A QUASI-NEWTON L_2-PENALTY METHOD FOR MINIMIZATION SUBJECT TO NONLINEAR EQUALITY CONSTRAINTS *

THOMAS F. COLEMAN† AND WEI YUAN‡

Abstract. We present a modified L_2 penalty function method for equality constrained optimization problems. The pivotal feature of our algorithm is that at every iterate we invoke a special change of variables to improve the ability of the algorithm to follow the constraint level sets. This change of variables gives rise to a suitable block diagonal approximation to the Hessian which is then used to construct a quasi-Newton method. We show that the complete algorithm is globally convergent with a local Q-superlinearly convergence rate. Preliminary computational results are given for a few problems.

1. Introduction. We construct a quasi-Newton L_2 penalty method for solving 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. This method possesses both strong global convergence properties and a local superlinear convergence rate by combining an L_2 penalty function method with reduced Hessian approaches.

We solve problem (1.1) by minimizing a sequence of penalty functions $\{p_{\mu_i}(x)\}$ as μ_i decreases to zero, where $\mu_i > 0$ and $p_\mu(x) = f(x) + \frac{1}{2\mu}\|c(x)\|_2^2$. It is well known that under reasonable assumptions a limit point of the sequence of solutions to $\min p_{\mu_i}(x)$ is a solution of (1.1). Unfortunately, ordinary unconstrained minimization techniques often converge slowly when applied to problem $\min p_{\mu_i}(x)$ due to the Maratos effect [18]. To overcome this difficulty, at each iterate x_k a change of variables is performed to improve the ability of the algorithm to follow the constraint level sets. As a result, the next iterate x_{k+1} approximately follows the level set $c(x) = c(x_k)$ so that the value of $\|c(x)\|$ increases only mildly.

Our algorithm attempts to keep the iterates in an envelope with width $\mathcal{O}(\mu_i)$ and containing surface $c(x) = 0$ when moving towards the solution. Whenever the current iterate is outside the envelope, a partition of the Hessian matrix allows for a safeguarded Newton step to reduce

---

* 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

the value of $\|c(x)\|$. For each μ_i certain criteria are incorporated to ensure that $\min p_{\mu_i}(x)$ is approximately solved. After the criteria are satisfied, μ_i is reduced to μ_{i+1}; a new penalty function $p_{\mu_{i+1}}(x)$ is then minimized. The above process is repeated as μ_i approaches zero. When μ_i tends to zero, our algorithm asymptotically assembles a reduced Hessian method.

Reduced Hessian methods determine a search direction d_k at each iterate x_k by solving

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

where $A_k = \nabla c(x_k)$, Z_k is a matrix whose columns form a basis for the null space of A_k, and B_k is an approximation to the reduced Hessian of the Lagrangian function. The next iterate has the form $x_{k+1} = x_k + \alpha_k d_k$ where $\alpha_k > 0$ is a step length obtained from a line search technique.

Reduced Hessian methods are proposed by Coleman and Conn [7], Gabay [11], Gilbert [12], and Murray and Wright [15]. The idea behind reduced Hessian methods is that it is reasonable to approximate the reduced Hessian $Z_k^T \nabla_x^2 L(x_k,\ \lambda_k) Z_k$ using a positive definite update formula since near the solution this matrix is positive definite. Thus matrix B_k in (1.2) is updated so that B_{k+1} is positive definite and $B_{k+1}h_k = y_k$, where $h_k = Z_k^T(x_{k+1} - x_k)$ and y_k is a secant approximation to $Z_k^T \nabla_x^2 L(x_k,\ \lambda_k) Z_k h_k$. In most reduced Hessian methods B_k is updated using Broyden's class of formulas. Coleman and Conn prove that if x_0 and B_0 are sufficiently close to the solution and to the reduced Hessian of the Lagrangian, respectively, then the iterates converge 2-step Q-superlinearly to the solution. Note that all the above mentioned analyses are local results. Recently Byrd and Nocedal [5] present some reduced Hessian methods for (1.1). Under certain conditions Byrd and Nocedal obtain local superlinear convergences for their methods and they also prove a global convergence result. Following Coleman and Conn [7], Byrd and Nocedal choose y_k:

$$y_k = Z_k^T[\nabla_x L(x_k + h_k, \lambda_k) - \nabla_x L(x_k, \lambda_k)]. \tag{1.3}$$

The maintenance of positive definite Hessian approximations is crucial to both the convergence results and the numerical performance of the reduced Hessian methods. Broyden's class of update formulas preserve the positive definiteness of matrices B_k provided $y_k^T h_k > 0$. In the unconstrained setting line search techniques guarantee that $y_k^T h_k > 0$. However, to our knowledge, there is no line search technique for (1.2) that guarantees $y_k^T h_k > 0$ unless it is assumed that the matrices $Z_k^T \nabla_x^2 L(x, \lambda_k) Z_k$ are uniformly positive definite for all x in the line search segment. This assumption does not hold for general nonlinear functions $f(x)$ and $c(x)$ when x_k is far away from the solution. Therefore, reduced Hessian methods which choose y_k as in (1.3) may fail in practice for general functions $f(x)$ and $c(x)$.

In our algorithm we compute the secant approximation vector y_k in a manner different from (1.3). The line search technique in our algorithm guarantees that $y_k^T h_k > 0$; hence, the positive definiteness of B_k is preserved while only assuming that $f(x)$ and $c(x)$ are smooth functions. Thus, our algorithm can be applied to general smooth functions.

Under certain assumptions we prove that our algorithm globally converges to a point x_* which satisfies the first order necessary optimality conditions to (1.1). We also prove that the local convergence rate of the algorithm is Q-superlinear.

Our algorithm is described in section 2. In section 3 we study its global convergence properties; local and superlinear convergence properties are proved in section 4. Our numerical results are presented and discussed in section 5. We make further discussion and concluding remarks in section 6.

2. Algorithm. Difficulties often occur when unconstrained minimization techniques are applied to minimize $p_\mu(x)$. One of those difficulties is that the Hessian matrix of $p_\mu(x)$ becomes singular as μ approaches zero. Another difficulty is referred to as the Maratos effect [18]. When μ and $c(x)$ are small and x is far away from a local minimizer (1.1), the Hessian of $p_\mu(x)$ is dominated, in some directions, by the constraint gradients. This domination causes most unconstrained minimization methods to compute steps almost entirely in the tangent space of the constraints. If the constraints are nonlinear then only small steps are accepted to ensure that $p_\mu(x)$ decreases. Therefore, convergence is unbarably slow.

In this section we construct a modified quasi-Newton method to minimize the L_2 penalty function $p_\mu(x)$. Instead of taking steps almost entirely in the tangent space of the constraints, our algorithm computes each step along a curve which follows a quadratic approximation to the constraint surface in order to avoid the Maratos effect.

Given a point x_k in $\Re^n$. Let $A_k \equiv \nabla c(x_k) \in \Re^{n\times m}$. Suppose that the QR factorization of A_k is

$$A_k = Q_k \bar{R}_k = [Y_k \; Z_k] \begin{bmatrix} R_k \\ 0 \end{bmatrix} = Y_k R_k, \tag{2.1}$$

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

$$w(h,v) \equiv x_k + s_k(h) + Y_k v$$

for $h \in R^{n-m}$ and $v \in R^m$, where

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

is a curve which follows approximately the level sets $c(x) = c(x_k)$. Note that (2.2) is a one-to-one mapping in a neighborhood of x_k. It follows from the chain rule that

$$\nabla_h c(u(0)) = 0 \qquad \text{and} \qquad \nabla_h^2 c(u(0)) = 0. \tag{2.3}$$

Thus

$$c(u(h)) = c(u(0)) + \mathcal{O}(\|h\|^3) = c(x_k) + \mathcal{O}(\|h\|^3) \qquad \text{as } h \to 0. \tag{2.4}$$

Equation (2.4) tells us that along the curve $u(h)$ the value of $\|c(u(h))\|$ changes very slightly. Simple calculations yield

$$\begin{cases} \nabla_h p_\mu(u(0)) = \nabla_h f(u(0)) = Z_k^T \nabla f(x_k) \\ \nabla_h^2 p_\mu(u(0)) = \nabla_h^2 f(u(0)) = Z_k^T \nabla^2 L(x_k) Z_k \equiv H_k, \end{cases} \tag{2.5}$$

where $\nabla^2 L(x) = \nabla^2 f(x) + \sum_{i=1}^m \lambda^i \nabla^2 c_i(x)$ and λ^i is the ith component of the Lagrangian multiplier λ. Note that $\lambda = \lambda_k = -R_k^{-1} Y_k^T \nabla f(x_k)$ at x_k.

Setting $\hat{p}_\mu(h,v) = p_\mu(w(h,v))$, it then follows 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

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

where $\nabla^2 \hat{L}(x_k) = \nabla^2 f(x_k) + \sum_{i=1}^m \frac{c_i(x_k)}{\mu} \nabla^2 c_i(x_k)$. Let us consider the system of linear equations from the Newton's method for $\nabla_{(h,v)} \hat{p}_\mu(h,v) = 0$:

$$\begin{bmatrix} Z_k^T \nabla^2 L(x_k) Z_k & Z_k^T \nabla^2 \hat{L}(x_k) Y_k \\ Y_k^T \nabla^2 \hat{L}(x_k) Z_k & Y_k^T \nabla^2 \hat{L}(x_k) Y_k + \frac{1}{\mu} R_k R_k^T \end{bmatrix} \begin{bmatrix} h \\ v \end{bmatrix} = - \begin{bmatrix} Z_k^T \nabla p_\mu(x_k) \\ Y_k^T \nabla p_\mu(x_k) \end{bmatrix}. \tag{2.6}$$

It follows from Taylor's theorem 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 the system (2.6) can be approximately written as $(Z_k^T \nabla^2 L(x_k) Z_k) h = -Z_k^T \nabla f(x_k + Y_k v)$.

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

$$Y_k^T \nabla p_\mu(x_k) = R_k R_k^{-1} Y_k^T [\nabla f(x_k) + A_k \frac{c(x_k)}{\mu}] = -\frac{1}{\mu} R_k [\mu\, \lambda_k - c(x_k)]. \tag{2.7}$$

It is clear that $\frac{1}{\mu} R_k R_k^T$ plays a dominant role in the lower part of (2.6) as μ tends to zero. 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 (2.6).

Therefore, the system (2.6) can be approximated by

$$\begin{cases} a) \quad R_k^T v = -c(x_k) \\ b) \quad B_k h = -Z_k^T \nabla f(x_k + Y_k v) \end{cases} \tag{2.8}$$

where B_k is an approximation to the reduced Hessian $H_k = Z_k^T \nabla^2 L(x_k) Z_k$. Unlike in (2.6), penalty parameter μ is absent from system (2.8). Thus, the singularity problem of Hessian of $p_\mu(x)$ disappears in (2.8). Note that (2.8 a) is Newton's method for nonlinear system $c(x) = 0$.

Accordingly, we minimize $p_\mu(x)$ using our modified quasi-Newton method as follows. At a point x_k, partition the step vector $(h^T,\ v^T)^T$ and compute v and h separately. First solve (2.8 a) for a vertical direction v, perform a line search along v to find a step length β as described below, and set $v = \beta v$. Then solve (2.8 b) for vector h, find a step length α by performing a line search along curve $u(h)$ as described below, and set $h_k = \alpha h$, where

$$u(h) = x_{k+} + s_k(h) = x_{k+} + Z_{k+}h + Y_{k+}R_{k+}^{-T}[c(x_{k+}) - c(x_{k+} + Z_k h)] \tag{2.9}$$

and $x_{k+} = x_k + Y_k v$. The line search procedures described below ensure that $p_\mu(x)$ decreases and $y_k^T\ h_k > 0$, where $y_k = \nabla_h p_\mu(u(h_k)) - \nabla_h p_\mu(u(0))$.

We use the BFGS formula

$$B_{k+1} = \mathrm{BFGS}(B_k,\ h_k,\ y_k) = B_k - \frac{B_k h_k h_k^T B_k}{h_k^T B_k h_k} + \frac{y_k y_k^T}{y_k^T h_k}$$

to update the Hessian approximation B_k. Since our line search technique guarantees that $y_k^T\ h_k > 0$, it is well known that the positive definiteness of B_k is inherited by B_{k+1}.

Our algorithm, Algorithm 2.1, is formally described in Figure 2.1. For a fixed penalty parameter $\mu_i > 0$, our modified quasi-Newton method is adopted to minimize $p_{\mu_i}(x)$ in an inner loop until the following criteria are satisfied.

$$\begin{cases} a) & \|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| \le \mu_i^{1/2}, \\ b) & \|c(x_i^{(l)})\| \le \Lambda_i^{(l)} \mu_i, \end{cases} \tag{2.10}$$

where $\Lambda_i^{(l)} = \max\{\|\lambda_i^{(l)}\|/\sigma,\ 1\}$. Upon satisfaction of (2.10) we reduce the penalty parameter to μ_{i+1} for the next inner loop. By the definition of μ_{i+1}, it follows that

$$\mu_i^{6/5} \le \mu_{i+1} \le \rho\mu_i, \tag{2.11}$$

where $\rho < 1$.

As mentioned above, at each iterate $x_i^{(l)}$ our modified quasi-Newton method determines a vertical direction v and a horizontal direction h separately. To economize, we employ a vertical step v only when $x_i^{(l)}$ is outside of an envelope of surface $c(x) = 0$. That is, when (2.10 b) does not hold, Newton's method (2.8 a) with a safeguard is employed to reduce the value of $\|c(x)\|$. When $x_i^{(l)}$ inhabits the envelope, i.e., when (2.10 b) holds, Algorithm 2.1 sets $v = 0$ and only (2.8 b) is solved. Since $\|c(x)\|$ varies only slightly along the curve $x = u(h)$, $x_i^{(l+1)}$ attempts to stay in the envelope.

Choose values $\mu_1 > 0$, $0 < \sigma < 1 - \frac{1}{\sqrt{2}}$, $0 < \omega < 1$;
Choose a point $x_0^* \in R^n$ and an $n \times n$ positive definite matrix $B_1^{(0)}$. Set $i := 1$.

while (μ_i is not small enough)
  Set $l := 0$, $x_i^{(0)} := x_{i-1}^*$;
  while either of the criteria in (2.10) does not hold
    $x_i^{(l+)} := x_i^{(l)}$;
    if (2.10 b) does not hold
      Solve $(R_i^{(l)})^T v_i^{(l)} = -c(x_i^{(l)})$;
      Find a $\beta_i^{(l)} > 0$ satisfying (2.14) by performing a line search along $v_i^{(l)}$;
      $x_i^{(l+)} := x_i^{(l)} + \beta_i^{(l)} Y_i^{(l)} v_i^{(l)}$;
    end;
    Solve $B_i^{(l)} h_i^{(l)} = -(Z_i^{(l+)})^T \nabla f(x_i^{(l+)})$;
    Find an $\alpha_i^{(l)} > 0$ satisfying (2.12) and (2.13) by doing a line search along $h_i^{(l)}$;
    $x_i^{(l+1)} := u(\alpha_i^{(l)} h_i^{(l)})$;
    $y_i^{(l)} := \nabla_h p_{\mu_i}(x_i^{(l+1)}) - \nabla_h p_{\mu_i}(x_i^{(l+)})$,
    $B_i^{(l+1)} := \mathrm{BFGS}(B_i^{(l)},\ \alpha_i^{(l)} h_i^{(l)},\ y_i^{(l)})$;
    $l := l + 1$;
  end;
  $x_i^* := x_i^{(l)}$; $\mu_{i+1} := \max\{\mu_i^{6/5},\ \rho \, \|(Z_i^*)^T \nabla f_i^*\|^2\}$;
  $i := i + 1$;
end;
Set $x^* := x_i^*$ and STOP;

Figure 2.1 : *Algorithm* 2.1

In Algorithm 2.1, the following line search approaches are used for choosing the step lengths $\alpha_i^{(l)}$ and $\beta_i^{(l)}$.

1. To search along $h_i^{(l)}$, find $\alpha > 0$, such that

$$p_{\mu_i}(u(\alpha h_i^{(l)})) - p_{\mu_i}(u(0)) \leq \sigma \alpha \nabla_h p_{\mu_i}(u(0))^T h_i^{(l)} \tag{2.12}$$

$$\nabla_h p_{\mu_i}(u(\alpha h_i^{(l)}))^T h_i^{(l)} \geq \omega \nabla_h p_{\mu_i}(u(0))^T h_i^{(l)} \tag{2.13}$$

   where $u(h) = x_i^{(l+)} + Y_i^{(l+)}(R_i^{(l+)})^{-T}[c(x_i^{(l+)}) - c(x_i^{(l+)} + Z_i^{(l+)}h)]$. $\alpha = 1$ is always taken whenever it satisfies (2.12) and (2.13).

2. Let $0 < \tau < \tau' < 1$ be given. We perform the line search along the direction $Y_i^{(l)} v_i^{(l)}$ as follows.

(a) Set $\beta := 1$;

(b) Until the line search condition

$$p_{\mu_i}(x_i^{(l)} + \beta Y_i^{(l)} v_i^{(l)}) - p_{\mu_i}(x_i^{(l)}) \le \sigma\beta \nabla p_{\mu_i}(x_i^{(l)})^T Y_i^{(l)} v_i^{(l)} \qquad (2.14)$$

is satisfied, choose a new $\beta \in [\tau\beta, \tau'\beta]$.

Similar to the proof of Theorems 6.3.2 and 6.3.3 in [9] by Dennis and Schnabel, we obtain the following lemma.

LEMMA 2.1. *Suppose the functions $f : \Re^n \to \Re$ and $c : \Re^n \to \Re^m$ are continuously differentiable on $\Re^n$. Suppose the sequences $\{x_i^{(l)}\}$, $\{h_i^{(l)}\}$ and $\{v_i^{(l)}\}$ are generated by Algorithm 2.1. There always exists an interval (α', α'') such that (2.12) and (2.13) are satisfied for all $\alpha \in (\alpha', \alpha'')$. There also always exists an interval (β', β'') such that (2.14) is satisfied for all $\beta \in (\beta', \beta'')$. Furthermore, if the conditions (2.12) and (2.13) hold, then for all i and l,*

$$(y_i^{(l)})^T h_i^{(l)} > 0. \qquad (2.15)$$

Proof. See Theorems 6.3.2 and 6.3.3 in [9]. ∎

Condition (2.15) ensures that the positive definiteness of matrices $\{B_i^{(l)}\}$ are preserved. It follows that Algorithm 2.1 can be applied to problem (1.1) as long as $f(x)$ and $c(x)$ are smooth functions.

The following lemma tells us that the sequence $\{p_{\mu_i}(x_i^{(l)})\}$ always decreases for a fixed μ_i.

LEMMA 2.2. *The sequence $\{x_i^{(l)}\}$ generated by Algorithm 2.1 always satisfies*

$$p_{\mu_i}(x_i^{(l+1)}) - p_{\mu_i}(x_i^{(l)}) \le 0. \qquad (2.16)$$

Furthermore, if

$$\|c(x_i^{(l)})\| > \Lambda_i^{(l)}\ \mu_i \qquad (2.17)$$

where $\Lambda_i^{(l)} = \max\{\frac{\|\lambda(x_i^{(l)})\|}{\sigma},\ 1\}$, *then*

$$\nabla p_{\mu_i}(x_i^{(l)})^T Y_i^{(l)} v_i^{(l)} = \lambda(x_i^{(l)})^T c(x_i^{(l)}) - \frac{1}{\mu_i}\|c(x_i^{(l)})\|^2. \qquad (2.18)$$

Proof. Since $B_i^{(l)} h_i^{(l)} = -\nabla_h p_{\mu_i}(u(0))$, (2.12) and the positive definiteness of $B_i^{(l)}$ imply that

$$p_{\mu_i}(x_i^{(l+1)}) - p_{\mu_i}(x_i^{(l+)}) \le -\sigma\alpha\ \nabla_h p_{\mu_i}(u(0))^T (B_i^{(l)})^{-1} \nabla_h p_{\mu_i}(u(0)) \le 0. \qquad (2.19)$$

If (2.17) does not hold, then $x_i^{(l+)} = x_i^{(l)}$ and (2.16) follows from (2.19).

If (2.17) holds, equation (2.18) follows from (2.7) and $v_i^{(l)} = (R_i^{(l)})^{-T} c_i^{(l)}$. Since $\sigma < 1$, it follows from (2.14) that

$$\begin{aligned} p_{\mu_i}(x_i^{(l+)}) - p_{\mu_i}(x_i^{(l)}) &\leq \sigma\beta\,[\lambda(x_i^{(l)})^T c(x_i^{(l)}) - \frac{1}{\mu_i}\|c(x_i^{(l)})\|^2] \\ &\leq \sigma\beta\,[\|\lambda(x_i^{(l)})\|\,\|c(x_i^{(l)})\| - \frac{1}{\mu_i}\|c(x_i^{(l)})\|^2] \\ &\leq \sigma\beta\,\frac{\sigma - 1}{\mu_i}\|c(x_i^{(l)})\|^2 \leq 0. \end{aligned} \tag{2.20}$$

Therefore (2.16) is true. ∎

3. Global Convergence.

In this section we analyze the global convergence of Algorithm 2.1. We call $x^* \in \Re^n$ a stationary point of problem (1.1) if it satisfies

$$Z(x^*)^T \nabla f(x^*) = 0 \qquad \text{and} \qquad c(x^*) = 0. \tag{3.1}$$

We prove in Lemma 3.1 that the step lengths $\beta_i^{(l)}$ are bounded away from zero. In Algorithm 2.1, for any fixed $\mu_i > 0$ a specially designed iterative method is adopted to minimize $p_{\mu_i}(x)$ until the criteria in (2.10) are satisfied. We show in Lemma 3.2 that these criteria will be satisfied after a finite number of inner iterations. Therefore, there exists a subsequence of $\{x_i^{(l)}\}$ which converges to x_*. Under a further assumption, we illustrate in Lemma 3.5 that all limit points of sequence $\{x_i^{(l)}\}$ are stationary points of problem (1.1). Finally, in Theorem 3.1 we establish that the whole sequence $\{x_i^{(l)}\}$ converges to x_* provided the number of the limit points of $\{x_i^{(l)}\}$ is finite.

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

1. *The functions $f: R^n \to R$, and $c: R^n \to R^m$ and their first and second derivatives are uniformly bounded in norm over D.*
2. *The matrix $A(x)$ has full column rank for all $x \in D$, and there is a constant γ_0 such that*

$$\|A(x)[A(x)^T A(x)]^{-1}\| \leq K_0 \tag{3.2}$$

for all $x \in D$.

Note that (3.2) implies that $\|Y_i^{(l)}(R_i^{(l)})^{-1}\| \leq K_0$ for all i and l.

To prove $\|c(x_i^{(l)})\| \leq \Lambda_i^{(l)}\mu_i$ for a fixed μ_i, we must ensure that the step lengths $\{\beta_i^{(l)}\}$ are uniformly bounded away from zero.

Lemma 3.1. *Suppose Assumption 3.1 is satisfied. Then there is a constant $\beta > 0$, such that*

$$\beta_i^{(l)} \geq \beta > 0.$$

holds for all i and l.

Proof. First, note that in Algorithm 2.1 $\beta_i^{(l)}$ is computed only when $\|c(x_i^{(l)})\| > \Lambda_i^{(l)}\mu_i$. Suppose that that $\beta_i^{(l)} < 1$. If step length $\tilde{\beta} \le 1$ is the most recent failure of (2.14), 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 matrices $\nabla^2 c_i(x)$ 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 $x_i^{(l)}$

$$\begin{aligned} p_{\mu_i}(x_i^{(l)} + \tilde{\beta} Y_i^{(l)} v) - p_{\mu_i}(x_i^{(l)}) &= \tilde{\beta}[\nabla p_{\mu_i}(x_i^{(l)})]^T Y_i^{(l)} v + \\ &\quad + \frac{\tilde{\beta}^2}{2}(Y_i^{(l)} v)^T[\nabla^2 p_{\mu_i}(\tilde{x}_i^{(l)})](Y_i^{(l)} v) \\ &\le \tilde{\beta}[\nabla p_{\mu_i}(x_i^{(l)})]^T Y_i^{(l)} v + \tilde{\beta}^2\frac{K}{\mu_i}\|Y_i^{(l)} v\|^2 \end{aligned} \tag{3.4}$$

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 (3.2) that

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

On the other hand, since $\|\lambda_i^{(l)}\| \le \Lambda_i^{(l)} < \frac{\|c(x_i^{(l)})\|}{\mu_i}$ whenever $\beta_i^{(l)}$ is computed, (2.18) implies that

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

Combining (3.5) and (3.6), we obtain that

$$\beta_i^{(l)} \ge \tau_1\,\tilde{\beta} > \tau_1\frac{(1-\sigma)^2}{KK_0^2}.$$

■

Under Assumption 3.1 and an additional condition in (3.7), we demonstrate in Lemma 3.2 that for any fixed $\mu_i > 0$ the criteria in Algorithm 2.1 can be satisfied after a finite number of iterations.

LEMMA 3.2. *Suppose Assumption 3.1 is satisfied and sequence $\{x_i^{(l)}\}$ is generated by Algorithm 2.1. Furthermore, assume that there exists a constant $M > 0$ such that*

$$\text{there are an infinite number of points in } \{x_i^{(l)}\} \text{ with } \|B_i^{(l)}\|\,\|(B_i^{(l)})^{-1}\| \le M. \tag{3.7}$$

Then for fixed $\mu_i > 0$, there exists an integer $l_i > 0$, such that for $l = l_i$

$$\|c(x_i^{(l)})\| \le \Lambda_i^{(l)} \mu_i \quad \text{and} \quad \|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\|^2 \le \mu_i. \tag{3.8}$$

Proof. First we prove by contradiction that there exists an integer $\bar{l}_i > 0$ such that

$$\|c(x_i^{(l)})\| \le \Lambda_i^{(l)} \mu_i \tag{3.9}$$

holds for all $l \ge \bar{l}_i$. If (3.9) does not hold, there exists a subsequence $\{l_s\}$ of $\{l\}$, such that

$$\|c(x_i^{(l_s)})\| > \Lambda_i^{(l)} \mu_i \ge \mu_i.$$

Thus, it follows from (2.16), (2.19) and (2.20) that

$$\begin{aligned} p_{\mu_i}(x_i^{(l_{s+1})}) - p_{\mu_i}(x_i^{(l_s)}) &\le p_{\mu_i}(x_i^{(l_s+)}) - p_{\mu_i}(x_i^{(l_s)}) \\ &\le -\sigma(1-\sigma)\beta_i^{(l_s)} \frac{\|c(x_i^{(l_s)})\|^2}{\mu_i} \\ &< -\sigma(1-\sigma)\beta_i^{(l_s)} \mu_i. \end{aligned}$$

Thus Lemma 3.1 implies that

$$p_{\mu_i}(x_i^{(l_{s+1})}) - p_{\mu_i}(x_i^{(l_s)}) \le -\sigma(1-\sigma)\beta\mu_i$$

which contradicts the fact that $p_{\mu_i}(x)$ is bounded below for any fixed $\mu_i > 0$. This contradiction indicates that (3.9) holds for all $l \ge \bar{l}_i$.

Because (3.9) holds for all $l \ge \bar{l}_i$, we have that $x_i^{(l+)} = x_i^{(l)}$ when $l \ge \bar{l}_i$. To prove (3.8) it suffices to show that for any fixed $\mu_i > 0$, there exists an integer l_i such that $\|Z_i(x_i^{(l_i)})^T \nabla f(x_i^{(l_i)})\| \le \mu_i^{1/2}$. We prove this inequality by contradiction. Assume that integer i_0 is the smallest number such that

$$\|(Z_{i_0}^{(l)})^T \nabla f(x_{i_0}^{(l)})\| > \mu_{i_0}^{1/2} \tag{3.10}$$

holds for all $l \ge 0$. Then, according to Algorithm 2.1, $\{x_i^{(l)}\}$ stays in the inner loop defined by $\mu = \mu_{i_0}$. Thus, except for a leading finite sequence points, sequence $\{x_i^{(l)}\}$ is the same as sequence $\{x_{i_0}^{(l)}\}$. From (3.7) we have that

$$\forall\, \bar{l} \ge 0, \ \exists\, l \ge \bar{l} \quad \text{such that} \quad \|B_{i_0}^{(l)}\| \, \|(B_{i_0}^{(l)})^{-1}\| \le M. \tag{3.11}$$

Inequality (2.12) and the positive definiteness of $B_i^{(l)}$ yield that

$$p_{\mu_i}(x_i^{(l+1)}) - p_{\mu_i}(x_i^{(l)}) \le \sigma \, [Z(x_i^{(l)})^T \nabla p_{\mu_i}(x_i^{(l)})]^T \, (\alpha_i^{(l)} h_i^{(l)}) \le -\sigma \, \alpha_i^{(l)} \, (h_i^{(l)})^T B_i^{(l)} h_i^{(l)} < 0.$$

Since $p_{\mu_i}(x)$ is bounded below, it follows that $\sum_{l=0}^{\infty} |[(Z_i^{(l)})^T \nabla f(x_i^{(l)})]^T\ (\alpha_i^{(l)} h)| < +\infty$. Therefore, for $i = i_0$, $\lim_{l\to\infty} |[(Z_i^{(l)})^T \nabla f(x_i^{(l)})]^T\ (\alpha_i^{(l)} h)| = 0$. Combining with (3.10), we get $\lim_{l\to\infty} \|\alpha_i^{(l)} h_i^{(l)}\| = 0$.

Inequality (2.13) implies that for $i = i_0$

$$\begin{aligned} -[Z(x_i^{(l)})^T \nabla f(x_i^{(l)})]^T\ (\alpha_i^{(l)} h_i^{(l)}) &\leq \frac{[\nabla_h p_{\mu_i}(u(\alpha_i^{(l)} h_i^{(l)})) - \nabla_h p_{\mu_i}(u(0))]^T [\alpha_i^{(l)} h_i^{(l)}]}{1-\omega} \\ &\leq \frac{\|\nabla_h p_{\mu_i}(u(\alpha_i^{(l)} h_i^{(l)})) - \nabla_h p_{\mu_{i_0}}(u(0))\| \|\alpha_i^{(l)} h\|}{1-\omega}. \quad (3.12) \end{aligned}$$

It follows from (3.10), (3.12) and the uniform continuity assumption that, as $l \to \infty$,

$$\frac{|\ [Z(x_{i_0}^{(l)})^T \nabla f(x_{i_0}^{(l)})]^T\ (\alpha_{i_0}^{(l)} h_{i_0}^{(l)})\ |}{\|Z(x_{i_0}^{(l)})^T \nabla f(x_{i_0}^{(l)})\|\ \|\alpha_{i_0}^{(l)} h_{i_0}^{(l)}\|} \leq \frac{\|\nabla_h p_{\mu_{i_0}}(u(\alpha_{i_0}^{(l)} h_{i_0}^{(l)})) - \nabla_h p_{\mu_{i_0}}(u(0))\|}{\mu_{i_0}^{1/2} \cdot (1-\omega)} \to 0. \quad (3.13)$$

On the other hand, since $Z(x_i^{(l)})^T \nabla f(x_i^{(l)}) = -B_i^{(l)} h_i^{(l)}$, it follows from (3.11) that for any $\bar{l} \geq 0$ there exists an integer $l \geq \bar{l}$ such that

$$-\frac{[Z(x_{i_0}^{(l)})^T \nabla f(x_{i_0}^{(l)})]^T\ (\alpha_{i_0}^{(l)} h_{i_0}^{(l)})}{\|(Z_{i_0}^{(l)})^T \nabla f(x_{i_0}^{(l)})\|\ \|\alpha_{i_0}^{(l)} h_{i_0}^{(l)}\|} \geq \frac{(h_{i_0}^{(l)})^T B_{i_0}^{(l)} h_{i_0}^{(l)}}{\|B_{i_0}^{(l)} h_{i_0}^{(l)}\| \|h_{i_0}^{(l)}\|} \geq \frac{1}{\|B_{i_0}^{(l)}\| \|[B_{i_0}^{(l)}]^{-1}\|} \geq \frac{1}{M} > 0.$$

This inequality contradicts (3.13). This contradiction proves (3.8). ∎

If the reduced Hessian matrix $Z(x_i^{(l)})^T \nabla^2 L(x, \lambda_i^{(l)}) Z(x_i^{(l)})$ is positive definite for all x in the line search segments, it is easy to establish that the condition in (3.7) holds in a manner similar to [5]. Therefore, Lemma 3.2 holds. That is, for any fixed μ_i there exists an integer $l_i > 0$ such that for $l = l_i$ the criteria in (2.10) are satisfied. Thus, we obtain the following results.

LEMMA 3.3. *Suppose that all the assumptions in Lemma 3.2 hold. For every integer i, let l_i denote the first integer $l \geq 0$ for which (2.10) is 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,*

$$\lim_{k\to\infty} [\|Z(x_k)^T \nabla f(x_k)\| + \|c(x_k)\|] = 0. \quad (3.14)$$

Proof. Since the criteria in (2.10) 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 (2.10) holds for $l = l_i$ and μ_i tends to zero, we get $\lim_{i\to\infty}[\|Z(x_i^{(l_i)})^T\nabla f(x_i^{(l_i)})\| + \|c(x_i^{(l_i)})\|] = 0$. Thus, Lemma 3.3 is proved. ∎

Lemma 3.3 says that, if Assumption 3.1 holds and the reduced Hessian matrix of the Lagrangian function is positive definite for all x in the line search segments, then there is a subsequence of $\{x_i^{(l)}\}$ which converges to a stationary point of problem (1.1). To prove that every limit point of sequence $\{x_i^{(l)}\}$ is a stationary point of problem (1.1), we need a stronger assumption than the condition in (3.7).

ASSUMPTION 3.2. *In Algorithm 2.1, for all $i \geq 1$ and $l \geq 0$ there is a constant $M > 0$ such that*

$$\|B_i^{(l)}\| \leq M \qquad \textit{and} \qquad \|[B_i^{(l)}]^{-1}\| \leq M. \tag{3.15}$$

Before we prove in Lemma 3.5 that all limit points of $\{x_k\}$ are stationary points, the following lemma is needed.

LEMMA 3.4. *Suppose that all the assumptions in Lemma 3.2 hold. Then*

$$\sum_{i=1}^{\infty}\sum_{l=0}^{l_i-1}[p_{\mu_i}(x_i^{(l)}) - p_{\mu_i}(x_i^{(l+1)})] < +\infty. \tag{3.16}$$

Proof. It follows from the Assumption 3.1 that there exists a constant $N_1 > 0$ such that for all integers $i > 0$ and $0 \leq l \leq l_i$, $|\Lambda_i^{(l)}| \leq N_1$. Thus

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

Notice that since $x_i^{(l_i)} = x_{i+1}^{(0)}$, it follows that

$$\begin{aligned}
\sum_{i=1}^{\infty}\sum_{l=0}^{l_i-1}[p_{\mu_i}(x_i^{(l)}) - p_{\mu_i}(x_i^{(l+1)})] &= \sum_{i=1}^{\infty}[p_{\mu_i}(x_i^{(0)}) - p_{\mu_i}(x_{i+1}^{(0)})] \\
&= \sum_{i=1}^{\infty}[p_{\mu_i}(x_i^{(0)}) - p_{\mu_{i-1}}(x_i^{(0)})] \\
&\qquad + \sum_{i=1}^{\infty}[p_{\mu_{i-1}}(x_i^{(0)}) - p_{\mu_i}(x_{i+1}^{(0)})] \\
&\leq \sum_{i=2}^{\infty}[p_{\mu_i}(x_i^{(0)}) - p_{\mu_{i-1}}(x_i^{(0)})] + N_2,
\end{aligned}$$

where $N_2 = p_{\mu_1}(x_1^{(0)}) - \inf\{p_{\mu_i}(x)\}$ is a constant since $p_{\mu_i}(x)$ is bounded below.

It follows from (3.15) and (2.11) that

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

Therefore, (2.11) implies that

$$\sum_{i=1}^{\infty}\sum_{l=0}^{l_i-1}[p_{\mu_i}(x_i^{(l)})-p_{\mu_i}(x_i^{(l+1)})]\le N_1^2\sum_{i=2}^{\infty}\mu_{i-1}^{4/5}+N_2\le\frac{N_1^2\mu_1^{4/5}}{1-\rho^{4/5}}+N_2.$$

■

Because of (3.16), we can prove in Lemma 3.5 that every limit point of the sequence $\{x_i^{(l)}\}$ is a stationary point of problem (1.1).

LEMMA 3.5. *Suppose that the conditions in Assumptions 3.1 and 3.2 are satisfied. Suppose that the sequence $\{x_k\}$ is the one described in Lemma 3.3. Then*

$$\lim_{k\to\infty}\|c(x_k)\|=0. \tag{3.17}$$

and

$$\lim_{k\to\infty}\|Z(x_k)^T\,\nabla f(x_k)\|=0. \tag{3.18}$$

Proof. To prove (3.17), we define

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

Therefore, $d_k\ge 0$. It is obvious from Lemma 3.1 and Lemma 2.2 that

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

which, with Lemma 3.3, implies that

$$\sum_{k=k_0}^{\infty}d_k\le\sum_{i=1}^{\infty}\sum_{l=0}^{l_i-1}d_i^{(l)}\le\frac{1}{\sigma\,\beta}\sum_{i=1}^{\infty}\sum_{l=0}^{l_i-1}[p_{\mu_i}(x_i^{(l)})-p_{\mu_i}(x_i^{(l+1)})]<+\infty.$$

Thus, $\lim_{k\to\infty}d_k=0$. Notice that since $\mu_i\to 0$ and $\|\lambda_i\|$ is bounded, it follows from the definition of $\{d_k\}$ that (3.17) holds.

To prove (3.18), note that from Lemma 3.3 it follows that for all $0\le l\le l_i$

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

Similar to the proof of Lemma 3.2, it follows that for all $0\le l\le l_i$

$$-[Z(x_i^{(l)})^T\nabla f(x_i^{(l)})]^T\,(\alpha_i^{(l)}h)\to 0 \qquad \text{as} \qquad i\to\infty.$$

Assuming (3.18) does not hold, then there exist an $\epsilon>0$ and a subsequence $\mathcal{S}$ such that

$$\|Z(x_k)^T\nabla f(x_k)\|\ge\epsilon \qquad \text{for } k\in\mathcal{S},$$

then

$$\|\alpha_k h_k\| \to 0 \qquad \text{for } k \in \mathcal{S}.$$

And since $y_k = \alpha_k \, B_{k+1} h_k$, similar to the proof of Lemma 3.2, it follows that for $k \in \mathcal{S}$

$$\begin{aligned} -\frac{[Z(x_k)^T \nabla f(x_k)]^T \, (\alpha_k h_k)}{\|Z(x_k)^T \nabla f(x_k)\| \|\alpha_k h_k\|} &\leq \frac{\|\nabla_h p_{\mu_i}(u(\alpha_k h_k)) - \nabla_h p_{\mu_i}(u(0))\|}{\epsilon(1-\omega)} \\ &= \frac{\|y_k\|}{\epsilon \, (1-\omega)} \leq \frac{M\|\alpha_k h_k\|}{\epsilon \, (1-\omega)} \to 0. \end{aligned}$$

Similar to the proof in Lemma 3.2, we can show that it contradicts the assumption in (3.16). ∎

Finally, if there are only a finite number of limit points to problem (1.1), then the sequence $\{x_i^{(l)}\}$ converges.

THEOREM 3.1. *Suppose that the conditions in Assumptions 3.1 and 3.2 are satisfied and that the sequence $\{x_k\}$ is the one described in Lemma 3.3. Moreover, suppose that every stationary point of (1.1) is isolated. Then*

$$\lim_{k\to\infty} x_k = x^* \tag{3.19}$$

holds, where x^* is a stationary point of (1.1).

Proof. Since $\{x_k\}$ is bounded, there exists a subsequence x_{k_j} of $\{x_k\}$ such that

$$\lim_{j\to\infty} x_{k_j} = x^*,$$

where $x^* \in D$ is an accumulation point of $\{x_k\}$. But by Lemma 3.5, (3.1) holds at x^*. That is, x^* is a stationary point of (1.1). Actually, by Lemma 3.5, every accumulation point x^* of $\{x_k\}$ will satisfy (3.1). Therefore, x^* is an isolated accumulation point of $\{x_k\}$.

Now we prove (3.19) by contradiction. Suppose $\{x_k\}$ does not converge. Since x^* is an isolated accumulation point of $\{x_k\}$, there exists a subsequence $\{x_{k_j}\}$ of $\{x_k\}$ and an $\epsilon > 0$ such that $\|x_{k_j+1} - x_{k_j}\| \geq \epsilon$ (Lemma 4.10, [14]). But using Lemma 3.5 it follows from (3.16) that $\lim_{k\to\infty} \|h_k\| = 0$ and $\lim_{k\to\infty} \|v_k\| = 0$. Hence, $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0$. This contradiction proves (3.19). ∎

4. Local Superlinear Convergence. Theorem 3.1 shows that the sequence $\{x_k\}$ converges to a local minimizer x_* of (1.1) under certain assumptions. In this section, we consider the local behavior of the sequence $\{x_k\}$. We prove under certain assumptions that superlinear convergence occurs in a neighborhood of x_*. In this section, we suppose that unit step lengths $\alpha_i^{(l)} = 1$ and $\beta_i^{(l)} = 1$ are taken. At the end of this section we will show that unit step lengths are admissible when $x_i^{(l)}$ is sufficiently close to the minimizer x_*.

ASSUMPTION 4.1. *The sequence* $\{x_i^{(l)}\}$ *generated by Algorithm 2.1 is contained in a bounded convex set D with the following properties:*

1. *The functions* $f: R^n \to R$, *and* $c: R^n \to R^m$ *are three times continuously differentiable in a neighborhood of* x_*.
2. *The matrix* $Z_*^T \nabla_{xx}^2 L(x^*, \lambda^*) Z^*$ *is positive definite.*
3. *The matrix* $A(x_*)$ *has full column rank.*

Note that Assumption 4.1 implies that there exists a constant K_0 such that $\|Y_i^{(l)}(R_i^{(l)})^{-1}\| \leq K_0$ when $x_i^{(l)}$ is in a neighborhood of x_*.

The following lemma, due to Byrd and Nocedal [5], is important to the local convergence of Algorithm 2.1. 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 for all* x *sufficiently close to* x_*,

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

for some positive constants K_1 *and* K_2.

Proof. See [5]. ■

To establish the local superlinear convergence, we need to show that

$$\theta_k \equiv \frac{\|(B_k - H_*)h_k\|}{\|h_k\|} \to 0 \qquad \text{as } k \to \infty \tag{4.1}$$

where $H_* = Z_*^T \nabla^2 L(x_*, \lambda_*) Z_*$. Similar to the proof technique used by many authors (e.g., Byrd and Nocedal [4] and Powell [17]), we can prove (4.1) by showing the two inequalities:

$$\|y_k - H_* h_k\| \leq K \max\{\|x_k - x_*\|, \|x_{k+1} - x_*\|\}\|h_k\| \tag{4.2}$$

and

$$\sum_{k=1}^{\infty} \|x_k - x_*\| < +\infty, \tag{4.3}$$

where $K > 0$ is a constant and $y_k = y_i^{(l)} = \nabla_h p_{\mu_i}(u(\alpha h)) - \nabla_h p_{\mu_i}(u(0))$.

To show (4.2) and (4.3), we define

$$\phi_{\mu_i}(x) = f(x) + \lambda(x)^T c(x) + \frac{1}{2\mu_i}\|c(x)\|^2 = p_{\mu_i}(x) + \lambda(x)^T c(x).$$

Function $\phi_{\mu_i}(x)$ has a nice property which is shown in Lemma 4.2. In Lemmas 4.3 – 4.5 we establish inequality (4.2). Then we prove in Lemmas 4.6 – 4.9 that inequality (4.3) holds.

Using inequalities (4.2) and (4.3), we demonstrate in Lemma 4.10 that (4.1) is true. Finally, we exhibit in Lemma 4.11 and Theorem 4.1 that the local convergence rate of Algorithm 2.1 is superlinear.

LEMMA 4.2. *Suppose Assumption 4.1 holds. Then there exist a neighborhood $U(x_*)$ of x_* and constants $K_3 > 0$ and $K_4 > 0$ such that for all $x \in U(x_*)$*

$$K_3\|x - x_*\|^2 \leq \phi_{\mu_i}(x) - \phi_{\mu_i}(x_*) \leq K_4\,[\|Z(x)^T\nabla f(x)\|^2 + \frac{1}{\mu_i}\|c(x)\|^2]$$

holds for all $x \in U(x_*)$.

Proof. To establish the left inequality, note that

$$\nabla\phi_{\mu_i}(x_*) = 0 \qquad \text{and} \qquad \nabla^2\phi_{\mu_i}(x_*) = \hat{G} + \frac{1}{\mu_i}A(x_*)A(x_*)^T,$$

where $\hat{G} = \nabla^2 L(x_*) + A(x_*)\lambda'(x_*) + \lambda'(x_*)^T A(x_*)^T$. Let $E \in R^{n\times m}$, with full column rank, such that $Z_*^T\hat{G}E = 0$. Thus, if $A(x_*)^T Ev = 0$ for some $v \in R^m$, then $Ev = Z_* w$ for some $w \in R^{n-m}$. But since $Z_*^T\hat{G}Z_* w = Z_*\hat{G}Ev = 0$, we have that $w = 0$, and hence $v = 0$. Therefore, $A(x_*)^T E$ is a nonsingular $m \times m$ matrix.

Now consider the $n \times n$ matrix

$$\begin{bmatrix} Z_*^T \\ E^T \end{bmatrix}[\hat{G} + \frac{1}{\mu_i}A_*A_*^T][Z_* \;\; E] = \begin{bmatrix} Z_*^T\hat{G}Z_* & 0 \\ 0 & E^T\hat{G}E + \frac{1}{\mu_i}E^TA_*A_*^TE \end{bmatrix}.$$

The matrix on the the right-hand side is positive definite if μ_i is small enough, say $\mu_i = \nu > 0$. Thus $\nabla^2\phi_\nu(x_*) = \hat{G} + \frac{1}{\nu}A_*A_*^T$ is positive definite for such ν.

Since $\nabla^2\phi_\nu(x)$ is continuous, there is a constant $K_3 > 0$ such that for all x in some neighborhood $U(x_*)$ of x_*, all eigenvalues of $\nabla^2\phi_\nu(x)$ are greater than $2K_3$. Since $\nabla\phi_\nu(x_*) = 0$, for $0 < \mu_i \leq \nu$

$$\phi_{\mu_i}(x) - \phi_{\mu_i}(x_*) \geq \phi_\nu(x) - \phi_\nu(x_*) \geq K_3\|x - x_*\|^2.$$

Now we consider the right hand inequality. If we define

$$\Phi(x) = f(x) + \lambda(x)^T c(x),$$

then $\phi_{\mu_i}(x) = \Phi(x) + \frac{1}{2\mu_i}\|c(x)\|^2$. Now since

$$\nabla\Phi(x_*) = 0 \qquad \text{and} \qquad \nabla^2\Phi(x_*) = \nabla^2 L(x_*),$$

it follows from Taylor's theorem and Lemma 4.1 that for x in some neighborhood of x_*,

$$\begin{aligned} \Phi(x) - \Phi(x_*) &= \frac{1}{2}(x - x_*)^T\nabla^2 L(x_*)(x - x_*) + o(\|x - x_*\|^2) \\ &\leq \frac{\|\nabla^2 L(x_*)\| + 1}{2}\|x - x_*\|^2 \\ &\leq \frac{\|\nabla^2 L(x_*)\| + 1}{K_1^2}\|Z(x)^T\nabla f(x)\|^2 + \frac{\|\nabla^2 L(x_*)\| + 1}{K_1^2}\|c(x)\|^2. \end{aligned}$$

Therefore

$$\begin{aligned}\phi(x)-\phi(x_*) &\leq \Phi(x)+\frac{1}{2\mu_i}\|c(x)\|^2-\Phi(x_*)\\ &\leq K_4[\|Z(x)^T\nabla f(x)\|^2+\frac{1}{2\mu_i}\|c(x)\|^2].\end{aligned}$$

■

Inequality (4.2) is not trivial to be proven because y_k consists of a factor $\frac{1}{\mu_i}\to\infty$. In Lemmas 4.3 and 4.4 we exhibits that $\|h_i^{(l)}\|$ and $\|c(x_i^{(l+)})\|$ are small. Then, based on Lemmas 4.3 and 4.4, we prove (4.2) in Lemma 4.5.

LEMMA 4.3. *Suppose Assumption 4.1 holds. Suppose the sequence generated by Algorithm 2.1 converges to x_*. Then there exist an integer $i_5>0$ and a constant $K_5>0$ such that for all $0\leq l\leq l_i$,*

$$\|h_i^{(l)}\|\leq K_5\ [\ \mu_{i-1}^{2/5}+\|c(x_i^{(l)})\|^{1/2}\ ] \tag{4.4}$$

whenever $i\geq i_5$.

Proof. It follows from (2.2) that

$$u(h)-u(0)=Zh+YR^{-1}[c(x_c)-c(x_c+Zh)],$$

which implies that $\|u(h)-u(0)\|\geq\|h\|$ since Y and Z are orthogonal. Thus Lemma 4.2 implies that

$$\begin{aligned}\|h_i^{(l)}\|^2 &\leq \|x_i^{(l+1)}-x_i^{(l+)}\|^2\\ &\leq 2\ [\|x_i^{(l+1)}-x_*\|^2+\|x_i^{(l+)}-x_*\|^2]\\ &\leq 2K_3^{-1}[\phi_{\mu_i}(x_i^{(l+1)})-\phi_{\mu_i}(x_*)+\phi_{\mu_i}(x_i^{(l+)})-\phi_{\mu_i}(x_*)].\end{aligned} \tag{4.5}$$

Since $\lambda(x)$ is bounded, let $K>0$ be a constant such that $\|\lambda(x)\|\leq K$. From the definition of $\phi_{\mu_i}(x)$ and the fact that $p(x_i^{(l)})$ is decreasing, we get

$$\begin{aligned}\phi_{\mu_i}(x_i^{(l+)})-\phi_{\mu_i}(x_*) &= p_{\mu_i}(x_i^{(l+)})-p_{\mu_i}(x_*)+\lambda(x_i^{(l+)})^Tc(x_i^{(l+)})\\ &\leq p_{\mu_i}(x_i^{(0)})-p_{\mu_i}(x_*)+\lambda(x_i^{(l+)})^Tc(x_i^{(l+)})\\ &\leq \phi_{\mu_i}(x_i^{(0)})-\phi_{\mu_i}(x_*)+K\|c(x_i^{(0)})\|+K\|c(x_i^{(l+)})\|.\end{aligned} \tag{4.6}$$

Similarly, we have

$$\phi_{\mu_i}(x_i^{(l+1)})-\phi_{\mu_i}(x_*)\leq\phi_{\mu_i}(x_i^{(0)})-\phi_{\mu_i}(x_*)+K\|c(x_i^{(0)})\|+K\|c(x_i^{(l+1)})\|. \tag{4.7}$$

According to Lemma 3.3, we obtain that $x_i^{(0)}=x_{i-1}^{(l_{i-1})}$ and, hence,

$$\|Z(x_i^{(0)})^T\nabla f(x_i^{(0)})\|^2\leq\mu_{i-1}\quad\text{and}\quad\|c(x_i^{(0)})\|\leq K\ \mu_{i-1}.$$

Thus, from (2.11) and Lemma 4.2, we get

$$\phi_{\mu_i}(x_i^{(0)}) - \phi_{\mu_i}(x_*) + K\|c(x_i^{(0)})\| \leq K_4(1+2K^2)\mu_{i-1}^{4/5}. \tag{4.8}$$

Inequalities (4.5) - (4.8) imply that

$$\|h_i^{(l)}\|^2 \leq \frac{4K_4(1+2K^2)}{K_3}\mu_{i-1}^{4/5} + \frac{2K}{K_3}(\|c(x_i^{(l+)})\| + \|c(x_i^{(l+1)})\|),$$

which yields that $\lim_{i\to\infty}\|h_i^{(l)}\| = 0$. Therefore, it follows from equation (2.4) that there exists a constant $N_1 > 0$ such that for sufficiently large i

$$\|c(x_i^{(l+1)})\| \leq \|c(x_i^{(l+)})\| + N_1 \, \|h_i^{(l)}\|^3. \tag{4.9}$$

Thus, we have

$$\|h_i^{(l)}\|^2 \leq \frac{4K_4(1+2K^2)}{K_3}\mu_{i-1}^{4/5} + \frac{4K}{K_3}\|c(x_i^{(l+)})\| + N_1 \, \|h_i^{(l)}\|^3.$$

Therefore, there exists a constant $K_5 > 0$ such that for sufficiently large i

$$\|h_i^{(l)}\|^2 \leq K_5^2[\mu_{i-1}^{4/5} + \|c(x_i^{(l+)})\|].$$

■

LEMMA 4.4. *Suppose Assumption 4.1 holds. Suppose the sequence generated by Algorithm 2.1 converges to x_*. Then there exists an integer $i_6 > 0$ and constants $K_6 > 0$ and $K_7 > 0$ such that all $0 \leq l \leq l_i$*

$$\|c(x_i^{(l+)})\| \leq K_6 \, \mu_i \tag{4.10}$$

and

$$\|h_i^{(l)}\| \leq K_7 \, \mu_{i-1}^{2/5} \tag{4.11}$$

whenever $i \geq i_6$.

Proof. First, by the assumption 4.1, there exists a constant K_6 such that $|\Lambda_i^{(l)}| \leq K_6$. Since $\mu_i \to 0$ and $\|c(x_i^{(l+)})\| \to 0$ as $i \to \infty$, it follows from (4.4), after some calculations, that for sufficiently large i

$$\|h_i^{(l)}\|^3 \leq \frac{5K_5^3}{2}\mu_{i-1}^{6/5} \; + \; \frac{1}{N_1}(\frac{1}{\sqrt{\rho}} - 1)\|c(x_i^{(l+)})\|. \tag{4.12}$$

Combining (4.9) and (4.12), we have that

$$\|c(x_i^{(l+1)})\| \leq K\mu_{i-1}^{6/5} \; + \; \frac{1}{\sqrt{\rho}}\|c(x_i^{(l+)})\|,$$

where $K = \frac{5N_1K_5^3}{2}$. If (2.17) does not hold, then, according to the algorithm, $x_i^{(l+)} = x_i^{(l)}$. It follows that

$$\|c(x_i^{(l+)})\| = \|c(x_i^{(l)})\| \le K_6\ \mu_i \le \rho K_6 \mu_{i-1}. \tag{4.13}$$

Therefore, (4.10) is true when (2.17) does not hold. If (2.17) does hold, then there exists a constant $N_2 > 0$ such that

$$\begin{aligned}\|c(x_i^{(l+)})\| &= \|c(x_i^{(l)} + Y_i^{(l)}v_i^{(l)})\| \\ &= \|c(x_i^{(l)}) + (A_i^{(l)})^T Y_i^{(l)} v_i^{(l)}\| + \mathcal{O}(\|v_i^{(l)}\|^2) \\ &\le \|c(x_i^{(l)}) + (R_i^{(l)})^T v_i^{(l)}\| + N_2\|c(x_i^{(l)})\|^2 \\ &= N_2\|c(x_i^{(l)})\|^2. \end{aligned} \tag{4.14}$$

If we can prove that for all $0 \le j < l_i$,

$$\|c(x_i^{(j)})\| \le K_6\ \mu_{i-1}, \tag{4.15}$$

then, since $\mu_{i-1}^{6/5} \le \mu_i$, it follows from (4.14) that (4.10) is true as long as μ_{i-1} is small enough that $N_2K_6^2\mu_{i-1}^{4/5} \le 1$. Now we prove (4.15) by induction. It is obvious from the algorithm that (4.15) is true for $j = 0$. It follows from (4.9), (4.13) and (4.14) that

$$\|c_i^{(l+1)}\| \le \begin{cases} (K\mu_{i-1}^{1/5} + \sqrt{\rho}K_6)\mu_{i-1} & \text{if (4.14) holds} \\ K\mu_{i-1}^{6/5} + \frac{N_2}{\sqrt{\rho}}\|c(x_i^{(l)})\|^2 & \text{otherwise.} \end{cases} \tag{4.16}$$

Let i be large enough that

$$\mu_{i-1} \le \min\{\ [\frac{(1-\sqrt{\rho})K_6}{K}]^5,\ \frac{\rho}{N_2K_6}\ \}. \tag{4.17}$$

Suppose (4.15) holds for $j = l$, then it follows from (4.16) and (4.17) that (4.15) holds for $j = l + 1$. Thus (4.15) holds for any integer $0 \le j \le l_i$ whenever $i \ge i_6$. Therefore (4.10) holds.

It follows from (4.4) and (4.10) that $\|h_i^{(l)}\| \le K_5[\mu_{i-1}^{2/5} + K_6\mu_i^{1/2}] \le K_5(1+\sqrt{\rho})\mu_{i-1}^{2/5}$. i.e., (4.11) holds with $K_7 = K_5(1+\sqrt{\rho})$. ∎

By the definition of $p_{\mu_i}(u(h))$, we have that

$$\nabla_h p_{\mu_i}(u(h)) = \nabla_h f(u(h)) + \frac{\nabla_h[c(u(h))^T c(u(h))]}{2\mu_i} = \nabla_h f(u(h)) + \frac{[\nabla_h c(u(h))]c(u(h))}{\mu_i}. \tag{4.18}$$

Using Taylor's theorem, it follows from (2.3) that

$$\nabla_h c((u(h)) = \nabla_h c((u(0)) + \nabla_h^2 c((u(0))h + \mathcal{O}(\|h\|^2) = \mathcal{O}(\|h\|^2). \tag{4.19}$$

Thus, it follows from (4.18), (2.4), (4.10), (4.11), (2.11) and (4.19) that there exists a constant $K_8 > 0$ such that

$$\begin{aligned} \|\nabla_h p_{\mu_i}(u(h_i^{(l)})) - \nabla_h f(u(h_i^{(l)}))\| &\leq \|\nabla_h c(u(h_i^{(l)}))\| \, \frac{\|c(u(h_i^{(l)}))\|}{\mu_i} \\ &\leq \|\nabla_h c(u(h_i^{(l)}))\| \, \frac{\|c(u(0))\| + \mathcal{O}(\|h_i^{(l)}\|^3)}{\mu_i} \\ &= \|\nabla_h c(u(h_i^{(l)}))\| \, \frac{\|c(x_i^{(l+)})\| + \mathcal{O}(\|h_i^{(l)}\|^3)}{\mu_i} \\ &\leq K_8 \|h_i^{(l)}\|^2. \end{aligned} \tag{4.20}$$

By inequality (4.20) we have established that $\nabla_h p_{\mu_i}(u(h_i^{(l)}))$ is a good approximation to $\nabla_h f(u(h_i^{(l)}))$. This is crucial to the proof of Lemma 4.5.

LEMMA 4.5. *Suppose Assumption 4.1 holds. Suppose 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 $0 \leq l \leq l_i$*

$$\|y_i^{(l)} - H_* \, h_i^{(l)}\| \leq K_9 \, [\max\{\|x_i^{(l+1)} - x_*\|, \ \|x_i^{(l)} - x_*\|\}] \|h_i^{(l)}\| \tag{4.21}$$

whenever $i \geq i_9$, where $H_* = Z(x_*)^T \nabla^2 L(x_*) Z(x_*)$.

Proof. Define $\hat{y}_i^{(l)} = \nabla_h f(u(h_i^{(l)})) - \nabla_h f(u(0))$. Since $\nabla_h p_{\mu_i}(u(0)) = \nabla_h f(u(0))$, it follows from (4.20) that

$$\|y_i^{(l)} - \hat{y}_i^{(l)}\| \leq K_8 \|h_i^{(l)}\|^2, \tag{4.22}$$

By the mean-value theorem, we have that

$$\hat{y}_i^{(l)} = G_i^{(l)} h_i^{(l)} \qquad \text{where} \qquad G_i^{(l)} = \int_0^1 \nabla_h^2 f(u(\tau h_i^{(l)})) d\tau.$$

By the continuity of $\nabla^2 L(x)$ and $u(h)$, there exists a constant N_1 such that for any $0 \leq \tau \leq 1$,

$$\begin{aligned} \|G_i^{(l)} - \nabla_h^2 f(u(0))\| &= \| \int_0^1 [\nabla_h^2 f(u(\tau h_i^{(l)})) - \nabla_h^2 f(u(0))] d\tau \| \\ &\leq \int_0^1 \|\nabla_h^2 f(u(\tau h_i^{(l)})) - \nabla_h^2 f(u(0))\| d\tau \\ &\leq \int_0^1 N_1 \|h_i^{(l)}\| d\tau = N_1 \|h_i^{(l)}\|. \end{aligned} \tag{4.23}$$

Since $u(0) = x_i^{(l+)}$, using the continuity of $\nabla^2 L(x)$ and $Z(x)$, it follows from (2.5) that there exists a constant N_2 such that

$$\begin{aligned} \|\nabla_h^2 f(u(0)) - H_*\| &= \|Z(x_i^{(l+)})^T \nabla^2 L(x_i^{(l+)}) Z(x_i^{(l+)}) - H_*\| \\ &= \mathcal{O}(\|x_i^{(l+)} - x_*\|) \leq N_2 [\|x_i^{(l+1)} - x_*\| + \|h_i^{(l)}\|]. \end{aligned} \tag{4.24}$$

It follows from (4.23) and (4.23) that

$$\begin{aligned}
\|y_i^{(l)} - H_* h_i^{(l)}\| &\leq \|\hat{y}_i^{(l)} - H_* h_i^{(l)}\| + \|y_i^{(l)} - \hat{y}_i^{(l)}\| \\
&\leq \|G_i^{(l)} - H_*\| \|h_i^{(l)}\| + K_8 \|h_i^{(l)}\|^2 \\
&\leq [\ \|G_i^{(l)} - \nabla_h^2 f(u(0))\| + \|\nabla_h^2 f(u(0)) - H_*\| + K_7 \|h_i^{(l)}\|\ ] \|h_i^{(l)}\| \\
&\leq [\ (N_1 + N_2 + K_7)\|h_i^{(l)}\| + N_2 \|x_i^{(l+1)} - x_*\|\ ] \|h_i^{(l)}\|.
\end{aligned}$$

Thus to prove (4.21), we only need to show that

$$\|h_i^{(l)}\| = \mathcal{O}(\|x_i^{(l+1)} - x_*\| + \|x_i^{(l)} - x_*\|). \tag{4.25}$$

Note that $c(x_c) - c(x_c + Zh) = \mathcal{O}(\|h\|^2)$ since $\nabla c(x_c) Z = AZ = 0$. It follows from the definition of $u(h)$ that $u(h) - u(0) = Zh + \mathcal{O}(\|h\|^2)$. Thus

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

Therefore, there exists constants K and $\epsilon > 0$ such that

$$\|x_i^{(l+1)} - x_i^{(l)}\| \geq \|h_i^{(l)}\| + \|v_i^{(l)}\| - K[\|h_i^{(l)}\|^2 + \|v_i^{(l)}\|^2] \geq \epsilon \|h_i^{(l)}\|,$$

which implies (4.25). Thus, Lemma 4.5 is proved. ∎

To establish inequality (4.3) we need one more assumption, stated as follows.

Assumption 4.2. *There is a constant $M > 0$ such that for all $s \in \Re^{n-m}$, $i \geq 1$ and $l \geq 0$*

$$s^T B_i^{(l)} s / s^T s \leq M. \tag{4.26}$$

As we will discuss in Section 4, Assumption 4.2 is weaker than often-used assumptions for the equality constrained optimization problem (1.1).

In Lemmas 4.6 and 4.7 we show that there exists a constant $K_{11} > 0$ such that for i large enough, if $l_i > 0$, then for $0 \leq l \leq l_i - 1$

$$\phi_{\mu_i}(x_i^{(l)}) - \phi_{\mu_i}(x_i^{(l+1)}) \geq K_{11}[\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\|^2 + \frac{1}{\mu_i}\|c(x_i^{(l)})\|^2]. \tag{4.27}$$

Using (4.27) and Lemma 4.2, we establish in Lemma 4.8 that there exists a constant r ($0 < r < 1$) such that

$$\phi_{\mu_i}(x_i^{(l+1)}) - \phi_{\mu_i}(x_*) \leq r^2\ [\phi_{\mu_i}(x_i^{(l)}) - \phi_{\mu_i}(x_*)]. \tag{4.28}$$

(4.28), with the left inequality in Lemma 4.2, yields (4.3).

LEMMA 4.6. *Suppose Assumption 4.1 holds. Assume the sequence generated by Algorithm 2.1 converges to x_*. Then there exist a constant $\kappa > 0$ such that for i large enough*

$$p_{\mu_i}(x_i^{(l+)}) - p_{\mu_i}(x_i^{(l+1)}) \geq K_{10}\|h_i^{(l)}\|^2. \tag{4.29}$$

Proof. Due to (4.11) and (2.11), we know that, for sufficiently large i, $\|h_i^{(l)}\| \leq K_7\mu_i^{1/3}$. Thus, since $u(0) = x_i^{(l+)}$, (2.4) and (4.10) yield that $\|c(u(h_i^{(l)}))\|^2 - \|c(u(0))\|^2 = \mathcal{O}(\mu_i\|h_i^{(l)}\|^3)$. Therefore, when i is sufficiently large, we have

$$[p_{\mu_i}(u_i^{(l)}(h_i^{(l)})) - p_{\mu_i}(u_i^{(l)}(0))] - [f(u_i^{(l)}(h_i^{(l)})) - f(u_i^{(l)}(0))] = \mathcal{O}(\|h_i^{(l)}\|^3). \tag{4.30}$$

Define $G_i^{(l)} = \int_0^1 \nabla_h^2 f(u(\tau h_i^{(l)}))(1-\tau)d\tau$. Then, similar to the proof of Lemma 4.5, we obtain that $\|G_i^{(l)} - H_*\| \leq \|G_i^{(l)} - H_i^{(l)}\| + \|H_i^{(l)} - H_*\| \to 0$. Since the matrix H_* is positive definite, there exists a constant $\kappa > 0$ such that, when i is sufficiently large, $h^T G_i^{(l)} h \geq \kappa\|h\|^2$ holds for all $h \in \Re^{n-m}$. Thus Taylor's theorem gives that

$$f(u(h_i^{(l+)})) - f(u(0)) = \nabla_h f(u(0))^T h_i^{(l)} + \frac{1}{2}(h_i^{(l)})^T G_i^{(l)} h_i^{(l)} \geq \nabla_h p_{\mu_i}(u(0))^T h_i^{(l)} + \frac{\kappa}{2}\|h_i^{(l)}\|^2. \tag{4.31}$$

Combining (4.31) with (4.30), we get that for sufficiently large i,

$$\begin{aligned} p_{\mu_i}(x_i^{(l+1)}) - p_{\mu_i}(x_i^{(l+)}) &= \nabla_h p_{\mu_i}(u(0))^T h_i^{(l)} + \frac{\kappa}{2}\|h_i^{(l)}\|^2 + \mathcal{O}(\|h_i^{(l)}\|^3) \\ &\geq \nabla_h p_{\mu_i}(u(0))^T h_i^{(l)} + \frac{\kappa}{4}\|h_i^{(l)}\|^2 \end{aligned} \tag{4.32}$$

where $u(0) = x_i^{(l+)}$. It follows from (2.12) that

$$p_{\mu_i}(x_i^{(l+1)}) - p_{\mu_i}(x_i^{(l+)}) \leq \sigma \nabla_h p_{\mu_i}(u(0))^T h_i^{(l)}. \tag{4.33}$$

Combining (4.32) and (4.33), we obtain that $(\sigma - 1)\nabla_h p_{\mu_i}(u(0))^T h_i^{(l)} \geq \frac{\kappa}{4}\|h_i^{(l)}\|^2$. It follows that $p_{\mu_i}(x_i^{(l+)}) - p_{\mu_i}(x_i^{(l+1)}) \geq -\sigma\nabla_h p_{\mu_i}(u(0))^T h_i^{(l)} \geq \frac{\sigma\kappa}{4(1-\sigma)}\|h_i^{(l)}\|^2$. Therefore, Lemma 4.6 is proved. ∎

LEMMA 4.7. *Suppose Assumption 4.1 and 4.2 hold. Suppose the sequence generated by Algorithm 2.1 converges to x_*. Then there exist $K_{11} > 0$ and an integer i_{11}, such that, whenever $i \geq i_{11}$, (4.27) holds for $0 \leq l < l_i$.*

Proof. Inequality (4.26) yields that $\text{eig}_{\min}([B_i^{(l)}]^{-1}) \geq 1/M$. Thus it follows from (2.19) that

$$p_{\mu_i}(x_i^{(l+)}) - p_{\mu_i}(x_i^{(l+1)}) \geq \frac{\sigma}{M}\|Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})\|^2. \tag{4.34}$$

Combining (4.29) and (4.34), we obtain that

$$p_{\mu_i}(x_i^{(l+)}) - p_{\mu_i}(x_i^{(l+1)}) \geq \frac{K_{10}}{2}\|h_i^{(l)}\|^2 + \frac{\sigma}{2M}\|Z(x_i^{(l+)})^T\nabla f(x_i^{(l+)})\|^2. \tag{4.35}$$

There are two possible cases: (2.17) holds, or (2.17) does not hold. In the first case, we have

$$\|c(x_i^{(l)})\| \leq \frac{\|c(x_i^{(l)})\|^2}{\mu_i} \qquad \text{and} \qquad \|\lambda(x_i^{(l)})\| \leq \frac{\|c(x_i^{(l)})\|}{\mu_i}. \tag{4.36}$$

It follows from (4.36) that

$$\begin{aligned} \frac{\sigma}{\mu_i}\|c(x_i^{(l)})\|^2 + (1-\sigma)c(x_i^{(l)})^T\lambda(x_i^{(l)}) &\geq \frac{\sigma}{\mu_i}\|c(x_i^{(l)})\|^2 - (1-\sigma)\|c(x_i^{(l)})\|\|\lambda(x_i^{(l)})\| \\ &\geq \frac{\sigma}{\mu_i}\|c(x_i^{(l)})\|^2 - (1-\sigma)\frac{\sigma}{\mu_i}\|c(x_i^{(l)})\|^2 \\ &= \frac{\sigma^2}{\mu_i}\|c(x_i^{(l)})\|^2. \end{aligned} \tag{4.37}$$

From (4.35) and (2.20), we get

$$\begin{aligned} p_{\mu_i}(x_i^{(l)}) - p_{\mu_i}(x_i^{(l+1)}) &\geq \frac{K_{10}}{2}\|h_i^{(l)}\|^2 + \frac{\sigma}{2M}\|Z(x_i^{(l+)})^T\nabla f(x_i^{(l+)})\|^2 + \\ &\quad +\sigma[\frac{1}{\mu_i}\|c(x_i^{(l)})\|^2 - c(x_i^{(l)})^T\lambda(x_i^{(l)})]. \end{aligned}$$

Thus, from the definition of $\phi_{\mu_i}(x)$ we obtain that

$$\begin{aligned} \phi_{\mu_i}(x_i^{(l)}) - \phi_{\mu_i}(x_i^{(l+1)}) &\geq \frac{K_{10}}{2}\|h_i^{(l)}\|^2 + \frac{\sigma}{2M}\|Z(x_i^{(l+)})^T\nabla f(x_i^{(l+)})\|^2 \\ &\quad +\frac{\sigma}{\mu_i}\|c(x_i^{(l)})\|^2 + (1-\sigma)\lambda(x_i^{(l)})^Tc(x_i^{(l)}) \\ &\quad -\lambda(x_i^{(l+1)})^Tc(x_i^{(l+1)}). \end{aligned} \tag{4.38}$$

Taylor's theorem yields that

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

Thus, it follows from (4.36) that

$$\begin{aligned} \|Z(x_i^{(l+)})^T\nabla f(x_i^{(l+)})\|^2 &= \|Z(x_i^{(l)})^T\nabla f(x_i^{(l)})\|^2 + o(\|c(x_i^{(l)})\|) \\ &= \|Z(x_i^{(l)})^T\nabla f(x_i^{(l)})\|^2 + o(\frac{\|c(x_i^{(l)})\|^2}{\mu_i}). \end{aligned} \tag{4.39}$$

It follows from (2.3), (4.14) and (4.36) that

$$c(x_i^{(l+1)}) = c(x_i^{(l+)}) + \mathcal{O}(\|h_i^{(l)}\|^3) = o(\|c(x_i^{(l)})\| + \|h_i^{(l)}\|^2) = o(\|h_i^{(l)}\|^2 + \frac{\|c(x_i^{(l)})\|^2}{\mu_i}). \tag{4.40}$$

Therefore, combining (4.38), (4.39), (4.37) and (4.40), we have that

$$\begin{aligned}\phi_{\mu_i}(x_i^{(l)}) - \phi_{\mu_i}(x_i^{(l+1)}) &\geq \frac{K_{10}}{2}\|h_i^{(l)}\|^2 + \frac{\sigma}{2M}\|Z(x_i^{(l)})^T\nabla f(x_i^{(l)})\|^2 + \frac{\sigma^2}{\mu_i}\|c(x_i^{(l)})\|^2 \\ &\quad + o(\|h_i^{(l)}\|^2 + \frac{\|c(x_i^{(l)})\|^2}{\mu_i}) \\ &\geq \frac{\sigma}{2M}\|Z(x_i^{(l)})^T\nabla f(x_i^{(l)})\|^2 + \frac{\sigma^2}{2\mu_i}\|c(x_i^{(l)})\|^2.\end{aligned}$$

Therefore, (4.27) holds with $K_{11} = \min\{\frac{1}{2M},\ \frac{1}{2}\}\ \sigma^2$.

In the second case, (2.17) does not hold. Since $0 \leq l < l_i$, it follows from Algorithm 2.1 that $x_i^{(l+)} = x_i^{(l)}$,

$$\|Z(x_i^{(l)})^T\nabla f(x_i^{(l)})\|^2 \geq \mu_i \qquad \text{and} \qquad \|c(x_i^{(l)})\| \leq \Lambda_i^{(l)}\ \mu_i. \tag{4.41}$$

It follows from (4.35) and (4.41) that for a constant $C_1 = \sigma/(4M|\Lambda_i^{(l)}|)$,

$$\begin{aligned}p_{\mu_i}(x_i^{(l)}) - p_{\mu_i}(x_i^{(l+1)}) &\geq \frac{K_{10}}{2}\|h_i^{(l)}\|^2 + \frac{\sigma}{2M}\|Z(x_i^{(l)})^T\nabla f(x_i^{(l)})\|^2 \\ &\geq \frac{K_{10}}{2}\|h_i^{(l)}\|^2 + \frac{\sigma}{4M}\|Z(x_i^{(l)})^T\nabla f(x_i^{(l)})\|^2 + \frac{\sigma}{4M}\frac{\mu_i^2}{\mu_i} \\ &\geq \frac{K_{10}}{2}\|h_i^{(l)}\|^2 + \frac{\sigma}{4M}\|Z(x_i^{(l)})^T\nabla f(x_i^{(l)})\|^2 + C_1\frac{\|c(x_i^{(l)})\|^2}{\mu_i}.\end{aligned} \tag{4.42}$$

Since $x_i^{(l+1)} - x_i^{(l)} = u(h_i^{(l)})$, (2.4) implies that for sufficiently large i

$$\begin{aligned}\lambda(x_i^{(l)})^Tc(x_i^{(l)}) - \lambda(x_i^{(l+1)})^Tc(x_i^{(l+1)}) &= [\lambda(x_i^{(l)}) - \lambda(x_i^{(l+1)})]^Tc(x_i^{(l)}) \\ &\quad + \lambda(x_i^{(l+1)})^T[c(x_i^{(l)}) - c(x_i^{(l+1)})] \\ &= \mathcal{O}(\|h_i^{(l)}\|\|c(x_i^{(l)})\| + \|h_i^{(l)}\|^3) \\ &\geq -C_2\|h_i^{(l)}\|\|c(x_i^{(l)})\| + \mathcal{O}(\|h_i^{(l)}\|^3) \\ &\geq -\frac{K_{10}}{3}\|h_i^{(l)}\|^2 - \frac{3C_2^2}{4K_{10}}\|c(x_i^{(l)})\|^2 + \mathcal{O}(\|h_i^{(l)}\|^3) \\ &\geq -\frac{K_{10}}{2}\|h_i^{(l)}\|^2 - \mu_i\frac{3C_2^2}{4K_{10}}\frac{\|c(x_i^{(l)})\|^2}{\mu_i},\end{aligned} \tag{4.43}$$

where $C_2 > 0$ is a constant. Adding (4.42) and (4.43) together, we obtain that (4.27) holds with $K_{11} = \max\{\frac{\sigma}{4M},\ \frac{C_1}{2}\}$. ∎

Using Lemmas 4.2 and 4.7, we prove (4.28) in Lemma 4.8.

LEMMA 4.8. *Suppose Assumption 4.1 and 4.2 hold. Suppose the sequence generated by Algorithm 2.1 converges to x_*. Then there exist an integer i_0 and a constant r $(0 < r < 1)$ such that, whenever $i \geq i_0$, inequality (4.28) holds for $0 \leq l < l_i$.*

Proof. By Lemma 4.2 and 4.9, when $i \geq i_0$, it follows that

$$\begin{aligned} \phi_{\mu_i}(x_i^{(l)}) - \phi_{\mu_i}(x_i^{(l+1)}) &\geq K_{11}[\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\|^2 + \frac{1}{\mu_i}\|c(x_i^{(l)})\|^2] \\ &\geq \frac{K_{11}}{K_4}[\phi_{\mu_i}(x_i^{(l)}) - \phi_{\mu_i}(x_*)] \end{aligned}$$

Adding $\phi_{\mu_i}(x_*) - \phi_{\mu_i}(x_i^{(l)})$, and then multiplying both side by -1, we obtain the inequality (4.28) with $r = (1 - \frac{K_{11}}{K_4})^{1/2}$. ∎

Since $c(x_*) = 0$, we have that $\phi_{\mu_i}(x_*) = \phi_{\mu_{i-1}}(x_*)$. Thus, it follows from Lemmas 4.2 and 4.11 that for i large enough

$$\begin{aligned} K_4^{1/2}\, \|x_i^{(l)} - x_*\| &\leq [\phi_{\mu_i}(x_i^{(l)}) - \phi_{\mu_i}(x_*)]^{1/2} \leq r^l\, [\phi_{\mu_i}(x_i^{(0)}) - \phi_{\mu_i}(x_*)]^{1/2} \\ &\leq r^l\, [\phi_{\mu_{i-1}}(x_{i-1}^{(l_{i-1})}) - \phi_{\mu_{i-1}}(x_*)]^{1/2} + r^l\, [\phi_{\mu_i}(x_i^{(0)}) - \phi_{\mu_{i-1}}(x_{i-1}^{(l_{i-1})})]^{1/2} \end{aligned}$$

If we can prove that the difference between $\phi_{\mu_i}(x_i^{(0)})$ and $\phi_{\mu_{i-1}}(x_{i-1}^{(l_{i-1})})$ is small enough, then we should be able to prove (4.2).

LEMMA 4.9. *Suppose Assumption 4.1 holds. Suppose the sequence generated by Algorithm 2.1 converges to x_*. Then there exists an integer $n_0 > 0$, such that*

$$\sum_{k=k_0}^{\infty} \|x_k - x_*\| < +\infty$$

where $k_0 = \sum_{i=0}^{i_0} l_i$ for some $i_0 > 0$.

Proof. Define

$$L_i = \begin{cases} (\frac{1}{2\mu_i} - \frac{1}{2\mu_{i-1}})^{1/2}\|c(x_i^{(0)})\| & \text{if } l_i > 0 \\ 0 & \text{if } l_i = 0 \end{cases}$$

Assumption 4.1 implies that $\Lambda_i^{(0)} \leq K_0$ for some constant $K_0 > 0$. Thus

$$\|c(x_i^{(0)})\| = \|c(x_{i-1}^{(l_{i-1})})\| \leq \Lambda_i^{(0)}\, \mu_{i-1} \leq K_0 \mu_{i-1}. \tag{4.44}$$

It follows from (2.11) and (4.44) that

$$0 \leq L_i \leq \frac{1}{\sqrt{2\mu_i}}\|c(x_i^{(0)})\| \leq \frac{\mu_{i-1}^{2/5}}{\sqrt{2}}K_0 \leq \frac{\rho^{2i/5}\mu_0}{\sqrt{2}}K_0 = \frac{\tilde{\rho}^i \mu_0}{\sqrt{2}}K_0,$$

where $\tilde{\rho} = \rho^{2/5} < 1$. Without loss of generality, we assume that $\tilde{\rho} \leq r$. Then we have that

$$\sum_{s=i_0+1}^{i} L_s r^{i-s} \leq \frac{\mu_0}{\sqrt{2}}K_0 \sum_{s=i_0+1}^{i} \tilde{\rho}^s r^{i-s} \leq \frac{\mu_0}{\sqrt{2}}K_0 r^i \sum_{s=i_0+1}^{i} (\frac{\tilde{\rho}}{r})^s \leq \frac{\mu_0}{\sqrt{2}}K_0\, i\, r^i. \tag{4.45}$$

Since $x_i^{(0)} = x_{i-1}^{(l_{i-1})}$ and $c(x_*) = 0$, the definition of $\phi_{\mu_i}(x)$ yields that

$$[\phi_{\mu_i}(x_i^{(0)}) - \phi_{\mu_i}(x_*)]^{1/2} \le [\phi_{\mu_{i-1}}(x_{i-1}^{(l_{i-1})}) - \phi_{\mu_i}(x_*)]^{1/2} + L_i. \tag{4.46}$$

From Lemma 4.11 and (4.45), we obtain that for $i \ge i_0$ and $0 \le j < l_i$,

$$\begin{aligned}
[\phi_{\mu_i}(x_i^{(j)}) - \phi_{\mu_i}(x_*)]^{1/2} &\le r^j [\phi_{\mu_{i-1}}(x_{i-1}^{(l_{i-1})}) - \phi_{\mu_{i-1}}(x_*)]^{1/2} + r^j L_i \\
&\le r^{j+l_{i-1}} [\phi_{\mu_{i-1}}(x_{i-1}^{(0)}) - \phi_{\mu_{i-1}}(x_*)]^{1/2} + r^j L_i. \\
&\le r^{j+l_{i-1}} [\phi_{\mu_{i-2}}(x_{i-2}^{(l_{i-2})}) - \phi_{\mu_{i-2}}(x_*)]^{1/2} + r^{j+1} L_{i-1} + r^j L_i \\
&\le \dots \qquad \dots \qquad \dots \\
&\le r^{j+l_{i-1}+\dots+l_{i_0}} [\phi_{\mu_{i_0}}(x_{i_0}^{(0)}) - \phi_{\mu_{i_0}}(x_*)]^{1/2} \\
&\qquad + r^{j+i-i_0-1} L_{i_0+1} + \dots + r^{j+1} L_{i-1} + r^j L_i \\
&\le r^{j+\sum_{t=i_0}^{i-1} l_t} N_2 + r^j \sum_{s=i_0+1}^{i} L_s r^{i-s}
\end{aligned} \tag{4.47}$$

where $N_2 = [\phi_{\mu_{i_0}}(x_{i_0}^{(0)}) - \phi_{\mu_{i_0}}(x_*)]^{1/2}$. Combining (4.45) and (4.47), we have that for $0 \le l < l_i$

$$[\phi_{\mu_i}(x_i^{(j)}) - \phi_{\mu_i}(x_*)]^{1/2} \le r^{j+\sum_{t=i_0}^{i-1} l_t} N_2 + \frac{\mu_0}{\sqrt{2}} K_0\, i\, r^{i+j}.$$

Thus, Lemma 4.2 implies that

$$\begin{aligned}
\sum_{j=0}^{l} \|x_i^{(j)} - x_*\| &\le K_3^{-1/2} \sum_{j=0}^{l} [\phi_{\mu_i}(x_i^{(j)}) - \phi_{\mu_i}(x_*)]^{1/2} \\
&\le K_3^{-1/2} N_2 \sum_{j=0}^{l} r^{j+\sum_{t=i_0}^{i-1} l_t} + K_3^{-1/2} \frac{\mu_0}{\sqrt{2}} K_0\, i\, r^i \sum_{j=0}^{l} r^j \\
&\le N_3 \sum_{j=0}^{l} r^{j+\sum_{t=i_0}^{i-1} l_t} + N_4\, i\, r^i
\end{aligned} \tag{4.48}$$

where $N_3 = K_3^{-1/2} N_2$ and $N_4 = \frac{K_3^{-1/2} \mu_0 K_0}{\sqrt{2}(1-r)}$.

According to Algorithm 2.1, for any integer I there exist integers j and i such that

$$\sum_{k=k_0}^{I} \|x_k - x_*\| = \sum_{j=0}^{l} \|x_i^{(j)} - x_*\| + \sum_{s=i_0}^{i-1} \sum_{j=0}^{l_s-1} \|x_s^{(j)} - x_*\| \le \sum_{s=i_0}^{i} \sum_{j=0}^{l_s-1} \|x_s^{(j)} - x_*\| \tag{4.49}$$

Thus, it follows from (4.48) and (4.49) that for any integer I,

$$\begin{aligned}
\sum_{k=k_0}^{I} \|x_k - x_*\| &\le N_3 \sum_{s=i_0}^{i} \sum_{j=0}^{l_s-1} r^{j+\sum_{t=i_0}^{s-1} l_t} + N_4 \sum_{s=i_0}^{i} (s\, r^s) \\
&\le N_3 \sum_{k=k_0}^{\infty} r^k + N_4 \sum_{s=i_0}^{i} (s\, r^s) \\
&\le \frac{N_3}{1-r} + \frac{r N_4}{(1-r)^2} < +\infty.
\end{aligned}$$

Therefore (4.3) holds. ■

Now we are ready to prove (4.1), essential to the establishment of local and superlinear convergence.

LEMMA 4.10. *Suppose Assumption 4.1 holds. Suppose the sequence generated by Algorithm 2.1 converges to x_*. Then*

$$\lim_{\substack{h_k \neq 0 \\ k \to \infty}} \frac{\|(B_k - H_*)h_k\|}{\|h_k\|} = 0.$$

Proof. The discussion in section 2 yields that for $h_k \neq 0$, $y_k^T h_k > 0$. It follows from (4.22) that

$$\frac{\|y_k - H_* h_k\|}{\|h_k\|} \leq \epsilon_k,$$

where $\epsilon_k = K_9\ [\max\{\|x_{k+1} - x_*\|,\ \|x_k - x_*\|\}]$. Theorem 4.1 implies that $\sum_{k=k_0}^{\infty} \epsilon_k < \infty$. Therefore, we can apply Theorem 3.2 in [4] by Byrd and Nocedal here so that for $h_k \neq 0$,

$$\theta_k \equiv \frac{\|(B_k - H_*)h_k\|}{\|h_k\|} \to 0. \tag{4.50}$$

■

Based on (4.50), we prove local and superlinear convergence of Algorithm 2.1 in Lemma 4.11 and Theorem 4.1.

LEMMA 4.11. *Suppose Assumption 4.1 and 4.2 hold. Suppose the sequence generated by Algorithm 2.1 converges to x_*. Then when $i > 0$ is sufficiently large,*

$$\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\| \leq \eta_i^{(l)}[\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| + \|c(x_i^{(l)})\|] + 2\|c(x_i^{(l+)})\| \tag{4.51}$$

where $\eta_i^{(l)} \to 0$.

Proof. If $h_i^{(l)} = 0$, then $x_i^{(l+1)} = x_i^{(l+)}$. It follows that

$$\|c(x_i^{(l+1)})\| = \|c(x_i^{(l+)})\| \leq 2\|c(x_i^{(l+)})\|$$

and

$$Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)}) = Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)}) = -B_i^{(l)} h_i^{(l)} = 0.$$

Therefore, (4.51) holds when $h_i^{(l)} = 0$.

If $h_i^{(l)} \neq 0$, then from (2.2), we know that

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

By Taylor's theorem, we get

$$\begin{aligned}
Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)}) &= Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)}) \\
&\qquad + [Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})]' s(h_i^{(l)}) + \mathcal{O}(\|h_i^{(l)}\|^2) \\
&= Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)}) \\
&\qquad + [Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})]' Z(x_i^{(l+)}) h_i^{(l)} + \mathcal{O}(\|h_i^{(l)}\|^2) \\
&= -B_i^{(l)} h_i^{(l)} + [Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})]' Z(x_i^{(l+)}) h_i^{(l)} + \mathcal{O}(\|h_i^{(l)}\|^2).
\end{aligned}$$

And by Coleman [6], we have that

$$[Z(x_i^{(l+)})^T \nabla f(x_i^{(l+)})]' = Z(x_i^{(l+)})^T \nabla^2 L(x_i^{(l+)}).$$

Thus, it follows from Theorem 4.2 that

$$\begin{aligned}
\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| &\le \| - B_i^{(l)} h_i^{(l)} + H_i^{(l+)} h_i^{(l)}\| + \mathcal{O}(\|h_i^{(l)}\|^2) \\
&\le \| \, [H_* - B_i^{(l)}] \, h_i^{(l)}\| + \| \, [H_i^{(l+)} - H_*] \, h_i^{(l)}\| + \mathcal{O}(\|h_i^{(l)}\|^2) \\
&= [\, \theta_i^{(l)} + \|H_i^{(l+)} - H_*\| \,] \, \|h_i^{(l)}\| + \mathcal{O}(\|h_i^{(l)}\|^2).
\end{aligned} \tag{4.52}$$

Equality (2.4) yields that

$$c(x_i^{(l+1)}) = c(x_i^{(l+)}) + \mathcal{O}(\|h_i^{(l)}\|^3). \tag{4.53}$$

Combining (4.52) and (4.53), we have that

$$\begin{aligned}
\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\| &= [\, \theta_i^{(l)} + \|H_i^{(l+)} - H_*\| \,] \, \|h_i^{(l)}\| + \\
&\qquad + \|c(x_i^{(l+)})\| + \mathcal{O}(\|h_i^{(l)}\|^2) \\
&\le \xi_i^{(l)} \|h_i^{(l)}\| + \|c(x_i^{(l+)})\|,
\end{aligned} \tag{4.54}$$

where $\xi_i^{(l)} = \theta_i^{(l)} + \|H_i^{(l+)} - H_*\| + C_1 \|h_i^{(l)}\| \to 0$ and C_1 is a constant. Using (4.25) and Lemma 4.1, from (4.54) we get

$$\begin{aligned}
\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\| &\le \xi_i^{(l)} K [\, \|x_i^{(l+1)} - x_*\| + \|x_i^{(l)} - x_*\| \,] + \|c(x_i^{(l+)})\| \\
&\le \frac{\xi_i^{(l)} K}{\sqrt{K_1}} [\, \|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\| \,] \\
&\qquad + \xi_i^{(l)} K \|x_i^{(l)} - x_*\| + \|c(x_i^{(l+)})\|,
\end{aligned} \tag{4.55}$$

where $k > 0$ is a constant. Due to Lemma 4.1 and (4.55), when i sufficiently large such that $\frac{\xi_i^{(l)} K}{\sqrt{K_1}} \le \frac{1}{2}$, inequality (4.55) yields that

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

Therefore (4.51) holds with $\eta_i^{(l)} = \frac{2\,\xi_i^{(l)} K}{\sqrt{K_1}}$. ∎

Now we obtain the local and superlinear convergence of our algorithm.

THEOREM 4.1. *Suppose Assumption 4.1 and 4.2 hold. Suppose the sequence $\{x_k\}$ generated by Algorithm 2.1 converges to x_*. Suppose that the unit lengths $\alpha_k = 1$ and $\beta_k = 1$ are admissible when x_k is sufficiently close to x_*. Then there exists a neighborhood U of x_* such that for $x_k \in U$, $\{x_k\}$ converges superlinearly.*

Proof. Notice that since $x_{i+1}^{(0)} = x_i^{(l_i)}$, it follows from Lemma 4.1 that it suffices to prove

$$\lim_{i\to\infty} \frac{\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\|}{\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| + \|c(x_i^{(l)})\|} = 0. \tag{4.56}$$

For l with $0 \le l < l_i$, if (2.17) holds, then

$$\|c(x_i^{(l+)})\| \le N_1 \|c(x_i^{(l)})\|^2.$$

Thus, inequality (4.51) implies that

$$\begin{aligned} \frac{\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\|}{\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| + \|c(x_i^{(l)})\|} &\le \eta_i^{(l)} + \frac{2N_1\|c(x_i^{(l)})\|^2}{\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| + \|c(x_i^{(l)})\|} \\ &\le \eta_i^{(l)} + 2N_1\|c(x_i^{(l)})\| \end{aligned}$$

from which (4.56) is implied since $\eta_i^{(l)}$ and $\|c(x_i^{(l)})\|$ go to zero.

For l with $0 \le l < l_i$, if (2.17) does not hold, then (4.41) must must hold. In this case, $x_i^{(l+)} = x_i^{(l)}$. It follows from (4.51) that

$$\frac{\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\|}{\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| + \|c(x_i^{(l)})\|} \le \eta_i^{(l)} + \frac{2\|c(x_i^{(l)}\|}{\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| + \|c(x_i^{(l)})\|}. \tag{4.57}$$

Since $|\Lambda_i^{(l)}|$ is bounded, (4.41) implies that

$$\|c(x_i^{(l)})\| \le K_6 \mu_i.$$

Therefore, inequality (4.57) becomes

$$\frac{\|Z(x_i^{(l+1)})^T \nabla f(x_i^{(l+1)})\| + \|c(x_i^{(l+1)})\|}{\|Z(x_i^{(l)})^T \nabla f(x_i^{(l)})\| + \|c(x_i^{(l)})\|} \le \eta_i^{(l)} + \frac{2K_6\mu_i}{\mu_i^{1/2}} = \eta_i^{(l)} + 2K_6\mu_i^{1/2},$$

which implies (4.56) since $\eta_i^{(l)}$ and μ_i tend to zero. ∎

Finally, we show in Lemmas 4.14 and 4.15 that after a certain number of steps the unit step lengths are admissible. Thus if we try to choose the step length $\alpha_i^{(l)} = 1$ and $\beta_i^{(l)} = 1$ as described in Remark 2.1, then $\alpha_k = 1$ and $\beta_k = 1$ when x_i is sufficiently close to x_*.

LEMMA 4.12. *Suppose $\{x_k\}$ converges to a point x_* at which $H(x_*)$ is positive definite, where*

$$H(x) = Z(x)^T \nabla^2 L(x) Z(x)$$

and if $\sigma < \frac{1}{2}$ and then there is an index $k_0 \geq 0$ such that for all $k \geq k_0$, $\alpha_k = 1$ is admissible.

Proof. First of all, it follows from the algorithm that

$$-\nabla_h p_{\mu_i}(x_i^{(l+)}) = B_i^{(l)} h_i^{(l)}, \tag{4.58}$$

and it follows from the proof of Lemma 4.6 that $\|G_i^{(l)} - H_*\| \to 0$ where $G_i^{(l)}$ is defined as in Lemma 4.6. It follow from (4.50) that there exists an integer i_0 such that for $i \geq i_0$,

$$\theta_i^{(l)} + \|G_i^{(l)} - H_*\| < \kappa \ \min\{\ \omega,\ 1 - 2\sigma\ \}, \tag{4.59}$$

where $\kappa > 0$ is defined in Lemma 4.7. From the mean value theorem,

$$p_{\mu_i}(x_i^{(l+)} + h_i^{(l)}) - p_{\mu_i}(x_i^{(l+)}) = \nabla_h p_{\mu_i}(x_i^{(l+)})^T h_i^{(l)} + \frac{1}{2}(h_i^{(l)})^T G_i^{(l)} h_i^{(l)}.$$

By (4.58), (4.50), (4.59) and Lemma 4.7, we have

$$\begin{aligned} p_{\mu_i}(x_i^{(l+)} + h_i^{(l)}) - p_{\mu_i}(x_i^{(l+)}) - \ & \tfrac{1}{2} \ \ \nabla_h p_{\mu_i}(x_i^{(l+)})^T h_i^{(l)} \\ & = \frac{1}{2}[\nabla_h p_{\mu_i}(x_i^{(l+)}) + G_i^{(l)} h_i^{(l)}]^T h_i^{(l)} \\ & \leq \frac{1}{2}[\ \theta_i^{(l)} + \|G_i^{(l)} - H_*\|\ ]\ \|h_i^{(l)}\|^2 \\ & \leq (\frac{1}{2} - \sigma)\kappa \|h_i^{(l)}\|^2 \\ & \leq -(\frac{1}{2} - \sigma)\nabla_h p_{\mu_i}(x_i^{(l+)})^T h_i^{(l)} \end{aligned} \tag{4.60}$$

Inequality (4.60) is equivalent to

$$p_{\mu_i}(x_i^{(l+)} + h_i^{(l)}) - p_{\mu_i}(x_i^{(l+)}) \leq \sigma \nabla_h p_{\mu_i}(x_i^{(l+)})^T h_i^{(l)}.$$

i.e., $\alpha_i^{(l)} = 1$ satisfies (2.12) for i sufficiently large.

To show that (2.13) holds for $\alpha_i^{(l)} = 1$ when $i \geq i_0$, we use the mean value theorem again to get that

$$\begin{aligned} |\nabla_h p_{\mu_i}(x_i^{(l+)} + h_i^{(l)})^T h_i^{(l)}| \ & = \ |[\nabla_h p_{\mu_i}(x_i^{(l+)}) + G_i^{(l)} h_i^{(l)}]^T h_i^{(l)}| \\ & \leq \ [\theta_i^{(l)} + \|G_i^{(l)} - H_*\|]\ \|h_i^{(l)}\|^2 \\ & \leq \ \omega\ \kappa \|h_i^{(l)}\|^2 \\ & \leq \ -\omega \nabla_h p_{\mu_i}(x_i^{(l+)})^T h_i^{(l)} \end{aligned} \tag{4.61}$$

because of (4.58), (4.50), (4.59) and Lemma 4.7. Inequality (4.61) yields that (2.13) holds for $\alpha_i^{(l)} = 1$ when $i \geq i_0$. ∎

LEMMA 4.13. *Suppose $\{x_n\}$ converges to a point x_* at which $c(x_*) = 0$. 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. Since $\mu_i \to 0$ and $\|v_i^{(l)}\| \to 0$, it suffices to show that when $\|v\| = \| - R^{-T} c(x)\|$ and μ are sufficiently small, we have

$$p_\mu(x + Yv) \leq p_\mu(x) + \sigma [Y^T \nabla p_\mu(x)]^T v$$

whenever $\|c(x)\| > \mu \max\{\frac{\|\lambda(x)\|}{\sigma}, 1\}$. It follows from Taylor's theorem that

$$\begin{aligned} p_\mu(x + Yv) &= p_\mu(x) + \nabla p_\mu(x)^T Y v + \frac{1}{2}(Y v)^T \nabla^2 p_\mu(x) Y v + o(\frac{\|v\|^2}{\mu}) \\ &= p_\mu(x) + \sigma [Y^T \nabla p_\mu(x)]^T v \\ &\qquad + (1-\sigma)\nabla p_\mu(x)^T Y v + \frac{1}{2}\frac{\|c(x)\|^2}{\mu} + o(\frac{\|c(x)\|^2}{\mu}). \end{aligned} \tag{4.62}$$

Similar to proof of Lemma 2.1, we have

$$\nabla p_\mu(x)^T Y v = \lambda(x)^T c(x) - \frac{\|c(x)\|^2}{\mu} \leq \|\lambda(x)\| \, \|c(x)\| - \frac{\|c(x)\|^2}{\mu} \leq (\sigma - 1)\frac{\|c(x)\|^2}{\mu}$$

and

$$\begin{aligned} \frac{1}{2} v^T \nabla_v^2 p_\mu(x) v &= \frac{1}{2} v^T Y^T [\nabla^2 f(x) + \sum_{i=1}^m \frac{c_i(x)}{\mu} \nabla^2 c_i(x)] Y v + \frac{1}{2}\frac{v^T R R^T v}{\mu} \\ &= \frac{1}{2}\frac{v^T R R^T v}{\mu} + \mathcal{O}((1 + \frac{\|c(x)\|}{\mu})\|v\|^2) \\ &= \frac{1}{2}\frac{\|c(x)\|^2}{\mu} + o(\frac{\|c(x)\|^2}{\mu}). \end{aligned}$$

It follows that

$$\begin{aligned} (1-\sigma)\nabla p_\mu(x)^T Y v + \frac{1}{2}\frac{\|c(x)\|^2}{\mu} &+ o(\frac{\|c(x)\|^2}{\mu}) \\ &\leq -(1-\sigma)^2 \frac{\|c(x)\|^2}{\mu} + \frac{1}{2}\frac{\|c(x)\|^2}{\mu} + o(\frac{\|c(x)\|^2}{\mu}) \\ &\leq (\frac{1}{2} - (1-\sigma)^2)\frac{\|c(x)\|^2}{\mu} + o(\frac{\|c(x)\|^2}{\mu}) \leq 0 \end{aligned} \tag{4.63}$$

since $\frac{1}{2} - (1-\sigma)^2 < 0$ by the assumption. Combining (4.62) and (4.63), we know that $\beta_i^{(l)} = 1$ satisfies (2.14) when i is sufficiently large. ∎

5. Numerical Results. In this section we present results of numerical experiments illustrating the performance of Algorithm 2.1. The problem set consists of a number of nonlinear equality constrained problems selected from the CUTE collection [3] and two problems generated by the authors. All numerical experiments discussed in this section 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	200	160	dense	quadratic
TEST2	200	160	dense	nonlinear

Table 1: Description of Problems

All testing problems are briefly described in Table 1. Most problems in Table 1 (all except TEST1 and TEST2) are from the CUTE collection [3]. Problem TEST1 is minimization of a Rosenbrock function [9] with quadratic equality constraints, i.e.,

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

where $a_i \in \Re^n$, $i = 1, 2, \dots, m$, are vectors, and $M_i \in \Re^{n\times n}$, $i = 1, 2, \dots, m$, are symmetric. The nonlinearity of problem TEST1 is high if the symmetric matrices M_i, $i = 1, 2, \dots, m$, are not extremely sparse. In problem TEST2 we add perturbation functions to both the objective and constraint functions in TEST1. Namely, problem TEST2 is

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

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, trigonometric 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 $A = [a_1, a_2, \ldots, a_m]$ and M_i, $i = 1, 2, \ldots, m$ are created randomly.

Problems	number of iterations	function evaluations		*error*
		f, c	g, A	
BT6	12	37	21	$O(10^{-6})$
BT11	9	25	18	$O(10^{-7})$
DIPIGRI	16	70	27	$O(10^{-6})$
DTOC2	12	17	17	$O(10^{-5})$
DTOC4	4	8	8	$O(10^{-5})$
DTOC6	11	18	17	$O(10^{-6})$
GENHS28	6	9	8	$O(10^{-6})$
HS100	17	75	29	$O(10^{-7})$
MWRIGHT	14	36	22	$O(10^{-5})$
ORTHREGA	83	883	91	$O(10^{-5})$
ORTHREGC	24	61	31	$O(10^{-5})$
ORTHREGD	21	90	38	$O(10^{-5})$
TEST1	104	443	116	$O(10^{-5})$
TEST2	149	526	161	$O(10^{-5})$

Table 2: Results Using Algorithm 2.1

When the penalty parameter $\mu > 0$ is large, there is no Maratos effect. Thus it is not necessary to use the curve technique when we minimize the penalty function $p_\mu(x)$ for such a μ. Therefore, we so not use the curve technique until $\mu > 0$ is large. Namely

$$x_{k+1} = x_k + Z_k h + Y_k v$$

when $\mu \geq 1.0$, and

$$x_{k+1} = x_k + s_k(h) + Y_k v$$

when $\mu < 1.0$, where $s_k(h) = Z_k h + Y_k R_k^{-T}[c(x_k) - c(x_k + Z_k h)]$.

When solving problems in Table 1 using Algorithm 2.1, we take $\mu_0 = 1$, $\rho = 0.1$ and $\sigma = 0.0001$. We set the stopping criterion to be $\mu < 10^{-8}$. For problems TEST1 and TEST2, the starting point is $x_0 = [.5, \ .5, \ \ldots, \ .5]^T$. For the testing problems drawn from the CUTE collection, we take the default values.

Table 2 illustrates the results of our numerical experiments for problems in Table 1. The first column gives the name of the problems we solved. The second column shows the number of iterations taken to reach the stopping criterion for different problems. Column "function evaluations" presents the number of function evaluations needed for problems in Table 1. Sub-column "f, c" indicates the number of function evaluations required for the functions

$f(x)$ and $c(x)$, "g, A" for their gradients. The last column shows how accurate Algorithm 2.1 reaches when applied to the testing problems, where the quantity "*error*" is defined as

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

6. Discussion and Concluding Remarks. In [5] Byrd and Nocedal propose algorithms based on reduced Hessian methods. Byrd and Nocedal prove that, for their algorithms,

$$\lim_{k \to \infty} [\|Z(x_k)^T \nabla f(x_k)\| + \|c(x_k)\|] = 0 \tag{6.1}$$

under an assumption stronger than condition (3.7). In particular, Byrd and Nocedal assume that there exists a $\gamma > 0$ such that

$$\mathrm{eig}_{\min}(Z_k^T \nabla^2 L(x, \lambda_k) Z_k) \geq \gamma, \quad \forall x \text{ in the line search segment.} \tag{6.2}$$

Moreover, algorithms in [5] cannot preserve the positive definiteness of B_k without assumption (6.2). However, assumption (6.2) is rarely satisfied when x_k is far away from the solution. Therefore, in contrast to Algorithm 2.1, algorithms in [5] may fail when applied to general nonlinear functions.

To prove our local convergence results, we need Assumption 4.2. Byrd and Nocedal [5] do not require this assumption for local convergence. However, for the Fletcher exact penalty function, Byrd and Nocedal assume that the Rayleigh quotients $s^T B_k s / s^T s$ are uniformly bounded away from zero for all k. This assumption is as strong as Assumption 4.2. Both Assumption 4.2 and the Byrd and Nocedal assumption in [5] are weaker than boundedness assumptions on $\|B_k\|$ and $\|B_k^{-1}\|$.

Boundedness assumptions on $\|B_k\|$ and $\|B_k^{-1}\|$ are required by Bertsekas [1], Boggs, Tolle and Wang [2], and Powell [17] for Fletcher's exact penalty function $\phi_\mu(x)$. This assumption is very often used for proving global and local convergence of algorithms for problem (1.1). The purpose of this assumption is to establish that the parameter μ in Fletcher's exact penalty function is fixed after a certain number of iterations. As we have shown in section 4, the measure function $\phi_\mu(x)$ decreases by a factor $r < 1$:

$$\phi_\mu(x_{k+1}) - \phi_\mu(x_*) \leq r\ [\phi_\mu(x_k) - \phi_\mu(x_*)].$$

If μ is finally fixed, then it follows that (4.3) holds and inequality (4.2) is easy to prove. Therefore (4.1) holds and superlinear convergence is established. In Algorithm 2.1 parameter μ cannot be fixed and the value of the penalty function increases whenever μ is reduced. Under the assumption that $s^T B_k s / s^T s$ is bounded, the penalty function decreases sufficiently for fixed μ, so that we can still achieve the local superlinear convergence results for Algorithm 2.1.

We have presented a quasi-Newton L_2-penalty method for solving equality constrained minimization problems. When quasi-Newton methods are applied to nonlinear equality constrained minimization problems, one of major difficulties is that positive definiteness of Hessian approximations may not be preserved. In addition, due to the effect of the penalty term, the L_2-penalty function often forces steps to be short when far from the solution. Thus, the L_2-penalty function is usually disregarded as a merit function for nonlinearly constrained minimization. In this paper we have proposed a new approach which not only maintains positive definite Hessian approximations, but also avoids unacceptably small steps when far from the solution.

The pivotal feature of our new technique is to invoke a change of variables at each iterate to improve the ability of the algorithm to follow the constraint level set. This local transformation gives rise to an appropriate block diagonal approximation to the Hessian which is then used to construct a quasi-Newton method. The local transformation also allows the algorithm to take "large" curved steps in the original variables, resulting in a mild increase in the penalty term.

We have established global convergence properties and superlinear convergence rate results for our algorithm. Numerical results of preliminary computational experiments indicate practical potential. Indeed, the theoretical properties along with our numerical results indicate that our algorithm has considerable potential for efficiently solving nonlinear equality constrained minimization problems.

REFERENCES

[1] D.P.Bertsekas, "Variable metric methods for constrained optimization using differentiable exact penalty function", *Proceedings of the 18th Allerton Conference Communication, Control and Computing* (University of Illinois Press, Champaign, 11, 1980) 584-591.

[2] P.T. Boggs, J.W. Tolls and P. Wang, "On the local convergence of quasi-Newton methods for constrained optimization", *SIAM Journal of Control and Optimization* 20 (1982) 161-171.

[3] 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).

[4] R.H. Byrd and J. Nocedal, "A tool for the analysis of quasi-Newton methods with application to unconstrained minimization", *SIAM Journal on Numerical Analysis* 26(1989) 727-739.

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

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

[7] T.F. Coleman and A.R. Conn, "On the local convergence of a quasi-Newton method for the nonlinear programming problem", *SIAM Journal on Numerical Analysis* 21(1984) 755-769.

[8] 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.

[9] J. E. Dennis, Jr. and R. B. Schnabel, "Numerical methods for unconstrained optimization and nonlinear equations", (Prentice-Hall, Englewood Cliffs, NJ, 1983).

[10] R.Fontecille, "Local convergence of secant methods for nonlinear constrained optimization", *SIAM Journal on Numerical Analysis* 25(1988) 692-712.

[11] D. Gabay, "Reduced quasi-Newton methods with feasibility improvement for nonlinearly constrained optimization", *Mathematical Programming Study* 16(1982) 18-44.

[12] J.C. Gilbert, "Maintaining the positive definiteness of the matrices in reduced Hessian methods for equality constrained optimization", International Institute for Applied System Analysis Technical Report WP 87-123 (Laxenberg, Austria, 1987).

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

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

[15] W. Murray and M.H.Wright, "Projected Lagrangian methods based on the trajectories of penalty and barrier functions", Systemic Optimization Laboratory Report 78-23, Stanford University (Stanford, CA, 1978).

[16] J. Nocedal and M. Overton, "Projected Hessian updating algorithms for nonlinearly constrained optimization", *SIAM Journal on Numerical Analysis* 22(1985) 821-850.

[17] M.J.D. Powell, "Some global convergence properties of a metric algorithm for minimization without exact line searches", in: R.W. Cottle and C.E. Lemke, eds., *Nonlinear Programming, SIAM-AMS Proceedings, Vol. IX* (Society for Industrial and Applied Mathematics, Philadelphia, 1976.)

[18] M.J.D. Powell, "Variable metric methods for constrained optimization", in: A. Bachem, M. Grotschel and B. Korte, eds., *Mathematical programming, the state of the art* (Springer-Verlaag, Berlin, 1983) pp. 288-311.

[19] M.J.D. Powell and Y. Yuan, "A recursive quadratic programming algorithm that uses differentiable exact penalty functions", *Mathematical Programming* 35 (1986) 265-278.