The Local Behavior of Newton's Method on Two Equivalent Systems from Linear Programming

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

Revised August, 2000

Abstract

Newton's method is a fundamental technique underlying many numerical methods for solving systems of nonlinear equations and optimization problems. However, it is often not fully appreciated that Newton's method can produce significantly different behavior when applied to equivalent systems. In this paper, we investigate differences in the local behavior of Newton's method when applied to two different but equivalent systems from linear programming: the optimality conditions of the logarithmic barrier function formulation, and the equations in the so-called perturbed optimality conditions. Through theoretical analysis and numerical results, we provide an explanation on why Newton's method performs more effectively on the latter system.

Key words. Newton's method; equivalent systems; 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]). It is often not fully appreciated, however, that Newton's method can exhibit significantly different local and global behavior on two equivalent systems. By equivalent systems, we refer to two systems of nonlinear equations that can be derived from one another and essentially share the same set of solutions (though some auxiliary variables/equations may be present in one but not in another). In this paper, we compare the behavior of Newton's method 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 equations in the

*Department of Mathematical Sciences, University of Texas at El Paso, El Paso, Texas, 79968-0514. This author was supported in part by the Center for Research and Parallel Computations, DOE DE-FG03-93ER25178, and the Sloan Foundation.

†Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 77005-1892.

‡Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 77005-1892. This author was supported in part by DOE Grant DE-FG03-97ER25331, DOE/LANL Contract 03891-99-23 and NSF Grant DMS-9973339.

perturbed first-order optimality conditions for the linear program. Though the two nonlinear systems have essentially the same set of solutions, El-Bakry, Tapia, Tsuchiya, and Zhang [2] 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.

Of these two equivalent systems, the perturbed first-order optimality conditions are widely used in interior-point methods for linear programming. However, the reasons for favoring this system have not been fully analyzed. In this paper, we provide an explanation on 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 and provide theoretical results on the radius of the sphere of convergence of Newton's method applied to the two equivalent systems. In Section 4 we present numerical results supporting the theory we developed in the previous section. Finally, we make some concluding remarks in Section 5.

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

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

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

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

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 $e = (1, ..., 1)^T \in R^n$.

2.1 Two Equivalent Systems

We derive one of the equivalent systems by formulating problem (1) in the logarithmic barrier framework. This framework, which was first 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

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

Assume that the feasible set $\{x : Ax = b, x \geq 0\}$ is bounded. Then for every value of $\mu > 0$, there exists a unique solution x_μ^* of the log-barrier subproblem. Under mild assumptions (see 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, and setting the gradient of the Lagrangian equal to zero. 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 $\mu > 0$. In applications of Newton's method near the solution, the Jacobian necessarily becomes ill-conditioned as μ approaches zero (see, [6, 9]).

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 the equivalent defining relation 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 $x, z > 0$. At a solution where x and z contain zero components, the Jacobian may or may not be nonsingular, depending on the degeneracy of the solution.

Kojima, Mizuno and Yoshise [5] first proposed to use system (4) to solve the linear program in a primal-dual interior-point method. 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.

Two things are worth noting in comparing the two systems (3) and (4). First, while F_B and F_B' are undefined for $\mu = 0$, F_P and F_P' are defined. Second, while F_B' is dependent on μ, F_P' is not. These differences will greatly affect the behavior of Newton's method when applied to the two systems as μ approaches zero.

We remark that although the Jacobian $F_P'(x, y, z; \mu)$ does not depend on μ, we will nevertheless use the present notation for the Jacobian to stress its association with system (4) for a given value of $\mu \geq 0$.

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.

2.2 Central Paths

Assume that the strictly feasible set $\{(x, y, z) : Ax = b;\ A^T y + z = c;\ x, z > 0\}$ is nonempty. 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 the set

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

The set of points in $\mathcal{C}_B$ forms a continuous path such that $\lim_{\mu \to 0}(x_\mu^*, y_\mu^*) = (x^*, y^*)$ ([3, 7]).

We remark that systems (3) and (4) are equivalent, in the sense that for $\mu > 0$, and $(x_\mu^*, y_\mu^*) \in \mathcal{C}_B$,

$$F_B(x_\mu^*, y_\mu^*; \mu) = 0 \Longleftrightarrow F_P(x_\mu^*, y_\mu^*, z_\mu^*; \mu) = 0$$

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 the set

$$\mathcal{C}_P = \{(x_\mu^*, y_\mu^*, z_\mu^*) : F_P(x_\mu^*, y_\mu^*, z_\mu^*; \mu) = 0, \quad \mu,\ x_\mu^*, z_\mu^* > 0\}. \tag{6}$$

2.3 Assumption and Notation

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

Nondegeneracy Assumption. Let the matrix A be of full rank m, and 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.

The nondegeneracy assumption guarantees that (x^*, y^*, z^*) is an isolated solution point in the primal-dual space. It is also well known that the pair (x^*, z^*) satisfies strict complementarity.

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 in which Newton's method is well-defined. More importantly, starting from any point in the neighborhood, Newton's method guarantees convergence to the solution. For systems (3) and (4), such a neighborhood also exists about the solution for any given $\mu > 0$ under our nondegeneracy assumption. 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. These results provide a theoretical explanation on why Newton's method is 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 system (3).

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, p. 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 $B(v^*; r) = \{v : \|v - v^*\| \le r\}$.

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 (with unit steplength) is well-defined and generates a sequence that converges to v^*.

Lemma 3.1.1 *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 $\mu \le \hat{\mu}$ there is a ball $B(x^*_\mu; \delta_\mu)$ such that for $x \in B(x^*_\mu; \delta_\mu)$, and $x > 0$, x satisfies*

$$\begin{array}{rcccl} C_1 & \le & x_i & \le & C_2, \quad i \in \mathcal{B} \\ C_3\mu & \le & x_i & \le & C_4\mu, \quad i \in \mathcal{N} \end{array} \tag{7}$$

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

Proof. Since x_i^*, $i \in \mathcal{B}$ is strictly positive, and $x_\mu^* \to x^*$ as $\mu \to 0$, there exist $\hat{\mu} > 0$ and constants $G_1, G_2 > 0$ such that for $\mu \leq \hat{\mu}$, we have $G_1 \leq (x_\mu^*)_i \leq G_2$ for $i \in \mathcal{B}$. Consider $\mu \leq \hat{\mu}$. Since x_μ^* is an interior point of $R_+^n \equiv \{x : x \geq 0\}$, there exist $\delta_\mu > 0$ such that $\delta_\mu \leq \min\{\beta\mu, G_1/2\}$, for some $\beta > 0$ and points $x \in R_+^n$ with $x > 0$ that satisfy

$$\left\| x - x_\mu^* \right\| \leq \delta_\mu \tag{8}$$

which implies that for $i = 1, \ldots, n$

$$| \, x_i - (x_\mu^*)_i \, | \leq \delta_\mu. \tag{9}$$

First we show that x_i for $i \in \mathcal{B}$ are bounded away from zero. From (9) for $i = 1, \ldots, n$

$$(x_\mu^*)_i - \delta_\mu \leq x_i \leq \delta_\mu + (x_\mu^*)_i. \tag{10}$$

Since $(x_\mu^*)_i$ for $i \in \mathcal{B}$ are bounded above and below and $\delta_\mu \leq G_1/2$, from (10) we obtain

$$0 < G_1/2 \leq x_i \leq G_1/2 + G_2, \qquad i \in \mathcal{B}.$$

Thus $C_1 \leq x_i \leq C_2$ for $i \in \mathcal{B}$ with $C_1 = G_1/2 > 0$ and $C_2 = G_1/2 + G_2 > 0$.

Now, we show the second part of the proof. By the nondegeneracy assumption, strict complementarity holds at the solution, which implies

$$\lim_{\mu \to 0} \frac{\mu}{(x_\mu^*)_i} > 0, \quad i \in \mathcal{N}.$$

Hence, for sufficiently small μ and for some constants $G_3, G_4 > 0$

$$\frac{1}{G_4} \leq \frac{\mu}{(x_\mu^*)_i} \leq \frac{1}{G_3}, \quad i \in \mathcal{N}.$$

Consequently,

$$G_3\mu \leq (x_\mu^*)_i \leq G_4\mu, \qquad i \in \mathcal{N}. \tag{11}$$

By (10) and (11) for $i \in \mathcal{N}$ we obtain

$$(G_3 - \delta_\mu/\mu)\mu \leq x_i \leq (\delta_\mu/\mu + G_4)\mu.$$

Since $\delta_\mu/\mu \leq \beta$, then for $i \in \mathcal{N}$

$$(G_3 - \beta)\mu \leq x_i \leq (\beta + G_4)\mu.$$

Therefore,

$$C_3\mu \leq x_i \leq C_4\mu, \qquad i \in \mathcal{N}$$

where $C_3 = G_3 - \beta$ and $C_4 = \beta + G_4 > 0$. We note that β can be chosen so that $C_3 > 0$. $\square$

Lemma 3.1.2 *Define*

$$P = XA^T(AX^2A^T)^{-1}AX \equiv \begin{bmatrix} P_{\mathcal{BB}} & P_{\mathcal{BN}} \\ P_{\mathcal{NB}} & P_{\mathcal{NN}} \end{bmatrix} \tag{12}$$

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 under the nondegeneracy assumption, 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 3.1.1 holds,*

$$\begin{array}{rclcrcl} \|P_{\mathcal{BB}} - I_m\| & \leq & C_6\mu^2, & & \|P_{\mathcal{BN}}\| & \leq & C_7\mu \\ \|P_{\mathcal{NB}}\| & \leq & C_8\mu, & & \|P_{\mathcal{NN}}\| & \leq & C_9\mu^2 \end{array} \tag{13}$$

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. Substituting $AX = [A_{\mathcal{B}}X_{\mathcal{B}}\ A_{\mathcal{N}}X_{\mathcal{N}}]$ in the definition of P we obtain

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

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

$$R \ = \ (X_{\mathcal{B}})^{-1}A_{\mathcal{B}}^{-1}A_{\mathcal{N}}X_{\mathcal{N}}. \tag{14}$$

Then P can be partitioned as follows

$$P = \begin{bmatrix} (I_m + RR^T)^{-1} & (I_m + RR^T)^{-1}R \\ R^T(I_m + RR^T)^{-1} & R^T(I_m + RR^T)^{-1}R \end{bmatrix}. \tag{15}$$

Applying the bounds in (7) to (14), 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 (15) 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 $C_6, C_7, C_8, C_9 > 0$. □

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

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

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

$$\|x^0 - x^*_\mu\| < K_1\mu. \tag{16}$$

Consider $\tilde{\mu}$ given in Lemma 3.1.2. Assume Newton's method is applied to system (3) for a particular value of $\mu \le \tilde{\mu}$. Denote (x, y) as the current Newton iterate where $x \in B(x^*_\mu; \delta_\mu)$ and x satisfies the conditions given in (7). Now, consider the next Newton iteration

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

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

$$\begin{aligned} \begin{pmatrix} x^+ - x^*_\mu \\ y^+ - y^*_\mu \end{pmatrix} &= F'_B(x, y; \mu)^{-1} \left[F_B(x^*_\mu, y^*_\mu; \mu) - F_B(x, y; \mu) - F'_B(x, y; \mu) \begin{pmatrix} x^*_\mu - x \\ y^*_\mu - y \end{pmatrix} \right] \\ &= F'_B(x, y; \mu)^{-1} \begin{bmatrix} \mu(X^*_\mu)^{-1}e - \mu X^{-1}e + \mu X^{-2}(x^*_\mu - x) \\ 0 \end{bmatrix}. \end{aligned} \tag{17}$$

By Taylor's Theorem,

$$\begin{aligned} (X^*_\mu)^{-1}e &= X^{-1}e - X^{-2}(x^*_\mu - x) + \hat{X}^{-3}(x - x^*_\mu)^2 \quad \text{or} \\ \hat{X}^{-3}(x - x^*_\mu)^2 &= (X^*_\mu)^{-1}e - X^{-1}e + X^{-2}(x^*_\mu - x) \end{aligned} \tag{18}$$

for some $\hat{x}_i \in [\min\{(x^*_\mu)_i, x_i\}, \max\{(x^*_\mu)_i, x_i\}]$, $i = 1, \ldots, n$. Substituting (18) into (17) we obtain

$$\begin{pmatrix} x^+ - x^*_\mu \\ y^+ - y^*_\mu \end{pmatrix} = F'_B(x, y; \mu)^{-1} \begin{bmatrix} \mu\hat{X}^{-3}(x - x^*_\mu)^2 \\ 0 \end{bmatrix}, \tag{19}$$

where

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

Making the above substitution for $(F'_B)^{-1}$ in (19) and multiplying the right-hand-side of (19) we obtain

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

Using the definition of P in (12), we rewrite

$$\begin{pmatrix} x^+ - x^*_\mu \\ y^+ - y^*_\mu \end{pmatrix} = \begin{bmatrix} X(P - I)X\hat{X}^{-3}(x - x^*_\mu)^2 \\ \mu(AX^2A^T)^{-1}AX^2\hat{X}^{-3}(x - x^*_\mu)^2 \end{bmatrix}. \tag{20}$$

We now consider first the vector $(x^+ - x^*_\mu)$ in (20). If we partition $(x^+ - x^*_\mu)$ into its basic and nonbasic components and use the notation for P in (12), then

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

which leads to

$$
\begin{aligned}
\|(x^+ - x^*_\mu)_{\mathcal{B}}\| &\le (\|X_{\mathcal{B}}(P_{\mathcal{BB}} - I_m)X_{\mathcal{B}}\hat{X}_{\mathcal{B}}^{-3}\| + \|X_{\mathcal{B}}P_{\mathcal{BN}}X_{\mathcal{N}}\hat{X}_{\mathcal{N}}^{-3}\|)\|x - x^*_\mu\|^2 \\
\|(x^+ - x^*_\mu)_{\mathcal{N}}\| &\le (\|X_{\mathcal{N}}P_{\mathcal{NB}}X_{\mathcal{B}}\hat{X}_{\mathcal{B}}^{-3}\| + \|X_{\mathcal{N}}(P_{\mathcal{NN}} - I_{n-m})X_{\mathcal{N}}\hat{X}_{\mathcal{N}}^{-3}\|)\|x - x^*_\mu\|^2.
\end{aligned}
$$

Applying the bounds given in (7) and (13) to the above, we obtain

$$\|x^+ - x^*_\mu\| \le \frac{C}{\mu}\|x - x^*_\mu\|^2, \tag{21}$$

for some constant $C > 0$. Recall that $\|x - x^*_\mu\| \le \delta_\mu \le \beta\mu$. It follows from (21) that if the initial iterate satisfies

$$\|x - x^*_\mu\| < \min\left\{\beta, \frac{1}{C}\right\}\mu \tag{22}$$

then the x-component of the Newton iteration sequence will converge to x^*_μ.

Now, consider the remaining m components of (20). Taking the norm and partitioning matrices, we obtain

$$
\begin{aligned}
&\|y^+ - y^*_\mu\| \\
\le\ & \|\mu(A_{\mathcal{B}}X^2_{\mathcal{B}}A^T_{\mathcal{B}} + A_{\mathcal{N}}X^2_{\mathcal{N}}A^T_{\mathcal{N}})^{-1}(A_{\mathcal{B}}X^2_{\mathcal{B}}\hat{X}_{\mathcal{B}}^{-3}(x - x^*_\mu)^2_{\mathcal{B}} + A_{\mathcal{N}}X^2_{\mathcal{N}}\hat{X}_{\mathcal{N}}^{-3}(x - x^*_\mu)^2_{\mathcal{N}})\| \\
\le\ & \mu\|(A_B X^2_B A^T_B + A_N X^2_N A^T_N)^{-1}\| \left(\|A_B X^2_B \hat{X}_B^{-3}\| + \|A_N X^2_N \hat{X}_N^{-3}\|\right) \|(x - x^*_\mu)^2\|.
\end{aligned}
$$

Applying (7), we have

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

Then it follows that for some constant $\hat{C} > 0$

$$\|y^+ - y^*_\mu\| \le \hat{C}\|x - x^*_\mu\|^2. \tag{23}$$

Thus, the y-component of the Newton iteration sequence converges to y^*_μ if (22) holds. In view of (22) and (23), we conclude that the Newton iteration sequence converges to (x^*_μ, y^*_μ) if the initial iterate x^0 satisfies (16) for all $\mu \le \tilde{\mu}$ and for K_1 defined as the constant in the right-hand side of (22). $\square$

The above lemma shows that the radius of the sphere of convergence of Newton's method satisfies $K_1\mu \le r_B(\mu)$. It establishes only a lower-bound result for the radius of the sphere of convergence of system (3). To establish that the radius of 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 3.2.4 *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 3.1.1 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, $r_B(\mu) \leq K_2\mu$. □

Now we are ready to give the main result for system (3).

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 3.2.3 and Lemma 3.2.4 produces the result. □

Since system (3) is not well-defined in a neighborhood of the solution for $\mu = 0$, it is not surprising that as $\mu \to 0$, the sphere of convergence would decrease to zero. However, it was previously not known that the radius of the sphere of convergence would decrease to zero at exactly the same rate as μ goes to zero. For the log-barrier formulation of the nonlinear program with inequality constraints, S. Wright [14] provided a lower-bound result for the radius of the sphere of convergence. In [14], 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

$$\|x^0 - x^*_\mu\| \leq \hat{C}\mu^\alpha \tag{24}$$

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 $R > 0$ 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*

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

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, y^0, z^0) satisfies

$$\left\| \begin{pmatrix} x^0 - x^*_\mu \\ y^0 - y^*_\mu \\ z^0 - z^*_\mu \end{pmatrix} \right\| \quad < \quad R, \tag{25}$$

which then implies that $r_P(\mu) \geq R > 0$.

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). Since $F'_P(x^*, y^*, z^*; 0)$ is nonsingular, by continuity there exist positive constants η and D such that

$$\|F'_P(x, y, z; \mu)^{-1}\| < D, \;\; \forall (x, y, z) \in B((x^*, y^*, z^*); 2\eta), \tag{26}$$

where $F'_P(x, y, z; \mu)$ is independent of μ.

Since $(x^*_\mu, y^*_\mu, z^*_\mu) \to (x^*, y^*, z^*)$, we can choose $\tilde{\mu}$ such that for all $\mu \leq \tilde{\mu}$,

$$\left\| \begin{pmatrix} x^*_\mu - x^* \\ y^*_\mu - y^* \\ z^*_\mu - z^* \end{pmatrix} \right\| \quad \leq \quad \eta.$$

Now let $(x, y, z) \in B((x^*_\mu, y^*_\mu, z^*_\mu); \eta)$ for $\mu \leq \tilde{\mu}$. Then

$$\left\| \begin{pmatrix} x - x^* \\ y - y^* \\ z - z^* \end{pmatrix} \right\| \leq \left\| \begin{pmatrix} x - x^*_\mu \\ y - y^*_\mu \\ z - z^*_\mu \end{pmatrix} \right\| + \left\| \begin{pmatrix} x^*_\mu - x^* \\ y^*_\mu - y^* \\ z^*_\mu - z^* \end{pmatrix} \right\| \leq \eta + \eta = 2\eta.$$

Hence, for such chosen μ and (x, y, z), the Jacobian $F'_P(x, y, z; \mu)$ is nonsingular and satisfies $\|F'_P(x, y, z; \mu)^{-1}\| < D$ in view of (26).

The Newton iterates are of the form

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

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

$$\begin{aligned} \begin{pmatrix} x^+ - x^*_\mu \\ y^+ - y^*_\mu \\ z^+ - z^*_\mu \end{pmatrix} \quad &= \quad F'_P(x, y, z; \mu)^{-1} \Big[F_P(x^*_\mu, y^*_\mu, z^*_\mu; \mu) - F_P(x, y, z; \mu) \\ &\quad\quad - F'_P(x, y, z; \mu)\{(x^*_\mu, y^*_\mu, z^*_\mu) - (x, y, z)\} \Big] \\ &= \quad F'_P(x, y, z; \mu)^{-1} \begin{bmatrix} 0 \\ 0 \\ (X - X^*_\mu)(Z - Z^*_\mu)e \end{bmatrix}. \end{aligned}$$

Hence,

$$\left\| \begin{pmatrix} x^+ - x^*_\mu \\ y^+ - y^*_\mu \\ z^+ - z^*_\mu \end{pmatrix} \right\| \leq \left\| F'_P(x,y,z;\mu)^{-1} \right\| \left\| (X - X^*_\mu)(Z - Z^*_\mu)e \right\| \leq D \left\| \begin{pmatrix} x - x^*_\mu \\ y - y^*_\mu \\ z - z^*_\mu \end{pmatrix} \right\|^2 . \qquad (27)$$

It follows from (27) that if the initial iterate (x^0, y^0, z^0) satisfies

$$\left\| \begin{pmatrix} x^0 - x^*_\mu \\ y^0 - y^*_\mu \\ z^0 - z^*_\mu \end{pmatrix} \right\| \ < \ 1/D$$

for any value of $\mu \in (0, \tilde{\mu})$, 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 of Newton's method is $R = \min\{\eta, 1/D\}$, which is independent of μ. □

Our analysis shows that the radius of the sphere of convergence is independent of μ and thus 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); that is, at least for small μ values, $r_B(\mu) \leq r_P(\mu)$. 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). In this section, we try to compute numerical upper-bound estimates on the radii of the spheres of convergence for Newton's method on systems (3) and (4). The purpose of these computations is not only to confirm our theory for nondegenerate problems for small values of μ, but also to obtain empirical information on degenerate problems and for relatively large values of μ.

4.1 Description of the Numerical Experiments

We note that (a) the variables are (x, y) for system (3), and (x, y, z) for (4); (b) for all $\mu > 0$, $(z^*_\mu)_i = \mu/(x^*_\mu)_i$; and (c) the variable y appears linearly in both systems (3) and (4). For the sake of comparison, we will only estimate the radii of the spheres of convergence for both systems in the x-space, using a fixed initial point for y. More specifically, for any given $\mu > 0$ and any chosen initial point x^0, we set $y^0 = 0$ and $z^0 = \mu(X^0)^{-1}e$. In the rest of the section, the term "sphere of convergence" is always restricted to the x-space only.

Our upper-bound estimates are based on the following simple idea. 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

$$x^0 = x^*_\mu + \lambda x_\alpha, \quad y^0 = 0, \qquad (28)$$

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 (x^*_μ, y^*_μ) for system (3) (or to $(x^*_\mu, y^*_\mu, \mu(X^*_\mu)^{-1}e)$ for system (4)), then obviously λ_α is an upper bound for the radius of the sphere of convergence of Newton's method at x^*_μ. This upper bound is the tightest possible in this particular direction if Newton's method converges to x^*_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, which includes negative values for x and z. Therefore, we can choose x_α to be any unit random vector. In our experiments on system (4), ten unit random vectors x_α are selected using the Matlab function `randn` followed by a normalization.

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 fact 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 where Newton's method is well-defined and convergent 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 estimate upper-bound for the radius of the *half-sphere of convergence* instead of the sphere of convergence. In our experiments on system (3), ten positive unit random vectors x_α are selected using the Matlab function `rand` followed by a normalization.

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

$$\mu = 50,\ 25,\ 1,\ 0.45,\ 0.25,\ 0.10,\ 0.05,\ 0.01,\ 0.0075,\ 0.005,\ 0.0005,\ 0.00005. \tag{29}$$

We include large values of μ ($\mu = 50, 25, 1$) in order to see 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 in (28) 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). Such residual definitions were designed to prevent the stopping criterion from being in favor of one system or another. Nonconvergence was recorded for a particular run with a given μ value and initial point of the form given in (28) 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 (29) and with a stopping tolerance of 10^{-8}. In particular, system (4) was solved using an interior-point primal-dual method.

For the given set of μ values, the estimates for the radii of the half-sphere or sphere of convergence were recorded as $\min_\alpha\{\lambda_\alpha\}$, where by our construction $\lambda_\alpha = \|x^0 - x^*_\mu\|$, and $\alpha = 1, 2, \cdots, 10$. We emphasize that in these experiments, we always used the pure Newton's method with the unit step-length.

In our implementation, we used a plain partial-pivoting Gaussian elimination (Matlab backslash) to solve all linear equations in computing Newton directions for both systems (3) and (4). This should minimize the effect of ill-conditioning caused by different elimination schemes that exploit sparsity.

4.2 Test Problems

Test problems consisted of six randomly generated problems `r1-r6` which are all nondegenerate, the Netlib nondegenerate problems: `scagr7, sc50b, share1b` and the Netlib degenerate problems: `adlittle, afiro, blend, sc50a,` and `share2b`. For the random problems, the data were generated from a uniform distribution on the interval $(0,1)$ using the Matlab function `rand`. For a given problem, the same ten unit random vectors x_α were used for all values of μ in (29). The problems were run on a Sun Ultra Sparc workstation using Matlab version 5.1. Test problem dimensions can be found in Table 1, where the first nine problems are nondegenerate and the last five are degenerate. We mention that some Netlib problems are not in the standard form and have inequality constraints, and the numbers of variables shown are before the addition of slack variables.

Table 1: Problem Test Set

Problem Number	Dimensions m, n
r1	13, 22
r2	34, 45
r3	13, 15
r4	14, 16
r5	23, 27
r6	26, 34
scagr7	130, 140
share1b	118, 225
sc50b	51, 48
adlittle	57, 97
afiro	28, 32
blend	75, 83
sc50a	51, 48
share2b	97, 79

4.3 Results for Nondegenerate Problems

We present numerical results for only four nondegenerate problems since we obtained similar results for the remaining 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 (29). 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' 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) appears to decrease to zero as $\mu \to 0$ in a linear fashion. Furthermore, our tests show a larger radius of the sphere of convergence of Newton's method on system (4) than on system (3) even before μ becomes small. In the case of problem `r2`, the radius for system (3) is noticeably larger than that for system (4) only when μ is very large.

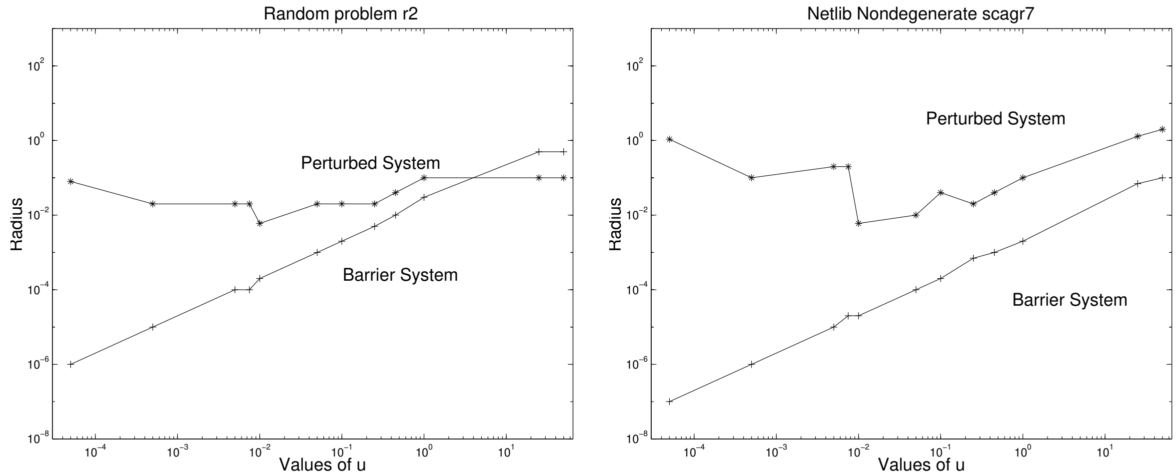

Figure 1: Radii of half-sphere and sphere of convergence, respectively, for Newton's method on systems (3) and (4)

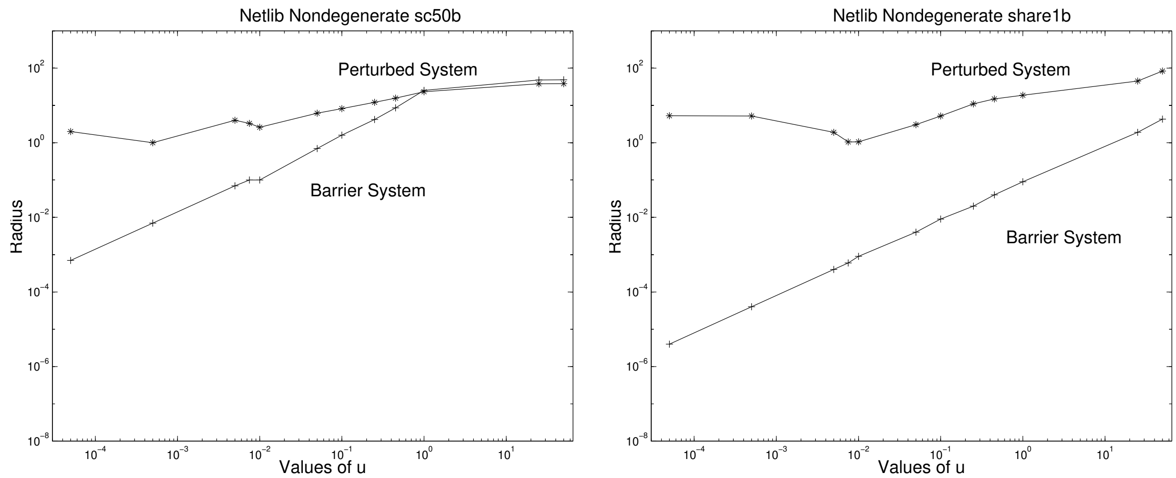

Figure 2: Radii of half-sphere and sphere of convergence, respectively, for Newton's method on systems (3) and (4)

4.4 Results for Degenerate Problems

For degenerate problems, we do not have a theory for the radius of the sphere of convergence of Newton's method on either of the two systems. We hope that numerical results would provide some empirical information on the behavior of Newton's method applied to these problems. We present numerical results on four of the five degenerate problems, as shown in Figures 3-4, omitting results for the problem `adlittle` because they are similar to the presented results.

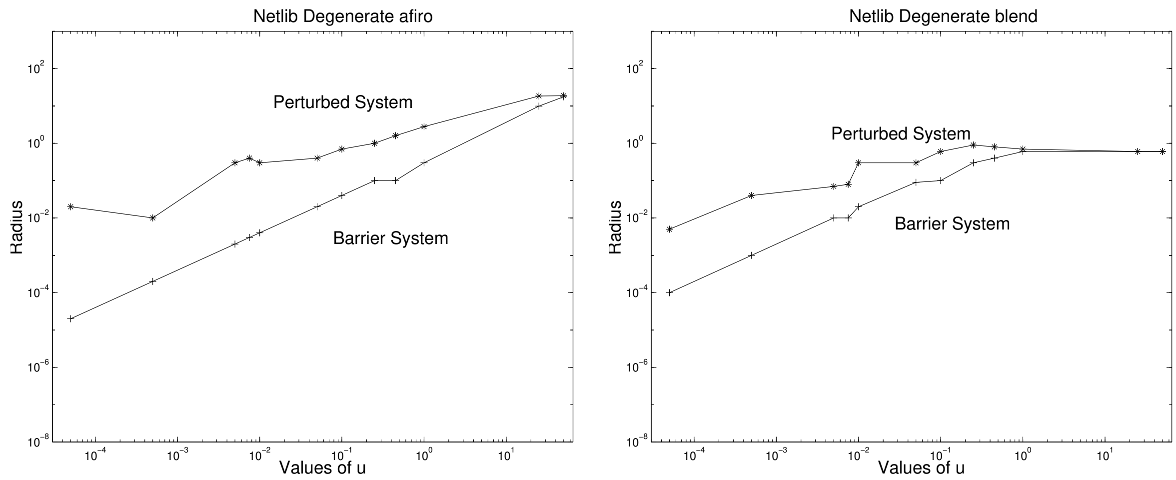

Figure 3: Radii of half-sphere and sphere of convergence, respectively, for Newton's method on systems (3) and (4)

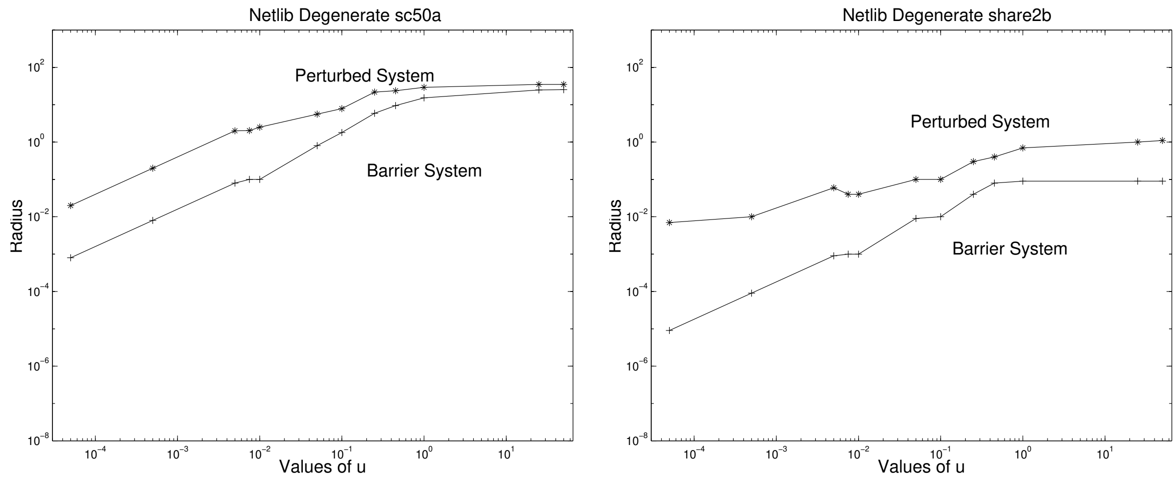

Figure 4: Radii of half-sphere and sphere of convergence, respectively, for Newton's method on systems (3) and (4)

The 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 observe that unlike the case of nondegenerate problems, the radius of the sphere of convergence on system (4) also appears to decrease to zero with μ.

In these tests, we observe that the radius of the sphere of convergence of system (4) is always larger than or equal to the radius of the sphere of convergence of system (3) for all the μ values given in (29). In particular, the radius associated with system (4) stays well above that for system

(3), by at least an order of magnitude, as $\mu \to 0$.

5 Conclusion

In this paper, we studied the local 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.

For nondegenerate problems, we have shown that the radius of the sphere of convergence of Newton's method on system (3) decreases to zero at exactly the same order as $\mu \to 0$, while the radius of the sphere of convergence associated with system (4) stays bounded away from zero as $\mu \to 0$. These theoretical results are established for exact arithmetics and hence are independent of the numerical conditioning of the Jacobian matrices for systems (3) and (4).

The numerical experiments have confirmed our theoretical results. Interestingly, on the majority of our test problems the estimated radius of 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 and large values of μ for which numerical ill-conditioning does not play a critical role.

There are multiple reasons why Newton's method performs more favorably on system (4) than on system (3) (see [11] for a recent work on this subject). Contrary to previous belief, M. Wright [10, 12] has shown that numerical ill-conditioning is not a determining factor. We believe that the results in this paper provide another fundamental reason why system (4) should be the system of choice to be used in an interior-point path-following framework. Similar results have been extended to the nonlinear program and will be reported in a subsequent paper.

Acknowledgments

We are grateful to two anonymous referees, as well as Richard Byrd, Amr El-Bakry and Masayuki Shida for their valuable comments and observations that have greatly helped improving 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] Lootsma, F., Hessian Matrices of Penalty Functions for Solving Constrained Optimization Problems. Philips Res. Repts 24, Eindhoven, The Netherlands, pp.322-331, 1969.

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

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

[9] Murray, W. Analytic Expressions for the Eigenvalues and Eigenvectors of Hessian Matrices of Barrier and Penalty Functions. J. Optim. Theory Appl. 7 (1971) 189-196.

[10] Wright, M.H., Some Properties of the Hessian in the Logarithmic Barrier Function, Mathematical Programming, 67 (1994) 265-295.

[11] Wright, M.H., Why a Pure Primal Newton Barrier Step May be Infeasible. SIAM Journal on Optimization, 5 (1995) 1-12.

[12] Wright, M.H., Ill-Conditioning and Computational Error in Primal-Dual Interior Methods for Nonlinear Programming. SIAM Journal on Optimization, 9 (1998) 84-111.

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

[14] 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.