SIAM J. NUMER. ANAL.
Vol. 21, No. 4, August 1984

009

ON THE LOCAL CONVERGENCE OF A QUASI-NEWTON METHOD FOR THE NONLINEAR PROGRAMMING PROBLEM*

THOMAS F. COLEMAN† AND ANDREW R. CONN‡

Abstract. In this paper we propose a new local quasi-Newton method to solve the equality constrained nonlinear programming problem. The pivotal feature of the algorithm is that a *projection* of the Hessian of the Lagrangian is approximated by a sequence of symmetric positive definite matrices. The matrix approximation is updated at every iteration by a projected version of the DFP or BFGS formula: this involves two evaluations of the Lagrangian gradient per iteration. We establish that the method is locally convergent and the sequence of x-values converges to the solution at a 2-step Q-superlinear rate.

1. Introduction. Quasi-Newton methods have had a large measure of success in the minimization of smooth nonlinear functions

$$f(x): R^n \to R^1.$$

In particular, the Davidon–Fletcher–Powell (DFP) and Broyden–Fletcher–Goldfarb–Shanno (BFGS) updating formulae have given solid numerical performances over the past decade and are generally accepted as the best rank-2 updating formulae (for dense problems). In addition to their numerical record, these methods have two significant theoretical properties: they yield locally Q-superlinearly convergent algorithms and their Hessian approximations remain positive definite.

It is difficult to fully explain the superior numerical performance of the DFP/BFGS method relative to other updates; however the maintenance of positive definite Hessian approximations seems crucial—it is also a "natural" property since the true Hessian at the solution will likely be positive definite (and will certainly be at least positive semi-definite). In addition, positive definiteness allows for a stable implementation (Gill and Murray (1974)) and ensures that search directions are also descent directions.

The situation for minimization in the presence of nonlinear constraints is less satisfactory. Successive quadratic programming (SQP) and projection approaches have recently been in vogue: however, a true Q-superlinear quasi-Newton method for the nonconvex case is unknown to the authors. Powell (1978) has adapted the BFGS formula to the nonlinearly constrained case. He gives sufficient conditions under which a successive quadratic programming approach will yield a 2-step Q-superlinear convergence rate (assuming convergence) but does not show that his modified BFGS method satisfies these conditions. Instead R-superlinear convergence is proven. Interestingly, the sufficiency conditions given by Powell necessitate that only a *projection* of the Lagrangian Hessian approximations be suitably accurate. We also note that Han (1976) has proven that this SQP/BFGS method exhibits Q-superlinear convergence for the convex case.

Other authors, Boggs, Tolle and Wang (1982), have given sufficient (and necessary) conditions for Q-superlinear convergence for the constrained problem, however we are unaware of an updating method which satisfies these conditions. Tanabe (1981) has proposed various projected updating schemes but, to our knowledge, has not established convergence properties.

In our opinion, a major difficulty with most of these approaches is that a full (n by n) positive definite Hessian approximation is required by each of the methods but

* Received by the editors August 24, 1982, and in final form November 11, 1983.
† Computer Science Department, Cornell University, Ithaca, New York 14853.
‡ Computer Science Department, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

only a *projection* of the Hessian of the Lagrangian need be positive definite at the solution. Therefore we feel that a more natural approach is to recur a positive definite approximation to the projection of the Hessian of the Lagrangian. Gill and Murray (1974) have followed such a strategy in the case where all constraints are linear, however there has been little work along these lines for the nonlinearly constrained problem. (Another possibility is to recur a positive definite approximation to the Hessian of an *augmented* Lagrangian function. Glad (1979), Han (1977) and Tapia (1977) have all established Q-superlinear convergence results for such methods. We will not discuss this type of method further in this paper.)

Coleman and Conn (1982a, b) have suggested Newton and discrete-Newton methods for nonlinearly constrained problems, which require only a projected Hessian approximation. The method we describe here is a direct extension of the local Newton method given by Coleman and Conn.

The motivating remarks given in these previous papers are applicable here also, however we present an alternative view. Consider the problem

$$\text{minimize } f(x), \quad \text{subject to } c_i(x)=0, \ i=1,\cdots,t,$$

where all functions are twice continuously differentiable. Suppose that our current estimate to the solution is x and let C be the n by t matrix of constraint gradients, evaluated at x. Let Z be an n by $(n-t)$ matrix, whose columns form an orthonormal basis for the null space of C^T (assume that C has rank t). Finally, let the correction to x, say δ, be defined as the solution to the following quadratic program:

$$\text{minimize } \nabla f(x)^T\delta+\tfrac{1}{2}\delta^T ZBZ^T\delta, \quad \text{subject to } C^T\delta+c(x)=0, \tag{1.1}$$

where B is an $(n-t)$ by $(n-t)$ positive definite matrix. Under the conditions mentioned above, (1.1) has a unique solution given by

$$\delta=h+v, \tag{1.2}$$

where

$$h=-ZB^{-1}Z^T\nabla f(x), \tag{1.3}$$

$$v=-C(C^TC)^{-1}c(x). \tag{1.4}$$

Note that B can be considered to be a positive definite approximation to

$$Z_*^T[\nabla^2 f(x^*)-\sum \lambda_i^*\nabla^2 c_i(x^*)]Z_*, \tag{1.5}$$

where $C_*^TZ_*=0$, $Z_*^TZ_*=I$, $C_*=(\nabla c_1(x^*),\cdots,\nabla c_t(x^*))$, $\nabla f(x^*)=C_*\lambda^*$ and $c(x^*)=0$.

Under second-order sufficiency conditions, the $(n-t)$ by $(n-t)$ matrix (1.5) is positive definite. The method we propose in § 3 uses a projected form of the DFP(BFGS) update to recur a positive definite approximation to (1.5) and involves two (Lagrangian) gradient evaluations per iteration. The correction to x that we analyze differs slightly from (1.2) in that (1.4) is replaced with

$$v=-C(C^TC)^{-1}c(x+h). \tag{1.6}$$

We emphasize that all results given in § 3 *are valid if* (1.4) *replaces* (1.6). We have carried out the analysis using (1.6) because of a result given in Coleman and Conn (1982b) which states that (1.3) together with (1.6) guarantee that a certain exact penalty function will decrease, provided x is sufficiently close to x^*. The result is valid in the discrete-Newton case and is not true if (1.4) is used instead of (1.6). We have not yet proven that a similar result is true for the case when B is a quasi-Newton approximation but it is this possibility which prompted the use of (1.6).

In § 2 we present conditions which are sufficient to give a 2-step Q-superlinear convergence rate (assuming convergence). These conditions are slightly more general than those given by Powell (1978) in that they do not presuppose a particular algorithm class. These conditions are in the spirit of the superlinearity characterization for unconstrained optimization given by Dembo, Eisenstat and Steihaug (1982).

In § 3 we describe the algorithm and then establish that the method is locally 2-step Q-superlinearly convergent. The method of proof is similar to that used by Broyden, Dennis, and Moré (1973) and Dennis and Moré (1974) for the unconstrained case.

In § 4 we give our concluding observations and discuss future work.

2. Sufficient conditions for 2-step Q-superlinear convergence. Consider the following equality constrained nonlinear programming problem:

$$\text{minimize } f(x), \quad \text{subject to } c_i(x)=0, i=1,\cdots,t,$$

where *all functions are twice continuously differentiable on an open convex set D of R^n and map $D \to R^1$*. Let $x^* \in D$ be a local solution to the equality constrained nonlinear programming problem. The question we address in this section is this: given that a sequence of points $\{x^k\}$ converges to x^*, when can we be assured that a 2-step Q-superlinear convergence rate is achieved? That is, what reasonable conditions ensure that

$$\|x^{k+1}-x^*\| = o\|x^{k-1}-x^*\| \quad ?$$

(We make extensive use of the "O" and "o" notation, where $\phi_k = O(\psi_k)$ means that the ratio ϕ_k/ψ_k remains bounded as k tends toward infinity and $\phi_k = o(\psi_k)$ means that the ratio ϕ_k/ψ_k tends to zero as k tends to infinity.)

Definitions and assumptions. Unless stated otherwise, the results given in this paper will all be subject to the following assumptions.

Let $C(x)$ denote the n by t matrix $(\nabla c_1(x),\cdots,\nabla c_t(x))$ and let $c(x)$ denote the vector $(c_1(x),\cdots,c_t(x))^T$. Define C_* and c^* to be $C(x^*)$ and $c(x^*)$ respectively. For any x in D, define $\lambda=\lambda(x)$ to be the vector $[C(x)^TC(x)]^{-1}C(x)^T\nabla f(x)$. *We will assume that $C(x^*)$ has full column rank. Since $C(x)$ is continuous it follows that there is an open convex set D containing x^* such that for all x in D, the singular values of $C(x)$ are uniformly bounded on D, above and below, by positive scalars.*

An n by $(n-t)$ matrix $Z(x)$ is defined to be a *Lipschitz continuous function of x in D* satisfying

$$Z(x)^TZ(x)=I \tag{2.1}$$

and

$$C(x)^TZ(x)=0, \tag{2.2}$$

where I represents the identity matrix. (Coleman and Sorensen (1982) have demonstrated that a suitable $Z(x)$ is well-defined and efficiently computable in a neighbourhood of x^*.) Uniquely define vectors $u(x)$ and $w(x)$ by

$$x-x^*=C(x)w(x)+Z(x)u(x). \tag{2.3}$$

Since x^* is a solution, it follows that the gradient of f can be expressed as a linear combination of the gradients of the constraint functions. That is, there exists a vector

$\lambda^* \in R^t$, such that

$$\nabla f(x^*) = C_* \lambda^*. \tag{2.4}$$

Define $L(x)$ to be the Lagrangian function $f(x) - c(x)^T \lambda^*$. *It will be assumed that the second-order sufficiency conditions hold at* x^*. *Thus the matrix*

$$H(x^*, x^*) = Z(x^*)^T [\nabla^2 f(x^*) - \sum \lambda_i^* \nabla^2 c_i(x^*)] Z(x^*),$$

is positive definite. We note that this implies that the eigenvalues of the $(n-t)$ by $(n-t)$ matrix $H(x, y)$, defined by

$$H(x, y) = Z(x)^T [\nabla^2 f(y) - \sum \lambda_i^* \nabla^2 c_i(y)] Z(x),$$

are uniformly bounded below by a positive scalar on an open convex region ((D, D), say) containing (x^*, x^*). The above implication is a consequence of the following: The eigenvalues of the matrix $H(x, y)$ are continuous functions of the elements of the matrix (see Ortega (1972, p. 45), for example). The elements of $H(x, y)$ vary continuously with (x, y) due to Z and $\nabla^2 f, \nabla^2 c_i$ being continuous functions of x and y respectively. Finally, the result follows from observing that $H(x^*, x^*)$ is positive definite. *We assume that the radius of D is sufficiently small so that the eigenvalues of the Hessian matrices*

$$\nabla^2 f(x), \nabla^2 c_i(x), \qquad i = 1, \cdots, t$$

are uniformly bounded above on D by a positive scalar and that the Hessian matrices satisfy a Lipschitz condition on D.

When the above quantities are evaluated at a particular point x^k, then the argument x^k will be abbreviated to a simple subscript or superscript. For example, $C(x^k)$ will be written C_k and $w(x^k)$ becomes w^k. We will denote $H(x^k, x^k)$ by H_k. The symbol "*" will be used to denote a function evaluated at x^*: for example, ∇f^* represents $\nabla f(x^*)$.

Let δ^k represent $x^{k+1} - x^k$, and define

$$r_L^k = Z_k^T \nabla f(x^k) + H_k Z_k^T \delta^k \quad \text{and} \quad r_c^k = c^k + C_k^T \delta^k.$$

Unless noted otherwise, the symbol $\|\cdot\|$ will denote the vector or matrix 2-norm. *One final assumption: we assume that finite convergence does not occur:* $x^k \neq x^*$, *for all k.*

Note that the residuals, r_L^k and r_c^k reflect the accuracy to which the systems $H_k Z_k^T \delta^k = -Z_k^T \nabla f(x^k)$ and $C_k^T \delta^k = -c^k$ are solved. (Alternatively, we can view r_L^k as reflecting the accuracy of the projected quasi-Newton approximation since $r_L^k = (B_k - H_k) B_k^{-1} Z_k^T \nabla f(x^k)$.) A natural approach is to show that if the residuals are suitably small at each iteration (i.e., the systems are solved to suitable accuracy) then a superlinear convergence rate is achieved. This view is compatible with the superlinear characterization, given by Dembo, Eisenstat and Steihaug (1982), for systems of nonlinear equations. The following result establishes a crucial link between residual size and 2-step Q-superlinear convergence.

THEOREM 2.1. *If* x^k *converges to* x^*, $\|x^{k+1} - x^*\| = O\|x^k - x^*\|$, *and*

$$\|r_L^k\| + \|r_c^k\| = o(\|Z_k^T \nabla f^k\| + \|c^k\|) \tag{2.5}$$

then

$$\|x^{k+1} - x^*\| = o\|x^{k-1} - x^*\|.$$

Proof. The proof is divided into three parts and uses w and v as defined in (2.3). In the first part it is established that $\|w^{k+1}\| = o\|x^k - x^*\|$; in Part 2, it is proven that $\|u^{k+1}\| = o\|x^{k-1} - x^*\|$; finally, in Part 3, the desired result is obtained.

Part 1. Clearly we can write

$$C_k^T(x^{k+1} - x^*) = C_k^T(x^k - x^*) - c^k + c^k + C_k^T\delta^k. \tag{2.6}$$

If we add $C_{k+1}^T(x^{k+1} - x^*)$ to both sides of (2.6), rearrange, and then take norms, we obtain

$$\begin{aligned}\|w^{k+1}\| \leqq \|(C_{k+1}^T C_{k+1})^{-1}\|[\|C_*^T(x^k - x^*) + (C_k - C_*)^T(x^k - x^*) - c^k\| \\ + \|c^k + C_k^T\delta^k\| + \|(C_{k+1} - C_k)^T(x^{k+1} - x^*)\|].\end{aligned} \tag{2.7}$$

By Taylor's theorem, and since $c^* = 0$,

$$c^k = C_*^T(x^k - x^*) + o\|x^k - x^*\|,$$

and therefore the first term in (2.7) is $o\|x^k - x^*\|$. By (2.5),

$$\|r_c^k\| = o(\|Z_k^T\nabla f^k\| + \|c^k\|). \tag{2.8}$$

Since $\nabla L(x)$ and $c(x)$ are Lipschitz continuous, $c^* = 0$, $Z_k^T\nabla f^k = Z_k^T\nabla L^k$ and $\nabla L^* = 0$, it follows that

$$\|Z_k^T\nabla f^k\| + \|c^k\| = O\|x^k - x^*\|, \tag{2.8.1}$$

and therefore, combining (2.8.1) with (2.8) it follows that the second term of (2.7) is $o\|x^k - x^*\|$. But clearly, by assumption, $\|x^{k+1} - x^*\| = O\|x^k - x^*\|$, and by the Lipschitz continuity of $C(x)$ we have $\|C_{k+1} - C_k\| = O\|x^{k+1} - x^k\|$. However,

$$\|x^{k+1} - x^k\| \leqq \|x^{k+1} - x^*\| + \|x^k - x^*\|$$

and hence $\|C_{k+1} - C_k\| = O\|x^k - x^*\|$. It follows that the third term of (2.7) is $o\|x^k - x^*\|$ and Part 1 is established: $\|w^{k+1}\| = o\|x^k - x^*\|$.

Part 2. Clearly we can write

$$Z_k^T(x^{k+1} - x^*) = H_k^{-1}[H_k Z_k^T(x^k - x^*) - Z_k^T(\nabla f^k - \nabla f^k) + H_k Z_k^T\delta^k]. \tag{2.9}$$

If we add $Z_{k+1}^T(x^{k+1} - x^*)$ to both sides, rearrange and take norms, we obtain

$$\|u^{k+1}\| \leqq T_1 + T_2 + T_3 + T_4, \tag{2.10}$$

where

$$T_1 = \|H_k^{-1}[-Z_k^T\nabla f^k + H(x^k, x^*)Z_k^T(x^k - x^*)]\|,$$

$$T_2 = \|-H_k^{-1}(H(x^k, x^*) - H_k)Z_k^T(x^k - x^*)\|,$$

$$T_3 = \|H_k^{-1}[Z_k^T\nabla f^k + H_k Z_k^T\delta^k]\|,$$

$$T_4 = \|(Z_{k+1} - Z_k)^T(x^{k+1} - x^*)\|.$$

By Taylor's theorem and $\nabla L^* = 0$,

$$\nabla L^k = \nabla^2 L^*(x^k - x^*) + o\|x^k - x^*\|,$$

and thus, using (2.3),

$$Z_k^T\nabla f^k = H(x^k, x^*)Z_k^T(x^k - x^*) + Z_k^T\nabla^2 L^* C_k w^k + o\|x^k - x^*\|.$$

But $\|\nabla^2 L^* C_k\|$ and $\|H_k^{-1}\|$ are bounded above and $\|w^k\|$ is $o\|x^{k-1} - x^*\|$, by Part 1; therefore, T_1 is $o\|x^{k-1} - x^*\|$. By convergence, $\|H(x^k, x^*) - H_k\| \to 0$, and this along with the fact that $\|H_k^{-1}\|$ is bounded above implies that T_2 is $o\|x^k-x^*\|$, and thus T_2 is $o\|x^{k-1}-x^*\|$.

Assumption (2.5) implies that

$$\|Z_k^T\nabla f^k+H_kZ_k^T\delta^k\|=o(\|Z_k^T\nabla f^k\|+\|c^k\|). \tag{2.11}$$

By (2.8.1) and the boundedness of H_k^{-1}, it follows that T_3 is $o\|x^k-x^*\|$ which implies that T_3 is $o\|x^{k-1}-x^*\|$. Finally, since $\|x^{k+1}-x^*\|$ is $O\|x^{k-1}-x^*\|$, and $\|Z_{k+1}-Z_k\|\to 0$, T_4 is $o\|x^{k-1}-x^*\|$. We have established that $T_i=o\|^{k-1}-x^*\|$ for $i=1,2,3,4$, which, in light of (2.10), implies that $\|u^{k+1}\|=o\|x^{k-1}-x^*\|$.

Part 3. By definition,

$$x^{k+1}-x^*=C_{k+1}w^{k+1}+Z_{k+1}u^{k+1},$$

which implies

$$\|x^{k+1}-x^*\|\leqq\|C_{k+1}\|\cdot\|w^{k+1}\|+\|Z_{k+1}\|\cdot\|u^{k+1}\|.$$

But $\|x^k-x^*\|$ is $O\|x^{k-1}-x^*\|$, and therefore, by Part 1, $\|w^{k+1}\|=o\|x^{k-1}-x^*\|$. Part 2 establishes that $\|u^{k+1}\|=o\|x^{k-1}-x^*\|$ and since $\|C_k\|$ and $\|Z_k\|$ are bounded above, it follows that $\|x^{k+1}-x^*\|=o\|x^{k-1}-x^*\|$. □

The conditions given here are closely related to those given by Powell (1978): however, the above conditions do not presuppose a particular algorithm class. In the next section we employ Theorem 2.1 to establish the local convergence rate property.

3. The algorithm and its properties. In this section we develop and analyze, in detail, a projected DFP updating procedure. We have chosen to focus on the DFP updating scheme, instead of the BFGS update, in order to follow more closely the results of Broyden, Dennis and Moré (1973) and Dennis and Moré (1974). It is not difficult to see, as Dennis and Moré (1974, § 4) indicate for the unconstrained case, that the results are equally true for the projected BFGS update specified below by (3.5.1).

The method we are concerned with is defined by:

(3.1) $h^k\leftarrow -Z_kB_k^{-1}Z_k^T\nabla f^k$

(3.2) $x^{k+}\leftarrow x^k+h^k$

If $(h^k=0)$ **go to** (3.6)

(3.3) $s^k\leftarrow Z_k^T(x^{k+}-x^k)$

(3.4) $y^k\leftarrow Z_k^T[(\nabla f^{k+}-C_{k+}\lambda^k)-\nabla f^k]$

(3.5) $$B_{k+1}\leftarrow B_k+\frac{[y^k-B_ks^k](y^k)^T+y^k[y^k-B_ks^k]^T}{(s^k)^Ty^k}-\frac{(s^k)^T(y^k-B_ks^k)y^k(y^k)^T}{((s^k)^Ty^k)^2}$$

(3.6) $v^k\leftarrow -C_k(C_k^TC_k)^{-1}c^{k+}$

(3.7) $x^{k+1}\leftarrow x^{k+}+v^k$

(3.8) $\lambda^{k+1}\leftarrow(C_{k+1}^TC_{k+1})^{-1}C_{k+1}^T\nabla f^{k+1}.$

Update (3.5) is just a projected version of the DFP formula. The corresponding projected BFGS formula is

$$B_{k+1} \leftarrow B_k + \frac{y^k(y^k)^T}{(y^k)^T s^k} - \frac{B_k s^k (s^k)^T B_k}{(s^k)^T B_k s^k}. \tag{3.5.1}$$

Note that if $h^k = v^k = 0$, then $x^k = x^*$. Once again we remark that (3.6) could be replaced with (1.4) and all results in this paper remain valid. Note also that the algorithm involves two evaluations of the gradient of the Lagrangian function per iteration. This contrasts with quasi-Newton methods for unconstrained optimization which involve one gradient evaluation per iteration. In the next three lemmas we establish some useful bounds.

LEMMA 3.1. *Provided $x^k \in D$ there exists a positive scalar K_0 such that.*

$$\|\lambda^k - \lambda^*\| \leqq K_0 \|x^k - x^*\|.$$

Proof.

$$\begin{aligned}
\|\lambda^k - \lambda^*\| &= \|(C_k^T C_k)^{-1} C_k^T \nabla f^k - (C_*^T C_*)^{-1} C_*^T \nabla f^*\| \\
&\leqq \|(C_*^T C_*)^{-1} C_*^T [\nabla f^k - \nabla f^*]\| \\
&\quad + \|(C_*^T C_*)^{-1} [C_k^T - C_*^T] \nabla f^k\| \\
&\quad + \|(C_k^T C_k)^{-1} [C_*^T C_* - C_k^T C_k] (C_*^T C_*)^{-1}\| \cdot \|C_k^T \nabla f^k\|.
\end{aligned}$$

But

$$\begin{aligned}
\|C_*^T C_* - C_k^T C_k\| &= \|(C_k - C_*)^T (C_k - C_*) - C_k^T (C_k - C_*) - (C_k - C_*)^T C_k\| \\
&\leqq \|C_k - C_*\|^2 + 2\|C_k\| \cdot \|C_k - C_*\|.
\end{aligned}$$

Hence, by considering that $C(x)$ and $\nabla f(x)$ are Lipschitz continuous on D, $\|(C(x)^T C(x))^{-1}\|$ and $\|C(x)\|$ are bounded above on D, and

$$\|\nabla f^k\| \leqq \|\nabla f^* \| \nabla f^* \| + O\|x^k - x^*\|,$$

we obtain the required result. □

LEMMA 3.2. *For some positive constant K_1 and any $Z(x)$, $x \in D$, $x^k \in D$,*

$$\|Z(x)^T (\nabla^2 f^k - \sum \lambda_i^k \nabla^2 c_i^k) Z(x) - H(x, x^*)\| \leqq K_1 \|x^k - x^*\|.$$

Proof. Clearly, by using $\|Z\| = 1$, the Lipschitz continuity of the Hessian matrices and the upper boundedness of $\nabla^2 c_i$, $\nabla^2 f$,

$$\begin{aligned}
\|Z(x)^T (\nabla^2 f^k - \sum \lambda_i^k \nabla^2 c_i^k) Z(x) - H(x, x^*)\| &\leqq \|(\nabla^2 f^k - \sum \lambda_i^k \nabla^2 c_i^k) - (\nabla^2 f^* - \sum \lambda_i^* \nabla^2 c_i^*)\| \\
&\leqq \tau_1 \|x^k - x^*\| + \tau_2 \|\lambda^k - \lambda^*\|,
\end{aligned}$$

for some positive scalars τ_1 and τ_2. Hence, by considering Lemma 3.1, the result follows. □

LEMMA 3.3. *For some positive constant K_2 and all x in D, $x^k \in D$,*

$$\|Z_k^T [\nabla^2 f(x) - \sum \lambda_i(x) \nabla^2 c_i(x)] Z_k - Z_*^T [\nabla^2 f(x) - \sum \lambda_i(x) \nabla^2 c_i(x)] Z_*\| \leqq K_2 \|x^k - x^*\|.$$

Proof. Let A denote the matrix $\nabla^2 f(x) - \sum \lambda_i(x) \nabla^2 c_i(x)$. Clearly,

$$\begin{aligned}
\|Z_k^T A Z_k - Z_*^T A Z_*\| &= \|(Z_k^T - Z_*) A (Z_k - Z_*) - 2 Z_*^T A Z_* + Z_k^T A Z_* + Z_*^T A Z_k\| \\
&\leqq \|Z_k - Z_*\|^2 \|A\| + 2\|Z_k - Z_*\| \cdot \|A\|.
\end{aligned}$$

But, by assumption, $Z(x)$ is Lipschitz continuous and since

$$A(x)=\nabla^2 L(x)+O\|\lambda(x)-\lambda^*\|,$$

the result follows from the boundedness of $\|\nabla^2 L(x)\|$ on D, and Lemma 3.1. □

The following result utilizes the bounds established in the previous three lemmas and, in conjunction with Lemma 3.5, will yield the convergence rate result.

LEMMA 3.4. *Assuming that* $\|x^{k+1}-x^*\|=O\|x^k-x^*\|$, *there exists a positive scalar* $\hat{\varepsilon}$ *such that if* $\|x^k-x^*\|\leqq\hat{\varepsilon}$, *then*

$$\|My^k-M^{-1}s^k\|\leqq\tfrac{1}{3}\|M^{-1}s^k\|,$$

where $M=H_*^{-1/2}$.

Proof. Clearly,

$$\|My^k-M^{-1}s^k\|\leqq\|M\|\cdot\|y^k-H_*s^k\|. \tag{3.9}$$

By Taylor's theorem, and Lipschitz continuity of $\nabla^2 f$, $\nabla^2 c_i$,

$$(\nabla f^{k+}-C_{k+}\lambda^k)=(\nabla f^k-C_k\lambda^k)+(\nabla^2 f^k-\sum\lambda_i^k\nabla^2 c_i^k)(x^{k+}-x^k)$$
$$+E_k(x^{k+}-x^k), \tag{3.9.1}$$

where $\|E_k\|=O\|x^{k+}-x^k\|$. But,

$$x^{k+}-x^k=Z_kZ_k^T(x^{k+1}-x^k)=Z_ks^k \tag{3.9.2}$$

and therefore, combining (3.9.1) and (3.9.2) and multiplying by Z_k^T establishes

$$y^k=Z_k^T[\nabla^2 f^k-\sum\lambda_i^k\nabla^2 c_i^k]Z_ks^k+Z_k^TE_kZ_ks^k. \tag{3.10}$$

But $(h^k)^Tv^k=0$ implies $\|x^{k+}-x^k\|\leqq\|x^{k+1}-x^k\|$. Hence,

$$\|E_k\|=O\|x^{k+1}-x^k\|=O(\max\{\|x^{k+1}-x^*\|,\|x^k-x^*\|\}).$$

Therefore, using (3.10), there exists a positive constant K_3 such that

$$\begin{aligned}\|y^k-H_*s^k\|&\leqq(\|Z_k^T[\nabla^2 f^k-\sum\lambda_i^k\nabla^2 c_i^k]Z_k-H_*\|\\&\quad+K_3\max\{\|x^{k+1}-x^*\|,\|x^k-x^*\|\})\cdot\|s^k\|\\&\leqq\|Z_k^T[\nabla^2 f^k-\sum\lambda_i^k\nabla^2 c_i^k]Z_k-Z_*^T[\nabla^2 f^k-\sum\lambda_i^k\nabla^2 c_i^k]Z_*\|\,\|s^k\|\\&\quad+\|Z_*^T[\nabla^2 f^k-\sum\lambda_i^k\nabla^2 c_i^k]Z_*-H_*\|\cdot\|s^k\|\\&\quad+K_3\max\{\|x^{k+1}-x^*\|,\|x^k-x^*\|\}\cdot\|s^k\|.\end{aligned}$$

Hence, in light of Lemmas 3.2, 3.3 and provided $\hat{\varepsilon}$ is sufficiently small,

$$\|y^k-H_*s^k\|\leqq(2K_1+2K_2+K_3)\max\{\|x^{k+1}-x^*\|,\|x^k-x^*\|\}\cdot\|s^k\|. \tag{3.10.1}$$

Since $\|x^{k+1}-x^*\|=O\|x^k-x^*\|$ (by assumption) it follows that for $\hat{\varepsilon}$ sufficiently small,

$$\|y^k-H_*s^k\|\leqq\frac{\|s^k\|}{3\|M\|^2},$$

which implies, by (3.9)

$$\|My^k-M^{-1}s^k\|\leqq\tfrac{1}{3}\|M^{-1}s^k\|.$$ □

Dennis and Moré (1974, Lemma 3.1) established the following "bounded deterioration" result. For completeness, we reproduce it here.

LEMMA 3.5. *Let M be a nonsingular symmetric matrix of order $n-t$ such that $\|My^k - M^{-1}s^k\| \leqq \frac{1}{3}\|M^{-1}s^k\|$ for some vectors y^k and s^k in R^{n-t} with $s^k \neq 0$. Then $(y^k)^T s^k > 0$ and thus B_{k+1} is well-defined by the update formula* (3.5). *Moreover, there are positive constants α_0, α_1 and α_2 (depending only on M and $n-t$) such that for any symmetric matrix A of order $n-t$,*

$$\|B_{k+1} - A\|_M \leqq \left[(1-\alpha_0\theta_k^2)^{1/2} + \frac{\alpha_1\|My^k - M^{-1}s^k\|}{\|M^{-1}s^k\|}\right] \cdot \|B_k - A\|_M + \alpha_2 \frac{\|y^k - As^k\|}{\|M^{-1}s^k\|},$$

where $\|Q\|_M = \|MQM\|_F$ (F denotes the Frobenius norm), $\alpha_0 \in (0, 1]$ and

$$\theta_k = \begin{cases} \dfrac{\|M[B_k - A]s^k\|}{\|B_k - A\|_M \|M^{-1}s^k\|} & \text{for } B_k \neq A, \\ 0 & \text{otherwise.} \end{cases}$$

We are now ready to prove, in Theorem 3.6, Lemma 3.7, and Theorem 3.8, that a 2-step Q-superlinear convergence rate is exhibited, provided we assume that the sequence converges. These results follow almost directly from the results of Dennis and Moré (1974) and Theorem 2.1.

THEOREM 3.6. *Assume that $\sum \|x^k - x^*\| < \infty$, $\|x^{k+1} - x^*\| = O\|x^k - x^*\|$, and that B_0 is symmetric positive definite. Then the algorithm defined by* (3.1)–(3.8) *produces a sequence of matrices B_k and vectors x^k which satisfy*

$$\frac{\|[B_k - H_*]Z_k^T(x^{k+1} - x^k)\|}{\|x^{k+1} - x^k\|} \to 0. \tag{3.11}$$

Proof. Initially, assume that $s^k \neq 0$ for all k. Clearly for k sufficiently large, Lemma 3.4 is applicable and therefore the assumptions of Lemma 3.5 are valid. But, for $M^{-2} = H_*$,

$$\|My^k - M^{-1}s^k\| \leqq \|M\| \cdot \|y^k - H_* s^k\|,$$

and using (3.10.1) and $\|x^{k+1} - x^*\| = O\|x^k - x^*\|$ (by assumption),

$$\|y^k - H_* s^k\| \leqq K_4 \|x^k - x^*\| \cdot \|s^k\|$$

for some positive K_4. Therefore, taking $A = M^{-2}$ in Lemma 3.5,

$$\|B_{k+1} - H_*\|_M \leqq [(1-\alpha_0\theta_k^2)^{1/2} + \alpha_1\sigma_k]\|B_k - H_*\|_M + \alpha_2\sigma_k,$$

where $\sigma_k = O\|x^k - x^*\|$,

$$\theta_k = \begin{cases} \dfrac{\|M[B_k - H_*]s^k\|}{\|B_k - H_*\|_M \|M^{-1}s^k\|} & \text{for } B_k \neq A, \\ 0 & \text{otherwise} \end{cases}$$

and $\alpha_0 \in (0, 1]$.

It is clear that Dennis and Moré (1974, Lemma 3.3, Thm. 3.4) can now be directly applied to establish that (3.11) is true if $s^k \neq 0$ for all k. In particular, since $(1-\alpha_0\theta_k^2)^{1/2} \leqq 1$, Dennis and Moré (1974, Lemma 3.3) guarantees the existence of

$$\lim_{k\to\infty} \|H_*^{-1/2} B_k H_*^{-1/2} - I\|_F.$$

The arguments of Dennis and Moré (1974, Thm. 3.4) can be applied without change to yield

$$\frac{\alpha_0}{2} \sum_{k=1}^{\infty} \theta_k^2 \|B_k - H_*\|_M < +\infty.$$

Either $\{\|B_k - H_*\|_M\}$ converges to zero, establishing (3.11) trivially, or $\{\theta_k\}$ converges to zero. In the latter case, we obtain

$$\frac{\|(B_k - H_*)s^k\|}{\|s^k\|} \to 0,$$

which clearly implies (3.11) since $s^k = Z_k^T(x^{k+1} - x^k)$ and $\|s^k\| \leqq \|x^{k+1} - x^k\|$. If $s^k = 0$ and $x^k \neq x^*$, then

$$\frac{\|(B_k - H_*)Z_k^T(x^{k+1} - x^k)\|}{\|x^{k+1} - x^k\|} = 0,$$

and the result is established. □

LEMMA 3.7. *Under the assumptions that* $\sum \|x^k - x^*\| < \infty$, $\|x^{k+1} - x^*\| = O\|x^k - x^*\|$, *and* B_0 *is positive definite, the algorithm given by* (3.1)–(3.8) *produces a sequence of iterates with the property*

$$\|r_L^k\| + \|r_c^k\| = o(\|Z_k^T \nabla f^k\| + \|c^k\|).$$

Proof. By definition,

$$w^k = (C_k^T C_k)^{-1} C_k^T (x^k - x^*),$$

and by Taylor's theorem and $c^* = 0$,

$$c^k = C_k^T(x^k - x^*) + o\|x^k - x^*\|.$$

Considering that $\|(C_k^T C_k)^{-1}\|$ is bounded above, it follows that

$$\|w^k\| = O\|c^k\| + o\|x^k - x^*\|. \tag{3.12}$$

Furthermore, by Taylor's theorem and using $\nabla L^* = 0$,

$$\nabla L^k = -\nabla^2 L^k (x^* - x^k) + o\|x^k - x^*\|,$$

which implies, using (2.3),

$$Z_k^T \nabla f^k = H_k u^k + Z_k^T \nabla^2 L^k C_k w^k + o\|x^k - x^*\|.$$

But $\|H_k^{-1}\|$ and $\|Z_k^T \nabla^2 L^k C_k\|$ are bounded above and therefore,

$$\|u^k\| = O(\|Z_k^T \nabla f^k\| + \|w^k\|) + o\|x^k - x^*\|. \tag{3.13}$$

Combining (3.12) and (3.13) produces, for k sufficiently large,

$$\|x^k - x^*\| = O(\|c^k\| + \|Z_k^T \nabla f^k\|), \tag{3.14}$$

since $\|C_k\|$ is bounded above. By definition,

$$\begin{aligned} r_L^k &= Z_k^T \nabla f^k + H_k Z_k^T (x^{k+1} - x^k) = Z_k^T \nabla f^k + H_* s^k - (H_* - H_k) s^k \\ &= -B_k s^k + H_* s^k + (H_k - H_*) s^k. \end{aligned}$$

Therefore, taking norms,

$$\|r_L^k\| \leqq \|[B_k - H_*] s^k\| + \|H_k - H_*\| \cdot \|s^k\|.$$

But $\|s^k\| = O\|x^k - x^*\|$ and $H_k \to H_*$, which along with Theorem 3.6 and (3.14) gives

$$\|r_L^k\| = o(\|Z_k^T \nabla f^k\| + \|c^k\|). \tag{3.15}$$

Finally, by definition,

$$r_c^k = c^k + C_k^T (x^{k+1} - x^k) = c^k + C_k^T (h^k + v^k).$$

But $C_k^T h^k = 0$ and $v^k = -C_k(C_k^T C_k)^{-1} c(x^k + h^k)$. It is now easy to verify, using Taylor's theorem, that

$$\|r_c^k\| = o\|x^k - x^*\|,$$

which implies, by (3.14),

$$\|r_c^k\| = o(\|Z_k^T \nabla f^k\| + \|c^k\|). \tag{3.16}$$

Clearly, by (3.15) and (3.16), the result is established. □

By Theorem 2.1 we have now established that the sequence x^k converges at a 2-step Q-superlinear rate (assuming $\sum \|x^k - x^*\| < \infty$ and $\|x^{k+1} - x^*\| = O\|x^k - x^*\|$). We state this formally in the following theorem.

THEOREM 3.8. *Under the assumptions that* $\sum \|x^k - x^*\| < \infty$, $\|x^{k+1} - x^*\| = O\|x^k - x^*\|$, *and* B_0 *is symmetric positive definite, algorithm* (3.1)–(3.8) *produces a sequence of iterates* $\{x^k\}$, *with the property*

$$\frac{\|x^{k+1} - x^*\|}{\|x^{k-1} - x^*\|} \to 0.$$

Proof. The result follows immediately from Lemma 3.7 and Theorem 2.1. □

The remaining results are needed to establish the local convergence properties: $\sum \|x^k - x^*\| < \infty$, and $\|x^{k+1} - x^*\| = O\|x^k - x^*\|$. First we establish two useful bounds in Lemmas 3.9 and 3.10.

LEMMA 3.9. *Assume that the smallest eigenvalue of* B_k *is greater than a positive scalar* K_5, *and that* $x^k \in D$. *Then there exists a positive scalar* K_6, *independent of k, such that*

$$\|h^k\| \leqq K_6 \|x^k - x^*\|.$$

Proof. By definition, $h^k = -Z_k B_k^{-1} Z_k^T \nabla L^k$, and since $\nabla L^* = 0$, it follows that

$$\|h^k\| \leqq \|Z_k B_k^{-1} Z_k^T [\nabla L^k - \nabla L^*]\|.$$

But since ∇L is Lipschitz continuous on D, $\|B_k^{-1}\|$ is bounded above, and $\|Z_k\| = 1$, the result follows. □

LEMMA 3.10. *Under the assumptions of Lemma* 3.9, *there exists a positive constant* K_7, *independent of k, such that*

$$\|v^k\| \leqq K_7 \|x^k - x^*\|.$$

Proof. By definition,

$$v^k = -C_k(C_k^T C_k)^{-1} c(x^k + h^k),$$

and since $C_k^T h^k = 0$,

$$c(x^k + h^k) = c^k + o\|h^k\|.$$

Clearly then,

$$\|v^k\| \leqq \|C_k\| \cdot \|(C_k^T C_k)^{-1}\| \cdot \{\|c^k\| + o\|h^k\|\}.$$

By the boundedness of $\|C_k\|$ and $\|(C_k^T C_k)^{-1}\|$, the fact that $c^* = 0$, the Lipschitz continuity of $c(x)$, and Lemma 3.9, the result follows. □

COROLLARY 3.11. *Under the assumptions of Lemma* 3.9, *there exists a positive constant* K_8, *independent of k, such that*

$$\|x^{k+1} - x^*\| \leqq K_8 \|x^k - x^*\|.$$

Proof. The result is an immediate consequence of Lemmas 3.9 and 3.10. □

We are now ready to show, in Lemmas 3.12, 3.13 and Corollary 3.14, that provided two consecutive points are sufficiently close to x^*, a (2-step) contraction is exhibited.

LEMMA 3.12. *Under the assumptions of Lemma 3.9 and provided $\|x^k - x^*\|$ is sufficiently small, there exists a positive constant K_9, independent of k, such that*

$$\|w^{k+1}\| \leqq K_9 \|x^k - x^*\|^2.$$

Proof. By Corollary 3.11 we can assume that $\|x^k - x^*\|$ is sufficiently small so that $x^{k+1} \in D$. It is easy to verify that, using $c^* = 0$ and $C_k^T h^k = 0$,

$$x^{k+1} = x^k - C_k(C_k^T C_k)^{-1} C_k^T (x^k - x^*) + h^k + p^k,$$

where p^k is a vector satisfying $\|p^k\| = O\|x^k - x^*\|^2$. Therefore, $C_k^T(x^{k+1} - x^*) = C_k^T p^k$. But, by definition,

$$w^{k+1} = (C_{k+1}^T C_{k+1})^{-1} C_{k+1}^T (x^{k+1} - x^*),$$

which implies that

$$w^{k+1} = (C_{k+1}^T C_{k+1})^{-1} [C_k^T p^k + (C_{k+1}^T - C_k^T)(x^{k+1} - x^*)].$$

But, $\|(C_{k+1}^T C_{k+1})^{-1}\|$, $\|C_k^T\|$ are bounded above, $\|p^k\|$ is $O\|x^k - x^*\|^2$, $\|x^{k+1} - x^*\|$ is $O\|x^k - x^*\|$ by Corollary 3.11, and $C(x)$ is Lipschitz continuous on D. The result follows immediately. □

LEMMA 3.13. *Provided the smallest eigenvalue of B_{k-1} and B_k is greater than a positive scalar K_5 then there exist positive scalars $\bar{\varepsilon}$ and $\bar{\Delta}$ such that if*

$$\|x^{k-1} - x^*\| \leqq \bar{\varepsilon}, \quad \|x^k - x^*\| \leqq \bar{\varepsilon}, \quad \|B_k^{-1} - H_*^{-1}\|_M \leqq \bar{\Delta},$$

then

$$\|u^{k+1}\| \leqq \tfrac{1}{4} \|x^{k-1} - x^*\|.$$

Proof. Initially choose $\bar{\varepsilon}$ so that $\|x - x^*\| \leqq \bar{\varepsilon}$ implies that x is in D. By Corollary 3.11 we can reduce $\bar{\varepsilon}$, if necessary, so that

$$\|x^{k-1} - x^*\| \leqq \bar{\varepsilon} \Rightarrow x^k \in D \text{ and } x^{k+1} \in D.$$

By (3.1)–(3.8),

$$x^{k+1} = x^k - Z_k H_*^{-1} Z_k^T \nabla L^k + Z_k [H_*^{-1} - B_k^{-1}] Z_k^T \nabla L^k + v^k.$$

However, subtracting x^* from both sides, multiplying by Z_k^T and using Lemma 3.3 yields, for $\bar{\varepsilon}$ sufficiently small,

$$u^{k+1} = A_k w^k + Z_k^T p^k + [H_*^{-1} - B_k^{-1}] Z_k^T \nabla L^k + (Z_{k+1} - Z_k)^T (x^{k+1} - x^*),$$

where $A_k = -H_*^{-1} Z_k^T \nabla^2 L^* C_k$, and p^k is a vector satisfying $\|p^k\| = o\|x^k - x^*\|$. But $\|A_k\|$ is bounded above, $\|Z_k^T \nabla L^k\| = O\|x^k - x^*\|$, and

$$\|Z_{k+1} - Z_k\| = O(\max\{\|x^{k+1} - x^*\|, \|x^k - x^*\|\}).$$

Therefore, by Corollary 3.11 and Lemma 3.12, there exists a positive constant K_{10} such that

$$\|u^{k+1}\| \leqq K_{10}[\|x^{k-1} - x^*\| + \|H_*^{-1} - B_k^{-1}\|] \|x^{k-1} - x^*\|.$$

Therefore, if $\max\{\bar{\varepsilon}, \bar{\Delta}\} \leqq 1/8K_{10}$, then

$$\|u^{k+1}\| \leqq \tfrac{1}{4} \|x^{k-1} - x^*\|,$$

which is the required result. □

COROLLARY 3.14. *Provided the smallest eigenvalue of B_{k-1} and B_k is greater than a positive scalar K_5, there exist positive scalars $\bar{\varepsilon}$ and $\bar{\Delta}$ such that if*

$$\|x^{k-1}-x^*\| \leqq \bar{\varepsilon}, \quad \|x^k-x^*\| \leqq \bar{\varepsilon}, \quad \|B_k^{-1}-H_*^{-1}\|_M \leqq \bar{\Delta},$$

then

$$\|x^{k+1}-x^*\| \leqq \tfrac{1}{2}\|x^{k-1}-x^*\|.$$

Proof. Initially let $\bar{\varepsilon}$ and $\bar{\Delta}$ be as defined in Lemma 3.13. Lemma 3.12 and the boundedness of $\|C_k\|$ allow $\bar{\varepsilon}$ to be further restricted, if necessary, until

$$\|C_{k+1}\| \cdot \|w^{k+1}\| \leqq \frac{1}{4K_8}\|x^k-x^*\|. \tag{3.17}$$

Combining (3.17) with Lemma 3.13 and Corollary 3.11 produces the desired inequality. □

Borrowing heavily from Broyden, Dennis and Moré (1973), we now establish the local convergence property.

THEOREM 3.15. *Suppose that the sequence $\{x^k, B_k\}$ is generated by algorithm (3.1)–(3.8) with starting pair $\{x^0, B_0\}$, and with the matrix B_0 being symmetric positive definite. Then there exist positive scalars ε_0 and Δ such that if $\|x^0-x^*\|<\varepsilon_0$, and $\|B_0-H_*\|_M<\Delta$, then*

$$\sum \|x^k-x^*\| < \infty.$$

Proof. Choose positive scalars ε_0 and Δ so that $\varepsilon_0 \leqq \bar{\varepsilon}$ and $\Delta \leqq \bar{\Delta}$, where $\bar{\varepsilon}$ and $\bar{\Delta}$ are as defined in the statement of Corollary 3.14. Further restrict Δ, if necessary, so that

$$2\rho\Delta\gamma \leqq \tfrac{1}{2}, \tag{3.18}$$

where for any matrix A, $\|A\| \leqq \rho\|A\|_M$, and $\gamma = \|H_*^{-1}\|$. But, by hypothesis, $\|B_0-H_*\| < \rho\Delta < 2\rho\Delta$, and by (3.18), the Banach perturbation lemma (Ortega and Rheinboldt (1970, p. 45)) can be applied to give

$$\|B_0^{-1}\| \leqq \frac{\gamma}{1-(2\rho\Delta\gamma)} \leqq 2\gamma. \tag{3.19}$$

Since $\|B_0^{-1}\|$ is bounded, Corollary 3.11 can be used, for $k=0$, to give

$$\|x^1-x^*\| = O(\varepsilon_0). \tag{3.20}$$

Let $\varepsilon_1 = \|x^1-x^*\|$ and set $\varepsilon = \max\{\varepsilon_0, \varepsilon_1\}$. Further restrict ε_0, if necessary, so that $\varepsilon \leqq \min\{\bar{\varepsilon}, \hat{\varepsilon}\}$. ($\hat{\varepsilon}$ is defined in Lemma 3.4.) If $s^0=0$ then $B_1=B_0$ and (3.21) is trivially true. Otherwise, the assumptions of Lemma 3.5 are valid here and

$$\|B_1-H_*\|_M - \|B_0-H_*\|_M \leqq \alpha_1 \frac{\|My^0-M^{-1}s^0\|}{\|M^{-1}s^0\|} \cdot 2\Delta + \alpha_2 \frac{\|y^0-H_*s^0\|}{\|M^{-1}s^0\|}, \tag{3.20.1}$$

where $M^2 = H_*^{-1}$. But

$$\|My^0-M^{-1}s^0\| \leqq \|M\| \cdot \|y^0-H_*s^0\|, \quad \text{and} \quad \|M^{-1}s^0\| \geqq \frac{\|s^0\|}{\|M\|}.$$

Hence, if we define

$$\alpha_3 = \alpha_1\|M\|^2[2(K_1+K_2)+K_3], \quad \text{and} \quad \alpha_4 = \alpha_2\|M\|[2(K_1+K_2)+K_3],$$

then (3.10.1) and (3.20.1) imply

$$\|B_1-H_*\|_M - \|B_0-H_*\|_M \leqq (2\alpha_3\Delta+\alpha_4)\varepsilon. \tag{3.21}$$

Further restrict ε_0, if necessary, so that

$$4(2\alpha_3\Delta+\alpha_4)\varepsilon \leqq \Delta, \tag{3.22}$$

which implies by (3.21) that

$$\|B_1-H_*\|_M \leqq 2\Delta. \tag{3.23}$$

Clearly, by (3.18) and (3.23) the Banach perturbation lemma can be applied again, to give

$$\|B_1^{-1}\| \leqq 2\gamma. \tag{3.24}$$

Now considering Corollary 3.14, we obtain

$$\|x^2-x^*\| \leqq \tfrac{1}{2}\|x^0-x^*\|. \tag{3.25}$$

We complete the proof with an induction step. Assume that

$$\|B_k-H_*\|_M \leqq 2\Delta, \quad \|B_k^{-1}\| \leqq 2\gamma, \quad \|x^{k+1}-x^*\| \leqq \tfrac{1}{2}\|x^{k-1}-x^*\| \quad \text{for } k=1,\cdots,m-1.$$

Clearly, for each k either Lemma 3.5 is applicable or $s^k=0$. In either case we obtain

$$\|B_{k+1}-H_*\|_M - \|B_k-H_*\|_M \leqq (2\alpha_3\Delta+\alpha_4)\cdot\varepsilon\cdot(\tfrac{1}{2})^{\lfloor k/2\rfloor}, \tag{3.26}$$

where $\lfloor p\rfloor$ represents the largest integer less than or equal to p. Therefore, summing both sides of (3.26) from $k=0$ to $k=m-1$ yields

$$\|B_m-H_*\|_M \leqq \|B_0-H_*\|_M+(2\alpha_3\Delta+\alpha_4)\cdot\varepsilon\cdot 4, \tag{3.27}$$

which, by (3.22) gives $\|B_m-H_*\|_M \leqq 2\Delta$. Therefore the Banach perturbation lemma will again give $\|B_m^{-1}\| \leqq 2\gamma$, and Corollary 3.11 will guarantee that $\|x^{m+1}-x^*\| \leqq \frac{1}{2}\|x^{m-1}-x^*\|$. It follows that $\sum \|x^k-x^*\| < \infty$. □

Theorem 3.8, Corollary 3.14 and Theorem 3.15 imply that (3.1)–(3.8) generates x-values which converge to x^* at a 2-step Q-superlinear rate. We state this formally in the following theorem.

THEOREM 3.16. *Suppose that the sequence* $\{x^k, B_k\}$ *is generated by algorithm* (3.1)–(3.8) *with starting pair* $\{x^0, B_0\}$, *and with the matrix* B_0 *being symmetric positive definite. Then, there exist positive scalars* ε_0 *and* Δ *such that if* $\|x^0-x^*\| \leqq \varepsilon_0$, *and* $\|B_0-H_*\|_M \leqq \Delta$, *then* $\{x^k\}$ *converges to* x^* *and does so at a* 2-*step* Q-*superlinear rate.*

Proof. The result follows immediately from Theorem 3.8, Corollary 3.11 and Theorem 3.15. □

4. Conclusions. We have proposed an adaptation of the DFP/BFGS formula to the nonlinearly constrained problem. The central feature of our approach is that a positive definite approximation to a *projected* Hessian is maintained. We have established, without assuming convexity, that the method is locally 2-step Q-superlinearly convergent. The performance of this method in practice is unknown and will be the subject of future work. A detailed discussion of implementation techniques is also postponed: we only remark that the conditions placed on $Z(x)$ can be realized in practise by using a careful implementation of the QR decomposition—details are given in Coleman and Sorensen (1982). (Of course the projected quasi-Newton step depends on the null space but not Z itself.)

For the inequality constrained problem, it is clear that once the active solution set (of constraints) is identified, either implicitly or explicitly, the results given here are directly applicable. However, the best way to modify projected approximations when the active set is changing is not presently known. Another subject of future work is how to adapt a line search algorithm and generally globalize the local procedure given here. In particular, it is not clear how to ensure $y_k^T s_k > 0$ (a necessary condition for the DFP/BFGS update) when x_k is not in D.

REFERENCES

P. T. BOGGS, J. W. TOLLE AND P. WANG (1982), *On the local convergence of quasi-Newton methods for constrained optimization*, SIAM J. Control and Optim., 20, pp. 161–171.

C. G. BROYDEN, J. E. DENNIS AND J. J. MORÉ (1973), *On the local and superlinear convergence of quasi-Newton methods*, J. Inst. Math. Appl., 12, pp. 223–245.

T. F. COLEMAN AND A. R. CONN (1982a), *Nonlinear programming via an exact penalty function: Asymptotic analysis*, Math. Programming, 24, pp. 123–136.

——— (1982b), *Nonlinear programming via an exact penalty function: Global analysis*, Math. Programming, 24, pp. 137–161.

T. F. COLEMAN AND D. C. SORENSEN (1982), *A note on the computation of an orthonormal basis for the null space of a matrix*, Tech. Rept. TR 82-510, Computer Science Dept., Cornell Univ., Ithaca, NY.

R. S. DEMBO, S. C. EISENSTAT AND T. STEIHAUG (1982), *Inexact Newton methods*, this Journal, 19, pp. 400–408.

J. E. DENNIS AND J. J. MORÉ (1974), *A characterization of superlinear convergence and its application to quasi-Newton methods*, Math. Comp., 28, pp. 549–560.

P. GILL AND W. MURRAY (1974), *Numerical Methods for Constrained Optimization*, Academic Press, London.

S. T. GLAD (1979), *Properties of updating methods for the multipliers in augmented Lagrangians*, J. Optim. Theory Appl., 28, pp. 135–156.

S.-P. HAN (1976), *Superlinearly convergent variable metric algorithms for general nonlinear programming problems*, Math. Programming, 11, pp. 263–282.

——— (1977), *Dual variable metric algorithms for constrained optimization*, SIAM J. Control Optim., 15, pp. 546–564.

J. M. ORTEGA AND W. C. RHEINBOLDT (1970), *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York.

J. M. ORTEGA (1972), *Numerical Analysis: A Second Course*, Academic Press, New York.

M. J. D. POWELL (1978), *The convergence of variable metric methods for nonlinearly constrained optimization calculations*, in O. L. Mangasarian, Nonlinear Programming 3, R. R. Meyer and S. M. Robinson, eds., Academic Press, New York, pp. 27–63.

K. TANABE (1981), *Feasibility-improving gradient-acute-projection methods: a unified approach to nonlinear programming*, Lecture Notes in Numerical and Applied Analysis 3, Springer-Verlag, New York, pp. 57–76.

R. A. TAPIA (1977), *Diagonalized multiplier methods and quasi-Newton methods for constrained optimization*, J. Optim. Theory Appl., 22, pp. 135–194.