{\bar{N}}^k\|} = 0. \tag{61}$$

Since $\{s_{\bar{N}}^k\}$ and $\{x_{\bar{N}}^k / \|x_{\bar{N}}^k\|\}$ are bounded sequences, by passing to a subsequence if necessary we may assume that

$$\lim_{k\in\mathcal{K}} s_{\bar{N}}^k = s_{\bar{N}}^\infty, \text{ and } \lim_{k\in\mathcal{K}} \frac{x_{\bar{N}}^k}{\|x_{\bar{N}}^k\|} = d_{\bar{N}}^\infty \tag{62}$$

for some $|\bar{N}|$-dimensional vectors $s_{\bar{N}}^\infty$ and $d_{\bar{N}}^\infty$. Clearly, $d_{\bar{N}}^\infty \geq 0$, $\|d_{\bar{N}}^\infty\| = 1$ and, by the second relation in (61), $s_{\bar{N}}^\infty d_{\bar{N}}^\infty = 0$. Let $N \equiv \{j \in \bar{N} \mid d_j^\infty = 0\}$ and $B \equiv \{1, \ldots, n\} \backslash N$. Since $\|d_{\bar{N}}^\infty\| = 1$ and $s_N^\infty \neq 0$, it follows that $N \subseteq \bar{N}$ properly, and hence that $B \supseteq \bar{B}$ properly. Using the first relation in (61), the second relation in (62) and the fact that $d_N^\infty = 0$ and $d_{\bar{N}\backslash N}^\infty > 0$, we can easily see that $\lim_{k\in\mathcal{K}} \Delta_N^k = 0$. Hence, it follows that $\liminf_{k\to\infty} \Delta_N^k = 0$, and since $N \subseteq \bar{N} \subseteq N_*$, we conclude that $N \in \Gamma$. But this together with the fact that N is contained $\bar{N}$ properly contradict the assumption that $\bar{N}$ is a minimal element of Γ. ∎

The next result plays an important role in several of the results that will be stated later. It is useful to recall the definition of the set $\mathcal{D}_B(x)$ given in (3).

Lemma 4.8 *The following statements hold:*

(a) $s^k_{B_*} = \mathcal{O}(\|X^k s^k\| \, \|x^{k+1} - x^k\|)$;

(b) there exists a sequence $\{(y^k_*, s^k_*)\} \subseteq \mathcal{D}_{B_*}(x^*)$ *such that*

$$\|s^k_{N_*} - (s^k_*)_{N_*}\| = \mathcal{O}(\|x^{k+1} - x^*\|), \tag{63}$$

$$\|X^k(s^k - s^k_*)\| = \mathcal{O}(\|X^k s^k\|). \tag{64}$$

Proof. Statement (a) follows from the fact that $(x^k_i)^{-1}$ is bounded above for every $i \in B_*$ and from the relation

$$s^k_i = (x^k_i)^{-2}(x^k_i - x^{k+1}_i)\frac{\|X^k s^k\|}{\beta}.$$

We next show (b). Consider the feasible system (e.g., if (y^*, s^*) is an accumulation point of $\{(y^k, s^k)\}$ then $(y, s) = (y^*, s^*)$ is a feasible solution of this system)

$$A^T y + s = g^*, \qquad s_{B_*} = 0 \tag{65}$$

(e.g., if (y^*, s^*) is an accumulation point of $\{(y^k, s^k)\}$ then $(y, s) = (y^*, s^*)$ is a feasible solution of this system). Using Lemma 4.1 with $z = (y^k, s^k)$ and the fact that $A^T y^k + s^k = g^{k+1}$, we conclude that for every $k \geq 0$, (65) has a solution (y^k_*, s^k_*) such that

$$\|y^k - y^k_*\| + \|s^k - s^k_*\| = \mathcal{O}(\|g^{k+1} - g^*\| + \|s^k_{B_*}\|).$$

Clearly, $\{(y^k_*, s^k_*)\} \subseteq \mathcal{D}_{B_*}(x^*)$, and (63) now follows from the above expression, statement (a) and the fact that $\|g^{k+1} - g^*\| = \|Q(x^{k+1} - x^*)\| = \mathcal{O}(\|x^{k+1} - x^*\|)$. We next show (64). First observe that (a) and the fact that $(s^k_*)_{B_*} = 0$ imply that

$$\|X^k_{B_*}(s^k - s^k_*)_{B_*}\| = \|X^k_{B_*} s^k_{B_*}\| = \mathcal{O}(\|x^{k+1} - x^k\| \, \|X^k s^k\|) \leq \mathcal{O}(\|X^k s^k\|).$$

By Lemma 4.2 and Lemma 2.4(c), it follows that

$$\|x^{k+1} - x^*\|^2 = \mathcal{O}(\|X^k s^k\|).$$

Using this relation, (5) and (63), we obtain

$$\begin{aligned} \|X^k_{N_*}(s^k - s^k_*)_{N_*}\| &\leq (1-\beta)^{-1}\|x^{k+1}_{N_*}\| \, \|(s^k - s^k_*)_{N_*}\| \\ &\leq \mathcal{O}\left(\|x^{k+1} - x^*\|^2\right) = \mathcal{O}(\|X^k s^k\|). \end{aligned}$$

■

Lemma 4.9 *Let* (B, N) *be a partition such that* $\mathcal{D}_B(x^*) \neq \emptyset$. *Then there exists a constant* $C_3 > 0$ *such that* $\|(P(x)Xg^*)_B\| \leq C_3 \, (\|x_N\| \Delta_N(x))$ *for every* $x \in \mathcal{P}_{++}$ *with* $\Delta_N(x)$ *sufficiently small.*

Proof. Since $\mathcal{D}_B(x^*) \neq \emptyset$, let $(\bar{y}, \bar{s}) \in \mathcal{D}_B(x^*)$ be given. We have $\bar{s}_B = 0$ and $P(x)Xg^* = P(x)X\bar{s}$. From Lemma 3.1, it follows that

$$P(x)X\bar{s} = \begin{pmatrix} \tilde{P}_N(x_N)X_N\bar{s}_N \\ \tilde{P}_B(x_B)X_B\bar{s}_B \end{pmatrix} + R(x)X\bar{s},$$

where $\|R(x)\| \leq C_1 \Delta_N(x)$ for every $x \in \mathcal{P}_{++}$ with $\Delta_N(x)$ sufficiently small. Since $\bar{s}_B = 0$, we obtain

$$\|(P(x)Xg^*)_B\| = \|(P(x)X\bar{s})_B\| \leq \|R(x)\|\|X_N\bar{s}_N\| \leq C_1\|\bar{s}_N\|\,\|x_N\|\Delta_N(x),$$

for every $x \in \mathcal{P}_{++}$ with $\Delta_N(x)$ sufficiently small. Letting $C_3 \equiv C_1\|\bar{s}_N\|$, the result follows. ∎

The main goal of the next three results is to prepare the reader for the proof that $\sum_{k=0}^{\infty} \Delta_{\bar{N}}^k < \infty$. This fact will be more easily proved by working with a related sequence of scalars, namely the sequence $\{\Psi_{\bar{B}}^k\}$ defined as

$$\Psi_{\bar{B}}^k \equiv \frac{f(x^k) - f(x^*)}{\min(x_{\bar{B}}^k)}, \qquad \forall k \geq 0. \tag{66}$$

We will show that $\sum_{k=0}^{\infty} \Psi_{\bar{B}}^k < \infty$ and this in turn will allow us to show that $\sum_{k=0}^{\infty} \Delta_{\bar{N}}^k < \infty$. For $\Delta_{\bar{N}}^k$ sufficiently small, the next three lemmas deals with three cases (not necessarily disjoint) that arise by considering the size of the two quantities:

$$\frac{\|x_{N_*\setminus\bar{N}}^k\|^2}{\|x_{\bar{N}}^k\|}, \qquad \frac{\min\,(x_{N_*\setminus\bar{N}}^k)}{\max\,(x_{N_*\setminus\bar{N}}^k)}.$$

Lemma 4.10 deals with the case where the first quantity is sufficiently small; Lemma 4.11 deals with the case where the second quantity is sufficiently small; finally, Lemma 4.12 deals with the case where the first and second quantities are bounded away from 0. In the proof of Theorem 4.13, we combine these three cases together to prove that $\sum_{k=0}^{\infty} \Delta_{\bar{N}}^k < \infty$.

Lemma 4.10 *Assume that $\bar{N}$ is a minimal element of Γ with respect to inclusion and let $\bar{B} = \{1,\ldots,n\}\setminus\bar{N}$. There exist constants $\delta_1 > 0$, $\tau_1 > 0$ and $\gamma_1 \in (0,1)$ such that the following implication holds for every k sufficiently large:*

$$\|x_{N_*\setminus\bar{N}}^k\|^2 \leq \tau_1\|x_{\bar{N}}^k\|, \quad \Delta_{\bar{N}}^k \leq \delta_1 \implies \Psi_{\bar{B}}^{k+1} \leq \gamma_1\Psi_{\bar{B}}^k.$$

Proof. Assume for contradiction that for some infinite index set $\mathcal{K}$, there holds

$$\lim_{k\in\mathcal{K}} \frac{\|x_{N_*\setminus\bar{N}}^k\|^2}{\|x_{\bar{N}}^k\|} = 0, \tag{67}$$

$$\lim_{k\in\mathcal{K}} \Delta_{\bar{N}}^k = 0, \tag{68}$$

$$\limsup_{k\in\mathcal{K}} \frac{\Psi_{\bar{B}}^{k+1}}{\Psi_{\bar{B}}^k} \geq 1. \tag{69}$$

Observe that (68) and Lemma 4.7 imply that

$$\|x_{\bar{N}}^k\| = \mathcal{O}(\|X^k s^k\|), \qquad \forall k \in \mathcal{K}. \tag{70}$$

Below we concentrate our effort in showing the following claim:

$$\lim_{k\in\mathcal{K}} (X_{\bar{B}}^k)^{-1}(x_{\bar{B}}^{k+1} - x_{\bar{B}}^k) = 0. \tag{71}$$

This relation clearly implies that

$$\lim_{k\in\mathcal{K}} \frac{\min(x^{k+1}_{\bar{B}})}{\min(x^k_{\bar{B}})} = 1,$$

which together with Lemma 2.4(b) in turn imply that $\limsup_{k\in\mathcal{K}} \Psi^{k+1}_{\bar{B}}/\Psi^k_{\bar{B}} < 1$, a fact that contradicts (69) and hence establishes the validity of the lemma.

To prove the claim, first observe that

$$(X^k)^{-1}(x^{k+1} - x^k) = \beta(p^k_1 + p^k_2), \tag{72}$$

where

$$p^k_1 \equiv P(x^k)\frac{X^k g^*}{\|X^k s^k\|}, \qquad p^k_2 \equiv P(x^k)\frac{X^k(g^{k+1} - g^*)}{\|X^k s^k\|}. \tag{73}$$

The claim follows once we show that $\lim_{k\in\mathcal{K}} (p^k_1)_{\bar{B}} = 0$ and $\lim_{k\in\mathcal{K}} p^k_2 = 0$. We first show that $\lim_{k\in\mathcal{K}} (p^k_1)_{\bar{B}} = 0$. Since $\bar{N} \in \Gamma$, it follows from Lemma 4.6(a) that $g^*_{\bar{B}} \in \mathrm{Range}(A^T_{\bar{B}})$, or equivalently, that $\mathcal{D}_{\bar{B}}(x^*) \neq \emptyset$. Hence, by (68), (70) and Lemma 4.9, we obtain

$$(p^k_1)_{\bar{B}} = \mathcal{O}\left(\frac{\|x^k_{\bar{N}}\|\,\Delta^k_{\bar{N}}}{\|X^k s^k\|}\right) \leq \mathcal{O}(\Delta^k_{\bar{N}}), \qquad \forall k \in \mathcal{K}, \tag{74}$$

from which we conclude that $\lim_{k\in\mathcal{K}} (p^k_1)_{\bar{B}} = 0$. We next show that $\lim_{k\in\mathcal{K}} p^k_2 = 0$. Indeed, recalling the definition of (y^k_*, s^k_*) in Lemma 4.8(b) and using the fact that $(y^k,\, s^k) \in \mathcal{D}(x^{k+1})$ and $(y^k_*,\, s^k_*) \in \mathcal{D}(x^*)$, we obtain

$$p^k_2 = P(x^k)\frac{X^k(s^k - s^k_*)}{\|X^k s^k\|}$$

and hence,

$$\|p^k_2\| \leq \frac{\|X^k(s^k - s^k_*)\|}{\|X^k s^k\|}. \tag{75}$$

Now, Lemma 4.8(a) and the fact that $(s^k_*)_{B_*} = 0$ imply

$$\frac{\|X^k_{B_*}(s^k_{B_*} - (s^k_*)_{B_*})\|}{\|X^k s^k\|} \leq \|x^k_{B_*}\|\frac{\|s^k_{B_*}\|}{\|X^k s^k\|} = \mathcal{O}(\|x^{k+1} - x^k\|), \qquad \forall k \in \mathcal{K}, \tag{76}$$

and, from Lemma 4.8(b), Theorem 4.4(b) and (70), we obtain

$$\begin{aligned}
\frac{\|X^k_{N_*}(s^k_{N_*} - (s^k_*)_{N_*})\|}{\|X^k s^k\|} &\leq \frac{(\|x^k_{\bar{N}}\| + \|x^k_{N_*\setminus\bar{N}}\|)\,\|s^k_{N_*} - (s^k_*)_{N_*}\|}{\|X^k s^k\|} \\
&\leq \mathcal{O}\left(\frac{(\|x^k_{\bar{N}}\| + \|x^k_{N_*\setminus\bar{N}}\|)}{\|x^k_{\bar{N}}\|^{1/2}}\,\frac{\|x^{k+1} - x^*\|}{\|x^k_{\bar{N}}\|^{1/2}}\right) \\
&\leq \mathcal{O}\left(\frac{\|x^k_{\bar{N}}\| + \|x^k_{N_*\setminus\bar{N}}\|}{\|x^k_{\bar{N}}\|^{1/2}}\right), \qquad \forall k \in \mathcal{K},
\end{aligned} \tag{77}$$

which converges to 0 in view of (67). It follows from (75), (76) and (77) that $\lim_{k\in\mathcal{K}} p^k_2 = 0$. We have thus shown that $\lim_{k\in\mathcal{K}} (X^k_{\bar{B}})^{-1}(x^{k+1}_{\bar{B}} - x^k_{\bar{B}}) = 0$. ∎

Lemma 4.11 *Assume that $\bar{N}$ is a minimal element of Γ with respect to inclusion and let $\bar{B} = \{1,\ldots,n\}\backslash\bar{N}$. There exist constants $\delta_2 > 0$, $M_0 > 0$ and $\gamma_2 \in (0,1)$ such that the following implication holds for every k sufficiently large:*

$$\frac{\max(x^k_{N_*\backslash\bar{N}})}{\min(x^k_{N_*\backslash\bar{N}})} \geq M_0, \quad \Delta^k_{\bar{N}} \leq \delta_2 \implies \Psi^{k+1}_{\bar{B}} \leq \gamma_2 \Psi^k_{\bar{B}}.$$

Proof. Assume for contradiction that for some infinite index set $\mathcal{K}$, there holds

$$\lim_{k\in\mathcal{K}} \frac{\max(x^k_{N_*\backslash\bar{N}})}{\min(x^k_{N_*\backslash\bar{N}})} = \infty, \tag{78}$$

$$\lim_{k\in\mathcal{K}} \Delta^k_{\bar{N}} = 0, \tag{79}$$

$$\limsup_{k\in\mathcal{K}} \frac{\Psi^{k+1}_{\bar{B}}}{\Psi^k_{\bar{B}}} \geq 1. \tag{80}$$

It is easy to see that there exist some infinite subset $\mathcal{K}_0$ of $\mathcal{K}$ and some set $N \subseteq \{1,\ldots,n\}$ such that

$$\limsup_{k\in\mathcal{K}_0} \frac{x^k_i}{\min(x^k_{N_*\backslash\bar{N}})} < \infty, \qquad \forall i \in N, \tag{81}$$

$$\lim_{k\in\mathcal{K}_0} \frac{x^k_i}{\min(x^k_{N_*\backslash\bar{N}})} = \infty, \qquad \forall i \notin N. \tag{82}$$

Using (79) and the fact that $\lim_{k\to\infty} x^k_{B_*} > 0$, it is easy to see that $\bar{N} \subseteq N \subseteq N_*$. Moreover, the definition of N, (78) and (79) imply that these two inclusions are proper. We claim that

$$\lim_{k\in\mathcal{K}_0} (X^k_{N\backslash\bar{N}})^{-1}(x^{k+1}_{N\backslash\bar{N}} - x^k_{N\backslash\bar{N}}) = 0. \tag{83}$$

As we will see, showing this relation is the major effort of the proof. As in (72), we know that $(X^k)^{-1}(x^{k+1} - x^k) = \beta(p^k_1 + p^k_2)$, where p^k_1 and p^k_2 are defined in (73). The claim follows if we show that $\lim_{k\in\mathcal{K}}(p^k_1)_{\bar{B}} = 0$ and $\lim_{k\in\mathcal{K}_0}(p^k_2)_N = 0$. The proof that $\lim_{k\in\mathcal{K}}(p^k_1)_{\bar{B}} = 0$ is exactly the same as in (74) of Lemma 4.10. We next show that $\lim_{k\in\mathcal{K}_0}(p^k_2)_N = 0$. First observe (81) and (82) imply that $\lim_{k\in\mathcal{K}_0}\Delta^k_N = 0$, where $N \equiv \{1,\cdots,n\}\backslash B$. Using Lemma 3.1, we obtain

$$\begin{aligned} p^k_2 &= \frac{1}{\|X^k s^k\|} P(x^k) X^k (s^k - s^k_*) \\ &= \frac{1}{\|X^k s^k\|} \begin{pmatrix} \tilde{P}_N(x^k_N) X^k_N (s^k - s^k_*)_N \\ \tilde{P}_B(x^k_B) X^k_B (s^k - s^k_*)_B \end{pmatrix} + R(x^k) \frac{X^k(s^k - s^k_*)}{\|X^k s^k\|}, \end{aligned}$$

where $\|R(x^k)\| = \mathcal{O}(\Delta^k_N)$. Using this relation, (63), (64) and (5), we obtain

$$\|(p^k_2)_N\| \leq \frac{\|\tilde{P}_N(x^k_N) X^k_N (s^k - s^k_*)_N\|}{\|X^k s^k\|} + \|R(x^k)\| \frac{\|X^k(s^k - s^k_*)\|}{\|X^k s^k\|}$$

$$
\begin{aligned}
&\leq \frac{\|x_N^k\|\,\|(s^k - s_*^k)_N\|}{\|X^k s^k\|} + \mathcal{O}(\Delta_N^k) \\
&\leq \mathcal{O}\left(\frac{\|x_N^k\|}{\|x^{k+1} - x^*\|}\,\frac{\|x^{k+1} - x^*\|^2}{\|X^k s^k\|}\right) + \mathcal{O}(\Delta_N^k) \\
&\leq \mathcal{O}\left(\frac{\|x_N^k\|}{\|x_{N_*\setminus N}^{k+1}\|}\right) + \mathcal{O}(\Delta_N^k) \\
&\leq \mathcal{O}\left(\frac{\|x_N^k\|}{\|x_{N_*\setminus N}^{k}\|}\right) + \mathcal{O}(\Delta_N^k) \\
&\leq \mathcal{O}(\Delta_N^k),
\end{aligned}
$$

where in the third inequality we used the fact that $\|x^{k+1} - x^*\|^2 = \mathcal{O}(\|X^k s^k\|)$, due to Lemma 2.4(c) and Lemma 4.2. Since $\{\Delta_N^k\}_{k\in\mathcal{K}_0}$ converges to 0, the last expression shows that $\lim_{k\in\mathcal{K}_0}(p_2^k)_N = 0$. Using (5), (81) and (82), it easy to see that for every $k \in \mathcal{K}_0$ sufficiently large,

$$
\min(x_{\bar{B}}^k) = \min(x_{N\setminus\bar{N}}^k), \qquad \min(x_{\bar{B}}^{k+1}) = \min(x_{N\setminus\bar{N}}^{k+1}). \tag{84}
$$

It then follows from (83) and (84) that

$$
\lim_{k\in\mathcal{K}_0}\left(\frac{\min(x_{\bar{B}}^{k+1})}{\min(x_{\bar{B}}^k)}\right) = \lim_{k\in\mathcal{K}_0}\left(\frac{\min(x_{N\setminus\bar{N}}^{k+1})}{\min(x_{N\setminus\bar{N}}^k)}\right) = 1.
$$

This relation and Lemma 2.4(b) then imply that $\limsup_{k\in\mathcal{K}_0} \Psi_{\bar{B}}^{k+1}/\Psi_{\bar{B}}^k < 1$, a fact that contradicts (80). ∎

Lemma 4.12 *Assume that $\bar{N}$ is a minimal element of Γ and let $\bar{B} = \{1,\ldots,n\}\setminus\bar{N}$. Let δ_0, τ_1 and M_0 be the constants introduced in Lemma 4.7, 4.10 and 4.11. Then there exist constants $\alpha_1 > 0$, $\alpha_2 > 0$ and $M_1 > 0$ such that the following two implications hold for every k sufficiently large:*

$$
\Delta_{\bar{N}}^k \leq \delta_0 \quad \Longrightarrow \quad \alpha_1 \Psi_{\bar{B}}^k \leq \Delta_{\bar{N}}^k \leq \alpha_2 \Psi_{\bar{B}}^k, \tag{85}
$$

$$
\left.\begin{aligned}
\Delta_{\bar{N}}^k &\leq \delta_0, \\
\|x_{N_*\setminus\bar{N}}^k\|^2 &\geq \tau_1 \|x_{\bar{N}}^k\|, \\
\max\,(x_{N_*\setminus\bar{N}}^k) &\leq M_0 \min\,(x_{N_*\setminus\bar{N}}^k)
\end{aligned}\right\} \quad \Longrightarrow \quad \Delta_{\bar{N}}^k \leq M_1 (f(x^k) - f(x^*))^{1/2}. \tag{86}
$$

Proof. We first show implication (85). Assume that k_0 is such that implication (60) holds for every $k \geq k_0$. Using (60), (12) and Theorem 4.4(a), we conclude that if $k \geq k_0$ and $\Delta_{\bar{N}}^k \leq \delta_0$ then

$$
\begin{aligned}
\|x_{\bar{N}}^k\| &\leq \frac{\|X_{\bar{N}}^k s_{\bar{N}}^k\|}{\eta_0} \leq \frac{\|X^k s^k\|}{\eta_0} \leq \frac{f(x^k) - f(x^{k+1})}{\beta\eta_0} \\
&\leq \frac{f(x^k) - f(x^*)}{\beta\eta_0} = \mathcal{O}(\|x_{\bar{N}}^k\|),
\end{aligned}
$$

or equivalently, that $\|x_{\bar{N}}^k\| \sim f(x^k) - f(x^*)$, for every k such that $\Delta_{\bar{N}}^k \leq \delta_0$. Using this fact, it is now straightforward to verify that there exist $\alpha_1 > 0$ and $\alpha_2 > 0$ such that (85) holds for every $k \geq k_0$.

We now show that (86) holds. Assume that $k \geq k_0$ is such that $\Delta_{\bar{N}}^k \leq \delta_0$, $\|x^k_{N_*\setminus\bar{N}}\|^2 \geq \tau_1 \|x^k_{\bar{N}}\|$ and $\max(x^k_{N_*\setminus\bar{N}}) \leq M_0 \min(x^k_{N_*\setminus\bar{N}})$. Then,

$$\begin{aligned}
\|(x^k_{\bar{B}})^{-1}\|\,\|x^k_{\bar{N}}\| &\leq \frac{\max((x^k_{\bar{B}})^{-1})\|x^k_{\bar{N}}\|}{\sqrt{n}} = \frac{\|x^k_{\bar{N}}\|}{\sqrt{n}\min(x^k_{\bar{B}})} \\
&\leq \frac{\|x^k_{\bar{N}}\|}{\sqrt{n}\min(x^k_{N_*\setminus\bar{N}})} \leq \frac{M_0\|x^k_{\bar{N}}\|}{\sqrt{n}\max(x^k_{N_*\setminus\bar{N}})} \\
&\leq \frac{M_0\|x^k_{\bar{N}}\|}{\|x^k_{N_*\setminus\bar{N}}\|} \leq \frac{M_0\,\|x^k_{\bar{N}}\|}{\sqrt{\tau_1\|x^k_{\bar{N}}\|}} \\
&= \frac{M_0}{\sqrt{\tau_1}}\|x^k_{\bar{N}}\|^{1/2} = \mathcal{O}(f(x^k) - f(x^*))^{1/2}.
\end{aligned}$$

This shows that there exists $M_1 > 0$ such that (86) holds for every $k \geq k_0$. ∎

The following result shows that the partition $(\bar{B}, \bar{N})$, where $\bar{N}$ is a minimal element of Γ, satisfies condition (A) of Assumption 3.2.

Theorem 4.13 *Assume that $\bar{N}$ is a minimal element of Γ and let $\bar{B} = \{1,\ldots,n\}\setminus\bar{N}$. Then, $\sum_{k=0}^{\infty}\Delta_{\bar{N}}^k < \infty$.*

Proof. First of all, it is useful to recall the content of Lemmas 4.10, 4.11 and 4.12 here. Namely, there exist positive constants δ_0, δ_1, δ_2, M_0, M_1, τ_1, α_1, α_2, $\gamma_1 < 1$ and $\gamma_2 < 1$ such that the following implications hold:

$$\|x^k_{N_*\setminus\bar{N}}\|^2 \leq \tau_1\|x^k_{\bar{N}}\|, \quad \Delta_{\bar{N}}^k \leq \delta_1 \implies \Psi_{\bar{B}}^{k+1} \leq \gamma_1\Psi_{\bar{B}}^k; \tag{87}$$

$$\frac{\max(x^k_{N_*\setminus\bar{N}})}{\min(x^k_{N_*\setminus\bar{N}})} \geq M_0, \quad \Delta_{\bar{N}}^k \leq \delta_2 \implies \Psi_{\bar{B}}^{k+1} \leq \gamma_2\Psi_{\bar{B}}^k; \tag{88}$$

$$\Delta_{\bar{N}}^k \leq \delta_0 \implies \alpha_1\Psi_{\bar{B}}^k \leq \Delta_{\bar{N}}^k \leq \alpha_2\Psi_{\bar{B}}^k; \tag{89}$$

$$\left.\begin{aligned}
\Delta_{\bar{N}}^k &\leq \delta_0, \\
\|x^k_{N_*\setminus\bar{N}}\|^2 &\geq \tau_1\|x^k_{\bar{N}}\|, \\
\max(x^k_{N_*\setminus\bar{N}}) &\leq M_0\min(x^k_{N_*\setminus\bar{N}})
\end{aligned}\right\} \implies \Delta_{\bar{N}}^k \leq M_1(f(x^k) - f(x^*))^{1/2}. \tag{90}$$

Observe that implications (87), (88) and (90) covers all possible situations for $\Delta_{\bar{N}}^k$ sufficiently small. Define

$$\epsilon_1 \equiv \frac{1-\beta}{1+\beta}\min\{\delta_0, \delta_1, \delta_2\}. \tag{91}$$

First, we claim that there exists an integer $k_0 \geq 0$ such that

$$\Delta_{\bar{N}}^k \leq \epsilon_1, \qquad \forall k \geq k_0. \tag{92}$$

Indeed, let

$$\epsilon_2 \equiv \frac{(1-\beta)\epsilon_1}{(1+\beta)\alpha_2}, \qquad \epsilon_3 \equiv \left(\frac{\alpha_1(1-\beta)\epsilon_2}{M_1}\right)^2. \tag{93}$$

Due to Lemmas 4.7, 4.10, 4.11 and 4.12 and the fact that $\liminf_{k\to\infty} \Delta_{\bar{N}}^k = 0$ and $\lim_{k\to\infty} f(x^k) = f(x^*)$, it is easy to show the existence of an integer $k_0 \geq 0$ such that the implications (87), (88), (89) and (90) hold for every $k \geq k_0$ and

$$\Delta_{\bar{N}}^{k_0} \leq \epsilon_1, \quad \Psi_{\bar{B}}^{k_0} \leq \epsilon_2, \quad f(x^{k_0}) - f(x^*) \leq \epsilon_3, \tag{94}$$

where we use (90) to obtain the second relation. We will prove by induction on k that

$$\Delta_{\bar{N}}^{k} \leq \epsilon_1, \quad \Psi_{\bar{B}}^{k} \leq \epsilon_2, \tag{95}$$

for all $k \geq k_0$, thereby proving our claim (92). In view of the first two inequalities in (94), k_0 satisfies (95). Assume that the index k satisfies (95). Since $\Delta_{\bar{N}}^k \leq \epsilon_1 \leq \min\{\delta_0, \delta_1, \delta_2\}$ and the implications (87), (88), (89) and (90) hold, we conclude that

$$\alpha_1 \Psi_{\bar{B}}^k \leq \Delta_{\bar{N}}^k \leq \alpha_2 \Psi_{\bar{B}}^k, \tag{96}$$

and that either

$$\Delta_{\bar{N}}^k \leq M_1 (f(x^k) - f(x^*))^{1/2} \tag{97}$$

holds, or the inequality

$$\Psi_{\bar{B}}^{k+1} \leq \bar{\gamma} \Psi_{\bar{B}}^k \tag{98}$$

holds, where $\bar{\gamma} \equiv \max\{\gamma_1, \gamma_2\} < 1$. Using (5), (96), the second inequality in (95) and (93), we obtain

$$\Delta_{\bar{N}}^{k+1} \leq \frac{1+\beta}{1-\beta} \Delta_{\bar{N}}^k \leq \frac{1+\beta}{1-\beta} \alpha_2 \Psi_{\bar{B}}^k \leq \frac{1+\beta}{1-\beta} \alpha_2 \epsilon_2 = \epsilon_1.$$

If inequality (97) holds then, using (5), (96), the last inequality in (94), (93), and the fact that $\{f(x^k) - f(x^*)\}$ decreases monotonically, we obtain

$$\Psi_{\bar{B}}^{k+1} \leq \frac{1}{1-\beta} \Psi_{\bar{B}}^k \leq \frac{1}{\alpha_1(1-\beta)} \Delta_{\bar{N}}^k \leq \frac{M_1}{\alpha_1(1-\beta)} (f(x^k) - f(x^*))^{1/2} \leq \frac{M_1}{\alpha_1(1-\beta)} \epsilon_3^{1/2} = \epsilon_2.$$

If instead (98) holds then, by the last inequality of (95) and the fact that $\bar{\gamma} < 1$, we obtain $\Psi_{\bar{B}}^{k+1} \leq \bar{\gamma} \Psi_{\bar{B}}^k \leq \Psi_{\bar{B}}^k \leq \epsilon_2$. We have thus shown that $\Delta_{\bar{N}}^{k+1} \leq \epsilon_1$ and $\Psi_{\bar{B}}^{k+1} \leq \epsilon_2$, and hence that (95) holds for every $k \geq k_0$.

We now show that $\sum_{k=0}^{\infty} \Delta_{\bar{N}}^k < \infty$. Indeed, since, for every $k \geq k_0$, $\Delta_{\bar{N}}^k \leq \min\{\delta_0, \delta_1, \delta_2\}$ and the implications (87), (88), (89) and (90) hold, it follows that (96) holds and that either (97) or (98) holds, for every $k \geq k_0$. Since (96) holds for every $k \geq k_0$, it is sufficient to show that $\sum_{k=0}^{\infty} \Psi_{\bar{B}}^k < \infty$. Let $\mathcal{K}$ denote those indices $k \geq k_0$ for which (97) holds. If $\mathcal{K}$ is finite then we have $\sum_{k=0}^{\infty} \Psi_{\bar{B}}^k < \infty$ since (98) holds for every k sufficiently large. So assume that $\mathcal{K}$ is infinite and, without loss of generality, that $k_0 \in \mathcal{K}$. Note that if $k' < k''$ are two consecutive indices in $\mathcal{K}$ then

$$\sum_{k=k'+1}^{k''} \Psi_{\bar{B}}^k \leq \frac{1}{1-\bar{\gamma}} \Psi_{\bar{B}}^{k'+1},$$

since (98) holds for every $k = k'+1, \ldots, k''-1$. Using this observation, (5), (96), the fact that (97) holds for every $k \in \mathcal{K}$ and Lemma 2.4(b), we obtain

$$\begin{aligned}\sum_{k=k_0+1}^{\infty} \Psi_{\bar{B}}^k &\leq \frac{1}{1-\bar{\gamma}} \sum_{k\in\mathcal{K}} \Psi_{\bar{B}}^{k+1} \leq \frac{1}{(1-\beta)(1-\bar{\gamma})} \sum_{k\in\mathcal{K}} \Psi_{\bar{B}}^k \\ &\leq \frac{1}{\alpha_1(1-\beta)(1-\bar{\gamma})} \sum_{k\in\mathcal{K}} \Delta_{\bar{N}}^k \leq \frac{M_1}{\alpha_1(1-\beta)(1-\bar{\gamma})} \sum_{k\in\mathcal{K}} (f(x^k)-f(x^*))^{1/2} \\ &\leq \frac{M_1}{\alpha_1(1-\beta)(1-\bar{\gamma})} \sum_{k=0}^{\infty} (f(x^k)-f(x^*))^{1/2} < \infty.\end{aligned}$$

We have thus proved that $\sum_{k=0}^{\infty} \Psi_{\bar{B}}^k < \infty$, and hence that $\sum_{k=0}^{\infty} \Delta_{\bar{N}}^k < \infty$. ∎

The following lemma is the last step towards the main result of this paper.

Lemma 4.14 *Assume that $\bar{N}$ is a minimal element of Γ and let $\bar{B} = \{1, \ldots, n\} \backslash \bar{N}$. Then the following statements hold:*

(a) For every $i \in \bar{N}$, there holds

$$x_i^k \sim \|x_{\bar{N}}^k\| \sim \|X^k s^k\| \sim \|X_{\bar{N}}^k s_{\bar{N}}^k\| \sim f(x^k) - f(x^*) \sim f(x^k) - f(x^{k+1});$$

(b) $\lim_{k\to\infty} \|X_{\bar{N}}^k (s^k - \tilde{s}^k)_{\bar{N}}\| / \|x_{\bar{N}}^k\| = 0$;

(c) $\liminf_{k\to\infty} (g^*)^T (x^k - x^{k+1}) / \|x_{\bar{N}}^k\| > 0$.

Proof. Due to relation (12), Lemma 4.7, Theorem 4.13 and Theorem 4.4(a), we conclude that for every $i \in \bar{N}$ and k sufficiently large,

$$\begin{aligned}x_i^k \leq \|x_{\bar{N}}^k\| &\leq \frac{\|X_{\bar{N}}^k s_{\bar{N}}^k\|}{\eta_0} \leq \frac{\|X^k s^k\|}{\eta_0} \leq \frac{f(x^k)-f(x^{k+1})}{\beta\eta_0} \\ &\leq \frac{f(x^k)-f(x^*)}{\beta\eta_0} = \mathcal{O}(x_i^k),\end{aligned}$$

which clearly implies statement (a). To prove (b), note that by (7) and (18), we have

$$X^k(s^k - \tilde{s}^k) = P(x^k) X^k (g^{k+1} - g^*) = P(x^k) X^k (s^k - s_*^k),$$

where $\{s_*^k\}$ is the sequence introduced in Lemma 4.8. Using Lemma 3.1, we obtain

$$X^k(s^k - \tilde{s}^k) = \begin{pmatrix} \tilde{P}_{\bar{B}}(x_{\bar{B}}^k) X_{\bar{B}}^k (s^k - s_*^k)_{\bar{B}} \\ \tilde{P}_{\bar{N}}(x_{\bar{N}}^k) X_{\bar{N}}^k (s^k - s_*^k)_{\bar{N}} \end{pmatrix} + R(x^k)[X^k(s^k - s_*^k)]$$

where $\|R(x^k)\| = \mathcal{O}(\Delta_{\bar{N}}^k)$. This relation together with (63), (64) and statement (a) imply

$$\begin{aligned}\frac{\|X_{\bar{N}}^k (s^k - \tilde{s}^k)_{\bar{N}}\|}{\|x_{\bar{N}}^k\|} &\leq \frac{\|\tilde{P}_{\bar{N}}(x_{\bar{N}}^k) X_{\bar{N}}^k (s^k - s_*^k)_{\bar{N}}\|}{\|x_{\bar{N}}^k\|} + \|R(x^k)\| \frac{\|X^k(s^k - s_*^k)\|}{\|x_{\bar{N}}^k\|} \\ &\leq \|(s^k - s_*^k)_{\bar{N}}\| + \mathcal{O}(\Delta_{\bar{N}}^k) = \mathcal{O}(\|x^{k+1} - x^k\| + \Delta_{\bar{N}}^k),\end{aligned}$$

from which (b) follows. We next prove (c). By (a), we have that $\liminf_{k\to\infty} \|X^k_{\bar{N}} s^k_{\bar{N}}\|/\|x^k_{\bar{N}}\| > 0$. Hence, by (b), we conclude that

$$\liminf_{k\to\infty} \|X^k_{\bar{N}} \tilde{s}^k_{\bar{N}}\|/\|x^k_{\bar{N}}\| > 0 \tag{99}$$

By Lemma 4.6, we have that $g^*_{\bar{B}} \in \text{Range}(A^T_{\bar{B}})$, or equivalently, that $\mathcal{D}_{\bar{B}} \neq \emptyset$. Let $(\bar{y}, \bar{s})$ be an arbitrary element of $\mathcal{D}_{\bar{B}}$. We have,

$$\begin{aligned}
(g^*)^T(x^k - x^{k+1}) &= \frac{\beta}{\|X^k s^k\|}(g^*)^T(X^k)^2 s^k \\
&= \frac{\beta}{\|X^k s^k\|}\left\{(g^*)^T(X^k)^2\tilde{s}^k + (g^*)^T(X^k)^2(s^k - \tilde{s}^k)\right\} \\
&= \frac{\beta}{\|X^k s^k\|}\left\{\|X^k\tilde{s}^k\|^2 + \bar{s}^T(X^k)^2(s^k - \tilde{s}^k)\right\} \\
&\geq \frac{\beta}{\|X^k s^k\|}\left\{\|X^k_{\bar{N}}\tilde{s}^k_{\bar{N}}\|^2 - \|X^k_{\bar{N}}\bar{s}_{\bar{N}}\|\,\|X^k_{\bar{N}}(s^k - \tilde{s}^k)_{\bar{N}}\|\right\}.
\end{aligned}$$

This inequality together with statements (a) and (b) imply that

$$\liminf_{k\to\infty} \frac{(g^*)^T(x^k - x^{k+1})}{\|x^k_{\bar{N}}\|} \geq \beta \liminf_{k\to\infty} \frac{\|X^k_{\bar{N}}\tilde{s}^k_{\bar{N}}\|^2}{\|X^k s^k\|\,\|x^k_{\bar{N}}\|},$$

which in view of (a) and (99) imply (c). ∎

We are now in a position to state the main result of this paper. Its content is very similar to the one in Theorem 3.6. The main difference lies in the fact that Theorem 4.15 establishes the existence of a (unique) partition $(\bar{B}, \bar{N})$ satisfying the Local Assumption 3.2 while Theorem 3.6 assumes its existence.

Theorem 4.15 *Assume that $\bar{N}$ is a minimal element of Γ and let $\bar{B} = \{1, \ldots, n\}\backslash\bar{N}$. Then the limit point x^* of the sequence $\{x^k\}$ generated by Algorithm AS is an optimal solution of problem (1) and the associated sequence of dual estimates $\{(y^k, s^k)\}$ converges to the optimal solution (y^*, s^*) of the dual problem (2) that solves the problem*

$$\begin{array}{ll}
\textit{maximize} & \sum_{i\in\bar{N}} \log s_i \\
\textit{subject to} & A^T y + s = g^*, \\
& s_{\bar{B}} = 0,\ s_{\bar{N}} > 0.
\end{array} \tag{100}$$

In particular, it follows that Γ has a unique minimal element, or equivalently, $\bar{N}$ is the smallest element of Γ (with respect to inclusion).

Proof. The result follows as an immediate consequence of Theorem 3.6 once we verify that the partition $(\bar{B}, \bar{N})$ satisfies Local Assumption 3.2. That $(\bar{B}, \bar{N})$ satisfies conditions (A) and (D) of Local Assumption 3.2 follows as a consequence of Theorem 4.13 and Lemma 4.5. We now show that $(\bar{B}, \bar{N})$ satisfies condition (B). Indeed, by Lemma 4.14(c), there exists $k_0 \geq 0$ such that $(g^*)^T(x^k - x^{k+1}) > 0$ for all $k \geq k_0$. This observation and the fact that $(g^*)^T(x^k - x^*) =$

$\sum_{l=k}^{\infty}(g^*)^T(x^l - x^{l+1})$ imply that $(g^*)^T(x^k - x^*) \geq (g^*)^T(x^k - x^{k+1})$ for every $k \geq k_0$. In view of Lemma 4.14(c), we then conclude that

$$\liminf_{k\to\infty} \frac{(g^*)^T(x^k - x^*)}{\|x^k_{\bar{N}}\|} \geq \liminf_{k\to\infty} \frac{(g^*)^T(x^k - x^{k+1})}{\|x^k_{\bar{N}}\|} > 0, \tag{101}$$

which shows that $(\bar{B}, \bar{N})$ satisfies condition (B). We now show that $(\bar{B}, \bar{N})$ satisfies condition (C). First observe that

$$u^k - \tilde{u}^k = \frac{X^k(s^k - \tilde{s}^k)}{(g^*)^T(x^k - x^*)} = \frac{P(x^k)X^k(g^{k+1} - g^*)}{(g^*)^T(x^k - x^*)} = \frac{P(x^k)X^k(s^k - s^k_*)}{(g^*)^T(x^k - x^*)}, \tag{102}$$

where $\{s^k_*\}$ is the sequence introduced in Lemma 4.8. Hence, by (64) and (101), we obtain

$$\begin{aligned} \limsup_{k\to\infty} \|u^k - \tilde{u}^k\| &\leq \limsup_{k\to\infty} \frac{\|X^k(s^k - s^k_*)\|}{(g^*)^T(x^k - x^*)} \leq \limsup_{k\to\infty} \frac{\mathcal{O}(\|X^k s^k\|)}{(g^*)^T(x^k - x^*)} \\ &\leq \limsup_{k\to\infty} \frac{\mathcal{O}(\|x^k_{\bar{N}}\|)}{(g^*)^T(x^k - x^*)} < \infty. \end{aligned} \tag{103}$$

This shows that $\{u^k - \tilde{u}^k\}$ is bounded. That the condition $\lim_{k\to\infty} \|u^k_{\bar{N}} - \tilde{u}^k_{\bar{N}}\| = 0$ is satisfied follows as an immediate consequence of the first equality in (102), Lemma 4.14(b), (101). ∎

Consider the following set associated with problem (1) and its dual (2):

$$\mathcal{S} \equiv \{(x, y, s) \mid x \in \mathcal{P}^+,\ (y, s) \in \mathcal{D}^+(x),\ x^T s = 0\}. \tag{104}$$

It is well known that $(x, y, s) \in \mathcal{S}$ if and only if x is an optimal solution of (1) and (x, y, s) is an optimal solution of (2). Define

$$\begin{aligned} B_{\mathcal{S}} &\equiv \{j \mid x_j > 0 \text{ for some } (x, y, s) \in \mathcal{S}\}; \\ N_{\mathcal{S}} &\equiv \{j \mid s_j > 0 \text{ for some } (x, y, s) \in \mathcal{S}\}; \\ J_{\mathcal{S}} &\equiv \{j \mid x_j = s_j = 0 \text{ for all } (x, y, s) \in \mathcal{S}\} = \{1, \ldots, n\} \backslash (B_{\mathcal{S}} \cup N_{\mathcal{S}}). \end{aligned}$$

It is well known that $(B_{\mathcal{S}}, N_{\mathcal{S}}, J_{\mathcal{S}})$ forms a partition of $\{1, \ldots, n\}$ and that there exists $(\bar{x}, \bar{y}, \bar{s}) \in \mathcal{S}$ such that $\bar{x}_{B_{\mathcal{S}}} > 0$ and $\bar{s}_{N_{\mathcal{S}}} > 0$.

Theorem 4.16 *Let $\bar{N}$ denote the smallest element of Γ. Then $\bar{N} = N_{\mathcal{S}}$.*

Proof. Observe that the limit point (x^*, y^*, s^*) is an element of $\mathcal{S}$ such that $s^*_{\bar{N}} > 0$. It follows from the definition of $N_{\mathcal{S}}$ that $\bar{N} \subseteq N_{\mathcal{S}}$. To prove the reverse inclusion, let $(\bar{x}, \bar{y}, \bar{s})$ be an element of $\mathcal{S}$ such that $\bar{x}_{B_{\mathcal{S}}} > 0$ and $\bar{s}_{N_{\mathcal{S}}} > 0$. Then, we have

$$f(x^k) - f(x^*) = f(x^k) - f(\bar{x}) \geq \nabla f(\bar{x})^T(x^k - \bar{x}) = \bar{s}^T(x^k - \bar{x}) = \bar{s}^T_{N_{\mathcal{S}}} x^k_{N_{\mathcal{S}}} \geq \bar{s}_j x^k_j,$$

and hence, $x^k_j = \mathcal{O}(f(x^k) - f(x^*))$ for every $j \in N_{\mathcal{S}}$. It follows from Lemma 4.14(a) that $x^k_i \sim x^k_j$ for every $i, j \in N_{\mathcal{S}}$. Since $\bar{N} \in \Gamma$, we have $\liminf_{k\to\infty} \Delta^k_{\bar{N}} = 0$ and, in view of the previous observation, this can only happens if $\bar{N} = N_{\mathcal{S}}$. ∎

Concluding Discussion

In this paper, we have established the global convergence of the affine scaling algorithm for convex QP without assuming any nondegeneracy assumption. We conclude the paper by pointing out a few problems which can be topics for future research.

The first problem is to remove the assumption that the set of optimal solutions of (1) is bounded. The only part we use this assumption in this paper is to show the convergence of the sequence $\{x^k\}$, where we use a result of Jie Sun [17]. The second problem is to establish that $B_S = \bar{B}$. This fact would imply that when the QP problem (1) has a pair of primal and dual solutions satisfying strictly complementarity (that is, $J_S = \emptyset$) then the limit points x^* and (y^*, s^*) of Theorem 4.15 would satisfy the strictly complementarity condition that $x^* + s^* > 0$. This would generalize a result obtained by Dikin [7] (see also Vanderbei and Lagarias [25]) for primal nondegenerate LP problems and by Tsuchiya and Muramatsu [24] for degenerate LP problems which establishes that the limit points x^* and (y^*, s^*) of Theorem 4.15 always satisfy the strictly complementarity condition. The last problem would be to show convergence of Algorithm AS for any value of $\beta \in (0, 1]$. Towards this end, a result established by Dikin [5] could be useful. For a homogeneous LP problem having 0 as the unique optimal solution, Dikin shows that each step of the affine scaling algorithm with $\beta = 1$ reduces the value of the potential function by a fixed constant.

Appendix A

In this appendix we establish some technical results that are used in the paper.

Lemma A.1: *If α_1 and α_2 are positive scalars then*

$$| \log \alpha_1 - \log \alpha_2 | \leq \frac{1}{\min\{\alpha_1, \alpha_2\}} |\alpha_1 - \alpha_2|$$

Lemma A.2: *Let an integer $q \geq 2$ be given. If $v \in \mathbb{R}^q$ is such that $e^T v = 0$ then*

$$\chi(v) \geq \frac{\|v\|}{\sqrt{q(q-1)}} \geq \frac{\|v\|}{q}.$$

Proof. If $v = 0$ then the result obviously hold. Assume then that $v \neq 0$ and let $I = \{i \mid v_i > 0\}$. Since $e^T v = 0$ and $v \neq 0$, it follows that $1 \leq |I| \leq q - 1$. Using the fact that $e^T v = 0$, we obtain

$$\begin{aligned} \chi(v)^2 &= \frac{1}{|I|} \sum_{i \in I} v_i^2 = \frac{\|v\|^2 - \sum_{i \notin I} v_i^2}{|I|} \geq \frac{\|v\|^2 - (\sum_{i \notin I} |v_i|)^2}{|I|} \\ &= \frac{\|v\|^2 - (\sum_{i \in I} v_i)^2}{|I|} \leq \frac{\|v\|^2 - |I|^2 \chi(v)^2}{|I|} \end{aligned}$$

Rearranging the above inequality, we obtain

$$\chi(v) \geq \frac{\|v\|}{\sqrt{|I|(|I|+1)}} \geq \frac{\|v\|}{\sqrt{q(q-1)}}$$

where the last inequality is due to the fact that $|I| \leq q-1$. ∎

The proof of the following result can be found in Tsuchiya and Muramatsu [24], Lemma 5.1 (Lemma 3.1 in the revised version).

Lemma A.3: *For every $(w,\rho) \in \Omega$, where*

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

there holds

$$q\log(1-\rho\|w\|^2) - \sum_{i=1}^{q}\log(1-\rho w_i) \leq \frac{q\rho}{q-\rho}\left\|w - \frac{1}{q}e\right\|^2\left(-q + \frac{\rho}{2\{1-\rho\chi[w]\}}\right).$$

Proposition A.4: *Assume that the sequences $\{\tilde{u}^k\} \subset \mathbb{R}^n$ and $\{\lambda_k\} \subset \mathbb{R}$ and the partition (B,N) of $\{1,\ldots,n\}$ satisfy:*

(a) there exists $\delta \in (0,1)$ such that $1 - \lambda_k\|\tilde{u}^k\|^2/\chi(\tilde{u}^k) \geq \delta$ for all $k \geq 0$;

(b) $e^T\tilde{u}^k_N - 1 \to 0$ as $k \to \infty$;

(c) $\|\tilde{u}^k_B\| \to 0$ as $k \to \infty$;

(d) $\{\tilde{u}^k\}$ is bounded;

(e) for every integer k sufficiently large,

$$\frac{\lambda_k}{2(1-\lambda_k)} \leq 1 + \eta_k, \quad \text{where} \quad \eta_k \equiv \left\|\frac{\tilde{u}^k_N}{e^T\tilde{u}^k_N} - \frac{1}{|N|}e\right\|.$$

Define

$$\phi(\tilde{u}^k,\lambda_k) \equiv |N|\,\log\left(1 - \frac{\lambda_k}{\chi(\tilde{u}^k)}\|\tilde{u}^k\|^2\right) - \sum_{i\in N}\log\left(1 - \frac{\lambda_k}{\chi(\tilde{u}^k)}\tilde{u}^k_i\right)$$

Then, there exists $\tau > 0$ such that, for every $k \geq 0$, $\phi(\tilde{u}^k,\lambda_k)$ is well-defined and

$$\phi(\tilde{u}^k,\lambda_k) \leq -\tau\lambda_k\eta_k^2\left[1 - \frac{\lambda_k}{2(1-\lambda_k)}(1+\eta_k)^{-1}\right] + \mathcal{O}\left(|e^T\tilde{u}^k_N - 1| + \|\tilde{u}^k_B\|^2\right). \tag{105}$$

Proof. Using (b), (d) and the definition of $\{\eta_k\}$, it is easy to see that $\{\eta_k\}$ is bounded. Due to (e), this implies that there exists $\delta' \in (0,1)$ such that

$$\lambda_k \leq 1 - \delta', \qquad \forall k \geq 0. \tag{106}$$

Well-definedness of $\phi(\tilde{u}^k,\lambda_k)$ follows from (a) and the fact that

$$1 - \frac{\lambda_k}{\chi(\tilde{u}^k)}\tilde{u}^k_i \geq 1 - \lambda_k \geq \delta', \qquad \forall k \geq 0.$$

Using conditions (b) and (c), it is easy to see that $\chi(\tilde{u}^k) = \chi(\tilde{u}^k_N)$ for every k sufficiently large. Let $w^k_N = \tilde{u}^k_N / e^T \tilde{u}^k_N$ for every $k \geq 0$. Since $e^T w^k_N = 1$, we have $\chi(w^k_N) \geq 1/|N|$ for every k. In view of (b) and the above observations, we obtain

$$\liminf \chi(\tilde{u}^k) = \liminf \chi(\tilde{u}^k_N) = \liminf \chi(w^k_N) \geq 1/|N|. \tag{107}$$

Using the fact that $\chi(\tilde{u}^k) = \chi(\tilde{u}^k_N)$ and the definition of w^k_N, we obtain

$$\frac{\lambda_k}{\chi(w^k_N)} w^k_j = \frac{\lambda_k}{\chi(\tilde{u}^k)} \tilde{u}^k_j, \qquad \forall j \in N \tag{108}$$

and

$$\begin{aligned} \frac{\lambda_k}{\chi(w^k_N)} \|w^k_N\|^2 - \frac{\lambda_k}{\chi(\tilde{u}^k)} \|\tilde{u}^k\|^2 &= \frac{\lambda_k \|\tilde{u}^k_N\|^2}{\chi(\tilde{u}^k)(e^T \tilde{u}^k_N)} - \frac{\lambda_k}{\chi(\tilde{u}^k)} (\|\tilde{u}^k_N\|^2 + \|\tilde{u}^k_B\|^2) \\ &= \frac{1 - e^T \tilde{u}^k_N}{e^T \tilde{u}^k_N} \frac{\lambda_k \|\tilde{u}^k_N\|^2}{\chi(\tilde{u}^k)(e^T \tilde{u}^k_N)} - \frac{\lambda_k}{\chi(\tilde{u}^k)} \|\tilde{u}^k_B\|^2 \\ &= \mathcal{O}(|e^T \tilde{u}^k_N - 1| + \|\tilde{u}^k_B\|^2). \end{aligned} \tag{109}$$

where the last equality follows from (b), (d), (106) and (107). Now, using (108), (109), Lemma A.1, Lemma A.3 and (a), we obtain

$$\begin{aligned} \phi(\tilde{u}^k, \lambda_k) &= |N| \log \left(1 - \frac{\lambda_k}{\chi(w^k_N)} \|w^k_N\|^2\right) - \sum_{i \in N} \log \left(1 - \frac{\lambda_k}{\chi(w^k_N)} w^k_i\right) \\ &\quad + |N| \left[\log \left(1 - \frac{\lambda_k}{\chi(\tilde{u}^k)} \|\tilde{u}^k\|^2\right) - \log \left(1 - \frac{\lambda_k}{\chi(w^k_N)} \|w^k_N\|^2\right)\right] \\ &\leq -\frac{|N| \, (\lambda_k / \chi(w^k_N))}{|N| - (\lambda_k / \chi(w^k_N))} \eta_k^2 \left(|N| - \frac{\lambda_k}{2(1 - \lambda_k)} \chi(w^k_N)^{-1}\right) \\ &\quad + \mathcal{O}\left(|e^T \tilde{u}^k_N - 1| + \|\tilde{u}^k_B\|^2\right). \end{aligned} \tag{110}$$

We now estimate the first term in the right hand side of (110). Using Lemma A.2 with $v = w^k_N - (1/|N|)e$, it is easy to verify that $\chi(w^k_N) \geq (1 + \eta_k)/|N|$. Hence,

$$|N| - \frac{\lambda_k}{2(1 - \lambda_k)} \chi(w^k_N)^{-1} \geq |N| \left(1 - \frac{\lambda_k}{2(1 - \lambda_k)} (1 + \eta_k)^{-1}\right) \geq 0, \tag{111}$$

where the last inequality follows from (e). Also, since $\chi(w^k_N) \leq \|w^k_N\|$ and $\{w^k_N\}$ is bounded, it follows that $\chi(w^k_N)^{-1} \geq \epsilon$, for some constant $\epsilon > 0$. Hence,

$$\frac{|N| \, (\lambda_k / \chi(w^k_N))}{|N| - (\lambda_k / \chi(w^k_N))} \geq \frac{\lambda_k}{\chi(w^k_N)} \geq \lambda_k \epsilon > 0. \tag{112}$$

Combining (110), (111) and (112) and defining $\tau \equiv \epsilon |N|$, we obtain (105). ∎

Appendix B

The objective of this section is to provide a proof of Lemma 3.1.

Lemma B.1: *Let $h \in \mathbf{R}^n$ and a partition (B,N) of $\{1,\ldots,n\}$ be given and define $p(x) = P(x)h$, $\tilde{p}_B(x_B) = \tilde{P}_B(x_B)h_B$ and $\tilde{p}_N(x_N) = \tilde{P}_N(x_N)h_N$. Then, the following statements hold:*

(a) The vector $\delta_B(x) = p_B(x) - \tilde{p}_B(x_B)$ is the unique optimal solution of the problem

$$\begin{array}{ll} \mathit{minimize}_{\delta_B} & \frac{1}{2}\|\delta_B\|^2 \\ \mathit{subject\ to} & A_B X_B \delta_B = -A_N X_N p_N(x); \end{array} \tag{113}$$

(b) The pair $(p_B(x), \delta_N(x))$, where $\delta_N(x) \equiv p_N(x) - \tilde{p}(x_N)$, is the unique optimal solution of the problem

$$\begin{array}{ll} \mathit{minimize}_{(p_B,\delta_N)} & \frac{1}{2}\|p_B\|^2 + \frac{1}{2}\|\delta_N\|^2 - h_B^T p_B \\ \mathit{subject\ to} & A_B X_B p_B + A_N X_N \delta_N = -A_N X_N \tilde{p}_N(x_N); \end{array} \tag{114}$$

Proof. We first show (a). Since $p(x)$ is the optimal solution of (13), we know that

$$p(x) - h \in \mathrm{Range}(XA^T), \qquad AXp(x) = 0. \tag{115}$$

Since $\tilde{p}_B(x_B)$ is the optimal solution of (15), we have

$$\tilde{p}_B(x_B) - h_B \in \mathrm{Range}(X_B A_B^T), \qquad A_B X_B \tilde{p}_B(x_B) = 0. \tag{116}$$

From (115) and (116), it follows that

$$p_B(x) - \tilde{p}_B(x_B) \in \mathrm{Range}(X_B A_B^T), \qquad A_B X_B [p_B(x) - \tilde{p}_B(x_B)] = -A_N X_N p_N(x),$$

and this shows that $\delta_B(x) = p_B(x) - \tilde{p}_B(x_B)$ satisfies the optimality conditions for (113). Hence, (a) follows. We next show (b). Since $\tilde{p}_N(x_N)$ is the optimal solution of (14), we conclude

$$\begin{pmatrix} 0 \\ \tilde{p}_N(x_N) - h_N \end{pmatrix} \in \mathrm{Range}(XA^T), \qquad A_N X_N \tilde{p}_N(x_N) \in \mathrm{Range}(A_B). \tag{117}$$

Hence, it follows from (115) and (117) that

$$\begin{pmatrix} p_B(x) - h_B \\ \delta_N(x) \end{pmatrix} \in \mathrm{Range}(XA^T), \qquad A_B X_B p_B(x) + A_N X_N \delta_N(x) = -A_N X_N \tilde{p}_N(x_N), \tag{118}$$

which shows that $(p_B(x), \delta_N(x))$ satisfies the optimality conditions for (114). ∎

Lemma B.2: *There exists a constant $K_0 > 0$ such that*

$$\|\delta_B(x)\| \le K_0 \Delta_N(x) \|p_N(x)\|, \qquad \forall x \in \mathcal{P}^{++}. \tag{119}$$

Proof. By (115), we know that $A_N X_N p_N(x) \in \text{Range}(A_B)$. Hence, Lemma 4.1 implies the existence of a constant $K_0 > 0$ with the property that for every $x \in \mathcal{P}^{++}$ there exists a vector $u_B(x)$ satisfying

$$A_B u_B(x) = -A_N X_N p_N(x), \qquad \|u_B(x)\| \leq K_0 \|X_N p_N(x)\|.$$

Thus, $X_B^{-1} u_B(x)$ is a feasible solution for (113). In view of Lemma B.1(a), this implies that for every $x \in \mathcal{P}^{++}$,

$$\|\delta_B(x)\| \leq \|X_B^{-1} u_B(x)\| \leq \|x_B^{-1}\| \|u_B(x)\| \leq K_0 \|x_B^{-1}\| \|X_N p_N(x)\| \leq K_0 \Delta_N(x) \|p_N(x)\|.$$

■

Lemma B.3: *There exist constants $K_1 > 0$ and $K_2 > 0$ with the property that, for every $x \in \mathcal{P}^{++}$ with $\Delta_N(x) \equiv \|x_B^{-1}\| \, \|x_N\|$ sufficiently small, there hold*

$$\|\delta_N(x)\| \leq K_1 \left[\Delta_N(x)^2 \|h_N\| + \Delta_N(x) \|h_B\| \right] \tag{120}$$

and

$$\|\delta_B(x)\| \leq K_2 \left[\Delta_N(x)^2 \|h_B\| + \Delta_N(x) \|h_N\| \right]. \tag{121}$$

Proof. We first prove that (120) holds. From (117) and (118), we have that $A_N X_N \delta_N(x) \in \text{Range}(A_B)$. By Lemma 4.1, there exists a constant $K_0 > 0$ (same as in Lemma B.2) such that, for every $x \in \mathcal{P}^{++}$, there exists a vector $v_B(x)$ satisfying

$$A_B v_B(x) = -A_N X_N \delta_N(x), \qquad \|v_B(x)\| \leq K_0 \|X_N \delta_N(x)\| \leq K_0 \|x_N\| \, \|\delta_N(x)\|. \tag{122}$$

Relation (118) and (122) imply that $A_B[X_B p_B(x) - v_B(x)] = -A_N X_N \tilde{p}_N(x_N)$, and hence that the pair $(p_B(x) - X_B^{-1} v_B(x), 0)$ is feasible for (114). This together with Lemma B.1(b) imply that

$$\frac{1}{2}\|p_B(x)\|^2 + \frac{1}{2}\|\delta_N(x)\|^2 - h_B^T p_B(x) \leq \frac{1}{2}\|p_B(x) - X_B^{-1} v_B(x)\|^2 - h_B^T [p_B(x) - X_B^{-1} v_B(x)].$$

Rearranging this last expression and using relation (122) and the fact that $\|u\|^2 - \|v\|^2 \leq \|u - v\| \, \|u + v\|$ for any two vectors u and v, we obtain

$$\begin{aligned}
\frac{1}{2}\|\delta_N(x)\|^2 &\leq \frac{1}{2}\left\{ \|p_B(x) - X_B^{-1} v_B(x)\|^2 - \|p_B(x)\|^2 \right\} + h_B^T [X_B^{-1} v_B(x)] \\
&\leq \frac{1}{2}\|2p_B(x) - X_B^{-1} v_B(x)\| \, \|X_B^{-1} v_B(x)\| + \|h_B\| \, \|X_B^{-1} v_B(x)\| \\
&\leq \|x_B^{-1}\| \, \|v_B(x)\| \left\{ \frac{1}{2}\|2p_B(x) - X_B^{-1} v_B(x)\| + \|h_B\| \right\} \\
&\leq K_0 \, \Delta_N(x) \|\delta_N(x)\| \left\{ \|p_B(x)\| + \frac{1}{2}\|x_B^{-1}\| \, \|v_B(x)\| + \|h_B\| \right\},
\end{aligned}$$

from which it follows that

$$\|\delta_N(x)\| \leq 2K_0 \, \Delta_N(x) \left\{ \|p_B(x)\| + \frac{1}{2}\|x_B^{-1}\| \, \|v_B(x)\| + \|h_B\| \right\}.$$

We now bound the term within the parentheses in the last expression. Using (122), Lemma B.2 and the fact that $\|\tilde{p}_B(x_B)\| \leq \|h_B\|$ and $\|\tilde{p}_N(x_N)\| \leq \|h_N\|$, we obtain

$$\begin{aligned}
&\|p_B(x)\| + \frac{1}{2}\|x_B^{-1}\|\,\|v_B(x)\| + \|h_B\| \\
&\leq \|\delta_B(x)\| + \|\tilde{p}_B(x_B)\| + \frac{1}{2}K_0\,\|x_B^{-1}\|\|X_N\|\|\delta_N(x)\| + \|h_B\| \\
&\leq K_0\,\Delta_N(x)\|p_N(x)\| + \frac{1}{2}K_0\,\Delta_N(x)\|\delta_N(x)\| + 2\|h_B\| \\
&\leq K_0\,\Delta_N(x)\left[\|p_N(x)\| + \frac{1}{2}\|\delta_N(x)\|\right] + 2\|h_B\| \\
&\leq K_0\,\Delta_N(x)\left[\|\tilde{p}_N(x)\| + \frac{3}{2}\|\delta_N(x)\|\right] + 2\|h_B\| \\
&\leq K_0\,\Delta_N(x)\left[\|h_N\| + \frac{3}{2}\|\delta_N(x)\|\right] + 2\|h_B\|
\end{aligned}$$

Combining the last two inequalities and rearranging, we have

$$\begin{aligned}
\|\delta_N(x)\| &\leq \frac{1}{1 - 3K_0^2\Delta_N(x)^2}\left[2K_0^2\Delta_N(x)^2\|h_N\| + 4K_0\,\Delta_N(x)\|h_B\|\right] \\
&\leq K_1\left[\Delta_N(x)^2\|h_N\| + \Delta_N(x)\|h_B\|\right],
\end{aligned}$$

for every $x \in \mathcal{P}^{++}$ such that $\Delta_N(x) \leq 1/\sqrt{6K_0}$, where $K_1 \equiv 4\max\{K_0^2, 2K_0\}$. We next show that (121) holds. Using Lemma B.2 and the fact that $\|\tilde{p}_N(x_N)\| \leq \|h_N\|$, we obtain

$$\begin{aligned}
\|\delta_B(x)\| &\leq K_0\Delta_N(x)\|p_N(x)\| \\
&\leq K_0\Delta_N(x)\,[\,\|\tilde{p}_N(x_N)\| + \|\delta_N(x)\|\,] \\
&\leq K_0\Delta_N(x)\left[\|h_N\| + K_1\Delta_N(x)^2\|h_N\| + K_1\Delta_N(x)\|h_B\|\right] \\
&\leq K_2\left[\Delta_N(x)^2\|h_B\| + \Delta_N(x)\|h_N\|\right],
\end{aligned}$$

for every $x \in \mathcal{P}^{++}$ such that $\Delta_N(x) \leq 1/\sqrt{K_1}$, where $K_2 \equiv \max\{2K_0, K_0K_1\}$. ∎

We are now ready to give the proof of Lemma 3.1.

Proof of Lemma 3.1: To show (a), let $x \in \mathcal{P}^{++}$ be given. The symmetry of the matrix $R(x)$ follows trivially from the fact that $P(x)$, $\tilde{P}_B(x_B)$ and $\tilde{P}_N(x_N)$ are symmetric matrices. From (117) and (118), we have that $A_N X_N \delta_N(x) \in \text{Range}(A_B)$. Hence, it follows from (14) that $\tilde{P}_N(x_N)\delta_N(x) = \delta_N(x)$, or equivalently,

$$\tilde{P}_N(x_N)R_{NB}(x)h_B + \tilde{P}_N(x_N)R_{NN}(x)h_N = R_{NB}(x)h_B + R_{NN}(x)h_N.$$

Since the last relation holds for every vector $h = (h_B, h_N)$, we conclude that $\tilde{P}_N(x_N)R_{NB}(x) = R_{NB}(x)$ and $\tilde{P}_N(x_N)R_{NN}(x) = R_{NN}(x)$. This shows the first relation in (17). Using the fact that $\tilde{P}_N(x_N)R_{NN}(x) = R_{NN}(x)$ and the fact that $R_{NN}(x)$ and $\tilde{P}_N(x_N)$ are symmetric matrices, we obtain

$$\tilde{P}_N(x_N)R_{NN}(x)\tilde{P}_N(x_N) = R_{NN}(x)\tilde{P}_N(x_N) = \left(\tilde{P}_N(x_N)R_{NN}(x)\right)^T = R_{NN}(x)^T = R_{NN}(x),$$

which shows the second relation in (17). We now show (b). Relation (120) implies that

$$\|\delta_B(x)\| = \|R_{BB}(x)h_B + R_{BN}(x)h_N\| \leq K_1(\Delta_N(x)^2\|h_B\| + \Delta_N(x)\|h_N\|),$$

for every $x \in \mathcal{P}^{++}$ with $\Delta_N(x)$ sufficiently small. Since this relation holds for every $h = (h_B, h_N)$, we conclude that $\|R_{BB}(x)\| \leq K_1\Delta_N(x)^2$ and $\|R_{BN}(x)\| \leq K_1\Delta_N(x)$, for every $x \in \mathcal{P}^{++}$ with $\Delta_N(x)$ sufficiently small. Similarly, using (121), it is easy to see that $\|R_{NN}(x)\| \leq K_2\Delta_N(x)^2$, for every $x \in \mathcal{P}^{++}$ with $\Delta_N(x)$ sufficiently small. Hence (b) follows if we let $K_0 \equiv \max\{K_1, K_2\}$.

References

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

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

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

[4] J. F. Bonnans and M. Bouhtou, *The trust region affine scaling interior point algorithm for convex and nonconvex quadratic programming*, manuscript, INRIA, B.P. 105, 78153, Rocquencourt, France, July 1993.

[5] I. I. Dikin, *Convergence of a sequence of dual variables at solution of a completely degenerate problem of linear programming.* Talk given at the 15th International Symposium on Mathematical Programming, An Arbor, Michigan, USA, August 15–19, 1994.

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

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

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

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

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

[11] C. C. Gonzaga and L. A. Carlos, *A primal affine–scaling algorithm for linearly constrained convex programs*, Technical Report ES–238/90, Dept. of Systems Engineering and Computer Science, COPPE Federal University of Rio de Janeiro, 21941 Rio de Janeiro, RJ, Brazil, December 1990.

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

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

[14] R. D. C. Monteiro, *Convergence and boundary behavior of the projective scaling trajectories for linear programming*, Mathematics of Operations Research, 16 (1991), pp. 842–858.

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

[16] R. Saigal, *A simple proof of primal affine scaling method*, Technical Report, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan, USA, December 1992.

[17] J. Sun, *A convergence proof for an affine-scaling algorithm for convex quadratic programming without nondegeneracy assumptions*, Mathematical Programming, 60 (1993), pp. 69–79.

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

[19] T. Tsuchiya, *Global convergence of the affine scaling algorithm for degenerate strictly convex quadratic programming problems.* Talk presented at TIMS/ORSA Joint Nationtal Meeting, Chicago, Illinois, May, 1993.

[20] ——, *Global convergence of the affine–scaling methods for degenerate linear programming problems*, Mathematical Programming, 52 (1991), pp. 377–404.

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

[22] ——, *Global convergence of the affine scaling algorithm for primal degenerate strictly convex quadratic programming problems*, Annals of Operations Research, 47 (1993), pp. 509–539.

[23] T. Tsuchiya and R. D. C. Monteiro, *Superlinear convergence of the affine scaling algorithm*, Technical Report, Dept. of Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721, USA, November 1992. Conditionally accepted for publication in *Mathematical Programming.*

[24] T. Tsuchiya and M. Muramatsu, *Global convergence of a long–step affine scaling algorithm for degenerate linear programming problems*, Research Memorandum 423, The Institute of Statistical Mathematics, 4–6–7 Minami–Azabu, Minato–ku, Tokyo 106, Japan, January 1992 (revised, September, 1992). (To appear in *SIAM Journal on Optimization*, Vol. 5, 1995).

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

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

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

[28] Y. Ye, *An extension of karmarkar's projective algorithm and the trust region method for quadratic programming*, in Progress in Mathematical Programming: Interior Point and Related Methods, N. Megiddo, ed., Springer, Berlin, 1989, pp. 49–63.

[29] ——, *On an affine scaling algorithm for nonconvex quadratic programming*, Mathematical Programming, 52 (1992), pp. 285–300.

[30] Y. Ye and E. Tse, *An extension of karmarkar's projective algorithm for convex quadratic programming*, Mathematical Programming, 44 (1989), pp. 157–180.