Appl. Math. Lett. Vol. 6, No. 3, pp. 37–42, 1993
Printed in Great Britain.
0893-9659/93 $6.00 + 0.00

THE LOCAL CONVERGENCE OF THE BYRD-SCHNABEL ALGORITHM FOR CONSTRAINED OPTIMIZATION

THOMAS F. COLEMAN
Computer Science Department, Cornell University
Ithaca, NY 14853, U.S.A.
AI-PING LIAO
School of Operations Research and Industrial Engineering, Cornell University
Ithaca, NY 14853, U.S.A.

(Received January 1993; accepted February 1993)

Abstract—Most reduced Hessian methods for equality constrained problems use a basis for the null space of the matrix of constraint gradients and possess superlinearly convergent rates under the assumption of continuity of the basis. However, computing a continuously varying null space basis is not straightforward. Byrd and Schnabel [1] propose an alternative implementation that is independent of the choice of null space basis, thus obviating the need for a continuously varying null space basis. In this note, we prove that the primary sequence of iterates generated by one version of their algorithm exhibits a local 2-step Q-superlinear convergence rate. We also establish that a sequence of "midpoints," in a closely related algorithm, is (1-step) Q-superlinearly convergent.

1. INTRODUCTION

The reduced Hessian methods for equality constrained optimization problems usually use a basis for the null space of the matrix of constraint gradients. Consider the problem

$$\begin{aligned} &\min f(x) \\ &\text{subject to } c(x) = 0, \end{aligned} \tag{1}$$

where $f : R^n \to R$ and $c : R^n \to R^t$ are smooth nonlinear functions. Suppose that $A(x)$ is the $n \times t$ matrix whose columns are the gradients of the constraint functions $c(x)$. We assume that $A(x)$ is of full column rank. Let $Z(x)$ be an orthonormal basis for the null space of $A(x)^T$; hence $Z(x)$ is an $n \times (n-t)$ full rank matrix satisfying $A(x)^T Z(x) = 0$. If $L(x, \lambda) = f(x) - c(x)^T \lambda$ is the Lagrangian for problem (1), then the reduced Hessian matrix can be expressed as $Z(x)^T \nabla_x^2 L(x, \lambda)\, Z(x)$. The reduced Hessian is dependent on the choice of null space basis $Z(x)$. Many reduced Hessian algorithms, e.g., Coleman and Conn [2], Nocedal and Overton [3], assume continuity of $Z(x)$. But, as pointed out by Coleman and Sorensen [4], the standard implementation of the QR factorization of $A(x)$ via Householder matrices does not necessarily yield a matrix $Z(x)$ with continuously varying elements. Coleman and Sorensen [4] propose factorization schemes which guarantee local continuity. In contrast, Byrd and Schnabel [1] propose an algorithm which is independent of the choice of the null space basis. In Section 2, we present the Byrd-Schnabel algorithm, and in Section 3, we prove that their algorithm is locally 2-step Q-superlinearly convergent.

2. THE BYRD-SCHNABEL ALGORITHM

In this section, we describe the Byrd-Schnabel algorithm.

ALGORITHM.

0. Choose an initial invertible matrix B_0 with the form $Z_0^T Q Z_0$, where Q is a symmetric matrix and Z_0 is a basis for the null space of $A(x_0)^T$ and an initial point x_0; let $k = 0$.

1. Compute

$$d_k = h_k + v_k, \tag{2}$$

where

$$h_k = -Z_k B_k^{-1} Z_k^T \nabla f(x_k),$$
$$v_k = -A_k (A_k^T A_k)^{-1} c(x_k).$$

Set $x_{k+1} := x_k + d_k$.

2. Compute Z_{k+1}, $T_k := Z_k^T Z_{k+1}$ and β_k.

3. Let

$$\bar{B}_k = T_k^T (B_k - \beta_k I) T_k + \beta_k I.$$

4. Compute

$$s_k := Z_{k+1}^T (x_{k+1} - x_k), \tag{3}$$
$$y_k := Z_{k+1}^T [\nabla_x L(x_{k+1}, \lambda_{k+1}) - \nabla_x L(x_{k+1} - Z_{k+1} Z_{k+1}^T d_k, \lambda_{k+1})], \tag{4}$$

where

$$\lambda_{k+1} = (A_{k+1}^T A_{k+1})^{-1} A_{k+1}^T \nabla f(x_{k+1}). \tag{5}$$

Update $\bar{B}_k$ using the DFP or BFGS update[1], $B_{k+1} = \mathcal{U}(\bar{B}_k; s_k, y_k)$, with secant equation $B_{k+1} s_k = y_k$.

5. Set k to $k+1$ and go to Step 1. ∎

We note that d_k is the solution to

$$\begin{array}{l} \min \nabla f(x_k)^T d + \frac{1}{2} d^T Z_k B_k Z_k^T d \\ \text{subject to } c(x_k) + A(x_k)^T d = 0. \end{array} \tag{6}$$

The scaling factor β_k can be regarded as an approximation to $\|\nabla_x^2 L(x_k, \lambda_k)\|$; for example, one can take $\beta_k = \|B_k\|$ (see Byrd and Schnabel [1]). Here we just assume that $\{\beta_k\}$ is bounded.

The algorithm we have described above is actually a member of the set of algorithms (or implementations) proposed by Byrd and Schnabel. In this set, Byrd and Schnabel allow for a variety of choices for s_k and y_k. We note that Byrd and Schnabel [1] do not give any convergence result for any member of their set of algorithms. In the next section, we prove that the algorithm described above, which we call the "Byrd-Schnabel algorithm," is locally 2-step Q-superlinearly convergent.

Next we note that if β_k is restricted to be positive, the update formula in this algorithm preserves positive definiteness.

THEOREM 1. *If B_k is positive definite and $y_k^T s_k > 0$, $\beta_k > 0$, then B_{k+1} is also positive definite.*

PROOF. The proof is straightforward: see Coleman and Liao [6]. ∎

We will show below that if we only assume that $\{\beta_k\}$ is bounded, then the update will preserve positive definiteness locally.

3. SUPERLINEAR CONVERGENCE OF THE BYRD-SCHNABEL ALGORITHM

In this section, we discuss the local properties of the Byrd-Schnabel algorithm. We assume that there is an open convex region, say D, containing a point x_* and the following statements hold:

[1] See, for example, Dennis and Schnabel [5].

A1: x_* is a local minimizer of problem (1).

A2: The functions f and c are twice continuously differentiable in a neighborhood of x_*.

A3: $A_* := A(x_*)$ is of full column rank t.

A4: $\nabla_x^2 L(x_*, \lambda_*)$ is positive definite on the null space of A_*^T, null(A_*^T).

Since the Byrd-Schnabel algorithm is independent of the choice of Z_k, we can assume that $Z_k = Z(x_k)$ in D where $Z(x)$ is a continuous differentiable function on D. We assume that $Z(x)$, $\nabla^2 f(x)$ and $\nabla^2 c(x)$ are Lipschitz continuous functions of x in D. We make extensive use of the "O" notation, where $\phi_k = O(\psi_k)$ means that the ratio ϕ_k/ψ_k remains bounded as k tends toward infinity. Coleman and Conn [2] prove the following result.

THEOREM 2. *If $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded, then $\|x_{k+1} - x_*\| = O(\|x_k - x_*\|)$ and there exist positive scalars K_0 and K_1, such that*
(i) $\|\lambda_k - \lambda_*\| \le K_0 \|x_k - x_*\|$,
(ii) $\|Z_k^T \nabla_x^2 L(x_k, \lambda_k) Z_k - H_*\| \le K_1 \|x_k - x_*\|$,
where λ_k is defined by (5). If, in addition, $x_k \longrightarrow x_*$ and

$$\frac{\|(B_k - H_*) Z_{k+1}^T (x_{k+1} - x_k)\|}{\|d_k\|} \longrightarrow 0, \tag{7}$$

where $d_k = x_{k+1} - x_k$ and $H_* := Z_*^T \nabla_x^2 L(x_*, \lambda_*) Z_*$, then $x_k \longrightarrow x_*$ 2-step superlinearly. ∎

LEMMA 3. *Assuming that $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded, s_k is given by (3) and y_k is given by (4), and there exists a positive scalar ε such that if $\|x_k - x_*\| \le \varepsilon$, then*

$$\|M y_k - M^{-1} s_k\| \le \frac{1}{3} \|M^{-1} s_k\|,$$

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

PROOF. First, we note that

$$\|M y_k - M^{-1} s_k\| \le \|M\| \cdot \|y_k - H_* s_k\|. \tag{8}$$

By Taylor's theorem,

$$\begin{aligned} \nabla_x L(x_{k+1}, \lambda_{k+1}) &- \nabla_x L(x_{k+1} - Z_{k+1} Z_{k+1}^T d_k, \lambda_{k+1}) \\ &= \nabla_x^2 L(x_{k+1}, \lambda_{k+1}) Z_{k+1} Z_{k+1}^T d_k + E_k Z_{k+1} Z_{k+1}^T d_k, \end{aligned} \tag{9}$$

where

$$\|E_k\| = O(\|Z_{k+1} Z_{k+1}^T d_k\|) = O(\|x_{k+1} - x_k\|) = O(\max\{\|x_{k+1} - x_*\|, \|x_k - x_*\|\}).$$

So

$$\begin{aligned} y_k &= Z_{k+1}^T (\nabla_x L(x_{k+1}, \lambda_{k+1}) - \nabla_x L(x_{k+1} - Z_{k+1} Z_{k+1}^T d_k, \lambda_{k+1})) \\ &= Z_{k+1}^T \nabla_x^2 L(x_{k+1}, \lambda_{k+1}) Z_{k+1} Z_{k+1}^T d_k + Z_{k+1}^T E_k Z_{k+1} Z_{k+1}^T d_k. \end{aligned}$$

Thus, by Theorem 2 and provided ε is sufficiently small,

$$\|y_k - H_* s_k\| \le (\|Z_{k+1}^T \nabla_x^2 L(x_{k+1}, \lambda_{k+1}) Z_{k+1} - H_*\| + \|Z_{k+1}^T E_k Z_{k+1}\|) \|s_k\| \tag{10}$$

$$\le (K_0 + K_1)\, O(\max\{\|x_{k+1} - x_*\|, \|x_k - x_*\|\}) \|s_k\|. \tag{11}$$

Hence, it follows that for ε sufficiently small,

$$\|y_k - H_* s_k\| \le \frac{\|s_k\|}{3\|M\|^2},$$

which implies, by (8)

$$\|M y_k - M^{-1} s_k\| \le \frac{1}{3} \|M^{-1} s_k\|.$$ ∎

LEMMA 4. *If $\|My_k - M^{-1}s_k\| \le \frac{1}{3}\|M^{-1}s_k\|$ with $s_k \neq 0$, then $y_k^T s_k > 0$ and thus, B_{k+1} is well-defined in this algorithm. Moreover, there are positive constants α_0, α_1 and α_2 such that*

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

where $\alpha_0 \in (0,1], \sigma_k := \max\{\|x_{k+1} - x_*\|, \|x_k - x_*\|\}$, and

$$\theta_k := \begin{cases} \dfrac{\|M[\bar{B}_k - H_*]s_k\|}{\|\bar{B}_k - H_*\|_M\|M^{-1}s_k\|} & \text{for } \bar{B}_k \neq H_*, \\ 0 & \text{otherwise.} \end{cases}$$

PROOF. We first note that

$$\|T_k - I\| = \|Z_k^T Z_{k+1} - Z_k^T Z_k\| = \|Z_k^T(Z_{k+1} - Z_k)\| = O(\sigma_k).$$

Thus,

$$\begin{aligned} \|\bar{B}_k - B_k\| &= \|T_k^T B_k T_k - B_k - \beta_k(T_k^T T_k - I)\| \le \|T_k^T B_k T_k - B_k\| + |\beta_k|\,\|T_k^T T_k - I\| \\ &= \|T_k^T B_k T_k - T_k^T B_k + T_k^T B_k - B_k\| + |\beta_k|\,\|T_k^T T_k - T_k + T_k - I\| \\ &\le (\|T_k^T B_k\| + \|B_k\|)\|T_k - I\| + |\beta_k|(\|T_k^T\| + 1)\|T_k - I\| \\ &= O(\sigma_k) + O(\sigma_k) = O(\sigma_k). \text{ (Since } \{\beta_k\} \text{ is bounded.)} \end{aligned} \tag{12}$$

This implies

$$\|\bar{B}_k - H_*\| \le \|B_k - H_*\| + O(\sigma_k). \tag{13}$$

Noting (11), this lemma thus follows from Lemma 3.1 of Dennis and Moré [7]. ∎

THEOREM 5. *Assume that $\sum\|x_k - x_*\| < \infty$, $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded. Then we have*

$$\frac{\|[B_k - H_*]s_k\|}{\|x_{k+1} - x_*\|} \longrightarrow 0.$$

PROOF. The argument is standard and derives from Dennis and Moré [7]. See Coleman and Liao [6] for details. ∎

From Theorem 2, we now need to show that $\sum\|x_k - x_*\| < \infty$ and $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded. The following lemma is Corollary 3.14 of Coleman and Conn [2].

LEMMA 6. *Provided the smallest eigenvalue of B_{k-1} and B_k is greater than a positive scalar K, there exist positive scalars ε and δ such that if*

$$\|x_{k-1} - x_*\| \le \varepsilon, \quad \|x_k - x_*\| \le \varepsilon, \quad \|B_k^{-1} - H_*^{-1}\|_M \le \delta,$$

then

$$\|x_{k+1} - x_*\| \le \frac{1}{2}\|x_{k-1} - x_*\|.$$ ∎

With the above lemma and Lemma 4, using the same technique employed in [2,8], we thus have the following result.

THEOREM 7. *Suppose that the sequence $\{x_k, B_k\}$ is generated by the algorithm with the initial quantities x_0, B_0, where B_0 is symmetric positive definite, and $\{\beta_k\}$ is bounded. Then there exist positive scalars ε and δ such that if*

$$\|x_0 - x_*\| \le \varepsilon, \quad \text{and} \quad \|B_0 - H_*\|_M \le \delta,$$

then $\|B_k - H_*\| \le 2\delta$, for $k = 0, 1, \ldots$, and

$$\sum\|x_k - x_*\| < \infty.$$ ∎

THEOREM 8. *Suppose that the sequence $\{x_k, B_k\}$ is generated by the algorithm with the initial quantities x_0, B_0, where $B_0 = Z_0^T Q Z_0$ and Q is a symmetric matrix, and $\{\beta_k\}$ is bounded. Then there exist positive scalars ε and δ such that if*

$$\|x_0 - x_*\| \le \varepsilon, \quad \text{and} \quad \|Q - \nabla_x^2 L(x_*, \lambda_*)\| \le \delta,$$

then $\|B_k - H_*\| \le 2\delta$, for $k = 0, 1, \ldots$, and $\{x_k\}$ converges to x_* at a 2-step Q-superlinear rate.

PROOF. Redefine ε if necessary so that

$$\|x_0 - x_*\| \le \varepsilon, \quad \text{and} \quad \|Q - \nabla_x^2 L(x_*, \lambda_*)\| \le \delta,$$

imply $\|B_0 - H_*\|_M \le \delta$. The result follows immediately from Theorems 2, 5 and 7. ∎

As a consequence of Theorem 8, we can further restrict ε and δ, if necessary, so that $(T_k x)^T B_k (T_k x) \ge \mu \|x\|^2$, for some $\mu > 0$, and $\|T_k x\| > (1 - \mu\kappa^{-1})^{1/2} \|x\|$, for all $k = 0, 1, \ldots,$ and $x \in R^n$, where $|\beta_k| \le \kappa$ and we can assume that $\kappa > 1$. Thus,

$$\begin{aligned} x^T \bar{B}_k x &= (T_k x)^T B_k (T_k x) + \beta_k (x^T x - (T_k x)^T (T_k x)) \\ &> (T_k x)^T B_k (T_k x) - \kappa (\|x\|^2 - (1 - \mu\kappa^{-1}) \|x\|^2) \\ &\ge \mu \|x\|^2 - \mu \|x\|^2 = 0. \end{aligned}$$

Therefore, if we assume that $\{\beta_k\}$ is bounded, then the update preserves positive definiteness locally.

4. CONCLUDING REMARKS

We note that Byrd and Schnabel [1] also suggest that one can take $d_k = h_k + \bar{v}_k$, where $\bar{v}_k = -A_k (A_k^T A_k)^{-1} c(x_k + h_k)$. Since $\|v_k - \bar{v}_k\| \le O(\|h_k\|^2)$, for this choice of d_k, by further restricting ε, if necessary, Lemma 3 holds and so do Lemma 4 and Theorem 5. Noting that Lemma 6 is valid for this choice of d_k (see Coleman and Conn [2]), Theorems 7 and 8 follow. Therefore, the algorithm is still 2-step Q-superlinearly convergent. Moreover, by Theorem 2.5 of Byrd [9], the sequence $\{x_k + h_k\}$ is (1-step) Q-superlinearly convergent.

Our result applies to our particular choices of s_k and y_k. However, other choices are also possible. For example, we can choose $s_k := Z_k^T (x_{k+1} - x_k)$ and $y_k := Z_k^T [\nabla_x L(x_k + h_k, \lambda_k) - \nabla_x L(x_k, \lambda_k)]$ as suggested by Coleman and Conn [2], and it is easy to prove that all the above results are also valid for this modification (provided the algorithm is changed by putting Step 4 before Step 2).

Finally, we note that Coleman [10] suggests a slight generalization of the Byrd-Schnabel algorithm: in Step 3 let

$$\bar{B}_k = T_k^T (B_k - C_k) T_k + C_k, \tag{14}$$

where C_k is symmetric but otherwise arbitrary. It is easy to show that, if $\{C_k\}$ is bounded, i.e., $\|C_k\| \le \kappa_c$, $k = 1, 2, \ldots$, for some $\kappa_c > 0$, then the algorithm is still locally 2-step Q-superlinearly convergent.

REFERENCES

1. R.H. Byrd and R.B. Schnabel, Continuity of the null space basis and constrained optimization, *Mathematical Programming* **35**, 32–41 (1986).
2. T.F. Coleman and A.R. Conn, On the local convergence of a quasi-Newton method for the nonlinear programming problem, *SIAM J. Numer. Anal.* **21**, 755–769 (1984).
3. J. Nocedal and M. Overton, Projected Hessian updating algorithms for nonlinearly constrained optimization, *SIAM J. Numer. Anal.* **22**, 821–850 (1985).
4. T.F. Coleman and D.C. Sorensen, A note on the computation of an orthonormal basis for the null space of a matrix, *Mathematical Programming* **29**, 234–242 (1984).
5. J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Inc., (1983).
6. T.F. Coleman and A.-P. Liao, On the local convergence of the Byrd-Schnabel algorithm for constrained optimization, Tech. Rep. 92-1, Cornell University, Ithaca, NY, (1992).

7. J.E. Dennis and J.J. Moré, A characterization of superlinear convergence and its application to quasi-Newton methods, *Math. Comp.* **28**, 549–560 (1974).
8. C.G. Broyden, J.E. Dennis and J.J. Moré, On the local and superlinear convergence of quasi-Newton methods, *J. Inst. Math. Appl.* **12**, 223–245 (1973).
9. R.H. Byrd, On the convergence of constrained optimization methods with accurate Hessian information on a subspace, *SIAM J. Numer. Anal.* **27**, 141–153 (1990).
10. T.F. Coleman, On characterizations of superlinear convergence for constrained optimization, In *Computational Solution of Nonlinear Systems of Equations*, (Edited by E.L. Allgower and K. Georg), pp. 113–133, Providence, RI, (1990).