The Behavior of Newton-type Methods on Two Equivalent Systems from Linear Programming

Maria Cristina Villalobos, Richard A. Tapia, and Yin Zhang

CRPC-TR98770-S
September 1998

Center for Research on Parallel Computation
Rice University
6100 South Main Street
CRPC - MS 41
Houston, TX 77005

Submitted September 1998; revised August 1999; available as Rice CAAM TR98-02

The Behavior of Newton-type Methods on Two Equivalent Systems from Linear Programming

Maria Cristina Villalobos* Richard A. Tapia† Yin Zhang‡

Abstract

Newton-type methods are fundamental techniques for solving systems of nonlinear equations. However, it is often not fully appreciated that these methods can produce significantly different behavior when applied to equivalent systems. In this paper, we investigate differences in local and global behavior of Newton-type methods when applied to two different but equivalent systems from linear programming: the optimality conditions of the logarithmic barrier formulation and the perturbed optimality conditions. Through theoretical analysis and numerical results, we show Newton-type methods perform more effectively on the latter system.

Key words. Newton's method; equivalent systems; interior-point methods; sphere of convergence; linear programming

1 Introduction

Newton's method is generally accepted as an effective tool for solving a system of nonlinear equations, $F(v) = 0$, where $F : R^n \to R^n$. It is a locally and quadratically convergent method under reasonable assumptions (see e.g. Dennis and Schnabel [1]). In many practical applications, globally convergent methods are required to solve nonlinear systems. As a result, Newton's method is modified and the resulting variant is termed a Newton-type method. It is often not fully appreciated that Newton-type methods can exhibit significantly different local and global behavior on two equivalent systems. In this paper, we compare the behavior of Newton-type methods applied to two well-known equivalent systems of nonlinear equations associated with linear programming.

The first of these equivalent systems consists of the first-order optimality conditions of the log-barrier formulation of the linear program. The second system consists of the perturbed first-order optimality conditions for the linear program. Though the two nonlinear systems have the same set of solutions, El-Bakry, Tapia, Tsuchiya, and Zhang [2]

*Dept of Computational and Applied Mathematics, Rice University, 6100 Main St-MS 134, Houston, TX, 77005-1892; cristina@caam.rice.edu. This author was supported in part by the Center for Research and Parallel Computations, DOE DE-FG03-93ER25178, and the Sloan Foundation,

†Dept of Computational and Applied Mathematics, Rice University, 6100 Main St-MS 134, Houston, TX, 77005-1892; rat@caam.rice.edu. This author was supported in part by DOE DE-FG05-86ER25017.

‡Department of Computational and Applied Mathematics, Rice University, Houston, TX, 77005-1892; zhang@caam.rice.edu. This author was supported in part by DOE DE-FG03-97ER25331.

show that Newton's method necessarily generates different iterates for the two systems. In this paper, we show that Newton's method applied to the perturbed optimality conditions for the linear program has a larger sphere of convergence than Newton's method applied to the optimality conditions of the log-barrier formulation of the linear program. In addition, we apply an interior-point path-following method to solve the two equivalent systems, and discuss several distinguishing properties in the behavior of the method on the two systems.

Of these two equivalent systems, the perturbed first-order optimality conditions are widely used to solve the linear program. However, the reasons for favoring this system have not been fully analyzed. In this paper, we explain why the system associated with the perturbed optimality conditions is the system of choice.

The paper is organized as follows. In section 2, we present the two equivalent nonlinear systems under consideration. In section 3, we introduce the notion of the sphere of convergence of Newton's method. We provide results on the radius of the sphere of convergence of Newton's method applied to the two equivalent systems. Section 4 presents numerical results supporting the theory we developed in the previous section. In section 5, we discuss the implementation of our interior-point path-following algorithm applied to the two equivalent systems. Section 6 describes the numerical behavior of our interior-point path-following method on the two equivalent systems. Finally, in section 7 we make some concluding remarks.

2 Two Equivalent Formulations

In this section, we introduce the linear programming problem and the two equivalent nonlinear systems under consideration. We consider the linear programming problem in the standard form

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

where $c, x \in R^n$, $b \in R^m$, $A \in R^{m \times n}$ with $m \leq n$, and $\text{rank}(A) = m$. The Lagrangian function associated with problem (1) is

$$L(x, y, z) = c^T x + y^T (Ax - b) - x^T z,$$

where $y \in R^m$ and $z \in R^n_+$ are, respectively, the vectors of Lagrange multipliers associated with the equality and the inequality constraints. The first-order optimality conditions for problem (1) are

$$F(x, y, z) \equiv \begin{bmatrix} A^T y + z - c \\ Ax - b \\ XZe \end{bmatrix} = 0 \quad (x, z \geq 0) \tag{2}$$

where $Z = diag(z), X = diag(x)$, and $e = (1, ..., 1)^T \in R^n$.

We derive one of the equivalent systems by formulating problem (1) in the logarithmic barrier framework. This framework, which was introduced by Frisch [4], consists of solving a sequence of equality constrained minimization problems with decreasing values of the barrier parameter $\mu > 0$. For problem (1) and a given value of $\mu > 0$, the log-barrier subproblem has the following form

$$\begin{array}{ll} \text{minimize} & c^T x - \mu \sum_{i=1}^{n} \log x_i \\ \text{subject to} & Ax = b \\ & (x, \mu > 0). \end{array}$$

Let x^*_μ denote the solution of the log-barrier problem for a given value of $\mu > 0$. Then under mild assumptions (se e.g. Fiacco and McCormick [3]), as $\mu \to 0$ the sequence of iterates $\{x^*_\mu\}$ converges to a solution x^* of problem (1), i.e. $lim_{\mu \to 0}\ x^*_\mu = x^*$.

The optimality conditions for the log-barrier subproblem are derived by differentiating the Lagrangian function,

$$\mathcal{L}(x, y; \mu) = c^T x - \mu \sum_{i=1}^{n} \log x_i - y^T (Ax - b)$$

where $y \in R^m$ is the vector of Lagrange multipliers associated with the equality constraints. Then the optimality conditions are

$$F_B(x, y; \mu) \equiv \begin{bmatrix} A^T y + \mu X^{-1} e - c \\ Ax - b \end{bmatrix} = 0 \qquad (\mu, x > 0). \tag{3}$$

Observe that the Jacobian of F_B is given by

$$F'_B(x, y; \mu) = \begin{bmatrix} -\mu X^{-2} & A^T \\ A & 0 \end{bmatrix}.$$

If $\text{rank}(A) = m$ and $x > 0$, then $F'_B(x, y; \mu)$ is nonsingular for $\mu > 0$. If any primal variables, x, are zero at the solution, then near the solution the Jacobian necessarily becomes ill-conditioned for μ close to zero.

Now we derive a nonlinear system equivalent to system (3). Consider the introduction of an auxiliary variable, $z \in R^n$, and define $z = \mu X^{-1} e$ which is written equivalently as $XZe = \mu e$. Substituting z into system (3) and adding our equivalent defining relations yields the system

$$F_P(x, y, z; \mu) \equiv \begin{bmatrix} A^T y + z - c \\ Ax - b \\ XZe - \mu e \end{bmatrix} = 0 \quad (\mu, x, z \geq 0). \tag{4}$$

The Jacobian of F_P is given by

$$F'_P(x, y, z; \mu) = \begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z & 0 & X \end{bmatrix}.$$

The Jacobian is nonsingular if $\text{rank}(A) = m$ and $x, z > 0$. System (4) can also be obtained by considering the first-order optimality conditions (2) of the linear program and perturbing the complementarity equation, $XZe = 0$, by μe.

Central Path

Let (x^*_μ, y^*_μ) denote the solution to system (3) for a particular value of $\mu > 0$, and similarly let $(x^*_\mu, y^*_\mu, z^*_\mu)$ denote the solution to system (4). Then by the central path for system (3), we mean

$$\mathcal{C}_B = \{(x^*_\mu, y^*_\mu) : F_B(x^*_\mu, y^*_\mu; \mu) = 0, \quad \mu,\ x^*_\mu > 0\}. \tag{5}$$

McLinden [6] shows that $\mathcal{C}_B$ is a continuous path such that $\lim_{\mu\to 0}(x^*_\mu, y^*_\mu) = (x^*, y^*)$. We remark that systems (3) and (4) are equivalent, in the sense that for $\mu > 0,$ and $(x^*_\mu, y^*_\mu) \in \mathcal{C}_B$,

$$F_B(x^*_\mu, y^*_\mu; \mu) = 0 \iff F_P(x^*_\mu, y^*_\mu, z^*_\mu; \mu) = 0$$

for $z^*_\mu = \mu(X^*_\mu)^{-1}e > 0$. For system (4), we have $\lim_{\mu\to 0}(x^*_\mu, y^*_\mu, z^*_\mu) = (x^*, y^*, z^*)$. Thus, for system (4) the central path is defined as

$$\mathcal{C}_P = \{(x^*_\mu, y^*_\mu, z^*_\mu) : F_P(x^*_\mu, y^*_\mu, z^*_\mu; \mu) = 0, \quad \mu,\ x^*_\mu, z^*_\mu > 0\}. \tag{6}$$

El-Bakry, Tapia, Tsuchiya, and Zhang [2] show that although systems (3) and (4) are equivalent, Newton's method necessarily generates different iterates for the two systems.

Preliminaries

Throughout the paper, we make use of the following assumption and notation.

Nondegeneracy Assumption. The matrix A has full rank m. Let (x^*, y^*, z^*) be a primal and dual nondegenerate solution of system (2). Without loss of generality, we assume that the first m components of x^* are positive, and the remaining $(n - m)$ components are zero.

Let $\mathcal{B} = \{i : (x^*)_i > 0\}$ and $\mathcal{N} = \{i : (x^*)_i = 0\}$. Then by the nondegeneracy assumption, $\mathcal{B} = \{1, \ldots, m\}$ and $\mathcal{N} = \{m + 1, \ldots, n\}$. The matrix A will be partitioned into $A = [A_\mathcal{B}\ A_\mathcal{N}]$ where $A_\mathcal{B}$ denotes the matrix consisting of the columns of A indexed by $\mathcal{B}$ and similarly for $A_\mathcal{N}$. Note that $\text{rank}(A_\mathcal{B}) = m$. If u is a vector, then its uppercase counterpart U will denote the diagonal matrix whose diagonal consists of the elements of u. For a vector $u \in R^n$, $u_\mathcal{B}$ is the vector of the first m components of u and $u_\mathcal{N}$ is the vector of the remaining $(n - m)$ components of u. The quantity u^2 represents the vector u whose components are individually squared. All norms $\|\cdot\|$ are assumed to be the Euclidean norm unless otherwise noted.

3 Sphere of Convergence: Analysis

Standard local theory of Newton's method applied to a nonlinear system (see e.g. [1]) provides the existence of a neighborhood about a solution that contains points which make Newton's method well-defined. Most importantly, starting from any point in the neighborhood, Newton's method guarantees convergence to the solution. For systems (3) and (4), a neighborhood also exists about the solution for any given $\mu > 0$. In this section, we introduce the notion of the sphere of convergence for Newton's method. We analyze the behavior of the radius of the sphere of convergence associated with systems (3) and (4) by considering Newton's method applied to these equivalent systems as $\mu \to 0$. Under the nondegeneracy assumption, our analysis shows that the radius of the sphere of convergence of Newton's method on system (3) decreases to zero in the same order as $\mu \to 0$. However, we show the radius of the sphere of convergence of Newton's method applied to system (4) has a lower-bound estimate independent of μ. We believe these results show Newton's method to be more efficient on system (4) than on system (3), at least for small values of $\mu > 0$.

3.1 Preliminaries

We introduce the notion of the sphere of convergence for Newton's method. Then, we present lemmas to be used in our analysis for the radius of the sphere of convergence of Newton's method on systems (3) and (4).

We remark that the notion of the sphere of convergence is not new. Several references can be found in the literature where this notion or similar concept is used, see [1, pg 91], for example. To conduct a rigorous study on the radius of convergence for Newton's method, we give a formal definition for the sphere of convergence below.

Definition We define the closed ball with radius r centered at v^* as $B(v^*; r) = \{v : \|v - v^*\| \leq r\}$.

Definition For a given nonlinear system, $F(v) = 0$, and a solution v^*, the *sphere of convergence* of Newton's method at v^* is defined as the largest closed ball centered at v^* such that starting from any interior point in the sphere, excluding v^*, Newton's method is well-defined and generates a sequence that converges to v^*.

Lemma 3.1.1 *Consider $\mu > 0$ and $(x^*_\mu, y^*_\mu, z^*_\mu)$ contained in $\mathcal{C}_P$. Then under the nondegeneracy assumption, there exists $\hat{\mu} > 0$ so that for $\mu \leq \hat{\mu}$ there is a ball $B((x^*_\mu, z^*_\mu); \delta_\mu)$ such that for $(x, z) \in B((x^*_\mu, z^*_\mu); \delta_\mu)$, and $(x, z) > 0$, (x, z) satisfy*

$$\begin{array}{llll} x_i \geq C_1 & \text{and} & z_i \leq C_2\mu & i \in \mathcal{B} \\ x_i \leq C_3\mu & \text{and} & z_i \geq C_4 & i \in \mathcal{N} \end{array} \tag{7}$$

for constants $C_1, C_2, C_3, C_4 > 0$.

Proof. Let $R^{2n}_+ = \{(x,z) : x \geq 0, z \geq 0\}$. For each $\mu > 0$, we have

$$(x^*_\mu)_i\,(z^*_\mu)_i = \mu \quad i = 1, \ldots, n. \tag{8}$$

We have that x^*_i, $i \in \mathcal{B}$ and z^*_i, $i \in \mathcal{N}$ are strictly positive, and $(x^*_\mu, z^*_\mu) \to (x^*, z^*)$ as $\mu \to 0$. Then there exist $\hat{\mu} > 0$ and $C > 0$ such that for $\mu \leq \hat{\mu}$, $(x^*_\mu)_i \geq C$ for $i \in \mathcal{B}$ and $(z^*_\mu)_i \geq C$ for $i \in \mathcal{N}$. Consider $\mu \leq \hat{\mu}$. Since (x^*_μ, z^*_μ) is an interior point of R^{2n}_+, there exist $\delta_\mu > 0$ such that $\delta_\mu \leq \min\{\beta\mu, C/2\}$, for $\beta > 0$ and points $(x,z) \in R^{2n}_+$ with $(x,z) > 0$ that satisfy

$$\left\| \begin{pmatrix} x - x^*_\mu \\ z - z^*_\mu \end{pmatrix} \right\| \leq \delta_\mu. \tag{9}$$

From (9), we obtain

$$|\, x_i - (x^*_\mu)_i \,| \leq \delta_\mu \qquad \text{and} \qquad |\, z_i - (z^*_\mu)_i \,| \leq \delta_\mu \qquad \text{for } i = 1, \ldots, n. \tag{10}$$

First we show that x_i for $i \in \mathcal{B}$ are bounded away from zero. From the first relation in (10) we have

$$(x^*_\mu)_i - \delta_\mu \leq x_i \leq \delta_\mu + (x^*_\mu)_i \qquad \text{for } i = 1, \ldots, n. \tag{11}$$

Since $(x^*_\mu)_i \geq C$ for $i \in \mathcal{B}$ and $\delta_\mu \leq C/2$, from (11) we obtain

$$0 < C/2 \leq C - \delta_\mu \leq x_i \qquad \text{for } i \in \mathcal{B}.$$

Thus $x_i \geq C_1$ for $i \in \mathcal{B}$ with $C_1 = C/2$. Similarly, we can show $z_i \geq C_4$ for $i \in \mathcal{N}$ and $C_4 > 0$.

Now, we show the second part of the proof. By (8) and (11) we obtain

$$x_i \leq \delta_\mu + \frac{\mu}{(z^*_\mu)_i} = (\delta_\mu/\mu + 1/(z^*_\mu)_i)\mu.$$

Now $\delta_\mu/\mu \leq \beta$, and for $i \in \mathcal{N}$, $(z^*_\mu)_i \geq C$, then

$$x_i \leq (\beta + \frac{1}{C})\mu = C_3\mu \text{ for } i \in \mathcal{N}$$

where $C_3 = \beta + 1/C > 0$. Therefore

$$x_i \leq C_3\mu \quad i \in \mathcal{N}.$$

Similarly, from the second relation in (10), we obtain that for some constant $C_2 > 0$, $z_i \leq C_2\mu$, for $i \in \mathcal{B}$. □

Lemma 3.1.2 *Consider $\mu > 0$ and (x^*_μ, y^*_μ) contained in $\mathcal{C}_B$. Then under the nondegeneracy assumption, there exists $\hat{\mu} > 0$ so that for $\mu \leq \hat{\mu}$ there is a ball $B(x^*_\mu; \delta_\mu)$ such that for $x \in B(x^*_\mu; \delta_\mu)$, and $x > 0$, x satisfies*

$$\begin{array}{lll} x_i \geq G_1 & and & i \in \mathcal{B} \\ x_i \leq G_2\mu & and & i \in \mathcal{N} \end{array} \tag{12}$$

for constants $G_1, G_2 > 0$.

Proof. The proof is similar to that of Lemma 3.1.1 so we omit the proof. □

Lemma 3.1.3 *Suppose the nondegeneracy assumption holds. Let $\bar{A} \equiv AW$ where $W = (X)^{\frac{1}{2}}(Z)^{-\frac{1}{2}}$. Define*

$$P = \bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{A} \equiv \begin{bmatrix} P_{\mathcal{BB}} & P_{\mathcal{BN}} \\ P_{\mathcal{NB}} & P_{\mathcal{NN}} \end{bmatrix} \tag{13}$$

where $P_{\mathcal{BB}} \in R^{m\times m}, P_{\mathcal{BN}} \in R^{m\times(n-m)}, P_{\mathcal{NB}} \in R^{(n-m)\times m}$, and $P_{\mathcal{NN}} \in R^{(n-m)\times(n-m)}$. Then there exists $\tilde{\mu} > 0$ such that for $0 < \mu \leq \tilde{\mu}$, and for $(x, z) \in B((x^*_\mu, z^*_\mu); \delta_\mu)$ where δ_μ is such that Lemma 3.1.1 holds,

$$\begin{array}{rclcrcl} \|P_{\mathcal{BB}} - I_m\| & \leq & C_6\mu^2, & & \|P_{\mathcal{BN}}\| & \leq & C_7\mu \\ \|P_{\mathcal{NB}}\| & \leq & C_8\mu, & & \|P_{\mathcal{NN}}\| & \leq & C_9\mu^2 \end{array} \tag{14}$$

for constants $C_6, C_7, C_8, C_9 > 0$.

Proof. Consider $0 < \mu \leq \hat{\mu}$ where $\hat{\mu}$ is such that Lemma 3.1.1 holds. Let $W_{\mathcal{B}} = (X_{\mathcal{B}}^{\frac{1}{2}} Z_{\mathcal{B}}^{-\frac{1}{2}})$ and $W_{\mathcal{N}} = (X_{\mathcal{N}}^{\frac{1}{2}} Z_{\mathcal{N}}^{-\frac{1}{2}})$. Note that $W_{\mathcal{B}}$ and $W_{\mathcal{N}}$ are nonsingular. Let $\bar{A} \equiv AW = [A_{\mathcal{B}}W_{\mathcal{B}}\ A_{\mathcal{N}}W_{\mathcal{N}}]$. Then substituting $\bar{A}$ in the definition of P we obtain

$$P = [A_{\mathcal{B}}W_{\mathcal{B}}\ A_{\mathcal{N}}W_{\mathcal{N}}]^T \left(A_{\mathcal{B}}(W_{\mathcal{B}})^2 A_{\mathcal{B}}^T + A_{\mathcal{N}}(W_{\mathcal{N}})^2 A_{\mathcal{N}}^T\right)^{-1} [A_{\mathcal{B}}W_{\mathcal{B}}\ A_{\mathcal{N}}W_{\mathcal{N}}].$$

Now, introduce the $m \times (n-m)$ matrix R where

$$R \quad = \quad (W_{\mathcal{B}})^{-1} A_{\mathcal{B}}^{-1} A_{\mathcal{N}} W_{\mathcal{N}}. \tag{15}$$

Then P can be partitioned as follows

$$P = \begin{bmatrix} (I_m + RR^T)^{-1} & (I_m + RR^T)^{-1}R \\ R^T(I_m + RR^T)^{-1} & R^T(I_m + RR^T)^{-1}R \end{bmatrix}. \tag{16}$$

Applying the bounds in (7) to (15), we obtain $\|R\| \leq C_5\mu$ for a constant $C_5 > 0$. Since $\|RR^T\| = O(\mu^2)$, then $\|RR^T\| \to 0$ as $\mu \to 0$. Then there exists $\tilde{\mu} > 0$ and $\tilde{\mu} \leq \hat{\mu}$ such that

for all $\mu \leq \tilde{\mu}$, we obtain $\|RR^T\| < 1$. Then using the Neumann series on $(I_m + RR^T)^{-1}$, we obtain from (16) and (13) that

$$\begin{array}{rclcrcl} \|P_{\mathcal{BB}} - I_m\| & \leq & C_6\mu^2, & & \|P_{\mathcal{BN}}\| & \leq & C_7\mu \\ \|P_{\mathcal{NB}}\| & \leq & C_8\mu, & & \|P_{\mathcal{NN}}\| & \leq & C_9\mu^2 \end{array}$$

for $\mu \leq \tilde{\mu}$ and constants $C_6, C_7, C_8, C_9 > 0$. □

Lemma 3.1.4 *Suppose the nondegeneracy assumption holds. Let $\bar{A} \equiv AX$. Define*

$$P = \bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{A} \equiv \begin{bmatrix} P_{\mathcal{BB}} & P_{\mathcal{BN}} \\ P_{\mathcal{NB}} & P_{\mathcal{NN}} \end{bmatrix} \tag{17}$$

where $P_{\mathcal{BB}} \in R^{m\times m}, P_{\mathcal{BN}} \in R^{m\times(n-m)}, P_{\mathcal{NB}} \in R^{(n-m)\times m}$, and $P_{\mathcal{NN}} \in R^{(n-m)\times(n-m)}$. Then there exists $\tilde{\mu} > 0$ such that for $0 < \mu \leq \tilde{\mu}$, and for $x \in B(x^*_\mu; \delta_\mu)$ where δ_μ is such that Lemma 3.1.2 holds,

$$\begin{array}{rclcrcl} \|P_{\mathcal{BB}} - I_m\| & \leq & G_3\mu^2, & & \|P_{\mathcal{BN}}\| & \leq & G_4\mu \\ \|P_{\mathcal{NB}}\| & \leq & G_5\mu, & & \|P_{\mathcal{NN}}\| & \leq & G_6\mu^2 \end{array} \tag{18}$$

for constants $G_3, G_4, G_5, G_6 > 0$.

Proof. The proof is similar to that of Lemma 3.1.3 so we omit the proof. □

3.2 Sphere of Convergence for System (3)

We provide a tight result showing that the radius of the sphere of convergence of Newton's method on system (3) decreases to zero in the same order that $\mu \to 0$. Our result follows from showing that a lower-bound and an upper-bound of order μ exist for the radius of the sphere of convergence.

Lemma 3.2.5 *Under the nondegeneracy assumption, there exist $\tilde{\mu} > 0$ and constant $K_1 > 0$ such that for any $\mu \leq \tilde{\mu}$, the radius of the sphere of convergence, $r_B(\mu)$, of Newton's method satisfies*

$$K_1\mu \leq r_B(\mu).$$

Proof. We will prove the above result by showing that the sequence of Newton iterates converges to the solution (x^*_μ, y^*_μ) if the initial point x^0 satisfies

$$\|x^0 - x^*_\mu\| \leq K_1\mu. \tag{19}$$

Consider $\tilde{\mu}$ given in Lemma 3.1.4. Suppose Newton's method is applied to system (3) for a particular value of $\mu \leq \tilde{\mu}$. Denote (x, y) as the current Newton iterate where $x \in B(x^*_\mu; \delta_\mu)$

and x satisfies the conditions given in (12). Now, consider the next Newton iteration to obtain

$$\begin{pmatrix} x^+ \\ y^+ \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} - F'_B(x,y;\mu)^{-1} F_B(x,y;\mu).$$

Using the fact that $F_B(x^*_\mu, y^*_\mu; \mu) = 0$ and evaluating F_B at the given points, we obtain

$$\begin{aligned} \begin{pmatrix} x^+ - x^*_\mu \\ y^+ - y^*_\mu \end{pmatrix} &= F'_B(x,y;\mu)^{-1} \left[F_B(x^*_\mu, y^*_\mu; \mu) - F_B(x,y;\mu) - F'_B(x,y;\mu) \begin{pmatrix} x^*_\mu - x \\ y^*_\mu - y \end{pmatrix} \right] \\ &= F'_B(x,y;\mu)^{-1} \begin{bmatrix} \mu (X^*_\mu)^{-1} e - \mu X^{-1} e + \mu X^{-2}(x^*_\mu - x) \\ 0 \end{bmatrix}. \end{aligned} \tag{20}$$

By Taylor's Theorem,

$$\begin{aligned} (X^*_\mu)^{-1} e &= X^{-1} e - X^{-2}(x^*_\mu - x) + \hat{X}^{-3}(x - x^*_\mu)^2 \quad \text{or} \\ \hat{X}^{-3}(x - x^*_\mu)^2 &= (X^*_\mu)^{-1} e - X^{-1} e + X^{-2}(x^*_\mu - x) \end{aligned} \tag{21}$$

for some $\hat{x}_i \in [\min\{(x^*_\mu)_i, x_i\}, \max\{(x^*_\mu)_i, x_i\}], i = 1 \ldots n$. Substitute (21) into (20) to obtain

$$\begin{pmatrix} x^+ - x^*_\mu \\ y^+ - y^*_\mu \end{pmatrix} = F'_B(x,y;\mu)^{-1} \begin{bmatrix} \mu \hat{X}^{-3}(x - x^*_\mu)^2 \\ 0 \end{bmatrix}, \tag{22}$$

where

$$F'_B(x,y;\mu)^{-1} = \begin{bmatrix} \frac{1}{\mu}[X^2 A^T (AX^2 A^T)^{-1} AX^2 - X^2] & X^2 A^T (AX^2 A^T)^{-1} \\ (AX^2 A^T)^{-1} AX^2 & \mu (AX^2 A^T)^{-1} \end{bmatrix}.$$

Making the above substitution for $(F'_B)^{-1}$ in (22) and multiplying the right-hand-side of (22) we obtain

$$\begin{pmatrix} x^+ - x^*_\mu \\ y^+ - y^*_\mu \end{pmatrix} = \begin{bmatrix} X[XA^T (AX^2 A^T)^{-1} AX - I] X \hat{X}^{-3}(x - x^*_\mu)^2 \\ \mu (AX^2 A^T)^{-1} AX^2 \hat{X}^{-3}(x - x^*_\mu)^2 \end{bmatrix}.$$

Now substitute in the definition of P in (17) to obtain

$$\begin{pmatrix} x^+ - x^*_\mu \\ y^+ - y^*_\mu \end{pmatrix} = \begin{bmatrix} X(P - I) X \hat{X}^{-3}(x - x^*_\mu)^2 \\ \mu (AX^2 A^T)^{-1} AX^2 \hat{X}^{-3}(x - x^*_\mu)^2 \end{bmatrix}. \tag{23}$$

We will consider first the vector $(x^+ - x^*_\mu)$ in (23). If we partition $(x^+ - x^*_\mu)$ into its basic and nonbasic components and use the notation for P in (17), then

$$\begin{aligned}(x^+ - x^*_\mu)_{\mathcal{B}} &= X_{\mathcal{B}}(P_{\mathcal{B}\mathcal{B}} - I_m)X_{\mathcal{B}}\hat{X}_{\mathcal{B}}^{-3}(x - x^*_\mu)^2_{\mathcal{B}} + X_{\mathcal{B}}P_{\mathcal{B}\mathcal{N}}X_{\mathcal{N}}\hat{X}_{\mathcal{N}}^{-3}(x - x^*_\mu)^2_{\mathcal{N}} \\ (x^+ - x^*_\mu)_{\mathcal{N}} &= X_{\mathcal{N}}P_{\mathcal{N}\mathcal{B}}X_{\mathcal{B}}\hat{X}_{\mathcal{B}}^{-3}(x^*_\mu - x)^2_{\mathcal{B}} + X_{\mathcal{N}}(P_{\mathcal{N}\mathcal{N}} - I_m)X_{\mathcal{N}}\hat{X}_{\mathcal{N}}^{-3}(x^*_\mu - x)^2_{\mathcal{N}}\end{aligned}$$

which leads to

$$\begin{aligned}\|(x^+ - x^*_\mu)_{\mathcal{B}}\| &\leq (\|X_{\mathcal{B}}(P_{\mathcal{B}\mathcal{B}} - I_m)X_{\mathcal{B}}\hat{X}_{\mathcal{B}}^{-3}\| + \|X_{\mathcal{B}}P_{\mathcal{B}\mathcal{N}}X_{\mathcal{N}}\hat{X}_{\mathcal{N}}^{-3}\|)\|x - x^*_\mu\|^2 \\ \|(x^+ - x^*_\mu)_{\mathcal{N}}\| &\leq (\|X_{\mathcal{N}}P_{\mathcal{N}\mathcal{B}}X_{\mathcal{B}}\hat{X}_{\mathcal{B}}^{-3}\| + \|X_{\mathcal{N}}(P_{\mathcal{N}\mathcal{N}} - I_m)X_{\mathcal{N}}\hat{X}_{\mathcal{N}}^{-3}\|)\|x - x^*_\mu\|^2.\end{aligned}$$

Applying the bounds given in (18) and (12), we obtain

$$\begin{aligned}\|(x^+ - x^*_\mu)_{\mathcal{B}}\| &\leq C\frac{1}{\mu}\|x - x^*_\mu\|^2, \\ \|(x^+ - x^*_\mu)_{\mathcal{N}}\| &\leq C\frac{1}{\mu}\|x - x^*_\mu\|^2\end{aligned} \tag{24}$$

for some constant $C > 0$. Since $\|x - x^*_\mu\| \leq \delta_\mu$ then using (24), the initial Newton point x^0 must satisfy

$$\|x^0 - x^*_\mu\| \leq \min\{\delta_\mu, \frac{1}{C}\mu\} \tag{25}$$

to obtain convergence to x^*_μ.

Now, consider the remaining m components of (23). Taking the norm and partitioning matrices, we obtain

$$\|y^+ - y^*_\mu\| \leq \|\mu(A_{\mathcal{B}}X^2_{\mathcal{B}}A^T_{\mathcal{B}} + A_{\mathcal{N}}X^2_{\mathcal{N}}A^T_{\mathcal{N}})^{-1}(A_{\mathcal{B}}X^2_{\mathcal{B}}\hat{X}_{\mathcal{B}}^{-3} + A_{\mathcal{N}}X^2_{\mathcal{N}}\hat{X}_{\mathcal{N}}^{-3})\|\,\|(x - x^*_\mu)^2\|,$$

and applying (12), we get

$$\|y^+ - y^*_\mu\| \leq \mu\left[O(1) + O(\mu^2)\right]^{-1}[O(1) + O(1/\mu)]\,\|(x - x^*_\mu)^2\|.$$

Then for a constant $\hat{C} > 0$, we obtain

$$\|y^+ - y^*_\mu\| \leq \hat{C}\|x - x^*_\mu\|^2. \tag{26}$$

Thus, the Newton sequence in y converges to y^*_μ if (25) holds. Then, using (25) and (26), Newton's method guarantees convergence to (x^*_μ, y^*_μ) if the initial iterate x^0 satisfies

$$\|x^0 - x^*_\mu\| \leq K_1\mu \tag{27}$$

for constant $K_1 > 0$ and $\mu \leq \tilde{\mu}$. □

The above lemma shows that the radius of the sphere of convergence of Newton's method satisfies $K_1\mu \leq r_B(\mu)$. It establishes only a lower-bound result for the radius of the sphere of convergence of system (3). To establish that the sphere of convergence decreases to zero at exactly the same order as $\mu \to 0$, we need an upper-bound of the same order. The following lemma establishes such an upper-bound.

Lemma 3.2.6 *Consider Newton's method applied to system (3). There exist constants $\hat{\mu} > 0$ and $K_2 > 0$ such that for any given $\mu \leq \hat{\mu}$, the radius of the sphere of convergence, $r_B(\mu)$, corresponding to this μ satisfies*

$$r_B(\mu) \leq K_2 \mu.$$

Proof. It suffices to show the existence of a point $x \geq 0$ with $\|x - x^*_\mu\| \leq K_2\mu$ from where Newton's method does not converge or is not defined. From Lemma 3.1.2 there exist $\hat{\mu} > 0$ and constant $K_2 > 0$ such that for $\mu \leq \hat{\mu}$ and for $i \in \mathcal{N}$, $(x^*_\mu)_i \leq K_2\mu$. Consider an $i \in \mathcal{N}$, and let

$$x = x^*_\mu - (x^*_\mu)_i e_i$$

where e_i is the ith canonical vector. Obviously, $\|x - x^*_\mu\| \leq K_2\mu$. Moreover, because $x_i = 0$, Newton's method is not defined at x. Therefore, $r_B(\mu) \leq K_2\mu$. □

Theorem 3.1 *There exist constants $\tilde{\mu} > 0$ and $K_1, K_2 > 0$ such that for $\mu \leq \tilde{\mu}$, the radius for the sphere of convergence, $r_B(\mu)$, of Newton's method applied to system (3) satisfies*

$$K_1\mu \leq r_B(\mu) \leq K_2\mu.$$

Proof. Application of Lemma 3.2.5 and Lemma 3.2.6 produces the result. □

For the log-barrier formulation of the nonlinear program with inequality constraints, S. Wright provides a lower-bound result for the radius of the sphere of convergence. In [10], it is shown that there exists a $\bar{\mu} > 0$ such that for $\mu \leq \bar{\mu}$ convergence to the solution x^*_μ can be obtained from any point x^0 that satisfies

$$\|x^0 - x^*_\mu\| \leq \hat{C}\mu^\alpha \tag{28}$$

for constant $\hat{C} > 0$ and $\alpha > 1$. In the case of linear programming, our result for system (3) is tight and shows that the radius of the sphere of convergence decreases in the same order as $\mu \to 0$, that is $\alpha = 1$ in our results.

3.3 Sphere of Convergence for System (4)

We now provide a lower-bound estimate for the radius of the sphere of convergence of Newton's method on system (4). We show the lower-bound estimate is independent of the value of μ. This result establishes that the sphere of convergence is bounded away from zero as $\mu \to 0$.

Theorem 3.3.1 *Under the nondegeneracy assumption, there exist constants $D_1 > 0$ and $\tilde{\mu} > 0$ such that for any $\mu \leq \tilde{\mu}$, the radius of the sphere of convergence, $r_P(\mu)$, of Newton's method satisfies*

$$D_1 \leq r_P(\mu).$$

Proof. We will show that Newton's method applied to system (4) generates iterates that converge to the solution $(x^*_\mu, y^*_\mu, z^*_\mu)$ if the initial point (x^0, z^0) satisfies

$$\left\| \begin{pmatrix} x^0 - x^*_\mu \\ z^0 - z^*_\mu \end{pmatrix} \right\| \leq D_1, \tag{29}$$

which then shows that $r_P(\mu) \geq D_1 > 0$. Consider $\mu \leq \tilde{\mu}$ where $\tilde{\mu}$ is given by Lemma 3.1.3. At a given value of μ, let (x, y, z) and $(x^*_\mu, y^*_\mu, z^*_\mu)$ denote respectively the current iterate and the solution of Newton's method applied to system (4). We have by Lemma 3.1.1 that $(x, z) \in B((x^*_\mu, z^*_\mu); \delta_\mu)$ and satisfies the conditions given in (7).

At the subsequent iteration, the Newton iterates are of the form

$$\begin{pmatrix} x^+ \\ y^+ \\ z^+ \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} - F'_P(x, y, z; \mu)^{-1} F_P(x, y, z; \mu).$$

Since $F_P(x^*_\mu, y^*_\mu, z^*_\mu; \mu) = 0$, we obtain

$$\begin{aligned} \begin{pmatrix} x^+ - x^*_\mu \\ y^+ - y^*_\mu \\ z^+ - z^*_\mu \end{pmatrix} &= F'_P(x, y, z; \mu)^{-1} \Big[F_P(x^*_\mu, y^*_\mu, z^*_\mu; \mu) - F_P(x, y, z; \mu) \\ &\qquad - F'_P(x, y, z; \mu)\{(x^*_\mu, y^*_\mu, z^*_\mu) - (x, y, z)\} \Big] \\ &= F'_P(x, y, z; \mu)^{-1} \begin{bmatrix} 0 \\ 0 \\ (X - X^*_\mu)(Z - Z^*_\mu)e \end{bmatrix} \qquad (30) \\ &= \begin{bmatrix} Z^{-1} \left[I_n - XA^T(AZ^{-1}XA^T)^{-1}AZ^{-1} \right] (X - X^*_\mu)(Z - Z^*_\mu)e \\ -(AZ^{-1}XA^T)^{-1}AZ^{-1}(X - X^*_\mu)(Z - Z^*_\mu)e \\ A^T(AZ^{-1}XA^T)AZ^{-1}(X - X^*_\mu)(Z - Z^*_\mu)e \end{bmatrix} \end{aligned}$$

where the last equality is obtained by multiplying the right hand side of (30).

We first consider convergence in the x variables. Let $W = Z^{-\frac{1}{2}}X^{\frac{1}{2}}$ be a diagonal matrix, then we obtain

$$(x^+ - x^*_\mu) = W \left[I_n - WA^T(AW^2A^T)^{-1}AW \right] WX^{-1}(X - X^*_\mu)(Z - Z^*_\mu)e. \tag{31}$$

Define $P = (WA^T(AW^2A^T)^{-1}AW)$ where P has the submatrix representation given in (13). If we partition the vectors and matrices in (31) into their basic and nonbasic components, we obtain

$$\begin{bmatrix} (x^+ - x^*_\mu)_{\mathcal{B}} \\ (x^+ - x^*_\mu)_{\mathcal{N}} \end{bmatrix} = \begin{bmatrix} W_{\mathcal{B}}(I_m - P_{\mathcal{B}\mathcal{B}})W_{\mathcal{B}}X_{\mathcal{B}}^{-1} & W_{\mathcal{B}}P_{\mathcal{B}\mathcal{N}}W_{\mathcal{N}}X_{\mathcal{N}}^{-1} \\ W_{\mathcal{N}}P_{\mathcal{N}\mathcal{B}}W_{\mathcal{B}}X_{\mathcal{B}}^{-1} & W_{\mathcal{N}}(I_{n-m} - P_{\mathcal{B}\mathcal{N}})W_{\mathcal{N}}X_{\mathcal{N}}^{-1} \end{bmatrix} \begin{bmatrix} (X - X^*_\mu)_{\mathcal{B}}(z - z^*_\mu)_{\mathcal{B}} \\ (X - X^*_\mu)_{\mathcal{N}}(z - z^*_\mu)_{\mathcal{N}} \end{bmatrix}.$$

Now, consider the basic components of $(x^+ - x^*_\mu)$. Substitute for W to obtain

$$\begin{aligned} (x^+ - x^*_\mu)_{\mathcal{B}} &= Z_{\mathcal{B}}^{-\frac{1}{2}}X_{\mathcal{B}}^{\frac{1}{2}}(I_m - P_{\mathcal{B}\mathcal{B}})Z_{\mathcal{B}}^{-\frac{1}{2}}X_{\mathcal{B}}^{\frac{1}{2}}X_{\mathcal{B}}^{-1}(X - X^*_\mu)_{\mathcal{B}}(z - z^*_\mu)_{\mathcal{B}} \\ &+ Z_{\mathcal{B}}^{-\frac{1}{2}}X_{\mathcal{B}}^{\frac{1}{2}}P_{\mathcal{B}\mathcal{N}}Z_{\mathcal{N}}^{-\frac{1}{2}}X_{\mathcal{N}}^{\frac{1}{2}}X_{\mathcal{N}}^{-1}(X - X^*_\mu)_{\mathcal{N}}(z - z^*_\mu)_{\mathcal{N}} \end{aligned}$$

and take the norm of $(x^+ - x^*_\mu)_{\mathcal{B}}$ to obtain

$$\begin{aligned} \|(x^+ - x^*_\mu)_{\mathcal{B}}\| &\leq (\|Z_{\mathcal{B}}^{-\frac{1}{2}}X_{\mathcal{B}}^{\frac{1}{2}}(I_m - P_{\mathcal{B}\mathcal{B}})Z_{\mathcal{B}}^{-\frac{1}{2}}X_{\mathcal{B}}^{\frac{1}{2}}X_{\mathcal{B}}^{-1}\| \\ &+ \|Z_{\mathcal{B}}^{-\frac{1}{2}}X_{\mathcal{B}}^{\frac{1}{2}}P_{\mathcal{B}\mathcal{N}}Z_{\mathcal{N}}^{-\frac{1}{2}}X_{\mathcal{N}}^{\frac{1}{2}}X_{\mathcal{N}}^{-1}\|)\|(X - X^*_\mu)(z - z^*_\mu)\|. \end{aligned}$$

Apply bounds (7) and (14) to the above inequality. Then for constants $D_1, D_2, D_3 > 0$, we obtain

$$\begin{aligned} \|(x^+ - x^*_\mu)_{\mathcal{B}}\| &\leq (D_1\mu + D_2)\|(X - X^*_\mu)(z - z^*_\mu)\| \\ &\leq D_3 \ \max\{\|x - x^*_\mu\|^2, \|z - z^*_\mu\|^2\}. \end{aligned} \tag{32}$$

If we do a similar analysis on $(x^+ - x^*_\mu)_{\mathcal{N}}$, then

$$\begin{aligned} (x^+ - x^*_\mu)_{\mathcal{N}} &= Z_{\mathcal{N}}^{-\frac{1}{2}}X_{\mathcal{N}}^{\frac{1}{2}}P_{\mathcal{N}\mathcal{B}}Z_{\mathcal{B}}^{-\frac{1}{2}}X_{\mathcal{B}}^{\frac{1}{2}}X_{\mathcal{B}}^{-1}(X - X^*_\mu)_{\mathcal{B}}(z - z^*_\mu)_{\mathcal{B}} \\ &+ Z_{\mathcal{N}}^{-\frac{1}{2}}X_{\mathcal{N}}^{\frac{1}{2}}(I_{n-m} - P_{\mathcal{B}\mathcal{N}})Z_{\mathcal{N}}^{-\frac{1}{2}}X_{\mathcal{N}}^{\frac{1}{2}}X_{\mathcal{N}}^{-1}(X - X^*_\mu)_{\mathcal{N}}(z - z^*_\mu)_{\mathcal{N}}. \end{aligned}$$

And hence for constants $D_4, D_5, D_6, D_7 > 0$, we obtain

$$\begin{aligned} \|(x^+ - x^*_\mu)_{\mathcal{N}}\| &\leq (D_4\mu + D_5 + D_6\mu)\ \max\{\|x - x^*_\mu\|^2, \|z - z^*_\mu\|^2\} \\ &\leq D_7 \ \max\{\|x - x^*_\mu\|^2, \|z - z^*_\mu\|^2\}. \end{aligned} \tag{33}$$

Combining (32) and (33) yields

$$\|x^+ - x^*_\mu\| \leq D_8 \ \max\{\|x - x^*_\mu\|^2, \|z - z^*_\mu\|^2\}. \tag{34}$$

where $D_8 = max\{D_3, D_7\}$. Through a similar argument on $(y^+ - y^*_\mu)$ and $(z^+ - z^*_\mu)$, we obtain

$$\|y^+ - y^*_\mu\| \leq D_9 \ \max\{\|x - x^*_\mu\|^2, \|z - z^*_\mu\|^2\} \qquad \text{and} \tag{35a}$$

$$\|z^+ - z^*_\mu\| \leq D_{10} \ \max\{\|x - x^*_\mu\|^2, \|z - z^*_\mu\|^2\} \tag{35b}$$

for constants $D_9, D_{10} > 0$. It follows from (34) and (35b) that if the initial iterate (x^0, z^0) satisfies

$$\begin{array}{rcll} \|x^0 - x^*_\mu\| & \leq & 1/D_8 & \text{and} \\ \|z^0 - z^*_\mu\| & \leq & 1/D_{10} & \end{array}$$

then Newton's method converges to the solution $(x^*_\mu, y^*_\mu, z^*_\mu)$. Thus, a lower bound estimate for the radius of the sphere of convergence is

$$\left\| \begin{pmatrix} x^0 - x^*_\mu \\ z^0 - z^*_\mu \end{pmatrix} \right\| \leq D_1 \quad \text{for } \mu \leq \tilde{\mu} \tag{36}$$

where $D_1 = min\{1/D_8, 1/D_{10}\}$. □

Our analysis shows that the radius of the sphere of convergence is independent of μ and stays bounded away from zero as $\mu \to 0$. This result indicates that the sphere of convergence associated with system (4) would eventually be larger than the sphere of convergence associated with system (3), at least for small μ values, so that $r_B(\mu) \leq r_P(\mu)$. In the next section, we show numerically that this is indeed the case.

4 Sphere of Convergence: Numerical Results

In Section 3, we provided bounds on the radii of the spheres of convergence of Newton's method on systems (3) and (4) under the nondegeneracy assumption. Our analysis shows that at least for small values of μ, the sphere of convergence for system (4) is larger than that for system (3). This result arises from the fact that system (3) is not well-defined in a neighborhood of the solution for $\mu = 0$. Therefore, as $\mu \to 0$, we expect the sphere of convergence to decrease to zero. However, it is not clear what will occur if we consider the half-sphere of convergence for system (3) which only contains points $x > 0$ from where Newton's method converges. In this section, we will obtain numerical upper-bound estimates on the radius of the half-sphere of convergence for Newton's method on system (3) and on the radius of the sphere of convergence for Newton's method on system (4).

Our upper-bound estimates are based on the following simple idea. In order to apply Newton's method with the same initial point for systems (3) and (4), the initial points were selected so that the corresponding equations in systems (3) and (4) produced the same residual, and the complementarity equation was met for system (4). So let $x_\alpha \in R^n$ be an arbitrary unit vector and $\lambda > 0$ be a scalar. Consider applying Newton's method to systems (3) and (4) starting from initial points of the form

$$x^0 = x^*_\mu + \lambda x_\alpha, \quad y^0 = 0, \tag{37}$$

and with $z^0 = \mu(X^0)^{-1}e$ for system (4). If for $\lambda = \lambda_\alpha > 0$, Newton's method does not converge to $v^*_\mu = (x^*_\mu, y^*_\mu, z^*_\mu)$ for system (4) (and to $v^*_\mu = (x^*_\mu, y^*_\mu, \mu(X^*_\mu)^{-1}e)$ for system (3)) , then obviously λ_α is an upper bound for the radius of the sphere of convergence of Newton's method at v^*_μ . This upper bound is the tightest possible in this particular direction if Newton's method converges to v^*_μ for any $\lambda \in (0, \lambda_\alpha)$. Numerically, this upper bound λ_α can be approximated by gradually increasing λ from zero by a small increment

until Newton's method fails to converge. We can generate a tighter upper bound by calculating λ_α for a set of random unit vectors $\{x_\alpha\}$, and then taking $\min_\alpha\{\lambda_\alpha\}$ as an upper bound.

Under the nondegeneracy assumption, for system (4) Newton's method is well-defined in a neighborhood of the solution to the linear program, corresponding to $\mu = 0$, which includes negative values for x and z. Therefore, we can choose x_α to be any unit random vector. For this purpose, ten unit random vectors x_α are selected using the Matlab function `randn`.

As we mentioned earlier, because of the presence of the term X^{-1}, system (3) is not well-defined for $\mu = 0$, and nor is Newton's method in any neighborhood of the solution to the linear program. This implies that the sphere of convergence of Newton's method shrinks to zero as $\mu \to 0$. However, it is not clear at all that the largest half-sphere inside the positive orthant should also shrink to zero as $\mu \to 0$. To be fair to system (3), we use only positive unit random vectors x_α. In this way, we actually obtain upper-bound estimates for the radius of the *half-sphere of convergence* instead of the sphere of convergence. For this purpose, ten unit random vectors x_α are selected using the Matlab function `rand` to ensure $x_\alpha \geq 0$.

To observe the behavior of the radii of the half-sphere of convergence for system (3) and for the sphere of convergence for system (4) as $\mu \to 0$, the numerical procedure described above was performed for various values of $\mu > 0$:

$$\mu = 50,\ 25,\ 1,\ 0.45,\ 0.25,\ 0.10,\ 0.05,\ 0.01,\ 0.0075,\ 0.005,\ 0.0005,\ 0.00005. \tag{38}$$

We chose to include large values of μ ($\mu = 50, 25, 1$) to determine the behavior of the radius of the sphere of convergence of Newton's method when far from the solution at $\mu = 0$. The parameter λ given in (37) was given an initial value of 10^{-10} and was incremented when the convergence criteria

$$\frac{\|v_\mu^* - v^k\|}{\|v_\mu^*\|} < tol$$

was satisfied at some iteration k, where $v^k = (x^k, y^k, \mu(X^k)^{-1}e)$ for system (3), and $v^k = (x^k, y^k, z^k)$ for system (4). Nonconvergence was recorded for a particular system with a given μ value and initial point of the form given in (37) if the maximum number of iterations, which we set to 50, was reached. The convergence tolerance was set to $tol = 10^{-8}$. The numerical solution v_μ^* was obtained by solving system (4) with a given value of μ in (38).

The half-sphere and sphere of convergence were recorded based on the theory provided in Section (3) for the radius of the sphere of convergence for system (3). Thus, we particularly focused on determining the half-sphere/sphere of convergence in terms of the primal variables, x. This test eliminated recording additional information, $\|z^0 - z_\mu^*\|$, provided by the theory in section (3) for system (4). Therefore, the half-sphere and sphere of convergence were recorded as

$$\min_\alpha\{\lambda_\alpha\},\ \text{for } \lambda_\alpha = \|x^0 - x_\mu^*\|.$$

We emphasize that in these experiments, we used the pure Newton's method, so the unit step-length was always taken.

Test problems consisted of six random nondegenerate problems `r1-r6`, the Netlib nondegenerate problems: `scagr7, sc50b, share1b` and the Netlib degenerate problems: `adlittle, afiro, blend, sc50a,` and `share2b`. The random data were generated from a uniform distribution on the interval $(0, 1)$. For a given problem, the same ten unit vectors x_α were used for all values of μ in (38). The problems were run on a Sun Ultra Sparc workstation using Matlab version 5.1. Test problem dimensions can be found in Table 1.

We have observed that for system (4) negative components in the iterates (x^k, z^k) did not preclude convergence; on the other hand, for system (3) negative components in x^k always led to nonconvergence.

4.1 Nondegenerate Problems

We performed our experiments on the set of six random nondegenerate problems and the three Netlib nondegenerate problems. Figures 1-2 show the radii of the half-sphere of convergence associated with system (3) and the sphere of convergence associated with system (4) graphed against the values of μ given in (38). Figure 1 contains the graph for a random problem, and the remaining graphs show results for the Netlib problems. The results show that the radius of the sphere of convergence of Newton's method on system (4) is bounded away from zero even for μ sufficiently small, but the radius of the half-sphere of convergence of Newton's method on system (3) decreases to zero as $\mu \to 0$. Furthermore, our tests show a larger radius of the sphere of convergence of Newton's method on system (4) than on system (3) as $\mu \to 0$. In some instances, the radius for system (3) is larger than that for system (4) for $\mu > 1$. This observation was particularly seen in four of the random problems and in problem `sc50b`, where the two radii slightly differ.

For the two equivalent systems, we observed that if Newton's method failed to converge for an initial point v^0 with parameter λ_α, the final Newton iterate had negative components. For system (3), if any components of x^k were negative the Jacobian matrix was highly ill-conditioned and convergence was precluded for Newton's method.

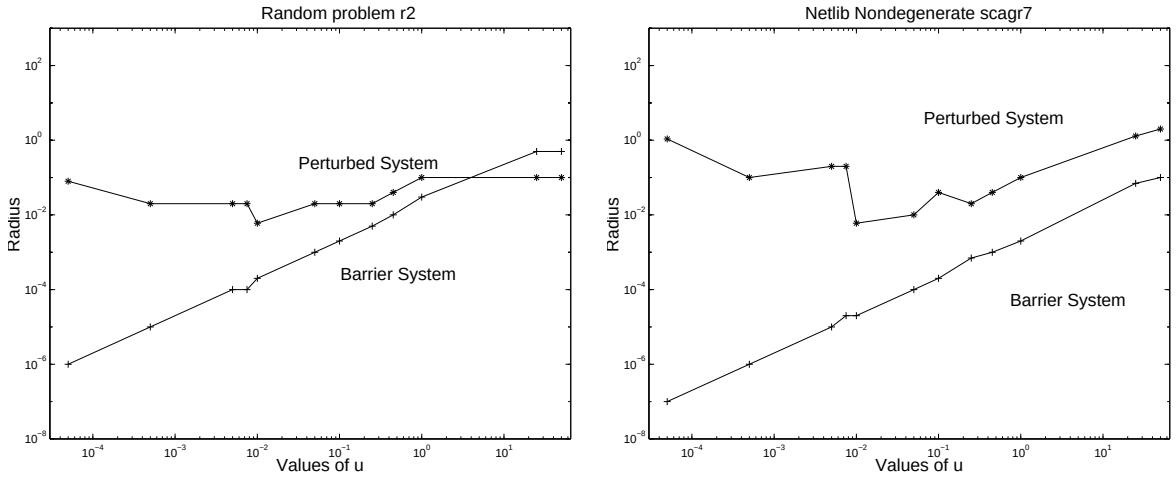

Figure 1: Radii of half-sphere and sphere of convergence, respectively, for Newton's method on systems (3) and (4)

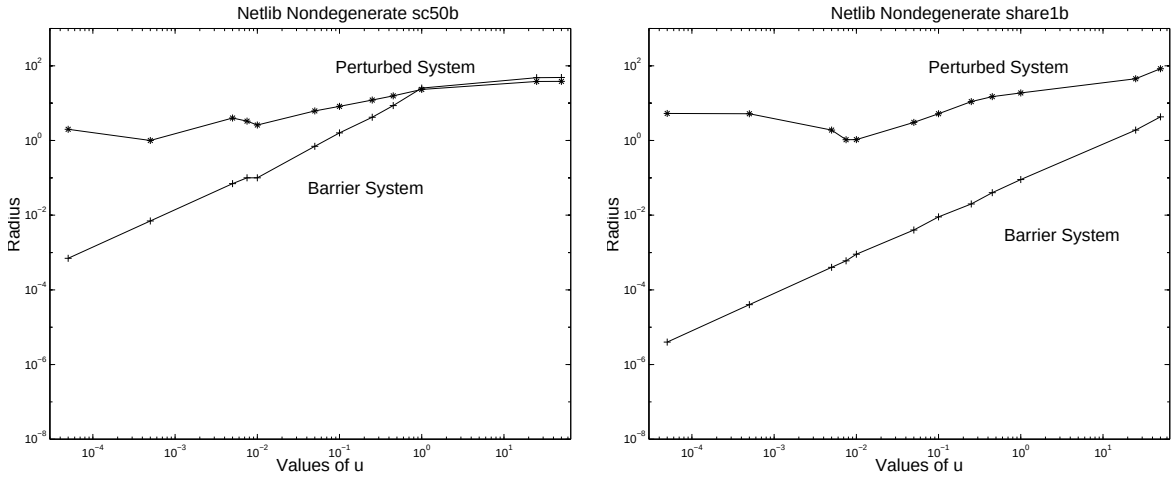

Figure 2: Radii of half-sphere and sphere of convergence, respectively, for Newton's method on systems (3) and (4)

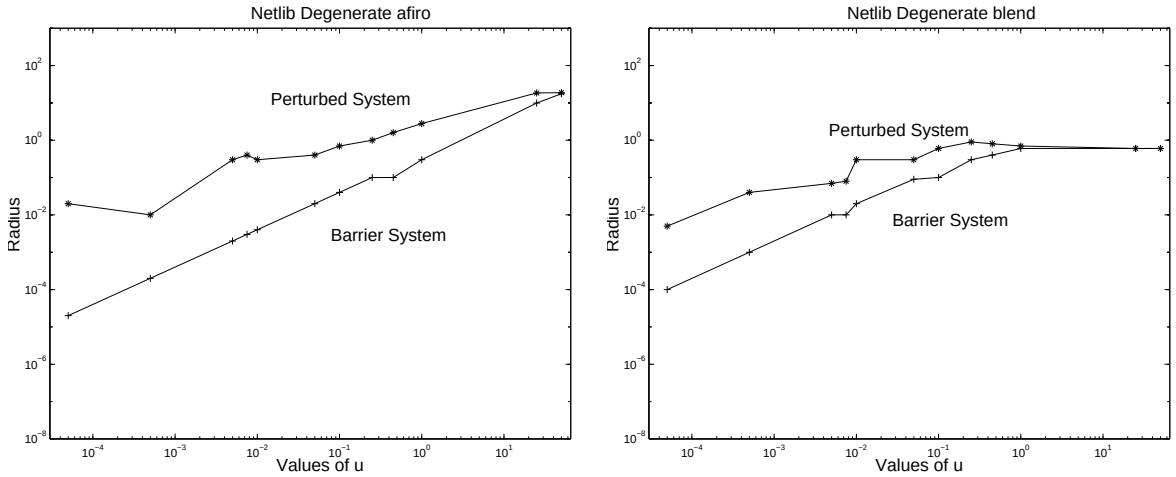

Figure 3: Radii of half-sphere and sphere of convergence, respectively, for Newton's method on systems (3) and (4)

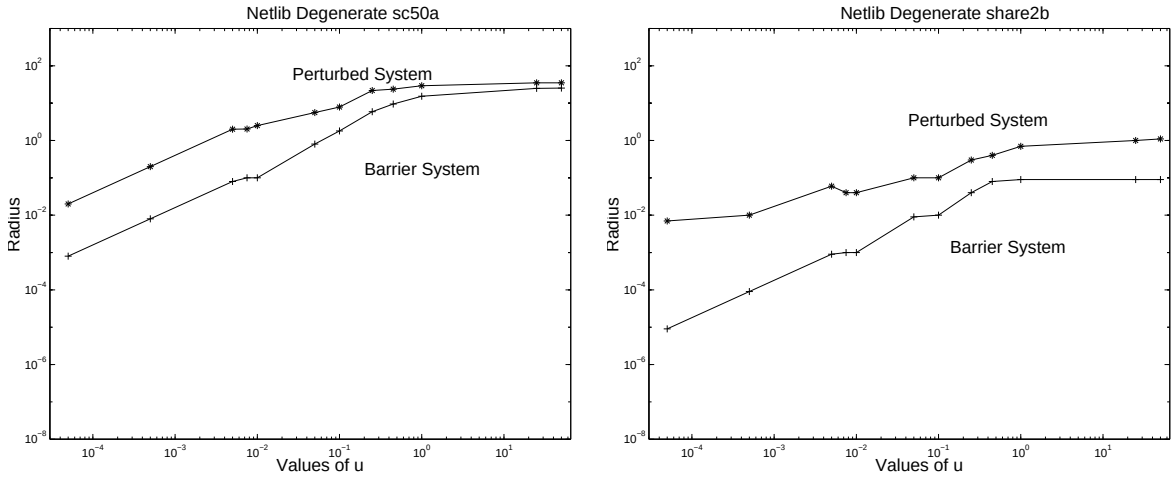

Figure 4: Radii of half-sphere and sphere of convergence, respectively, for Newton's method on systems (3) and (4)

4.2 Degenerate Problems

In addition, we investigated the half-sphere and sphere of convergence, respectively, of Newton's method on systems (3) and (4) for degenerate problems. We have not obtained any theory for the spheres of convergence of Newton's method on degenerate problems but present only numerical results. The same values of μ given in (38) were used. Our numerical results are shown in Figures 3-4.

Our results show that the radius of the half-sphere of convergence of Newton's method on system (3) appears to decrease to zero as μ approaches zero, as in the case with the nondegenerate problems. We also observe that the radius of the sphere of convergence of system (4) decreases to zero as μ decreases to zero, contrary to the results obtained with the nondegenerate problems. However, the radius associated with system (4) stays well above that for system (3), by at least a magnitude of ten, as $\mu \to 0$. In these tests, we observe that the radius of system (4) is always larger than or equal to the radius of system (3) for all the μ values given in (38).

When Newton's method failed to converge, the final Newton iterate contained negative components. Also, if any component of the iterate x^k was negative for system (3), the Jacobian became highly ill-conditioned and convergence of Newton's method was not obtained.

5 Interior-Point Path-Following Algorithm

In the previous two sections, we investigated differences in the behavior of the sphere of convergence of Newton's method when applied to the two equivalent systems (3) and (4). The notion of the sphere/half-sphere of convergence is considered to be a local feature of Newton's method. As such, our theoretical and numerical investigations were both done for fixed μ values, even though we did consider the effects of μ approaching zero. Now, we study the global behavior of a Newton-type method applied to systems (3) and (4) as μ decreases to zero.

As is well known, Newton's method is only locally convergent in general. In most applications, a "globally convergent" method is desired. For linear programming, interior-point path-following algorithms are such "globally convergent" methods. These algorithms generate iterates that closely or loosely follow the central path parameterized by μ to a solution as μ decreases to zero. The central path can be defined through either system (3) or (4), see (5) and (6). For a given value of μ, a damped Newton's method is used to drive the iterates towards the point on the central path corresponding to the given μ value and to keep the nonnegative variables, x in (3) and (x, z) in (4), strictly positive as well.

A very simple interior path-following algorithm is given below as Algorithm IPF. We will apply this same algorithm to both systems (3) and (4) and determine its performance on these two equivalent systems. Since damped Newton is the underlining method used to solve these systems, any performance discrepancy in Algorithm IPF should be due to the different behavior of (damped) Newton's method on the two equivalent systems. In our numerical experiments, we will pay particular attention to the behavior of Algorithm IPF when the iterates are required to follow the central path loosely or closely.

We emphasize that Algorithm IPF is not at all an efficient algorithm for system (4), for which there exist more efficient interior-point path-following algorithms (for example, Kojima et. al's primal-dual interior-point algorithm [5], and Mehrotra's predictor-corrector algorithm [7]). For system (3), however, it has not been shown that one can do significantly better than Algorithm IPF.

Algorithm IPF *Let* $v^k = (x^k, y^k)$ *for system (3), and* $v^k = (x^k, y^k, z^k)$ *for system (4). Given initial iterate* v^0 *with positive components corresponding to the nonnegative variables, and parameters* $\mu, \mu tol > 0$, $\alpha^k, tol \in (0, 1]$, *and* $\sigma, \delta(\mu) \in (0, 1)$.

While $\|F(v^k; \mu)\| + \mu > tol$

- *Solve for* Δv^k *in* $F'(v^k; \mu)\Delta v^k = -F(v^k; \mu)$
- *Form the new iterate*
  $v^{k+1} \leftarrow v^k + \alpha^k \Delta v^k$
- *Update* μ *and* μtol
  if $\|F(v^{k+1}; \mu)\| < \mu tol$
  $\mu \leftarrow \sigma\mu$
  $\mu tol \leftarrow \delta(\mu)$
  end
- *Increment iteration number*
  $k \leftarrow k + 1$

End

The algorithm's goal lies in decreasing μ to zero and in guiding the iterates along the central path to the solution. The Newton step is damped to keep the iterates $x^k > 0$ for system (3) and $x^k, z^k > 0$ for system (4). Damping is performed by selecting α_k to be a given fraction of the steplength to the boundary of the positive orthant from the point v^k along the direction Δv^k. The parameter μ and possibly the tolerance μtol are updated once the iterates are in a neighborhood of the solution (x^*_μ, y^*_μ) for system (3) and $(x^*_\mu, y^*_\mu, z^*_\mu)$ for system (4). When this update occurs, μ is decreased to μ^+ and another nonlinear system is solved having μ^+ as its parameter. In the case of decreasing μtol, the subsequent iterates are required to follow the central path more closely. An approximate solution is obtained when the convergence test is satisfied, that is, when the sum of μ and the Euclidean norms of the residuals of $F(v^k; \mu)$ fall below a tolerance that is close to zero.

6 Numerical Results

In this section, we describe the numerical results obtained from applying Algorithm IPF on systems (3) and (4). Attention is given to explain the role of the centrality tolerance, μtol, as well as, the effect of ill-conditioning on the behavior of Algorithm IPF applied to the two equivalent systems. Depending on the value of μtol and on the ill-conditioning of the nonlinear system, the iterates may not converge to a solution of prescribed accuracy, as we will demonstrate.

Before proceeding to explain the numerical results, we explain the parameters and tolerances used in Algorithm IPF. As mentioned earlier, $\|F(v^k;\mu)\|$ denotes the sum of the Euclidean norms of the residuals of $F(v^k;\mu)$. We let $\sigma = 0.2$ and describe the centrality tolerance, μtol by a given formula $\delta(\mu)$. To keep the new iterates positive, we chose the following steplength calculation of α^k for system (3) and system (4)

$$\alpha^k = min(0.95\frac{-1}{min((U^k)^{-1}\Delta u^k, -0.5)}, 1)$$

where

$$U^k = \begin{cases} X^k & \text{for system (3),} \\ \begin{pmatrix} X^k \\ Z^k \end{pmatrix} & \text{for system (4)} \end{cases}$$

and similarly for Δu^k.

Two sets of experiments were conducted to observe any change in the behavior of Algorithm IPF. In the first set, μtol was fixed for Algorithm IPF, i.e. $\delta(\mu) \equiv \mu tol$. Three tests were conducted depending on the value of μtol given by

$$\{10^{-6}, 10^{-4}, 10^{-2}\}. \tag{39}$$

When $\mu tol = 10^{-6}$, the iterates were required to follow the central path closely, but when $\mu tol = 10^{-2}$ the iterates followed the central path loosely. In the second set of experiments, μtol was initialized with $\mu tol = 0.1$ and updated as follows

$$\delta(\mu) = \frac{1}{10} min(1, \mu). \tag{40}$$

That is, when $\mu \geq 1$, then $\mu tol = 0.1$; when $\mu < 1$, μtol is set to one-tenth the current μ value. As μ decreases to zero, Algorithm IPF is required to follow the central path more and more closely.

Nonconvergence of Algorithm IPF was recorded if the maximum number of iterations, which was set to 300, was reached.

Test problems consisted of six randomly generated, nondegenerate problems `r1-r6`, three nondegenerate problems from Netlib: `scagr7, sc50b, share1b`, and five degenerate problems from Netlib: `adlittle, afiro, blend, sc50a,` and `share2b`. For the random test problems, the data was generated from a uniform distribution on the interval $(0, 1)$; the initial point x^0 was randomly generated, $z^0 = x^0$ and $y^0 = 0$. For the Netlib problems, the initial point was supplied by LIPSOL. For each problem, we used the same initial point for all numerical tests performed. The parameter μ was given an initial value of $\|x^0\|^2/n$.

In the first set of experiments, presented in Tables 1-4, the parameter μtol is fixed at one of the values listed in (39). Tables 1 and 2 show results for $tol = 10^{-6}$ with respect to $\mu tol = 10^{-2}$ and $\mu tol = 10^{-6}$. Tables 3 and 4 show results for a larger convergence tolerance of $tol = 10^{-2}$ with respect to $\mu tol = 10^{-2}$ and $\mu tol = 10^{-6}$.

From Tables 1-4, we see that Algorithm IPF applied to system (4) terminated successfully for all the test problems and for all the tested tol and μtol values . These results

suggest that Algorithm IPF does not need to follow the central path closely in order to obtain high accuracy solutions. In fact, fewer iterations are required if the iterates are allowed to follow the central path loosely. Furthermore, for most problems the condition number of the Jacobian matrix, $F'_P(v^k, \mu)$, remained moderate near the solution as μ approached zero.

When applied to system (3), Algorithm IPF exhibited very different behavior. With $tol = 10^{-6}$ and fixed $\mu tol = 10^{-2}$, we observe from Table 1 that Algorithm IPF does not reach the prescribed accuracy in all the test cases except one (similar results were obtained for $\mu tol = 10^{-4}$). Nonconvergence was due to failure in reaching the prescribed accuracy in dual feasibility $\|A^T y^k + \mu (X^k)^{-1} - c\|$, which usually fell between 10^{-3} and 10^{-6} at the end, short of the required accuracy of 10^{-6}. In addition, we observed that the condition number of the Jacobian matrix $F'_B(v^k, \mu)$ tends to blow up at the end. It is worth noting, however, that when $\mu tol = 10^{-6}$ the Jacobian matrix exhibited better conditioning near the solution, and the algorithm was able to reach the $tol = 10^{-6}$ accuracy. These results strongly suggest that the severe ill-conditioning of $F'_B(v, \mu)$ near the solution can be alleviated if Algorithm IPF is forced to follow the central path closely.

Table 1: Results for $tol = 10^{-6}$ and fixed $\mu tol = 10^{-2}$

Problem	Dimensions	NPDM		NLBM	
Number	m, n	Iters	cond(F'_P)	Iters	cond(F'_B)
r1	13, 22	24	7.3e+03	† 300	∗ Inf
r2	34, 45	28	8.6e+04	300	Inf
r3	13, 15	15	2.5e+03	300	Inf
r4	14, 16	16	2.7e+03	300	Inf
r5	23, 27	27	2.3e+04	300	Inf
r6	26, 34	20	1.3e+04	300	3.2e+51
scagr7	130, 140	63	1.9e+07	300	5.1e+33
share1b	118, 225	75	1.4e+11	300	2.4e+33
sc50b	51, 48	32	1.3e+05	300	2.6e+41
adlittle	57, 97	57	5.8e+12	183	8.5e+33
afiro	28, 32	34	4.9e+12	300	4.4e+55
blend	75, 83	37	1.7e+08	300	4.6e+58
sc50a	51, 48	32	1.4e+07	300	2.1e+77
share2b	97, 79	45	1.5e+11	300	2.2e+45
Total Iters		505		4083	

† refers to the maximum number of iterations reached,
∗ Inf refers to Matlab's representation of positive infinity.

Table 2: Results for $tol = 10^{-6}$ and fixed $\mu tol = 10^{-6}$

Problem Number	NPDM		NLBM	
	Iters	cond(F'_P)	Iters	cond(F'_B)
r1	30	7.3e+03	79	1.2e+11
r2	34	8.6e+04	81	2.0e+11
r3	24	2.5e+03	64	2.2e+08
r4	22	2.7e+03	73	4.3e+10
r5	32	2.3e+04	83	1.3e+12
r6	29	1.3e+04	70	2.9e+08
scagr7	79	1.9e+07	152	8.0e+15
share1b	91	1.4e+11	138	2.6e+14
sc50b	43	1.3e+05	87	1.4e+07
adlittle	68	5.8e+12	145	1.1e+23
afiro	44	4.9e+12	103	8.8e+18
blend	55	1.7e+08	108	3.0e+14
sc50a	46	1.4e+07	89	6.4e+08
share2b	60	1.8e+11	102	2.4e+19
Total Iters	657		1374	

Table 3: Results for $tol = 10^{-2}$ and fixed $\mu tol = 10^{-2}$

Problem Number	NPDM		NLBM	
	Iters	cond(F'_P)	Iters	cond(F'_B)
r1	19	7.3e+03	36	4.1e+07
r2	23	8.6e+04	38	6.5e+07
r3	10	2.4e+03	24	1.6e+05
r4	11	2.7e+03	27	2.8e+06
r5	22	2.3e+04	40	4.2e+08
r6	15	1.5e+04	31	1.1e+05
scagr7	58	1.7e+08	93	2.6e+13
share1b	69	3.4e+13	81	2.8e+13
sc50b	26	3.4e+05	32	8.3e+03
adlittle	52	1.9e+09	91	1.2e+16
afiro	28	3.1e+08	51	3.7e+10
blend	30	8.0e+04	52	1.0e+06
sc50a	26	9.2e+05	32	2.2e+04
share2b	37	8.1e+06	50	2.5e+12
Total Iters	426		678	

Table 4: Results for $tol = 10^{-2}$ and fixed $\mu tol = 10^{-6}$

Problem Number	NPDM		NLBM	
	Iters	cond(F'_P)	Iters	cond(F'_B)
r1	23	7.3e+03	39	4.1e+07
r2	27	8.6e+04	41	6.5e+07
r3	15	2.5e+03	28	1.6e+05
r4	15	2.7e+03	30	2.8e+06
r5	26	2.3e+04	42	4.2e+08
r6	20	1.4e+04	34	1.1e+05
scagr7	70	1.8e+08	111	2.6e+13
share1b	81	3.4e+13	101	2.8e+13
sc50b	33	3.4e+05	43	8.6e+03
adlittle	61	1.9e+09	104	1.2e+16
afiro	35	3.1e+08	60	3.7e+10
blend	39	8.0e+04	64	1.0e+06
sc50a	36	8.9e+05	45	2.3e+04
share2b	46	8.2e+06	58	4.2e+10
Total Iters	527		800	

In order to minimize the effects of ill-conditioning for system (3) in our comparison, we also implemented the algorithm with a much relaxed stopping tolerance of $tol = 10^{-2}$ to observe the behavior of Algorithm IPF away from the solution. In this experiment, Algorithm IPF terminated in all test cases with low accuracy solutions for the both systems. As can be seen from Tables 3 and 4, the Jacobian matrix exhibited smaller condition numbers compared to the previous results when $tol = 10^{-6}$, especially for system (3). In this case, we still observe that considerably fewer iterations are required by Algorithm IPF on system (4) than on (3).

In the second set of experiments, μtol was initialized to $\mu tol = 0.1$ and updated as given in (40). Thus, the iterates were required to follow the central path more closely as μ decreased to zero. In this instance, Algorithm IPF applied to systems (3) and (4) converged for all the test cases. Again, Algorithm IPF required fewer iterations on (4) than on (3) as demonstrated in Table 5.

Table 5: Results for $tol = 10^{-6}$ with initial $\mu tol = 10^{-1}$ and updated

Problem Number	NPDM		NLBM	
	Iters	cond(F'_P)	Iters	cond(F'_B)
r1	24	7.3e+03	74	1.2e+11
r2	29	8.6e+04	77	2.0e+11
r3	15	2.5e+03	59	2.2e+08
r4	17	2.7e+03	66	8.6e+09
r5	27	2.3e+04	77	1.3e+12
r6	23	1.3e+04	67	2.9e+08
scagr7	64	1.9e+07	127	8.0e+15
share1b	75	1.4e+11	112	2.6e+14
sc50b	31	1.3e+05	73	1.4e+07
adlittle	55	5.8e+12	128	1.1e+23
afiro	33	4.9e+12	89	8.8e+18
blend	44	1.7e+08	92	3.0e+14
sc50a	35	1.4e+07	72	6.4e+08
share2b	46	1.8e+11	89	2.4e+19
Total Iters	518		1202	

In summary, we observed that for the tested values of tol or μtol, Algorithm IPF is more robust and more efficient using system (4) than using system (3). Using the former system, Algorithm IPF does not need to follow the central path closely and requires consistently less iterations. We believe that this phenomenon can be partly explained by the difference in the sizes of the spheres of convergence for the two equivalent systems, as studied in Sections 3 and 4.

In the case of system (3), our numerical results suggest that in order to obtain high accuracy solutions, Algorithm IPF needs to follow the central path closely to alleviate the problem of ill-conditioning. For a related work on ill-conditioning of primal-dual systems, we refer the reader to the recent paper by M. Wright [8].

7 Conclusion

In this paper, we studied the local and global behavior of Newton's method on two equivalent systems from linear programming: the optimality system (3) for the log-barrier formulation of the linear program and the perturbed optimality system (4) for the linear program itself.

On the issue of local behavior, we proved that for nondegenerate problems, the radius of the sphere of convergence of Newton's method on system (3) decreases to zero at exactly the same order as $\mu \to 0$. On the other hand, the radius of the sphere of convergence associated with system (4) stays bounded away from zero as $\mu \to 0$. Our numerical experiments confirm these results. Interestingly, for the majority of our tests, the sphere of convergence of Newton's method was consistently larger on system (4) than on system (3) not only for small values of μ, but also for medium values as well. In addition, we provided numerical results for the case of degenerate problems. Our results show that both radii seem to decrease to zero as $\mu \to 0$; however, the radius associated with system (4) is still larger than that of system (3).

It is then not surprising that the superior local behavior of Newton's method on system (4) will be reflected in a global setting. To test the global behavior of (damped) Newton's method on the two equivalent systems, we applied a simple interior-point path-following algorithm for solving linear programs using the two equivalent systems. Our numerical results showed that to obtain convergence, the algorithm using system (3) needs to follow the central path closely to alleviate the ill-conditioning of the Jacobian matrix. However, convergence is always obtained with the algorithm using system (4). Moreover, even when both algorithms followed the central path closely and converged, the algorithm using system (4) required fewer iterations; thus it is more efficient. In fact, far more efficient algorithms exist for system (4) than the simple path-following algorithm used in our test.

We believe that the results in this paper not only confirm that the linear programming system (4) is the system of choice to be used in an interior-point path-following framework, but also provide an explanation. Similar results have been extended to the nonlinear program and will be reported in a subsequent paper.

Acknowledgments

The authors thank Richard Byrd and Amr El-Bakry for their insightful comments and observations on the topics of the paper.

References

[1] Dennis, J.E., and Schnabel, R.B., Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, NJ, 1983. Reprinted by SIAM Publications, 1996.

[2] El-Bakry A.S., Tapia, R.A., Tsuchiya, T., Zhang, Y., On the Formulation and Theory of the Newton Interior-Point Method for Nonlinear Programming, J. Optim. Theory Appl. 89 (3) (1996) 507-541.

[3] Fiacco, A.V., and McCormick, G.P., Nonlinear Programming, Sequential Unconstrained Minimization Techniques, Wiley, New York, 1968. Reprinted by SIAM Publications, 1990.

[4] Frisch, R., The Logarithmic Potential Method of Convex Programming, Technical Report, University Institute of Economics, Oslo, Norway, (1955).

[5] Kojima, M., Mizuno, S., and Yoshise, A., A Primal-Dual Interior Point Algorithm for Linear Programming, Progress in Mathematical Programming, Interior-Point and Related Methods, Springer-Verlag, New York, 1989.

[6] McLinden, L., An Analogue of Moreau's Proximation Theorem, with Application to the Nonlinear Complementarity Problem, Pacific J. Mathematics 88 (1) (1980) 101-161.

[7] Mehrotra, S., On the Implementation of a Primal-Dual Interior Point Method, SIAM J. Optim. 2 (1992) 575-601.

[8] Wright, M.H., Ill-Conditioning and Computational Error in Primal-Dual Interior Methods for Nonlinear Programming, Technical Report 97-4-04, Computing Science Research Center, Bell Laboratories, Murray Hill, New Jersey, February 1997.

[9] Wright, S.J., Primal-Dual Interior-Point Methods, SIAM Publications, 1997.

[10] Wright, S.J., On the convergence of the Newton/log-barrier method, Technical Report ANL/MCS-P681-0897, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois, August 1997.

[11] Zhang, Y., LIPSOL, version 0.4., Department of Computational and Applied Mathematics, Rice University.

[12] Zhang, Y., Tapia, R.A., and Dennis, J.E., On the Superlinear and Quadratic Convergence of Primal-Dual Interior Point Linear Programming Algorithms, SIAM J. Optim. 2 (1992) 304-324.