On the Superlinear and Quadratic Convergence of Primal-Dual Interior Point Linear Programming Algorithms

Y. Zhang
R. A. Tapia
J. E. Dennis

Center for Research on Parallel Computa:
Rice University
P.O. Box 1892
Houston, TX 77251-1892

On the Superlinear and Quadratic Convergence
of Primal-Dual Interior Point
Linear Programming Algorithms

Yin Zhang
R.A. Tapia
J.E. Dennis

January, 1990
(revised April, 1991)

TR90-6

On the Superlinear and Quadratic Convergence

of Primal-Dual Interior Point

Linear Programming Algorithms [*][†]

Yin Zhang [‡]Richard A. Tapia [§]and John E. Dennis [¶]

January, 1990

(Revised April, 1991)

Abstract

This paper presents a convergence rate analysis for interior point primal-dual linear programming algorithms. Conditions that guarantee Q-superlinear convergence are identified in two distinct theories. Both state that, under appropriate assumptions, Q-superlinear convergence is achieved by asymptotically taking the step to the boundary of the positive orthant and letting the barrier parameter approach zero at a rate that is superlinearly faster than the convergence of the duality gap to zero. The first theory makes no nondegeneracy assumption and explains why in recent numerical experimentation Q-superlinear convergence was always observed. The second theory requires the

1

restrictive assumption of primal nondegeneracy. However, it gives the surprising result that Q-superlinear convergence can still be attained even if centering is not phased out, provided the iterates asymptotically approach the central path. The latter theory is extended to produce a satisfactory Q-quadratic convergence theory. It requires that the step approach the boundary as fast as the duality gap approaches zero and the barrier parameter approach zero as fast as the square of the duality gap approaches zero.

Keywords: Linear programming, Primal-dual interior point algorithms, Duality-gap-reducing and centering, Newton's method, Q-superlinear and Q-quadratic convergence.

Abbreviated Title: Superlinear and quadratic convergence of primal-dual algorithms

1   Introduction

This paper considers linear programs in the standard form:

$$
\begin{aligned}
\text{minimize} \quad & c^T x \\
\text{subject to} \quad & Ax = b, \\
& x \geq 0,
\end{aligned}
\tag{1.1}
$$

where $c, x \in \mathbf{R}^n$, $b \in \mathbf{R}^m$, $A \in \mathbf{R}^{m \times n} (m < n)$ and A has full rank m. The dual linear program of (1.1) can be expressed in the following symmetric form

$$
\begin{aligned}
\text{minimize} \quad & d^T y \\
\text{subject to} \quad & By = Bc, \\
& y \geq 0,
\end{aligned}
\tag{1.2}
$$

where $y \in \mathbf{R}^n$ is the vector of dual slack variables, $d = A^T(AA^T)^{-1}b$, $B \in \mathbf{R}^{(n-m) \times n}$ has full row rank and $AB^T = 0$ (i.e., the columns of B^T form a basis for the null space of A). This form of the dual was introduced by Todd and Ye in [20]. A pair (x, y) is called strictly feasible if x and y are feasible for (1.1) and (1.2), respectively, and are positive as well.

The weak duality theorem says that the duality gap $x^T y$ is non-negative for any feasible pair (x, y). We will assume that the primal feasibility set contains strictly feasible points and

2

that the set of optimal solutions for the primal linear program is nonempty and bounded. For any optimal feasible pair (x_*, y_*), the duality gap is closed, i.e., $x_*^T y_* = 0$.

Primal-dual interior point algorithms attempt to solve the primal and dual linear programs simultaneously by generating a sequence of strictly feasible pairs $\{(x_k, y_k)\}$ (and often another dual variable vector – the Lagrange multipliers associated with the primal constraints $Ax = b$) that converges to an optimal feasible pair (x_*, y_*). The objective of such algorithms is to drive the duality gap $x_k^T y_k$ to zero. Primal-dual approachs of this form were first introduced by Megiddo [13] using a logarithmic barrier function method. Megiddo's idea was developed by Kojima, Mizuno and Yoshise [8] into a full algorithm with a polynomial complexity bound. A conceptually different approach was proposed by Todd and Ye [20] based on reducing a primal-dual potential function which is analogous to the Karmarkar primal potential function [7]. Other works on primal-dual interior point algorithms include Monteiro and Adler [16], Lustig [10, 9], Gonzaga and Todd [5], Huang and Kortanek [6], Choi, Monma and Shanno [2], McShane, Monma and Shanno [12], and Lustig, Marsten and Shanno [11].

The above works can be classified roughly into two groups. Papers in the first group ([8, 20], for example) focused on designing algorithms with polynomial complexity bounds. Papers in the second group ([2, 11, 12], for example) were more concerned with computational and implementational issues. Unfortunately, there is a discrepancy between the two groups. That is, the algorithms that were described in the second group and were shown to have good practical performance are not those that were studied in the first group and were shown to possess polynomial complexity bounds. This discrepancy is understandably due to the limitation of the worst case analysis used in deriving polynomial complexity bounds. Recently, there have been works aimed at narrowing this discrepancy from a probabilistic point of view, see Mizuno, Todd and Ye [14, 15]. In the current work, we try to shed light on another fundamental aspect of continuous optimization algorithms; namely, the blending of two often conflicting objectives: global convergence and fast local convergence. A convergence rate analysis for algorithms that belong to a very general class of primal-

dual interior point methods is presented. This theory shows how superlinear and quadratic convergence can be attained from primal-dual interior point algorithms.

It is well understood, in the continuous optimization community, that fast local convergence is an important factor in evaluating the efficiency of an iterative method. Moreover, while interior-point algorithms for linear programming are certainly iterative methods, local convergence properties have not received much attention. A plausible explanation for this lack of attention is the common belief that interior-point algorithms essentially possess finite termination. That is, once one gets close enough to the optimal solution set, the interior-point method can be terminated and available information (mainly the zero-nonzero structure of an optimal solution) can be used to obtain an optimal solution through some finite procedure. In the context of this guessing strategy, it is natural to question the value of fast local convergence in linear programming applications. However, our computational experience has taught us that although a correct early guess, on occasion, is certainly possible, especially in the case of a nondegenerate optimal vertex, in general one needs to be very close to the solution set in order to *guarantee* a correct guess. In addition, fast convergence usually occurs much earlier than the standard Newton's method theory predicts; a property often referred to in nonlinear applications as the semi-local behavior of Newton's method. Therefore, the construction of algorithms with fast local convergence can be an important and beneficial activity even in linear programming applications. However, in the interest of conciseness we have decided to present only theory in the present study. A comprehensive numerical investigation is the subject of a current study.

The concept of the central path (trajectory) plays an important role in designing and analyzing interior point algorithms. It was first studied in linear programming by Sonnevend [17] and by Bayer and Lagarias [1], see also Megiddo [13]. The central path can be expressed in several ways. Perhaps the simplest is that a strictly feasible pair (x, y) is on the central path if and only if it satisfies

$$[x]_1[y]_1 = [x]_2[y]_2 = \ldots = [x]_n[y]_n$$

4

where $[x]_i$ ($[y]_i$) is the i-th element of x (y), or equivalently,

$$[x]_i[y]_i = x^T y/n, \quad , i = 1, 2, \ldots, n. \tag{1.3}$$

This paper is organized as follows. In Section 2, we describe a general primal-dual interior point algorithmic framework. Then in Section 3, we present our superlinear convergence rate analysis and in Section 4, we present our quadratic convergence rate analysis. Concluding remarks are given in Section 5.

2 A Primal-Dual Algorithmic Framework

In this section, we describe a general primal-dual interior point algorithmic framework. This general framework can also be derived from the point of view of barrier function methods or potential function reduction methods, as was done, for example, in [8] and [20]. We hope that our somewhat different approach adds new insight to these algorithms.

If the primal variables and the dual slack variables are updated at a given strictly feasible pair (x, y) by the formulas

$$x_+ = X(e + \alpha p) \quad \text{and} \quad y_+ = Y(e + \alpha q), \tag{2.1}$$

where $X = diag(x)$, $Y = diag(y)$, $e \in \mathbf{R}^n$ has all components equal to one, $p, q \in \mathbf{R}^n$ and $\alpha > 0$ is the steplength, then in order for x_+ and y_+ to be strictly feasible, p, q and α must satisfy

$$AXp = 0 \quad \text{and} \quad e + \alpha p > 0, \tag{2.2}$$

$$BYq = 0 \quad \text{and} \quad e + \alpha q > 0. \tag{2.3}$$

We will consider projected-gradient type methods. Namely, the feasible directions p and q are obtained by projecting the negative gradients of relevant functions into the null spaces of AX and BY, respectively. Therefore, we first need to construct two $n \times n$ projection matrices H_p and H_q such that $AXH_p = 0$ and $BYH_q = 0$. If A and B were not scaled by X and Y, respectively, then it would be sufficient to define $H_q = P_A$ and $H_p = I - P_A$, where

$P_A = A^T(AA^T)^{-1}A$. This definition would give $AH_p = 0$ and $BH_q = 0$ because $A^T \perp B^T$. Obviously, in this case both H_p and H_q would be orthogonal projections and therefore would be symmetric and positive semi-definite. The symmetry and positive semi-definiteness of H_p (H_q) is important because for any function $\phi : \mathbf{R}^n \to \mathbf{R}$, the projected negative gradient $-H_p\nabla\phi$ ($-H_q\nabla\phi$) will be not only a primal (dual) feasible direction but also a descent direction for ϕ as long as $H_p\nabla\phi \neq 0$ ($H_q\nabla\phi \neq 0$). Furthermore, it is worth noting that one would only need to compute either H_p or H_q because $H_p + H_q = I$.

Even though the matrices A and B are scaled by X and Y, respectively, it is still possible to construct two projection matrices H_p and H_q based on just one orthogonal projection matrix (though H_p and H_q themselves will not be orthogonal projections) and obtain the desirable property that both H_p and H_q are symmetric, positive semi-definite. Consider the following matrices that we will call *scaled projections*:

$$H_p = \hat{D}(I - \hat{P})\hat{D} \quad \text{and} \quad H_q = \hat{D}\hat{P}\hat{D}.$$

Here $\hat{D}$ is a positive-definite diagonal matrix and $\hat{P}$ is an orthogonal projection matrix, both contained in $\mathbf{R}^{n \times n}$. The equations $A(XH_p) = 0$ and $B(YH_q) = 0$ and the fact $A^T \perp B^T$ imply that $H_pXYH_q = 0$, which in turn requires that $\hat{P}(\hat{D}XY\hat{D})(I - \hat{P}) = 0$. The last equation will hold for any orthogonal projection matrix $\hat{P}$ if $\hat{D}XY\hat{D} = I$. This leads to the following choice for $\hat{D}$,

$$\hat{D} = (XY)^{-\frac{1}{2}}.$$

It now follows from $AXH_p = 0$ that $(AX^{\frac{1}{2}}Y^{-\frac{1}{2}})(I - \hat{P}) = 0$. Hence we need to define the orthogonal projection matrix $\hat{P}$ as the orthogonal projection into the range space of $X^{\frac{1}{2}}Y^{-\frac{1}{2}}A^T$, namely,

$$\hat{P} = X^{\frac{1}{2}}Y^{-\frac{1}{2}}A^T(AXY^{-1}A^T)^{-1}AX^{\frac{1}{2}}Y^{-\frac{1}{2}}. \tag{2.4}$$

This definition of $\hat{P}$ gives not only $AXH_p = 0$, but also $BYH_q = 0$. Therefore, we finally conclude that the choices for the two scaled projection matrices H_p and H_q should be

$$H_p = (XY)^{-\frac{1}{2}}(I - \hat{P})(XY)^{-\frac{1}{2}}, \tag{2.5}$$

$$H_q = (XY)^{-\frac{1}{2}}\hat{P}(XY)^{-\frac{1}{2}}, \tag{2.6}$$

where $\hat{P}$ is defined by (2.4). The proof of the following proposition is now straightforward.

Proposition 2.1 *If H_p and H_q are defined by (2.5) and (2.6), respectively, then*

1. *Both H_p and H_q are symmetric, positive semi-definite;*

2. *$AXH_p = 0$ and $BYH_q = 0$;*

3. *$H_pXYH_q = 0$;*

4. *$H_p + H_q = (XY)^{-1}$.*

Obviously, the scaled projection H_p (H_q) will project the negative gradient into a primal (dual) feasible direction which is also a descent direction (provided that the projection is nonzero). It is worth noting that in order to construct the two scaled projections we only need to calculate one orthogonal projection matrix $\hat{P}$.

To derive the directions p and q in (2.1), we first define a function

$$\phi(u,v) = (e+u)^T XY(e+v). \tag{2.7}$$

Obviously, if $x_+ = X(e+u)$ and $y_+ = Y(e+v)$ are primal and dual feasible, respectively, then $\phi(u,v) = x_+^T y_+ \geq 0$ represents the duality gap at the updated pair (x_+, y_+). It is easy to see that

$$\nabla_u \phi(0,0) = \nabla_v \phi(0,0) = XYe. \tag{2.8}$$

Now define

$$p_\phi = -H_p\nabla_u\phi(0,0) = -[(XY)^{-\frac{1}{2}}(I-\hat{P})(XY)^{-\frac{1}{2}}]XYe, \tag{2.9}$$

$$q_\phi = -H_q\nabla_v\phi(0,0) = -[(XY)^{-\frac{1}{2}}\hat{P}(XY)^{-\frac{1}{2}}]XYe. \tag{2.10}$$

From 1 and 2 of Proposition 2.1, the above defined (p_ϕ, q_ϕ) is clearly a feasible descent direction for $\phi(u,v)$ at the current point $(0,0)$. We call (p_ϕ, q_ϕ) the duality-gap-reducing direction.

Using the formulas (2.1), we define the barrier function at the given strictly feasible pair (x, y) as

$$\psi(u, v) = -\sum_{i=1}^{n} \ln([X(e + u)]_i [Y(e + v)]_i), \tag{2.11}$$

where $[a]_i$ denotes the i-th element of the vector a. The gradient of $\psi(u, v)$ at the current point $(u, v) = (0, 0)$ satisfies

$$\nabla_u \psi(0, 0) = \nabla_v \psi(0, 0) = -e. \tag{2.12}$$

The scaled projections of the components of the negative gradient direction of ψ into the primal and dual feasible spaces are, respectively,

$$p_\psi = -H_p \nabla_u \psi(0, 0) = [(XY)^{-\frac{1}{2}}(I - \hat{P})(XY)^{-\frac{1}{2}}]e, \tag{2.13}$$

$$q_\psi = -H_q \nabla_v \psi(0, 0) = [(XY)^{-\frac{1}{2}}\hat{P}(XY)^{-\frac{1}{2}}]e. \tag{2.14}$$

The direction (p_ψ, q_ψ) defined above is a descent direction for the barrier function $\psi(u, v)$ at the current point $(0, 0)$; thus it pulls the next iterate towards the interior of the primal and dual feasible sets. We will call (p_ψ, q_ψ) the centering direction.

In almost every primal-dual interior point algorithm, the step direction in the primal or dual space is a linear combination of the duality-gap-reducing direction and the centering direction. More specifically, for some $\sigma \in [0, 1)$,

$$p = p_\phi + \sigma \frac{x^T y}{n} p_\psi = -H_p (XYe - \sigma \frac{x^T y}{n} e), \tag{2.15}$$

$$q = q_\phi + \sigma \frac{x^T y}{n} q_\psi = -H_q (XYe - \sigma \frac{x^T y}{n} e), \tag{2.16}$$

In the sequel, we will use the notation:

$$\min(u) = \min_{1 \leq i \leq n} [u]_i, \quad \min(u, v) = \min_{1 \leq i \leq n} ([u]_i, [v]_i)$$

for $u, v \in \mathbf{R}^n$; the corresponding quantities for the maximums are similarly defined.

The following proposition can be easily verified using Proposition 2.1 and direct substitution.

Proposition 2.2 *If p and q are defined by (2.15) and (2.16), respectively, then*

1. $AXp = 0$ *and* $BYq = 0;$

2. $p^T XYq = 0;$

3. $p + q = -e + \sigma \frac{x^T y}{n} (XY)^{-1} e;$

4. $(e + \alpha p)^T XY(e + \alpha q) = x^T y [1 - \alpha(1 - \sigma)].$

We define the steplength α in (2.1) by the formula

$$\alpha = \frac{-\tau}{\min(p, q)}, \quad \tau \in (0, 1). \tag{2.17}$$

These choices of p, q and α guarantee that the new primal and dual variables x_+ and y_+ obtained from formulas (2.1) will remain strictly feasible.

We now state an algorithmic framework for interior point primal-dual algorithms.

Algorithm 1 *Given a strictly feasible pair* (x_0, y_0). *For* $k = 0, 1, 2, \ldots,$ *let*

$$x_{k+1} = X_k(e + \alpha_k p_k) \quad and \quad y_{k+1} = Y_k(e + \alpha_k q_k), \tag{2.18}$$

where p_k, q_k *and* α_k *are defined by (2.15), (2.16) and (2.17), respectively, and all the quantities involved (including* σ *and* τ) *are indexed by* k.

This algorithm generates strictly feasible sequences $\{x_k\}$ and $\{y_k\}$. It is a descent algorithm for the duality gap which is reduced at iteration k by a factor $1 - \alpha_k(1 - \sigma_k) < 1$. Almost all the existing primal-dual algorithms that use only one projection per iteration fit into the above algorithmic framework with different choices for the parameters σ_k and τ_k.

For example, in the primal-dual algorithm of Kojima, Mizuno and Yoshise [8], at each iteration a constant σ_k is chosen from $(0, 1)$ and, depending on this value of σ_k, restrictions are put on the parameter τ_k to ensure a polynomial complexity bound. In similar primal-dual algorithms implemented by Choi *et. al.* [2], McShane *et. al.* [12] and Lustig *et. al.* [11], very small values of σ_k were used and long steps were taken. Impressive numerical results have been obtained for these implementations though a polynomial complexity bound is no longer known.

9

Other examples include Todd and Ye's primal-dual potential reduction algorithm [20] and Monteiro and Adler's path-following primal-dual algorithms [16]. Todd and Ye's primal-dual potential function is

$$\Phi_\rho(x,y) = (n + \rho)\ln(\text{trace}[XY]) - \ln(\det[XY]).$$

This choice was motivated by the Karmarkar primal potential function [7]. At a given strictly feasible pair (x,y), if we define $\hat{\Phi}_\rho(u,v) = \Phi_\rho(X(e+u), Y(e+v))$, then we can see, though this was not the way the authors derived their algorithm, that the scaled projected negative gradient direction of $\hat{\Phi}_\rho(u,v)$ at $(0,0)$, gives the updating directions for (x,y) proposed by Todd and Ye and they are of the form of (2.15) and (2.16). Todd and Ye used $\rho = \nu\sqrt{n}$ in their algorithm where ν is a positive constant. This choice of ρ leads, at each iteration, to the choice

$$\sigma_k = \frac{\sqrt{n}}{\sqrt{n} + \nu}$$

in (2.15) and (2.16). In Monteiro and Adler's path-following primal-dual algorithms [16], one can show that

$$\sigma_k = 1 - \frac{\delta}{\sqrt{n}}$$

where δ is chosen to be a number in $(0, \sqrt{n})$ subject to a certain restriction. The restriction is such that δ is bounded above as $n \to \infty$ (Monteiro and Adler actually chose $\delta = 0.35$ in their analysis).

3 Superlinear Convergence

We first introduce two quantities defined at each iteration of Algorithm 1. At the k-th iteration, let

$$\theta_k = \frac{x_k^T y_k/n}{\max(X_k Y_k e)} \quad \text{and} \quad \eta_k = \frac{x_k^T y_k/n}{\min(X_k Y_k e)}.$$

Since $x_k^T y_k/n$ is the average value of the elements of $X_k Y_k e$, it is clear that $\theta_k \leq 1$ and $\eta_k \geq 1$. Moreover, it follows from (1.3) that the pair (x_k, y_k) is on the central path if and only if $\theta_k = 1$ or equivalently $\eta_k = 1$.

10

In this section, we present two distinct Q-superlinear convergence theories, namely, Theorem 3.1 and Theorem 3.3. Our first Q-superlinear convergence theory is quite general and makes no nondegeneracy assumption. Some relevant comments will follow its proof.

Theorem 3.1 Let $\{x_k\}$ and $\{y_k\}$ be generated by Algorithm 1, $x_k \to x_*$ and $y_k \to y_*$. Assume (i) strict complementarity, (ii) the sequence $\{\eta_k\}$ is bounded, and (iii) $\tau_k \to 1$ and $\sigma_k \to 0$. Then the duality gap sequence $\{x_k^T y_k\}$ converges to zero Q-superlinearly. That is, the Q_1-factor

$$Q_1 = \lim_{k \to \infty} \sup \frac{x_{k+1}^T y_{k+1}}{x_k^T y_k} = 0. \tag{3.1}$$

Proof: From 4 of Proposition 2.2, we have

$$Q_1 = 1 - \lim_{k \to \infty} \inf \alpha_k (1 - \sigma_k).$$

Since $\sigma_k \to 0$, $Q_1 = 0$ if and only if $\liminf_{k \to \infty} \alpha_k = 1$. We will prove that $\alpha_k \to 1$.

Multiply both sides of the equation in 3 of Proposition 2.2 by $(X_k Y_k)^{\frac{1}{2}}$ and consider the square of the ℓ_2-norm of both sides. From 2 of Proposition 2.2 we have

$$\|(X_k Y_k)^{\frac{1}{2}} p_k\|_2^2 + \|(X_k Y_k)^{\frac{1}{2}} q_k\|_2^2 = x_k^T y_k (1 - 2\sigma_k + \sigma_k^2 \frac{x_k^T y_k}{n} \frac{e^T (XY)^{-1} e}{n});$$

or equivalently,

$$\|T_k^{-\frac{1}{2}} p_k\|_2^2 + \|T_k^{-\frac{1}{2}} q_k\|_2^2 = n(1 - 2\sigma_k + \sigma_k^2 \frac{e^T T_k e}{n}), \tag{3.2}$$

where $T_k = (x_k^T y_k / n)(XY)^{-1}$. Assumption (ii) implies that $\{T_k\}$ is bounded above and $\{T_k^{-\frac{1}{2}}\}$ is bounded away from zero. Therefore, from (3.2) both $\{p_k\}$ and $\{q_k\}$ are bounded. It follows from (2.17) that $\{\alpha_k\}$ is bounded away from zero.

Now assume $[x_*]_i > 0$. Obviously,

$$1 = \lim_{k \to \infty} \frac{[x_{k+1}]_i}{[x_k]_i} = \lim_{k \to \infty} (1 + \alpha_k [p_k]_i).$$

This implies $[p_k]_i \to 0$, because $\{\alpha_k\}$ is bounded away from zero. Since $\sigma_k \to 0$, from 3 of Proposition 2.2 we have $(p_k + q_k) \to -e$. Hence, $[q_k]_i \to -1$. On the other hand, if $[x_*]_i = 0$, then $[y_*]_i > 0$ by strict complementarity. The same argument, interchanging the roles of p_k

11

and q_k, gives $[q_k]_i \to 0$ and $[p_k]_i \to -1$. Therefore, the components of p_k and q_k converge to either 0 or -1. Consequently, from (2.17) $\alpha_k \to 1$ since $\tau_k \to 1$. This completes the proof. $\square$

In Theorem 3.1, a source of concern has been the compatibility of Assumptions (ii) and (iii). On the surface, it seems as if letting $\tau_k \to 1$ and $\sigma_k \to 0$ might force $\eta_k \to \infty$. However, our numerical experience has shown this not to be the case. Indeed, Theorem 3.1 was the direct consequence of a rather extensive numerical experimentation. The superlinear convergence theory presented in the first draft of this paper consisted of only Theorem 3.3 and required the assumption that $\{\eta_k\}$ be bounded. This assumption has been removed in the present version. In subsequent numerical studies with highly degenerate Netlib problems, we let $\tau_k \to 1$ and $\sigma_k \to 0$ and always observed strict complementarity, $\{\eta_k\}$ bounded, $\alpha_k \to 1$ and Q-superlinear convergence. This phenomenon motivated us to search for a theory that could explain this occurrence and consequently led to the discovery of Theorem 3.1. We feel that Theorem 3.1 offers a satisfactory explanation of what we observed in practice. (In a more recent study, Zhang and Tapia [21] have proved that it is possible to choose $\sigma_k \to 0$ and $\tau_k \to 1$ while maintaining global convergence and the boundedness of $\{\eta_k\}$. Thus, the compatibility of the assumptions in Theorem 3.1 has been demonstrated.)

In numerical computation, the boundedness of $\{\eta_k\}$ requires some qualification because an algorithm is always stopped in a finite number of iterations. In our numerical experiments, we did not observe the trend of continued growth in the values of η_k as our algorithm was about to stop; while the observed convergence was clearly Q-superlinear and $\alpha_k \to 1$. Of course, the behavior of $\{\eta_k\}$ varies with several factors including how fast $\{\tau_k\}$ converges to one and $\{\sigma_k\}$ to zero. We do not imply that unbounded $\{\eta_k\}$ can never occur. Instead, we feel that it appears to be more an exception than the rule. This topic undoubtedly merits further study.

In the following development, we show that if we assume nondegeneracy, then we can obtain Q-superlinear convergence without assuming the boundedness of $\{\eta_k\}$. The following theorem concerns the Q_1 factor of the duality gap sequence.

Theorem 3.2 *Let $\{x_k\}$ and $\{y_k\}$ be generated by Algorithm 1, $x_k \to x_*$ and $y_k \to y_*$. Assume (i) strict complementarity and (ii) x_* is a nondegenerate vertex. Then the duality gap sequence $\{x_k^T y_k\}$ converges to zero and the Q_1-factor is*

$$Q_1 = \lim_{k \to \infty} \sup \frac{x_{k+1}^T y_{k+1}}{x_k^T y_k} = 1 - \lim_{k \to \infty} \inf \frac{\tau_k(1 - \sigma_k)}{1 - \theta_k \sigma_k}. \tag{3.3}$$

To prove Theorem 3.2, we need the following two lemmas. The first lemma has been proved in [19] under slightly different assumptions. For the sake of completeness, we include its proof here.

Lemma 3.1 *Let $\hat{P}_k$ be defined by (2.4) with X and Y indexed by k. Without loss of generality, assume that the first m elements of x_* are positive. Then under the assumptions of Theorem 3.2,*

$$\lim_{k \to \infty} \hat{P}_k = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.4}$$

Proof: Let

$$d_k = X_k^{\frac{1}{2}} Y_k^{-\frac{1}{2}} e \text{ and } D_k = diag(d_k).$$

Then

$$\hat{P}_k = D_k A^T (A D_k^2 A^T)^{-1} A D_k.$$

By our assumptions, we have

$$[y_k]_i \to 0, \ i = 1, 2, \ldots, m$$

and

$$[x_k]_i \to 0, \ i = m + 1, m + 2, \ldots, n.$$

It then follows from the definition of d_k that

$$[d_k]_i \to +\infty, \ i = 1, 2, \ldots, m$$

and

$$[d_k]_i \to 0, \ i = m + 1, m + 2, \ldots, n.$$

13

Now let A_1 be the m by m submatrix of A consisting of its first m columns and A_0 be the m by $n - m$ submatrix of A consisting of its last $n - m$ columns. Clearly, A_1 is nonsingular. Similarly, let D_k^∞ and D_k^0 be the diagonal matrices of dimensions m and $n - m$, respectively, with the first m and the last $n - m$ elements of d_k on their diagonals, respectively. Evidently, D_k^∞ is nonsingular for all k and $\{D_k^0\}$ converges to zero.

Substituting $AD_k = [A_1 D_k^\infty \quad A_0 D_k^0]$ for $AX_k^{\frac{1}{2}} Y_k^{-\frac{1}{2}}$ in (2.4), we obtain

$$\hat{P}_k = [A_1 D_k^\infty \quad A_0 D_k^0]^T [A_1 (D_k^\infty)^2 A_1^T + A_0 (D_k^0)^2 A_0^T]^{-1} [A_1 D_k^\infty \quad A_0 D_k^0].$$

Note that A_1 is nonsingular and let

$$R_k = (D_k^\infty)^{-1} A_1^{-1} A_0 D_k^0. \tag{3.5}$$

We have

$$\hat{P}_k = \begin{bmatrix} (I_m + R_k R_k^T)^{-1} & (I_m + R_k R_k^T)^{-1} R_k \\ R_k^T (I_m + R_k R_k^T)^{-1} & R_k^T (I_m + R_k R_k^T)^{-1} R_k. \end{bmatrix}. \tag{3.6}$$

Since $D_k^0 \to 0$ and $(D_k^\infty)^{-1} \to 0$, so does R_k. Now it is evident that

$$\lim_{k \to \infty} \hat{P}_k = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix},$$

which completes the proof. □

As mentioned above, our next lemma will be used not only in the proof of Theorem 3.2, but also in our quadratic convergence theory.

Lemma 3.2 *Let $\hat{P}_k$, p_k and q_k be defined by (2.4), (2.15) and (2.16), respectively, with X and Y indexed by k. Under the assumptions of Lemma 3.1,*

$$p_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ \vdots \\ -1 \end{pmatrix} + \sigma_k \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \frac{x_k^T y_k / n}{[X_k Y_k e]_{m+1}} \\ \vdots \\ \frac{x_k^T y_k / n}{[X_k Y_k e]_n} \end{pmatrix} + O(x_k^T y_k),$$

14

and

$$q_k = \begin{pmatrix} -1 \\ \vdots \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \sigma_k \begin{pmatrix} \frac{x_k^T y_k/n}{[X_k Y_k e]_1} \\ \vdots \\ \frac{x_k^T y_k/n}{[X_k Y_k e]_m} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + O(x_k^T y_k),$$

where the number of zeros is m in p_k, and $n - m$ in q_k.

Proof: Since $R_k \to 0$,

$$(I_m + R_k R_k^T)^{-1} = I_m + O(\|R_k\|^2).$$

Hence, from (3.6)

$$\hat{P}_k = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix} + E_k,$$

where

$$E_k = \begin{bmatrix} O(\|R_k\|^2) & R_k + O(\|R_k\|^3) \\ R_k^T + O(\|R_k\|^3) & O(\|R_k\|^2) \end{bmatrix}.$$

It follows from the definition of p_k and q_k (see (2.15) and (2.16)),

$$p_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ \vdots \\ -1 \end{pmatrix} + \sigma_k \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \frac{x_k^T y_k/n}{[X_k Y_k e]_{m+1}} \\ \vdots \\ \frac{x_k^T y_k/n}{[X_k Y_k e]_n} \end{pmatrix} - r_k^p,$$

where

$$r_k^p = (X_k Y_k)^{-\frac{1}{2}} E_k (X_k Y_k)^{-\frac{1}{2}} (X_k Y_k e + \sigma_k \frac{x_k^T y_k}{n} e).$$

By strict complementarity, $[X_k Y_k e]_i = O([x_*]_i)$ when $[x_*]_i = 0$ and $[X_k Y_k e]_i = O([y_k]_i)$ when $[x_*]_i > 0$. Also note that $x_k^T y_k = \|X_k Y_k e\|_1$. From the definition of D_k, we have

15

$D_k^0 = O(\sqrt{x_k^T y_k})$ and $(D_k^\infty)^{-1} = O(\sqrt{x_k^T y_k})$. Thus, $R_k = O(x_k^T y_k)$. It is easy to see that the diagonal blocks of $(X_k Y_k)^{-\frac{1}{2}} E_k (X_k Y_k)^{-\frac{1}{2}}$ are $O(x_k^T y_k)$. Therefore

$$r_k^p = (X_k Y_k)^{-\frac{1}{2}} \begin{bmatrix} 0 & R_k \\ R_k^T & 0 \end{bmatrix} ((X_k Y_k)^{\frac{1}{2}} e + \sigma_k \frac{x_k^T y_k}{n} (X_k Y_k)^{-\frac{1}{2}} e) + O(x_k^T y_k).$$

A straightforward matrix-vector multiplication shows that

$$[r_k^p]_i = \begin{cases} \frac{1}{[x_k]_i} \sum_{j=m+1}^{n} [A_1^{-1} A_0]_{i,j-m} ([x_k]_j - \sigma_k \frac{x_k^T y_k}{n} \frac{1}{[y_k]_j}) + O(x_k^T y_k), & 1 \le i \le m \\ \frac{1}{[y_k]_i} \sum_{j=1}^{m} [A_0^T A_1^{-T}]_{i-m,j} ([y_k]_j - \sigma_k \frac{x_k^T y_k}{n} \frac{1}{[x_k]_j}) + O(x_k^T y_k), & m < i \le n \end{cases}$$

Since

$$\lim_{k \to \infty} [x_k]_i = \begin{cases} [x_*]_i > 0, & 1 \le i \le m, \\ 0, & m < i \le n, \end{cases} \quad \text{and} \quad \lim_{k \to \infty} [y_k]_i = \begin{cases} 0, & 1 \le i \le m, \\ [y_*]_i > 0, & m < i \le n, \end{cases}$$

it is evident that $r_k^p = O(x_k^T y_k)$. This proves the first equality for p_k. Similarly, we can prove the second equality for q_k. $\square$

Now we are ready to prove Theorem 3.2.

Proof of Theorem 3.2: Without loss of generality, we assume that the first m components of x_* are positive and consequently the remaining $n - m$ components are zero.

It follows from Lemma 3.1, Lemma 3.2 and (2.17) that

$$\alpha_k = \frac{-\tau_k}{-1 + \sigma_k \frac{x_k^T y_k / n}{\max(X_k Y_k e)} + O(x_k^T y_k)} = \frac{\tau_k}{1 - \sigma_k \theta_k + O(x_k^T y_k)}. \tag{3.7}$$

From 4 of Proposition 2.2,

$$\frac{x_{k+1}^T y_{k+1}}{x_k^T y_k} = 1 - \alpha_k (1 - \sigma_k) = 1 - \frac{\tau_k (1 - \sigma_k)}{1 - \sigma_k \theta_k + O(x_k^T y_k)}.$$

Now (3.3) follows immediately. This completes the proof. $\square$

Observe that $\tau_k \in (0, 1)$, $\sigma_k \in [0, 1)$ and $\theta_k \in (0, 1]$. Therefore, for all k

$$\frac{\tau_k (1 - \sigma_k)}{1 - \theta_k \sigma_k} < 1.$$

Thus from (3.3), $Q_1 = 0$ if and only if

$$\lim_{k \to \infty} \tau_k \frac{1 - \sigma_k}{1 - \theta_k \sigma_k} = 1. \tag{3.8}$$

By examining (3.8), we have the following corollary. Its proof should be straightforward.

Corollary 3.1 *Under the assumptions of Theorem 3.2:*

1. *If* $\lim_{k \to \infty} \tau_k < 1$, *then* $Q_1 > 0$.

2. *If* $\lim_{k \to \infty} \tau_k = 1$, *then* $Q_1 = 0$ *if and only if*

$$\lim_{k \to \infty} \frac{1 - \sigma_k}{1 - \theta_k \sigma_k} = 1.$$

In particular, the above limit is 1 if $\sigma_k \to 0$ *or* $\theta_k \to 1$.

To emphasize the significance of Corollary 3.1, we formally state its interpretation as the following theorem. It is important to remember that $\{\sigma_k\}$ and $\{\tau_k\}$ are directly under our control, but $\{\theta_k\}$ is not.

Theorem 3.3 *Under the assumptions of Theorem 3.2, the duality gap sequence* $\{x_k^T y_k\}$ *generated by Algorithm 1 converges to zero Q-superlinearly if the sequence* $\{\tau_k\}$ *converges to 1 and either of the following two conditions holds:*

1. *The centering step is phased out asymptotically, i.e.,* $\lim_{k \to \infty} \sigma_k = 0$.

2. *The convergence of the primal-dual sequence* $\{(x_k, y_k)\}$ *to* $(x_\bullet, y_\bullet)$ *is along the central path, i.e.,* $\lim_{k \to \infty} \theta_k = 1$.

The convergence of $\{x_k^T y_k\}$ *is no better than Q-linear if* $\lim_{k \to \infty} \tau_k < 1$.

It is interesting to compare Theorem 3.1 and Theorem 3.3. The assumptions for the two theorems are different. In the proofs of the two theorems, we used different approaches and obtained distinct results.

Theorem 3.3 states that it is not necessary to have $\sigma_k \to 0$ in order to attain superlinear convergence. Admittedly, the case where the iterates converge asymptotically along the central path is a very special and perhaps unlikely case.

Observe that $\lim_{k \to \infty} \tau_k = 1$ means that our step asymptotically approaches the boundary of the positive orthant. Another interesting observation from (3.3) is that assuming $\lim_{k \to \infty} \tau_k = 1$,

$$\lim_{k \to \infty} \sup \frac{(1 - \theta_k)\sigma_k}{1 - \sigma_k \theta_k} = Q_1 \le \lim_{k \to \infty} \sup \sigma_k. \tag{3.9}$$

Therefore, even in the case of linear convergence, in general the smaller σ_k is, the faster the convergence will be. This may in part explain why good numerical performance was obtained from the implementations of primal-dual algorithms by Choi *et. al.* [2], McShane *et. al.* [12] and Lustig *et. al.* [11] where very small values of σ ($\sigma = 1/n$ or $1/\sqrt{n}$) were used.

If the Todd and Ye potential function method [20] is used to generate updating directions with the choice $\rho = n + \nu\sqrt{n}$, then as previously mentioned

$$\sigma = \frac{\sqrt{n}}{\sqrt{n} + \nu}.$$

Evidently, σ approaches 1 rapidly as n increases. Since the left-hand side of (3.9) tends to 1 as $\sigma \to 1$, unless $\theta_k \to 1$, the Q-linear convergence rate for this choice of σ will generally deteriorate towards 1 with the increase of n. Here we see clearly an inverse relationship between a good polynomial complexity bound (Todd and Ye proved that their algorithm converges in $O(\sqrt{n}L)$ iterations) and a good Q-convergence rate. Such a relationship also exists in Monteiro and Adler's $O(\sqrt{n}L)$-iteration path-following algorithms [16] where

$$\sigma = 1 - \frac{\delta}{\sqrt{n}}$$

and δ is bounded. Clearly, their path-following algorithms also show a deterioration of Q-convergence rate as the problem size increases. However, it is quite possible that the above mentioned two algorithms can still have reasonable R-behavior.

Now we prove a stronger convergence result for those primal and dual variables that converge to zero.

Theorem 3.4 *Let $\{x_k\}$ and $\{y_k\}$ be generated by Algorithm 1, $x_k \to x_*$ and $y_k \to y_{**}$. Assume (i) strict complementarity, (ii) $\tau_k \to 1$ and (iii) $\sigma_k \eta_k \to 0$ or $\theta_k \to 1$. If $x_k^T y_k \to 0$ Q-superlinearly, then the primal and dual variables that converge to zero do so Q-superlinearly.*

Proof: From (2.18), we have

$$X_k^{-1} x_{k+1} = e + \alpha_k p_k \quad \text{and} \quad Y_k^{-1} y_{k+1} = e + \alpha_k q_k.$$

Hence, by 3 of Proposition 2.2 we have

$$X_k^{-1} x_{k+1} + Y_k^{-1} y_{k+1} = (2 - \alpha_k)e + \alpha_k \sigma_k \frac{x_k^T y_k}{n}(X_k Y_k)^{-1} e. \tag{3.10}$$

18

If $\sigma_k \eta_k \to 0$, the second term in the right-hand side of (3.10) vanishes in the limit (notice that $\eta_k = \|(x_k^T y_k/n)(X_k Y_k)^{-1} e\|_\infty$). Also, $\alpha_k \to 1$. Therefore,

$$\lim_{k \to \infty} (X_k^{-1} x_{k+1} + Y_k^{-1} y_{k+1}) = e. \qquad (3.11)$$

On the other hand, if $\theta_k \to 1$, the second term in the right-hand side of (3.10) converges to $\alpha_k \sigma_k e$. Meanwhile, it follows from (3.7) that $\alpha_k(1 - \sigma_k) \to 1$. Hence, (3.11) also holds.

If $[x_*]_i = 0$, then by strict complementarity, $[y_*]_i > 0$ and $[y_{k+1}]_i/[y_k]_i \to 1$. It follows from (3.11) that $[x_{k+1}]_i/[x_k]_i \to 0$. Therefore, $[x_k]_i \to 0$ Q-superlinearly. By the symmetry of the relation (3.11), we have $[y_k]_j \to 0$ Q-superlinearly if $[y_*]_j = 0$. $\qquad \square$

Note that it is easy to enforce $\sigma_k \eta_k \to 0$ since we have direct control over σ_k and we can compute η_k before we set σ_k.

Since $x_k^T y_k = \|X_k Y_k e\|_1$, it is evident that when $\{x_k^T y_k\}$ converges to zero Q-superlinearly so does the sequence $\{X_k Y_k e\}$. We now demonstrate that this superlinear convergence is actually component-wise.

Corollary 3.2 *Under the assumptions of Theorem 3.4, if $x_k^T y_k \to 0$ Q-superlinearly, then the sequence $\{X_k Y_k e\}$ converges to zero Q-superlinearly component-wise.*

Proof: By strict complementarity, either $[x_k]_i \to 0$ or $[y_k]_i \to 0$ for each index i. From Theorem 3.4, we have either

$$\lim_{k \to \infty} \frac{[x_{k+1}]_i}{[x_k]_i} = 0 \quad \text{and} \quad \lim_{k \to \infty} \frac{[y_{k+1}]_i}{[y_k]_i} = 1$$

or

$$\lim_{k \to \infty} \frac{[x_{k+1}]_i}{[x_k]_i} = 1 \quad \text{and} \quad \lim_{k \to \infty} \frac{[y_{k+1}]_i}{[y_k]_i} = 0.$$

In either case,

$$\lim_{k \to \infty} \frac{[x_{k+1}]_i [y_{k+1}]_i}{[x_k]_i [y_k]_i} = \lim_{k \to \infty} \frac{[X_{k+1} Y_{k+1} e]_i}{[X_k Y_k e]_i} = 0.$$

This completes the proof of the component-wise Q-superlinear convergence of $\{X_k Y_k e\}$. $\qquad \square$

In 1980, Tapia [18] (see Theorem 3) pointed out that an algorithm which at each iteration satisfies the Taylor linearization of the complementarity equation has the property

that the variables that converge to zero do so Q-superlinearly. This result assumed strict complementarity and steplength one. Observe that (3.11) is equivalent to

$$X_k Y_k e + Y_k(x_{k+1} - x_k) + X_k(y_{k+1} - y_k) \to 0.$$

We see that the Taylor linearization of complementarity is satisfied asymptotically in our situation.

We close this section by commenting that taking different steplengths in the primal space and in the dual space may result in a larger reduction in the duality gap locally, i.e., at any given iteration; however, it seems unlikely that superlinear convergence could be achieved without both step lengths approaching one asymptotically.

4   Quadratic Convergence

In this section, we show that under the assumptions of Theorem 3.2 quadratic convergence can be achieved by primal-dual algorithms if we both phase out the centering direction and let the steps approach the boundary at a sufficiently fast rate. In contrast to the analysis of superlinear convergence, which is done in a scaled gradient-projection framework, the study of quadratic convergence will be in the framework of Newton's method.

We first reformulate Algorithm 1 as a perturbed and damped Newton's method.

It is well known that at optimality the primal, dual and dual slack variables x, λ and y satisfy

$$\begin{pmatrix} Ax - b \\ A^T\lambda + y - c \\ XYe \end{pmatrix} = 0, \tag{4.1}$$

$x \geq 0$ and $y \geq 0$. To eliminate the dual variables λ from the above system, we pre-multiply the second equation by the nonsingular matrix $[A^T \ B^T]^T$. Noticing that $BA^T = 0$, we obtain

$$0 = \begin{bmatrix} A \\ B \end{bmatrix} (A^T\lambda + y - c) = \begin{pmatrix} AA^T\lambda + A(y - c) \\ By - Bc \end{pmatrix}.$$

Since AA^T is nonsingular, λ is uniquely determined once y is known. Removing the equation for λ, we arrive at the following $2n$ by $2n$ system consisting of primal feasibility (see (1.1)), dual feasibility (see (1.2)) and complementarity:

$$F(x,y) = \begin{pmatrix} Ax - b \\ By - Bc \\ XYe \end{pmatrix} = 0, \tag{4.2}$$

as well as the non-negativity constraints for (x, y).

Similarly, we can show that a strictly feasible pair (x, y) on the central path satisfies

$$\hat{F}(x,y,\mu) = \begin{pmatrix} Ax - b \\ By - Bc \\ XYe - \mu e \end{pmatrix} = 0. \tag{4.3}$$

for some $\mu > 0$. Evidently, $\hat{F}(x, y, \mu) = 0$ is a perturbation of the system $F(x, y) = 0$ with the perturbation term $-\mu e$ added to the nonlinear portion of $F(x, y)$. It is also obvious that $\hat{F}(x, y, 0) = F(x, y)$.

The following proposition relates the search direction (p, q) in Algorithm 1 to a perturbed Newton's direction $(\Delta x, \Delta y)$.

Proposition 4.1 *Let (x, y) be a strictly feasible pair and let p and q be defined by (2.15) and (2.16). Then p and q satisfy*

$$\begin{pmatrix} Xp \\ Yq \end{pmatrix} = \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \equiv -[F'(x,y)]^{-1}\hat{F}(x,y,\mu) \tag{4.4}$$

for the choice $\mu = \sigma \frac{x^T y}{n}$.

Proof: Notice that

$$\hat{F}'_{(x,y)}(x,y,\mu) = F'(x,y) = \begin{bmatrix} A & 0 \\ 0 & B \\ Y & X \end{bmatrix},$$

consequently, we have

$$
\begin{bmatrix} A & 0 \\ 0 & B \\ Y & X \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} 0 \\ 0 \\ w \end{pmatrix},
$$

where $w = XYe - \sigma \frac{x^T y}{n} e$. Thus $A \Delta x = 0$, so $\Delta x = B^T u$, where $u \in \mathbf{R}^{n-m}$. Similarly, $\Delta y = A^T v$, where $v \in \mathbf{R}^m$. Substituting Δx and Δy into the third equation block of the system and multiplying both sides by AY^{-1}, we obtain

$$
AXY^{-1}A^T v = -AY^{-1}w.
$$

Thus,

$$
\Delta y = A^T v = -A^T (AXY^{-1}A^T)^{-1} AY^{-1}w.
$$

It is now straightforward to verify from (2.16) that

$$
Y^{-1}\Delta y = -H_q w = q.
$$

Consequently, by 4 of Proposition 2.2,

$$
X^{-1}\Delta x = -(XY)^{-1}w - Y^{-1}\Delta y = -H_p w = p.
$$

This completes the proof. □

We can therefore view a primal-dual algorithm as a perturbed and damped Newton's method. At the k-th iteration the iterate is obtained from the perturbed system $\hat{F}(x, y, \mu_k) = 0$. The sequence of the perturbation parameters $\{\mu_k\}$ converges to zero as $x_k^T y_k \to 0$. We use the qualifier damped because at each iteration the steplength is determined by formula (2.17) in order to keep the iterates in the interior of the feasibility set. The positivity requirements for x and y in general prevent a full Newton step from being taken. It is well known that taking full steps asymptotically is a critical ingredient for the Q-quadratic convergence of Newton's method (see Dennis and Moré [3, Corollary 2.3]).

We now rewrite Algorithm 1 in the following equivalent form of a perturbed and damped Newton's method.

Algorithm 2 *Given a strictly feasible pair* (x_0, y_0). *For* $k = 0, 1, 2, \ldots,$ *let*

$$x_{k+1} = x_k + \alpha_k \Delta x_k \quad and \quad y_{k+1} = y_k + \alpha_k \Delta y_k, \tag{4.5}$$

where Δx_k *and* Δy_k *are defined by (4.4), and* α_k *by (2.17), and all the quantities involved (including* σ *and* τ*) are indexed by* k.

To establish Q-quadratic convergence for Algorithm 2, we need to address the following three issues:

1. Is the Jacobian matrix $F'(x, y)$ nonsingular at optimality?

2. How fast must the centering direction – the perturbation controlled by μ – be phased out?

3. Can full Newton steps be taken asymptotically and at a rate that ensures quadratic convergence?

The following lemma answers the first question.

Lemma 4.1 *Let* (x_*, y_*) *be an optimal pair for the linear programs (1.1) and (1.2). Under the assumptions of Theorem 3.2, the* $2n \times 2n$ *matrix*

$$F'(x_*, y_*) = \begin{bmatrix} A & 0 \\ 0 & B \\ Y_* & X_* \end{bmatrix}$$

is nonsingular.

Proof: It can be shown that Assumptions (i) and (ii) of Theorem 3.2 imply that y_* is a nondegenerate vertex of the dual (1.2). Without loss of generality, we assume that the first m components of x_* are positive and consequently the remaining $n - m$ components are zero. Let

$$x_*^+ = \begin{pmatrix} [x_*]_1 \\ \vdots \\ [x_*]_m \end{pmatrix} \quad and \quad y_*^+ = \begin{pmatrix} [y_*]_{m+1} \\ \vdots \\ [y_*]_n \end{pmatrix}.$$

By our assumptions, $x_*^+ > 0$ and $y_*^+ > 0$.

Let A_1 be the m by m submatrix of A consisting of its first m columns and A_0 be the m by $n - m$ submatrix of A consisting of its last $n - m$ columns. Clearly, A_1 is nonsingular because its columns form the optimal basis for the primal linear program (1.1). The same ordering also leads to $B = [B_0\ B_1]$ where $B_1 \in \mathbf{R}^{(n-m) \times (n-m)}$ is nonsingular and its columns form the optimal basis for the dual linear program (1.2). Using the above introduced notation, we have

$$F'(x_*, y_*) = \begin{bmatrix} A_1 & A_0 & 0 & 0 \\ 0 & 0 & B_0 & B_1 \\ 0 & 0 & X_*^+ & 0 \\ 0 & Y_*^+ & 0 & 0 \end{bmatrix}. \tag{4.6}$$

By examining blocks of this matrix, we can easily see that it is indeed nonsingular. □

Further examination of (4.6) reveals that Lemma 4.1 is sharp in the sense that $F'(x_*, y_*)$ will be singular if the number of nonzeros in x_* (y_*) is not m $(n - m)$.

An answer to the second question is not hard to find. From standard analysis for Newton-like methods, an $O(\|F(x, y)\|^2)$ perturbation term does not destroy quadratic convergence. In our context, this is equivalent to the requirement $\mu_k = O((x_k^T y_k)^2)$ since for any feasible pair (x, y)

$$x^T y = \|XYe\|_1 = \|F(x, y)\|_1.$$

An answer to the third question requires further analysis. For the ease of notation, let us denote the pair (x, y) by $z \in \mathbf{R}^{2n}$. In a damped Newton method for $F(z) = 0$, if Δz_k is the full Newton step at the k-th iteration and α_k is the steplength, then

$$z_{k+1} = z_k + \alpha_k \Delta z_k = z_k + \Delta z_k - (1 - \alpha_k)\Delta z_k.$$

From standard analysis,

$$\|z_{k+1} - z_*\| \le O(\|z_k - z_*\|^2) + |1 - \alpha_k| \|\Delta z_k\|.$$

Since

$$\|\Delta z_k\| = O(\|F(z_k)\|) = O(\|z_k - z_*\|),$$

24

it is clear that if

$$|1 - \alpha_k| = O(\|z_k - z_*\|),$$

then quadratic convergence will be achieved.

From (3.7), we see that the steplength α_k depends on τ_k, σ_k and an $O(x_k^T y_k)$ term. Since $x_k^T y_k = O(\|z_k - z_*\|)$, in order to ensure $|1 - \alpha_k| = O(\|z_k - z_*\|)$, we see that it is sufficient to have $1 - \tau_k$ and σ_k be $O(\|z_k - z_*\|)$. If we take $\sigma_k = O(x_k^T y_k)$, then we have

$$\sigma_k = O(\|z_k - z_*\|) \quad \text{and} \quad \mu_k = O(\|z_k - z_*\|^2).$$

Moreover, we can easily enforce the requirement $1 - \tau_k = O(x_k^T y_k) = O(\|z_k - z_*\|)$.

Now we are in a position to prove the following quadratic convergence theorem. Its proof is basically a rigorous and detailed treatment of the above discussion. As a by-product, we also obtain a local convergence result.

Theorem 4.1 *Let $\{(x_k, y_k)\}$ be generated by Algorithm 2. Assume (i) strict complementarity, (ii) x_* is a nondegenerate vertex, and (iii) the choices of σ_k and τ_k satisfy at each iteration*

$$0 \le \sigma_k \le \min(\sigma, c_1 x_k^T y_k) \quad \text{and} \quad \max(\tau, 1 - c_2 x_k^T y_k) \le \tau_k < 1 \qquad (4.7)$$

where $\sigma \in [0, 1)$, $\tau \in (0, 1)$ and $c_1, c_2 > 0$. Then

1. *whenever $\{(x_k, y_k)\}$ converges to (x_*, y_*), it does so Q-quadratically, i.e., there exist a constant $\gamma > 0$ such that for k sufficiently large,*

$$\|(x_{k+1}, y_{k+1}) - (x_*, y_*)\| \le \gamma \|(x_k, y_k) - (x_*, y_*)\|^2; \qquad (4.8)$$

2. *there exists a number $\delta > 0$ such that whenever $\|(x_0, y_0) - (x_*, y_*)\| \le \delta$, then $\{(x_k, y_k)\}$ converges to (x_*, y_*).*

Proof: Again we use the notation $z = (x, y)$. Also let $\hat{e} = (0 \ldots 0 \ 1 \ldots 1)^T \in \mathbf{R}^{2n}$ where the numbers of zeros and ones are both n. As mentioned in Section 2 after (2.17), the sequence $\{z_k\}$ is always well-defined and remains strictly feasible.

Following the standard analysis for Newton-like methods (see Dennis and Schnabel [4]. for example), we have

$$z_{k+1} - z_* = z_k - z_* - \alpha_k [F'(z_k)]^{-1} \hat{F}(z_k, \mu_k)$$

$$= [F'(z_k)]^{-1} \{ [F(z_*) - F(z_k) - F'(z_k)(z_* - z_k)] + (1 - \alpha_k) F(z_k) + \alpha_k \mu_k \hat{e} \}.$$

Therefore,

$$\|z_{k+1} - z_*\| \leq \|[F'(z_k)]^{-1}\| (\|F(z_*) - F(z_k) - F'(z_k)(z_* - z_k)\|$$

$$+ |1 - \alpha_k| \|F(z_k)\| + \alpha_k \mu_k \|e\|). \tag{4.9}$$

Note $z_k = (x_k, y_k)$ is strictly feasible and $x_k^T y_k = \|F(z_k)\|_1 = \|F(z_k) - F(z_*)\|_1$. There exists $\delta_1 > 0$ and $c_3 > 0$ such that if $\|z_k - z_*\| \leq \delta_1$, then

$$x_k^T y_k \leq c_3 \|z_k - z_*\| \quad \text{and} \quad \|F(z_k)\| \leq c_3 \|z_k - z_*\|. \tag{4.10}$$

This follows from the fact that $F(z)$ is continuously differentiable. Also note that $F(z)$ is a quadratic, hence there exists $c_4 > 0$ such that for any k

$$\|F(z_*) - F(z_k) - F'(z_k)(z_* - z_k)\| \leq c_4 \|z_k - z_*\|^2. \tag{4.11}$$

In view of the continuity and nonsingularity of $F'(z)$ at z_*, there exist $\delta_2 > 0$ and $c_5 > 0$ such that if $\|z - z_*\| \leq \delta_2$, then

$$\|[F'(z)]^{-1}\| \leq c_5. \tag{4.12}$$

In addition, there exist $\delta_3 > 0$ and $c_6 > 0$ such that if z_k satisfies $\|z_k - z_*\| \leq \delta_3$, then from (3.7)

$$|1 - \alpha_k| = \left| \frac{(1 - \tau_k) - \theta_k \sigma_k + O(x_k^T y_k)}{1 - \theta_k \sigma_k + O(x_k^T y_k)} \right|$$

$$\leq 2|(1 - \tau_k) - \theta_k \sigma_k + c_6 x_k^T y_k|$$

$$\leq 2(c_2 + \theta_k c_1 + c_6) x_k^T y_k.$$

Here we assumed that δ_3 is sufficiently small so that $1 - \theta_k \sigma_k - c_6 x_k^T y_k \geq 1/2$ and (4.10) holds. We also used the assumptions $\sigma_k \leq c_1 x_k^T y_k$ and $1 - \tau_k \leq c_2 x_k^T y_k$. Using (4.10) and noting $\theta_k \leq 1$, we have

$$|1 - \alpha_k| \leq c_7 \|z_k - z_*\|, \tag{4.13}$$

26

where $c_7 = 2c_3(c_2 + c_1 + c_6)$.

It follows from (4.7), (4.9), (4.10), (4.11), (4.12) and (4.13) that if z_k satisfies

$$\|z_k - z_*\| \leq \min(\delta_1, \delta_2, \delta_3),$$

then (4.8) holds with

$$\gamma = c_5(c_4 + c_7 c_3^2 + c_1 c_3^2 \|e\|/n).$$

Inequality (4.8) implies that if $\{z_k\}$ converges to z_*, then it does so Q-quadratically. This proves the first statement.

Now we only need to prove the second statement – the convergence of $\{z_k\}$. Let

$$\delta = \min(\delta_1, \delta_2, \delta_3, r/\gamma)$$

for some $r \in (0, 1)$. If $\|z_0 - z_*\| \leq \delta$, then

$$\|z_1 - z_*\| \leq \gamma \|z_0 - z_*\|^2 \leq r \|z_0 - z_*\|.$$

So $\|z_1 - z_*\| \leq r\delta \leq \delta$. Now we proceed by induction. This establishes the convergence of $\{z_k\}$ to z_*.   □

In our numerical experimentation, we found that even for highly degenerate problems the observed convergence was effectively Q-quadratic until the iterates got too close to a solution and the singularity of the Jacobian matrix was encountered. This curious but pleasing phenomenon is the subject of further investigation.

5   Concluding Remarks

The rich structure present in the primal-dual formulation has led us to establish some rather strong convergence rate results.

No superlinear convergence results have been established so far for either primal or dual interior point algorithms. In fact, Gonzaga and Todd [5] showed that an algorithm that takes either primal or dual steps and reduces the Todd-Ye primal-dual potential function can not

have an R convergence rate greater than one (independent of n). Thus from the viewpoint of convergence rate, our results suggest that primal-dual algorithms should be preferred to either primal or dual algorithms. Combined with the favorable numerical results obtained by a number of authors (Choi *et. al.* [2], McShane *et. al.* [12] and Lustig *et. al.* [11]), this preference for primal-dual algorithms seems to be well founded.

We have shown that for the class of primal-dual algorithms studied, approximate centering should be viewed as a globalization strategy for Newton's method. Like other globalization strategies, it may improve the global behavior of the algorithm; but if not properly implemented, it will destroy fast local convergence. This fact lends credibility to the belief that polynomiality alone does not guarantee that local convergence rate properties have not been compromised or that the algorithm necessarily is fast. The algorithms of Kojima, Mizuno and Yoshise [8]), Monteiro and Adler [16] and Todd and Ye [20] possess polynomiality but can not have fast Q-convergence.

Our preliminary numerical experimentation has shown that even without centering the damped Newton algorithms that take steps close to the boundary of the positive orthant still have reasonable global behavior; although centering usually helps. This should not be totally unexpected since we are applying the damped Newton's method to a mildly nonlinear problem (see (4.2)).

One of the key components of this research is equation (3.7) which shows that in the damped Newton's method one can asymptotically make the steplength approach one at a rate that guarantees the fast convergence of Newton's method.

It seems to be difficult and costly, if at all possible, to ensure that the sequence $\{(x_k, y_k)\}$ converges to (x_*, y_*) along the central path. Therefore, it is our belief that at this stage the only viable strategy for designing a Q-superlinearly or Q-quadratically convergent primal-dual interior point algorithm is to phase out the centering step at the specified speed. The effect of degeneracy on the quadratic rate of convergence and the development of a quadratically convergent practical algorithm are the subjects of current research.

Acknowledgment

References

[1] D. Bayer and J.C. Lagarias. The non-linear geometry of linear programming, I. Affine and projective scaling, II. Legendre transformation coordinates, III. Central trajectories., 1986. Preprints, AT & T Bell Lab, Murray Hill, NJ.

[2] I. C. Choi, C.L. Monma, and D.F. Shanno. Further development of a primal-dual interior point method for linear programming. Technical Report RRR# 60-88, RETCOR, Rutgers University, 1988.

[3] J. E. Dennis Jr. and J. J. Moré. A characterization of superlinear convergence and its application to quasi-Newton methods. *Math. Comp*, 28:549–560, 1974.

[4] J. E. Dennis Jr. and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall Inc., Englewood cliffs, New Jersey, 1983.

[5] C. C. Gonzaga and M.J. Todd. An $O(\sqrt{n}L)$-iteration large-step primal-dual affine algorithm for linear programming. Technical Report 862, School of Operations Research and Industrial Engineering, Cornell University, 1989.

[6] Siming Huang and K.O. Kortanek. A simultaneous primal-dual potential reduction algorithm for linear programming. Technical Report 89-2, Dept. Management Sciences, The University of Iowa, 1989.

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

[8] M. Kojima, S. Mizuno, and A. Yoshise. A primal-dual interior point method for linear programming. In Nimrod Megiddo, editor, *Progress in Mathematical programming, interior-point and related methods*, pages 29–47. Springer-Verlag, New York. 1989.

[9] I.J. Lustig. Feasibility issues in an interior point method for linear programming. Technical Report SOR 88-9, Dept. Civil Eng. and O.R., Princeton University, 1988.

[10] I.J. Lustig. A generic primal-dual interior point algorithm. Technical Report SOR 88-3. Dept. Civil Eng. and O.R., Princeton University, 1988.

[11] I.J. Lustig, R.E. Marsten, and D.F. Shanno. Computational experience with a primal-dual interior point method for linear programming. Technical Report SOR 89-17, Dept. Civil Eng. and O.R., Princeton University, 1989.

[12] K.A. McShane, C.L. Monma, and D.F. Shanno. An implementation of a primal-dual interior point method for linear programming. *ORSA J. Computing*, 1:70–83, 1989.

[13] N. Megiddo. Pathways to the optimal set in linear programming. In Nimrod Megiddo, editor, *Progress in Mathematical programming, interior-point and related methods*, pages 131–158. Springer-Verlag, New York, 1989.

[14] S. Mizuno, M.J. Todd, and Y. Ye. Anticipated behavior of long-step algorithms for linear programming. Technical Report Working paper Series No. 89-33, College of Business Administration, The University of Iowa, 1989.

[15] S. Mizuno, M.J. Todd, and Y. Ye. Anticipated behavior of path-following algorithms for linear programming. Technical Report 878, School of Operations Research and Industrial Engineering, Cornell University, 1989.

[16] R.C. Monteiro and I. Adler. Interior path-following primal-dual algorithms. Part I: linear programming. *Math. Prog.*, 44:27–41, 1989.

[17] G. Sonnevend. An analytic center for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In A. Prekopa, editor. *Lecture Notes in Control and Inf. Sci.*, pages 866–876. Springer, Berlin, 1985, 1984.

[18] R. A. Tapia. On the role of slack variables in quasi-Nweton methods for constrained optimization. In L. C. W. Dixon and G. P. Szegö, editors, *Numerical Optimization of dynamic systems*, pages 235–246. North-Holland, 1980.

[19] R. A. Tapia and Y. Zhang. An optimal basis identification technique for interior point linear programming methods. To appear in the special issue of *Linear Algebra and its Applications* on interior point methods for linear programming, 1991.

[20] M.J. Todd and Y. Ye. A centered projective algorithm for linear programming. *Math. of O.R.*, 15:508–529,1990.

[21] Y. Zhang and R. A. Tapia. A quadratically convergent polynomial primal-dual interior-point algorithm for linear programming. Technical Report No. 90-40, Dept. of Mathematical Sciences, Rice University, 1990.