Superlinear Primal-Dual Affine Scaling Algorithms for LCP

R. D. C. Monteiro* S. J. Wright†

Abstract

We describe an interior-point algorithm for monotone linear complementarity problems in which primal-dual affine scaling is used to generate the search directions. The algorithm is shown to have global and superlinear convergence with Q-order up to (but not including) two. The technique is shown to be consistent with a potential-reduction algorithm, yielding the first potential-reduction algorithm that is both globally and superlinearly convergent.

Key words: interior-point methods, primal-dual affine scaling, linear programming, linear complementarity

1 Introduction

During the past three years, we have seen the appearance of several papers dealing with primal-dual interior-point algorithms for linear programs (LP) and monotone linear complementarity problems (LCP) that are superlinearly or quadratically convergent. For LP, these works include McShane [7], Mehrotra [8], Tsuchiya [15], Ye [17], Ye et al. [19], Zhang and Tapia [21], and Zhang, Tapia, and Dennis [22]. For LCP, we mention Ji, Potra, and Huang [1], Ji et al. [2], Kojima, Kurita, and Mizuno [3], Kojima, Meggido, and Noma [4], Ye and Anstreicher [18], and Zhang, Tapia, and Potra [23]. The introductory section of Ye and Anstreicher [18] contains an interesting discussion of the historical development of superlinearly convergent primal-dual interior-point algorithms.

*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332. The work of this author was based on research supported by the National Science Foundation under grant DDM-9109404 and the Office of Naval Research under grant N00014-93-1-0234. This work was done while the first author was a faculty member of the Systems and Industrial Engineering Department at the University of Arizona.

†Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439. The work of this author was based on research supported by the Office of Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38.

In this paper, we discuss superlinearly convergent primal-dual affine scaling methods for solving the monotone LCP. This problem consists of finding a vector pair $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ such that

$$y = Mx + q, \tag{1a}$$
$$x \geq 0, \quad y \geq 0, \tag{1b}$$
$$x^T y = 0, \tag{1c}$$

where $q \in \mathbb{R}^n$, and $M \in \mathbb{R}^{n \times n}$ is positive semidefinite. In subsequent discussion, we say that a point (x, y) is *feasible* if it satisfies the equations (1a) and (1b), and *strictly feasible* if (1a) and (1b) are satisfied with $x > 0$, $y > 0$. We refer to (x, y) as a *solution* only if all three conditions in (1) hold.

Previous superlinearly convergent algorithms for (1) have required all iterates to belong to a neighborhood of the central path defined by either

$$\mathcal{N}_2(\beta) = \left\{ (x, y) \;\; \text{feasible} \;\mid\; \|XYe - (x^T y/n)e\|_2 \leq \beta(x^T y/n) \right\}$$

or

$$\mathcal{N}_{-\infty}(\beta) = \left\{ (x, y) \;\; \text{feasible} \;\mid\; x_i y_i \geq (1 - \beta)(x^T y/n), \;\; \forall i = 1, \ldots, n \right\},$$

where

$$X = \operatorname{diag}(x), \qquad Y = \operatorname{diag}(y), \qquad e = (1, 1, \ldots, 1)^T,$$

and $\beta \in [0, 1)$ is a constant. For example, the predictor-corrector algorithms of Ye et al. [19] for LP and Ji, Potra, and Huang [1] and Ye and Anstreicher [18] for LCP use the neighborhood $\mathcal{N}_2$, while the linear programming algorithm of Zhang and Tapia [20] uses $\mathcal{N}_{-\infty}$. In this paper, we use a different neighborhood defined with respect to two parameters $\delta \geq 0$ and $\eta > 0$ by

$$\mathcal{N}(\delta, \eta) = \left\{ (x, y) \;\; \text{strictly feasible} \;\mid\; x_i y_i \geq \eta (x^T y)^{1+\delta}, \;\; \forall i = 1, \ldots, n \right\}. \tag{2}$$

Clearly, $\mathcal{N}(\delta, \eta)$ is equal to the neighborhood $\mathcal{N}_{-\infty}(1 - n\eta)$ when $\delta = 0$. The parameters δ and η that define $\mathcal{N}(\delta, \eta)$ do not need to be changed as the solution is approached to obtain rapid local convergence. In this respect, the neighborhood (2) differs from $\mathcal{N}_2$ and $\mathcal{N}_{-\infty}$, which need to be expanded during the final stages of the algorithm to achieve superlinear convergence (see Ye [17], Ye and Anstreicher [18]).

Our algorithm uses primal-dual affine scaling search directions. These directions are simply Newton steps for the system of nonlinear equations formed by (1a) and the complementarity condition $XYe = 0$. Because of the connection to Newton's method, we would expect such an algorithm to be quadratically convergent if started close to a unique nondegenerate solution and allowed to take full steps. In this paper, we show that by judicious choice of the step size, all iterates will remain in $\mathcal{N}(\delta, \eta)$ while simultaneously achieving fast local convergence. Depending on the choice of δ, the Q-order of the local convergence can lie anywhere in the range $(1, 2)$. Moreover, our nondegeneracy assumption requires only that one of the solutions (x^*, y^*) has $x^* + y^* > 0$, not that the solution is unique.

In Section 2, we define the search directions and find bounds on the components of these directions, in terms of the complementarity gap $x^T y$ and the parameters δ and η that define the neighborhood $\mathcal{N}(\delta,\eta)$. We pay special attention to the case of M skew-symmetric, which occurs when (1) is derived from a linear programming problem. In this case, the bounds on the search directions are a little tighter and are *global*; that is, they hold everywhere in the relative interior of the set of feasible points and not just in the neighborhood $\mathcal{N}(\delta,\eta)$. In Section 3, we show that a specific choice of step length yields a globally and superlinearly convergent algorithm. For particular choices of the parameter δ, the number of iterates is polynomial in the size of the problem. Finally, in Section 4, we show that the algorithm of Section 3 is consistent with a potential reduction algorithm based on the Tanabe-Todd-Ye potential function

$$\psi_q(x,y) = q\log x^T y - \sum_{i=1}^{n} \log x_i y_i, \qquad (q > n), \tag{3}$$

where an Armijo line search with a well-chosen initial trial step length is used. The resulting algorithm is again globally and superlinearly convergent.

Tunçel [16] introduced an algorithm for linear programming that uses affine scaling search directions in conjunction with a penalty function of the form

$$\Psi_\delta(x,y) = (\delta+1)\log(x^T y/n) - \log(\min_j\{x_j y_j\}), \tag{4}$$

for $\delta > 0$. Step lengths are chosen to keep Ψ_δ *constant* from iteration to iteration. This function is closely related to our neighborhood $\mathcal{N}(\delta,\eta)$ since

$$(x,y) \in \mathcal{N}(\delta,\eta) \quad \Leftrightarrow \quad \Psi_\delta(x,y) \le -\log\eta - (1+\delta)\log n.$$

Tunçel [16] proves global convergence but has no superlinear convergence result. In Section 5, we prove as a consequence of our results that Tunçel's method is superlinear for $\delta \in (0,1)$. Mizuno and Nagasawa [9] describe a method for linear programming which also uses affine scaling search directions and the potential function (3). They prove complexity results for an algorithm that takes a steplength greater than a specified minimum value (defined by a formula not unlike our (35)) which does not increase (3).

The following notational conventions are used in the remainder of the paper: Unless otherwise specified, $\|.\|$ denotes the Euclidean norm. For a general vector $z \in \mathbb{R}^n$ and index set $B \subseteq \{1,\ldots,n\}$, z_B denotes the vector made up of components z_i for $i \in B$. If $M \in \mathbb{R}^{n\times n}$ and $B, N \subseteq \{1,\ldots,n\}$, then $M_{B.}$ refers to the submatrix of M consisting of the rows $i \in B$. Similarly, $M_{.N}$ denotes the submatrix of M corresponding to the columns $j \in N$. If $D \in \mathbb{R}^{n\times n}$ is diagonal, we write $D > 0$ when all the diagonal elements are strictly positive, and use the notation D_B to denote the diagonal submatrix constructed from D_{ii}, $i \in B$. We say that (B,N) is a partition of $\{1,\ldots,n\}$ if $B \cup N = \{1,\ldots,n\}$ and $B \cap N = \emptyset$.

2 Technical Results

In this section, we state our assumptions and derive bounds on components of the iterates (x^k, y^k) and the steps $(\Delta x^k, \Delta y^k)$ for $k = 0, 1, \cdots$. It is assumed throughout that (x^k, y^k)

lies in the neighborhood $\mathcal{N}(\delta, \eta)$.

We make use of the following assumptions. The first assumption is implicit throughout the paper; the second is invoked explicitly where needed.

Assumption 1 (Existence of a strictly feasible point) *The set of strictly feasible points for (1) is nonempty.*

Assumption 2 (Nondegeneracy) *There exists a solution (x^*, y^*) of (1) such that $x^*+y^* > 0$.*

When (x^*, y^*) is the vector pair from Assumption 2, we can define a partition (B, N) of $\{1, \ldots, n\}$ by

$$B = \{i \mid x_i^* > 0\}, \qquad N = \{i \mid s_i^* > 0\}.$$

It is easy to show that all solutions to (1) have $x_N^* = 0$ and $y_B^* = 0$.

We consider in this paper the following class of primal-dual affine scaling algorithms.

Algorithm PDA

initially: Let (x^0, y^0) be a strictly feasible point;

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

Let $(\Delta x^k, \Delta y^k)$ denote the solution of the linear system

$$\begin{aligned} Y\Delta x + X\Delta y &= -YXe, && \text{(5a)} \\ \Delta y - M\Delta x &= 0, && \text{(5b)} \end{aligned}$$

where $(x, y) = (x^k, y^k)$, $X = diag(x)$, and $Y = diag(y)$;

Choose $\alpha_k > 0$ such that $(x^k + \alpha_k\Delta x^k, y^k + \alpha_k\Delta y^k)$ is strictly feasible;

Set $(x^{k+1}, y^{k+1}) = (x^k + \alpha_k\Delta x^k, y^k + \alpha_k\Delta y^k)$;

end for

We start by stating a well-known result from linear algebra.

Lemma 2.1 *Let $F \in \mathbb{R}^{p\times q}$ be given. Then there exists a nonnegative constant $C = C(F)$ with the following property: for $g \in \mathbb{R}^p$ such that the system $Fw = g$ is feasible, there exists a solution $\bar{w}$ of $Fw = g$ such that*

$$\|\bar{w}\| \le C\|g\|.$$

The following technical lemma unifies Theorem 2.5 and Lemma A.1 of Monteiro, Tsuchiya and Wang [13], which in turn are based on Theorem 2 of Tseng and Luo [14].

Lemma 2.2 *Let $f \in \mathbb{R}^q$ and $H \in \mathbb{R}^{p \times q}$ be given. Then there exists a nonnegative constant $L = L(f, H)$ with the property that for any diagonal matrix $D > 0$ and any vector $h \in$ Range(H), the (unique) optimal solution $\bar{w} = \bar{w}(D, h)$ of*

$$\min_w \; f^T w + \frac{1}{2} \|Dw\|^2, \qquad \text{subject to } Hw = h, \tag{6}$$

satisfies

$$\|\bar{w}\|_\infty \le L \left\{ |f^T \bar{w}| + \|h\|_\infty \right\}. \tag{7}$$

Proof. We first show that $|f^T \bar{w}| + \|h\|_\infty = 0$ if and only if $\bar{w} = 0$. The reverse assertion is clear, since from (6) we have $h = H\bar{w} = 0$. For the forward assertion, note that the optimality conditions for (6) are have $D^2 \bar{w} + f + H^T u = 0$ for some $u \in \mathbb{R}^p$. Since $f^T \bar{w} = 0$ and $h = 0$, we have

$$\bar{w}^T D^2 \bar{w} = -f^T \bar{w} - \bar{w}^T H^T u = 0.$$

Hence, since $D > 0$, the result follows. Clearly the inequality (7) holds in this case. In the remainder of the proof we show that L can be chosen such that (7) also holds for nonzero solutions to (6).

Assume for contradiction that there exists a sequence of diagonal matrices $\{D^k\}$ with $D_k > 0$ and a sequence $\{h^k\} \subseteq$ *Range*(H) such that $|f^T w^k| + \|h^k\|_\infty > 0$ for every k and

$$\lim_{k \to \infty} \frac{\|w^k\|_\infty}{|f^T w^k| + \|h^k\|_\infty} = \infty,$$

where w^k is the (unique) optimal solution of (6) with $D = D^k$ and $h = h^k$. By taking subsequences of subsequences, if necessary, we can identify a constant $L_1 > 0$ and a nonempty index set $\mathcal{J} \subseteq \{1, \ldots, q\}$ such that

$$\frac{|w_j^k|}{|f^T w^k| + \|h^k\|_\infty} \;\le\; L_1, \qquad \forall j \notin \mathcal{J}; \tag{8}$$

$$\lim_{k \to \infty} \frac{|w_j^k|}{|f^T w^k| + \|h^k\|_\infty} \;=\; \infty, \qquad \forall j \in \mathcal{J}. \tag{9}$$

Let us define the following linear system

$$f^T w = f^T w^k, \tag{10a}$$
$$Hw = h^k, \tag{10b}$$
$$w_j = w_j^k, \qquad \forall j \notin \mathcal{J}, \tag{10c}$$

and note that w^k is a solution of this system. By Lemma 2.1, (10) has a solution $\hat{w}^k$ such that

$$\|\hat{w}^k\|_\infty \le L_2 \left\{ |f^T w^k| + \|h^k\|_\infty + \max_{j \notin \mathcal{J}} (\,|w_j^k|\,) \right\},$$

(the last term being omitted if $\mathcal{J} = \{1, \ldots, q\}$), where $L_2 \geq 0$ is a constant depending only on f, H, and $\mathcal{J}$. By substituting (8) into this bound, we obtain

$$\begin{aligned} \|\hat{w}^k\|_\infty &\leq L_2 \left\{ |f^T w^k| + \|h^k\|_\infty + L_1 \left(|f^T w^k| + \|h^k\|_\infty \right) \right\} \\ &= L_3 \left(|f^T w^k| + \|h^k\|_\infty \right), \end{aligned} \tag{11}$$

where $L_3 \equiv L_2 (1 + L_1)$. From (9), there exists $K \geq 0$ such that for all $k \geq K$ we have

$$|w_j^k| > L_3(|f^T w^k| + \|h^k\|_\infty), \qquad \forall j \in \mathcal{J}. \tag{12}$$

Combining (11) and (12), we have

$$|w_j^k| > \|\hat{w}^k\|_\infty, \qquad \forall j \in \mathcal{J},\ \forall k \geq K.$$

From this relation and the fact that $\hat{w}^k$ satisfies (10c), we obtain

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

Hence, since $\hat{w}^k$ satisfies (10a), we have

$$f^T \hat{w}^k + \frac{1}{2}\|D^k \hat{w}^k\|^2 < f^T w^k + \frac{1}{2}\|D^k w^k\|^2, \qquad \forall k \geq K. \tag{13}$$

This relation together with the fact that $\hat{w}^k$ satisfies (10b) contradicts the assertion that w^k is an optimal solution of (6) with $D = D^k$ and $h = h^k$. ■

Our main aim now is to derive bounds on the components of the search directions generated by Algorithm PDA. We first deal with the case of M skew-symmetric. This special case is of interest mainly because the linear programming formulation

$$\min_w c^T w, \qquad \text{subject to } Aw \geq b,\ w \geq 0,$$

can, with the introduction of a Lagrange multiplier λ, be posed in the form (1) with

$$M = \begin{bmatrix} 0 & A \\ -A^T & 0 \end{bmatrix}, \qquad q = \begin{bmatrix} -b \\ c \end{bmatrix}.$$

The following two lemmas lay the foundations for a global bound on $(\Delta x, \Delta y)$ which is found in Theorem 2.5.

Lemma 2.3 *Assume that the matrix M in (1) is skew-symmetric, that is, $M = -M^T$, and that Assumption 1 holds. Then for any strictly feasible point (x, y), the solution $(\Delta x, \Delta y)$ of (5) solves the quadratic program*

$$\min_{(w,z)} q^T w + \frac{1}{2}\|D^{-1}w\|^2 + \frac{1}{2}\|Dz\|^2, \qquad \text{subject to } z - Mw = 0, \tag{14}$$

where $D \equiv X^{1/2}Y^{-1/2}$.

Proof. From (5b), $(\Delta x, \Delta y)$ is clearly feasible for (14). It remains to verify that the optimality conditions hold, that is,

$$q + D^{-2}\Delta x = M^T u, \qquad D^2 \Delta y = -u, \tag{15}$$

for some $u \in \mathbb{R}^n$. By using relations (5a), (5b), and (1a) and the fact that $M = -M^T$, we obtain

$$q + D^{-2}\Delta x = q + X^{-1}Y\Delta x = q - y - \Delta y = -M(x + \Delta x) = M^T(x + \Delta x)$$

and

$$D^2\Delta y = Y^{-1}X\Delta y = -x - \Delta x.$$

Hence, $(\Delta x, \Delta y)$ satisfies (15) with $u = x + \Delta x$. ∎

Lemma 2.4 *Assume that M in (1) is skew-symmetric. Let a strictly feasible point (x, y) be given, and consider the solution $(\Delta x, \Delta y)$ of (5). Then,*

$$q^T x = x^T y = -q^T \Delta x. \tag{16}$$

Proof. We have

$$q^T x = (y - Mx)^T x = y^T x - x^T M x = y^T x,$$

where the last equality is due to fact that $M = -M^T$. Using this fact again, together with (5a) and (1a), we obtain

$$-x^T y = y^T \Delta x + x^T \Delta y = (Mx + q)^T \Delta x + x^T M \Delta x = q^T \Delta x + x^T M^T \Delta x - x^T M^T \Delta x = q^T \Delta x.$$

∎

Theorem 2.5 *Assume that $M = -M^T$. Then, for every strictly feasible point (x, y), the solution $(\Delta x, \Delta y)$ of (5) satisfies*

$$\|(\Delta x, \Delta y)\|_\infty \le -C_1\, q^T \Delta x = C_1\, x^T y, \tag{17}$$

where $C_1 \ge 0$ is a constant independent of (x, y).

Proof. The bound follows directly from Lemmas 2.2, 2.3, and 2.4. ∎

We stress that this result holds when (x, y) is *any* strictly feasible point, not just when (x, y) lies in the neighborhood $\mathcal{N}(\delta, \eta)$. An interesting consequence of this result is that the sequence $\{(x^k, y^k)\}$ generated by Algorithm PDA when M is skew-symmetric always converges, regardless of the choice of step sizes $\{\alpha_k\}$, though not necessarily to a solution of (1). We offer a formal proof as our next result.

Corollary 2.6 *Assume that $M = -M^T$. Then the sequence $\{(x^k, y^k)\}$ generated by Algorithm PDA converges.*

Proof. From (5) and the skew-symmetry of M, we have

$$0 \le {x^{k+1}}^T y^{k+1} = {x^k}^T y^k (1 - \alpha_k) + \alpha_k^2 {\Delta x^k}^T M \Delta x^k = {x^k}^T y^k (1 - \alpha_k) \le {x^k}^T y^k.$$

Hence, by Lemma 2.4, the sequence $\{{x^k}^T y^k\} = \{q^T x^k\}$ is monotonically decreasing and bounded below by zero and therefore convergent. Using Theorem 2.5, we obtain

$$\|x^{k+1} - x^k\| = \alpha_k \|\Delta x^k\| \le -C_1 \alpha_k (q^T \Delta x^k) = C_1 (q^T x^k - q^T x^{k+1}). \tag{18}$$

Since the sequence $\{q^T x^k\}$ is convergent, the above relation implies that $\{x^k\}$ is a Cauchy sequence and therefore convergent. Therefore $\{y^k\} = \{M x^k + q\}$ is also convergent, and we have the result. ∎

We now turn to the case of general positive semidefinite matrices M. We start with the following technical result which provides bounds on several quantities involving the direction $(\Delta x, \Delta y)$. Its proof is well known and can be found in several papers (see for example Kojima, Mizuno, and Yoshise [6], Mizuno, Todd, and Ye [10], and Monteiro and Adler [12]).

Lemma 2.7 *Let (x, y) be a strictly feasible point and $(\Delta x, \Delta y)$ be the solution of (5). Then*

(a) $\max_{1 \le i \le n} |\Delta x_i \Delta y_i| \le x^T y / 4$;

(b) $0 \le \Delta x^T \Delta y \le x^T y / 4$;

(c) $\max \{\|D^{-1} \Delta x\|, \|D \Delta y\|\} \le (x^T y)^{1/2}$, *where* $D \equiv X^{1/2} Y^{-1/2}$.

We next state some simple results concerning boundedness of certain components of (x, y) and $(\Delta x, \Delta y)$.

Lemma 2.8 *Suppose that Assumption 2 holds. Then there exists a constant $r > 0$ such that*

$$x_i \le (x^T y)/r, \qquad \forall i \in N, \tag{19a}$$
$$y_i \le (x^T y)/r, \qquad \forall i \in B, \tag{19b}$$

for every feasible point (x, y).

Proof. Consider a solution (x^*, y^*) with $x^* + y^* > 0$, as in Assumption 2. Because of monotonicity, we obtain

$$0 \le (x - x^*)^T M (x - x^*) = (x - x^*)^T (y - y^*) = x^T y - x^{*T} y - x^T y^*,$$

where the last equality is due to the fact that $x^{*T} y^* = 0$. Hence, $x^T y^* \le x^T y$, which implies that

$$x_i \le (x^T y)/y_i^*, \qquad \forall i \in N.$$

Similarly, we obtain

$$y_i \le (x^T y)/x_i^*, \qquad \forall i \in B,$$

and the result follows by choosing r to be the smallest component of (x_B^*, y_N^*). ∎

Lemma 2.9 *Suppose that Assumption 2 holds and that $\delta \ge 0$ and $\eta > 0$ are given. Then for any $(x, y) \in \mathcal{N}(\delta, \eta)$ with corresponding $(\Delta x, \Delta y)$ defined by (5), we have*

$$|\Delta x_i| \le \frac{(x^T y)^{1-\frac{\delta}{2}}}{r\eta^{1/2}}, \qquad \forall i \in N, \tag{20a}$$

$$|\Delta y_i| \le \frac{(x^T y)^{1-\frac{\delta}{2}}}{r\eta^{1/2}}, \qquad \forall i \in B, \tag{20b}$$

and

$$x_i \ge r\eta (x^T y)^{\delta}, \qquad \forall i \in B, \tag{21a}$$

$$y_i \ge r\eta (x^T y)^{\delta}, \qquad \forall i \in N, \tag{21b}$$

where r is the constant from Lemma 2.8.

Proof. Let $D = X^{1/2} Y^{-1/2}$, and note from Lemma 2.7(c) that

$$\|D^{-1} \Delta x\| \le (x^T y)^{1/2}. \tag{22}$$

Since $(x, y) \in \mathcal{N}(\delta, \eta)$, we have

$$x_i y_i \ge \eta (x^T y)^{1+\delta}, \qquad \forall i = 1, \ldots, n. \tag{23}$$

Using relations (22), (19a), and (23), we obtain

$$|\Delta x_i| \le x_i \left(\frac{x^T y}{x_i y_i}\right)^{1/2} \le \left(\frac{x^T y}{r}\right) \left(\frac{1}{\eta (x^T y)^{\delta}}\right)^{1/2} = \frac{(x^T y)^{1-\frac{\delta}{2}}}{r\eta^{1/2}}, \qquad \forall i \in N,$$

yielding (20a). Relation (20b) follows similarly. To show (21b), we observe that relations (23) and (19a) imply

$$y_i \ge \frac{\eta (x^T y)^{1+\delta}}{x_i} \ge r\eta (x^T y)^{\delta}, \qquad \forall i \in N,$$

yielding (21b). Similarly, (21a) follows from (23) and (19b). ∎

We now provide upper bounds for the remaining components of the search directions, namely, Δx_B and Δy_N. This part of the development is based on the approach of Ye and Anstreicher [18] and, in particular, the following lemma.

Lemma 2.10 (Ye and Anstreicher [18, Lemma 3.5]) *Suppose that Assumption 2 holds and let $(\Delta x, \Delta y)$ be the solution of (5). Then $(u,v) = (\Delta x_B, \Delta y_N)$ solves the problem*

$$\min_{(u,v)} \tfrac{1}{2}\|D_B^{-1}u\|^2 + \tfrac{1}{2}\|D_N v\|^2,$$

subject to

$$I_{.N}v - M_{.B}u = M_{.N}\Delta x_N - I_{.B}\Delta y_B. \tag{24}$$

Our bounds for Δx_B and Δy_N are given in the following result.

Lemma 2.11 *Suppose that Assumption 2 holds and that $\delta \geq 0$ and $\eta > 0$ are given. Then there exists a constant $C_2 > 0$ independent of δ and η such that for any $(x,y) \in \mathcal{N}(\delta,\eta)$ with corresponding $(\Delta x, \Delta y)$ defined by (5), we have*

$$\|\Delta x_B\|_\infty \leq \frac{C_2}{\eta^{1/2}}(x^T y)^{1-\frac{\delta}{2}}, \qquad \|\Delta y_N\|_\infty \leq \frac{C_2}{\eta^{1/2}}(x^T y)^{1-\frac{\delta}{2}}. \tag{25}$$

Proof. Applying Lemma 2.2 to problem (24), we conclude that there exists a constant $C_3 > 0$ which depends only on the matrix M and the index sets B and N such that

$$\begin{aligned}
\|(\Delta x_B, \Delta y_N)\|_\infty &\leq C_3\|M_{.N}\Delta x_N - I_{.B}\Delta y_B\|_\infty \\
&\leq C_3\left\{\|M_{.N}\|_\infty\|\Delta x_N\|_\infty + \|I_{.B}\|_\infty\|\Delta y_B\|_\infty\right\} \\
&\leq C_3 \max\{\|M_{.N}\|_\infty, \|I_{.B}\|_\infty\}\frac{(x^T y)^{1-\frac{\delta}{2}}}{r\eta^{1/2}},
\end{aligned}$$

where the last inequality is due to relations (20a) and (20b). The result now follows by setting

$$C_2 \stackrel{\triangle}{=} (C_3/r)\max\{\|M_{.N}\|_\infty, \|I_{.B}\|_\infty\}.$$

∎

We note that the result of Lemma 2.11 with $\delta = 0$ is slightly stronger than the one obtained by Ye and Anstreicher [18, Theorem 3.6] in the sense that the constant C_2 does not depend on the size of (x,y). Lemma 2.2 plays a crucial role in deriving this stronger version.

We can now merge the results of Theorem 2.5 with Lemmas 2.9 and 2.11 to obtain the following theorem.

Theorem 2.12 *Suppose that Assumption 2 holds and that $\delta > 0$ and $\eta > 0$ are given. Then there exists a constant $C_4 \geq 1$ independent of δ and η such that for any $(x,y) \in \mathcal{N}(\delta,\eta)$ with $(\Delta x, \Delta y)$ defined by (5), we have*

$$\|\Delta x\|_\infty \leq \frac{C_4}{\eta^\rho}(x^T y)^{1-\rho\delta}, \qquad \|\Delta y\|_\infty \leq \frac{C_4}{\eta^\rho}(x^T y)^{1-\rho\delta},$$

where $\rho = 0$ if M is skew-symmetric and $\rho = 1/2$ otherwise.

3 The Basic Algorithm

In this section, we develop a special version of a primal-dual affine scaling algorithm for which all iterates lie in a neighborhood $\mathcal{N}(\delta,\eta)$, for some $\delta > 0$ and $\eta > 0$. We show that when $\delta \in (0, 1/2)$ ($\delta \in (0,1)$, if M is skew-symmetric), the algorithm is superlinearly convergent with q-order equal to $2 - 2\delta$ (respectively, $2-\delta$). We also derive a bound on the number of iterations to reduce the duality gap below a specified tolerance. For certain choices of δ, the algorithm is shown to have a polynomial bound on the total number of iterations.

The following notation and definitions will be used in the remainder of the paper. Given the strictly feasible point (x, y) and the search direction $(\Delta x, \Delta y)$ from (5), define

$$
\begin{aligned}
(x(\alpha), y(\alpha)) &\triangleq (x,y) + \alpha(\Delta x, \Delta y); \\
\phi &\triangleq \frac{\Delta x^T \Delta y}{x^T y}; \\
\chi &\triangleq \max_{\substack{i=1,\cdots,n \\ \Delta x_i \Delta y_i < 0}} \left(\frac{|\Delta x_i \Delta y_i|}{x_i y_i} \right),
\end{aligned}
\tag{26}
$$

where we assume that the maximum over the empty set is 0. We use ϕ_k and χ_k for the values of ϕ and χ at $(x,y) = (x^k, y^k)$ and $(\Delta x, \Delta y) = (\Delta x^k, \Delta y^k)$.

Lemma 3.1 *Let (x,y) be a strictly feasible point and $(\Delta x, \Delta y)$ be the solution of (5). Then the following statements hold for all $\alpha \in [0,1]$:*

(a) $x(\alpha)^T y(\alpha) = x^T y(1 - \alpha + \alpha^2 \phi)$;

(b) $x_i(\alpha) y_i(\alpha) \ge x_i y_i (1 - \alpha - \alpha^2 \chi)$;

(c) $(1-\alpha) x^T y \le x(\alpha)^T y(\alpha) \le (1 - \alpha/2)^2 x^T y$;

(d) $\Delta x_i \Delta y_i = 0$ *for all* $i = 1, \cdots, n$ *if and only if* $(x(1), y(1))$ *is a solution of (1) with* $x(1) + y(1) > 0$.

Proof. Statements (a) and (b) follow immediately from (5a) and (26). Lemma 2.7(b) implies that $0 \le \phi \le 1/4$, which together with (a) implies statement (c).

To prove (d), we assume first that $\Delta x_i \Delta y_i = 0$ for all $i = 1, \cdots, n$, so that $\phi = 0$. Using (5a), we can partition $\{1, \cdots, n\}$ into index sets B and N such that

$$
\begin{aligned}
i \in B &\Rightarrow \Delta y_i = -y_i, \quad \Delta x_i = 0, \\
i \in N &\Rightarrow \Delta x_i = -x_i, \quad \Delta y_i = 0.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
i \in B &\Rightarrow y_i(1) = 0, \quad x_i(1) = x_i > 0, \\
i \in N &\Rightarrow x_i(1) = 0, \quad y_i(1) = y_i > 0,
\end{aligned}
$$

and so $(x(1), y(1))$ satisfies (1b) and (1c) and $x(1) + y(1) > 0$, The equations $y = Mx + q$ and (5b) imply that (1a) is also satisfied, so the forward implication is true. For the converse, assume $(x(1), y(1))$ solution with $x(1) + y(1) > 0$. Partitioning $\{1, \cdots, n\}$ into $B = \{i \,|\, x_i(1) > 0\}$ and $N = \{i \,|\, y_i(1) > 0\}$, we have using (5a) that

$$\begin{aligned} i \in B &\;\Rightarrow\; \Delta y_i = -y_i, \;\; \Delta x_i = 0, \\ i \in N &\;\Rightarrow\; \Delta x_i = -x_i, \;\; \Delta y_i = 0. \end{aligned}$$

Hence, the converse is proved, and so (d) holds. ■

Lemma 3.2 *If $u \le 1$ and $\delta \in (0, 1]$, then $(1-u)^{1+\delta} \le 1 - (1+\delta)u + \delta u^2$.*

Proof. By the mean value theorem, we have that

$$(1-u)^{\delta} = 1 - \delta u + (\delta - 1)\delta(1 - \bar{u})^{\delta - 2}\frac{u^2}{2}$$

for some $\bar{u}$ between 0 and u. The last term in the above expression is nonpositive, and therefore $(1-u)^{\delta} \le 1 - \delta u$. Hence,

$$(1-u)^{1+\delta} = (1-u)(1-u)^{\delta} \le (1-u)(1-\delta u) = 1 - (1+\delta)u + \delta u^2,$$

giving the result. ■

Lemma 3.3 *Let $\delta \in (0, 1]$ and $\eta > 0$ be given, and assume that $(x, y) \in \mathcal{N}(\delta, \eta)$. Let $(\Delta x, \Delta y)$ be defined by (5), and suppose that $\Delta x_i \Delta y_i \ne 0$ for some $i = 1, \cdots, n$, so that $(x(1), y(1))$ does not solve (1). Then, for every*

$$\alpha \in J \triangleq \left[0, \frac{\delta}{\delta + \chi + (1+\delta)\phi}\right],$$

the following statements hold:

(a) $1 - \alpha - \alpha^2 \chi > 0$;

(b) $(x(\alpha), y(\alpha)) \in \mathcal{N}(\delta, \eta)$.

Proof. Since $\Delta x_i \Delta y_i \ne 0$ for some i, we can easily see that either $\chi > 0$ or $\phi > 0$, or both. Hence, $J \subseteq [0, 1)$. In view of Lemma 3.1(c) and $x^T y > 0$, this implies that $x(\alpha)^T y(\alpha) > 0$ for all $\alpha \in J$. Using Lemma 3.1(a), Lemma 3.2, and the inclusions $\phi \in [0, 1/4]$, $\alpha \in [0, 1)$, we obtain

$$\begin{aligned} 0 &< \eta \left[x(\alpha)^T y(\alpha)\right]^{1+\delta} \\ &= \eta (x^T y)^{1+\delta}[1 - \alpha(1 - \alpha\phi)]^{1+\delta} \\ &\le \eta (x^T y)^{1+\delta}[1 - (1+\delta)\alpha(1-\alpha\phi) + \delta\alpha^2(1-\alpha\phi)^2] \\ &\le \eta (x^T y)^{1+\delta}[1 - (1+\delta)\alpha(1-\alpha\phi) + \delta\alpha^2] \\ &\le \eta (x^T y)^{1+\delta}[1 - \alpha - \alpha^2\chi], \end{aligned} \tag{27}$$

where in the last inequality we used the fact that the interval J is exactly the set of all $\alpha \geq 0$ for which

$$1 - \alpha(1+\delta)(1-\alpha\phi) + \delta\alpha^2 \leq 1 - \alpha - \alpha^2\chi.$$

Clearly, (27) implies statement (a).

Since $(x, y) \in \mathcal{N}(\delta, \eta)$, we have

$$x_i y_i \geq \eta(x^T y)^{1+\delta}, \qquad \forall i = 1, \cdots, n. \tag{28}$$

Using this relation together with Lemma 3.1(b) and statement (a), we obtain

$$x_i(\alpha)y_i(\alpha) \geq x_i y_i(1 - \alpha - \alpha^2\chi) \geq \eta(x^T y)^{1+\delta}(1 - \alpha - \alpha^2\chi), \qquad \forall i = 1, \cdots, n. \tag{29}$$

Hence, it follows from (27) and (29) that

$$\min_{i=1,\cdots,n} \{x_i(\alpha)y_i(\alpha)\} \geq \eta[x(\alpha)^T y(\alpha)]^{1+\delta} > 0, \qquad \forall \alpha \in J.$$

Thus, $(x(\alpha), y(\alpha)) \in \mathcal{N}(\delta, \eta)$ for all $\alpha \in J$, and the result follows. ∎

Lemma 3.4 *Assume that $(x, y) \in \mathcal{N}(\delta, \eta)$, where $\delta \in (0, 1]$ and $\eta > 0$ are given constants. Then*

$$\eta(x^T y)^\delta \leq \frac{1}{n}. \tag{30}$$

Proof. Since $(x, y) \in \mathcal{N}(\delta, \eta)$, we have

$$x^T y = \sum_{i=1}^n x_i y_i \geq n\eta(x^T y)^{1+\delta},$$

giving the result. ∎

Lemma 3.5 *Suppose that $\delta \in (0, 1]$ and $\eta > 0$ are given. There exists a constant $\bar{C} > 0$ independent of δ and η such that if $(x, y) \in \mathcal{N}(\delta, \eta)$, then*

$$\max\{\phi, \chi\} \leq \bar{C}\frac{(x^T y)^{1-\delta\xi}}{3\eta^\xi}, \tag{31}$$

where $\xi = 1$ if the matrix M is skew-symmetric and $\xi = 2$ otherwise.

Proof. Consider the constant C_4 as in the statement of Theorem 2.12, and define $\bar{C} = 3C_4^2$. We will show that $\bar{C}$ fulfills the requirements of the lemma. For any $(x, y) \in \mathcal{N}(\delta, \eta)$, Theorem 2.12 and Lemma 3.4 imply that

$$\begin{aligned} \phi &= \frac{\Delta x^T \Delta y}{x^T y} \\ &\leq \frac{nC_4^2\,[\,(x^T y)^{2(1-\rho\delta)}\,]/\eta^{2\rho}}{x^T y} \end{aligned}$$

$$\begin{aligned} &= n\eta(x^Ty)^\delta C_4^2 \frac{(x^Ty)^{1-(2\rho+1)\delta}}{\eta^{2\rho+1}} \\ &= n\eta(x^Ty)^\delta C_4^2 \frac{(x^Ty)^{1-\delta\xi}}{\eta^{\xi}} \\ &\leq \bar{C}\frac{(x^Ty)^{1-\delta\xi}}{3\eta^{\xi}}, \end{aligned} \tag{32}$$

where in the last equality we have used the fact that $2\rho+1=\xi$. For a bound on χ, we use $(x,y)\in\mathcal{N}(\delta,\eta)$ and Theorem 2.12 to deduce that

$$\chi \leq \max_{1\leq i\leq n}\left\{\frac{|\Delta x_i\Delta y_i|}{x_iy_i}\right\} \leq \frac{C_4^2\,[\,(x^Ty)^{2(1-\rho\delta)}\,]/\eta^{2\rho}}{\eta(x^Ty)^{1+\delta}} = C_4^2\frac{(x^Ty)^{1-(2\rho+1)\delta}}{\eta^{2\rho+1}} = \bar{C}\frac{(x^Ty)^{1-\delta\xi}}{3\eta^{\xi}}. \tag{33}$$

Clearly, (32) and (33) imply the result. ■

We now state the main theorem concerning convergence of the algorithm. For the purpose of this theorem and the next section we define a measure of centrality of the initial point as follows:

$$\pi_0 = \frac{\min_{i=1,\cdots,n}\{x_i^0y_i^0\}}{{x^0}^Ty^0/n}. \tag{34}$$

Theorem 3.6 *Let $\delta\in(0,1/\xi)$ and the strictly feasible point (x^0,y^0) be given, where ξ is the constant defined in Lemma 3.5. Suppose that Algorithm PDA is applied with step lengths*

$$\alpha_k = \frac{\delta}{\delta+\chi_k+(1+\delta)\phi_k}. \tag{35}$$

Then

(a) the sequence $\{{x^k}^Ty^k\}$ converges monotonically to 0;

(b) given any $\epsilon\in(0,{x^0}^Ty^0)$, the algorithm achieves ${x^k}^Ty^k\leq\epsilon$ in at most

$$O\left(\left(\frac{{x^0}^Ty^0}{\epsilon}\right)^\delta\left(\frac{n}{\delta\pi_0}\right)\log\frac{{x^0}^Ty^0}{\epsilon}\right) \tag{36}$$

iterations;

(c) either the algorithm terminates with ${x^k}^Ty^k=0$ in a finite number of iterations, or else the sequence $\{{x^k}^Ty^k\}$ converges to zero superlinearly with Q-order $2-\delta\xi$.

Proof. First, we observe that there is an index $K\geq 0$ with $\Delta x_i^K\Delta y_i^K=0$ for all $i=1,\cdots,n$ if and only if (x^{K+1},y^{K+1}) is a solution of (1) (see Lemma 3.1(d)). Moreover, using the proof of statement (b) below, one can easily see that if such an index K exists, then it must be of the order in (36). We will henceforth assume that finite termination does not occur and so for all $k \geq 0$, there is at least one $i = 1, \cdots, n$ such that $\Delta x_i^k \Delta y_i^k \neq 0$.

(a) Choosing any $\eta > 0$ such that $(x^0, y^0) \in \mathcal{N}(\delta, \eta)$, we have from relation (35) and Lemma 3.3 that $(x^k, y^k) \in \mathcal{N}(\delta, \eta)$ for all $k \geq 0$. We now find a lower bound on the step length α_k. By Lemma 3.1(c) and the fact that $\alpha_k \in [0, 1]$ for all $k \geq 0$, we have that $\{{x^k}^T y^k\}$ is a decreasing sequence. Hence, using (31), we obtain

$$\max\{\phi_k, \chi_k\} \leq \bar{C} \frac{({x^k}^T y^k)^{1-\delta\xi}}{3\eta^\xi} \leq \bar{C} \frac{({x^0}^T y^0)^{1-\delta\xi}}{3\eta^\xi}, \qquad \forall k \geq 0,$$

which, in view of (35), implies

$$\alpha_k \geq \bar{\alpha} \triangleq \frac{\delta}{\delta + \bar{C}({x^0}^T y^0)^{1-\delta\xi}/\eta^\xi} > 0, \qquad \forall k \geq 0.$$

Using Lemma 3.1(c) again, we get ${x^{k+1}}^T y^{k+1} \leq (1 - \bar{\alpha}/2)^2 {x^k}^T y^k$, $\forall k \geq 0$, which clearly implies statement (a).

(b) Let

$$\eta = \frac{\pi_0}{n({x^0}^T y^0)^\delta}, \tag{37}$$

where π_0 is defined in (34). Note that η is the largest possible choice for which $(x^0, y^0) \in \mathcal{N}(\delta, \eta)$. By relation (35) and Lemma 3.3, we know that $(x^k, y^k) \in \mathcal{N}(\delta, \eta)$ for all $k \geq 0$. Assuming that ${x^k}^T y^k \geq \epsilon$, we now seek bounds on ϕ_k and χ_k that are independent of the somewhat murky constant $\bar{C}$. Using Lemma 2.7(a) and the fact that $(x^k, y^k) \in \mathcal{N}(\delta, \eta)$, we obtain

$$\chi_k \leq \max_{1 \leq i \leq n} \left\{ \frac{|\Delta x_i^k \Delta y_i^k|}{x_i^k y_i^k} \right\} \leq \frac{{x^k}^T y^k / 4}{\eta({x^k}^T y^k)^{1+\delta}} = \frac{1}{4\eta({x^k}^T y^k)^\delta}.$$

Hence, from (37) and the assumption that ${x^k}^T y^k \geq \epsilon$, we obtain

$$\chi_k \leq \frac{1}{4\eta({x^k}^T y^k)^\delta} \leq \frac{n}{4\pi_0(\epsilon/{x^0}^T y^0)^\delta}.$$

Moreover, Lemma 2.7(b) implies that $\phi_k \leq 1/4$. These two last relations together with (35) then imply that

$$\alpha_k \geq \frac{\delta}{\delta + n/[4\pi_0(\epsilon/{x^0}^T y^0)^\delta] + (1/2)} \geq \frac{\delta\pi_0(\epsilon/{x^0}^T y^0)^\delta}{2n},$$

where in the last inequality we used the fact that $\pi_0 \leq 1$ and $\epsilon < {x^0}^T y^0$. Hence, by Lemma 3.1(c), we conclude that

$$\frac{{x^{k+1}}^T y^{k+1}}{{x^k}^T y^k} \leq \left(1 - \frac{\delta\pi_0(\epsilon/{x^0}^T y^0)^\delta}{4n}\right)^2. \tag{38}$$

We have thus shown that (38) holds whenever ${x^k}^T y^k \geq \epsilon$. Now, let K be the smallest nonnegative integer k for which ${x^{k+1}}^T y^{k+1} \leq \epsilon$. Assume for contradiction that

$$K > \left(\frac{{x^0}^T y^0}{\epsilon} \right)^{\delta} \left(\frac{2n}{\delta \pi_0} \right) \log \frac{{x^0}^T y^0}{\epsilon}.$$

By taking logarithms and using the inequality $\log(1-\beta) \leq -\beta$ for $\beta \in (0,1)$, we can show that

$$\left(1 - \frac{\delta \pi_0 (\epsilon / {x^0}^T y^0)^{\delta}}{4n} \right)^{2K} < \frac{\epsilon}{{x^0}^T y^0}.$$

Using this relation and the fact that (38) holds for every $k \leq K$, we obtain

$${x^K}^T y^K \leq \left(1 - \frac{\delta \pi_0 (\epsilon / {x^0}^T y^0)^{\delta}}{4n} \right)^{2K} {x^0}^T y^0 < \epsilon,$$

which contradicts the fact that ${x^K}^T y^K \geq \epsilon$. Therefore (36) holds, and we have proved statement (b).

(c) Choose η as in the proof of part (a). From Lemma 3.1(a), (35), and (31), we have

$$\begin{aligned} {x^{k+1}}^T y^{k+1} &= {x^k}^T y^k (1 - \alpha_k + \alpha_k^2 \phi_k) \\ &\leq {x^k}^T y^k \left(\frac{\chi_k + (1+\delta)\phi_k}{\delta + \chi_k + (1+\delta)\phi_k} + \phi_k \right) \\ &\leq {x^k}^T y^k \left(\frac{\bar{C}({x^k}^T y^k)^{1-\delta\xi}/\eta^{\xi}}{\delta} + \frac{\bar{C}({x^k}^T y^k)^{1-\delta\xi}}{3\eta^{\xi}} \right) \\ &\leq \frac{4\bar{C}}{3\delta\eta^{\xi}} \left({x^k}^T y^k \right)^{2-\delta\xi}, \end{aligned}$$

which implies statement (c). ■

We observe that for certain values of δ, the algorithm of Theorem 3.6 has a polynomial bound on the number of iterations. To see this, assume that $L > 0$ is such that $2^L \geq \max\{{x^0}^T y^0, \epsilon^{-1}\}$. Assume also that the initial point (x^0, y^0) is chosen in such a way that $\pi_0 = O(1)$ is independent of n and L. Then, in terms of n, L, and δ, the algorithm of Theorem 3.6 terminates in $O(nL2^{2L\delta}/\delta)$ iterations. If we choose $\delta = O(1/L)$, then the number of iterations is $O(nL^2)$.

4 A Superlinearly Convergent Potential Reduction Algorithm

Although our algorithm has excellent local convergence properties, preliminary computations have shown that its behavior at points remote from the solution is poor. It is therefore worthwhile to merge our method with other methods with more attractive global convergence properties. In this section, we embed our method in a potential reduction method. The resulting method retains the global convergence behavior of potential reduction methods while exhibiting the fast local convergence associated with the algorithm of Section 3. We are also motivated by a desire to show that potential reduction methods can be superlinearly convergent.

Our potential reduction algorithm is based on the Tanabe-Todd-Ye potential function

$$\psi_q(x,y) = q \log x^T y - \sum_{i=1}^{n} \log x_i y_i, \qquad (q > n). \tag{39}$$

We start by specifying the algorithm and stating a global convergence theorem. Then we show that if q lies in the range $(n, n+1/\xi)$, the algorithm becomes compatible with the method described in the previous section and is superlinearly convergent.

We start by specifying the algorithm.

Algorithm PDPR

initially: Choose fixed constants $\beta > 1$ and $\mu \in (0,1)$. Let $\{\tau_k\}$ be a scalar sequence satisfying $\tau_k \geq 1/2$, and let (x^0, y^0) be strictly feasible;

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

Find $(\Delta x^k, \Delta y^k)$ by solving (5) with $(x,y) = (x^k, y^k)$;

Define $\alpha_k = \beta^{-m_k}\tau_k$, where m_k is the smallest nonnegative integer for which $(x^k(\beta^{-m_k}\tau_k), y^k(\beta^{-m_k}\tau_k)) > 0$ and

$$\psi_q(x^k(\beta^{-m_k}\tau_k), y^k(\beta^{-m_k}\tau_k)) \leq \psi_q(x^k, y^k) + \mu(\beta^{-m_k}\tau_k)\nabla\psi_q(x^k, y^k)^T \begin{bmatrix} \Delta x^k \\ \Delta y^k \end{bmatrix}; \tag{40}$$

Set $(x^{k+1}, y^{k+1}) = (x^k(\alpha_k), y^k(\alpha_k))$;

end for

This algorithm is closely related to the algorithm of Monteiro [11] for convex programming. Its global convergence properties can be analyzed by using techniques like those of Monteiro [11] and Kojima et al. [5]. We omit the details of this analysis and simply state the final theorem.

Theorem 4.1 *Suppose that Assumption 2 holds. Then the sequence of iterates $\{(x^k, y^k)\}$ generated by Algorithm PDPR satisfies*

$$\lim_{k\to\infty} {x^k}^T y^k = 0.$$

The following assumption ensures that Algorithm PDPR enters a superlinear phase in which the local convergence rate of the algorithm of Theorem 3.6 is attained.

Assumption 3 *There is an integer $K > 0$ such that the iterate (x^K, y^K) of Algorithm PDPR satisfies*

$$\min_i x_i^K y_i^K \geq \left(\frac{\bar{C}}{\delta}\right)^{1/\xi} \left(x^{K^T} y^K\right)^{1+1/\xi}, \tag{41}$$

where $\bar{C}$ and ξ are defined in Lemma 3.5.

Suppose that for some $\delta \in (0, 1/\xi)$, the sequence of initial step sizes $\{\tau_k\}$ is selected as

$$\tau_k = \max\left(\frac{1}{2}, \frac{\delta}{\delta + \chi_k + (1+\delta)\phi_k}\right), \qquad \forall k \geq 0. \tag{42}$$

We have the following result.

Lemma 4.2 *Suppose that Assumptions 2 and 3 hold, and let $\{(x^k, y^k)\}$ be the sequence of iterates generated by Algorithm PDPR with the initial step size τ_k defined by (42) for all $k \geq 0$. Define*

$$\eta = \frac{\min x_i^K y_i^K}{(x^{K^T} y^K)^{1+\delta}}, \tag{43}$$

where K is the index in Assumption 3. Then, for all $k \geq K$, we have

(a) $(x^k, y^k) \in \mathcal{N}(\delta, \eta)$; and

(b) $\tau_k = \delta/(\delta + \chi_k + (1+\delta)\phi_k)$.

Proof. Recall that by Lemma 3.1(c), the sequence $\{x^{k^T} y^k\}$ decreases. Using this fact together with relations (41) and (43), one can easily verify that

$$\frac{\bar{C}(x^{k^T} y^k)^{1-\xi\delta}}{\eta^\xi} \leq \delta, \qquad \forall k \geq K. \tag{44}$$

We show first that (a) implies (b) for each $k \geq K$. Assuming (a), we have from (31) and (44) that

$$\frac{\chi_k + (1+\delta)\phi_k}{\delta + \chi_k + (1+\delta)\phi_k} \leq \frac{\bar{C}(x^{k^T} y^k)^{1-\xi\delta}/\eta^\xi}{\delta + \bar{C}(x^{k^T} y^k)^{1-\xi\delta}/\eta^\xi} \leq \frac{\delta}{2\delta} = \frac{1}{2},$$

and therefore, from (42), statement (b) holds.

We now prove (a) by induction. Clearly, by the definition (43), (a) holds for $k = K$. Assume now that (a) is satisfied at some arbitrary $k \geq K$. Then (b) also holds at the index k, and we can use Lemma 3.3(b) and the fact that $\alpha_k \in (0, \tau_k]$ to conclude that $(x^{k+1}, y^{k+1}) \in \mathcal{N}(\delta, \eta)$, giving the result. ■

A result like Theorem 3.6(c) ensures superlinear convergence provided we show that $m_k = 0$ for all k sufficiently large, that is, $\alpha_k = \tau_k$. We do so in the following theorem.

Theorem 4.3 *Suppose that Assumptions 2 and 3 hold, and let $\{(x^k, y^k)\}$ be the sequence of iterates generated by Algorithm PDPR with the initial step size τ_k defined by (42) for some $\delta \in (0, 1/\xi)$. Assume also that $\tau_k < 1$ for all k. Then the sequence $\{{x^k}^T y^k\}$ converges to zero superlinearly with Q-order $2 - \delta\xi$.*

Proof. Global convergence of the sequence $\{{x^k}^T y^k\}$ to zero follows from Theorem 4.1. We next show that $\alpha_k = \tau_k$ for all $k \geq K$ sufficiently large, from which superlinear convergence of $\{{x^k}^T y^k\}$ to zero follows as a consequence of Lemma 4.2(b) and Theorem 3.6(c). To show that $\alpha_k = \tau_k$, we need only show that

$$(x^k(\tau_k), y^k(\tau_k)) > 0, \tag{45}$$

and

$$\psi_q(x^k(\tau_k), y^k(\tau_k)) - \psi_q(x^k, y^k) \leq \mu\tau_k \nabla\psi_q(x^k, y^k)^T \left[\begin{array}{c} \Delta x^k \\ \Delta y^k \end{array} \right]. \tag{46}$$

Lemma 4.2(a) implies that (45) holds for all $k \geq K$. To prove that (46) holds for all $k \geq K$ sufficiently large, we start by finding a lower bound on its right-hand side. Using (39) and (5a), we have

$$\begin{aligned}
&\mu\tau_k \nabla\psi_q(x^k, y^k)^T \left[\begin{array}{c} \Delta x^k \\ \Delta y^k \end{array} \right] \\
&= \mu\tau_k \left\{ \left[\frac{q}{{x^k}^T y^k} y^k - (X^k)^{-1} e \right]^T \Delta x^k + \left[\frac{q}{{x^k}^T y^k} x^k - (Y^k)^{-1} e \right]^T \Delta y^k \right\} \\
&= \mu\tau_k \left\{ \frac{q}{{x^k}^T y^k} ({y^k}^T \Delta x^k + {x^k}^T \Delta y^k) - \sum_{i=1}^n \left[\frac{\Delta x_i^k}{x_i^k} + \frac{\Delta y_i^k}{y_i^k} \right] \right\} \\
&= -\mu\tau_k (q - n) \geq -\mu(q - n).
\end{aligned} \tag{47}$$

To find an upper bound for the left-hand side of (46), note that for all $k \geq K$, we have

$$\begin{aligned}
\frac{\tau_k^2 \max(\phi_k, \chi_k)}{1 - \tau_k} &= \left(\frac{\delta}{\delta + \chi_k + (1+\delta)\phi_k} \right)^2 \left(\frac{\delta + \chi_k + (1+\delta)\phi_k}{\chi_k + (1+\delta)\phi_k} \right) \max(\phi_k, \chi_k) \\
&= \left(\frac{\delta^2}{\delta + \chi_k + (1+\delta)\phi_k} \right) \frac{\max(\phi_k, \chi_k)}{\chi_k + (1+\delta)\phi_k} \\
&\leq \delta.
\end{aligned}$$

Using this bound together with statements (a) and (b) of Lemma 3.1, we obtain

$$\begin{aligned}
&\psi_q(x^k(\tau_k), y^k(\tau_k)) - \psi_q(x^k, y^k) \\
&= q \log \left[\frac{x^k(\tau_k)^T y^k(\tau_k)}{{x^k}^T y^k} \right] - \sum_{i=1}^n \log \left[\frac{x_i^k(\tau_k) y_i^k(\tau_k)}{x_i^k y_i^k} \right]
\end{aligned}$$

$$\begin{aligned}
&\leq\ q\log(1-\tau_k+\tau_k^2\phi_k)-\sum_{i=1}^{n}\log(1-\tau_k-\tau_k^2\chi_k)\\
&=\ (q-n)\log(1-\tau_k)+q\log\left(1+\frac{\tau_k^2\phi_k}{1-\tau_k}\right)-n\log\left(1-\frac{\tau_k^2\chi_k}{1-\tau_k}\right)\\
&\leq\ (q-n)\log(1-\tau_k)+q\log(1+\delta)-n\log(1-\delta). \qquad (48)
\end{aligned}$$

Since ${x^k}^T y^k \to 0$, it follows from (31) and (42) that $1-\tau_k \to 0$. Therefore, from (48), the left-hand side of (46) approaches $-\infty$ as $k\to\infty$. Hence, in view of (47), the inequality (46) will hold for all sufficiently large $k \geq K$.

The proof of superlinearity now follows as in Theorem 3.6(c). ∎

Note that if $\tau_k = 1$ for some k, we have $\Delta x_i^k \Delta y_i^k = 0$ for all i; therefore, from Lemma 3.1(d), (x^{k+1}, y^{k+1}) is an exact solution of (1). In this case, Algorithm PDPR will reject the step $\alpha_k = \tau_k$ because (45) is not satisfied! Any implementation would surely recognize this special case, so we have avoided the complication that it introduces into the analysis above.

Obviously, we cannot explicitly identify the index K required by Assumption 3, since we do not know the value of $\bar{C}$ in general. However, we can be sure that Assumption 3 holds if we choose q to be in the interval $(n, n+1/\xi)$, as we show in the following theorem.

Theorem 4.4 *Suppose that Assumptions 1 and 2 hold and that $q \in (n, n+1/\xi)$. Then there is an index K such that (41) holds at the iterate (x^K, y^K) generated by Algorithm PDPR.*

Proof. Let us define Q and $\bar{Q}$ as

$$Q \stackrel{\triangle}{=} \psi_q(x^0, y^0), \qquad \bar{Q} \stackrel{\triangle}{=} Q-(n-1)\log(n-1).$$

Since ψ_q is reduced at each iteration of Algorithm PDPR, we have $\psi_q(x^k, y^k) \leq Q$ for all $k \geq 0$. Hence, for all $k \geq 0$ and $i = 1, \cdots, n$, we have

$$\begin{aligned}
&(q-n+1)\log {x^k}^T y^k - \log x_i^k y_i^k\\
&\quad\leq\ Q-(n-1)\log {x^k}^T y^k + \sum_{j\neq i}\log x_j^k y_j^k\\
&\quad\leq\ Q-(n-1)\log\left({x^k}^T y^k - x_i^k y_i^k\right) + \sum_{j\neq i}\log x_j^k y_j^k\\
&\quad\leq\ Q-(n-1)\log(n-1)\ =\ \bar{Q},
\end{aligned}$$

where the third inequality follows from the arithmetic-geometric mean inequality, namely: $(\sum_{i=1}^p a_i)/p \geq (\Pi_{i=1}^p a_i)^{1/p}$ for any positive scalars $a_1, \ldots, a_p$. Hence,

$$\log\frac{x_i^k y_i^k}{\left({x^k}^T y^k\right)^{q-n+1}} \geq -\bar{Q} \ \Rightarrow\ x_i^k y_i^k \geq e^{-\bar{Q}}\left({x^k}^T y^k\right)^{q-n+1},$$

and so

$$x_i^k y_i^k \geq \left(\frac{\bar{C}}{\delta}\right)^{1/\xi}\left({x^k}^T y^k\right)^{1+1/\xi}\left\{e^{-\bar{Q}}\left(\frac{\delta}{\bar{C}}\right)^{1/\xi}\left({x^k}^T y^k\right)^{q-n-1/\xi}\right\}.$$

Since $q - n - 1/\xi < 0$ and ${x^k}^T y^k \to 0$, there is a $K > 0$ such that the last bracketed term in the expression is greater than 1 for all $k \geq K$, giving the result. ■

Finally, we combine the results of the last two theorems to obtain the following result.

Corollary 4.5 *Suppose that Assumption 2 holds, that $q \in (n, n + 1/\xi)$, that the sequence of initial step sizes $\{\tau_k\}$ is defined by (42) for some $\delta \in (0, 1/\xi)$, and that $\tau_k < 1$ for all k. Then the sequence $\{{x^k}^T y^k\}$ generated by Algorithm PDPR converges to zero superlinearly with Q-order $2 - \delta\xi$.*

5 Concluding Remarks

In this concluding section, we discuss the close relationship that exists between the algorithm of Theorem 3.6 and the one presented by Tunçel [16]. We also show that Tunçel's algorithm is superlinearly convergent whenever the parameter δ that appears in the potential function (4) lies in the interval $(0, 1)$.

For simplicity, we assume that the affine scaling direction $(\Delta x, \Delta y)$ at any strictly feasible point satisfies $\Delta x_i \Delta y_i \neq 0$, for some $i \in \{1, \cdots, n\}$, so that finite termination of the algorithm never occurs. For the purpose of this section, we also assume that the matrix M in (1) is skew-symmetric. Since Tunçel's algorithm is for linear programs, it fits into this framework.

We describe Tunçel's algorithm in an equivalent but slightly different way from [16] in order to better point out its connection to the neighborhood $\mathcal{N}(\delta, \eta)$. Let the strictly feasible point (x^k, y^k) be the k-th iterate of the algorithm. Tunçel's algorithm sets $(x^{k+1}, y^{k+1}) = (x^k(\alpha_k), y^k(\alpha_k))$, where the stepsize $\alpha_k \in (0, 1)$ is such that the point $(x^k(\alpha_k), y^k(\alpha_k))$ satisfies the equation in (x, y) defined by

$$\frac{\min_{i=1,\cdots,n}\{x_i y_i\}}{(x^T y)^{1+\delta}} = \frac{\min_{i=1,\cdots,n}\{x_i^k y_i^k\}}{({x^k}^T y^k)^{1+\delta}}.$$

Tunçel shows that such a stepsize always exists and is unique. In the proof of this fact, he also shows that $(x^k(\alpha), y^k(\alpha))$ satisfies the strict inequality

$$\frac{\min_{i=1,\cdots,n}\{x_i^k(\alpha) y_i^k(\alpha)\}}{[x^k(\alpha)^T y^k(\alpha)]^{1+\delta}} > \frac{\min_{i=1,\cdots,n}\{x_i^k y_i^k\}}{({x^k}^T y^k)^{1+\delta}}$$

for all $\alpha \in (0, \alpha_k)$. Observe that the iterates of this algorithm satisfy

$$\frac{\min_{i=1,\cdots,n}\{x_i^k y_i^k\}}{({x^k}^T y^k)^{1+\delta}} = \frac{\min_{i=1,\cdots,n}\{x_i^0 y_i^0\}}{({x^0}^T y^0)^{1+\delta}} \triangleq \eta, \qquad \forall k \geq 0,$$

which, in terms of the potential function (4), is equivalent to $\Psi_\delta(x^k, y^k) = \Psi_\delta(x^0, y^0)$ for all $k \geq 0$. It is now easy to see that Tunçel's step size is the largest $\alpha > 0$ for which $(x^k(\alpha), y^k(\alpha)) \in \mathcal{N}(\delta, \eta)$. In view of Lemma 3.3(b), it follows that our step size (35) is less than or equal to Tunçel's step size. Therefore Tunçel's algorithm achieves larger or equal reduction of the duality gap at each iteration while generating all points within the neighborhood $\mathcal{N}(\delta, \eta)$. Hence, the same proof given for Theorem 3.6(c) can be used to show that Tunçel's algorithm is superlinearly convergent with q-order of convergence equal to $2-\delta$, whenever $\delta \in (0,1)$.

Acknowledgment

Part of this work was done while the first author was visiting the Mathematics and Computer Science Division of Argonne National Laboratory. He is grateful to Argonne for its financial support and its congenial scientific atmosphere.

References

[1] J. Ji, F. A. Potra, and S. Huang, *A predictor-corrector method for linear complementarity problems with polynomial complexity and superlinear convergence*, Technical Report 18, Department of Mathematics, University of Iowa, Iowa City, Iowa, August 1991.

[2] J. Ji, F. A. Potra, R. A. Tapia, and Y. Zhang, *An interior-point algorithm with polynomial complexity and superlinear convergence for linear complementarity problems*, Technical Report 91–23, Department of Mathematical Sciences, Rice University, Houston, TX, July 1991.

[3] M. Kojima, Y. Kurita, and S. Mizuno, *Large–step interior point algorithms for linear complementarity problems*, SIAM Journal on Optimization, 3 (1993), pp. 398–412.

[4] M. Kojima, N. Megiddo, and T. Noma, *Homotopy continuation methods for nonlinear complementarity problems*, Mathematics of Operations Research, 16 (1991), pp. 754–774.

[5] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise, *A unified approach to interior-point algorithms for linear complementarity problems*, no. 538 in Lecture Notes in Computer Science, Springer-Verlag, Berlin, 1991.

[6] M. Kojima, S. Mizuno, and A. Yoshise, *A polynomial-time algorithm for a class of linear complementarity problems*, Mathematical Programming, 44 (1989), pp. 1–26.

[7] K. A. McShane, *A superlinearly convergent $O(\sqrt{n}L)$ iteration primal–dual linear programming algorithm*, SIAM Journal on Optimization, 4 (1994), pp. 247–261.

[8] S. Mehrotra, *Quadratic convergence in a primal-dual method*, Technical Report 91–15, Dept. of Industrial Engineering and Management Science, Northwestern University, Evanston, IL 60208, USA, September 1991.

[9] S. Mizuno and A. Nagasawa, *A primal-dual affine scaling potential reduction algorithm for linear programming*, Research Memorandum 427, The Institute of Statistical Mathematics, Minami-Azabu, Minato-ku, Tokyo, Japan, February 1992.

[10] S. Mizuno, M. Todd, and Y. Ye, *On adaptive step primal-dual interior-point algorithms for linear programming*, Mathematics of Operations Research, 18 (1993), pp. 964–981.

[11] R. D. C. Monteiro, *A globally convergent interior point algorithm for convex programming*, Technical Report 91–021, Dept. of Systems and Industrial Engineering, University of Arizona, July 1991.

[12] R. D. C. Monteiro and I. Adler, *Interior path-following primal-dual algorithms. Part I: Linear programming*, Mathematical Programming, 44 (1989), pp. 27–41.

[13] 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.

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

[15] T. Tsuchiya, *Quadratic convergence of Iri and Imai's method for degenerate linear programming problems*, Research Memorandum 412, The Institute of Statistical Mathematics, 4–6–7 Minami–Azabu, Minato–ku, Tokyo 106, Japan, 1991.

[16] L. Tuncel, *Constant potential primal-dual algorithms: A framework*, Tech. Rep. 1018, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, August 1992.

[17] Y. Ye, *On the Q-order of convergence of interior-point algorithms for linear programming*, in Proceedings of the 1992 Symposium on Applied Mathematics, W. Fang, ed., Institute of Applied Mathematics, Chinese Academy of Sciences, 1992, pp. 534–543.

[18] Y. Ye and K. Anstreicher, *On quadratic and $O(\sqrt{n}L)$ convergence of a predictor-corrector algorithm for LCP*, Mathematical Programming, Series A, 62 (1993), pp. 537–551.

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

[20] Y. Zhang and R. A. Tapia, *Superlinear and quadratic convergence of primal–dual interior–point methods for linear programming revisited*, Journal of Optimization Theory and Applications, 73 (1992), pp. 229–242.

[21] ——, *A superlinearly convergent polynomial primal–dual interior–point algorithm for linear programming*, SIAM Journal on Optimization, 3 (1993), pp. 118–133.

[22] Y. Zhang, R. A. Tapia, and J. E. Dennis, *On the superlinear and quadratic convergence of primal-dual interior point linear programming algorithms*, SIAM Journal on Optimization, 2 (1992), pp. 304–324.

[23] Y. Zhang, R. A. Tapia, and F. A. Potra, *On the superlinear convergence of interior point algorithms for a general class of problems*, SIAM Journal on Optimization, 3 (1993), pp. 413–422.