A NEW TRUST REGION ALGORITHM FOR EQUALITY CONSTRAINED OPTIMIZATION *

THOMAS F. COLEMAN† AND WEI YUAN‡

Abstract. We present a new trust region algorithm for solving nonlinear equality constrained optimization problems. At each iterate a change of variables is performed to improve the ability of the algorithm to follow the constraint level sets. The algorithm employs L_2 penalty functions for obtaining global convergence. Under certain assumptions we prove that this algorithm globally converges to a point satisfying the second order necessary optimality conditions; the local convergence rate is quadratic. Results of preliminary numerical experiments are presented.

1. Introduction. We consider the equality constrained optimization problem

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

where $x \in \Re^n$ and $f : \Re^n \to \Re$, and $c : \Re^n \to \Re^m$ are smooth nonlinear functions.

Problem (1.1) is often solved by successive quadratic programming (SQP) methods. At a current point $x_k \in \Re^n$, SQP methods determine a search direction d_k by solving a quadratic programming problem

$$\begin{array}{ll} \text{minimize} & \nabla f(x_k)^T d + \frac{1}{2} d^T M_k d \\ \text{subject to} & c(x_k) + A_k^T d = 0, \end{array} \tag{1.2}$$

where $A_k = \nabla c(x_k)$ and M_k is the Hessian of the Lagrangian at x_k or an approximation to it. The new iterate is given by $x_{k+1} = x_k + \alpha_k d_k$ where $\alpha_k > 0$ is chosen to drive the iterates towards the solution of (1.1).

In this paper we propose a new algorithm for (1.1) based, in part, on the trust region idea. Trust region methods have proved to be very successful for unconstrained optimization problems (e.g., Fletcher [13], Gay [14], Moré and Sorensen [19], Powell [20], Shultz, Schnabel and Byrd [22]). Trust region methods usually have stronger convergence properties than line search methods. Trust region methods can also be applied to equality constrained problems.

* This research was partially supported by the Applied Mathematical Sciences Research Program (KC-04-02) of the Office of Energy Research the U.S. Department of Energy under grant DE-FG02-86ER25013.A000, and by the Computational Mathematics Program of the National Science Foundation under grant DMS-8706133, and by the Advanced Computing Research Institute, a unit of 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.

† Department of Computer Science, Cornell University, Ithaca, New York 14853

‡ Center for Applied Mathematics, Cornell University, Ithaca, New York 14853

A natural way to bring trust region ideas into problem (1.1) is to add a trust region constraint in problem (1.2). i.e., at a current iterate x_k, solve the following subproblem:

$$\begin{array}{ll} \text{minimize} & \nabla f(x_k)^T d + \frac{1}{2} d^T M_k d \\ \text{subject to} & c(x_k) + A_k^T d = 0 \\ & \|d\|_2 \le \Delta_k, \end{array} \tag{1.3}$$

for a direction $d = d_k$, where $\Delta_k > 0$. However, problem (1.3) may not have a feasible region if $c(x_k) \neq 0$. To overcome this difficulty, Vardi [23], Byrd, Schnabel and Shultz [3], and Celis, Dennis and Tapia [5] propose different approaches.

In [3] and [23] the authors replace the equality constraint in (1.3) by a parameterized constraint, $\alpha\, c(x_k) + A_k^T d = 0$, where $0 < \alpha \le 1$. The resultant subproblem is always feasible if α is chosen properly. However, it is not clear how to choose parameter α effectively in practice. In [5] the equality constraint in (1.3) is relaxed differently. The authors of [5] substitute an inequality constraint $\|c(x_k) + A_k^T d\|_2 \le \theta_k$ for the equality constraint in (1.3), where $\theta_k > 0$ tends to zero as $k \to \infty$. Powell and Yuan present a similar method in [21]. Unfortunately this approach is mainly of theoretical interest because there is no efficient procedure for solving the subproblem derived from the above relaxation.

Instead of dealing with the infeasibility problem in (1.3) at each iteration, our algorithm solves (1.1) by minimizing a series of L_2 penalty functions $\{p_{\mu_i}(x)\}$ as $\mu_i > 0$ decreases to zero, where $p_{\mu_i}(x) = f(x) + \frac{1}{2\mu_i}\|c(x)\|_2^2$. To avoid the Maratos effect [18], we use a specially designed trust region method to minimize $p_{\mu_i}(x)$ for a fixed μ_i. At each iterate we introduce a change of variables that forces the iterates to approximately follow the constraint level sets. At each iteration our proposed change of variables causes the iterates to move along a curve which is a quadratic approximation to $c(x) = c(x_k)$. Furthermore, following the curve the value of $\|c(x)\|_2$ increases only slightly.

The idea of the change of variables is due to Coleman [6]. Hempel [17] presents an algorithm that uses this idea to formulate two coupled trust region subproblems and to update the current iterate by solving the two subproblems. Under some strict assumptions, the algorithm in [17] is globally convergent with a local R-superlinear convergence rate.

The change of variables also leads to an efficient block diagonal approximation to the Hessian matrix of the L_2 penalty function. As a consequence, our algorithm obtains the next iterate at each iteration in two parts: a Newton step with a line search is used in one subspace and a trust region subproblem is (approximately) solved in the orthogonal space. We solve the trust region subproblem approximately using a two-dimensional subspace approach [4]. Improvement is measured using the L_2 penalty function $p_{\mu_i}(x)$: From one iterate to the next the algorithm always reduces the value of the L_2 penalty function for a fixed μ_i. The decrease of $p_{\mu_i}(x)$ promotes the global convergence of the algorithm.

For simplicity, we use $\|\cdot\|$ to denotes the 2-norm $\|\cdot\|_2$ throughout this paper. The outline of this paper is as follows. In section 2 we introduce a change of variables at each iterate and describe our algorithm. Under appropriate assumptions, we show in section 3 that the sequence generated by our algorithm converges to a point satisfying the second order necessary optimality conditions. In section 4 we further assume that at the limit point the second order sufficient optimality conditions are satisfied and we emonstrate that convergence rate is quadratic in a neighborhood of the optimal point. In section 5 some preliminary numerical results are illustrated and discussed. Finally we make some concluding remarks in section 6.

2. Algorithm. As we mentioned in §1, our strategy for solving problem (1.1) is to minimize a sequence of penalty functions $\{p_{\mu_i}(x)\}$ as μ_i tends to zero. Under reasonable assumptions the limit point of the sequence of minimizers of $p_{\mu_i}(x)$ is a local minimizer of (1.1). Unfortunately, due to the Maratos effect [18], unconstrained minimization techniques often converge slowly when applied to minimize $p_{\mu_i}(x)$. The reason for the slow convergence is that the Hessian of $p_{\mu_i}(x)$ is dominated, in some directions, by the constraint gradients when μ_i and $\|c(x)\|$ are small and x is far away from a local minimizer of (1.1). This causes most unconstrained minimization methods to compute steps almost entirely in the tangent space of $c(x)$. Since $\|c(x)\|$ increases in the tangent space of $c(x)$, when $c(x)$ is nonlinear, methods are forced to take very small steps to ensure that $p_{\mu_i}(x)$ decreases. Therefore convergence is unacceptably slow.

In this section, we develop a specially designed trust region method for minimizing $p_{\mu_i}(x)$. At each iterate we introduce a change of variables that forces the next iterate to move along a curve and this curve approximately follows the constraint level sets. The Maratos effect is thereby avoided.

We introduce the change of variables as follows. At a current point $x_k \in \Re^n$, let $A_k = \nabla c(x_k) \in \Re^{n \times m}$. Suppose that the QR factorization of A_k is

$$A_k = Q\bar{R}_k = [Y_k \ Z_k] \begin{bmatrix} R_k \\ 0 \end{bmatrix} = Y_k R_k,$$

where $Q_k \in \Re^{n\times n}$ is orthogonal and R_k is an $m \times m$ upper triangular matrix. We assume that A_k has full column rank so that R_k is nonsingular. Define

$$w(h, v) \equiv u(h) + Y_k v = x_k + s(Z_k h) + Y_k v$$

for $h \in \Re^{n-m}$ and $v \in \Re^m$, where $u(h)$ is a quadratic approximation to a curve along the level sets $c(x) = c(x_k)$. More precisely, $u(h)$ is a smooth one-to-one mapping satisfying

$$|\ u(h) - [\ x_k + Z_k h - \frac{1}{2} Y_k R_k^{-T}(h^T Z_k^T \nabla^2 c(x_k) Z_k h)\ ]\ | = \mathcal{O}(\|h\|^3) \tag{2.1}$$

in a neighborhood of x_k. For example, both

$$u(h) = x_k + s(Z_k h) = x_k + Z_k h + Y_k R_k^{-T}[c(x_k) - c(x_k + Z_k h)] \tag{2.2}$$

and

$$u(h) = x_k + s(Z_k h) = x_k + Z_k h - \frac{1}{2} Y_k R_k^{-T} [h^T Z_k^T \nabla^2 c(x_k) Z_k h] \tag{2.3}$$

are such mappings. If second order information is not available, we can adopt the form of $u(h)$ in (2.2). Otherwise, we may define $u(h)$ via (2.3) since then evaluation of $u(h)$ uses only the function evaluations at the current point x_k.

It follows from the chain rule that

$$\nabla_h c(u(0)) = 0 \qquad \text{and} \qquad \nabla_h^2 c(u(0)) = 0. \tag{2.4}$$

Thus

$$c(u(h)) = c(u(0)) + \mathcal{O}(\|h\|^3) = c(x_k) + \mathcal{O}(\|h\|^3) \qquad \text{as } h \to 0. \tag{2.5}$$

Equality (2.5) indicates that along the curve $u(h)$ the value of $\|c(u(h))\|$ varies only modesty around $h = 0$. Simple calculations yield

$$\begin{cases} \nabla_h p_\mu(u(0)) = \nabla_h f(u(0)) = Z_k^T \nabla f(x_k) \\ \nabla_h^2 p_\mu(u(0)) = \nabla_h^2 f(u(0)) = Z_k^T \nabla^2 L(x_k) Z_k \equiv H_k, \end{cases} \tag{2.6}$$

where $\nabla_x^2 L(x) = \nabla_x^2 f(x) + \sum_{j=1}^m \lambda^j \nabla_x^2 c_j(x)$ and $\lambda = -R_k^{-1} Y_k^T \nabla f(x_k) \equiv \lambda_k$ at x_k.

Setting $\hat{p}_\mu(h, v) = p_\mu(w(h, v))$, from the chain rule, we obtain that

$$\nabla_{h,v} \hat{p}_\mu(0,0) = \begin{bmatrix} \nabla_h \hat{p}(0,0) \\ \nabla_v \hat{p}(0,0) \end{bmatrix} = \begin{bmatrix} Z_k^T \nabla p_\mu(x_k) \\ Y_k^T \nabla p_\mu(x_k) \end{bmatrix} = \begin{bmatrix} Z_k^T \nabla f(x_k) \\ Y_k^T \nabla p_\mu(x_k) \end{bmatrix},$$

and

$$\nabla_{h,v}^2 \hat{p}_\mu(0,0) = \begin{bmatrix} Z_k^T \nabla^2 L(x_k) Z_k & Z_k^T \nabla^2 \hat{L}(x_k) Y_k \\ Y_k^T \nabla^2 \hat{L}(x_k) Z_k & Y_k^T \nabla^2 \hat{L}(x_k) Y_k + \frac{1}{\mu} R_k R_k^T \end{bmatrix},$$

where $\nabla^2 \hat{L}(x_k) = \nabla^2 f(x_k) + \sum_{j=1}^m \frac{c_j(x_k)}{\mu} \nabla^2 c_j(x_k)$. The standard trust region subproblem associated with $p_\mu(x)$ in terms of (h, v) is

$$\begin{array}{ll} \text{minimize} & \nabla_{h,v} \hat{p}(0,0)^T \ \hat{s} + \frac{1}{2} \hat{s}^T \ \nabla_{h,v}^2 \hat{p}(0,0) \ \hat{s} \\ \text{subject to} & \|\hat{s}\|_2 \leq \Delta \end{array} \tag{2.7}$$

where $\hat{s} = (h^T, v^T)^T$. It follows from [14] that problem (2.7) is equivalent to a linear system of the form

$$\begin{bmatrix} Z_k^T \nabla^2 L(x_k) Z_k + \alpha I_{n-m} & Z_k^T \nabla^2 \hat{L}(x_k) Y_k \\ Y_k^T \nabla^2 \hat{L}(x_k) Z_k & Y_k^T \nabla^2 \hat{L}(x_k) Y_k + \frac{1}{\mu} R_k R_k^T + \alpha I_m \end{bmatrix} \begin{bmatrix} h \\ v \end{bmatrix}$$
$$= - \begin{bmatrix} Z_k^T \nabla p_\mu(x_k) \\ Y_k^T \nabla p_\mu(x_k) \end{bmatrix}. \tag{2.8}$$

where $\alpha \geq 0$ is chosen so that the coefficient matrix $(\nabla^2_{h,v}\hat{p}(0,0) + \alpha I_n)$ of system (2.8) is positive semi-definite and $\|\hat{s}\|_2 \leq \Delta$.

By Taylor's theorem, we get that $Z_k^T \nabla p_\mu(x_k) + Z_k^T \nabla^2 \hat{L}(x_k) Y_k v \approx Z_k^T \nabla p_\mu(x_k + Y_k v) \approx Z_k^T \nabla f(x_k + Y_k v)$. Thus, the upper part of system (2.8) can be approximately written as $(Z_k^T \nabla^2 L(x_k) Z_k + \alpha I_{n-m})h = -Z_k^T \nabla f(x_k + Y_k v)$.

Because $A_k = Y_k R_k$ and $\lambda_k = -R_k^{-1} Y_k^T \nabla f(x_k)$, we have

$$Y_k^T \nabla p_\mu(x_k) = R_k R_k^{-1} Y_k^T [\nabla f(x_k) + A_k \frac{c(x_k)}{\mu}] = -\frac{1}{\mu} R_k [\mu \ \lambda_k - c(x_k)]. \tag{2.9}$$

As $\mu \to 0$, it is clear that matrix $\frac{1}{\mu} R_k R_k^T$ plays a dominant role in the lower part of system (2.8). Thus, system $\frac{1}{\mu} R_k R_k^T v = -Y_k^T \nabla p_\mu(x_k) = \frac{1}{\mu} R_k [\mu \ \lambda_k - c(x_k)] \approx -\frac{1}{\mu} R_k \, c(x_k)$ is an approximation to the lower part of system (2.8).

Therefore, system (2.8), hence problem (2.7), can be approximated by

$$\begin{bmatrix} B_k + \alpha I & 0 \\ 0 & R_k^T \end{bmatrix} \begin{bmatrix} h \\ v \end{bmatrix} = - \begin{bmatrix} Z_k^T \nabla f(x_k + Y_k v) \\ c(x_k) \end{bmatrix} \tag{2.10}$$

where B_k is the reduced Hessian $H_k = Z_k^T \nabla^2 L(x_k) Z_k$ or an approximation to it. Note that the penalty parameter μ is absent from system (2.10).

System (2.10) suggests that trust region subproblem (2.7) can be (approximately) solved as follows. Partition the step vector $\hat{s} = (h^T, v^T)$ and compute v and h separately. First solve

$$R_k^T v = -c(x_k) \tag{2.11}$$

for a vertical direction v, set $v = \beta v$ as described below, and then solve

$$\begin{array}{ll} \text{minimize} & g(x_k + Y_k v)^T h + \frac{1}{2} h^T \ B_k \ h \\ \text{subject to} & \|h\|_2 \leq \Delta_k \end{array} \tag{2.12}$$

where $\Delta_k > 0$, $g(x) = Z(x)^T \nabla f(x)$. Note that the solution v to (2.11) is the Newton step at x_k for the equation $c(x) = 0$.

To ensure that the penalty function $p_{\mu_i}(x)$ decreases, we perform the following backtracking line search approach to choose the step length β along the direction $Y_k v$ at x_k.

1. Set $\beta := 1$;
2. Until the line search condition

$$p_{\mu_i}(x_k + \beta Y_k v) - p_{\mu_i}(x_k) \leq \sigma \beta \nabla p_{\mu_i}(x_k)^T Y_k v \tag{2.13}$$

   is satisfied, choose a new $\beta \in [\tau_1 \beta, \tau_2 \beta]$.

where $0 < \tau_1 < \tau_2 < 1$ are given. As we can see in Lemma 2.1, this vertical step procedure guarantees that the penalty function $p_{\mu_i}(x)$ decreases if $\|c(x_k)\| > \mu_i \max\{\|\lambda_k\|/\sigma,\ 1\}$.

To economize we employ a vertical step v, as described above, only when x_k is outside an envelope around the surface $c(x) = 0$. That is, when $\|c(x_k)\| > \mu_i \cdot \max\{\|\lambda_k\|/\sigma,\ 1\}$, we compute v as described above. When $\|c(x_k)\| \leq \mu_i \cdot \max\{\|\lambda_k\|/\sigma,\ 1\}$, i.e., when x_k is inside the envelope, we do not solve (2.11) and do not use the vertical step v. In this case (2.12) is solved for a step h and the next iterate has the form $x_{k+1} = u(h) = x_k + s(Z_k h)$. Due to (2.5), $\|c(x)\|$ varies only mildly along the curve $u(h)$ beginning at $h = 0$.

At the beginning of this section we mentioned that the columns of the matrix Z_k form an orthonormal basis for the null space of A_k^T. However, this does not completely specify Z_k. Moreover, if the choice of null space changes significantly from one iterate to the next, convergence of our algorithm cannot be achieved. Coleman and Sorensen [9] expose this issue and suggest several procedures for computing Z_k when performing a QR factorization of A_k. Those procedures guarantee that $Z(x)$ varies smoothly, locally. In our algorithm, we use the approach of Coleman and Sorensen to compute Z_k. For details, see [9].

Our algorithm, Algorithm 2.1, is described in Figure 2.1, where $x_i^{(l)}$, instead of x_k, denotes the iterates. For a fixed parameter μ_i an inner loop is employed to approximately minimize $p_{\mu_i}(x)$ until the following conditions are satisfied.

$$\begin{cases} a) & \|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| \leq \mu_i^{1/2}, \\ b) & \|c(x_i^{(l)})\| \leq \Lambda_i^{(l)} \mu_i, \\ c) & Z(x_i^{(l)})^T \nabla^2 L(x_i^{(l)}) Z(x_i^{(l)}) + \theta_i I \geq 0, \end{cases} \tag{2.14}$$

where $\Lambda_i^{(l)} = \max\{\|\lambda_i^{(l)}\|/\sigma,\ 1\}$ and $\theta_i > 0$ tends to zero as $i \to \infty$. After (2.14) is satisfied, the parameter μ_i is reduced to μ_{i+1} and a new inner loop starts. This process continues as the parameter μ_i goes to zero.

In Algorithm 2.1, $q_i^{(l)}(h) = g(x_i^{(l+)})^T h + \frac{1}{2} h^T B_i^{(l+)} h$, where $B_i^{(l+)}$ is an approximation to the reduced Hessian matrix at $x_i^{(l+)}$. In each inner loop we compute a Newton step with a line search when (2.14 b) does not hold. Next we approximately solve trust region subproblem $\min\{q_i^{(l)}(h) : \|h\|_2 \leq \Delta_i^{(l)}\}$ by computing an $h = h_i^{(l)}$ with $\|h\| \leq \Delta_i^{(l)}$ satisfying (2.15) – (2.17).

Condition 1: There is a $\gamma_1 > 0$ such that

$$q_i^{(l)}(h) \leq -\gamma_1 \|g(x_i^{(l+)})\| \ \min\{\Delta_i^{(l)},\ \|g(x_i^{(l+)})\|/\|B_i^{(l+)}\|\} \tag{2.15}$$

where $g(x) = Z(x)^T \nabla f(x)$.

Condition 2: There is a $\gamma_2 > 0$ such that

$$q_i^{(l)}(h) \leq (\Delta_i^{(l)})^2 \cdot \mathrm{eig}_{\min}(B_i^{(l+)})/\gamma_2 \tag{2.16}$$

where $\mathrm{eig}_{\min}(B_i^{(l+)})$ is the minimum eigenvalue of the matrix $B_i^{(l+)}$.

Condition 3: When $B_i^{(l+)}$ is positive definite and $\|(B_i^{(l+)})^{-1}\bar{g}\| \le \Delta_i^{(l)}$ holds for all $\bar{g} \in \Re^{n-m}$ satisfying $\|g(x_i^{(l+)}) - \bar{g}\| \le \epsilon\, \|g(x_i^{(l+)})\|$,

$$\|B_i^{(l+)}h + g(x_i^{(l+)})\| \le \epsilon\, \|g(x_i^{(l+)})\|, \tag{2.17}$$

where $0 \le \epsilon < 1/5$ is given.

Set $i = 1$. Choose parameters $0 \le \mu_{\min} << \mu_0,\ \ \rho < 1,\ \ \kappa > 0,$ and $0 < \eta_1 < \eta_2 < 1$;
Choose an initial point $x_0^* \in \Re^n$ and an initial trust radius $\Delta_0^* > 0$.
`while` $\mu_i > \mu_{\min}$
  Set $l := 0$, $x_i^{(0)} := x_{i-1}^*$; and $\Delta_i^{(0)} = \min\{2\Delta_{i-1}^*,\ \kappa\mu_{i-1}^{2/5}\}$;
  `while` any of the criteria in (2.14) does not hold;
    $x_i^{(l+)} := x_i^{(l)}$;
    `if` (2.14 b) does not hold;
      Solve $(R_i^{(l)})^T v_i^{(l)} = -c(x_i^{(l)})$;
      Find the step length $\beta_i^{(l)} > 0$ by performing line search along $v_i^{(l)}$;
      $x_i^{(l+)} := x_i^{(l)} + \beta_i^{(l)}\, Y_i^{(l)} v_i^{(l)}$;
    `end;`
    Compute $h_i^{(l)}$ satisfying (2.15) – (2.17) in Conditions 1 - 3;
    Calculate the ratio $r_i^{(l)} = [p_{\mu_i}(u(h_i^{(l)})) - p_{\mu_i}(u(0))] \,/\, q_i^{(l)}(h_i^{(l)})$;
    `while` $r_i^{(l)} < \eta_1$
      Set $\Delta_i^{(l)} := \Delta_i^{(l)}/4$;
      Compute $h_i^{(l)}$ satisfying (2.15) – (2.17) in Conditions 1 - 3;
      Calculate the ratio $r_i^{(l)}$;
    `end;`
    `if` $r_i^{(l)} > \eta_2$,
      $\Delta_i^{(l)} := \min\{2\Delta_i^{(l)}, \kappa\mu_{i-1}^{2/5}\}$;
    `end;`
    $\Delta_i^{(l+1)} = \Delta_i^{(l)}$;
    $x_i^{(l+1)} = x_i^{(l+)} + s(Z_i^{(l+)} h_i^{(l)})$;
    $l := l + 1$;
  `end;`
  $x_i^* := x_i^{(l)}$; $\Delta_i^* := \Delta_i^{(l)}$; $\mu_{i+1} := \min\{\rho\mu_i,\ \mu_i^{6/5}\}$; $i := i + 1$;
`end;`
Set $x^* := x_i^*$ and STOP;

Figure 2.1 : *Algorithm* 2.1

Note that, if $\epsilon = 0$ in (2.17), h is the exact solution to linear system $B_i^{(l+)} h = -g(x_i^{(l+)})$. As showed by Shultz, Schnabel and Byrd in [22], when $\epsilon = 0$, any h satisfying Conditions 1, 2 and 3 approximately solves the trust region problem. And, of course, the exact solution to the trust region subproblem satisfies these conditions.

From Algorithm 2.1 we see that

$$\|h_i^{(l)}\| \leq \Delta_i^{(l)} \leq \kappa \mu_{i-1}^{2/5}, \tag{2.18}$$

and

$$\mu_i = \mu_{i-1}^{6/5} \qquad \text{for } i \text{ large enough} \quad \text{and} \qquad \mu_i \leq \rho\ \mu_{i-1}. \tag{2.19}$$

As we can see in the proof of Lemma 4.4, (2.18) guarantees that $q_i^{(l)}(h)$ is a good approximation to $p_{\mu_i}(u_i^{(l)}(h)) - p_{\mu_i}(u_i^{(l)}(0))$ no matter how small the value of μ_i is.

Suppose $\{x_i^{(l)}\}$ is the sequence generated by Algorithm 2.1, the following lemma says that the sequence $\{p_{\mu_i}(x_i^{(l)})\}$ always decreases for any fixed μ_i.

LEMMA 2.1. *The sequence $\{x_i^{(l)}\}$ generated by Algorithm 2.1 always satisfies*

$$p_{\mu_i}(x_i^{(l+1)}) - p_{\mu_i}(x_i^{(l)}) \leq 0. \tag{2.20}$$

Furthermore, if (2.14 b) does not holds, then

$$\nabla p_{\mu_i}(x_i^{(l)})^T Y_i^{(l)} v_i^{(l)} = c(x_i^{(l)})^T \lambda_i^{(l)} - \frac{1}{\mu_i}\|c(x_i^{(l)})\|^2. \tag{2.21}$$

Proof. Since $r_i^{(l)} > 0$, then it follows from the definition of $r_i^{(l)}$ and (2.15) that

$$p_{\mu_i}(x_i^{(l+1)}) - p_{\mu_i}(x_i^{(l+)}) \leq r_i^{(l)}\ q_i^{(l)}(h_i^{(l)}) \leq 0. \tag{2.22}$$

If (2.14 b) holds, then $x_i^{(l+)} = x_i^{(l)}$. Thus, inequality (2.22) yields (2.20).

If (2.14 b) does not hold, since $v_i^{(l)} = (R_i^{(l)})^{-T} c_i^{(l)}$, equality (2.21) follows from (2.9). Since $\sigma < 1$, it follows from (2.13) that

$$\begin{aligned} p_{\mu_i}(x_i^{(l+)}) - p_{\mu_i}(x_i^{(l)}) &\leq \sigma\beta_i^{(l)}\ [\ c(x_i^{(l)})^T \lambda_i^{(l)} - \frac{1}{\mu_i}\|c(x_i^{(l)})\|^2\ ] \\ &\leq \sigma\beta_i^{(l)}\ [\ \|c(x_i^{(l)})\|\ \|\lambda_i^{(l)}\| - \frac{1}{\mu_i}\|c(x_i^{(l)})\|^2\ ] \\ &\leq \sigma\beta_i^{(l)}\ \frac{\sigma - 1}{\mu_i}\|c(x_i^{(l)})\|^2 < 0. \end{aligned} \tag{2.23}$$

Combining inequalities (2.22) and (2.23), we obtain (2.20). ∎

3. Global Convergence. In this section we analyze global convergence properties of Algorithm 2.1. We begin this section with our assumptions. Under these assumptions we show that all limit points of the sequence generated by Algorithm 2.1 satisfy the first order necessary optimality conditions to (1.1). Then we prove that if the number of limit points is finite, the primary sequence converges to a point satisfying the second order necessary optimality conditions.

ASSUMPTION 3.1. *The sequence $\{x_i^{(l)}\}$ generated by Algorithm 2.1 is contained in a bounded convex set D and the following properties hold.*

1. *The functions $f : \Re^n \to \Re$, and $c : \Re^n \to \Re^m$ and their first and second derivatives are uniformly bounded in norm over D.*
2. *The matrix $A(x)$ has full column rank, and there is a constant K_0 such that*

$$\|A(x)[A(x)^T A(x)]^{-1}\| \le K_0 \tag{3.1}$$

for all $x \in D$.

Note that (3.1) implies that $\|Y_i^{(l)}(R_i^{(l)})^{-T}\| \le K_0$ for all i and l.

In Algorithm 2.1, for any fixed μ_i the penalty function $p_{\mu_i}(x)$ is minimized until (2.14) is satisfied. In Lemmas 3.2 and 3.3 we illustrate that for any fixed μ_i, (2.14) holds after a finite number of iterations. Before doing that, we show in Lemma 3.1 that the step lengths $\{\beta_i^{(l)}\}$ are uniformly bounded away from zero.

LEMMA 3.1. *Under Assumption 3.1, there is a constant $\beta > 0$, such that for all i and l, the step length $\beta_i^{(l)}$ satisfies*

$$\beta_i^{(l)} \ge \beta > 0. \tag{3.2}$$

Proof. First, note that in Algorithm 2.1 $\beta_i^{(l)}$ is computed only when condition (2.14 b) is not satisfied. Suppose that $\beta_i^{(l)} < 1$. Let step length $\tilde\beta \le 1$ be the most recent failure of (2.13). Then,

$$p_{\mu_i}(x_i^{(l)} + \tilde\beta Y_i^{(l)} v) - p_{\mu_i}(x_i^{(l)}) > \sigma\tilde\beta[\nabla p_{\mu_i}(x_i^{(l)})]^T Y_i^{(l)} v \tag{3.3}$$

and $\tau_1\tilde\beta \le \beta_i^{(l)}$. Since $\nabla^2 c_i(x)$ and $\nabla^2 f(x)$ are bounded and

$$\nabla^2 p_{\mu_i}(x) = \nabla^2 f(x) + \sum_{i=1}^{m} \frac{c_i(x)}{\mu_i}\nabla^2 c_i(x) + \frac{1}{\mu_i}A(x)A(x)^T,$$

Taylor's theorem yields that for some $\tilde x_i^{(l)}$ near $x_i^{(l)}$

$$p_{\mu_i}(x_i^{(l)} + \tilde\beta Y_i^{(l)} v) - p_{\mu_i}(x_i^{(l)}) \;\; = \;\; \tilde\beta[\nabla p_{\mu_i}(x_i^{(l)})]^T Y_i^{(l)} v +$$

$$+\frac{\tilde{\beta}^2}{2}(Y_i^{(l)}v)^T[\nabla^2 p_{\mu_i}(\tilde{x}_i^{(l)})](Y_i^{(l)}v)$$
$$\leq \tilde{\beta}[\nabla p_{\mu_i}(x_i^{(l)})]^T Y_i^{(l)}v + \tilde{\beta}^2\frac{K}{\mu_i}\|Y_i^{(l)}v\|^2, \qquad (3.4)$$

where K is a constant independent of μ_i. Noting that $Y_i^{(l)}v = -Y_i^{(l)}(R_i^{(l)})^{-T}c(x_i^{(l)})$, it follows from (3.3), (3.4), and (3.1) that

$$-(1-\sigma)[\nabla p_{\mu_i}(x_i^{(l)})]^T Y_i^{(l)}v < \tilde{\beta}\frac{K}{\mu_i}\|Y_i^{(l)}v\|^2 \leq \tilde{\beta}\frac{KK_0^2}{\mu_i}\|c(x_i^{(l)})\|^2. \qquad (3.5)$$

On the other hand, since $\|\lambda_i^{(l)}\| \leq \Lambda_i^{(l)} < \frac{\|c(x_i^{(l)})\|}{\mu_i}$ whenever $\beta_i^{(l)}$ is computed, (2.21) implies that

$$\nabla p_{\mu_i}(x_i^{(l)})^T Y_i^{(l)}v \leq \|c(x_i^{(l)})\|\|\lambda_i^{(l)}\| - \frac{\|c(x_i^{(l)})\|^2}{\mu_i} \leq -(1-\sigma)\frac{\|c(x_i^{(l)})\|^2}{\mu_i}. \qquad (3.6)$$

Combining (3.5) and (3.6), we obtain $\beta_i^{(l)} \geq \tau_1\,\tilde{\beta} > \tau_1\frac{(1-\sigma)^2}{KK_0^2}$. ■

From (2.20) and (2.23), for any $\bar{l} > l$,

$$p_{\mu_i}(x_i^{(\bar{l})}) - p_{\mu_i}(x_i^{(l)}) \leq p_{\mu_i}(x_i^{(l+)}) - p_{\mu_i}(x_i^{(l)}) \leq -\sigma(1-\sigma)\beta_i^{(l)}\frac{\|c(x_i^{(l)})\|^2}{\mu_i}. \qquad (3.7)$$

Thus, inequality (2.14 b) holds for fixed μ_i after a finite number of iterations. Otherwise, from (3.7) and Lemma 2.1, we have

$$p_{\mu_i}(x_i^{(l_{s+1})}) - p_{\mu_i}(x_i^{(l_s)}) \leq -\sigma(1-\sigma)\beta_i^{(l_s)}(\Lambda_i^{(l_s)})^2\mu_i \leq -\sigma(1-\sigma)\;\beta\;\mu_i < 0. \qquad (3.8)$$

Inequality (3.8) contradicts the fact that $p_{\mu_i}(x)$ is bounded below for any fixed μ_i. Therefore we obtain the following result.

LEMMA 3.2. *Suppose Assumption 3.1 is satisfied. For any fixed μ_i, suppose $\{x_i^{(l)}\}$ is generated by Algorithm 2.1. Then there exists an integer $\bar{l}_i > 0$, such that for all $l \geq \bar{l}_i$*

$$\|c(x_i^{(l)})\| \leq \Lambda_i^{(l)}\;\mu_i. \qquad (3.9)$$

Let $\Delta > 0$ be a trust radius. Suppose that h satisfies Conditions 1, 2, and 3 in section 2 with $\Delta_i^{(l)} = \Delta$. Define

$$r_i^{(l)}(\Delta) = [p_{\mu_i}(u_i^{(l)}(h)) - p_{\mu_i}(u_i^{(l)}(0))]\;/\;q_i^{(l)}(h).$$

Using Taylor's theorem, we obtain that

$$\begin{aligned} p_{\mu_i}(u_i^{(l)}(h)) - p_{\mu_i}(u_i^{(l)}(0)) &= \nabla_h p_{\mu_i}(u_i^{(l)}(0))^T h + h^T[\int_0^1 \nabla_h^2 p_{\mu_i}(u_i^{(l)}(\tau h))(1-\tau)d\tau]\,h \\ &= q_i^{(l)}(h) + h^T\{\int_0^1[\nabla_h^2 p_{\mu_i}(u_i^{(l)}(\tau h)) \\ &\qquad -\nabla_h^2 p_{\mu_i}(u_i^{(l)}(0))](1-\tau)d\tau\}h. \end{aligned}$$

Subtracting $q_i^{(l)}(h)$ from both sides and then dividing both sides by $q_i^{(l)}(h)$, this becomes

$$\begin{aligned} |r_i^{(l)}(\Delta) - 1| &= | \, [p_{\mu_i}(u_i^{(l)}(h)) - p_{\mu_i}(u_i^{(l)}(0))] - q_i^{(l)}(h) \, | \, / \, |q_i^{(l)}(h)| \\ &\leq \frac{\|h\|^2}{|q_i^{(l)}(h)|} \| \int_0^1 [\nabla_h^2 p_{\mu_i}(u_i^{(l)}(\tau h)) - \nabla_h^2 p_{\mu_i}(u_i^{(l)}(0))](1-\tau) d\tau \|. \end{aligned} \tag{3.10}$$

Based on (3.10), in Lemma 3.3 we show that (2.14 a) and (2.14 c) hold for some integer l.

LEMMA 3.3. *Suppose Assumption 3.1 is satisfied. For fixed $\mu_i > 0$, suppose $\{x_i^{(l)}\}$ is generated by Algorithm 2.1. Then there exists a subsequence of $\{x_i^{(l)}\}$, $l \in \mathcal{L}$, such that*

$$\lim_{\substack{l \in \mathcal{L} \\ l \to \infty}} \|g(x_i^{(l)})\| = 0 \tag{3.11}$$

and

$$\lim_{\substack{l \in \mathcal{L} \\ l \to \infty}} eig_{\min}(H(x_i^{(l)})) \geq 0, \tag{3.12}$$

where $g(x) = Z(x)^T \nabla f(x)$, $H(x) = Z(x)^T \nabla^2 L(x) Z(x)$ and $eig_{\min}(H)$ denotes the smallest eigenvalue of a matrix H.

Proof. Due to Lemma 3.2, we can assume that $x_i^{(l+)} = x_i^{(l)}$ without loss of generality. There are two possible cases: either

1. $\inf_l \Delta_i^{(l)} = \bar{\Delta} > 0$, or
2. There exists an integer set $\bar{\mathcal{L}}$ such that $\Delta_i^{(l+1)} < \Delta_i^{(l)}$ and $\Delta_i^{(l)} \to 0$ for $l \in \bar{\mathcal{L}}$.

In case 1, suppose there exists an $\epsilon > 0$ such that, for all l, $\|g(x_i^{(l)})\| \geq \epsilon$. Then, because $\|B_i^{(l)}\| \leq K$, it follows from (2.15) in Condition 1 that

$$q_i^{(l)}(h_i^{(l)}) \leq -\gamma_1 \epsilon \min\{\bar{\Delta}, \ \epsilon/K\} \equiv -\bar{\epsilon}. \tag{3.13}$$

Recall that according to Algorithm 2.1, we have that for all l

$$p_{\mu_i}(x_i^{(l+1)}) - p_{\mu_i}(x_i^{(l)}) \leq r_i^{(l)} \ q_i^{(l)}(h_i^{(l)}) \leq \eta_1 \ q_i^{(l)}(h_i^{(l)}) \leq -\eta_1 \bar{\epsilon} < 0.$$

But by the definition of $p_{\mu_i}(x)$ we know that $p_{\mu_i}(x)$ is bounded below, thus

$$p_{\mu_i}(x_i^{(l+1)}) - p_{\mu_i}(x_i^{(l)}) \to 0 \qquad \text{as} \qquad l \to \infty.$$

This contradiction establishes that there exists an integer set $\tilde{\mathcal{L}}$ such that

$$\lim_{l \to \infty} \{g(x_i^{(l)})\}_{l \in \tilde{\mathcal{L}}} = 0.$$

Now, suppose there exists an $\epsilon > 0$ such that, for all $l \in \tilde{\mathcal{L}}$, $\text{eig}_{\min}(H(x_i^{(l)})) \leq -\epsilon < 0$. Then, (2.16) in Condition 2 yields $q_i^{(l)}(h_i^{(l)}) \leq -\frac{(\Delta_i^{(l)})^2}{\gamma_2}\epsilon \leq -\frac{\bar{\Delta}^2}{\gamma_2}\epsilon < 0$. Similar to (3.13), this inequality contradicts the fact that $p_{\mu_i}(x)$ is bounded below. Therefore (3.11) and (3.12) hold with an integer set $\mathcal{L} \subset \tilde{\mathcal{L}}$.

In case 2, by Assumption 3.1, $\{x_i^{(l)}\}_{l\in\bar{\mathcal{L}}}$ is in a compact set D. Thus there exists a convergent subsequence. Without loss of generality, we assume that

$$\Delta_i^{(l)} \to 0 \qquad \text{and} \qquad x_i^{(l)} \to \bar{x} \qquad \text{for} \qquad l \in \bar{\mathcal{L}}.$$

Recall that in Algorithm 2.1 if the radius $\Delta_i^{(l)}$ decreases, there must be a previously tried radius $\bar{\Delta}_i^{(l)}$ for which

$$\Delta_i^{(l)} = \frac{1}{4}\bar{\Delta}_i^{(l)} \qquad \text{and} \qquad r_i^{(l)}(\bar{\Delta}_i^{(l)}) \le \eta_1.$$

Since $\Delta_i^{(l)} \to 0$, we have

$$r_i^{(l)}(\bar{\Delta}_i^{(l)}) \le \eta_1 \qquad \text{and} \qquad \bar{\Delta}_i^{(l)} \to 0 \qquad \text{for} \qquad l \in \bar{\mathcal{L}}. \tag{3.14}$$

Suppose $\|g(\bar{x})\| = 2\epsilon > 0$, then for $l \in \bar{\mathcal{L}}$ sufficiently large, $\|g(x_i^{(l)})\| = \epsilon$. It follows from (2.15) that for $l \in \bar{\mathcal{L}}$

$$q_i^{(l)}(\bar{h}_i^{(l)}) \le -\gamma_1\epsilon \min\{\bar{\Delta}_i^{(l)},\ \epsilon/K\} \le -\gamma_1\epsilon\bar{\Delta}_i^{(l)}$$

since $\bar{\Delta}_i^{(l)} \to 0$. Thus, from (3.10), for $l \in \bar{\mathcal{L}}$,

$$|r_i^{(l)}(\bar{\Delta}_i^{(l)}) - 1| \le \frac{\|\bar{\Delta}_i^{(l)}\|}{\gamma_1\epsilon}\int_0^1 \|\nabla_h^2 p_{\mu_i}(u_i^{(l)}(\tau\bar{h}_i^{(l)}) - \nabla_h^2 p_{\mu_i}(u_k^{(l)}(0))\|(1-\tau)d\tau \to 0.$$

This contradicts (3.14). Thus (3.11) holds with $\mathcal{L} = \bar{\mathcal{L}}$.

Suppose $\text{eig}_{\min}(H(\bar{x})) \le -2\epsilon < 0$, then for $l \in \bar{\mathcal{L}}$ sufficiently large, $\text{eig}_{\min}(H(x_i^{(l)})) \le \epsilon$. Thus, (2.16) implies that for $l \in \bar{\mathcal{L}}$

$$q_i^{(l)}(\bar{h}_i^{(l)}) \le -\frac{(\bar{\Delta}_i^{(l)})^2}{\gamma_2}\epsilon.$$

It follows from (3.10) that for $l \in \bar{\mathcal{L}}$

$$|r_i^{(l)}(\bar{\Delta}_i^{(l)}) - 1| \le \gamma_2\int_0^1 \|\nabla_h^2 p_{\mu_i}(u_i^{(l)}(\tau\bar{h}_i^{(l)}) - \nabla_h^2 p_{\mu_i}(u_i^{(l)}(0))\|(1-\tau)d\tau \to 0$$

since $\|\bar{h}_i^{(l)}\| \le \|\bar{\Delta}_i^{(l)}\| \to 0$. This also contradicts (3.14). Therefore (3.12) holds with $\mathcal{L} = \bar{\mathcal{L}}$. ∎

Lemmas 3.2 and 3.3 indicate that for any fixed μ_i there exists an integer $l_i > 0$ such that for $l = l_i$ the criteria in (2.14) are satisfied. Thus, we obtain the following results.

LEMMA 3.4. *Suppose Assumption 3.1 is satisfied. For every integer i, let l_i denote the first integer $l \geq 0$ for which (2.14) are satisfied. Then Algorithm 2.1 generates a sequence $\{x_k\}$ such that $x_k = x_i^{(l)}$, where $k = k(i,l) = \sum_{j=1}^{i-1} l_j + l$ and $l \leq l_i$. Furthermore,*

$$\underline{\lim}_{k\to\infty} [\|Z(x_k)^T \nabla f(x_k)\| + \|c(x_k)\|] = 0,$$

and

$$\underline{\lim}_{k\to\infty} eig_{\min}[Z(x_k)^T \nabla^2 L(x_k) Z(x_k) + \theta_i I] \geq 0.$$

Proof. Since the criteria in (2.14) are satisfied for $l = l_i$, according to Algorithm 2.1, we reduce the penalty parameter μ_i to μ_{i+1}, set $x_{i+1}^{(0)} = x_i^{(l_i)}$, and start another inner loop with new parameter μ_{i+1}. Therefore, Algorithm 2.1 generates a sequence

$$x_0^{(0)}, \ldots, x_0^{(l_0-1)}, x_1^{(0)}, \ldots, x_1^{(l_1-1)}, x_2^{(0)}, \ldots\ldots\ldots, x_{i-1}^{(l_{i-1}-1)}, x_i^{(0)}, \ldots, x_i^{(l_i)}, \ldots.$$

We can reindex the sequence $\{x_i^{(l)}\}$ as $\{x_k\}$ such that $x_k = x_i^{(l)}$, where $k = k(i,l) = \sum_{j=1}^{i-1} l_j + l$ with $l \leq l_i$.

Since (2.14) holds for $l = l_i$ and μ_i tends to zero, we get $\lim_{i\to\infty}[\|Z(x_i^{(l_i)})^T \nabla f(x_i^{(l_i)})\| + \|c(x_i^{(l_i)})\|] = 0$ and $\lim_{i\to\infty} \mathrm{eig}_{\min}[Z(x_i^{(l_i)})^T \nabla^2 L(x_i^{(l_i)}) Z(x_i^{(l_i)}) + \theta_i I] \geq 0$. Thus, Lemma 3.3 is proved. ∎

From the proof of Lemma 3.3 we see that Condition 1 yields (3.11), and Condition 2 implies (3.12). Thus, if $h_i^{(l)}$ does not satisfy Conditions 2 and 3, we can still establish (3.11) from Condition 1. As we will see in the rest of this section, x_k converges to a point satisfying the first order necessary optimality conditions if $h_i^{(l)}$ in Algorithm 2.1 satisfies Condition 1. Before we prove in Lemmas 3.6 and 3.7 that all limit points of $\{x_k\}$ satisfy the first order necessary conditions, we need the following lemma.

LEMMA 3.5. *Suppose Assumption 3.1 is satisfied. Assume $\{x_i^{(l)}\}$ is generated by Algorithm 2.1. Then*

$$\sum_{i=1}^{\infty} \sum_{l=0}^{l_i-1} [p_{\mu_i}(x_i^{(l)}) - p_{\mu_i}(x_i^{(l+1)})] < +\infty. \tag{3.15}$$

Proof. Due to (2.19), we assume that i_0 is an integer such that $\mu_i = \mu_{i-1}^{6/5}$ for all $i \geq i_0$. Since $x_i^{(0)} = x_{i-1}^{(l_{i-1})}$, we have that

$$\sum_{i=i_0+1}^{\infty} \sum_{l=0}^{l_i-1} [p_{\mu_i}(x_i^{(l)}) - p_{\mu_i}(x_i^{(l+1)})] \quad = \quad \sum_{i=i_0+1}^{\infty} [p_{\mu_i}(x_i^{(0)}) - p_{\mu_i}(x_i^{(l_i)})]$$

$$\begin{aligned} &= \sum_{i=i_0+1}^{\infty} [p_{\mu_i}(x_i^{(0)}) - p_{\mu_{i-1}}(x_i^{(0)})] \\ &\qquad + \sum_{i=i_0+1}^{\infty} [p_{\mu_{i-1}}(x_{i-1}^{(l_{i-1})}) - p_{\mu_i}(x_i^{(l_i)})] \\ &\le \sum_{i=i_0+1}^{\infty} [p_{\mu_i}(x_i^{(0)}) - p_{\mu_{i-1}}(x_i^{(0)})] + N_2 \qquad (3.16) \end{aligned}$$

where $N_2 = p_{\mu_{i_0}}(x_{i_0}^{(l_{i_0})}) - \inf_i p_{\mu_i}(x_i^{(l_i)})$ is a constant since $p_{\mu_i}(x)$ is bounded below.

It follows from the Assumption 3.1 that there exists a constant $N_1 > 0$ such that for all integers $i > 0$ and $0 \le l \le l_i$, $\Lambda_i^{(l)} \le \max\{\ K_0\|\nabla f(x_i^{(l)})\|/\sigma,\ 1\ \} \le N_1$. Thus

$$\|c(x_i^{(0)})\| \le \Lambda_i^{(l)} \mu_{i-1} \le N_1\ \mu_{i-1}. \qquad (3.17)$$

From (3.17) and $\mu_i = \mu_{i-1}^{6/5}$, we get that

$$p_{\mu_i}(x_i^{(0)}) - p_{\mu_{i-1}}(x_i^{(0)}) = [\frac{1}{\mu_i} - \frac{1}{\mu_{i-1}}]\|c(x_i^{(0)})\|^2 \le \frac{\mu_{i-1}^2}{\mu_i} N_1^2 \le \mu_{i-1}^{4/5} N_1^2.$$

Thus, (3.16) yields

$$\sum_{i=i_0}^{\infty} \sum_{l=0}^{l_i-1} [p_{\mu_i}(x_i^{(l)}) - p_{\mu_i}(x_i^{(l+1)})] \le N_1^2 \sum_{i=1}^{\infty} \mu_{i-1}^{4/5} + N_2 \le N_1^2 \frac{\mu_0^{4/5}}{1-\rho^{4/5}} + N_2.$$

Therefore, (3.15) is established. ∎

Because of (3.15), we are able to prove in Lemmas 3.6 and 3.7 that any limit points of sequence $\{x_k\}$ described in Lemma 3.4 satisfies the first order necessary optimality conditions.

Lemma 3.6. *Suppose Assumption 3.1 is satisfied. Suppose $\{x_k\}$ is the sequence described in Lemma 3.4. Then*

$$\lim_{k\to\infty} \|c(x_k)\| = 0. \qquad (3.18)$$

Proof. To establish (3.18), we define

$$d_k = \begin{cases} \frac{\|c(x_k)\|^2}{\mu_i} - c(x_k)^T\lambda_k & \text{if } \|c(x_k)\| > \Lambda_k\mu_i; \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $d_k \ge 0$. It is obvious from Lemma 3.1 and Lemma 2.1 that

$$\sigma\beta\ d_k \equiv \sigma\beta\ d_i^{(l)} \le p_{\mu_i}(x_i^{(l)}) - p_{\mu_i}(x_i^{(l+)}) \le p_{\mu_i}(x_i^{(l)}) - p_{\mu_i}(x_i^{(l+1)})$$

which, with (3.15), implies that

$$\sum_{k=k_0}^{\infty} d_k = \sum_{i=1}^{\infty}\sum_{l=0}^{l_i-1} d_i^{(l)} < +\infty. \tag{3.19}$$

Therefore, $\lim_{k\to\infty} d_k = 0$.

Notice that since $\sigma\Lambda(x_k) \geq \|\lambda_k\|$, we have that if $\|c(x_k)\| > \Lambda_k\mu_i$,

$$\begin{aligned} \|c(x_k)\| &\leq \frac{1}{(1-\sigma)\Lambda_k}[\Lambda_k\|c(x_k)\| - \sigma\Lambda_k\|c(x_k)\|] \\ &\leq \frac{1}{(1-\sigma)\Lambda_k}[\frac{\|c(x_k)\|^2}{\mu_i} - \|\lambda_k\|\|c(x_k)\|] \\ &\leq \frac{d_k}{(1-\sigma)\Lambda_k} \leq \frac{d_k}{1-\sigma}. \end{aligned} \tag{3.20}$$

Therefore, $\|c(x_k)\| \leq \max\{\frac{d_k}{1-\sigma},\ \Lambda_k\mu_i\}$, which implies (3.18) since $\mu_i \to 0$ and $d_k \to 0$. ∎

LEMMA 3.7. *Suppose Assumption 3.1 is satisfied. Suppose $\{x_k\}$ is the sequence described in Lemma 3.4. Then*

$$\lim_{k\to\infty} \|Z(x_k)^T\nabla f(x_k)\| = 0. \tag{3.21}$$

Proof. Let $g(x) = Z(x)^T\nabla f(x)$ and $x_i^{(l_i+)}$ denote $x_{i+1}^{(0+)}$. We first prove by contradiction that $\|g(x_i^{(l+)})\| \to 0$. Without loss of generality, we suppose that for any i there is an integer k_i $(1 \leq k_i < l_i)$ such that $\|g(x_i^{(k_i+)})\| \geq \epsilon > 0$. Because of (3.1) and the definition of $v_i^{(l)}$, it follows from (3.18) and Taylor's theorem that for all $0 \leq l < l_i$ and all i

$$\|g(x_i^{(l+)})\| = \|g(x_i^{(l)} + \beta_i^{(l)}Y_i^{(l)}v_i^{(l)})\| = \|g(x_i^{(l)})\| + \mathcal{O}(\|c(x_i^{(l)})\|). \tag{3.22}$$

Since $\|g(x_i^{(l_i)})\| \leq \mu_i^{1/2}$ and $\mu_i \to 0$, (3.18) implies that $\|g(x_i^{(l_i+)})\| < \epsilon/2$ for i sufficiently large. Assume that $l = j_i$ is the nearest index to k_i in integer set $\{l \in (k_i, l_i] : \|g(x_i^{(l+)})\| < \epsilon/2\}$. Then $\|g(x_i^{(l+)})\| \geq \epsilon/2$ for l satisfying $k_i \leq l < j_i$. Hence, from (2.15), for $k_i \leq l \leq j_i - 1$,

$$p_{\mu_i}(x_i^{(l)}) - p_{\mu_i}(x_i^{(l+1)}) \geq -r_i^{(l)}\, q_i^{(l)}(h_i^{(l)}) \geq -\eta_1\, q_i^{(l)}(h_i^{(l)}) \geq \frac{\eta_1\gamma_1\epsilon}{2}\min\{\Delta_i^{(l)}, \frac{\epsilon}{2K}\}. \tag{3.23}$$

Inequality (3.23) leads that

$$\begin{aligned} p_{\mu_i}(x_i^{(k_i)}) - p_{\mu_i}(x_i^{(j_i)}) &= \sum_{l=k_i}^{j_i-1}[p_{\mu_i}(x_i^{(l)}) - p_{\mu_i}(x_i^{(l+1)})] \\ &\geq \frac{\eta_1\gamma_1\epsilon}{2}\min\{\sum_{l=k_i}^{j_i-1}\Delta_i^{(l)}, \frac{\epsilon}{2K}\} \\ &\geq \frac{\eta_1\gamma_1\epsilon}{2}\min\{\sum_{l=k_i}^{j_i-1}\|h_i^{(l)}\|, \frac{\epsilon}{2K}\}. \end{aligned} \tag{3.24}$$

Combining (3.15) and (3.24), we have that $\sum_{l=k_i}^{j_i-1} \|h_i^{(l)}\| \to 0$ as $i \to \infty$. Because of Assumption 3.1 (2), (3.19) and (3.20) yield

$$\sum_{l=k_i}^{j_i} \|v_i^{(l)}\| \le K_0 \sum_{l=k_i}^{j_i} \|c(x_i^{(l)})\| \le \frac{K_0}{1-\sigma} \sum_{l=k_i}^{j_i} d_i^{(l)} \to 0, \qquad \text{as } i \to \infty.$$

Therefore, the continuity of $u(h)$ implies that there exists a constant $K > 0$ such that

$$\|x_i^{(k_i+)} - x_i^{(j_i+)}\| \le \sum_{l=k_i}^{j_i-1} \|x_i^{(l+)} - x_i^{(l+1+)}\| \le K(\sum_{l=k_i}^{j_i-1} \|h_i^{(l)}\| + \sum_{l=k_i+1}^{j_i} \|v_i^{(l)}\|) \to 0 \tag{3.25}$$

as $i \to \infty$. Due to the continuity of $Z(x)$ and $\nabla f(x)$ and (3.25), we have that

$$\|g(x_i^{(k_i+)}) - g(x_i^{(j_i+)})\| \le \epsilon/4$$

for all i sufficiently large. Therefore when i is sufficiently large

$$\|g(x_i^{(k_i+)})\| \le \|g(x_i^{(k_i+)}) - g(x_i^{(j_i+)})\| + \|g(x_i^{(j_i+)})\| \le \epsilon/4 + \epsilon/2 = 3\epsilon/4.$$

This contradicts the assumption that $\|g(x_i^{(k_i+)})\| \ge \epsilon$. Since $\|g(x_i^{(l+)})\| \to 0$, it follows from (3.22) and (3.18) that (3.21) holds. ■

Now we further assume that sequence $\{x_k\}$ has only a finite number of limit points. We end this section by showing in Theorem 3.1 that in this case there is actually only one limit point x_* for $\{x_k\}$ and at x_* the second order necessary optimality conditions are satisfied.

THEOREM 3.1. *Suppose Assumptions 3.1 is satisfied, $\{x_k\}$ is the sequence described in Lemma 3.4, and that there exist only finite number of limit points to $\{x_k\}$. Then*

$$\lim_{k\to\infty} x_k = x_*$$

where x_* is a point at which the second order necessary optimality conditions are satisfied.

Proof. We first prove by contradiction that the sequence converges. Suppose $\{x_k\}$ does not converge. Since there are only finite number of limit points to $\{x_k\}$, every limit point is an isolated one. Thus, Lemma 4.10 of [19] yields that there exists a subsequence $\{x_{k_j}\}$ of $\{x_k\}$ and an $\epsilon > 0$ such that $\|x_{k_j+1} - x_{k_j}\| \ge \epsilon$ for all j. Inequality (3.25) shows that $\|v_k\|$ tends to zero, which implies that $\|x_{k+1} - x_k\| \to 0$ because of (2.18). This contradiction shows that $\{x_k\}$ converges.

Note that θ_i goes to zero in criterion (2.14, c). Since $\{x_k\}$ converges to a point x_*, from Lemma 3.4, we know that the second order necessary optimality conditions are satisfied at x_*. ■

4. Local Quadratic Convergence. Theorem 3.1 shows that, under certain assumptions, the sequence $\{x_k\}$ converges to a point x_* satisfying the second order necessary conditions. We show in this section that the convergence rate is quadratic in a neighborhood of x_*. Throughout this section we let $B_i^{(l+)}$ denote the reduced Hessian matrix $H(x_i^{(l+)})$.

ASSUMPTION 4.1. *The point x_* is a local minimizer of problem (1.1) where the following conditions hold.*

1. *The functions $\nabla^2 f(x)$ and $\nabla^2 c_i(x)$, $i = 1, \ldots, m$, are Lipschitz continuous in a neighborhood of x_*.*
2. *The matrix $A(x_*)$ has full column rank.*
3. *The reduced Hessian matrix $Z_*^T \nabla_x^2 L(x^*, \lambda^*) Z^*$ is positive definite.*

Note that the full column rank assumption in Assumption 4.1 implies that there is constant K_0 such that $\|Y_i^{(l)}(R_i^{(l)})^{-T}\| \leq K_0$ when $x_i^{(l)}$ is in a neighborhood of x_*.

In Lemmas 4.1, 4.2 and 4.3, we establish some preliminary results. Lemma 4.1 is due to Byrd and Nocedal [2]. It says that when x is sufficiently close to x_*, the distance between x and x_* is "equivalent" to the sum of the norms of the reduced gradient of $f(x)$ and the constraint function $c(x)$.

LEMMA 4.1. *Suppose Assumption 4.1 holds. Then*

$$K_1\|x - x_*\|^2 \leq \|Z(x)^T \nabla f(x)\|^2 + \|c(x)\|^2 \leq K_2\|x - x_*\|^2 \tag{4.1}$$

for all x sufficiently near x_*.

Proof. See [2]. ∎

In Lemma 4.2 we demonstrate that the unit step length $\beta_i^{(l)} = 1$ is admissible after a certain number of iterations.

LEMMA 4.2. *Suppose Assumption 4.1 holds. Suppose the sequence generated by Algorithm 2.1 converges to x_*. If $\sigma < 1 - \frac{1}{\sqrt{2}}$, then there exists an integer i_1 such that for $i \geq i_1$ and $0 \leq l \leq l_i$*

$$\beta_i^{(l)} = 1.$$

Proof. We need to prove that for sufficiently large i

$$p_{\mu_i}(x_k + Y_k v_k) \leq p_{\mu_i}(x_k) + \sigma \nabla p_{\mu_i}(x_k)^T Y_k v_k \tag{4.2}$$

whenever (2.14 b) does not hold. Lemma 3.6 tells us that both $\|c(x_k)\|$ and d_k tend to zero as $k \to \infty$. Thus, from the definition of v_k and d_k, we get that $\|Y_k v_k\|^2/\mu_i \leq K_0^2\|c(x_k)\|^2/\mu_i \to$

0. Since $\|c(x_k)\| \to 0$ and $\mu_i \to 0$, a simple calculation yields

$$v_k^T Y_k^T \nabla^2 p_{\mu_i}(x_k)\ Y_k v_k = \frac{\|c(x_k)\|^2}{\mu_i} + o(\frac{\|Y_k v_k\|^2}{\mu_i}) = \frac{\|c(x_k)\|^2}{\mu_i} + o(\frac{\|c(x_k)\|^2}{\mu_i}).$$

From Taylor's theorem, we obtain that

$$\begin{aligned} p_{\mu_i}(x_k + Y_k v_k) &= p_{\mu_i}(x_k) + \nabla p_{\mu_i}(x_k)^T\ Y_k v_k \\ &\quad + \frac{1}{2} v_k^T Y_k^T\ \nabla^2 p_{\mu_i}(x_k)\ Y_k v_k + o(\frac{\|Y_k v_k\|^2}{\mu_i}) \\ &= p_{\mu_i}(x_k) + \sigma \nabla p_{\mu_i}(x_k)^T Y_k v_k \\ &\quad + (1-\sigma)\nabla p_{\mu_i}(x_k)^T\ Y_k v_k + \frac{\|c(x_k)\|^2}{2\mu_i} + o(\frac{\|c(x_k)\|^2}{\mu_i}). \end{aligned} \tag{4.3}$$

It follows from (2.21) and (2.14 b) that

$$\nabla p_{\mu_i}(x_k)^T Y_k v_k \leq \|\lambda_k\|\ \|c(x_k)\| - \frac{1}{\mu_i}\|c(x_k)\|^2 \leq -(1-\sigma)\frac{\|c(x_k)\|^2}{\mu_i}. \tag{4.4}$$

Since $-(1-\sigma)^2 + \frac{1}{2} < 0$ by assumption, (4.4) implies that

$$\begin{aligned} (1-\sigma)\nabla p_{\mu_i}(x_k)^T Y_k v_k + \frac{\|c(x_k)\|^2}{2\mu_i} &+ o(\frac{\|c(x_k)\|^2}{\mu_i}) \\ &\leq [-(1-\sigma)^2 + \frac{1}{2}]\frac{\|c(x_k)\|^2}{\mu_i} + o(\frac{\|c(x_k)\|^2}{\mu_i}) \\ &\leq 0. \end{aligned} \tag{4.5}$$

Combining (4.3) and (4.5), we obtain (4.2). ∎

Due to Lemma 4.2, we assume that $\beta_i^{(l)} = 1$. Therefore, we can establish two important inequalities in Lemma 4.3.

LEMMA 4.3. *Suppose Assumption 4.1 holds. Assume the sequence generated by Algorithm 2.1 converges to x_*. Then there exists an integer $i_3 > 0$ and a constant $K_3 > 0$ such that all $0 \leq l \leq l_i$*

$$\|c(x_i^{(l)})\| \leq K_3\ \mu_{i-1} \tag{4.6}$$

and

$$\|c(x_i^{(l+)})\| \leq K_3\ \mu_i \tag{4.7}$$

whenever $i \geq i_3$.

Proof. First , by Assumption 4.1, there exists a constant $K_3 \geq 1$ such that $\Lambda_i^{(l)} \leq K_3$. If (2.14 b) holds, then, by Algorithm 2.1, $x_i^{(l+)} = x_i^{(l)}$. Hence

$$\|c(x_i^{(l+)})\| = \|c(x_i^{(l)})\| \leq \Lambda_i^{(l)}\ \mu_i \leq K_3\ \mu_i \leq \rho K_3 \mu_{i-1}. \tag{4.8}$$

which means that both (4.6) and (4.7) are true.

If (2.14 b) does not hold, then, since $\beta_i^{(l)} = 1$ when i is sufficiently large, Taylor's theorem yields that there exists a constant $N_2 > 0$ such that

$$\begin{aligned}\|c(x_i^{(l+)})\| &= \|c(x_i^{(l)} + Y_i^{(l)} v_i^{(l)})\| \\ &= \|c(x_i^{(l)}) + (A_i^{(l)})^T Y_i^{(l)} v_i^{(l)}\| + \mathcal{O}(\|v_i^{(l)}\|^2) \\ &\le \|c(x_i^{(l)}) + (R_i^{(l)})^T v_i^{(l)}\| + N_2 \|c(x_i^{(l)})\|^2 \\ &= N_2 \|c(x_i^{(l)})\|^2. \end{aligned} \tag{4.9}$$

We prove (4.6) by induction as follows. Algorithm 2.1 shows that (4.6) is true for $l = 0$. Suppose that (4.6) holds for $l = j$. It follows from (2.5) and (2.18) that there exists a constant $K > 0$ such that

$$\|c(x_i^{(j+1)})\| \le \|c(x_i^{(j+)})\| + \frac{K}{\kappa^3}\|h_i^{(j)}\|^3 \le \|c(x_i^{(j+)})\| + K\mu_{i-1}^{6/5}. \tag{4.10}$$

Thus, combining (4.8) (4.9), and (4.10), we obtain that

$$\|c(x_i^{(j+1)})\| \le \begin{cases} (\rho K_3 + K\mu_{i-1}^{1/5})\mu_{i-1} & \text{if (2.14 b) holds for } l = j, \\ N_2\|c(x_i^{(j)})\|^2 + K\mu_{i-1}^{6/5} & \text{otherwise.} \end{cases} \tag{4.11}$$

If i is sufficiently large such that

$$\mu_{i-1} \le \min\{\ [\frac{(1-\rho)K_3}{K}]^5,\ \frac{\rho}{N_2 K_3}\ \},$$

then (4.11) yields that (4.6) holds for $l = j+1$. Thus (4.6) holds for any integer $0 \le l \le l_i$ whenever $i \ge i_3$.

Since $\mu_{i-1}^{6/5} = \mu_i$, it follows from (4.6) and (4.9) that (4.7) is true as long as i is sufficiently large such that $N_2 K_3 \mu_{i-1}^{4/5} \le 1$. ∎

In Algorithm 2.1, we use $q_i^{(l)}(h)$ as a quadratic model for function $p_{\mu_i}(u_i^{(l)}(h))$. We show in Lemma 4.4 that $q_i^{(l)}(h)$ is a good model. That is, $q_i^{(l)}(h)$ is a good approximation to $p_{\mu_i}(u_i^{(l)}(h)) - p_{\mu_i}(u_i^{(l)}(0))$ even if μ_i tends to zero. Based on Lemma 4.4, we illustrate in Lemma 4.5 that $h_i^{(l)}$ is eventually an approximate Newton step, i.e., $\|H(x_i^{(l+)})h_i^{(l)} + g(x_i^{(l+)})\| \le \epsilon\|g(x_i^{(l+)})\|$, where $g(x) = Z(x)^T\nabla f(x)$. This is the key to the proof of our local convergence results. In Lemma 4.6 we establish a relation between the value $\|Z(x)^T\nabla f(x)\| + \|c(x)\|$ at x_{k+1} and that at x_k. Finally, we obtain the local convergence rate of Algorithm 2.1 in Theorem 4.1.

LEMMA 4.4. *Suppose Assumption 4.1 holds. Suppose the sequence $\{x_i^{(l)}\}$ is generated by Algorithm 2.1. Then there exist i_6 and K_6 such that for all $i \ge i_6$ and $0 \le l \le l_i$*

$$|r_i^{(l)} - 1| \le K_6\|h_i^{(l)}\|^3/|q_i^{(l)}(h_i^{(l)})| \tag{4.12}$$

where K_6 is a constant independent of μ_i.

Proof. Due to (2.18) and (2.19), we know that, for i large enough, $\|h_i^{(l)}\| \le \kappa\mu_{i-1}^{2/5} \le \kappa\mu_i^{1/3}$. Thus, since $u_i^{(l)}(0) = x_i^{(l+)}$, (2.5) and (4.7) imply that

$$\begin{aligned}
\|c(u_i^{(l)}(h_i^{(l)}))\|^2 &= \|c(u_i^{(l)}(0))\|^2 + \mathcal{O}(\|c(u_i^{(l)}(0))\| \, \|h_i^{(l)}\|^3 + \|h_i^{(l)}\|^6) \\
&= \|c(u_i^{(l)}(0))\|^2 + \mathcal{O}(\mu_i \, \|h_i^{(l)}\|^3).
\end{aligned}$$

Therefore, there exists an integer $i_5 > 0$ and a constant K_5, independent of μ_i, such that for $i \ge i_5$ and $0 \le l \le l_i$

$$\begin{aligned}
|\,[p_{\mu_i}(u_i^{(l)}(h_i^{(l)})) - p_{\mu_i}(u_i^{(l)}(0)) &- [f(u_i^{(l)}(h_i^{(l)})) - f(u_i^{(l)}(0))]\,| \\
&= |\frac{\|c(u_i^{(l)}(h_i^{(l)}))\|^2 - \|c(u_i^{(l)}(0))\|^2}{2\mu_i}| \\
&\le \frac{\mathcal{O}(\mu_i\|h_i^{(l)}\|^3)}{2\mu_i} \le K_5\|h_i^{(l)}\|^3. \qquad (4.13)
\end{aligned}$$

Define

$$G_i^{(l)} = \int_0^1 \nabla_h^2 f(u(\tau h_i^{(l)}))(1-\tau)d\tau \qquad \text{and} \qquad H_i^{(l+)} = Z(x_i^{(l+)})^T \nabla^2 L(x_i^{(l+)}) Z(x_i^{(l+)}).$$

Since $u(0) = x_i^{(l+)}$ and using (2.6), we have that $H_i^{(l+)} = \nabla_h^2 f(u_i^{(l)}(0))$. By the Lipschitz continuity of $\nabla^2 L(x)$ and $u(h)$, there exists a constant N_1 such that for any $0 \le \tau \le 1$,

$$\|\nabla_h^2 f(u_i^{(l)}(\tau h_i^{(l)})) - H_i^{(l+)}\| = \|\nabla_h^2 f(u_i^{(l)}(\tau h_i^{(l)})) - \nabla_h^2 f(u(0))\| \le N_1\|h_i^{(l)}\|.$$

Thus, we have

$$\|G_i^{(l)} - \frac{1}{2}H_i^{(l+)}\| = \|\int_0^1 [\nabla_h^2 f(u(\tau h_i^{(l)})) - H_i^{(l+)}](1-\tau)d\tau\| \le N_1\|h_i^{(l)}\|. \qquad (4.14)$$

Using Taylor's theorem, it follows from (4.14) that

$$\begin{aligned}
f(u_i^{(l)}(h_i^{(l)})) - f(u_i^{(l)}(0)) &= g(x_i^{(l+)})^T h_i^{(l)} + (h_i^{(l)})^T G_i^{(l)} h_i^{(l)} \\
&\le q_i^{(l)}(h_i^{(l)}) + \|G_i^{(l)} - \frac{1}{2}H_i^{(l+)}\|\|h_i^{(l)}\|^2 \\
&\le q_i^{(l)}(h_i^{(l)}) + N_1\|h_i^{(l)}\|^3. \qquad (4.15)
\end{aligned}$$

Combining (4.13) and (4.15), we obtain that

$$\begin{aligned}
|\,[p_{\mu_i}(u_i^{(l)}(h_i^{(l)})) - p_{\mu_i}(u_i^{(l)}(0))] &- q_i^{(l)}(h_i^{(l)})\,| \\
&\le |\,[f(u_i^{(l)}(h_i^{(l)})) - f(u_i^{(l)}(0))] - q_i^{(l)}(h_i^{(l)})\,| + K_5\,\|h_i^{(l)}\|^3 \\
&\le (N_1 + K_5)\|h_i^{(l)}\|^3.
\end{aligned}$$

Therefore, (4.12) holds with $K_6 = N_1 + K_5$. ∎

When $x_i^{(l+)}$ is sufficiently close to x_*, $H(x_i^{(l+)})$ is positive definite. By means of Lemma 4.4 we prove in Lemma 4.5 that, if (4.16) is satisfied, then an approximate Newton step is taken. i.e., inequality (4.17) holds. It is obvious from Algorithm 2.1 that (4.16) is true for $l = 0$. In Lemma 4.6 we exhibit inductively that (4.16) holds for all l with $0 \leq l < l_i$ when i is sufficiently large. Thus, approximate Newton steps are eventually taken for all $0 \leq l < l_i$ when i is sufficiently large.

LEMMA 4.5. *Suppose Assumption 4.1 holds. Assume the sequence generated by Algorithm 2.1 converges to x_*. Then there exists an integer i_7 such that for all $i \geq i_7$ and $0 \leq l \leq l_i$, if*

$$\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| \leq \mu_{i-1}^{1/2}, \tag{4.16}$$

then

$$\|H_i^{(l+)} h_i^{(l)} + Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})\| \leq \epsilon \, \|Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})\|. \tag{4.17}$$

Proof. We first show that, regardless of inequality (4.16), $\Delta_i^{(l)} = \kappa \mu_i^{2/5}$ for i sufficiently large. It follows from Algorithm 2.1 and (4.6) that

$$\|x_i^{(l+)} - x_i^{(l)}\| \leq \|Y_i^{(l)} v\| \leq K_0 \|c(x_i^{(l)})\| \leq K_0 K_3 \mu_{i-1},$$

which yields that $\|x_i^{(l+)} - x_i^{(l)}\| \to 0$. Thus $x_i^{(l+)}$ converges to x_* since $x_i^{(l)}$ tends to x_*. Since $H(x_*) = Z_*^T \nabla^2 L(x_*, \lambda_*) Z_*$ is positive definite, $H(x_i^{(l+)})$ is positive definite for i sufficiently large. Due to the continuity of $H(x)$, it follows that $\|H(x_i^{(l+)})\| \leq 2\|H(x_*)\|$ and $\|H(x_i^{(l+)})^{-1}\| \leq 2\|H(x_*)^{-1}\|$ when i is sufficiently large.

According to Algorithm 2.1, when $H(x_i^{(l+)})$ is positive definite, Conditions 3 in section 2 yields that either

$$\exists\, \bar{g} \in \Re^{n-m} \text{ satisfying } \|g(x_i^{(l+)}) - \bar{g}\| \leq \epsilon \, \|g(x_i^{(l+)})\| \text{ and } \|H(x_i^{(l+)})^{-1} \bar{g}\| > \Delta_i^{(l)} \tag{4.18}$$

or

$$\|H(x_i^{(l+)}) h_i^{(l)} + g(x_i^{(l+)})\| \leq \epsilon \, \|g(x_i^{(l+)})\|. \tag{4.19}$$

If (4.18) holds, then $\|h_i^{(l)}\| \leq \Delta_i^{(l)} < \|H(x_i^{(l+)})^{-1}\| \, [\, \|\bar{g} - g(x_i^{(l+)})\| + \|g(x_i^{(l+)})\| \,]$. Otherwise, (4.19) is true. Note that $\|h_i^{(l)}\| \leq \|H(x_i^{(l+)})^{-1}\| \, \| \, [\, H(x_i^{(l+)}) h_i^{(l)} + g(x_i^{(l+)}) \,] - g(x_i^{(l+)}) \, \|$. Thus, in either case, we have

$$\|h_i^{(l)}\| \leq \|H(x_i^{(l+)})^{-1}\| \, [\epsilon \, \|g(x_i^{(l+)})\| + \|g(x_i^{(l+)})\| \,] \leq 2 \, \|H(x_i^{(l+)})^{-1}\| \, \|g(x_i^{(l+)})\|.$$

From (2.15), we get that

$$\begin{aligned} q_i^{(l)}(h_i^{(l)}) &\leq -\gamma_1 \|g(x_i^{(l+)})\| \min\{\Delta_i^{(l)},\ \|g(x_i^{(l+)})\|/\|H(x_i^{(l+)})\|\} \\ &\leq -\gamma_1 \frac{\|h_i^{(l)}\|}{\|H(x_i^{(l+)})^{-1}\|} \min\{\|h_i^{(l)}\|,\ \frac{\|h_i^{(l)}\|}{\|H(x_i^{(l+)})\|\|H(x_i^{(l+)})^{-1}\|}\}. \\ &\leq -\frac{\gamma_1\ K_7}{16} \frac{\|h_i^{(l)}\|^2}{\|H(x_*)^{-1}\|}, \end{aligned}$$

where $K_7 = 1/[\|H(x_*)\|\|H(x_*)^{-1}\|]$. It follows from (3.10) and (4.12) that as $i \to \infty$,

$$|r_i^{(l)} - 1| \leq \frac{\|h_i^{(l)}\|^3}{|q_i^{(l)}(h_i^{(l)})|} \leq \frac{16 K_6 \|H(x_*)^{-1}\|}{K_7 \gamma_1} \|h_i^{(l)}\| \to 0.$$

Therefore, when i is sufficiently large, for all l

$$r_i^{(l)} \geq \eta_2.$$

According to Algorithm 2.1, when $r_i^{(l)} \geq \eta_2$, we enlarge the radius $\Delta_i^{(l)}$ until it is equal to $\kappa\mu_{i-1}^{2/5}$. Since μ_i decreases, there exists i_7 such that $\Delta_i^{(l)} = \kappa\mu_{i-1}^{2/5}$ holds for $i \geq i_7$.

Using equality $\Delta_i^{(l)} = \kappa\mu_{i-1}^{2/5}$, we prove now that (4.17) holds if inequality (4.16) is true. Since $v_i^{(l)} = -(R_i^{(l)})^{-T} c(x_i^{(l)})$, Taylor's theorem implies that

$$\begin{aligned} \|Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})\| &\leq \|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| + K\|Y_i^{(l)} v_i^{(l)}\| \\ &\leq \|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| + \hat{K}\|c(x_i^{(l)})\| \end{aligned} \tag{4.20}$$

where K and $\hat{K}$ are constants. If (4.16) holds, it follows from (4.6) that for i sufficiently large

$$\|Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})\| \leq \mu_{i-1}^{1/2} + \hat{K} K_3 \mu_{i-1} \leq 2\mu_{i-1}^{1/2}. \tag{4.21}$$

That is, $\|g(x_i^{(l+)})\| \leq 2\mu_{i-1}^{1/2}$ where $g(x)$ denotes $Z(x)^T \nabla f(x)$.

Since $\|H(x_i^{(l+)})^{-1}\| \leq K$ for a constant $K > 0$, $\Delta_i^{(l)} = \kappa\mu_{i-1}^{2/5}$, and μ_i tends to zero, we get that

$$\|H(x_i^{(l+)})^{-1} g(x_i^{(l+)})\| \leq K\|g(x_i^{(l+)})\| \leq 2K\mu_{i-1}^{1/2} \leq \kappa\mu_{i-1}^{2/5} = \Delta_i^{(l)}, \tag{4.22}$$

when i is large enough such that $\mu_{i-1} \leq (\kappa/2K)^{10}$. Thus, it follows from Condition 3 in section 2 that (4.17) holds. ∎

To prove Lemma 4.6, we need inequality (4.24). By the definition of $u(h)$ in (2.3), we have

$$u(h) - u(0) = s(Zh) = Zh - \frac{1}{2} Y R^{-1}[h^T Z^T \nabla^2 c(x_c) Zh].$$

It follows from the boundedness of $\|Z_i^{(l+)}\|$, $\|Y_i^{(l+)}(R_i^{(l+)})^{-1}\|$ and $\nabla^2 c(x_i^{(l+)})$ that

$$x_i^{(l+1)} - x_i^{(l+)} = s(Z_i^{(l+)} h_i^{(l)}) = Z_i^{(l+)} h_i^{(l)} + \mathcal{O}(\|h_i^{(l)}\|^2). \tag{4.23}$$

Since $x_i^{(l+)} - x_i^{(l)} = Y_i^{(l)} v_i^{(l)}$, (4.23) yields

$$\begin{aligned} x_i^{(l+1)} - x_i^{(l)} &= x_i^{(l+1)} - x_i^{(l+)} + Y_i^{(l)} v_i^{(l)} \\ &= Z_i^{(l+)} h_i^{(l)} + Y_i^{(l)} v_i^{(l)} + \mathcal{O}(\|h_i^{(l)}\|^2) \\ &= Z_i^{(l+)} h_i^{(l)} + Y_i^{(l+)} v_i^{(l)} + (Y_i^{(l)} - Y_i^{(l+)}) v_i^{(l)} + \mathcal{O}(\|h_i^{(l)}\|^2) \\ &= Z_i^{(l+)} h_i^{(l)} + Y_i^{(l+)} v_i^{(l)} + o(\|h_i^{(l)}\| + \|v_i^{(l)}\|). \end{aligned}$$

By the definition of Y and Z, it is easy to see that

$$\begin{aligned} \|x_i^{(l+1)} - x_i^{(l)}\| &\geq \|Z_i^{(l+)} h_i^{(l)} + Y_i^{(l+)} v_i^{(l)}\| - \frac{1}{4}[\|h_i^{(l)}\| + \|v_i^{(l)}\|] \\ &\geq (\|h_i^{(l)}\|^2 + \|v_i^{(l)}\|^2)^{1/2} - \frac{1}{4}[\|h_i^{(l)}\| + \|v_i^{(l)}\|] \\ &\geq \frac{1}{4}\|h_i^{(l)}\|. \end{aligned}$$

Thus, there exist an integer i_8 and a constant K_8 such that for $i \geq i_8$

$$\|h_i^{(l)}\|^2 \leq K_8[\|x_i^{(l+1)} - x_*\|^2 + \|x_i^{(l)} - x_*\|^2]. \tag{4.24}$$

Before proving Lemma 4.6, we define

$$\delta_i^{(l)} = H(x_i^{(l+)}) h_i^{(l)} + g(x_i^{(l+)}).$$

That is, $\delta_i^{(l)}$ is the residual when linear system $H(x_i^{(l+)}) h_i^{(l)} = -g(x_i^{(l+)})$ is solved.

LEMMA 4.6. *Suppose Assumption 4.1 holds. Assume the sequence generated by Algorithm 2.1 converges to x_*. Then there exist an integer i_9 and a constant $K_9 > 0$ such that for $0 \leq l < l_i$, if*

$$\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| \leq \mu_{i-1}^{1/2}, \tag{4.25}$$

then

$$\begin{aligned} \|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\| \leq\; & K_9[\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\|^2 + \|c(x_i^{(l)})\|^2] \\ & + 2[\|c(x_i^{(l+)})\| + \|\delta_i^{(l)}\|] \end{aligned} \tag{4.26}$$

and

$$\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| \leq \mu_{i-1}^{1/2} \tag{4.27}$$

whenever $i \geq i_9$.

Proof. Due to the definition of $\delta_i^{(l)}$ and equality $x_i^{(l+1)} = x_i^{(l+)} + s(Z_i^{(l+)} h_i^{(l)})$, Taylor's theorem and (4.23) yields

$$\begin{aligned} Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)}) &= Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)}) \\ &\quad + [Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})] s(Z_i^{(l+)} h_i^{(l)}) + \mathcal{O}(\|h_i^{(l)}\|^2) \\ &= -H_i^{(l+)} h_i^{(l)} + [Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})]'_x \, Z(x_i^{(l+)}) h_i^{(l)} \\ &\quad + \delta_i^{(l)} + \mathcal{O}(\|h_i^{(l)}\|^2). \end{aligned} \tag{4.28}$$

It follows from [7] that $[Z(x)^T \nabla f(x)]'_x = Z(x)^T \nabla^2 L(x)$. Thus, by the definition of $H_i^{(l+)}$ and (4.28), we have that $\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| = \|\delta_i^{(l)}\| + \mathcal{O}(\|h_i^{(l)}\|^2)$. In addition to (2.5) in which $c(x_i^{(l+1)}) = c(x_i^{(l+)}) + \mathcal{O}(\|h_i^{(l)}\|^3)$, we obtain that there exists a constant $K > 0$ such that

$$\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\| \le \|c(x_i^{(l+)})\| + \|\delta_i^{(l)}\| + K \, \|h_i^{(l)}\|^2. \tag{4.29}$$

Inequalities (4.24) and (4.1) imply that for i large enough

$$\begin{aligned} \|h_i^{(l)}\|^2 &\le \xi_i^{(l)} \|x_i^{(l+1)} - x_*\| + K_8 \|x_i^{(l)} - x_*\|^2 \\ &\le \frac{\xi_i^{(l)}}{\sqrt{K_1}} [\, \|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\| \,] + K_8 \|x_i^{(l)} - x_*\|^2 \\ &\le \frac{1}{2K} [\, \|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\| \,] + K_8 \|x_i^{(l)} - x_*\|^2, \end{aligned} \tag{4.30}$$

where $\xi_i^{(l)} = K_8 \|x_i^{(l+1)} - x_*\| \to 0$. Combining (4.29) and (4.30), it follows from (4.1) that for i large enough

$$\begin{aligned} \frac{1}{2} [\, \|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\| \,] &\le \|c(x_i^{(l+)})\| + \|\delta_i^{(l)}\| + K K_8 \|x_i^{(l)} - x_*\|^2 \\ &\le \frac{K K_8}{K_1} [\, \|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\|^2 + \|c(x_i^{(l)})\|^2 \,] \\ &\quad + \|c(x_i^{(l+)})\| + \|\delta_i^{(l)}\|. \end{aligned}$$

Therefore (4.26) holds with $K_9 = \frac{2KK_8}{K_1}$.

From (4.25), similar to the proof of Lemma 4.5, we obtain that (4.21) and (4.17) hold. i.e., $\|g(x_i^{(l+)})\| \le 2 \, \mu_{i-1}^{1/2}$ and $\|\delta_i^{(l)}\| \le \epsilon \, \|g(x_i^{(l+)})\| \le 2 \, \epsilon \, \mu_{i-1}^{1/2}$ with $0 \le \epsilon < 1/5$. Therefore, it follows from (4.26), (4.25), (4.6), (4.7), and (4.21) that

$$\begin{aligned} \|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| &\le K_9 [\, \mu_{i-1} + K_3^2 \mu_{i-1}^2 \,] + 2K_3 \mu_i + 2\|\delta_i^{(l)}\| \\ &\le (K_9 + K_3^2 \mu_{i-1} + 2K_3) \mu_{i-1} + 4\epsilon \, \mu_{i-1}^{1/2} \\ &\le \mu_{i-1}^{1/2}. \end{aligned}$$

when i is large enough. ∎

The results in Lemma 4.6 pave the way for quadratic convergence. Since inequality (4.25) is always true for $l = 0$, it follows inductively from Lemma 4.6 that (4.25) and (4.26) hold for all l with $0 \le l < l_i$ when i is sufficiently large.

THEOREM 4.1. *Suppose Assumption 4.1 holds. Suppose the sequence generated by Algorithm 2.1 converges to x_*. Then there exists a neighborhood N of x_* such that for $x_k \in N$, $\{x_k\}$ converges to x_* superlinearly if*

$$\delta_i^{(l)} = o(\|Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})\|). \tag{4.31}$$

Moreover, $\{x_k\}$ converges to x_* quadratically if

$$\delta_i^{(l)} = \mathcal{O}(\|Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})\|^2). \tag{4.32}$$

Proof. Since $x_i^{(l)} \to x_*$, it follows from Assumption 4.1 that for i sufficiently large, the reduced Hessian matrix at $x_i^{(l)}$ is positive definite. Therefore, inequality (2.14 c) is always true.

To establish the local convergence rate, we first prove that there exists a constant $K > 0$ such that when i is sufficiently large

$$\|c(x_i^{(l+)})\| \le K[\ \|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\|^2 + \|c(x_i^{(l)})\|^2\ ] \tag{4.33}$$

for $0 \le l < l_i$. Actually, if (2.14 b) does not hold, then, similar to the proof of inequality (4.9), we know that there exists a constant $N_2 > 0$ such that $\|c(x_i^{(l+)})\| \le N_2 \|c(x_i^{(l)})\|^2$, which implies (4.33). On the other hand, when (2.14 b) holds with $0 \le l < l_i$, we are in the inner loop. According to Algorithm 2.1, at least one inequality in (2.14) does not hold when we are in the inner loop. Because inequality (2.14 c) is always true when i is sufficiently large, inequality (2.14 a) must not hold for l with $0 \le l < l_i$. That is, $\mu_i < \|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\|^2$. Thus, inequality (4.8) yields that $\|c(x_i^{(l+)})\| \le K_3 \mu_i \le K_3 \|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\|^2$.

Therefore, inequality (4.33) is always true no matter whether (2.14 b) holds or not. Thus, combining (4.26) and (4.33), we have

$$\begin{aligned} \|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\| \quad &\le \quad (K_9 + 2K)[\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\|^2 \\ &\qquad + \|c(x_i^{(l)})\|^2] + 2\|\delta_i^{(l)}\|. \end{aligned} \tag{4.34}$$

If (4.31) holds, then it follows from (4.20) and (4.34) that

$$\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\| = o(\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| + \|c(x_i^{(l)})\|).$$

Thus, Lemma 4.1 implies that Algorithm 2.1 has a superlinear convergence rate.

If (4.32) holds, then it follows from (4.20) and (4.34) that

$$\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\| = \mathcal{O}(\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\|^2 + \|c(x_i^{(l)})\|^2).$$

Thus, Algorithm 2.1 has a quadratic convergence rate. ■

5. Numerical Experiments. In the final section, we present preliminary computational results to illustrate the performance of Algorithm 2.1. We test implementations of Algorithm 2.1 on a set of nonlinear equality constrained problems from the CUTE collection [1] and our own test problems. To exhibit the important role the change of variables plays in Algorithm 2.1, we compare the implementations of Algorithm 2.1 to a variation of our algorithm without change of variables. All our experiments were performed in MATLAB Version 4.1 on a Sun 4/670 workstation.

problems	n	m	$nnz(A)$	constraints
BT6	5	2	5	nonlinear
BT11	5	3	8	nonlinear
DIPIGRI	7	4	19	nonlinear
DTOC2	58	36	144	nonlinear
DTOC4	29	18	65	nonlinear
DTOC6	21	10	31	nonlinear
GENHS28	300	298	894	linear
HS100	7	4	19	nonlinear
MWRIGHT	5	3	8	nonlinear
ORTHREGA	517	256	1792	nonlinear
ORTHREGC	505	250	1750	nonlinear
ORTHREGD	203	100	500	nonlinear
TEST1	500	300	2308	quadratic
TEST2	500	300	2661	nonlinear

Table I: Description of Problems

Table I gives a brief description of our test problem set. Most problems in Table I (all except TEST1 and TEST2) are selected from the CUTE collection [1]. Problems TEST1 and TEST2 are our own test problems. Problem TEST1 is to minimize a Rosenbrock function [11] subject to some quadratic equality constraints, i.e.,

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n-1}[(1-x_i)^2 + 100(x_{i+1} - x_i^2)^2] \\ \text{subject to} & a_i^T x + .5 x^T M_i x = 0, \qquad i = 1, \ldots, m, \end{array} \tag{5.1}$$

where $a_i \in \Re^n$, $i = 1, 2, \ldots, m$, are vectors, and $M_i \in \Re^{n \times n}$, $i = 1, 2, \ldots, m$, are symmetric. To increase nonlinearity, in the TEST2 problems we add some perturbation functions the objective function and the quadratic constraint functions in (5.1). Namely, we solve the following problem:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n-1}[(1-x_i)^2 + 100(x_{i+1}-x_i^2)^2] + \delta_0(x) \\ \text{subject to} & a_i^T x + .5x^T M_i x + \delta_i(x) = 0, \quad i = 1,\dots,m, \end{array} \tag{5.2}$$

where $\delta_0(x)$, $\delta_i(x)$, $i = 1, 2, \dots, m$, are perturbation functions. The perturbation functions are generated randomly to be linear combinations of polynomials, trigonometry functions, logarithmic functions and exponential functions. For example, $\delta_0(x)$ could be

$$\delta_0(x) = (x_1^2 + x_4)^2 + 1.0 + \log(1 + x_2^2 + x_3^2) + 10\sin(2\pi x_5)\cos(2\pi x_6) - e^{-(x_5-x_3)^2} + \dots.$$

In problems TEST1 and TEST2, the matrices $[a_1, a_2, \dots, a_m]$ and M_i, $i = 1, 2, \dots, m$, were created randomly.

As we mentioned in section 2, we approximately solve subproblem $\min\{q_i^{(l)}(h) \ : \ \|h\|_2 \le \Delta_i^{(l)}\}$ for $Z_i^{(l+)}h$ by considering a 2-dimensional subproblem at each iteration:

$$\min\{q_i^{(l)}(h) \ : \ \|h\|_2 \le \Delta_i^{(l)}, \ h \in [g(x_i^{(l+)}), \ d] \ \}, \tag{5.3}$$

where

$$d = \begin{cases} d_{newton} & \text{if } H_i^{(l+)} \text{ is positive definite,} \\ d_{neg} & \text{otherwise} \end{cases}$$

$$H_i^{(l+)} d_{newton} + g(x_i^{(l+)}) = \delta_i^{(l)}, \tag{5.4}$$

and

$$d_{neg}^T \ H_i^{(l+)} \ d_{neg} \le \gamma \ \mathrm{eig}_{\min}(H_i^{(l+)}) \ \|d_{neg}\|_2^2. \tag{5.5}$$

In (5.5), $\gamma > 0$ is a constant and $\mathrm{eig}_{\min}(H_i^{(l+)})$ denotes the smallest eigenvalue of $H_i^{(l+)}$. In [4] it is proved that the solution to (5.3), $h = h_i^{(l)}$, satisfies conditions 1, 2 and 3 stated in section 2. In practice, inequality (5.5) is difficult to satisfy. We relax (5.5) and require only $d_{neg}^T \ H_i^{(l+)} \ d_{neg} \le 0$. However, this relaxation does have a (theoretical) cost; the resultant algorithm no longer guarantees convergence to a point satisfying the second order necessary optimality conditions.

We compute the direction d by attempting the Cholesky factorization of $H(x_i^{(l+)})$. If $H(x_i^{(l+)})$ is positive definite, we solve (5.4) for a Newton direction d_{newton}. A negative curvature direction d_{neg} is obtained when the Cholesky factorization fails (see, e.g., Section 4.4 in [15]). Such a direction d_{neg} satisfies $d_{neg}^T \ H_i^{(l+)} \ d_{neg} \le 0$.

We use the curve technique, $x_{k+1} = x_{k+} + s(Z_{k+}h)$, only when necessary. In particular, we take

$$x_{k+1} = x_{k+} + Z_{k+}h \tag{5.6}$$

when $r_k \geq \eta$. Otherwise, we use the curve technique to update the iterates:

$$x_{k+1} = x_{k+} + s(Z_{k+}h) \tag{5.7}$$

where $s(Z_{k+}h) = Z_{k+}h + Y_{k+}R_{k+}^{-T}[c(x_{k+}) - c(x_{k+} + Z_{k+}h)]$.

When solving problems in Table I using Algorithm 2.1, we set the stopping criterion to be $\mu < \mu_{\min} = 10^{-7}$. We take $\mu_0 = 1.0$, $\Delta_0 = \|Z(x_0)^T \nabla f(x_0)\|^{1/2}$, $\rho = 0.1$, and $\sigma = 0.0001$. For both TEST1 and TEST2, the starting points are $x_0 = [.5,\ .5,\ \ldots,\ .5]^T$. For the test problems drawn from the CUTE collection, we take the CUTE default values.

Problems	number of iterations	function evaluations $f,\ c$	$g,\ A$	H	*error*
BT6	10	69	20	10	$O(10^{-7})$
BT11	6	19	13	6	$O(10^{-9})$
DIPIGRI	12	39	19	12	$O(10^{-7})$
DTOC2	12	32	17	9	$O(10^{-6})$
DTOC4	3	7	7	3	$O(10^{-6})$
DTOC6	6	20	13	6	$O(10^{-10})$
GENHS28	2	4	4	2	$O(10^{-14})$
HS100	11	27	20	11	$O(10^{-7})$
MWRIGHT	8	20	16	8	$O(10^{-9})$
ORTHREGA	32	295	49	30	$O(10^{-8})$
ORTHREGC	8	24	17	8	$O(10^{-10})$
ORTHREGD	6	12	10	3	$O(10^{-8})$
TEST1	15	29	25	15	$O(10^{-5})$
TEST2	10	25	21	10	$O(10^{-7})$

Table II: Results Using Algorithm 2.1

Table II illustrates the results of our numerical experiments for problems in Table I. The first column states the names of the problems. The second column shows the number of iterations needed to reach the stopping criteria. Column "function evaluations" gives the number of function evaluations needed for different problems. Sub-columns "f, c" indicate the number of function evaluations needed for the functions $f(x)$ and $c(x)$, "g, A" for their gradients, and "H" for the Hessians of the Lagrangian functions. On the sixth column the quantity "*error*" is defined as

$$error = \sqrt{\|Z(x)^T \nabla f(x)\|_2^2 + \|c(x)\|_2^2}, \tag{5.8}$$

where x is the approximate solution to (1.1).

We also solved the test problems from the CUTE collection in Table I using LANCELOT. When running LANCELOT, we used second derivatives and set the stopping criterion to be $error < 10^{-5}$ where *error* is defined in (5.8). We present the results of our LANCELOT experiments in Table III.

Problems	number of iterations	function evaluations f, c	g, A	H	*error*
BT6	28	28	25	28	$O(10^{-5})$
BT11	19	19	20	19	$O(10^{-6})$
DIPIGRI	52	52	48	52	$O(10^{-6})$
DTOC2	18	18	18	18	$O(10^{-6})$
DTOC4	17	17	17	17	$O(10^{-7})$
DTOC6	22	22	23	22	$O(10^{-5})$
GENHS28	6	6	7	6	$O(10^{-7})$
HS100	52	52	47	52	$O(10^{-6})$
MWRIGHT	19	19	19	19	$O(10^{-6})$
ORTHREGA	92	92	84	92	$O(10^{-5})$
ORTHREGC	30	30	28	30	$O(10^{-5})$
ORTHREGD	44	44	40	44	$O(10^{-5})$

Table III: Results Using LANCELOT

Tables II and III suggest that Algorithm 2.1 is very promising. We can see in Tables II and III that for almost all test problems Algorithm 2.1 takes fewer iterations and function evaluations than LANCELOT. For a few problems (BT6, DTOC2 and ORTHREGA) Algorithm 2.1 takes more function evaluations for "f" and "c" than LANCELOT. This is mainly due to the line search procedure in Algorithm 2.1. But for the same problems (BT6, DTOC2 and ORTHREGA) Algorithm 2.1 takes fewer function evaluations for "g", "A" and "H" than LANCELOT.

Finally, we use problem DTOC6 as an example to illustrate the essential role the change of variables plays in Algorithm 2.1. We use three variations of Algorithm 2.1 to solve problem DTOC6. In the first variation we perform the change of variables at every iteration. The second variation is the same one we used to obtain results in Table II: we try form (5.6) first and perform the change of variables when form (5.6) fails to provide good trust region ratios. In the third variation we do not perform any change of variables at all and adopt form (5.6) for all iterations. The results are presented in Table IV.

Table IV illustrates that the change of variables is significant to implementations of Algorithm 2.1. From Table IV we can see that the first and second variations work well and the third one takes significantly longer to converge due to the Maratos effect. Actually, for almost all problems we tested, the third variation works poorly. Table IV also indicates that the first variation is more efficient than the second for the particular problem DTOC6. However, for most problems in Table I, the second variation is the best.

Variations	number of iterations	function evaluations			error
		f, c	g, A	H	
Variation 1	11	33	19	11	$O(10^{-7})$
Variation 2	50	126	59	50	$O(10^{-5})$
Variation 3	1170	1252	1182	1170	$O(10^{-4})$

Table IV: Comparison of Different Variations

6. Concluding Remarks. The L_2-penalty function is usually dismissed as a merit function for nonlinearly constrained minimization for two reasons: The Hessian matrix of the penalty function becomes singular as the solution is approached, and short steps must often be taken, when far from the solution, due to the effect of the penalty term. Coleman and Hempel [8] and Gould [16] have indicated how to circumvent potential problems due to Hessian singularity at the solution. In this paper we have proposed a new technique which not only avoids problems due to Hessian singularity at the solution, but also avoids excessively small steps when far from the solution.

The crux of our new approach is a local transformation, defined at each iteration, and the approximate solution of a trust region problem in the new (local) variables which allows for the efficient computation of a 'large' curved step in the original variables. This curved step follows a quadratic approximation to the current constraint contour and therefore causes little increase in the penalty term.

We have established global convergence properties and second-order convergence rate results; we have demonstrated how this technique can be efficiently implemented. Finally, we have provided computational results of preliminary numerical experiments. The computational and theoretical results indicate that this approach has considerable practical potential for solving nonlinear equality-constrained minimization problems.

Acknowledgements. The authors would like to thank Dr. Yuying Li for her helpful comments and suggestions on this paper.

REFERENCES

[1] I. Bongartz, A.R. Conn, N.I.M. Could and Ph.L. Toint, "CUTE: constrained and unconstrained testing environment", Research Report, IBM T.J. Watson Research Center, Yorktown Heights, New York, (1993).

[2] R.H. Byrd and J. Nocedal, "An analysis of reduced Hessian methods for constrained optimization", *Mathematical Programming* 49 (1991) 285-323.

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

[4] R.H. Byrd, R.B. Schnabel and G.A. Shultz, "Approximate solution of the trust region problem by minimization over two-dimensional subspaces", *Mathematical Programming* 40 (1988) 247-263.

[5] M.R. Celis, J.E. Dennis and R.A. Tapia, "A trust region strategy for nonlinear equality constrained optimization," in: P.T. Boggs, R.H. Byrd and R.B. Schnabel, eds., *Numerical Optimization 1984* (SIAM, Philadelphia, 1985) 71-82.

[6] T.F. Coleman, private communication.

[7] T.F. Coleman, "On characterizations of superlinear convergence for constrained optimization", *Lectures in Applied Mathematics* 26 (1990) 113-133.

[8] T.F.Coleman and C. Hempel, "Computing a trust region step for a penalty function", *SIAM Journal on Scientific and Statistical Computing* 11 (1990) 180-201.

[9] T.F. Coleman and 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. Could and Ph.L. Toint, "LANCELOT: a FORTRAN package for large-scale nonlinear optimization (Release A)", *No. 17 in Springer Series in Computational Mathematics*, Springer-Verlag, New York, (1992).

[11] J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* , Prentice-Hall, Englewood Cliffs, New Jersey, (1983).

[12] M. El-Alem, "A global convergence theory for the Celis-Dennis-Tapia algorithm for constrained optimization," Tech. Report 88-19, Department of Mathematical Sciences, Rice University, Houston, TX, 1988.

[13] R. Fletcher, *Practical Methods of Optimization*, John Wiley and Sons, Chichester, (1987), 2nd ed.

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

[15] P.E. Gill, W. Murray, and M.H. Wright, *Practical Optimization*, Academic Press, London and New York, (1981).

[16] N.I.M. Gould, "On the accurate determination of search directions for simple differentiable penalty functions", *IMA Journal of Numerical Analysis*, 6 (1986), 357-372.

[17] C.G. Hempel, "A globally convergent quadratic penalty function method with fast local convergence properties", Ph.D Dissertation, Center for Applied Mathematics, Cornell University (1990).

[18] N. Maratos, "Exact penalty function algorithms for finite dimensional and control optimization problems", Ph.D. thesis, University of London (1978).

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

[20] M.J.D. Powell, "Convergence properties of a class of minimization algorithms", in: O.L. Mangasarian, R.R. Meyer and S.M. Robinson, eds., *Nonlinear Programming*, Vol. 2 (Academic Press, New York, 1975) 1-27.

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

[22] G.A. Shultz, R.B. Schnabel and 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.

[23] A. Vardi, "A trust region algorithm for equality constrained minimization: convergence properties and implementation," *SIAM Journal of Numerical Analysis* 22 (1985) 575-591.