Journal of Computational Mathematics, Vol.20, No.3, 2002, 225–244.

AN INTERIOR TRUST REGION ALGORITHM FOR NONLINEAR MINIMIZATION WITH LINEAR CONSTRAINTS*[1)]

Jian-guo Liu
(*Department of Mathematics, University of North Texas, Denton, TX 76203, USA*)

Abstract

An interior trust-region-based algorithm for linearly constrained minimization problems is proposed and analyzed. This algorithm is similar to trust region algorithms for unconstrained minimization: a trust region subproblem on a subspace is solved in each iteration. We establish that the proposed algorithm has convergence properties analogous to those of the trust region algorithms for unconstrained minimization. Namely, every limit point of the generated sequence satisfies the Krush-Kuhn-Tucker (KKT) conditions and at least one limit point satisfies second order necessary optimality conditions. In addition, if one limit point is a strong local minimizer and the Hessian is Lipschitz continuous in a neighborhood of that point, then the generated sequence converges globally to that point in the rate of at least 2-step quadratic. We are mainly concerned with the theoretical properties of the algorithm in this paper. Implementation issues and adaptation to large-scale problems will be addressed in a future report.

Key words: Nonlinear programming, Linear constraints, Trust region algorithms, Newton methods, Interior algorithms, Quadratic convergence.

1. Introduction

Consider the following linearly constrained minimization problem:

$$\min_{x\in\Re^n} f(x) : \Re^n \Rightarrow \Re \tag{1.1}$$

$$\text{subject to } Ax = b \text{ and } x \geq 0,$$

where f is assumed to be twice continuously differentiable, $A \in \Re^{m\times n}$, and $b \in \Re^m$. We are interested in locating a local minimizer and call such a minimizer a solution to (1.1).

We propose an interior trust region based algorithm. Starting from a strictly feasible point x^0 (or, interior point, i.e., $Ax^0 = b$ and $x^0 > 0$), a sequence $\{x^k\}$ is generated and every x^k remains interior. In each iteration, a trust region subproblem is solved and the iterate x^k is updated. A projected gradient is used as a watch dog to guarantee global convergence. Under certain assumptions, we establish that the algorithm has convergence properties analogous to those of the trust region algorithms for unconstrained minimization (see, e.g., [28]). Namely,

* Received March 23, 1999; Final revised September 26, 2001.
1) Research partially supported by the Faculty Research Grant RIG-35547 and ROG-34628 of the University of North Texas, and in part by the Cornell Theory Center which receives major funding from the National Science Foundation and IBM Corporation, with additional support from New York State and members of its Corporate Research Institute. This work was partially done while the author was visiting Cornell University and the University of California, San Diego.

every limit point of the generated sequence satisfies the Krush-Kuhn-Tucker (KKT) conditions and at least one limit point satisfies second order necessary optimality conditions. In addition, if one limit point is a strong local minimizer and the Hessian of f is Lipschitz continuous in a neighborhood of that point, then the sequence converges globally to that point and the rate of convergence is at least 2-step quadratic.

Trust region algorithms for unconstrained minimization have been studied by many authors (see, e.g., [14], [21], [23], [27], and [28]). A trust region algorithm can be briefly described as follows.

Let x be the current approximation to a solution of $\min_x f(x)$ and let $\delta > 0$ be the current trust region radius. A solution to the model trust region subproblem

$$\min_{\Delta x} \{q(\Delta x) := f + \nabla f^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f \Delta x \; : \; \|\Delta x\| \le \delta\} \tag{1.2}$$

is computed. If the actual reduction on the objective function, $f(x) - f(x+\Delta x)$, is satisfactory comparing with $f(x) - q(\Delta x)$, the reduction predicted by the quadratic model q in (1.2), $x + \Delta x$ is taken as the next approximation, and the trust region radius δ is updated for the next iteration. Otherwise the trust region radius is reduced and a new trust region subproblem is solved. A reduction is satisfactory if

$$\frac{f(x) - f(x + \Delta x)}{f(x) - q(\Delta x)} \ge \eta,$$

where $\eta \in (0, 1)$ is a given constant.

Excellent global convergence properties of the trust region algorithms for unconstrained minimization have been established. In [28] for example, it is shown that every limit point of the generated sequence satisfies first order conditions and at least one limit point satisfies second order necessary optimality conditions. In addition, if there is one limit point which is a strong local minimizer, then the whole sequence converges globally to this point and the rate of convergence is quadratic. In [27], it is proved that if a reasonable strategy for increasing the trust region radius is imposed, then every limit point satisfies second order necessary conditions.

Trust region algorithms have also been applied to minimization problems with simple bounds (see, e.g., [6] and [10]), to equality constrained minimization (see, e.g., [2], [4], [24], and [29]), and to linearly constrained minimization (see, e.g., [5], [11], and [15]). Similar convergence results are also obtained.

Among the algorithms for linearly constrained minimization, many are based on the active set method, such as [3], [12], [13], [16], [18], [19], [22], [25], [26], and [30]. But some are of interior point type, e.g., [11] and [15].

Interior algorithms have been extensively studied and applied to many optimization problems, including linear programming, linear complementarity problems, and quadratic programming. A main feature of the interior algorithms is that the number of iterations is not very sensitive to problem dimension. For references, see, e.g., [20], [31], and [32].

This paper is organized as follows. In Section 2, we first give the optimality conditions for (1.1) and then motivate and define our algorithm. Convergence properties are established in Section 3 and 4. Finally, we have some concluding remarks in Section 5.

We use $\nabla f(x)$ to denote the gradient of $f(x)$ and $\nabla^2 f(x)$ to denote the Hessian. Given x^* and x^k, we write $f^* = f(x^*)$, $f^k = f(x^k)$, and use similar notations for ∇f and $\nabla^2 f$. We use superscripts to denote the iteration counts and use subscripts to indicate the indices of vector components. We occasionally drop the superscripts when there is no confusion. The norm $\|\cdot\|$ used in this paper is l_2 norm unless otherwise specified. Sets will be denoted by calligraphic capital letters. Given a vector $x \in \Re^n$, the notation $x > 0$ means $x_i > 0$ for every $1 \le i \le n$ and $x \ge 0$ means $x_i \ge 0$ for every $1 \le i \le n$. We call x *a feasible point* if $Ax = b$ and $x \ge 0$. We call x *an interior point* if x is feasible and $x > 0$, and we call x *a boundary point* if x is feasible and $x_i = 0$ for some $1 \le i \le n$. When $M \in \Re^{n\times n}$ is a square matrix, the notation $M > 0$ indicates that M is positive definite and the notation $M \ge 0$ indicates that M is positive semidefinite. If x denotes a vector, then the corresponding capital letter $X = diag(x)$ will denote the diagonal matrix whose diagonal entries are the components of x. Finally, if $x = [x_1, x_2, \cdots, x_n]^T$ and $M = (m_{ij}) \in \Re^{n\times n}$, then $|x| = [|x_1|, |x_2|, \cdots, |x_n|]^T$ and $|M| = (|m_{ij}|) \in \Re^{n\times n}$.

2. The Algorithm IPTR

In this section, we first introduce the optimality conditions for (1.1). Then we describe how to compute the updating steps and give the motivation. Finally, we define the algorithm.

Conditions for a point $x \in \Re^n$ to be a local minimizer of (1.1) are well-known and a set of numerically verifiable conditions can be phrased as follows (see for example [18]): if x^* is a solution to (1.1), then there exists $w^* \in \Re^m$ such that

$$\begin{aligned} &\text{feasibility:} && Ax^* = b \text{ and } x^* \ge 0, && (2.1)\\ &\text{complementarity:} && X^*(\nabla f^* + A^T w^*) = 0, && (2.2)\\ &\text{sign condition:} && x_i^* = 0 \Longrightarrow (\nabla f^* + A^T w^*)_i \ge 0 \ \ (1 \le i \le n), && (2.3)\\ &\text{positive semidefiniteness:} && p^T \nabla^2 f^* p \ge 0 \ \text{ for every } p \in \mathcal{N}(x^*), && (2.4) \end{aligned}$$

where for a given feasible point x,

$$\mathcal{N}(x) = \{\, p \in \Re^n \ :\ Ap = 0 \ \text{ and } \ p_i = 0 \text{ for every } i \in \mathcal{A}(x)\,\},$$

and

$$\mathcal{A}(x) = \{\, i \ :\ x_i = 0\,\}. \qquad (2.5)$$

Conditions (2.1) – (2.3) are first order necessary conditions and are known as the Karush-Kuhn-Tucker (KKT) conditions. Conditions (2.1) – (2.4) are called the second order necessary conditions. Second-order sufficiency conditions for optimality are obtained by replacing "$\ge$" with "$>$" in (2.3) and (2.4). A point x^* is called a strong local minimizer if x^* satisfies these second order sufficiency conditions.

The two equations in the KKT conditions form a nonlinear system

$$F(x,w) = \begin{bmatrix} X(\nabla f(x) + A^T w) \\ Ax - b \end{bmatrix} = 0 \qquad (2.6)$$

which will be useful in the proof of quadratic convergence.

2.1. Projected Trust Region Subproblem and Projected Gradient

In our algorithm, each updating step involves two vectors: the solution of a projected trust region subproblem and a projected gradient. The projected trust region subproblem is motivated by the Newton system of (2.6): its solution is ultimately the Newton direction (with respect to x) of (2.6) (see Section 4). More specifically, for given x, we define

$$\begin{aligned} g &= \nabla f(x) + A^T w, \quad \text{(where } w \text{ is defined later in (2.21))} && (2.7)\\ M &= \nabla^2 f(x) + X^{-1}|G|, \quad \text{(where } |G| = diag(|g_1|, |g_2|, \cdots, |g_n|)) && (2.8) \end{aligned}$$

and solve

$$\min_{\Delta x} \{\psi(\Delta x) = \frac{1}{2}\Delta x^T M \Delta x + \Delta x^T \nabla f(x) \,:\, A\Delta x = 0, \|X^{-\frac{1}{2}}\Delta x\| \le \delta\}, \tag{2.9}$$

where for some given $\delta_u > 0$, $\delta \in (0, \delta_u]$ is updated in each iteration subject to a ratio test similar to one used in trust region algorithms for unconstrained minimization (e.g., [21]) and for minimization with simple bounds (e.g., [6]).

Let Δx_{tr} denote the solution of (2.9) (the subscript *tr* stands for *trust region*). Due to the choices of M in (2.8) and the scaling in (2.9), the solution Δx_{tr} has two important properties. First, it is a feasible descent direction for f subject to the constraints (i.e., $A\Delta x_{tr} = 0$ and $\nabla f(x)^T \Delta x_{tr} < 0$, see next section for the proof). This is important for global convergence. Second, Δx_{tr} is ultimately the Newton direction (with respect to x) for system (2.6). Therefore, local quadratic convergence can be expected.

Problem (2.9) can be described in a different way. Let

$$D = X^{\frac{1}{2}}, \quad \bar{A} = AD, \quad \bar{M} = DMD, \quad \text{and} \quad \bar{g} = Dg. \tag{2.10}$$

We may write Δx_{tr} as

$$\Delta x_{tr} = D\bar{Z}\Delta\bar{x}_{tr}, \tag{2.11}$$

where $\bar{Z} = \bar{Z}(x)$ is a matrix whose columns form an orthonormal basis for the null space of $\bar{A}$ and $\Delta\bar{x}_{tr}$ denotes the solution to

$$\min_{\Delta\bar{x}} \{\bar{\psi}(\Delta\bar{x}) = \frac{1}{2}\Delta\bar{x}^T \bar{Z}^T \bar{M} \bar{Z} \Delta\bar{x} + \Delta\bar{x}^T \bar{Z}^T \bar{g} \,:\, \|\Delta\bar{x}\| \le \delta\}. \tag{2.12}$$

Clearly,

$$\psi(\Delta x_{tr}) = \bar{\psi}(\Delta\bar{x}_{tr}). \tag{2.13}$$

From the results in [14] and [28], there exists $\lambda_{tr} \ge 0$ such that

$$\begin{aligned} (\bar{Z}^T\bar{M}\bar{Z} + \lambda_{tr} I)\Delta\bar{x}_{tr} &= -\bar{Z}^T\bar{g}, && (2.14)\\ \bar{Z}^T\bar{M}\bar{Z} + \lambda_{tr} I &\ge 0, && (2.15)\\ \lambda_{tr}(\delta - \|\Delta\bar{x}_{tr}\|) &= 0, && (2.16) \end{aligned}$$

and

$$\lambda_{tr} \le \frac{\|\bar{g}\|}{\delta} + \|\bar{M}\|. \tag{2.17}$$

By the definition of $\bar{Z}$, there exists Δw_{tr} such that

$$(\bar{M} + \lambda_{tr} I)\bar{Z}\Delta\bar{x}_{tr} + \bar{g} + \bar{A}^T \Delta w_{tr} = 0. \tag{2.18}$$

Or equivalently,

$$(M+\lambda_{tr}X^{-1})\Delta x_{tr}+g+A^T\Delta w_{tr}=0 \text{ and } A\Delta x_{tr}=0. \tag{2.19}$$

When $\bar{A}\bar{A}^T$ is nonsingular,

$$\Delta w_{tr}=-(\bar{A}\bar{A}^T)^{-1}\bar{A}\,(\,(\bar{M}+\lambda_{tr}I)\bar{Z}\Delta\bar{x}_{tr}+\bar{g}\,). \tag{2.20}$$

In our algorithm, Δw_{tr} will not be computed but it will be useful in the convergence analysis.

The matrix $\bar{Z}$ in (2.12) is not specified. In this paper we assume $\bar{Z}(x)$ is continuous in an appropriate region. This may not be true for an arbitrary choice of $\bar{Z}$. For discussions on this regard see [1], [9], and [17].

The step Δx_{tr} will be calculated by (2.11) and (2.12) instead of (2.9) since some entries of the matrix M may approach infinity as some diagonal components of X go to zero. The matrix $\bar{M}$ does not have this disturbing property.

In order to ensure that Δx_{tr} converges to the Newton step (with respect to x) of system (2.6), we need to update w appropriately. A reasonable way to update w is based on condition (2.2), i.e., at a local minimizer

$$X\,(\nabla f(x)+A^Tw)=0.$$

Therefore, if AXA^T is nonsingular, we may compute w by

$$w=-(AXA^T)^{-1}AX\,\nabla f(x). \tag{2.21}$$

When w is computed by (2.21), it is a function of x.

Trust region algorithms have strong convergence properties and have exhibited robust performance in unconstrained minimization (see [14], [21], [27], and [28]). In our case, however, Δx_{tr} may not always be a good choice for updating step due to the constraints. Similar to the *dogleg* algorithm [23] and the algorithms in [6] and [8], we follow a hybrid strategy. We choose the step by combining Δx_{tr} with a projected gradient defined by

$$\Delta x_g=\mu_g D\frac{\bar{g}}{\|\bar{g}\|}, \tag{2.22}$$

where μ_g is the solution to the following one-dimensional problem:

$$\min_{\mu}\{\psi_g(\mu)=\frac{\mu^2}{2}\frac{\bar{g}^T}{\|\bar{g}\|}\bar{M}\frac{\bar{g}}{\|\bar{g}\|}+\mu\frac{\bar{g}^T}{\|\bar{g}\|}\bar{g}\;:\;|\mu|\le\delta\,\}. \tag{2.23}$$

The solution μ_g is given by

$$\mu_g=\begin{cases}-\delta & \text{if } \bar{g}^T\bar{M}\bar{g}\le 0,\\ \max(-\delta,-\frac{\|\bar{g}\|^3}{\bar{g}^T\bar{M}\bar{g}}) & \text{otherwise.}\end{cases} \tag{2.24}$$

In addition, similar to (2.14) – (2.16), there exists $\lambda_g\ge 0$ such that

$$(\tfrac{\bar{g}^T}{\|\bar{g}\|}\bar{M}\tfrac{\bar{g}}{\|\bar{g}\|}+\lambda_g I)\mu_g=-\tfrac{\bar{g}^T}{\|\bar{g}\|}\bar{g}, \tag{2.25}$$

$$\tfrac{\bar{g}^T}{\|\bar{g}\|}\bar{M}\tfrac{\bar{g}}{\|\bar{g}\|}+\lambda_g\ge 0, \tag{2.26}$$

$$\lambda_g(\delta-|\mu_g|)=0, \tag{2.27}$$

and

$$\lambda_g \le \frac{\|\bar{g}\|}{\delta} + \|\bar{M}\|. \tag{2.28}$$

It is easy to verify that if w is defined by (2.21) then

$$\psi(\Delta x_g) = \psi_g(\mu_g), \quad A\Delta x_g = 0, \tag{2.29}$$

and Δx_g is a projection of $\nabla f(x)$ onto the null space of A. Moreover, Δx_g is a feasible descent direction for f subject to the constraints (i.e., $A\Delta x_g = 0$ and $\nabla f(x)^T \Delta x_g < 0$). We describe in Section 2.3 how to combine Δx_{tr} and Δx_g to form the updating step. The motivation is that the projected gradient will be used in a way as a watch dog to guarantee global convergence.

2.2. Maintaining Feasibility

Let x^k be the current iterate, an interior point, and let Δx be a feasible descent direction for f at x^k subject to the constraints, i.e., $A\Delta x = 0$ and $\nabla f(x^k)^T \Delta x < 0$. When moving in the direction Δx from x^k, a variable may reach a bound, i.e., $(x^k + \alpha \Delta x)_i = 0$ for some $1 \le i \le n$ and for some $\alpha > 0$. Therefore, to maintain strict feasibility and yet allow the solution (which may have some of the variables at their bounds $x_i^* = 0$) to be approached asymptotically and sufficiently fast, we define the step length as follows.

For each k, let

$$\theta^k = \frac{\|\tilde{X}^k g^k\| + |\psi^k(\Delta x_{tr}^k)|}{1 + \|\tilde{X}^k g^k\| + |\psi^k(\Delta x_{tr}^k)|}, \tag{2.30}$$

where $\tilde{x}^k$ is a function of x^k defined as follows: for each $i = 1, 2, \cdots, n$,

$$\tilde{x}_i = \begin{cases} x_i & \text{if } g_i \ge 0 \text{ or } x_i > \|\bar{g}\|, \\ -\max(1, x_i) & \text{otherwise.} \end{cases} \tag{2.31}$$

It is clear that $\tilde{X}g = 0$ if and only if the KKT conditions are satisfied with (x, w). In addition, $\tilde{x} = x$ in a small neighborhood of a nondegenerate solution of (1.1) (see Section 3).

Let

$$\sigma^k = \max(\tau_\sigma, 1 - \theta^k) \text{ for some given } \tau_\sigma \in (0, 1), \tag{2.32}$$

$$\beta_{tr}^k = \min_{1 \le i \le n} \{-\frac{x_i^k}{(\Delta x_{tr}^k)_i} > 0\}, \ ((x^k + \beta_{tr}^k \Delta x_{tr}^k)_i = 0 \text{ for some } 1 \le i \le n) \tag{2.33}$$

$$\alpha_{tr}^k = \min(\tau_\alpha, \sigma^k \beta_{tr}^k) \text{ for some given } \tau_\alpha \in (1, 2), \tag{2.34}$$

and

$$\beta_g^k = \min_{1 \le i \le n} \{-\frac{x_i^k}{(\Delta x_g^k)_i} > 0\}, \tag{2.35}$$

$$\alpha_g^k = \min(\tau_\alpha, \sigma^k \beta_g^k). \tag{2.36}$$

The step lengths for Δx_{tr}^k and Δx_g^k will be α_{tr}^k and α_g^k, respectively. In Theorem 4.6, we show that $\theta^k > 0$ unless x^k satisfies the second order necessary conditions (2.1) – (2.4). Therefore, if x^k is an interior feasible point not satisfying (2.1) – (2.4), then both $x^k + \alpha_{tr}^k \Delta x_{tr}^k$ and $x^k + \alpha_g^k \Delta x_g^k$ are interior feasible points. In addition, Theorem 4.6 shows that θ^k will not be very small unless x^k is sufficiently close to optimality. Hence components of x^k will be prevented from prematurely getting too close to zero. Moreover, in Theorem 4.4, we will show that $\theta^k \longrightarrow 0$ and $\alpha_{tr}^k \longrightarrow 1$ as x^k converges to a solution of (1.1), thus the sequence $\{x^k\}$ will be allowed to approach a solution sufficiently fast.

2.3. The Algorithm

We can now state our algorithm.

Algorithm. IPTR

Let x^0 be an interior feasible point. Let $\delta_u > 0$.

Let $\tau_s,\ \epsilon_s \in (0,\,1)$ and $\delta^0 \in (0,\,\delta_u]$ be given.

For $k = 0, 1, 2, \cdots$ until "convergence"

1). Let $D^k = (X^k)^{\frac{1}{2}}$ and compute $w^k = -(AX^kA^T)^{-1}AX^k\,\nabla f^k$;

2). Let $g^k = \nabla f^k + A^Tw^k$ and $\bar{g}^k = D^k\,g^k$;

3). Determine $\tilde{x}^k = \tilde{x}(x^k,\,w^k)$ (see (2.31)) and let $\gamma^k = \frac{\|\,\bar{g}^k\,\|}{\|\,|\tilde{X}^k|^{\frac{1}{2}}\,g^k\,\|}$;

4). Compute Δx_{tr}^k, α_{tr}^k, Δx_g^k, α_g^k and δ^{k+1} by Procedure TR (defined next)

5). If $\frac{\psi^k(\gamma^k\alpha_{tr}^k\Delta x_{tr}^k)}{\psi^k(\gamma^k\alpha_g^k\Delta x_g^k)} \ge \tau_s$ and $\tilde{x}_i^k = x_i^k$ for every $x_i^k \le \epsilon_s$

$$s^k = s_{tr}^k = \gamma^k\alpha_{tr}^k\Delta x_{tr}^k;$$

else

$$s^k = s_g^k = \gamma^k\alpha_g^k\Delta x_g^k;$$

6). Update $x^{k+1} = x^k + s^k$.

Procedure TR.

Let $0 < \eta_1 < \eta_2 < 1$ and $0 < \tau_1 < \tau_2 < 1 < \tau_3$ be given.

a). Compute $\Delta x_{tr}^k = \Delta x_{tr}^k(\delta^k)$ (see (2.11)) and α_{tr}^k (see (2.34)); Let

$$s_{tr}^k = \gamma^k\alpha_{tr}^k\Delta x_{tr}^k;$$

b). Compute $\rho_{tr}^k = \frac{f(x^k+s_{tr}^k)-f(x^k)+\frac{1}{2}(s_{tr}^k)^T(X^k)^{-1}|G^k|s_{tr}^k}{\psi^k(s_{tr}^k)}$;

c). If $\rho_{tr}^k \le \eta_1$ then $\delta^k := \delta \in [\tau_1\delta^k,\,\tau_2\delta^k]$ and go to a);

d). If $\rho_{tr}^k \le \eta_2$ then $\delta^{k+1} \in [\tau_2\delta^k,\,\delta^k]$ else $\delta^{k+1} \in [\delta^k,\ \min(\delta_u,\,\tau_3\delta^k)]$;

e). Compute $\Delta x_g^k = \Delta x_g^k(\delta^k)$ (see (2.22)) and α_g^k (see (2.36)); Let

$$s_g^k = \gamma^k\alpha_g^k\Delta x_g^k.$$

Note that the trust region radius δ^k is updated in Step a) through Step d). Step e) uses the radius δ^k that is determined in Step a) through Step d). The ratio tests are similar to the one used in [6]. In the next two sections, we consider the convergence properties of Algorithm IPTR.

3. Convergence Properties of Algorithm IPTR

In this section we establish the convergence properties of Algorithm IPTR. These convergence properties are proved under the following assumptions:

(A1) The level set $\mathcal{L} = \{x\,:\,x$ is feasible and $f(x) \le f(x^0)\}$ is compact and $f(x)$ is twice continuously differentiable on $\mathcal{L}$.

(A2) AXA^T is nonsingular for every $x \in \mathcal{L}$.

(A3) For every feasible point x satisfying $X(\nabla f(x) + A^T w) = 0$ for some $w \in \Re^m$,
$x_i = 0 \Longrightarrow (\nabla f(x) + A^T w)_i \neq 0 \;\; (1 \leq i \leq n)$.

Assumption **(A2)** is known as primal nondegeneracy. Assumption **(A3)** says that for every feasible point satisfying the complementarity condition, strict complementarity holds. Note that assumptions **(A2)** and **(A3)** are different from the primal and dual nondegeneracy assumptions for linear programming in that **(A2)** and **(A3)** do not restrict a feasible point x satisfying the complementarity condition to be a vertex. That is important since for a nonlinear problem, solutions are not necessarily vertices.

By **(A1)** and **(A2)**, there exists $C_1 > 0$ such that

$$\|(AXA^T)^{-1}\| \leq C_1 \quad \text{for every } \; x \in \mathcal{L}. \tag{3.1}$$

Recall that in each iteration,

$$\delta^k \in (0,\, \delta_u], \tag{3.2}$$

where $\delta_u > 0$ is a given scalar.

The first lemma shows that Procedure TR will terminate in finitely many iterations for every k.

Lemma 3.1. *For every k, Procedure TR terminates in finitely many iterations.*

Proof. By (2.31), it is clear that $\gamma^k \leq 1$ for every k. By **(A1)** and the fact that $\alpha_{tr}^k < 2$ and $\alpha_g^k < 2$, we have $\|\alpha_{tr}^k \Delta x_{tr}^k(\delta^k)\| \leq C_2\, \delta^k$ and $\|\alpha_g^k \Delta x_g^k(\delta^k)\| \leq C_2\, \delta^k$ for some $C_2 > 0$. Moreover, by Taylor's Theorem,

$$f(x+s) - f(x) + \frac{1}{2} s^T X^{-1} |G| s = \psi(s) + \frac{1}{2} s^T (\nabla^2 f(x+ts) - \nabla^2 f(x)) s \tag{3.3}$$

for some $t \in (0,\, 1)$ where $\psi(s)$ is defined in (2.9). Therefore, when δ^k is sufficiently small, the inequality $\rho_{tr}^k > \eta_1$ will be satisfied since f is assumed to be twice continuously differentiable. Then Procedure TR will terminate.

The following lemma defines a few basic equalities used throughout the remainder of this paper.

Lemma 3.2. *Let each Δx_{tr}^k and Δx_g^k be defined by (2.11) and (2.22). Let s^k be defined as in Algorithm IPTR. Then*

$$f(x^k) - f(x^k + s^k) \geq -\eta_1\, \psi^k(s^k) + \frac{1}{2} (s^k)^T (X^k)^{-1} |G^k| s^k; \tag{3.4}$$

$$\begin{aligned} \psi^k(t\, \Delta x_{tr}^k) &= \bar{\psi}^k(t\, \Delta \bar{x}_{tr}^k) \\ &= -t\,(1 - \frac{t}{2})\, (\Delta \bar{x}_{tr}^k)^T ((\bar{Z}^k)^T \bar{M}^k \bar{Z}^k + \lambda_{tr}^k I) \Delta \bar{x}_{tr}^k - \frac{t^2}{2}\, \lambda_{tr}^k\, \|\Delta \bar{x}_{tr}^k\|^2 \leq 0 \\ &\qquad \forall t \in [0,\, \min(2, \beta_{tr}^k)]; \end{aligned} \tag{3.5}$$

$$\begin{aligned} \psi(t\, \Delta x_g^k) &= \psi_g^k(t\, \mu_g^k) \\ &= -t\,(1 - \frac{t}{2})\, (\frac{(\bar{g}^k)^T}{\|\bar{g}^k\|} \bar{M}^k \frac{\bar{g}^k}{\|\bar{g}^k\|} + \lambda_g^k)(\mu_g^k)^2 - \frac{t^2}{2}\, \lambda_g^k\, (\mu_g^k)^2 \leq 0 \\ &\qquad \forall t \in [0,\, \min(2, \beta_g^k)]. \end{aligned} \tag{3.6}$$

Proof. The relation (3.4) is a direct consequence of Lemma 3.1 and the ratio test in Procedure TR, while (3.5) and (3.6) follow by (2.14), (2.15), (2.25), and (2.26).

Lemma 3.3. *For any subsequence $\{k_j\}$ such that $\{\|\bar{g}^{k_j}\|\}$ is bounded away from zero, the subsequence $\{\alpha_g^{k_j}\}$ must be bounded away from zero.*

Proof. Suppose $\|\bar{g}^{k_j}\| \geq \epsilon > 0$. If $\{\alpha_g^{k_j}\}$ is not bounded away from zero then there is a subsequence, still denoted by $\{\alpha_g^{k_j}\}$, that converges to zero. It follows from (2.36) that $\beta_g^{k_j} \longrightarrow 0$. Since n is finite, from the definition of β_g^k (see (2.35)), we may (without loss of generality) assume that $\beta_g^{k_j} = \frac{-x_1^{k_j}}{(\Delta x_g^{k_j})_1}$. Then by (2.22),

$$\frac{\|\bar{g}^{k_j}\|}{\mu_g^{k_j} g_1^{k_j}} = \frac{x_1^{k_j}}{(\Delta x_g^{k_j})_1} \longrightarrow 0. \tag{3.7}$$

Therefore, we must have $\|\bar{g}^{k_j}\| \longrightarrow 0$ since $|\mu_g^{k_j}| \leq \delta_u$ for every k_j and by **(A1)** $\{|g_1^{k_j}|\}$ is bounded above. The assumption that $\{\alpha_g^{k_j}\}$ is not bounded away from zero has led to a contradiction.

The next two theorems are similar to Theorem 4.1 and Theorem 4.5 in [28] and our proofs follow quite similarly to the proofs of those two results.

Theorem 3.4. *Let $\{x^k\}$ be generated by Algorithm IPTR. Then $\{f(x^k)\}$ converges and*

$$\liminf_{k\to\infty}\{\bar{g}^k := D^k(\nabla f^k + A^T w^k)\} = 0. \tag{3.8}$$

Proof. Since $x^{k+1} = x^k + s^k$, by (3.4) and the definition of s^k in Algorithm IPTR,

$$f(x^k) - f(x^{k+1}) \geq -\eta_1\,\tau_s\,\psi^k(\gamma^k \alpha_g^k \Delta x_g^k) \geq 0, \tag{3.9}$$

where η_1, τ_s, and γ^k are given in Algorithm IPTR. Hence $\{f(x^k)\}$ is monotonically decreasing. Therefore, $\{f(x^k)\}$ converges by **(A1)**.

The proof of (3.8) is by contradiction. Suppose (3.8) is false, then there exists $\epsilon > 0$ such that

$$\|\bar{g}^k\| \geq \epsilon. \tag{3.10}$$

Using (3.9), (3.6), and letting $y^k = \frac{\bar{g}^k}{\|\bar{g}^k\|}$, we have

$$\gamma^k \alpha_g^k\,(1 - \frac{\gamma^k \alpha_g^k}{2})\,((y^k)^T \bar{M}^k y^k + \lambda_g^k)(\mu_g^k)^2 \longrightarrow 0. \tag{3.11}$$

From (2.25),

$$((y^k)^T \bar{M}^k y^k + \lambda_g^k)\mu_g^k = -\|\bar{g}^k\|. \tag{3.12}$$

By Lemma 3.3, $\{\alpha_g^k\}$ is bounded away from zero. Therefore it follows from (3.11) and (3.12) that $-\mu_g^k\|\bar{g}^k\| = ((y^k)^T \bar{M}^k y^k + \lambda_g^k)(\mu_g^k)^2 \longrightarrow 0$. That would give $\mu_g^k \to 0$ which, by (3.12), (3.10), and Assumption **(A1)**, implies that $\{\lambda_g^k\}$ is unbounded above. Using (2.27), we see that

$$\delta^k = -\mu_g^k \to 0. \tag{3.13}$$

By (2.14), we have

$$\delta^k \geq \|\Delta \bar{x}_{tr}^k\| \geq \frac{\|(\bar{Z}^k)^T \bar{g}^k\|}{\|(\bar{Z}^k)^T \bar{M}^k \bar{Z}^k\| + \lambda_{tr}^k} \to 0.$$

Therefore we must have either $\|(\bar{Z}^k)^T\bar{g}^k\| \to 0$ or $\lambda_{tr}^k \to \infty$.

Suppose $\|(\bar{Z}^k)^T\bar{g}^k\| \to 0$ and assume $x^k \to x_*$ (or we may consider a subsequence). Then $\bar{Z}_*^T\bar{g}_* = \bar{Z}(x_*)^T\bar{g}(x_*) = 0$. Since $\bar{g}_* = D_*(\nabla f_* + A^T w_*) = D_*\nabla f_* + \bar{A}_*^T w_*$ and $\bar{A}_*\bar{Z}_* = 0$, we have $\bar{Z}_*^T D_*\nabla f_* = 0$. In other words, $D_*\nabla f_* \in null(\bar{Z}_*^T) = range(\bar{A}_*^T)$ which implies that $\bar{A}_*^T\hat{w} = D_*\nabla f_*$ for some $\hat{w} \in \Re^m$. By Assumption **(A2)** and the definition of w in Algorithm IPTR, $\hat{w} = -w_*$. Therefore, $\bar{g}^k \to \bar{g}_* = 0$ which is a contradiction to (3.10).

Now suppose $\lambda_{tr}^k \to \infty$. Then by (3.5) and the fact that $\alpha_{tr}^k < 2$, we have

$$|\psi(\gamma^k\alpha_{tr}^k\Delta x_{tr}^k)| = -\psi(\gamma^k\alpha_{tr}^k\Delta x_{tr}^k) \geq \frac{(\gamma^k\alpha_{tr}^k)^2}{2}\lambda_{tr}^k\|\Delta\bar{x}_{tr}^k\|^2. \tag{3.14}$$

It follows from the definition of ρ_{tr} in Procedure TR, (3.3), (2.11), Assumption **(A1)**, and (3.14) that $|\rho_{tr}^k - 1| \to 0$. Therefore, there exists an $K > 0$ such that $\rho_{tr}^k > \eta_2$ for all $k \geq K$. The updating rule for δ^k in Procedure TR would give $\delta^k \geq \delta_* > 0$. That is a contradiction to $\delta^k \to 0$.

Therefore, (3.8) must be true. That completes the proof.

Theorem 3.5. *Let $\{x^k\}$ be generated by Algorithm IPTR. Then*

$$\bar{g}^k = D^k(\nabla f^k + A^T w^k) \longrightarrow 0. \tag{3.15}$$

Consequently, every limit point x^* of $\{x^k\}$ satisfies

$$X^*(\nabla f^* + A^T w^*) = 0 \ \textit{where}\ w^* = -(AX^*A^T)^{-1}AX^*\nabla f^*. \tag{3.16}$$

Proof. Suppose there is a subsequence $\{x^{k_j}\} \subset \{x^k\}$ such that $\|\bar{g}^{k_j}\| \geq \epsilon > 0$ for all $j = 1, 2, \cdots$. Due to Theorem 3.5, we may select an integer l_j corresponding to each j such that

$$l_j = max\{l \in [k_j, k_{j+1}) \mid \|\bar{g}^i\| \geq \frac{\epsilon}{2},\ k_j \leq i \leq l\} \tag{3.17}$$

and

$$\|\bar{g}^{l_j+1}\| < \frac{\epsilon}{2} \quad \text{(otherwise we may consider a subsequence of } \{x^{k_j}\}\text{)}. \tag{3.18}$$

Using the same arguments from (3.10) to (3.13) and Lemma 3.3, we have

$$\alpha_g^l \geq \epsilon_1 > 0 \ \text{ and }\ \delta^l = -\mu_g^l \to 0 \ \text{ (as } j \to \infty) \ \text{ for all }\ k_j \leq l \leq l_j,\ j = 1, 2, \cdots. \tag{3.19}$$

Therefore, for all $k_j \leq l \leq l_j,\ j = 1, 2, \cdots$,

$$\begin{aligned} f_l - f_{l+1} &\geq \eta_1\tau_s\frac{(\gamma^l\alpha_g^l)^2}{2}\lambda_g^l(\mu_g^l)^2 \ \text{ (by (3.9) and (3.6))} \\ &= \eta_1\tau_s\frac{(\gamma^l\alpha_g^l)^2}{2}(-\|\bar{g}^l\| - \mu_g^l(y^l)^T\bar{M}^l y^l)\mu_g^l \ \text{ (by (3.12))} \\ &\geq \epsilon_2\,\delta^l \ \text{ for some }\ \epsilon_2 > 0 \ \text{ (Assumption } \textbf{(A1)} \text{ is used).} \end{aligned} \tag{3.20}$$

From (3.20) and the convergence of $\{f^k\}$ it follows that

$$f^{k_j} - f^{l_j+1} = \Sigma_{l=k_j}^{l_j}(f_l - f_{l+1}) \geq \epsilon_2\,\Sigma_{l=k_j}^{l_j}\delta^l \to 0 \ \text{ as }\ j \to \infty. \tag{3.21}$$

From (3.21) it follows that $\|x^{k_j} - x^{l_j+1}\| \to 0$ as $j \to \infty$. By Assumption **(A1)** and **(A2)**, $\bar{g}(x)$ is uniformly continuous on $\mathcal{L}$ and it follows that $\|\bar{g}^{k_j} - \bar{g}^{l_j+1}\| < \frac{\epsilon}{2}$ for all j sufficiently large. Therefore

$$\|\bar{g}^{k_j}\| \le \|\bar{g}^{k_j} - \bar{g}^{l_j+1}\| + \|\bar{g}^{l_j+1}\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for all j sufficiently large. The assumption that $\|\bar{g}^{k_j}\| \ge \epsilon$ has led to a contradiction and we must conclude that $\bar{g}^k \to 0$.

Next, we show that every limit point of $\{x^k\}$ satisfies the KKT conditions. We break the proofs into several lemmas. The first lemma says that any two limit points of $\{x^k\}$ have the same zero components.

Lemma 3.6. *Let x^* and y^* be any two limit points of $\{x^k\}$. Then $\mathcal{A}(x^*) = \mathcal{A}(y^*)$, where the set $\mathcal{A}$ is defined in (2.5).*

Proof. Let $\mathcal{S}$ be the set of all limit points of $\{x^k\}$. For each $x \in \mathcal{S}$, define

$$\mathcal{E}(x) = \{\, y \in \mathcal{S} \,:\, \mathcal{A}(y) = \mathcal{A}(x) \,\}.$$

We first show that each $\mathcal{E}(x)$ is a closed set. Suppose $\{y^k\} \subset \mathcal{E}(x)$ and $y^k \longrightarrow y$. It is easy to see that $y \in \mathcal{S}$. For $i \in \mathcal{A}(x)$, $y_i^k = 0$ for every k so $y_i = 0$. For $i \notin \mathcal{A}(x)$, $(\nabla f(y^k) + A^T w(y^k))_i = 0$ for every k. Hence $(\nabla f(y) + A^T w(y))_i = 0$ and therefore, $Y(\nabla f(y) + A^T w(y)) = 0$. Then **(A3)** implies that $y_i \ne 0$ for every $i \notin \mathcal{A}(x)$. Therefore $y \in \mathcal{E}(x)$ and $\mathcal{E}(x)$ is closed.

It is clear that if $y \in \mathcal{E}(x)$, then $\mathcal{E}(y) = \mathcal{E}(x)$. Therefore, the number of distinct elements in the set $\{\,\mathcal{E}(x) \,:\, x \in \mathcal{S}\,\}$ is finite since n is finite. Using the closedness of each $\mathcal{E}(x)$ $(x \in \mathcal{S})$, we see that

$$\epsilon_1 = \inf\{\, \|x - y\| \,:\, x, y \in \mathcal{S},\ \mathcal{E}(x) \ne \mathcal{E}(y) \,\} > 0. \qquad (3.22)$$

Now let x^* be any limit point of $\{x^k\}$. By (3.22), there exists $\epsilon \in (0, \epsilon_1)$ and $k_1 > 0$ such that for every $k \ge k_1$,

$$\text{either } (x_i^k)^{\frac{1}{2}} < \tfrac{\epsilon}{3(1+2\delta_u)} \text{ for every } i \in \mathcal{A}(x^*), \text{ or } x_i^k > \tfrac{2}{3}\epsilon \text{ for some } i \in \mathcal{A}(x^*). \qquad (3.23)$$

Since x^* is a limit point, there exists $k_2 \ge k_1$ such that

$$(x_i^{k_2})^{\frac{1}{2}} < \frac{\epsilon}{3(1+2\delta_u)} \text{ for every } i \in \mathcal{A}(x^*). \qquad (3.24)$$

Notice that for every k, $\|(D^k)^{-1}\Delta x_{tr}^k\| \le \delta_u$, $\|(D^k)^{-1}\Delta x_g^k\| \le \delta_u$, $\gamma^k \le 1$, $\alpha_{tr}^k < 2$, and $\alpha_g^k < 2$, hence

$$\|(D^k)^{-1} s^k\| \le 2\,\delta_u \text{ for every k}, \qquad (3.25)$$

which implies that

$$|s_i^{k_2}| \le 2\,\delta_u\,(x_i^{k_2})^{\frac{1}{2}} \text{ for every } i \in \mathcal{A}(x^*). \qquad (3.26)$$

Then (3.24) and (3.26) yield

$$x_i^{k_2+1} \le (1 + 2\,\delta_u)(x_i^{k_2})^{\frac{1}{2}} \le \frac{\epsilon}{3} \text{ for every } i \in \mathcal{A}(x^*). \qquad (3.27)$$

Therefore, by (3.23),

$$(x_i^{k_2+1})^{\frac{1}{2}} < \frac{\epsilon}{3(1+2\delta_u)} \text{ for every } i \in \mathcal{A}(x^*). \qquad (3.28)$$

By induction, we have that for every $k \geq k_2$,

$$x_i^k \leq (x_i^k)^{\frac{1}{2}} < \frac{\epsilon}{3(1+2\delta_u)} \leq \frac{\epsilon}{3} < \frac{\epsilon_1}{3} \text{ for every } i \in \mathcal{A}(x^*). \tag{3.29}$$

Thus $\mathcal{A}(x^*) \subset \mathcal{A}(x)$ for every $x \in \mathcal{S}$. Since x^* is an arbitrary limit point, the proof is complete.

Again, let $\mathcal{S}$ be the set of all limit points of $\{x^k\}$. Let $sign(t)$ denotes the sign of $t \in \Re$. For each $x \in \mathcal{S}$, define

$$\mathcal{T}(x) = \{\, y \in \mathcal{S} \,:\, sign(g_i(y)) = sign(g_i(x)) \text{ for every } i \in \mathcal{A}(x) \,\}, \tag{3.30}$$

where $g(x) = \nabla f(x) + A^T w(x)$ with $w(x) = -(AXA^T)^{-1}AX\,\nabla f(x)$, and $g(y)$ is similarly defined. Then we have the following result.

Lemma 3.7. *The number of distinct elements in the set* $\{\, \mathcal{T}(x) \,:\, x \in \mathcal{S} \,\}$ *is finite. Consequently,*

$$\epsilon_2 = \inf\{\, \|x-y\| \,:\, x, y \in \mathcal{S},\ \mathcal{T}(x) \neq \mathcal{T}(y) \,\} > 0. \tag{3.31}$$

Proof. Similar to the proof of Lemma 3.6, we can first show that each $\mathcal{T}(x)$ ($x \in \mathcal{S}$) is closed. It is clear that $\mathcal{T}(y) = \mathcal{T}(x)$ if $y \in \mathcal{T}(x)$. Then the lemma follows.

Theoram 3.8. *Let* $\{x^k\}$ *be generated by Algorithm IPTR. Then every limit point* x^* *of* $\{x^k\}$ *satisfies the KKT conditions.*

Proof. First, by **(A1)**, (2.11), and (2.22), there exists $C_3 > 0$ such that for every k,

$$\|s^k\| \leq C_3\, \gamma^k. \tag{3.32}$$

Let x^* be a limit point of $\{x^k\}$. By (3.16), $X^*g^* = 0$ where $g^* = \nabla f(x^*) + A^T w^*$. It suffices to show that $g_i^* \geq 0$ for every $i \in \mathcal{A}(x^*)$.

Assume the contrary, that is, $g_i^* < 0$ for some $i \in \mathcal{A}(x^*)$. Fix i and let

$$\epsilon^* = \min\{\, |g_i(x)| \,:\, x \in \mathcal{T}(x^*) \,\}, \tag{3.33}$$

which is well-defined since $\mathcal{T}(x^*)$ is compact. By **(A3)**, $\epsilon^* > 0$.

By Lemma 3.7 and Theorem 3.5, there exists $\epsilon \in (0, \epsilon_2)$ (ϵ_2 is defined in (3.31)) and $k_1 > 0$ such that for every $k \geq k_1$,

$$\text{either } \|x^k - x\| \leq \tfrac{\epsilon}{3} \text{ for some } x \in \mathcal{T}(x^*) \text{ or } \|x^k - x\| \geq \tfrac{2\epsilon}{3} \text{ for every } x \in \mathcal{T}(x^*), \tag{3.34}$$

and

$$\|\bar{g}^k\| \leq \frac{\epsilon\, \epsilon^*}{12\, C_3}. \tag{3.35}$$

Since $i \in \mathcal{A}(x^*)$, by Lemma 3.6 and the definition of ϵ^* (3.33), there exists $k_2 \geq k_1$ such that if $k \geq k_2$ and $\|x^k - x\| \leq \frac{\epsilon}{3}$ for some $x \in \mathcal{T}(x^*)$, then

$$g_i^k \leq -\frac{\epsilon^*}{2},\ \ (x_i^k)^{\frac{1}{2}} \leq \frac{\epsilon^*}{2},\ \text{ and } x_i^k \leq \epsilon_s. \tag{3.36}$$

For such k, using (2.31) and the first two inequalities in (3.36), we have $\tilde{x}_i^k = -1$. Then using the third inequality in (3.36) and the definitions of γ^k and s^k in Algorithm IPTR, we have

$$s^k = \gamma^k \alpha_g^k \Delta x_g^k \text{ and } \gamma^k \leq \frac{\epsilon}{6\, C_3} \tag{3.37}$$

for every $k \geq k_2$ such that $\|x^k - x\| \leq \frac{\epsilon}{3}$ for some $x \in \mathcal{T}(x^*)$.

Let $k_3 \geq k_2$ such that $\|x^{k_3} - x^*\| \leq \frac{\epsilon}{3}$ (such k_3 exists since x^* is a limit point). Then $\|s^{k_3}\| \leq \frac{\epsilon}{6}$ by (3.32) and the fact that $\gamma^{k_3} \leq \frac{\epsilon}{6\,C_3}$. Hence

$$\|x^{k_3+1} - x\| \leq \|x^{k_3} - x\| + \|s^{k_3}\| \leq \frac{\epsilon}{2}. \tag{3.38}$$

Then (3.34) yields

$$\|x^{k_3+1} - x\| \leq \frac{\epsilon}{3} \text{ for some } x \in \mathcal{T}(x^*).$$

By induction, we have for every $k \geq k_3$,

$$\|x^k - x\| \leq \frac{\epsilon}{3} \leq \frac{\epsilon_2}{3} \text{ for some } x \in \mathcal{T}(x^*).$$

Therefore, $\mathcal{T}(x) = \mathcal{T}(x^*)$ for every $x \in \mathcal{S}$ which implies that $\mathcal{T}(x^*) = \mathcal{S}$. Then $g_i^k < 0$ for every k sufficiently large. Noticing that $\mu_g^k < 0$ by (2.23), we have $(\Delta x_g^k)_i > 0$ by (2.22). Hence (3.37) implies that $s_i^k > 0$ for every k sufficiently large. Therefore, $\{x_i^k\}$ is bounded away from zero, which contradicts the fact that $i \in \mathcal{A}(x^*)$. Hence x^* satisfies the KKT conditions.

Corollary 3.9. *For every k sufficiently large,*

$$\tilde{x}^k = x^k \text{ and } \gamma^k = 1, \tag{3.39}$$

where γ^k is defined in Algorithm IPTR.

Proof. It is a direct consequence of Lemma 3.6, **(A3)**, and Theorem 3.8.

Next, we turn to second order optimality conditions. We show that the sequence $\{x^k\}$ generated by Algorithm IPTR has a limit point satisfying the second order necessary optimality conditions.

Lemma 3.10. *There exists $\bar{\alpha} > 0$ such that $\alpha_{tr}^k \geq \bar{\alpha}$ for every k.*

Proof. If the lemma is false, then there exists a subsequence $\{k_j\}$ such that

$$\alpha_{tr}^{k_j} \longrightarrow 0,$$

which by (2.34) implies $\beta_{tr}^{k_j} \longrightarrow 0$. Since n is finite, from the definition of β_{tr}^k (see (2.33)), we may (without loss of generality) assume that $\beta_{tr}^{k_j} = \frac{-x_1^{k_j}}{(\Delta x_{tr}^{k_j})_1}$. Since $\|\Delta x_{tr}^k\| \leq \delta_u \|D^k\|$ by (2.11), we see by **(A1)** and (2.11) that $\{\|\Delta x_{tr}^k\|\}$ is bounded above. So $x_1^{k_j} \longrightarrow 0$.

Let x^* be a limit point of $\{x^{k_j}\}$. To simplify notations, assume $x^{k_j} \longrightarrow x^*$. Then $w^{k_j} \longrightarrow w^* = -(AX^*A^T)^{-1}AX^*\nabla f(x^*)$ and $x_1^* = 0$. By **(A3)**, $|g_1^*| > 0$. Now multiplying the first equation of (2.19) with X^k gives

$$X^k \nabla^2 f^k \Delta x_{tr}^k + |G^k| \Delta x_{tr}^k + \lambda_{tr}^k \Delta x_{tr}^k + X^k g^k + X^k A^T \Delta w_{tr}^k = 0, \tag{3.40}$$

which yields

$$\frac{|g_1^{k_j}| + \lambda_{tr}^{k_j}}{(\nabla^2 f^{k_j} \Delta x_{tr}^{k_j})_1 + g_1^{k_j} + (A^T \Delta w_{tr}^{k_j})_1} = \frac{-x_1^{k_j}}{(\Delta x_{tr}^{k_j})_1} = \beta_{tr}^{k_j} \longrightarrow 0.$$

But this is impossible since $g_1^{k_j} \longrightarrow g_1^* > 0$ and by **(A1)**, (2.20), and (3.1), there exist C_4, $C_5 > 0$ such that $|(\nabla^2 f^{k_j} \Delta x_{tr}^{k_j})_1 + g_1^{k_j} + (A^T \Delta w_{tr}^{k_j})_1| \leq C_4 + C_5 \lambda_{tr}^{k_j}$ for every k_j. Therefore, the lemma holds.

The following lemma reveals the relationship between the positive definiteness of $\nabla^2 f^*$ in $\mathcal{N}^*$ and that of $(\bar{Z}^*)^T \bar{M}^* \bar{Z}^*$. The proof can be found in [7].

Lemma 3.11. *Let x^* be any feasible point satisfying the complementarity condition (2.2). Then*

(i) $(\bar{Z}^*)^T \bar{M}^* \bar{Z}^* > 0$ *if and only if* $p^T \nabla^2 f^* p > 0$ *for every* $p \in \mathcal{N}^*$ *and* $p \neq 0$.

(ii) $(\bar{Z}^*)^T \bar{M}^* \bar{Z}^* \geq 0$ *if and only if* $p^T \nabla^2 f^* p \geq 0$ *for every* $p \in \mathcal{N}^*$.

The next result is motivated by and similar to Theorem 4.7 in [21].

Theorem 3.12. *The sequence $\{x^k\}$ generated by Algorithm IPTR has a limit point x^* which satisfies the second order necessary conditions (2.1) – (2.4).*

Proof. By Theorem 3.8, it suffices the show that $\{x^k\}$ has a limit point x^* which satisfies (2.4).

Clearly, if there exists a subsequence $\{\lambda_{tr}^{k_j}\} \longrightarrow 0$, then by (2.15), there exists a limit point x^* such that $(\bar{Z}^*)^T \bar{M}^* \bar{Z}^* \geq 0$, and the desired result follows by Lemma 3.11.

We can show that $\{\lambda_{tr}^k\}$ is not bounded away from zero by contradiction. Suppose $\lambda_{tr}^k \geq \epsilon > 0$ for every k sufficiently large. By (3.4), (3.5), (3.39), and the convergence of $\{f(x^k)\}$,

$$f(x^k) - f(x^{k+1}) \geq -\eta_1 \, \psi^k(\alpha_{tr}^k \Delta x_{tr}^k) \geq \eta_1 \, \frac{(\alpha_{tr}^k)^2}{2} \lambda_{tr}^k \|\Delta \bar{x}_{tr}^k\|^2 \longrightarrow 0.$$

Then Lemma 3.10 implies $\|\Delta \bar{x}_{tr}^k\| \longrightarrow 0$. Using (2.16), we have $\delta^k \longrightarrow 0$. By **(A1)** and (2.11), there exists $C_6 > 0$ such that $\|\Delta x_{tr}^k\| \leq C_6 \, \|\Delta \bar{x}_{tr}^k\|$ for every k. So by (3.5),

$$|\psi^k(\alpha_{tr}^k \Delta x_{tr}^k)| \geq \frac{(\alpha_{tr}^k)^2}{2} \lambda^k \|\Delta \bar{x}_{tr}^k\|^2 \geq \frac{\bar{\alpha}^2 \epsilon}{2 C_6^2} \|\Delta x_{tr}^k\|^2.$$

Now, the definition of ρ_{tr}^k in Algorithm IPTR and a standard estimate give (letting $s_{tr}^k = \alpha_{tr}^k \Delta x_{tr}^k$)

$$\begin{aligned} |\rho_{tr}^k - 1| &= |f(x^k + s_{tr}^k) - f(x^k) + \frac{1}{2}(s_{tr}^k)^T (X^k)^{-1} |G^k| s_{tr}^k - \psi^k(s_{tr}^k)| \, / \, |\psi^k(s_{tr}^k)| \qquad (3.41) \\ &\leq \frac{1}{2} \frac{\|s_{tr}^k\|^2}{|\psi^k(s_{tr}^k)|} \max_{0 \leq \xi \leq 1} \|\nabla^2 f(x^k + \xi s_{tr}^k) - \nabla^2 f(x^k)\| \\ &\leq \frac{4 C_6^2}{\bar{\alpha}^2 \epsilon} \max_{0 \leq \xi \leq 1} \|\nabla^2 f(x^k + \xi s_{tr}^k) - \nabla^2 f(x^k)\|, \end{aligned}$$

where we used the fact that $\alpha_{tr}^k < 2$. Therefore, the uniform continuity of $\nabla^2 f$ on $\mathcal{L}$ and the fact that $\|\Delta x_{tr}^k\| \leq C_6 \, \|\Delta \bar{x}_{tr}^k\| \longrightarrow 0$ imply $\rho_{tr}^k > \eta_1$ for all k sufficiently large. Then the updating rules for δ^k yield that $\{\delta^k\}$ is bounded away from zero which is a contradiction to the fact that $\delta^k \longrightarrow 0$. That completes the proof.

Following the proof of Theorem 4.11 in [21], we can easily obtain the next result.

Theorem 3.13. *Every isolated limit point of $\{x^k\}$ satisfies the second order necessary conditions (2.1) – (2.4). Therefore, if $\{x^k\}$ converges to x^* say, then x^* satisfies the second order necessary conditions.*

As remarked in [28], failure of convergence of $\{x^k\}$ will require an extremely pathological situation. It is easy to see that f must have the same value at each of the limit points of

$\{x^k\}$. Moreover, at least one of the limit points satisfies the second order necessary optimality conditions.

4. Quadratic Convergence of Algorithm IPTR

In this section we consider the local convergence rate properties of Algorithm IPTR: we establish superlinear and quadratic convergence results when there exists a limit point which is a strong local minimizer. Recall that under assumptions **(A1)**, **(A2)**, and **(A3)**, a limit point x^* is a strong local minimizer if the following holds:

$$p^T \nabla^2 f(x^*) p > 0 \text{ for every } p \in \mathcal{N}(x^*) \text{ and } p \neq 0. \tag{4.1}$$

We first give some preliminary results.

Lemma 4.1. *Let x^* be any limit point of $\{x^k\}$. Then*

$$v^T \nabla f^* = 0 \ \textit{for every} \ v \in \mathcal{N}(x^*). \tag{4.2}$$

Proof. By (3.16), $X^*(\nabla f^* + A^T w^*) = 0$. Then by **(A3)**, we may define $y^* \in \Re^n$ by $y_i^* = 1$ if $i \in \mathcal{A}(x^*)$ and $y_i^* = 1/x_i^*$ otherwise. Thus $v^T Y^* X^* = v^T$ for every $v \in \mathcal{N}(x^*)$, which implies that

$$v^T \nabla f^* = v^T(\nabla f^* + A^T w^*) = v^T Y^* X^* (\nabla f^* + A^T w^*) = 0.$$

Lemma 4.2. *Every limit point x^* satisfying (4.1) is an isolated limit point.*

Proof. We prove the lemma by contradiction. Assume there is a sequence of limit points $\{x^l\}$ which converges to x^* and $x^l \neq x^*$ for every l. By Lemma 3.6, $\mathcal{A}(x^l) = \mathcal{A}(x^*)$ for every l where $\mathcal{A}$ is defined in (2.5). Then $\mathcal{N}(x^l) = \mathcal{N}(x^*)$ and $x^l - x^* \in \mathcal{A}(x^*)$ for every l. Therefore, by Lemma 4.1, $(x^l - x^*)^T(\nabla f^l - \nabla f^*) = 0$. But on the other hand,

$$\nabla f^l - \nabla f^* = \int_0^1 \nabla^2 f(x^* + t(x^l - x^*))(x^l - x^*)\, dt,$$

and for l sufficiently large, by (4.1) and the continuity of $\nabla^2 f(x)$,

$$(x^l - x^*)^T(\nabla f^l - \nabla f^*) = \int_0^1 [\,(x^l - x^*)^T \nabla^2 f(x^* + t(x^l - x^*))\,(x^l - x^*)\,]\, dt > 0.$$

This is a contradiction to that $(x^l - x^*)^T(\nabla f^l - \nabla f^*) = 0$. Therefore, x^* is an isolated limit point.

Theorem 4.3. *Suppose there exists a limit point x^* which satisfies (4.1). Then $x^k \longrightarrow x^*$.*

Proof. We first establish that

$$|\mu_g^k| + \|\Delta x_g^k\| + \|\Delta \bar{x}_{tr}^k\| + \|\Delta x_{tr}^k\| \longrightarrow 0. \tag{4.3}$$

In fact, by (4.1) and Lemma 3.11, there exists $\epsilon > 0$ such that the least eigenvalue of $(\bar{Z}^k)^T \bar{M}^k \bar{Z}^k$ is greater than or equal to ϵ for every k sufficiently large. Moreover, since $\bar{A}^k \bar{g}^k = 0$, there exists y^k with $\|y^k\| = 1$ such that $\frac{\bar{g}^k}{\|\bar{g}^k\|} = \bar{Z}^k y^k$ for every k. It follows from (2.25) and (3.15) that $\mu_g^k \longrightarrow 0$. Hence $\Delta x_g^k \longrightarrow 0$ by **(A1)** and (2.22). Similarly, we can show that $\Delta \bar{x}_{tr}^k \longrightarrow 0$ and $\Delta x_{tr}^k \longrightarrow 0$. So (4.3) holds. Therefore, $s^k \longrightarrow 0$ and the theorem follows by Lemma 4.2.

Lemma 4.4. *Suppose there exists a limit point x^* which satisfies (4.1). Then each of the following holds.*

(i) *There exists* $\delta_l > 0$ *such that* $\delta^k \geq \delta_l$ *for every* k.

(ii) $\lambda_{tr}^k = 0$ *for every* k *sufficiently large.*

(iii) $\Delta w_{tr}^k \longrightarrow 0$.

(iv) $\theta^k \longrightarrow 0$.

(v) $\alpha_{tr}^k \longrightarrow 1$.

(vi) $s^k = \alpha_{tr}^k \Delta x_{tr}^k$ *for every* k *sufficiently large.*

Proof. Proof of (i). By (4.1) and Lemma 3.11, there exists $\epsilon > 0$ such that the least eigenvalue of $(\bar{Z}^k)^T \bar{M}^k \bar{Z}^k$ is greater than or equal to ϵ for every k sufficiently large. It follows from (3.5), Lemma 3.10, **(A1)**, and (2.11) that $|\psi^k(\alpha_{tr}^k \Delta x_{tr}^k)| \geq C_7 \|\Delta x_{tr}^k\|^2$ for some $C_7 > 0$ and for every k sufficiently large. Therefore, similar to (3.41), we have

$$|\rho_{tr}^k - 1| \longrightarrow 0,$$

which and the updating rules for δ^k imply that $\delta^k \geq \delta_l$ for some $\delta_l > 0$ and for every k.

Proof of (ii). It follows from (2.16), (4.3) (particularly, $\|\Delta \bar{x}_{tr}^k\| \longrightarrow 0$), and (i) of this lemma.

Proof of (iii). It follows from (2.20), (3.15), (4.3), and (ii) of this lemma.

Proof of (iv). It follows from (2.30), (3.5), (4.3), (3.15), and (3.39).

Proof of (v). By definition, $x_i^* > 0$ for every $i \notin \mathcal{A}(x^*)$. So by (4.3)

$$-\frac{x_i^k}{(\Delta x_{tr}^k)_i} \longrightarrow \infty \text{ for every } i \notin \mathcal{A}(x^*).$$

By **(A3)** and Theorem 3.8, $g_i^* > 0$ for every $i \in \mathcal{A}(x^*)$. Using (3.40), (4.3), (ii) and (iii) of this lemma, and the convergence of $\{x^k\}$, we have

$$-\frac{x_i^k}{(\Delta x_{tr}^k)_i} = \frac{|g_i^k| + \lambda_{tr}^k}{(H\Delta x_{tr}^k)_i + g_i^k + (A^T \Delta w_{tr}^k)_i} \longrightarrow 1 \text{ for every } i \in \mathcal{A}(x^*). \tag{4.4}$$

Therefore, by (2.32), (2.33), (2.34), and (iv) of this lemma, we have $\beta_{tr}^k \longrightarrow 1$, $\rho_{tr}^k \longrightarrow 1$, and $\alpha_{tr}^k \longrightarrow 1$.

Proof of (vi). By Corollary 3.9 and the definition of s^k in Algorithm IPTR, we need only to show that

$$\frac{\psi^k(\alpha_{tr}^k \Delta x_{tr}^k)}{\psi^k(\alpha_g^k \Delta x_g^k)} \geq \tau_s \quad \text{for every } k \text{ sufficiently large.} \tag{4.5}$$

In fact, let $\mu_g(\delta)$ be the solution to (2.23). Then $\psi_g(\mu_g(\delta_1)) \leq \psi_g(\mu_g(\delta_2))$ if $\delta_1 \geq \delta_2$. Therefore, if for each k we let δ^k be the trust region radius determined by steps a) to d), and δ_g^k be the trust region radius determined by steps e) to g) in Algorithm IPTR, then

$$\psi_g^k(\mu_g^k(\delta^k)) \leq \psi_g^k(\mu_g^k(\delta_g^k)) \tag{4.6}$$

since $\delta^k \geq \delta_g^k$.

On the other hand, similar to (i) and (ii) of this lemma, it can be shown that $\{\delta_g^k\}$ is bounded away from zero and $\lambda_g^k = 0$ for every k sufficiently large. Hence by (3.6),

$$\psi_g^k(\mu_g^k(\delta_g^k)) \leq \psi_g^k(\alpha_g^k\, \mu_g^k(\delta_g^k)) = \psi^k(\alpha_g^k\, \Delta x_g^k). \tag{4.7}$$

Moreover, since $\Delta\bar{x}_{tr}^k$ solves (2.12) (with δ^k) and $\mu_g^k(\delta^k)$ solves (2.23), we have

$$\bar{\psi}^k(\Delta\bar{x}_{tr}^k) \le \psi_g^k(\mu_g^k(\delta^k)). \tag{4.8}$$

It follows from (4.6), (4.7), and (4.8) that

$$\bar{\psi}^k(\Delta\bar{x}_{tr}^k) \le \psi^k(\alpha_g^k\,\Delta x_g^k).$$

Therefore, by (v) of this lemma and (3.5), (notice that $\psi^k(\alpha_g^k\,\Delta x_g^k) < 0$)

$$\frac{\psi^k(\alpha_{tr}^k\Delta x_{tr}^k)}{\psi^k(\alpha_g^k\Delta x_g^k)} = \frac{\alpha_{tr}^k(2-\alpha_{tr}^k)\bar{\psi}^k(\Delta\bar{x}_{tr}^k)}{\psi^k(\alpha_g^k\Delta x_g^k)} \ge \alpha_{tr}^k(2-\alpha_{tr}^k) \longrightarrow 1.$$

Since $\tau_s < 1$, we see that (4.5) is true.

It is clear that $F(x^*,\,w^*) = 0$ where F is defined by (2.6). Similar to the results established in [8] (Theorem 4.4 to Theorem 4.11 in [8]), we can show that Δx_{tr}^k is ultimately the Newton direction (with respect to x) for (2.6), and consequently, we can establish the rate of convergence for Algorithm IPTR as follows.

Theorem 4.5. *Suppose there exists a limit point x^* which satisfies (4.1). Suppose $\nabla^2 f(x)$ is Lipschitz continuous in a neighborhood of x^*. Then there exists $C_8 > 0$ such that for every k sufficiently large,*

$$\|x^{k+1} - x^*\| \le C_8\,\|x^{k-1} - x^*\|\,\|x^k - x^*\|, \tag{4.9}$$
$$\|(x^{k+1}, w^{k+1}) - (x^*, w^*)\| \le C_8\,\|(x^{k-1}, w^{k-1}) - (x^*, w^*)\|\,\|(x^k, w^k) - (x^*, w^*)\|. \tag{4.10}$$

In other words, the sequences $\{x^k\}$ and $\{(x^k,\,w^k)\}$ converge to x^* and $(x^*,\,w^*)$ in the rate of at least 2-step quadratic, respectively. Moreover, similar to Theorem 4.9 in [8], We can show that the sequence $\{(x^k,\,\hat{w}^k)\}$ converges to $(x^*,\,w^*)$ quadratically, where $\{\hat{w}^k\}$ is defined by

$$\hat{w}^k = w^{k-1} + \Delta w_{tr}^{k-1} \ \ (k \ge 1), \tag{4.11}$$

and $\{\Delta w_{tr}^k\}$ is defined by (2.20).

To conclude this section, we justify the statement about θ^k given at the end of Section 2.2.

Theorem 4.6. *Let θ^k be defined by (2.30). Let ν^k denote the least eigenvalue of $(\bar{Z}^k)^T\bar{M}^k\bar{Z}^k$. Then*

(i) $\|\tilde{X}^k g^k\| \le 2\,\theta^k$ *and* $-\frac{4\,\theta^k}{\delta_l^2} \le \nu^k$ *for every k such that $\theta^k \le \frac{1}{2}$, where $\delta_l > 0$ is the constant given in* (i) *of Lemma 4.4.*

(ii) $\theta^k = 0$ *if and only if x^k satisfies the second order necessary conditions (2.1) – (2.4).*

Proof. Proof of (i). By (2.30), it is clear that $\|\tilde{X}^k g^k\| \le 2\,\theta^k$. In addition,

$$|\bar{\psi}^k(\Delta\bar{x}_{tr}^k)| = |\psi^k(\Delta x_{tr}^k)| \le 2\,\theta^k,$$

which by (3.5) implies that

$$\frac{1}{2}\lambda_{tr}^k\|\Delta\bar{x}_{tr}^k\|^2 \le 2\,\theta^k.$$

If $\|\Delta\bar{x}^k_{tr}\| < \delta^k$, then (2.16) yields $\lambda^k_{tr} = 0$ which implies that $\nu^k \geq 0 \geq -\frac{4\theta^k}{\delta_l^2}$. Otherwise, $\|\Delta\bar{x}^k_{tr}\| = \delta^k \geq \delta_l$ and

$$\lambda^k_{tr} \leq \frac{4\theta^k}{\delta_l^2}.$$

Therefore, by (2.15),

$$\nu^k \geq -\lambda^k_{tr} \geq -\frac{4\theta^k}{\delta_l^2}.$$

Proof of (ii). Assume first that $\theta^k = 0$. Then it follows from Part (i) that x^k satisfies the second order conditions (2.1) – (2.4).

Now we assume that x^k satisfies conditions (2.1) – (2.4). Then $\tilde{X}^k g^k = 0$ and $\bar{g}^k = D^k g^k = 0$. So (2.14) implies

$$(\Delta\bar{x}^k_{tr})^T(\bar{Z}^k)^T\bar{M}^k\bar{Z}^k\Delta\bar{x}^k_{tr} + \lambda^k_{tr}\|\Delta\bar{x}^k_{tr}\|^2 = 0. \tag{4.12}$$

On the other hand, by Lemma 3.11, condition (2.4) implies that $(\Delta\bar{x}^k_{tr})^T(\bar{Z}^k)^T\bar{M}^k\bar{Z}^k\Delta\bar{x}^k_{tr} \geq 0$. Then, by (4.12) and (2.16), $\lambda^k_{tr} = 0$. Therefore, using (2.14) and the fact that $\bar{g}^k = 0$, we have

$$\psi^k(\Delta x^k_{tr}) = \bar{\psi}^k(\Delta\bar{x}^k_{tr}) = \frac{1}{2}(\Delta\bar{x}^k_{tr})^T(\bar{Z}^k)^T\bar{M}^k\bar{Z}^k\Delta\bar{x}^k_{tr} = 0.$$

Then $\theta^k = 0$.

5. Concluding Remarks

We have proposed an interior projected trust region algorithm for linearly constrained optimization. Under compactness of the level set and nondegeneracy assumptions, convergence results analogous to trust region algorithms for unconstrained minimization are obtained. Preliminary numerical experiments indicate that the algorithm works well for problems with small number of variables. However, as it stands, the algorithm is not suitable for large-scale problems since it requires full-dimensional trust region computations and several matrix factorizations in each major iteration. In response to this we are currently investigating a modification to this approach involving iterative and approximate linear solvers. This will be the topic of a future report.

6. Acknowledgments

The author thanks Tom Coleman at Cornell for his encouragement and for carefully reading the drafts of this paper, and thanks Philip Gill at UC San Diego and Yuying Li at Cornell for many helpful discussions.

References

[1] R.H. Byrd, R.B. Schnabel, Continuity of the null space and constrained optimization, *Mathematical Programming,* **35** (1986), 32-41.

[2] R.H. Byrd, R.B. Schnabel, G.A. Shultz, A trust region algorithm for nonlinearly constrained optimization, *SIAM Journal on Numerical Analysis,* **24** (1987), 1152-1170.

[3] P.H. Calamai, J.J. More, Projected gradient methods for linearly constrained problems, *Mathematical Programming,* **39** (1987), 93-116.

[4] M.R. Celis, J.E. Dennis, R.A. Tapia, A trust region strategy for nonlinear equality constrained optimization, in Numerical Optimization 1984, P. T. Boggs, R. H. Byrd and R. B. Schnabel, eds., SIAM, Philadelphia, PA, pp. 71-82.

[5] T.F. Coleman, Y. Li, A trust region and affine scaling interior point method for nonconvex minimization with linear inequality constraints, *Mathematical Programming,* **88** (2000), 1-31.

[6] T.F. Coleman, Y. Li, An interior trust region approach for nonlinear minimization subject to bounds, *SIAM Journal on Optimization,* **6** (1996), 418-445.

[7] T.F. Coleman, J. Liu, An interior Newton method for quadratic programming, Tech. Rep. TR 93-1388, Computer Science Department, Cornell University, 1993.

[8] T.F. Coleman, J. Liu, An interior Newton method for quadratic programming, *Mathematical Programming,* **85** (1999), 491-523.

[9] T.F. Coleman, D. Sorensen, A note on the computation of an orthonormal basis for the null space of a matrix, *Mathematical Programming,* **29** (1984), 234-242.

[10] A.R. Conn, N.I.M. Gould, P.L. Toint, Global convergence of a class of trust region algorithms for optimization with simple bounds, *SIAM Journal on Numerical Analysis,* **25** (1988), 433-460.

[11] A.R. Conn, N.I.M. Gould, P.L. Toint, A primal-dual algorithm for minimizing a nonconvex function subject to bound and linear equality constraints, Nonlinear optimization and related topics (Erice, 1998), 15–49, Appl. Optim., 36, Kluwer Acad. Publ., Dordrecht, 2000.

[12] R. Fletcher, Practical Methods of Optimization, John Wiley and Sons, New York, second edition, 1987.

[13] A. Forsgren, W. Murray, Newton methods for large-scale linearly-constrained minimization, *SIAM J. Matrix Anal. Appl.,* **14** (1993), 560–587.

[14] D.M. Gay, Computing optimal locally constrained steps, *SIAM Journal on Scientific and Statistical Computing,* **2** (1981), 186-197.

[15] D.M. Gay, A trust-region approach to linearly constrained optimization, In D. F. Griffiths, editor, Numerical Analysis, number 1066 in Lecture Notes in Mathematics, pp. 72-105. Springer-Verlag, Berlin, 1984.

[16] P.E. Gill, W. Murray, Newton type methods for unconstrained and linearly constrained optimization, *Mathematical Programming,* **7** (1974), 311-350.

[17] P.E. Gill, W. Murray, M.A. Saunders, G.W. Stewart, M.H. Wright, Properties of representation of a basis for the null space, *Mathematical Programming,* **33** (1985), 172-186.

[18] P.E. Gill, W. Murray, M.H. Wright, Practical Optimization, Academic Press, London, 1981.

[19] D. Goldfarb, Extension of Davidon's variable metric method to maximization under linear inequality and equality constraints, *SIAM Journal on Applied Mathematics,* **17** (1969), 739-764.

[20] C. Gonzaga, Path-following mathods for linear programming, *SIAM Review,* **34** (1992), 167-224.

[21] J.J. More, D. Sorensen, Computing a trust region step, *SIAM Journal on Scientific and Statistical Computing,* **4** (1983), 553-572.

[22] B.A. Murtagh, M.A. Saunders, Large-scale linearly constrained optimization, *Mathematical Programming,* **14** (1978), 41-72.

[23] M.J.D. Powell, A new algorithm for unconstrained optimization, in Nonlinear Programming, J. B. Rosen, O. L. Mangasarian and K. Ritter, eds., Academic Press, New York, 1970, pp. 31-65.

[24] M.J.D. Powell, Y. Yuan, A trust region algorithm for equality constrained optimization, *Mathematical Programming,* **49** (1991), 189-211.

[25] K. Ritter, Convergence and superlinear convergence of algorithms for linearly constrained minimization problems, In L.C.W. Dixon, E. Spedicato, and G.P. Szego, editors, Nonlinear Optimization: Theory and Algorithm, Part II, pp. 221-251. Birkhauser, Boston, Massachusetts, 1980.

[26] J.B. Rosen, The gradient projection method for nonlinear programming. Part I. Linear constraints, *SIAM Journal on Applied Mathematics,* **8** (1960), 181-217.

[27] G.A. Schultz, R.B. Schnabel, R.H. Byrd, A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties, *SIAM Journal on Numerical Analysis,* **22** (1985), 47-67.

[28] D. Sorensen, Trust region methods for unconstrained optimization, *SIAM Journal on Numerical Analysis,* **19** (1982), 409-426.

[29] A. Vardi, A trust region algorithm for equality constrained minimization: Convergence properties and implementation, *SIAM J. Numer. Anal.,* **22** (1985), 575-591.

[30] S.J. Wright, An interior-point algorithm for linearly constrained optimization, *SIAM Journal on Optimization,* **2** (1992), 450-473.

[31] Y. Ye, On affine scaling algorithms for nonconvex quadratic programming, *Mathematical Programming,* **56** (1992), 285-300.

[32] Y. Zhang, On the convergence of a class of infeasible interior-point methods for the horizontal linear complementarity problem, *SIAM Journal on Optimization,* **4** (1994), 208-227.