

The Sphere of Convergence of Newton's Method on Two Equivalent Systems from Nonlinear Programming

María Cristina Villalobos* Richard A. Tapia[†] Yin Zhang[‡]

May 1999
(Revised September, 1999)

Abstract

We study a local feature of a Newton logarithmic barrier function method and a Newton primal-dual interior-point method for the nondegenerate inequality constrained optimization problem. In particular, we study the radius of the sphere of convergence of Newton's method applied to the nonlinear systems of equations associated with the two aforementioned interior-point methods. Our theoretical and numerical results are clearly in favor of using Newton primal-dual methods for solving the optimization problem. This work is an extension of the authors' earlier work [13] on linear programming problems.

Key words. Newton logarithmic barrier method; Newton primal-dual interior-point method; equivalent systems, sphere of convergence

1 Introduction

Linear and nonlinear programming problems are often solved by formulating and solving a sequence of parameterized systems of nonlinear equations associated with the optimization problem. Two popular approaches that have been a subject of study for solving the nonlinear program include the Newton logarithmic barrier (log-barrier) function method and the Newton primal-dual interior-point method. Each method applies Newton's method to a parameterized system of nonlinear equations and damps the step, if necessary, to produce strictly feasible points. The Newton log-barrier method can be viewed as applying Newton's method to the system of equations (*barrier system*) arising from the log-barrier formulation of the optimization problem. Similarly, the Newton primal-dual method can be viewed as applying Newton's method to the system of equations (*perturbed system*) obtained from the perturbed optimality conditions of the optimization problem.

*Department of Computational and Applied Mathematics, Rice University, Houston, TX, 77005. This author was supported in part by the Center for Research and Parallel Computations, DOE DE-FG03-93ER25178, and the Sloan Foundation. (cris@caam.rice.edu)

[†]Department of Computational and Applied Mathematics, Rice University, Houston, TX, 77005. This author was supported in part by DOE DE-FG05-86ER25017 and DOE/LANL Contract 03891-99-23. (rat@caam.rice.edu)

[‡]Department of Computational and Applied Mathematics, Rice University, Houston, TX, 77005. This author was supported in part by DOE Grant DE-FG03-97ER25331, DOE/LANL Contract 03891-99-23, and NSF Grant DMS-9973339. (zhang@caam.rice.edu)

Recent work, such as [9, 14, 15, 17], has focused on studying the performance of Newton log-barrier methods for the inequality constrained optimization problem. Since Newton primal-dual methods have been a success for solving the linear program, current investigation [3, 5, 11, 12] has focused on extending these methods to solve the nonlinear program.

Recently, Villalobos, Tapia, and Zhang [13] showed dissimilar behavior in the radius of the sphere of convergence associated with the Newton log-barrier and the Newton primal-dual methods for nondegenerate linear programming problems. Their results imply that fewer Newton iterations would be required by the Newton primal-dual method as the solution to the optimization problem is approached. This result provides an explanation for the success of Newton primal-dual methods in linear programming.

In this paper, we show distinct local behavior between the Newton log-barrier method and the Newton primal-dual method for nondegenerate inequality constrained optimization problems. Our study shows why Newton primal-dual methods should be applied to solve the nonlinear program. We extend our work in [13] and study the radius of the sphere of convergence of Newton's method applied to the barrier and perturbed systems for the nondegenerate nonlinear program. S. J. Wright [15] establishes a lower-bound, which is dependent on μ and has order greater than one, for the radius of the sphere of convergence associated with the barrier system. However, we prove a stronger result showing that the radius of the sphere of convergence associated with the barrier system decreases to zero with the same order as μ decreases to zero.

The paper is organized as follows. In Section 2, we present background material discussing the optimality conditions for the inequality constrained optimization problem. Then we briefly discuss the Newton log-barrier and the Newton primal-dual interior-point methods for the nonlinear program. Section 3 discusses the sphere of convergence of Newton's method. In Section 4 we present theoretical results for the radius of the sphere of convergence of Newton's method applied to the barrier and perturbed systems, and Section 5 presents numerical results supporting our theoretical findings. Finally, some concluding remarks are presented in Section 6.

2 Background

In this section, we present the inequality constrained optimization problem and then discuss two commonly studied interior-point methods. We conclude the section with a discussion on the trajectory of solutions produced by the two equivalent systems under consideration.

2.1 Optimality Conditions

We consider the following nonlinear inequality constrained optimization problem

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && g(x) \geq 0 \end{aligned} \tag{1}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are twice Lipschitz continuously differentiable. We will use $g_i(x)$ to denote the i th component of $g(x)$. The optimality conditions of problem (1) are formed by considering the Lagrangian function

$$L(x, z) = f(x) - \sum_{i=1}^m z_i g_i(x),$$

where the elements of $z \in \mathbb{R}^m$ are the Lagrangian multipliers associated with the inequality constraints.

Let x^* be a local solution of problem (1). Let $\mathcal{B} = \{i : g_i(x^*) = 0\}$, i.e. \mathcal{B} denotes the set of indices corresponding to the active constraints. Regularity holds at x^* if the set of active constraint gradients, that is, $\{\nabla g_i(x^*) : i \in \mathcal{B}\}$ is linearly independent. If f and g are differentiable and regularity holds at x^* , then the following Karush-Kuhn-Tucker (KKT) conditions are satisfied for x^* and a multiplier z^*

$$\nabla_x L(x^*, z^*) = 0, \quad Z^* g(x^*) = 0, \quad g(x^*) \geq 0, \quad z^* \geq 0 \quad (2)$$

where $Z^* = \text{diag}(z^*)$ and $\nabla_x L(x^*, z^*) = \nabla f(x^*) - \sum_{i=1}^m z_i^* \nabla g_i(x^*)$ is the gradient of the Lagrangian function. We collect the equations in (2) to form the system

$$F(x, z) \equiv \begin{bmatrix} \nabla f(x) - \sum_{i=1}^m z_i \nabla g_i(x) \\ Zg(x) \end{bmatrix} = 0. \quad (3)$$

Strict complementarity holds if for all i , $z_i^* + g_i(x^*) > 0$. By the second-order condition for problem (1), we mean that for $\eta \neq 0$ and $\nabla g_i(x^*)^T \eta = 0, i \in \mathcal{B}$

$$\eta^T \nabla_x^2 L(x^*, z^*) \eta > 0.$$

Then the sufficiency conditions are satisfied at x^* if the KKT conditions and the second-order condition hold. For further details on Lagrange multiplier theory, see Avriel [1].

2.2 Newton Logarithmic Barrier Function Method

We now discuss the Newton logarithmic barrier function method and the barrier system, which is associated with the optimality conditions of the logarithmic barrier function formulation for the nonlinear program.

The log-barrier framework was introduced implicitly by Frisch [6] and consists of solving problem (1) through a sequence of log-barrier subproblems for decreasing values of $\mu > 0$. For a given $\mu > 0$, the log-barrier subproblem is

$$\begin{aligned} \text{minimize} \quad \beta(x) &\equiv f(x) - \mu \sum_{i=1}^m \log g_i(x) \\ &(g(x) > 0) \end{aligned} \quad (4)$$

where $\beta(x)$ is the barrier function and μ is the barrier parameter. The first-order optimality conditions for problem (4) are

$$F_B(x; \mu) \equiv \nabla_x \beta(x) = 0, \quad (g(x) > 0) \quad (5)$$

where

$$\nabla_x \beta(x) = \nabla f(x) - \sum_{i=1}^m \frac{\mu}{g_i(x)} \nabla g_i(x) \quad (6)$$

is the gradient of the barrier function with respect to x . We will call the nonlinear system of equations in (5) the *barrier system* for problem (1). The Jacobian of system (5) is given by

$$F'_B(x; \mu) \equiv \nabla^2 f(x) - \sum_{i=1}^m \frac{\mu}{g_i(x)} \nabla^2 g_i(x) + \sum_{i=1}^m \frac{\mu}{g_i^2(x)} \nabla g_i(x) \nabla^T g_i(x).$$

Let x^* denote a local solution for the inequality constrained problem (1), and let x_μ^* denote a solution for system (5), whenever it exists, for $\mu > 0$. The Jacobian is positive definite at x_μ^* for small $\mu > 0$ when strict complementarity and the second-order sufficiency conditions hold at x^* . Therefore, x_μ^* is an unconstrained minimizer of the log-barrier subproblem (4). Under mild conditions [4], the solutions x_μ^* converge to x^* as μ decreases to zero.

A Newton log-barrier function method for the nonlinear program applies Newton's method to approximately solve the barrier system. In order to stay inside the feasible region of the inequality constraints, the Newton step is damped, if necessary. The procedure is continued for decreasing values of $\mu > 0$ until an approximate solution to the nonlinear program is obtained.

2.3 Newton Primal-Dual Interior-Point Method

We present the perturbed system which forms the basis of a Newton primal-dual interior-point method for the nonlinear program. The equivalence of the barrier and perturbed systems is described in Section 2.4.

Let us introduce an auxiliary variable $z \in \mathbb{R}^m$ and define z_i to be the term $\mu/g_i(x)$ that appears in the gradient (6) of the barrier function. Then $z_i = \mu/g_i(x)$ can be written equivalently as $Zg(x) = \mu e$ where $Z = \text{diag}(z)$. These equivalent defining relations yield the system

$$F_P(x, z; \mu) \equiv \begin{bmatrix} \nabla f(x) - \sum_{i=1}^m z_i \nabla g_i(x) \\ Zg(x) - \mu e \end{bmatrix} = 0 \quad (z, g(x) > 0). \quad (7)$$

The system $F_P(x, z; \mu)$ will be referred to as the *perturbed system* for problem (1). The Jacobian of $F_P(x, z; \mu)$ is given by

$$F'_P(x, z; \mu) \equiv \begin{bmatrix} \nabla_x^2 \ell(x, z) & -\nabla g(x) \\ Z \nabla^T g(x) & G(x) \end{bmatrix}$$

where $G(x) = \text{diag}(g(x))$, $\nabla g(x)$ is an $n \times m$ matrix whose i th column consists of the gradient of $g_i(x)$, and $\nabla_x^2 \ell(x, z)$ denotes the Hessian with respect to x of the Lagrangian function

$$\ell(x, z) = f(x) - \sum_{i=1}^m z_i g_i(x).$$

The Jacobian $F'_P(x^*, z^*; 0)$ is nonsingular if regularity, strict complementarity, and the second-order sufficiency conditions for problem (1) are satisfied. We note that system (7) can also be obtained by deriving the optimality conditions of problem (1) and perturbing the complementarity equation, $Zg(x) = 0$, by μe .

A Newton primal-dual interior-point method for the nonlinear program (1) consists of applying Newton's method to the perturbed system and damping the Newton step, if necessary, to strictly satisfy the inequality constraints, $z, g(x) \geq 0$. The procedure is continued with decreasing values of $\mu > 0$ until an approximate solution to the nonlinear program is obtained.

The algorithms for the Newton log-barrier method and the Newton primal-dual interior-point method are similar. The main distinction lies in the different nonlinear system that is associated with each method.

2.4 Trajectory of Solutions

Systems (5) and (7) are parameterized by $\mu > 0$. For a particular value of $\mu > 0$, let x_μ^* denote the solution to system (5), and similarly let (x_μ^*, z_μ^*) denote the solution to system (7). Under mild conditions, for μ small and positive, the Jacobians of $F_B(x; \mu)$ and $F_P(x, z; \mu)$ are nonsingular at the solutions x_μ^* and (x_μ^*, z_μ^*) , respectively. In addition, the solution (x^*, z^*) at $\mu = 0$ exists for system (7).

Assume f and g are twice continuously differentiable and x^* satisfies the sufficiency conditions and regularity. Then a trajectory of solutions exists for system (5) which is described by

$$\mathcal{C}_B = \{x_\mu^* : F_B(x_\mu^*; \mu) = 0, g(x_\mu^*) > 0, 0 < \mu \leq \hat{\mu}\} \quad (8)$$

for some $\hat{\mu} > 0$ (see [4]). In particular $\lim_{\mu \rightarrow 0} x_\mu^* = x^*$, where x^* is a solution of problem (1). McLinden [8] was the first to show the global existence and smoothness of the trajectory of solutions for a convex nonlinear program.

Systems (5) and (7) are equivalent, in the sense that for $0 < \mu \leq \hat{\mu}$ and $x_\mu^* \in \mathcal{C}_B$,

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

for $(z_\mu^*)_i = \mu/g(x_\mu^*)_i$, $i = 1, \dots, m$. For system (7) we also have $\lim_{\mu \rightarrow 0} (x_\mu^*, z_\mu^*) = (x^*, z^*)$. For the perturbed system the trajectory is defined as

$$\mathcal{C}_P = \{(x_\mu^*, z_\mu^*) : F_P(x_\mu^*, z_\mu^*; \mu) = 0, 0 \leq \mu \leq \hat{\mu}, z_\mu^*, g(x_\mu^*) > 0\}. \quad (9)$$

We refer the reader to the proofs given by Fiacco and McCormick [4] for further details on the existence and smoothness of the trajectories (8) and (9).

3 Sphere of Convergence

Under a set of standard assumptions, local convergence theory of Newton's method (see for example [2]) provides the existence of a neighborhood about a solution where Newton's method is well-defined. More importantly, starting from any point in the neighborhood, Newton's method converges to the solution of the nonlinear system. The local convergence analysis of Newton's method can be applied to the barrier and perturbed systems for a fixed value of μ . In this manner, we also obtain a neighborhood about each point on the trajectory, equivalently given by (8) and (9).

We begin now by providing a definition for the sphere of convergence below.

Definition The closed ball with radius r centered at v^* is defined as $B(v^*, r) = \{v : \|v - v^*\|_2 \leq r\}$.

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, other than v^* , Newton's method is well-defined and generates a sequence that converges to v^* .

We note that the Jacobian of $F(v)$ is allowed to be undefined or singular at v^* since we restrict Newton's method to be defined elsewhere in the neighborhood except at v^* .

We are primarily concerned with the behavior of the radii of the spheres of convergence of a pure Newton's method applied to the barrier and perturbed systems as μ decreases to zero. The asymptotic behavior of the radii will provide an explanation for the performance of the Newton log-barrier and the Newton primal-dual interior-point methods. It is our belief that if the radius for the sphere of convergence of Newton's method applied to the barrier or perturbed system is bounded away from zero as μ decreases to zero, then this implies that few iterations would be required by Newton's method to converge to a solution on the trajectory (8) or (9). However, if the radius decreases to zero, then more iterations would be required by a Newton interior-point method to converge to a solution on the trajectory and, in general, to obtain a solution to the nonlinear program.

4 Theory for the Radius of the Sphere of Convergence

We analyze the behavior of the radius of the sphere of convergence of Newton's method applied to the barrier and perturbed systems of the inequality constrained optimization problem (1) as μ decreases to zero. For μ sufficiently small, S. J. Wright [15] established a lower-bound of $C\mu^\alpha$ where $\alpha > 1$ and $C > 0$ for the radius of the sphere of convergence of Newton's method applied to the barrier system.

In Section 4.2, we provide a sharper result for the radius of the sphere of convergence of Newton's method on the barrier system showing that the radius is bounded below and above in $O(\mu)$. This result implies that the radius of the sphere of convergence associated with the barrier system decreases to zero in the same order as μ decreases to zero. Our proof uses the same assumptions on the functions f and g of problem (1) as S. J. Wright [15]. We also consider range and null subspace information of the active constraint gradients.

Finally in Section 4.3, we conduct a similar analysis on the radius of the sphere of convergence of Newton's method applied to the perturbed system as μ decreases to zero.

We note that in the proof of Lemma 4.1.4 an upper-bound $\hat{\mu}$ is shown to exist to prove the boundedness of the constraints $g(x)$ close to the solution. To simplify our notation and proofs, we will make extensive use of $\hat{\mu}$ and without loss of generality, we will assume the subsequent lemmas and theorems hold for $\mu \leq \hat{\mu}$.

4.1 Preliminaries

We present some preliminary results and lemmas for the analysis of the radius of the sphere of convergence of Newton's method applied to the barrier system. A nondegeneracy assumption is assumed to hold for the lemmas and theorems in the remainder of the section. We state some results on functions given by Lemmas 4.1.1 and 4.1.2. We will make much use of Lemma 4.1.3, which is taken from [2] and is presented here for completeness and used in the proof of Lemma 4.2.7 for the barrier system.

Recall that $\mathcal{B} = \{i : g_i(x^*) = 0\}$, i.e. \mathcal{B} is the set of indices of the active constraints at x^* . All norms, $\|\cdot\|$, are understood to be the Euclidean norm unless otherwise stated. Now, let $U_{\mathcal{B}}(x)$ denote an orthonormal matrix that spans the range space of the active constraint gradients at x . Let $U_{\mathcal{N}}(x)$ denote an orthonormal matrix that spans the left null space of the active constraint gradients at x . Then $[U_{\mathcal{B}}(x) \ U_{\mathcal{N}}(x)]$ is orthogonal.

Nondegeneracy Assumption At the solution x^* , regularity, strict complementarity, and the second-order sufficiency conditions are satisfied.

Lemma 4.1.1 *Given $c : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, where D is a compact set. If c is Lipschitz continuous over D with Lipschitz constant $M_1 > 0$, then cc^T is Lipschitz continuous over D with Lipschitz constant*

$$M = 2M_1 \tau_c$$

where $\tau_c = \max\{\|c(x)\| : x \in D\}$.

Proof Let $z, y \in D$. Then

$$\begin{aligned} \|c(z)c(z)^T - c(y)c(y)^T\| &= \|c(z)c(z)^T - c(z)c(y)^T + c(z)c(y)^T - c(y)c(y)^T\| \\ &= \|c(z)(c(z)^T - c(y)^T) + (c(z) - c(y))c(y)^T\| \\ &\leq \|c(z)\| \|c(z)^T - c(y)^T\| + \|c(z) - c(y)\| \|c(y)^T\|. \end{aligned} \quad (10)$$

Since c is bounded and is Lipschitz continuous with constant $M_1 > 0$, then from (10) we obtain

$$\|c(z)c(z)^T - c(y)c(y)^T\| \leq M \|z - y\|,$$

where $M = 2M_1 \tau_c$.

□

Lemma 4.1.2 Given $h : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^{m \times n}$ and $v : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$, where D is a compact and convex set. Assume h is Lipschitz continuous over D with Lipschitz constant $M_1 > 0$, and v is continuously differentiable with $v(x) \neq 0$ on D . Then for a positive integer p , $h(x)/v^p(x)$ is Lipschitz continuous in D with Lipschitz constant

$$M = \frac{M_1}{\kappa^p} + \frac{p \eta \tau \gamma^{p-1}}{\kappa^{2p}}$$

where $\eta = \max\{\|\nabla v(x)\| : x \in D\}$, $\tau = \max\{\|h(x)\| : x \in D\}$, $\gamma = \max\{|v(x)| : x \in D\}$, and $\kappa = \min\{|v(x)| : x \in D\}$.

Proof Let $z, y \in D$. Then

$$\begin{aligned} \left\| \frac{h(z)}{v^p(z)} - \frac{h(y)}{v^p(y)} \right\| &= \left\| \frac{v^p(y)h(z) - v^p(z)h(y)}{v^p(z)v^p(y)} \right\| \\ &= \left\| \frac{v^p(z)(h(z) - h(y)) - h(z)(v^p(z) - v^p(y))}{v^p(z)v^p(y)} \right\| \\ &\leq \frac{|v^p(z)| \|h(z) - h(y)\|}{|v^p(z)v^p(y)|} + \frac{\|h(z)\| |v^p(z) - v^p(y)|}{|v^p(z)v^p(y)|}. \end{aligned} \quad (11)$$

Using the Lipschitz continuity of h and the first-order Taylor Series of $v^p(z)$, we obtain from (11) that for ξ between z and y , and $M_1 > 0$

$$\left\| \frac{h(z)}{v^p(z)} - \frac{h(y)}{v^p(y)} \right\| \leq \frac{M_1 \|z - y\|}{|v^p(y)|} + \frac{\|h(z)\| |p v^{p-1}(\xi) \nabla v(\xi)^T (z - y)|}{|v^p(z)v^p(y)|}.$$

Since ∇v is bounded above, then $|\nabla v(\xi)^T (z - y)| \leq \eta \|z - y\|$, and since h and v are bounded, we obtain

$$\left\| \frac{h(z)}{v^p(z)} - \frac{h(y)}{v^p(y)} \right\| \leq M \|z - y\|$$

for constant $M = M_1/\kappa^p + p \eta \tau \gamma^{p-1}/\kappa^{2p}$.

□

Lemma 4.1.3 Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be continuously differentiable in the open convex set $D \subset \mathbb{R}^n$. Let F' be Lipschitz continuous at $x \in D$ under a vector norm and the induced matrix operator norm with Lipschitz constant γ . Then, for any $x + p \in D$,

$$\|F(x + p) - F(x) - F'(x)p\| \leq \frac{\gamma}{2} \|p\|^2.$$

The following lemma provides bounds on the values that the inequality constraints $g_i(x)$ can take for points x that lie close to the trajectory (9), which is parameterized by μ . Recall $C_P = \{(x_\mu^*, z_\mu^*) : F_P(x_\mu^*, z_\mu^*; \mu) = 0, z_\mu^* > 0, 0 \leq \mu \leq \hat{\mu}\}$ (see (9)).

Lemma 4.1.4 Consider $\mu > 0$ and (x_μ^*, z_μ^*) contained in \mathcal{C}_P . Assume $g : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ is continuously differentiable on the compact, convex set D . Then under the nondegeneracy assumption, there exist $\hat{\mu} > 0$ and constants $K, J_1, J_2, J_3 > 0$ so that for $\mu \leq \hat{\mu}$ and $\|x - x_\mu^*\| \leq K\mu$ where $x \in D$, we have

$$J_1\mu \leq g_i(x) \leq J_2\mu \quad i \in \mathcal{B}, \quad (12a)$$

$$g_i(x) \geq J_3 \quad i \notin \mathcal{B}. \quad (12b)$$

Proof For $(x_\mu^*, z_\mu^*) \in \mathcal{C}_P$, we have

$$g_i(x_\mu^*) (z_\mu^*)_i = \mu \quad i = 1, \dots, m. \quad (13)$$

Under our assumption, we have $(x_\mu^*, z_\mu^*) \rightarrow (x^*, z^*)$ as $\mu \rightarrow 0$ (see [4]). Because g is continuous,

$$g_i(x_\mu^*) \rightarrow g_i(x^*).$$

By the strict complementarity condition at x^* , there exist $\hat{\mu} > 0$ and $J > 0$ such that for $\mu \leq \hat{\mu}$

$$g_i(x_\mu^*) \geq J \quad i \notin \mathcal{B} \text{ and } (z_\mu^*)_i \geq J, \quad i \in \mathcal{B}. \quad (14)$$

Since $z_\mu^* \rightarrow z^*$, without loss of generality for $\mu \leq \hat{\mu}$, we have

$$(z_\mu^*)_i \leq 2z_i^*, \quad i \in \mathcal{B}. \quad (15)$$

Let

$$\begin{aligned} M &= \max\{\|\nabla g_i(x)\| : x \in D, i \notin \mathcal{B}\}, \\ \omega &= \max\{z_i^* : i \in \mathcal{B}\}, \\ K &\leq \frac{1}{2M\omega}, \text{ and} \\ \hat{\mu} &\leq \frac{J}{2MK}. \end{aligned} \quad (16)$$

Consider $\mu \leq \hat{\mu}$. Let x be such that $\|x - x_\mu^*\| \leq K\mu$. By Taylor's Theorem, there exists ξ_i between x and x_μ^* such that

$$g_i(x) = g_i(x_\mu^*) + \nabla g_i(\xi_i)^T (x - x_\mu^*). \quad (17)$$

If we apply the Cauchy-Schwarz inequality to (17) and the quantities in (15) and (16), we have for $i \in \mathcal{B}$

$$\begin{aligned}
g_i(x) &\geq g_i(x_\mu^*) - \|\nabla g_i(\xi_i)\| \|x - x_\mu^*\| \\
&\geq \frac{\mu}{2z_i^*} - MK\mu \\
&= J_1\mu
\end{aligned}$$

where $J_1 = (\frac{1}{2\omega} - MK) > 0$.

Now we establish the second part of the inequality in (12a). In a similar manner, we obtain from (17) that for, $i \in \mathcal{B}$

$$\begin{aligned}
g_i(x) &\leq g_i(x_\mu^*) + \|\nabla g_i(\xi_i)\| \|x - x_\mu^*\| \\
&\leq \mu/(z_\mu^*)_i + MK\mu.
\end{aligned} \tag{18}$$

Applying (14) to (18), we have that

$$g_i(x) \leq J_2\mu$$

for $J_2 = (1/J + MK) > 0$.

Now consider $i \notin \mathcal{B}$. By (14) and (17), we have

$$\begin{aligned}
g_i(x) &\geq J + \nabla g_i(\xi_i)^T(x - x_\mu^*) \\
&\geq J - \|\nabla g_i(\xi_i)\| \|x - x_\mu^*\| \\
&\geq J - MK\mu \\
&\geq J - MK\hat{\mu} \\
&\geq J/2,
\end{aligned}$$

where the last inequality arises from the upper-bound on $\hat{\mu}$ in (16). Then, we have for $i \in \mathcal{B}$ that $g_i(x) \geq J_3$ for $J_3 = J/2 > 0$.

□

4.2 Barrier System

In this section, we prove additional results required for the analysis of the radius of the sphere of convergence of Newton's method applied to the barrier system. Then we derive lower- and upper-bound results for the radius of the sphere of convergence in Lemmas 4.2.8 and 4.2.9, respectively.

S. J. Wright [15] provides a representation for $F'_B(x; \mu)^{-1}$ considering that μ is sufficiently small and that x is close to x_μ^* . We have provided additional information on an upper-bound for μ , given in Lemma 4.1.4, to achieve a similar lemma. The proof for Lemma 4.2.5 below is omitted, and the reader is referred to S. J. Wright's proof [15] for details.

Lemma 4.2.5 *Let the nondegeneracy assumption hold. Then for $\mu \leq \hat{\mu}$ and $\|x - x_\mu^*\| \leq K\mu$, where $\hat{\mu}$ and K are given in Lemma 4.1.4, $F'_B(x; \mu)^{-1}$ can be written in the following manner*

$$F'_B(x; \mu)^{-1} = \begin{bmatrix} U_{\mathcal{B}}(x) & U_{\mathcal{N}}(x) \end{bmatrix} \begin{bmatrix} H_{11}(x; \mu) & H_{12}(x; \mu) \\ H_{12}^T(x; \mu) & H_{22}(x; \mu) \end{bmatrix} \begin{bmatrix} U_{\mathcal{B}}^T(x) \\ U_{\mathcal{N}}^T(x) \end{bmatrix} \tag{19}$$

where $H_{11}(x; \mu) = O(\mu)$, $H_{12}(x; \mu) = O(\mu)$, and $H_{22}(x; \mu) = O(1)$.

S. J. Wright and Jarre [17] use the implicit function theorem to describe a continuous trajectory of solutions for the system given in (3) with the right-hand side modified, that is,

$$F(x, z; \lambda, \xi) = \begin{bmatrix} \lambda \\ \xi e \end{bmatrix}.$$

where $\lambda \in \mathbb{R}^n$ and $\xi \in \mathbb{R}$. We make use of their lemma, modifying it with specific choices of λ and ξ to suit our purposes. For a proof of Lemma 4.2.6, we refer the reader to S. J. Wright and Jarre [17].

Lemma 4.2.6 *Under the nondegeneracy assumption, let the vector pair $(\tilde{x}(\mu), \tilde{z}(\mu))$ denote the solution of the nonlinear system*

$$F(\tilde{x}, \tilde{z}) = \begin{bmatrix} 0 \\ \mu e \end{bmatrix}$$

for given $\mu \geq 0$ and F defined as in (3). Then there are positive constants $\hat{\mu} > 0$ and $M > 0$ such that

(i) $(\tilde{x}(\mu), \tilde{z}(\mu))$ is a C^2 function of μ in the neighborhood defined by

$$\mathcal{N}_{\hat{\mu}} = \{\mu : 0 \leq \mu \leq \hat{\mu}\}.$$

(ii) For $\mu_1, \mu_2 \in \mathcal{N}_{\hat{\mu}}$, we have

$$\begin{bmatrix} \tilde{x}(\mu_1) \\ \tilde{z}(\mu_1) \end{bmatrix} - \begin{bmatrix} \tilde{x}(\mu_2) \\ \tilde{z}(\mu_2) \end{bmatrix} = F'(\tilde{x}(\mu_1), \tilde{z}(\mu_1))^{-1} \begin{bmatrix} 0 \\ (\mu_1 - \mu_2)e \end{bmatrix} + r, \quad (20)$$

where

$$\|r\| \leq M(\mu_1 - \mu_2)^2.$$

Since $F'(\tilde{x}, \tilde{z})$ is continuous and nonsingular close to (x^*, z^*) , then for $\hat{\mu}$ small, we obtain

$$\left\| \begin{pmatrix} \tilde{x}(\mu_1) \\ \tilde{z}(\mu_1) \end{pmatrix} - \begin{pmatrix} \tilde{x}(\mu_2) \\ \tilde{z}(\mu_2) \end{pmatrix} \right\| = O(|\mu_1 - \mu_2|).$$

Therefore, if we let $\mu_1 = \mu$ and $\mu_2 = 0$, then

$$(\tilde{x}(\mu), \tilde{z}(\mu)) = (x_\mu^*, z_\mu^*) \quad \text{and} \quad (\tilde{x}(\mu_2), \tilde{z}(\mu_2)) = (x^*, z^*).$$

It follows that for some constant $M > 0$

$$\left\| \begin{pmatrix} x_\mu^* - x^* \\ z_\mu^* - z^* \end{pmatrix} \right\| \leq M\mu. \quad (21)$$

The following notation and quantities will be used in the remaining lemmas. The next result, Lemma 4.2.7, will be needed in the proof of Lemma 4.2.8 to obtain a lower-bound for the radius of the sphere of convergence of Newton's method on the barrier system.

For our analysis on the radius of the sphere of convergence, we consider

$$x_+ - x_\mu^* = F'_B(x; \mu)^{-1} [F_B(x_\mu^*; \mu) - F_B(x; \mu) - F'_B(x; \mu)(x_\mu^* - x)], \quad (22)$$

where x is the initial Newton iterate, x_+ is the subsequent iterate, and x_μ^* is the solution to the barrier system. We obtain upper-bounds on the quantities in the right-hand side of (22). So define R as the vector on the right-hand side of (22), that is,

$$\begin{aligned} R(x; \mu) &\equiv F_B(x_\mu^*; \mu) - F_B(x; \mu) - F'_B(x; \mu)(x_\mu^* - x) \\ &\equiv R_1(x; \mu) + R_2(x; \mu) + R_3(x; \mu) \end{aligned} \quad (23)$$

where

$$\begin{aligned} R_1(x; \mu) &= \nabla f(x_\mu^*) - \nabla f(x) - \nabla^2 f(x)(x_\mu^* - x), \\ R_2(x; \mu) &= \sum_{i \notin \mathcal{B}} \left(-\frac{\mu}{g_i(x_\mu^*)} \nabla g_i(x_\mu^*) + \frac{\mu}{g_i(x)} \nabla g_i(x) \right. \\ &\quad \left. - \frac{\mu}{g_i^2(x)} \nabla g_i(x) \nabla g_i(x)^T (x_\mu^* - x) + \frac{\mu}{g_i(x)} \nabla^2 g_i(x)(x_\mu^* - x) \right), \\ R_3(x; \mu) &= \sum_{i \in \mathcal{B}} \left(-\frac{\mu}{g_i(x_\mu^*)} \nabla g_i(x_\mu^*) + \frac{\mu}{g_i(x)} \nabla g_i(x) \right. \\ &\quad \left. - \frac{\mu}{g_i^2(x)} \nabla g_i(x) \nabla g_i(x)^T (x_\mu^* - x) + \frac{\mu}{g_i(x)} \nabla^2 g_i(x)(x_\mu^* - x) \right). \end{aligned} \quad (24)$$

Let $\hat{\mu}$ and K be given as in the proof of Lemma 4.1.4. For all $0 < \mu \leq \hat{\mu}$ and for some constant $\rho > 0$, let

$$\|x - x_\mu^*\| \leq K\mu \subset B(x^*; \rho).$$

We define the following terms for $i = 1, \dots, m$

$$\begin{aligned} \tau_i &= \max\{\|\nabla g_i(x) \nabla g_i(x)^T\| : x \in B(x^*; \rho)\}, \\ \alpha_i &= \max\{\|\nabla^2 g_i(x)\| : x \in B(x^*; \rho)\}, \\ \beta_i &= \max\{\|\nabla g_i(x)\| : x \in B(x^*; \rho)\}, \\ \gamma_i &= \max\{g_i(x) : x \in B(x^*; \rho)\}, \text{ and} \\ \kappa_i &= \min\{g_i(x) : \|x - x_\mu^*\| \leq K\mu\}. \end{aligned} \quad (25)$$

Lemma 4.2.7 *Assume f and g are twice Lipschitz continuously differentiable over $B(x^*; \rho)$. Let R be given as defined in (23) with domain $B(x^*; \rho)$. Assume $\hat{\mu}$ and K are given as in the proof of Lemma 4.1.4. Then under the nondegeneracy assumption, for $\mu \leq \hat{\mu}$ and constants $C_1, C_2 > 0$, it follows that for $\|x - x_\mu^*\| \leq K\mu$,*

$$\|U_B^T(x)R(x; \mu)\| \leq C_1 \frac{1}{\mu^2} \|x - x_\mu^*\|^2 \text{ and} \quad (26)$$

$$\|U_{\mathcal{N}}^T(x)R(x; \mu)\| \leq C_2 \frac{1}{\mu} \|x - x_\mu^*\|^2. \quad (27)$$

Proof Consider $\mu \leq \hat{\mu}$ and $\|x - x_\mu^*\| \leq K\mu$. Let the functions R_1, R_2 , and R_3 defined in (24) and $U_{\mathcal{N}}^T$ and $U_{\mathcal{B}}^T$ be evaluated at x for a given parameter $\mu \leq \hat{\mu}$. Multiplying $U_{\mathcal{B}}^T(x)$ with R , yields

$$\|U_{\mathcal{B}}^T(x)R(x; \mu)\| \leq \|U_{\mathcal{B}}^T\| [\|R_1\| + \|R_2\| + \|R_3\|]. \quad (28)$$

We will obtain an upper-bound for $\|U_{\mathcal{B}}^T R\|$ by obtaining upper-bounds for each of the four terms in (28).

Since $\|U_{\mathcal{B}}^T\|$ is independent of μ , we have $\|U_{\mathcal{B}}^T\| \leq C$ for a constant $C > 0$. Now for R_1 we obtain

$$\|R_1\| = \|\nabla f(x_\mu^*) - \nabla f(x) - \nabla^2 f(x)(x_\mu^* - x)\|.$$

Applying Lemma 4.1.3 to the above term yields

$$\|R_1\| \leq \frac{D_1}{2} \|x_\mu^* - x\|^2 \quad (29)$$

where $D_1 > 0$ is the Lipschitz constant of $\nabla^2 f$.

To determine upper-bounds for the terms, $\|R_2\|$ and $\|R_3\|$, we will proceed as in the proof of Lemma 4.1.3 to obtain the Lipschitz constants for $\nabla g \nabla g^T / g^2$ and $\nabla^2 g / g$ that depend on μ . Let $x_t = x + tp$ for $t \in [0, 1]$ and let p be the direction $(x_\mu^* - x)$. Applying Lemma 4.1.3 to R_2 , we obtain

$$\begin{aligned} \|R_2\| &\leq \sum_{i \notin \mathcal{B}} \int_0^1 \left\| \frac{\mu}{g_i(x_t)^2} \nabla g_i(x_t) \nabla g_i(x_t)^T - \frac{\mu}{g_i(x_t)} \nabla^2 g_i(x_t) \right. \\ &\quad \left. - \frac{\mu}{g_i(x)^2} \nabla g_i(x) \nabla g_i(x)^T + \frac{\mu}{g_i(x)} \nabla^2 g_i(x) \right\| \|x_\mu^* - x\| dt \\ &\leq \sum_{i \notin \mathcal{B}} \int_0^1 \left\| \frac{\mu}{g_i(x_t)^2} \nabla g_i(x_t) \nabla g_i(x_t)^T - \frac{\mu}{g_i(x)^2} \nabla g_i(x) \nabla g_i(x)^T \right\| \|x_\mu^* - x\| dt \\ &\quad + \sum_{i \notin \mathcal{B}} \int_0^1 \left\| \frac{\mu}{g_i(x_t)} \nabla^2 g_i(x_t) - \frac{\mu}{g_i(x)} \nabla^2 g_i(x) \right\| \|x_\mu^* - x\| dt. \end{aligned} \quad (30)$$

Since $\nabla g \nabla g^T$ is Lipschitz continuous by Lemma 4.1.1, then applying Lemma 4.1.2 to (30) yields

$$\begin{aligned} \|R_2\| &\leq \sum_{i \notin \mathcal{B}} \int_0^1 \left[\frac{\mu M_i}{\kappa_i^2} + \frac{2\mu \eta_i \tau_i \gamma_i}{\kappa_i^4} \right] \|x_t - x\| \|x_\mu^* - x\| dt \\ &\quad + \sum_{i \notin \mathcal{B}} \int_0^1 \left[\frac{\mu F_i}{\kappa_i} + \frac{\mu \rho_i \alpha_i}{\kappa_i^2} \right] \|x_t - x\| \|x_\mu^* - x\| dt \end{aligned}$$

where $\eta_i, \rho_i > 0$ and $\tau_i, \alpha_i, \gamma_i$, and κ_i are given in (25), and M_i and F_i are, respectively, the Lipschitz constants for $\nabla g_i \nabla g_i^T$ and $\nabla^2 g_i$. By (12), $J_3 \leq \kappa_i$ for $i \notin \mathcal{B}$ and we obtain

$$\begin{aligned} \|R_2\| &\leq \sum_{i \notin \mathcal{B}} \int_0^1 \left[\frac{\mu M_i}{J_3^2} + \frac{2\mu \eta_i \tau_i \gamma_i}{J_3^4} \right] \|x_t - x\| \|x_\mu^* - x\| dt \\ &\quad + \sum_{i \notin \mathcal{B}} \int_0^1 \left[\frac{\mu F_i}{J_3} + \frac{\mu \rho_i \alpha_i}{J_3^2} \right] \|x_t - x\| \|x_\mu^* - x\| dt. \end{aligned} \quad (31)$$

From the definition of x_t and p , we obtain that $\|x_t - x\| = t\|x_\mu^* - x\|$. Making the substitution of $\|x_t - x\|$ in (31) and integrating with respect to t , we obtain

$$\begin{aligned} \|R_2\| &\leq \sum_{i \notin \mathcal{B}} \left[\frac{\mu M_i}{2J_3^2} + \frac{2\mu\eta_i\tau_i\gamma_i}{2J_3^4} \right] \|x_\mu^* - x\|^2 + \left[\frac{\mu F_i}{2J_3} + \frac{\mu\rho_i\alpha_i}{2J_3^2} \right] \|x_\mu^* - x\|^2 \\ &\leq D_2\mu \|x_\mu^* - x\|^2 \end{aligned} \quad (32)$$

for some constant $D_2 > 0$.

Similarly, we obtain for R_3 that

$$\begin{aligned} \|R_3\| &\leq \sum_{i \in \mathcal{B}} \int_0^1 \left[\frac{\mu M_i}{\kappa_i^2} + \frac{2\mu\eta_i\tau_i\gamma_i}{\kappa_i^4} \right] \|x_t - x\| \|x_\mu^* - x\| dt \\ &\quad + \sum_{i \in \mathcal{B}} \int_0^1 \left[\frac{\mu F_i}{\kappa_i} + \frac{\mu\rho_i\alpha_i}{\kappa_i^2} \right] \|x_t - x\| \|x_\mu^* - x\| dt. \end{aligned} \quad (33)$$

Now, by (12a), $J_1\mu \leq \kappa_i$ and $\gamma_i \leq J_2\mu$ for $i \in \mathcal{B}$. Making these substitutions in (33) and integrating we obtain

$$\|R_3\| \leq \sum_{i \in \mathcal{B}} \left[\frac{\mu M_i}{2J_1^2\mu^2} + \frac{2\mu\eta_i\tau_i J_2\mu}{2J_1^4\mu^4} \right] \|x_\mu^* - x\|^2 + \left[\frac{\mu F_i}{2J_1\mu} + \frac{\mu\rho_i\alpha_i}{2J_1^2\mu^2} \right] \|x_\mu^* - x\|^2.$$

Therefore,

$$\|R_3\| \leq D_3 \frac{1}{\mu^2} \|x_\mu^* - x\|^2$$

for some constant $D_3 > 0$. Collecting the upper-bounds obtained on the terms, $\|U_{\mathcal{B}}^T\|$, $\|R_1\|$, $\|R_2\|$, and $\|R_3\|$, we obtain

$$\|U_{\mathcal{B}}^T R(x; \mu)\| \leq C_1 \frac{1}{\mu^2} \|x_\mu^* - x\|^2$$

for a constant $C_1 > 0$.

For the second part of the proof, we work with $U_{\mathcal{N}}$ and again will determine an upper-bound for $\|U_{\mathcal{N}}^T R\|$. We have

$$\begin{aligned} \|U_{\mathcal{N}}^T R\| &\leq \|U_{\mathcal{N}}^T R_1\| + \|U_{\mathcal{N}}^T R_2\| + \|U_{\mathcal{N}}^T R_3\| \\ &\leq \|U_{\mathcal{N}}^T\| [\|R_1\| + \|R_2\|] + \|U_{\mathcal{N}}^T R_3\|. \end{aligned}$$

Since $\|U_{\mathcal{N}}^T\|$ is independent of μ , we have $\|U_{\mathcal{N}}^T\| \leq G$ for some constant $G > 0$. We can use the upper-bound estimates obtained in the first part of the proof for $\|R_1\|$ and $\|R_2\|$. However, the situation changes for estimating $\|U_{\mathcal{N}}^T R_3\|$, the term associated with the active constraints. Following the proof as in Lemma 4.1.3, we have

$$\begin{aligned} U_{\mathcal{N}}^T R_3 &= \sum_{i \in \mathcal{B}} \left[\int_0^1 \frac{\mu}{g_i(x_t)^2} U_{\mathcal{N}}^T(x) \nabla g_i(x_t) \nabla^T g_i(x_t) - \frac{\mu}{g_i(x_t)} U_{\mathcal{N}}^T(x) \nabla^2 g_i(x_t) \right. \\ &\quad \left. - \frac{\mu}{g_i(x)^2} U_{\mathcal{N}}^T(x) \nabla g_i(x) \nabla g_i^T(x) + \frac{\mu}{g_i(x)} U_{\mathcal{N}}^T(x) \nabla^2 g_i(x) \right] (x - x_\mu^*) dt. \end{aligned}$$

Then

$$\begin{aligned}
\|U_{\mathcal{N}}^T R_3\| &\leq \sum_{i \in \mathcal{B}} \int_0^1 \left\| \frac{\mu}{g_i(x_t)^2} U_{\mathcal{N}}^T(x) \nabla g_i(x_t) \nabla^T g_i(x_t) - \frac{\mu}{g_i(x_t)} U_{\mathcal{N}}^T(x) \nabla^2 g_i(x_t) \right. \\
&\quad \left. - \frac{\mu}{g_i(x)^2} U_{\mathcal{N}}^T(x) \nabla g_i(x) \nabla g_i^T(x) + \frac{\mu}{g_i(x)} U_{\mathcal{N}}^T(x) \nabla^2 g_i(x) \right\| \|x - x_\mu^*\| dt \\
&\leq \sum_{i \in \mathcal{B}} (T_{1i} + T_{2i})
\end{aligned} \tag{34}$$

where

$$T_{1i} = \int_0^1 \left\| \frac{\mu}{g_i(x_t)^2} U_{\mathcal{N}}^T(x) \nabla g_i(x_t) \nabla g_i(x_t)^T - \frac{\mu}{g_i(x)^2} U_{\mathcal{N}}^T(x) \nabla g_i(x) \nabla g_i(x)^T \right\| \|x - x_\mu^*\| dt$$

and

$$T_{2i} = \int_0^1 \left\| \frac{\mu}{g_i(x_t)} U_{\mathcal{N}}^T(x) \nabla^2 g_i(x_t) - \frac{\mu}{g_i(x)} U_{\mathcal{N}}^T(x) \nabla^2 g_i(x) \right\| \|x - x_\mu^*\| dt.$$

We will work with T_{1i} and T_{2i} individually and obtain upper-bounds for each. Since $U_{\mathcal{N}}(x)$ is orthogonal to the set of active constraint gradients at x , we obtain for T_{1i} that

$$\begin{aligned}
T_{1i} &= \int_0^1 \left\| \frac{\mu}{g_i(x_t)^2} U_{\mathcal{N}}^T(x) \nabla g_i(x_t) \nabla g_i(x_t)^T \right\| \|x - x_\mu^*\| dt \\
&= \int_0^1 \left\| \frac{\mu}{g_i(x_t)^2} U_{\mathcal{N}}^T(x) [\nabla g_i(x_t) \nabla g_i(x_t)^T - \nabla g_i(x) \nabla g_i(x)^T] \right\| \|x - x_\mu^*\| dt.
\end{aligned}$$

By Lemma 4.1.4, $g_i(x_t) \geq J_1 \mu$. Hence, we obtain

$$T_{1i} \leq \frac{\mu}{J_1 \mu^2} \|U_{\mathcal{N}}^T(x)\| \int_0^1 \|\nabla g_i(x_t) \nabla g_i(x_t)^T - \nabla g_i(x) \nabla g_i(x)^T\| \|x - x_\mu^*\| dt.$$

By Lemma 4.1.1, $\nabla g_i \nabla g_i^T$ is Lipschitz continuous with Lipschitz constant $2S\beta_i$ where $S > 0$ is the Lipschitz constant of ∇g_i . Then

$$\begin{aligned}
T_{1i} &\leq \frac{1}{J_1 \mu} \|U_{\mathcal{N}}^T(x)\| \int_0^1 2S\beta_i \|x_t - x\| \|x - x_\mu^*\| dt \\
&\leq \frac{1}{J_1 \mu} S\beta_i \|U_{\mathcal{N}}^T(x)\| \|x - x_\mu^*\|^2 \\
&\leq \frac{G_1}{\mu} \|x - x_\mu^*\|^2
\end{aligned} \tag{35}$$

for a constant $G_1 > 0$.

Now, we work with the second term T_{2i} which is bounded by

$$T_{2i} \leq \|U_{\mathcal{N}}^T\| \int_0^1 \left\| \frac{\mu}{g_i(x_t)} \nabla^2 g_i(x_t) - \frac{\mu}{g_i(x)} \nabla^2 g_i(x) \right\| \|x - x_\mu^*\| dt. \quad (36)$$

Applying Lemma 4.1.2 to the second norm in (36) and the bounds (12) given in Lemma 4.1.4, we obtain

$$\begin{aligned} T_{2i} &\leq \|U_{\mathcal{N}}^T\| \int_0^1 \left[\frac{\mu F_i}{J_1 \mu} + \frac{\mu \rho_i \alpha_i}{J_1^2 \mu^2} \right] \|x_t - x\| \|x - x_\mu^*\| dt \\ &\leq \|U_{\mathcal{N}}^T\| \left[\frac{\mu F_i}{2J_1 \mu} + \frac{\mu \rho_i \alpha_i}{2J_1^2 \mu^2} \right] \|x - x_\mu^*\|^2 \\ &\leq \frac{G_2}{\mu} \|x - x_\mu^*\|^2 \end{aligned} \quad (37)$$

where $\rho_i, G_2 > 0$ and F_i is the Lipschitz constant for $\nabla^2 g_i$.

Collecting the upper-bounds (35) and (37) for T_{1i} and T_{2i} , respectively, we obtain that for some $G_3 > 0$

$$\|U_{\mathcal{N}}^T R_3\| \leq \frac{G_3}{\mu} \|x - x_\mu^*\|^2. \quad (38)$$

Using the upper-bounds on $\|R_1\|$ and $\|R_2\|$ given in (29) and (32) and the upper-bound for $\|U_{\mathcal{N}}^T R_3\|$ in (38), we obtain

$$\begin{aligned} \|U_{\mathcal{N}}^T R(x; \mu)\| &\leq \|U_{\mathcal{N}}^T\| \left[\frac{D_1}{2} \|x_\mu^* - x\|^2 + D_2 \mu \|x_\mu^* - x\|^2 \right] + \frac{G_3}{\mu} \|x - x_\mu^*\|^2 \\ &\leq C_2 \frac{1}{\mu} \|x - x_\mu^*\|^2 \end{aligned}$$

for some constant $C_2 > 0$.

□

The next two lemmas establish lower- and upper-bound results for the radius of the sphere of convergence of Newton's method applied to the barrier system. The two results are combined to produce Theorem 4.1, which proves that the radius of the sphere of convergence of Newton's method on the barrier system decreases to zero linearly as μ decreases to zero. Our result is stronger than S. J. Wright's result [15] which shows that the radius of the sphere of convergence is bounded below by $C\mu^\alpha$, where $C > 0$ and $\alpha > 1$.

Lemma 4.2.8 *Under the nondegeneracy assumption, there exists $\hat{\mu} > 0$ and $K_1 > 0$ such that the radius of the sphere of convergence, $r_B(\mu)$, of Newton's method applied to the barrier system $F_B(x; \mu) = 0$, given by (5), satisfies*

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

for all $\mu \leq \hat{\mu}$.

Proof Let $\hat{\mu}$ and K be given by the proof of Lemma 4.1.4. Consider $\mu \leq \hat{\mu}$ and let x satisfy $\|x - x_\mu^*\| \leq K\mu$. We will prove the above result by showing that whenever the initial Newton iterate x^0 satisfies

$$\|x^0 - x_\mu^*\| \leq K_1\mu$$

for $K_1 > 0$ and $\mu \leq \hat{\mu}$, then Newton's method will converge to x_μ^* . Consider Newton's method applied to $F_B(x; \mu) = 0$. Let x and x_μ^* respectively denote the current iterate and the solution to the system. Then at the following Newton iteration we obtain

$$\begin{aligned} x_+ - x_\mu^* &= F'_B(x; \mu)^{-1} [F_B(x_\mu^*; \mu) - F_B(x; \mu) - F'_B(x; \mu)(x_\mu^* - x)] \\ &= F'_B(x; \mu)^{-1} R(x; \mu). \end{aligned}$$

where x_+ is the subsequent Newton iterate, and R is defined in (23). Substituting (19) for the inverse of the Jacobian and multiplying it with R produces

$$x_+ - x_\mu^* = \begin{bmatrix} U_B(x) & U_N(x) \end{bmatrix} \begin{bmatrix} O(\mu) & O(\mu) \\ O(\mu) & O(1) \end{bmatrix} \begin{bmatrix} U_B^T(x)R(x; \mu) \\ U_N^T(x)R(x; \mu) \end{bmatrix},$$

and taking norms yields

$$\|x_+ - x_\mu^*\| \leq \|O(\mu)U_B^T R + O(\mu)U_N^T R\| + \|O(\mu)U_B^T R + O(1)U_N^T R\|.$$

Substituting upper-bounds (26) and (27) for $\|U_B^T R\|$ and $\|U_N^T R\|$, respectively, we obtain

$$\begin{aligned} \|x_+ - x_\mu^*\| &\leq O(\mu)C_1 \frac{1}{\mu^2} \|x_\mu^* - x\|^2 + O(\mu)C_2 \frac{1}{\mu} \|x_\mu^* - x\|^2 + GC_2 \frac{1}{\mu} \|x_\mu^* - x\|^2 \\ &\leq W \frac{1}{\mu} \|x_\mu^* - x\|^2 \end{aligned}$$

for constants $G, W > 0$. Therefore, if the initial iterate x^0 satisfies

$$\|x^0 - x_\mu^*\| < K_1\mu,$$

where $K_1 = \min\{1/W, K\}$, we obtain convergence of the Newton sequence to x_μ^* . □

Lemma 4.2.9 *Under the nondegeneracy assumption, we have that for $\mu \leq \hat{\mu}$, there exists $K_2 > 0$ such that the radius of the sphere of convergence, $r_B(\mu)$, of Newton's method applied to the barrier system $F_B(x; \mu)$, given by (5), satisfies*

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

Proof It suffices to show that there exists an x satisfying $\|x - x_\mu^*\| \leq K\mu$, where K is given in Lemma 4.1.4, from where Newton's method does not converge or is not defined. From S. J. Wright's result (21), we have that $\|x^* - x_\mu^*\| \leq M\mu$ for small μ and $M > 0$. We will demonstrate two cases regarding the relationship of K and M to prove our result. If $M < K$, then $\|x^* - x_\mu^*\| \leq K\mu$ and $g_i(x^*) = 0$ for $i \in \mathcal{B}$. Thus, we can take $K_2 \equiv M$. Now assume $K < M$. Choose $\tilde{\mu} \leq \hat{\mu}$ such that $\|x^* - x_\mu^*\| \leq K\tilde{\mu}$. Then for all $\mu \leq \tilde{\mu}$, we have $\|x^* - x_\mu^*\| \leq K\mu$ and $g_i(x^*) = 0, i \in \mathcal{B}$. In this case, we let $K_2 \equiv K$. Clearly, $F_B(x; \mu)$ and $F'_B(x; \mu)$ are undefined whenever $g_i(x) = 0$ for any i .

□

Theorem 4.1 *Under the nondegeneracy assumption, there exist constants $\tilde{\mu}$ and $K_1, K_2 > 0$ such that for $\mu \leq \tilde{\mu}$, the radius of the sphere of convergence, $r_B(\mu)$, of Newton's method applied to the barrier system $F_B(x; \mu) = 0$, given by (5), satisfies*

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

Proof Application of Lemmas 4.2.8 and 4.2.9 produces the result with $\tilde{\mu} \leq \hat{\mu}$.

□

4.3 Perturbed System

Now we investigate the behavior of the radius of the sphere of convergence of Newton's method applied to the perturbed system as μ decreases to zero.

Lemma 4.2.6 provides the existence of $F'_P(x_\mu^*, z_\mu^*; \mu)^{-1}$ for $0 \leq \mu \leq \hat{\mu}$. Since (x_μ^*, y_μ^*) exists, and $F'_P(x, z; \mu)$ is Lipschitz continuous and nonsingular at (x_μ^*, z_μ^*) the standard assumptions for Newton's method are satisfied at (x_μ^*, z_μ^*) . Hence, there exists a ball centered at (x_μ^*, z_μ^*) such that starting from any point in the ball Newton's method will converge to the solution (x_μ^*, z_μ^*) . The following theorem states a new result showing that the radius of the sphere of convergence of Newton's method applied to the perturbed system is bounded away from zero as μ decreases to zero.

Theorem 4.2 *Under the nondegeneracy assumption, there exist constants $\hat{\mu}, D > 0$ such that for all $0 \leq \mu \leq \hat{\mu}$, the radius of the sphere of convergence, $r_P(\mu)$, of Newton's method applied to the perturbed system $F_P(x, z; \mu) = 0$, given by (7), is bounded away from zero, that is,*

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

Proof We will prove the above result by showing that the Newton iterates converge to the solution (x_μ^*, z_μ^*) if the initial Newton iterate (x^0, z^0) satisfies

$$\left\| \begin{pmatrix} x^0 - x_\mu^* \\ z^0 - z_\mu^* \end{pmatrix} \right\| \leq D \quad (39)$$

for some constant $D > 0$. Consider $0 \leq \mu \leq \hat{\mu}$. Assume Newton's method is applied to the perturbed system $F_P(x, z; \mu) = 0$ for a fixed μ . Let (x, z) and (x_μ^*, z_μ^*) respectively denote the current Newton iterate and the solution to the perturbed system for a given value of μ . Then at the subsequent Newton iteration, we have

$$\begin{pmatrix} x^+ - x_\mu^* \\ z^+ - z_\mu^* \end{pmatrix} = F'_P(x, z; \mu)^{-1} [F_P(x_\mu^*, z_\mu^*; \mu) - F_P(x, z; \mu) - F'_P(x, z; \mu)\{(x_\mu^*, z_\mu^*) - (x, z)\}]. \quad (40)$$

Taking norms in (40), we have

$$\left\| \begin{pmatrix} x^+ - x_\mu^* \\ z^+ - z_\mu^* \end{pmatrix} \right\| \leq \|F'_P(x, z; \mu)^{-1}\| \times \left\| F_P(x_\mu^*, z_\mu^*; \mu) - F_P(x, z; \mu) - F'_P(x, z; \mu) \left\{ \begin{pmatrix} x_\mu^* - x \\ z_\mu^* - z \end{pmatrix} \right\} \right\|.$$

Since $\|F'_P(x, z; \mu)^{-1}\| \leq M$ for some $M > 0$ then a straightforward application of Lemma 4.1.3 yields that for some constant $N > 0$

$$\left\| \begin{pmatrix} x^+ - x_\mu^* \\ z^+ - z_\mu^* \end{pmatrix} \right\| \leq \frac{MN}{2} \left\| \begin{pmatrix} x - x_\mu^* \\ z - z_\mu^* \end{pmatrix} \right\|^2.$$

To obtain convergence of the Newton sequence in (x, z) to (x_μ^*, z_μ^*) , it is sufficient to have

$$\left\| \begin{pmatrix} x^0 - x_\mu^* \\ z^0 - z_\mu^* \end{pmatrix} \right\| \leq \frac{2}{MN}. \quad (41)$$

Therefore, $D \leq r_P(\mu)$ where $0 < D \leq \frac{1}{MN}$.

□

5 Numerical Experimentation on the Radius of the Sphere of Convergence

In this section, we obtain numerical upper-bound estimates for the radius of the sphere of convergence of Newton's method on the two equivalent systems for various values of $\mu > 0$.

Our numerical experiments were conducted similar to the implementation done in [13] for the linear program. We briefly describe the algorithm and new modifications.

We applied Newton's method to the barrier and perturbed systems for various values of $\mu > 0$. Initial points for Newton's method were of the form

$$v^0 = v_\mu^* + \lambda v_\alpha,$$

where v_α was chosen using the Matlab function `rand`, which selects random entries from a uniform distribution on the unit interval $(0, 1)$. The point v_μ^* is the solution x_μ^* and (x_μ^*, z_μ^*) for systems (5) and (7), respectively. The initial point v^0 was merely x^0 for Newton's method on the barrier system. The first n components of the initial point v^0 for the perturbed system composed the point x^0 and the remaining components composed the point z^0 for Newton's method.

The experiments were begun with initial points v^0 close to v_μ^* , that is, λ was initialized close to zero. Full Newton steps were always taken. If for a given $\lambda > 0$ Newton's method generated an iterate that satisfied a convergence criterion, then λ was incremented and Newton's method was started with a new initial point v^0 . The convergence criterion was given by

$$\frac{\|x_\mu^* - x_k\|}{1 + \|x_\mu^*\|} + \frac{\|z_\mu^* - z_k\|}{1 + \|z_\mu^*\|} \leq tol \quad (42)$$

where $(z_k)_i = \mu/g_i(x_k)$ for the barrier system for $i = 1, \dots, m$, and the convergence tolerance was set to $tol = 10^{-8}$. Nonconvergence of Newton's method was recorded when the maximum number of iterations, which was set to 100, was reached.

Ten unit random directions v_α having positive and negative components were generated. For each direction, v_α , Newton's method was applied to the barrier and perturbed systems until the convergence criteria (42) was met at some λ_α . For a given value of $\mu > 0$, the numerical upper-bound estimate for the radius of the sphere of convergence of Newton's method applied to the barrier or perturbed system was recorded as the minimum of the ten λ_α 's, that is,

$$\min_{\alpha} \{\lambda_\alpha\},$$

where $\lambda_\alpha = \|x^0 - x_\mu^*\|$ for the barrier system and $\lambda_\alpha = \|v^0 - v_\mu^*\|$ for the perturbed system. Newton's method was applied to the barrier and perturbed systems with the following parameterized values of μ :

$$\mu = 0.45, 0.25, 0.10, 0.05, 0.01, 0.0075, 0.005, 0.0005, 0.00005 \quad (43)$$

as in the experiments for the linear program [13].

A subset of nondegenerate problems were considered from the Hock and Schittkowski test set [7]. Table 1 shows the problem numbers in the first column followed by the number of variables and inequality constraints. The problems were run on a Sun Ultra Sparc workstation using Matlab version 5.2.

Table 1: Test Problems

Problem	n	m
HS 10	2	1
HS 11	2	1
HS 12	2	1
HS 16	2	5
HS 17	2	5
HS 22	2	2
HS 33	3	5
HS 34	3	5
HS 43	4	3

Now we present several numerical results on upper-bounds for the radius of the sphere of convergence of Newton's method. We show only four of the problems whose results are typical to the other problems that are not presented. Figures 1 - 2 show the radius of the sphere of

convergence associated with the two equivalent systems graphed against the values of μ given in (43).

The numerical experiments show the radius of the sphere of convergence of Newton’s method on the barrier system decreasing towards zero for small values of $\mu > 0$. However, the radius associated with the perturbed system clearly stays away from zero. In addition, the radius associated with the perturbed system is also larger than the radius associated with the barrier system for all μ values. These numerical results confirm our theory for the radius of the sphere of convergence of Newton’s method on these two equivalent systems for nondegenerate problems.

In the case when the maximum number of iterations was reached, we observed that the final Newton iterate was infeasible for the two equivalent systems. As in the linear programming case [13], we observed also that infeasible Newton iterates did not always preclude convergence to the solution of the perturbed system. However, convergence was precluded for Newton’s method on the barrier system if an iterate became infeasible. As a result, the Jacobian became increasingly ill-conditioned.

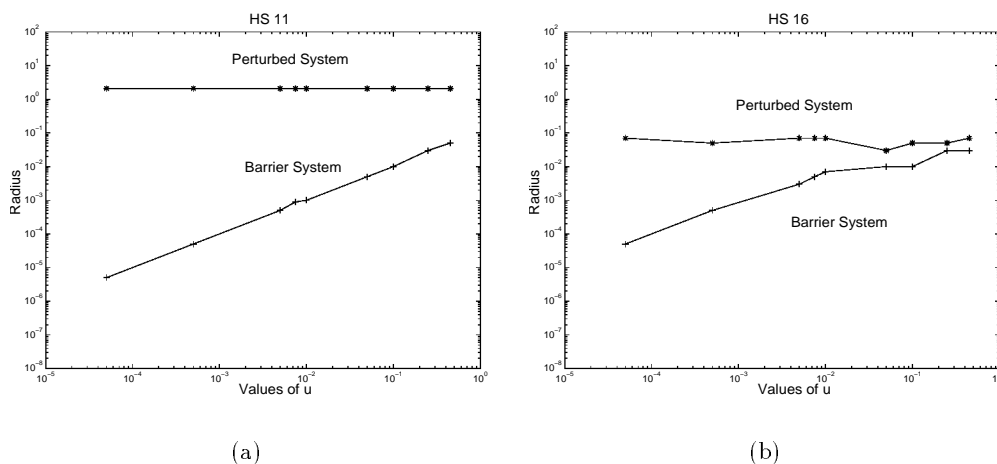


Figure 1: Radii of the spheres of convergence for Newton’s method applied to the barrier and perturbed systems

6 Conclusions

We studied an aspect of the local behavior of a Newton log-barrier function method and a Newton primal-dual interior-point method. The Newton log-barrier method can be viewed as applying damped Newton’s method to the barrier system. Similarly, the Newton primal-dual method can be viewed as applying damped Newton’s method to the perturbed system of the optimization problem. In particular, we analyzed the radius of the sphere of convergence of Newton’s method applied to the barrier and perturbed systems for nondegenerate problems as the barrier parameter μ decreases to zero. Previous results [15] showed that the radius of the sphere of convergence of Newton’s method on the barrier system was bounded below in $O(\mu^\alpha)$ for $\alpha > 1$, but no upper-bound was obtained. Our theoretical results are stronger and establish that the radius of the sphere of convergence of

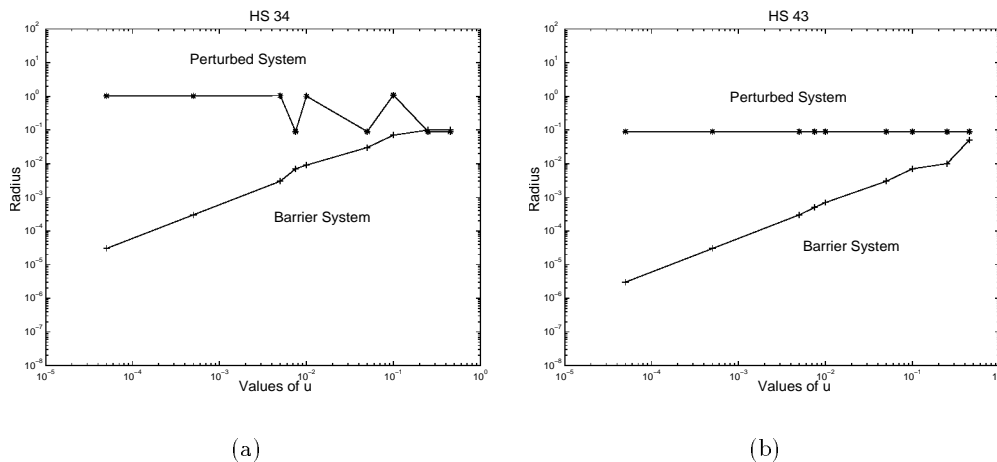


Figure 2: Radii of the spheres of convergence for Newton's method applied to the barrier and perturbed systems

Newton's method on the barrier system decreases to zero with the same order as μ decreases to zero.

In addition, we presented numerical experiments on the radius of the sphere of convergence of Newton's method applied to the barrier and perturbed systems for nondegenerate problems. Our experiments reinforce the theoretical results obtained for the radius of the sphere of convergence of Newton's method applied to the barrier and perturbed systems. The numerical and theoretical results imply that the radius of the sphere of convergence of Newton's method applied to the perturbed system is larger than the radius associated with the barrier system, at least for small μ values. As a consequence, we expect fewer Newton iterations to obtain an approximate solution to the nonlinear program. Our results clearly favor the use of Newton primal-dual interior-point methods for solving the nonlinear program.

References

- [1] M. Avriel. *Nonlinear Programming: Analysis and Methods*. Prentice-Hall, Englewood Cliffs, NJ, 1976.
- [2] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, NJ, 1983. Reprinted by SIAM Publications, 1996.
- [3] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang. On the formulation and theory of the Newton interior-point method for nonlinear programming. *Journal of Optimization Theory and Applications*, 89(3):507–541, 1996.
- [4] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming, Sequential Unconstrained Minimization Techniques*. Wiley, New York, 1968. Reprinted by SIAM Publications, 1990.

- [5] A. Forsgren and P. E. Gill. Primal-dual interior methods for nonconvex nonlinear programming. Technical Report NA-3, Department of Mathematics, University of California at San Diego, 1996.
- [6] R. Frisch. The logarithmic potential method of convex programming. Technical report, University Institute of Economics, Oslo, Norway, 1955.
- [7] W. Hock and K. Schittkowski. Test examples for nonlinear programming codes. In *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, 1981.
- [8] L. McLinden. An analogue of Moreau's proximation theorem, with application to the nonlinear complementarity problem. *Pacific Journal of Mathematics*, 88(1):101–161, 1980.
- [9] S. G. Nash and A. Sofer. Why extrapolation helps barrier methods. Technical report, Operations Research and Engineering Department, George Mason University, 1998.
- [10] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, Inc., San Diego, CA, 1970.
- [11] Z. Parada and R. A. Tapia. Computational experience with a modified augmented lagrangian merit function in a primal-dual interior-point method. Technical report, Department of Computational and Applied Mathematics, Rice University, 1995.
- [12] T. Urban, A. Tits, and C. T. Lawrence. A primal-dual interior point method for nonconvex optimization with multiple logarithmic barrier parameters and with strong convergence properties. Technical report, University of Maryland, College Park, 1998.
- [13] M. C. Villalobos, R. A. Tapia, and Y. Zhang. The behavior of Newton-type methods on two equivalent systems from linear programming. Technical Report CRPC-TR98770-S, Department of Computational and Applied Mathematics, Rice University, 1998.
- [14] M. H. Wright. Why a pure primal Newton barrier step may be infeasible. *SIAM Journal of Optimization*, 5(1):1–12, 1995.
- [15] S. J. Wright. On the convergence of the Newton/log-barrier method. Technical Report P681-0897, Mathematics and Computer Science Division, Argonne National Laboratory, 1997.
- [16] S. J. Wright. *Primal-Dual Interior-Point Methods*. SIAM Publications, Philadelphia, 1997.
- [17] S. J. Wright and F. Jarre. The role of linear objective functions in barrier methods. Technical Report P485-1294, Mathematics and Computer Science Division, Argonne National Laboratory, 1997.
- [18] Y. Zhang. Lipsol, version 0.4. Department of Computational and Applied Mathematics, Rice University.
- [19] Y. Zhang, R. A. Tapia, and J. E. Dennis. On the superlinear and quadratic convergence of primal-dual interior point linear programming algorithms. *SIAM Journal of Optimization*, pages 304–324, 1992.