On the Local Behavior of an Interior Point Method for Nonlinear Programming

Richard H. Byrd* Guanghui Liu † Jorge Nocedal‡

January 24, 1998

Report OTC 98/02

OPTIMIZATION TECHNOLOGY CENTER

Abstract

We study the local convergence of a primal-dual interior point method for nonlinear programming. A linearly convergent version of this algorithm has been shown in [2] to be capable of solving large and difficult non-convex problems. But for the algorithm to reach its full potential, it must converge rapidly to the solution. In this paper we describe how to design the algorithm so that it converges superlinearly on regular problems.

Key words: constrained optimization, interior point method, large-scale optimization, nonlinear programming, primal method, primal-dual method, successive quadratic programming.

*Computer Science Department, University of Colorado, Boulder CO 80309. This author was supported by ARO grant DAAH04-94-0228, and AFOSR grant F49620-97-1-0164.

†IEMS Department, Northwestern University, Evanston Il 60208 This author was supported by by National Science Foundation grant CCR-9625613 and Department of Energy grant DE-FG02-87ER25047-A004.

‡ECE Department, Northwestern University, Evanston Il 60208. This author was supported by National Science Foundation grant CCR-9625613 and by Department of Energy grant DE-FG02-87ER25047-A004.

1. Introduction

There is at present a great deal of interest in the development of interior point methods for solving nonlinear programming problems of the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0 \\ & g(x) \leq 0, \end{array} \tag{1.1}$$

where $f : \mathbf{R}^n \to \mathbf{R}$, $h : \mathbf{R}^n \to \mathbf{R}^q$, and $g : \mathbf{R}^n \to \mathbf{R}^t$ are smooth functions.

The case when (1.1) is a convex program can be solved by algorithms that are, to a great extent, direct extensions of primal-dual interior point methods for linear programming. However, the presence of non-convexity raises a number of obstacles that can be tackled only by sophisticated methods that include many features not present in linear programming algorithms. These include the exploitation of negative curvature directions, the handling of ill-conditioning in the objective function, and the avoidance of stationary points that are not minimizers.

Several algorithms have recently been proposed to solve non-convex nonlinear programs (see [2], [3], [7], [8]). In this paper we study one of the most promising of these algorithms, which is described in Byrd, Hribar and Nocedal [2], and which is based on a general framework proposed by Byrd, Gilbert and Nocedal [1]. It has been shown in [2] that this algorithm is robust, can make effective use of second derivative information and sparsity in the problem, and that it is competitive with the LANCELOT algorithm in terms of function evaluations. No attempt was made in [2] to make this new algorithm rapidly convergent, and the goal of this paper is to develop the theoretical foundation for an implementation that gives a superlinear rate of convergence.

The algorithm proposed in [2] is a barrier method. To describe it we begin by adding positive slack variables s to the inequalities $g(x) \leq 0$, and by adding the penalty term $-\mu \sum_{i=1}^{t} \ln(s_i)$ to the objective function $f(x)$, so as to prevent the slack variables from becoming negative. We thus obtain the barrier subproblem

$$\begin{array}{ll} \text{minimize} & \psi_\mu(x,s) \\ \text{s.t.} & c(x,s) = 0 \\ & s > 0, \end{array} \tag{1.2}$$

where $\mu > 0$ is the barrier parameter and where

$$\psi_\mu(x,s) = f(x) - \mu \sum_{i=1}^{t} \ln(s_i), \qquad c(x,s) = \begin{pmatrix} h(x) \\ g(x) + s \end{pmatrix}. \tag{1.3}$$

The Lagrangian function associated with this barrier problem is

$$L(x, s, \lambda_E, \lambda_I) = \psi_\mu(x,s) + \lambda_E^T h(x) + \lambda_I^T (g(x) + s). \tag{1.4}$$

The interior point algorithm uses an SQP method [9],[5] with trust region techniques to find an approximate solution of the barrier problem (1.2). Once this is done, the barrier parameter μ is decreased, and a new barrier problem is approximately solved by the SQP method.

Each step d of the SQP method is obtained by approximately solving the quadratic program

$$\begin{aligned} \text{minimize} \quad & \nabla\psi_\mu(x,s)^T d + \frac{1}{2} d^T W d \\ \text{s.t.} \quad & A(x,s)^T d + c(x,s) = b \\ & \|d\|_S \leq \Delta, \qquad d_s \geq -\tau s. \end{aligned} \tag{1.5}$$

Here

$$A(x,s) = \left[\begin{array}{cc} A_h(x) & A_g(x) \\ 0 & I \end{array} \right] \tag{1.6}$$

is the matrix of constraint gradients, $S = diag(s_1, ..., s_t)$, and b is a residual vector designed to ensure that the constraints of (1.5) are compatible. The vector b is chosen to be as small as possible, as described in [2], and its precise definition is not important in the discussion that follows. We have written the step d in terms of its x and s-components,

$$d = \left(\begin{array}{c} d_x \\ d_s \end{array} \right).$$

The trust region constraint makes use of the scaling matrix S, and is defined by

$$\|d\|_S^2 \equiv \|d_x\|^2 + \|S^{-1} d_s\|^2 \leq \Delta^2, \tag{1.7}$$

where $\|\cdot\|$ denotes the Euclidean norm. We have also included the *fraction to the boundary rule* $d_s \geq -\tau s$ which is commonly used in linear programming; here τ is a parameter close to 1 whose choice will be studied in this paper. The trust region radius Δ is chosen to ensure decrease of the merit function

$$\phi(x,s;\nu) = \psi_\mu(x,s) + \nu \|c(x,s)\|_2, \qquad \nu > 0. \tag{1.8}$$

The choice of the matrix W in the quadratic program (1.5) is important. If we define it as

$$W_P = \left(\begin{array}{cc} \nabla^2_{xx} L(x,s,\lambda_E,\lambda_I) & 0 \\ 0 & \mu S^{-2} \end{array} \right),$$

then W is the Hessian of the Lagrangian (1.4) with respect to x and s, and the method will generate primal search directions. On the other hand, if we define W to be

$$W_{PD} = \left(\begin{array}{cc} \nabla^2_{xx} L(x,s,\lambda_E,\lambda_I) & 0 \\ 0 & S^{-1}\Lambda \end{array} \right), \tag{1.9}$$

where $\Lambda = diag(\lambda_1, ..., \lambda_t)$, then the search direction will have primal-dual characteristics. Since the primal-dual step is known to be better able to approach the solution, we focus on it in this paper and we will assume that W is given by (1.9). The Lagrange multipliers λ_E

and λ_I are least squares estimates based on minimizing $\|\nabla_{(x,s)}L(x_k, s_k, \lambda_E, \lambda_I)\|_2$. The way in which these multipliers are computed is not important here, but we should note that we may adjust λ_I to make it positive.

As mentioned before, the step d of the SQP method is an approximate solution of the quadratic program (1.5). It is computed by means of a projected conjugate gradient iteration on (1.5). Indefiniteness is handled by stopping the conjugate gradient iteration as soon as a direction of negative curvature is found, and by following this negative curvature direction to the boundary of the trust region [11]. Once a step d is computed, we test if it provides sufficient reduction in the merit function ϕ. If not, we reduce the trust region and compute a new step by resolving the quadratic program (1.5) with this smaller trust region. But if the step provides a sufficient reduction in the merit function, we set

$$x^+ = x + d_x \ ; \quad s^+ = s + d_s, \tag{1.10}$$

where the superscript + denotes the new iterate. We then check whether the barrier problem has been approximately solved by testing the inequality

$$\|F_\mu(x^+, s^+, \lambda_E(x^+, s^+), \lambda_I(x^+, s^+))\| \le \epsilon_\mu, \tag{1.11}$$

for some stopping tolerance ϵ_μ, where

$$F_\mu(z) = \begin{pmatrix} \nabla f(x) + A_h(x)\lambda_E + A_g(x)\lambda_I \\ S\lambda_I - \mu e \\ h(x) \\ g(x) + s \end{pmatrix}, \quad z = (x, s, \lambda_E, \lambda_I) \tag{1.12}$$

with $e = (1, 1, ..., 1)^T$, and where the functions $\lambda_E(x, s)$ and $\lambda_I(x, s)$ are least squares estimates of the multipliers. The barrier parameter μ is decreased as soon as (1.11) is satisfied. Note that $F_\mu(z)$ represents the optimality conditions for the barrier problem (1.2). We should emphasize that the iterates of the interior point algorithm are given by (1.10); this means that each step of the SQP iteration defines a step of the algorithm.

This broad outline of the algorithm will be sufficient for our purposes because we will study here only its local convergence properties. Byrd, Gilbert and Nocedal [1] have analyzed the global behavior of the primal version of this method, and found it to be quite satisfactory. Since we expect that the good global convergence properties of the interior point method can produce an iterate close to a KKT point, we will assume in this paper that an iterate with associated multipliers,

$$z_k = (x_k, s_k, \lambda_E, \lambda_I) \tag{1.13}$$

has been generated in a neighborhood of a minimizer z^* satisfying the following conditions.

Assumptions I.

1. (Stationarity) The vector $z^* = (x^*, s^*, \lambda_E^*, \lambda_I^*)$ is a KKT point for the nonlinear program (1.1), i.e. $F_0(z^*) = 0$, $\lambda_I^* \ge 0$ and $s^* \ge 0$.

2. (Smoothness) The Hessian matrices $\nabla^2 f(x)$, $\nabla^2 h_i(x)$, $i = 1, \cdots, q$, $\nabla^2 g_j(x)$, $j = 1, \cdots, t$, are locally Lipschitz continuous at x^*.

3. (Regularity) The active constraint gradients

$$\{\nabla h_i(x^*), i = 1, \cdots, q\} \cup \{\nabla g_j(x^*),\ j \in \mathcal{B}\}$$

are linearly independent, where $\mathcal{B} = \{j : g_j(x^*) = 0\}$.

4. (Optimality) The second order sufficiency conditions for optimality are satisfied at z^*: for all $v \neq 0$ satisfying $\nabla h_i(x^*)^T v = 0$, $i = 1, \cdots, q$, and $\nabla g_j(x^*)^T v = 0$, $j \in \mathcal{B}$, we have that $v^T \nabla^2_{xx} L(z^*) v > 0$.

5. (Strict complementarity) $s^* + \lambda_I^* > 0$.

As the iterates z_k generated by the interior point algorithm given in [2] approach the solution z^*, the step produced by the algorithm will approach more and more the x and s-components of the *pure primal-dual* step given by

$$F'_\mu(z) d_N = -F_\mu(z), \tag{1.14}$$

where F_μ is defined by (1.12), and

$$F'_\mu(z) = \begin{pmatrix} \nabla^2_{xx} L & 0 & A_h(x) & A_g(x) \\ 0 & \Lambda_I & 0 & S \\ A_h^T(x) & 0 & 0 & 0 \\ A_g^T(x) & I & 0 & 0 \end{pmatrix}. \tag{1.15}$$

This is of course a Newton step on the system $F_\mu(z) = 0$. Even though the interior point algorithm proposed in [2] does not compute a step d by directly solving this system, but by means of a projected conjugate gradient method applied to (1.5), one can show that near the solution z^*, its step d satisfies

$$F'_\mu(z) d = -F_\mu(z) + r_\mu, \tag{1.16}$$

for a certain residual vector r_μ. Therefore d is an inexact Newton step, and is related to the pure primal-dual step by

$$d = d_N + [F'_\mu(z)]^{-1} r_\mu. \tag{1.17}$$

The magnitude of the residual vector must be controlled carefully. In the implementation described in [2], the CG iteration is stopped as soon as $\|r_\mu\| \leq \mu$. But, as we will see, to obtain a faster rate of convergence we will need to decrease the magnitude of r_μ more rapidly.

Our analysis will focus on the iteration $z^+ = z + d$, where d is given by (1.16). This is an idealization of the iteration of the interior point algorithm described in [2], but the results obtained here will be directly applicable to that algorithm. To partly bridge this gap, we will also consider the effect of resetting λ_E^+ and λ_I^+ based on least squares estimates at (x^+, s^+).

Notation. Throughout the paper z and z^+ will denote the current and new iterates, respectively. Assumptions I imply that, in a neighborhood of the minimizer z^*, and for all sufficiently small values of μ, the barrier problem (1.2) associated with each barrier function ψ_μ has a unique minimizer which we denote by $z^*(\mu)$. We will write

$$F_0(z) = F(z),$$

and since $F'(z) = F'_\mu(z)$ for any μ, we will use the former for simplicity. Throughout the paper $\|\cdot\|$ denotes the Euclidean norm.

2. Some Conditions for Superlinear Convergence

Since the Jacobian $F'_\mu(z)$ defined by (1.15) is independent of the barrier parameter μ, so is the region of convergence of the Newton iteration (1.14). As a result, the algorithm can be designed so that near the solution $z^*(\mu)$ of the barrier problem, each step d converges superlinearly to $z^*(\mu)$. The goal of this section is to show that the step also converges superlinearly to the solution z^* of the nonlinear program (1.1). We begin by stating the following well-known result.

Lemma 2.1 *([6] and page 73 of [10]). If Assumptions I hold, there exists a neighborhood $N(z^*) = \{z : \|z - z^*\| \leq \delta\}$ such that for all $z \in N(z^*)$ the Jacobian $F'(z)$ is invertible and*

$$||[F'(z)]^{-1}|| \leq M, \tag{2.1}$$

for some constant $M > 0$. Moreover for all $z, z' \in N(z^*)$ we have that

$$\|F'(z)(z - z') - F(z) + F(z')\| \leq L(\|z - z'\|^2), \tag{2.2}$$

for some constant $L > 0$.

The following lemma is a direct result of the implicit function theorem. It states that the solution $z^*(\mu)$ of the barrier problem (1.2) is a Lipschitz continuous function of μ.

Lemma 2.2 *(page 128 of [10]) Suppose that Assumptions I hold. Then there is $\bar{\mu} > 0$ such that for all $\mu \leq \bar{\mu}$, the system $F_\mu(z) = 0$ has a solution $z^*(\mu) \in N(z^*)$, where $N(z^*)$ is defined in Lemma 2.1. Moreover*

$$\|z^*(\mu) - z^*\| \leq C\mu, \tag{2.3}$$

where

$$C = \max_{z \in N(z^*)} \|[F'(z)]^{-1} \frac{\partial F_\mu(z)}{\partial \mu}\|$$

is a constant independent of μ.

The next result gives a standard inexact Newton analysis of the step generated by the interior point method.

Theorem 2.3 *Suppose that Assumptions I hold; let $N(z^*)$ be the neighborhood defined in Lemma 2.1, and $\bar{\mu}$ be the threshold value defined in Lemma 2.2. Then for any $z \in N(z^*)$ and $\mu < \bar{\mu}$,*

$$\|z^+ - z^*(\mu)\| \le ML\|z - z^*(\mu)\|^2 + M\|r_\mu\|, \tag{2.4}$$

where M and L are defined by (2.1)-(2.2), *and r_μ is the residual of the CG iteration and is defined by* (1.17).

Proof. Since $z^*(\mu)$ is the minimizer of the barrier problem associated with ψ_μ, we have that $F_\mu(z^*(\mu)) = 0$, and thus

$$F(z^*(\mu)) = \mu(0, e, 0, 0)^T.$$

Using this, recalling (1.14), (1.17) and the equality $F'_\mu = F'$, we have that

$$\begin{aligned} z^+ - z^*(\mu) &= z + d - z^*(\mu) \\ &= [F'(z)]^{-1} F'(z)(z - z^*(\mu)) + d \\ &= [F'(z)]^{-1} F'(z)(z - z^*(\mu)) + d_N + [F'(z)]^{-1} r_\mu \\ &= [F'(z)]^{-1}\left[F'(z)(z - z^*(\mu)) - F(z) + \mu(0, e, 0, 0)^T\right] \\ &\quad + [F'(z)]^{-1} r_\mu \\ &= [F'(z)]^{-1}\left[F'(z)(z - z^*(\mu)) - F(z) + F(z^*(\mu))\right] \\ &\quad + [F'(z)]^{-1} r_\mu. \end{aligned} \tag{2.5}$$

By Lemma 2.2, for all $0 < \mu < \bar{\mu}$ the barrier minimizers satisfy $z^*(\mu) \in N(z^*)$. We therefore have from (2.5), (2.1) and (2.2) that if $z \in N(z^*)$ then (2.4) holds. □

The following technical lemma states that $\|F_\mu(z)\|$ can be used as a measure of the distance between z and the solution $z^*(\mu)$ of the barrier problem.

Lemma 2.4 *Let M and $N(z^*)$ be defined in Lemma 2.1 and $\bar{\mu}$ be defined in Lemma 2.2. Then $\forall z$ sufficiently close to $z^*(\mu)$, and $\forall \mu < \bar{\mu}$,*

$$\|z - z^*(\mu)\| \le 2M\|F_\mu(z)\|, \quad \|F_\mu(z)\| \le 2Q\|z - z^*(\mu)\|, \tag{2.6}$$

where

$$Q \equiv \sup_{z \in N(z^*)} \|F'(z)\|. \tag{2.7}$$

Proof. By Taylor's theorem, we have

$$F_\mu(z) = F_\mu(z) - F_\mu(z^*(\mu)) = F'(z^*(\mu))(z - z^*(\mu)) + o(\|z - z^*(\mu)\|). \tag{2.8}$$

Using this, (2.1) and (2.7), and Lemma 2.2, it follows that $\forall z \in N(z^*)$ and $\forall \mu < \bar{\mu}$,

$$\begin{aligned} \|F_\mu(z)\| &\le \|F'(z^*(\mu))\|\|z - z^*(\mu)\| + o(\|z - z^*(\mu)\|) \\ &\le Q\|z - z^*(\mu)\| + o(\|z - z^*(\mu)\|) \end{aligned} \tag{2.9}$$

and

$$\begin{aligned}
\|z - z^*(\mu)\| &= \|[F'(z^*(\mu))]^{-1}[F_\mu(z) + o(\|z - z^*(\mu)\|)]\| \\
&\leq M\,[\|F_\mu(z)\| + o(\|z - z^*(\mu)\|)] \\
&= M\|F_\mu(z)\| + o(M\|z - z^*(\mu)\|). \qquad (2.10)
\end{aligned}$$

We can assume, without loss of generality, that

$$o(\|z - z^*(\mu)\|) \leq \min\{Q, \frac{1}{2M}\}\|z - z^*(\mu)\|.$$

Substituting this into (2.9) and (2.10), we complete the proof.

□

The following theorem presents a strategy for decreasing the barrier parameter and controlling the size of the residual, so as to achieve a fast rate of convergence near the solution z^*.

Theorem 2.5 *Suppose that Assumptions I hold, and let z be an iterate sufficiently close to z^* at which the barrier parameter is decreased from μ to μ^+. Suppose that the residual in* (1.16) *satisfies*

$$\|r_{\mu^+}\| \leq M_1\|F_{\mu^+}(z)\|^{1+\alpha}, \qquad (2.11)$$

for some positive constants M_1, α. Then if $\mu^+ = o(\|F(z)\|)$ and $\alpha > 0$, the step will be superlinearly convergent to z^*; moreover, if $\mu^+ = O(\|F(z)\|^2)$ and $\alpha \geq 1$, the step will be quadratically convergent.

Proof.

$$\begin{aligned}
\|F_{\mu^+}(z)\| &= \|F(z) - \mu^+(0, e, 0, 0)^T\| \\
&\leq \|F(z)\| + \sqrt{t}\mu^+ \\
&= \|F(z)\|\left(1 + \frac{\sqrt{t}\mu^+}{\|F(z)\|}\right).
\end{aligned}$$

Since the assumption $\mu^+ = O(\|F(z)\|^2)$ or $\mu^+ = o(\|F(z)\|)$ implies that

$$\frac{\sqrt{t}\mu^+}{\|F(z)\|} \to 0,$$

we can assume that

$$\frac{\sqrt{t}\mu^+}{\|F(z)\|} \leq 1.$$

Then

$$\|F_{\mu^+}(z)\| \leq 2\|F(z)\|. \qquad (2.12)$$

Using this, Theorem 2.3 and Lemma 2.2, we deduce that

$$\begin{aligned}
\|z^+ - z^*\| &\leq \|z^+ - z^*(\mu^+)\| + \|z^*(\mu^+) - z^*\| \\
&\leq ML\|z - z^*(\mu^+)\|^2 + M\|r_{\mu+}\| + C\mu^+ \\
&\leq ML\left[\|z - z^*\| + \|z^* - z^*(\mu^+)\|\right]^2 + MM_1\|F_{\mu^+}(z)\|^{1+\alpha} + C\mu^+ \\
&\leq ML\left(2\|z - z^*\|^2 + 2\|z^* - z^*(\mu^+)\|^2\right) + MM_1\|F_{\mu^+}(z)\|^{1+\alpha} \\
&\quad +C\mu^+ \\
&\leq 2ML\|z - z^*\|^2 + 2MLC^2(\mu^+)^2 + 2^{1+\alpha}MM_1\|F(z)\|^{1+\alpha} \\
&\quad +C\mu^+ \\
&\leq 2ML\|z - z^*\|^2 + 2^{1+\alpha}MM_1\|F(z)\|^{1+\alpha} + \\
&\quad (2MLC\bar{\mu} + 1)C\mu^+. \qquad (2.13)
\end{aligned}$$

The second inequality in (2.6) with $\mu = 0$ gives

$$\|F(z)\| \leq 2Q\|z - z^*\|. \qquad (2.14)$$

Using this and (2.13) it follows that

$$\begin{aligned}
\frac{\|z^+ - z^*\|}{\|z - z^*\|} &\leq 2ML\|z - z^*\| + 2^{1+\alpha}MM_1\frac{\|F(z)\|^{1+\alpha}}{\|z - z^*\|} \\
&\quad +(2MLC\bar{\mu} + 1)C\frac{\mu^+}{\|z - z^*\|} \\
&\leq 2ML\|z - z^*\| + (4Q)^{1+\alpha}MM_1\|z - z^*\|^\alpha \\
&\quad +(2MLC\bar{\mu} + 1)C\frac{\mu^+}{\|z - z^*\|} \\
&\leq 2ML\|z - z^*\| + (4Q)^{1+\alpha}MM_1\|z - z^*\|^\alpha \\
&\quad +2QC(2MLC\bar{\mu} + 1)\frac{\mu^+}{\|F(z)\|}.
\end{aligned}$$

Using this, we see that the conditions

$$\alpha > 0,\ \mu^+ = o(\|F(z)\|), \qquad (2.15)$$

imply that the right hand side is $o(1)$ so that the step is superlinearly convergent. Moreover, (2.14) and the conditions

$$\alpha \geq 1,\ \mu^+ = O(\|F(z)\|^2), \qquad (2.16)$$

imply that the right hand side is $O(\|z - z^*\|)$, so that convergence is quadratic. □

The conditions given in this theorem are not the only possible (or practical) ones. One can show, for example, that the theorem is valid if the residual satisfies the bound

$$\|r_{\mu^+}\| \leq M_1(\mu^+)^\alpha,$$

for some positive constants α, M_1.

Now let us consider the strategy used in [2] and described in Section 1, where given a primal vector (x, s), we compute least squares multiplier estimates $\lambda_E(x, s)$ and $\lambda_I(x, s)$, prior to computing the new step d. Assuming that the barrier parameter is decreased from μ to μ^+ at this point, the multipliers are obtained by minimizing $\|F_{\mu^+}(z)\|$. Then Assumption I-3 implies that, near (x^*, s^*), the resulting vector $z = (x, s, \lambda_E, \lambda_I)$ satisfies $\|z - z^*\| = O(\|(x, s) - (x^*, s^*)\| + \mu^+)$. This fact immediately gives us the following statement about the convergence of that algorithm.

Corollary 2.6 *2.1 Suppose that Assumptions I hold, that each step is computed using values of* (λ_E, λ_I) *based on least squares estimates, and that the residual in* (1.16) *satisfies* (2.11). *Let* (x, s) *be a primal vector sufficiently close to* (x^*, s^*) *at which the barrier parameter has been decreased from* μ *to* μ^+. *Then if* $\mu^+ = o(\|F(z)\|)$ *and* $\alpha > 0$, *the step from* (x, s) *to* (x^+, s^+) *will be superlinearly convergent to* (x^*, s^*)*; moreover, if* $\mu^+ = O(\|F(z)\|^2)$ *and* $\alpha \geq 1$, *the step will be quadratically convergent.*

The results presented in this section ignore the fact that the slack variables must remain positive. How to ensure this is the focus of the next section.

3. Feasibility

In primal-dual methods for linear programming, the step generated by the algorithm may violate the bounds on the slacks, and it is thus common to use a backtracking line search to ensure that the slacks remain sufficiently positive [13]. The interior point algorithm for nonlinear programming [2] which is the focus of this paper employs trust regions and computes the search direction by means of the conjugate gradient (CG) method. Because this CG iteration has the property that its estimates of the step d are of increasing norm [11], one can safely terminate the CG iteration if any of these estimates lies outside the trust region. It is desirable to preserve this feature and to avoid the need for a backtracking line search. To this end we will study under what conditions can we guarantee that the step keeps the slacks sufficiently positive, and can therefore be used directly in the trust region method.

Throughout this section we will assume that at the current iterate z, the barrier optimality test (1.11) is satisfied, and will analyze the step d taken from z. This step attempts to minimize the barrier problem associated with ψ_{μ^+}. We begin by specializing the estimate given by Theorem 2.3 in this case.

Theorem 3.1 *Suppose that Assumptions I hold. If* z *is sufficiently close to* z^* *and* μ^+ *is sufficiently small, then*

$$\|z^+ - z^*(\mu^+)\| \leq M'(\epsilon_\mu + \mu)^2 + M\|r_{\mu^+}\|, \tag{3.1}$$

where M' *is a positive constant and* M *is the constant defined in Lemma 2.1.*

Proof. By Theorem 2.3, there exists a neighborhood $N(z^*)$ and a value $\bar{\mu} > 0$ such that $\forall z \in N(z^*)$ and for $\mu^+ < \bar{\mu}$,

$$\begin{aligned} \|z^+ - z^*(\mu^+)\| &\leq ML\|z - z^*(\mu^+)\|^2 + M\|r_{\mu^+}\| \\ &\leq ML\left[\|z - z^*(\mu)\| + \|z^*(\mu) - z^*(\mu^+)\|\right]^2 \\ &\quad +M\|r_{\mu^+}\|. \end{aligned} \tag{3.2}$$

By Lemma 2.2 the minimizers of the barrier problems associated with ψ_μ and ψ_{μ^+} satisfy

$$\begin{aligned} \|z^*(\mu) - z^*(\mu^+)\| &\leq \|z^*(\mu) - z^*\| + \|z^*(\mu^+) - z^*\| \\ &\leq C(\mu + \mu^+) \\ &\leq 2C\mu, \end{aligned} \tag{3.3}$$

where the last inequality follows from the fact that the sequence of barrier parameters is decreasing. Using this in (3.2), recalling the stopping test (1.11), and (2.6), we have that

$$\begin{aligned} \|z^+ - z^*(\mu^+)\| &\leq ML\left(\|z - z^*(\mu)\| + 2C\mu\right)^2 + M\|r_{\mu^+}\| \\ &\leq ML\left(2M\|F_\mu(z)\| + 2C\mu\right)^2 + M\|r_{\mu^+}\| \\ &\leq ML\left(2M\epsilon_\mu + 2C\mu\right)^2 + M\|r_{\mu^+}\| \\ &\leq M'(\epsilon_\mu + \mu)^2 + M\|r_{\mu^+}\|, \end{aligned}$$

where $M' = 4ML\left(\max\{C, M\}\right)^2$.

□

This theorem and the definition (1.13) of z implies, in particular, that

$$\left| s_i^+ - \frac{\mu^+}{\lambda_i^*(\mu^+)} \right| \leq M'(\epsilon_\mu + \mu)^2 + M\|r_{\mu^+}\|,$$

or

$$s_i^+ \geq \frac{\mu^+}{\lambda_i^*(\mu^+)} - \left[M'(\epsilon_\mu + \mu)^2 + M\|r_{\mu^+}\|\right], \tag{3.4}$$

where $\lambda_i^*(\mu^+)$ is the optimal inequality constraint multiplier for the barrier problem associated with ψ_{μ^+}, i.e. $\lambda_i^*(\mu^+) = [\lambda_I^*(\mu^+)]_i$. Thus if the stopping tolerance ϵ_μ and the residual $\|r_{\mu^+}\|$ are sufficiently small, the slacks are forced to be sufficiently positive. Note that (3.4) suggests that a quadratic decrease in μ may lead to a violation of feasibility. To see this, suppose that the last term in (3.4) is of order $O(\mu^2)$. Then

$$s_i^+ \geq \frac{\mu^+}{\lambda_i^*(\mu^+)} - O(\mu^2), \tag{3.5}$$

and if we choose $\mu^+ = O(\mu^2)$, then right hand side in (3.5) could become negative. Of course this argument is not rigorous because the constant implicit in the $O(\mu^2)$ term is unknown, and because the inequalities leading to (3.4) may not be tight.

We can now give conditions on the choice of the parameter τ which controls the fraction to the boundary rule, and on the strategy for decreasing μ, that ensure that the new slacks are sufficiently positive.

Theorem 3.2 *Suppose that Assumptions I hold, and that the parameters* $\mu^+, \epsilon_\mu, r_{\mu^+}$ *and* τ *are chosen so that*

$$\frac{(\mu+\epsilon_\mu)^2 + \|r_{\mu^+}\|}{\mu^+} \to 0, \tag{3.6}$$

and

$$\limsup\, (1-\tau)\left(\frac{\mu+\epsilon_\mu}{\mu^+}\right) < 1. \tag{3.7}$$

Then if z *is sufficiently close to* z^**, if* $\|F_\mu(z)\| \le \epsilon_\mu$*, and if* μ *is sufficiently small, the step* d *will satisfy the fraction to the boundary rule* $s^+ \ge (1-\tau)s$*.*

Proof. The limit (3.7) implies that for sufficiently small μ,

$$(1-\tau)\left(\frac{\mu+\epsilon_\mu}{\mu^+}\right) \le (1-\sigma)^2, \tag{3.8}$$

for some constant $0<\sigma<1$. Since $\mu^+ < \mu$ and $\epsilon_\mu > 0$, (3.8) implies

$$1-\tau < (1-\sigma)^2. \tag{3.9}$$

Given an index i, we divide our analysis into two cases.

Case I: Suppose that $\lambda_i^* = 0$, where λ_i denotes the i-th component of λ_I. In this case Assumption I-5 implies that $s_i^* > 0$. By Lemma 2.2 and Theorem 3.1,

$$\begin{aligned} s_i^+ &\ge s_i^* - |s_i^* - s_i^+| \\ &\ge s_i^* - \|z^* - z^*(\mu^+)\| - \|z^+ - z^*(\mu^+)\| \\ &\ge s_i^* - \left[C\mu^+ + M'(\epsilon_\mu+\mu)^2 + M\|r_{\mu^+}\|\right]. \end{aligned}$$

For μ and thus μ^+ sufficiently small, (3.6) implies that $\|r_{\mu^+}\|$ is small, and we can assume that the term inside the square brackets is less than σs_i^*. In addition, we can assume that for all z sufficiently close to z^*,

$$s_i \le s_i^*/(1-\sigma).$$

Therefore

$$\frac{s_i^+}{s_i} \ge (1-\sigma)^2, \tag{3.10}$$

and thus by (3.9), s_i^+ satisfies the fraction to the boundary rule.

Case II: Suppose that $\lambda_i^* > 0$. The step to the new value s^+ is determined by (1.15)-(1.16), and the second block of that equation reads

$$\lambda_i d_{s_i} + s_i d_{\lambda_i} = \mu^+ - \lambda_i s_i + r_i^+, \tag{3.11}$$

or

$$\lambda_i s_i^+ = \mu^+ - s_i d_{\lambda_i} + r_i^+, \tag{3.12}$$

where r_i^+ denotes the i-th component of r_{μ^+}. Assumption I-5 implies that $s_i^* = 0$; using this, Lemma 2.3 and the fact that the barrier test is satisfied at z, we have

$$s_i = s_i - s_i^* \leq \|z - z^*\| \leq 2M\|F(z)\| \leq 2M(\|F_\mu(z)\| + \sqrt{t}\mu) \leq 2M(\epsilon_\mu + \sqrt{t}\mu).$$

This and (3.12) give

$$\begin{aligned} \frac{s_i^+}{s_i} &= \frac{\mu^+ - s_i d_{\lambda_i} + r_i^+}{\lambda_i s_i} && (3.13) \\ &\geq \frac{\mu^+ - 2M(\epsilon_\mu + \sqrt{t}\mu)\|d\| - \|r_{\mu^+}\|}{\lambda_i s_i}. && (3.14) \end{aligned}$$

Now using (1.17), (1.14), (1.12), (2.1), (1.11) and Lemma 2.1 we can bound $\|d\|$ by

$$\begin{aligned} \|d\| &= \|[F'(z)]^{-1}\left(-F_{\mu^+}(z) + r_{\mu^+}\right)\| \\ &\leq \|[F'(z)]^{-1}\|\left[\| - F_\mu(z) + (\mu^+ - \mu)(0, e, 0, 0)^T\| + \|r_{\mu^+}\|\right] \\ &\leq \|[F'(z)]^{-1}\|\left[\| - F_\mu(z)\| + \|(\mu^+ - \mu)(0, e, 0, 0)^T\| + \|r_{\mu^+}\|\right] \\ &\leq M\left(\epsilon_\mu + 2\sqrt{t}\mu + \|r_{\mu^+}\|\right). \end{aligned}$$

Substituting this in (3.14) yields

$$\begin{aligned} \frac{s_i^+}{s_i} &\geq \frac{\mu^+ - M_3(\epsilon_\mu + \mu)^2 - M_4\|r_{\mu^+}\|}{\lambda_i s_i} \\ &\geq \frac{\mu^+ - \max\{M_3, M_4\}\left[(\epsilon_\mu + \mu)^2 + \|r_{\mu^+}\|\right]}{\lambda_i s_i} \\ &= \left(\frac{\mu^+}{\lambda_i s_i}\right)\left[1 - \max\{M_3, M_4\}\frac{(\mu + \epsilon_\mu)^2 + \|r_{\mu^+}\|}{\mu^+}\right], && (3.15) \end{aligned}$$

where $M_3 = 2M^2 \max\{4t, 1\}$, and $M_4 = 1 + 2M^2 \sup(\epsilon_\mu + \sqrt{t}\mu)$. If (3.6) holds, then we can assume that

$$\max\{M_3, M_4\}\frac{(\mu + \epsilon_\mu)^2 + \|r_{\mu^+}\|}{\mu^+} \leq \sigma.$$

Substituting this into (3.15), we have

$$\frac{s_i^+}{s_i} \geq \left(\frac{\mu^+}{\lambda_i s_i}\right)(1 - \sigma) \geq \left(\frac{\mu^+}{\lambda_i s_i}\right)(1 - \sigma)^2.$$

Now, by (1.12) and the stopping rule

$$\lambda_i s_i \leq \mu + \|F_\mu(z)\| < \mu + \epsilon_\mu, \qquad (3.16)$$

so that

$$\frac{s_i^+}{s_i} \geq \left(\frac{\mu^+}{\mu + \epsilon_\mu}\right)(1 - \sigma)^2. \qquad (3.17)$$

This and (3.8) show that the fraction to the boundary rule is satisfied.

$\square$

4. Solving the Barrier Problem in One Step

We now describe conditions that ensure that the barrier stopping test is satisfied after just one step $z^+ = z + d$, where d is defined by (1.16). We assume that z satisfies $\|F_\mu(z)\| \le \epsilon_\mu$, and therefore that at z the barrier parameter is reduced to μ^+.

Using Lemma 2.4 and Theorem 3.1 we have that for all $z \in N(z^*)$ and $\mu^+ < \bar{\mu}$

$$\|F_{\mu^+}(z^+)\| \le 2QM'(\epsilon_\mu + \mu)^2 + 2QM\|r_{\mu^+}\|.$$

The barrier stopping test will be satisfied if the right hand side is less than ϵ_{μ^+}. Since

$$2QM'(\epsilon_\mu + \mu)^2 + 2QM\|r_{\mu^+}\| \le \max\{2QM', 2QM\} \left(\frac{(\epsilon_\mu + \mu)^2 + \|r_{\mu^+}\|}{\epsilon_{\mu^+}} \right) \epsilon_{\mu^+},$$

we can simply impose the condition

$$\frac{(\epsilon_\mu + \mu)^2 + \|r_{\mu^+}\|}{\epsilon_{\mu^+}} \to 0. \tag{4.1}$$

Suppose, for example that the tolerance ϵ_μ is proportional to μ, i.e. $\epsilon_\mu = \theta\mu$, $\epsilon_{\mu^+} = \theta\mu^+$ for some positive constant θ. Then (4.1) is equivalent to

$$\frac{\mu^2 + \|r_{\mu^+}\|}{\mu^+} \to 0. \tag{4.2}$$

Note that this limit will not hold if the barrier parameter decreases quadratically.

Finally, we consider the variant to the algorithm where, after computing x^+ and s^+ from (1.16) we compute least-squares Lagrange multiplier estimates $\lambda_E(x^+, s^+)$ and $\lambda_I(x^+, s^+)$ by minimizing $\|F_{\mu^+}\|$. It is clear that the stopping test is still satisfied since the least squares criterion implies

$$\|F_{\mu^+}(x^+, s^+, \lambda_E(x^+, s^+), \lambda_I(x^+, s^+))\| \le \|F_{\mu^+}(z^+)\|.$$

5. Practical Implementations

We would like the step of the interior point algorithm to satisfy the following three conditions near the solution z^* of the nonlinear programming problem: (i) it should keep the slack variables sufficiently positive, (ii) it should satisfy the barrier stopping test, (iii) it should provide a superlinear rate of convergence.

We have seen in Theorem 3.2 that feasibility is obtained if

$$\frac{(\mu + \epsilon_\mu)^2 + \|r_{\mu^+}\|}{\mu^+} \to 0, \tag{5.1}$$

and

$$\limsup\ (1-\tau)\left(\frac{\mu+\epsilon_\mu}{\mu^+}\right) < 1. \tag{5.2}$$

In the previous section we have shown that the barrier stopping test is satisfied after one step if (4.1) holds:

$$\frac{(\epsilon_\mu+\mu)^2+\|r_{\mu^+}\|}{\epsilon_{\mu^+}} \to 0. \tag{5.3}$$

We have also shown in Theorem 2.5 that superlinear convergence is obtained if

$$\mu^+ = o(\|F(z)\|), \qquad \|r_{\mu^+}\| \le M_1\|F_{\mu^+}(z)\|^{1+\alpha}, \quad \alpha > 0. \tag{5.4}$$

Therefore the main goal of this section is to find strategies that will satisfy (5.1)-(5.4). The free parameters in the algorithm are ϵ_μ (for the barrier stopping test), $\|r_\mu\|$ (for controlling the inexactness of the step computation), τ (for controlling the fraction to the boundary), and the rate at which the barrier parameter μ is decreased. For simplicity we will assume that $\epsilon_\mu, \epsilon_{\mu^+}$ are given by

$$\epsilon_\mu = \theta\mu, \quad \epsilon_{\mu^+} = \theta\mu^+, \quad \theta \in [0, \sqrt{t}). \tag{5.5}$$

In this case (5.1) and (5.3) are both implied by the condition

$$\frac{\mu^2+\|r_{\mu^+}\|}{\mu^+} \to 0, \tag{5.6}$$

and (5.2) becomes

$$\limsup\ \left(\frac{\mu}{\mu^+}\right)(1-\tau)(1+\theta) < 1. \tag{5.7}$$

We can define μ^+ as a function of the previous barrier parameter, $\mu^+ = \mu^+(\mu)$, or as a function of the current iterate, $\mu^+ = \mu^+(z)$. We will consider both options. Before we describe superlinearly convergent strategies we consider the following parameter settings based on a linear decrease in μ, and which are similar to those used in [2].

Strategy 1: (Linear decrease of μ.)

We select $\theta = 1$ in (5.5), and set

$$\mu^+ = \gamma\mu, \quad \gamma \in (0,1), \qquad \tau \in (0,1);$$

i.e. μ decreases by a constant factor at every iteration, and τ is constant. We control the residual by means of

$$\|r_{\mu^+}\| \le M_1\mu^{1+\alpha}, \quad M_1 > 0,\ \alpha > 0.$$

Let us verify that this ensures that the step is feasible and satisfies the barrier stopping test. Since $\mu \to 0$, we have

$$\begin{aligned}\frac{\mu^2+\|r_{\mu^+}\|}{\mu^+} &\le \frac{\mu^2+M_1\mu^{1+\alpha}}{\gamma\mu} \\ &= \frac{1}{\gamma}[\mu+M_1\mu^\alpha] \to 0.\end{aligned} \tag{5.8}$$

Therefore (5.6) is satisfied. Also since

$$\left(\frac{\mu}{\mu^+}\right)(1-\tau)(1+\theta) = \frac{(1-\tau)(1+\theta)}{\gamma}, \tag{5.9}$$

we can satisfy (5.7) by forcing the right hand side of (5.9) to be strictly less than 1. This can be done, for example, by choosing $\theta = 1$, $\tau = 0.995$ and $\gamma = 0.2$, which are the parameter settings in [2]. This strategy will therefore produce a step that, near the solution, keeps the slack sufficiently positive and satisfies the barrier stopping test. The rate of convergence will, however, be only linear.

Strategy 2: (μ^+ as a function of μ.)

Here we set

$$\mu^+ = \mu^{1+\delta}, \quad 0 < \delta < 1, \qquad \tau = 1 - \mu^\beta, \quad \beta > \delta,$$

and control the residual by

$$\|r_{\mu^+}\| = \mu^{1+\alpha}, \quad \alpha > \delta. \tag{5.10}$$

We first verify that conditions (5.6) and (5.7) are satisfied. Since $0 < \delta < \beta$ and $\mu \to 0$, we have

$$\left(\frac{\mu}{\mu^+}\right)(1-\tau)(1+\theta) = (1+\theta)\mu^{\beta-\delta} \to 0.$$

Also since $\alpha > \delta$, $\delta < 1$, and $\mu \to 0$, we have

$$\begin{aligned} \frac{\mu^2 + \|r_{\mu^+}\|}{\mu^+} &= \frac{\mu^2 + \mu^{1+\alpha}}{\mu^{1+\delta}} \\ &\leq \mu^{1-\delta} + \mu^{\alpha-\delta} \to 0. \end{aligned}$$

Therefore we have ensured feasibility and satisfaction of the barrier stopping test.

Now we show that superlinear convergence is obtained. Since the stopping test (1.11) is satisfied at z with $\epsilon_\mu = \theta\mu$, we have

$$\begin{aligned} \|F(z)\| &\geq \|\mu(0,e,0,0)^T\| - \|F_\mu(z)\| \\ &\geq \sqrt{t}\mu - \epsilon_\mu \\ &= (\sqrt{t} - \theta)\mu. \end{aligned} \tag{5.11}$$

Hence $\mu = O(\|F(z)\|)$, and since $\mu^+ = o(\mu)$, we have that the first of the conditions in (5.4) is satisfied. Further by (5.10) and (5.11) we have that

$$\|r_{\mu^+}\| \leq \left(\frac{\|F(z)\|}{\sqrt{t} - \theta}\right)^{1+\alpha}. \tag{5.12}$$

We also have that

$$\begin{aligned} \|F_{\mu^+}(z)\| &\geq \|F(z)\| - \|\mu^+(0,e,0,0)^T\| \\ &= \|F(z)\| - \sqrt{t}\mu^+ \\ &= \|F(z)\|\left(1 - \frac{\sqrt{t}\mu^+}{\|F(z)\|}\right). \end{aligned}$$

By the first of the conditions in (5.4), we can assume that

$$\frac{\sqrt{t}\mu^+}{\|F(z)\|} < \frac{1}{2},$$

so that

$$\|F_{\mu^+}(z)\| \geq \frac{\|F(z)\|}{2}.$$

Using this in (5.12), we have

$$\|r_{\mu^+}\| \leq \left(\frac{2}{\sqrt{t}-\theta}\right)^{1+\alpha} \|F_{\mu^+}(z)\|^{1+\alpha},$$

which completes the proof of the second of the conditions in (5.4). Therefore, the step is superlinearly convergent.

The next strategy seems to be more attractive because the barrier parameter is selected according to how much reduction one has made in the optimality conditions of the nonlinear program.

Strategy 3: (μ^+ is a function of z.)

Choose

$$\mu^+ = \|F(z)\|^{1+\delta}, \quad 0 < \delta < 1, \qquad \tau = 1 - \|F(z)\|^\beta, \quad \beta > \delta,$$

and control the residual by means of

$$\|r_{\mu^+}\| = \|F_{\mu^+}(z)\|^{1+\alpha}, \quad \alpha > \delta.$$

We first verify that (5.6) holds. Since the barrier stopping test is satisfied at z with $\epsilon_\mu = \theta\mu$, we have

$$\begin{aligned} \|F(z)\| &\leq \|F_\mu(z)\| + \|\mu(0,e,0,0)^T\| \\ &\leq \epsilon_\mu + \sqrt{t}\mu \\ &\leq 2\sqrt{t}\mu. \end{aligned} \tag{5.13}$$

Using this and (2.12) we have

$$\begin{aligned} \frac{\|r_{\mu^+}\|}{\mu^+} &\leq \frac{2^{1+\alpha}\|F(z)\|^{1+\alpha}}{\|F(z)\|^{1+\delta}} \\ &= 2^{1+\alpha}\|F(z)\|^{\alpha-\delta} \\ &\leq 2^{1+\alpha}(2\sqrt{t})^{\alpha-\delta}\mu^{\alpha-\delta}. \end{aligned} \tag{5.14}$$

This converges to zero as $\mu \to 0$, because $\alpha > \delta$. Using (5.11) and recalling that $\delta < 1$, we also have that

$$\begin{aligned} \frac{\mu^2}{\mu^+} &= \frac{\mu^2}{\|F(z)\|^{1+\delta}} \\ &\leq \frac{1}{(\sqrt{t}-\theta)^{1+\delta}}\mu^{1-\delta} \to 0, \end{aligned}$$

which, together with (5.14), gives (5.6).

Next, from (5.13) and $\beta > \delta$, we have

$$\begin{aligned} \left(\frac{\mu}{\mu^+}\right)(1-\tau)(1+\theta) &= \mu \|F(z)\|^{\beta-1-\delta}(1+\theta) \\ &\leq (2\sqrt{t})^{\beta-1-\delta}\mu^{\beta-\delta}(1+\theta) \to 0, \end{aligned}$$

which gives (5.7).

It is easy to see that (5.4) holds. In conclusion, Strategy 3 ensures that the step is feasible, that the barrier stopping test is satisfied and that the rate of convergence is superlinear.

Rather than basing the choice of the barrier parameter on the norm of the optimality conditions, $\|F(z)\|$, we can make it depend only on the portion corresponding to complementarity. More specifically, if the stopping tolerance is given by (5.5) with $\theta < 1$, and if τ and r_{μ^+} are defined as in Strategy 3, then we could set

$$\mu^+ = (\max_{1\leq i\leq t} \lambda_i s_i)^{1+\delta}, \quad \text{or} \quad \mu^+ = [(\sum_{i=1}^{t} \lambda_i s_i)/t]^{1+\delta}. \tag{5.15}$$

These are similar to some of the rules used in linear programming algorithms [13]. That (5.15) ensures feasibility of the step, satisfaction of the barrier stooping test, and superlinear convergence follows from the fact that (1.11) forces each term $\lambda_i s_i$ to be of magnitude μ; it is then simple to verify that the arguments given for Strategy 3 hold.

6. Final Remarks

We have presented sufficient conditions for achieving superlinear convergence in the context of the interior point method described in [2]. The analysis differs from [12, 4] in that we wish to ensure that the step generated by the algorithm keeps the slacks sufficiently positive (so that a backtracking line search is therefore not needed) and in that we focus on how to satisfy an inner barrier stopping test.

Our analysis assumes that the regularity conditions stated in Assumptions I hold. We have taken care, however, to make specific reference to the constants M, Q, L, C, which characterize the condition number of the KKT system and other problem characteristics. This will enable us, in a future study, to investigate the effect that degeneracy, near-degeneracy and ill-conditioning have on various aspects of the algorithm.

7. *

References

[1] R.H. Byrd, J.C. Gilbert, and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming", Technical Report OTC 96/02, Optimization Technology Center, Northwestern University (1996).

[2] R.H. Byrd, M.E. Hribar, and J. Nocedal, "An interior point algorithm for large scale nonlinear programming", Technical Report OTC 97/05, Optimization Technology Center, Northwestern University (1997).

[3] A.R. Conn, N.I.M. Gould and Ph.L. Toint, "A primal-dual algorithm for minimizing a non-convex function subject to bound and linear equality constraints", Report RC 20639, IBM T.J. Watson Research Center, Yorktown Heights, New York, 1997.

[4] 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", J. Optim. Theory. Appl., 89, pp. 507–5451, 1996.

[5] R. Fletcher, "Practical Methods of Optimization", Second Edition, John Wiley & Sons, New York, 1990

[6] A.V. Fiacco and G.P. McCormick, "Nonlinear Programming: Sequential Unconstrained Minimization Techniques", Wiley & Sons, 1968.

[7] A. Forsgren and P.E. Gill, "Primal-dual interior methods for nonconvex nonlinear programming", Technical Report, University of California at San Diego, 1996.

[8] D.M. Gay, M.L. Overton and M.H. Wright, "A primal-dual interior point method for non-convex nonlinearly constrained optimization", to appear.

[9] P.E. Gill, W. Murray, and M.H. Wright, "Practical Optimization", Academic Press, Inc., 1981.

[10] J.M. Ortega and W.C. Rheinboldt, "Iterative Solution of Nonlinear Equations in Several Variables", Academic Press, New York, 1970.

[11] T. Steihaug, "The conjugate gradient method and trust regions in large scale optimization", *SIAM J. Numer. Anal*, 20, pp. 626-637, 1983.

[12] H. Yamashita and H. Yabe, "Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization", Mathematical Programming, 75, pp. 377–397, 1996.

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