LOCAL CONVERGENCE OF THE SYMMETRIC RANK-ONE ITERATION*

C.T. KELLEY† AND E.W. SACHS‡

Abstract. We consider conditions under which the SR1 iteration is locally convergent. We apply the result to a pointwise structured SR1 method that has been used in optimal control.

Key words. SR1 update, Pointwise quasi-Newton method, optimal control

AMS(MOS) subject classifications. 47H17, 49K15, 65F10, 65H10, 49M15, 65J15, 65K10

1. Introduction.

The symmetric rank-one (SR1) update [1] is a quasi-Newton method that preserves symmetry of an approximate Hessian (optimization problems) or Jacobian (nonlinear equations). The analysis in this paper is from the nonlinear equations point of view. Our purpose is to prove a local convergence result using the concept of uniform linear independence from [5], extend that result to structured updates where part of the Jacobian can be computed exactly, and then apply those results to the pointwise SR1 update considered in [14] in the context of optimal control.

We we begin with a nonlinear equation

$$F(x) = 0 \tag{1.1}$$

in R^N. We make the standard assumptions in nonlinear equations.

ASSUMPTION 1.1. *F has a root x^*. There is $\delta > 0$ such that the Jacobian $F'(x)$ exists and is Lipschitz continuous in the set*

$$\mathcal{B}(\delta) = \{x \mid \|x - x^*\| < \delta\}$$

with Lipschitz constant γ. $F'(x^*)$ is nonsingular.

Later in this paper we will assume that $F'(x)$ is symmetric near x^* and use the SR1 update to maintain symmetric approximations to $F'(x^*)$.

From current approximations (x_c, B_c) to $(x^*, F'(x^*))$ the SR1 iteration computes a new point x_+ by computing a search direction,

$$B_c d = -F(x_c), \tag{1.2}$$

and updating x_c,

$$x_+ = x_c + \lambda d, \tag{1.3}$$

* Version of May 17, 1995.

† North Carolina State University, Department of Mathematics and Center for Research in Scientific Computation, Box 8205, Raleigh, N. C. 27695-8205 (Tim_Kelley@ncsu.edu). The research of this author was supported by National Science Foundation grant #DMS-9321938 and North Atlantic Treaty Organization grant #CRG 920067. Computing was partially supported by an allocation of time from the North Carolina Supercomputing Center.

‡ Universität Trier, FB IV – Mathematik and Graduiertenkolleg Mathematische Optimierung, 54296 Trier, Germany (sachs@uni-trier.de). The research of this author was supported by North Atlantic Treaty Organization grant #CRG 920067.

where $0 < \lambda \leq 1$ is a steplength parameter. B_c is updated to form B_+ by setting

$$y = F(x_+) - F(x_c), s = x_+ - x_c,$$

and, if $(y - B_c s)^T s \neq 0$, updating B_c to obtain

$$B_+ = B_c + \frac{(y - B_c s)(y - B_c s)^T}{(y - B_c s)^T s}. \tag{1.4}$$

If $(y - B_c s)^T s = 0$, the update for B is skipped, so $B_+ = B_c$.

Observations have been reported [15], [4], [5], [19], [10], [18], [11], [20], [14], that indicate that SR1 can outperform BFGS in the context of optimization, where either the approximate Hessians can be expected to be positive definite or a trust-region framework is used [3], [4], [5].

One modification of SR1 that allows for a convergence analysis has been proposed in [17].

The SR1 update was considered in [2] where it was shown that the local superliner convergence theory in that paper did not apply, because the updates could become undefined.

Preservation of symmetry is an attractive feature, as is the possibility of storing one vector per iterate in a matrix-free limited memory implementation. However, implementations that store a single vector for each iteration are known for Broyden's method [8], [13], and the BFGS method [21], which have much better understood convergence properties. The advantage of the SR1 method over others is in a reduction in the number of iterations. Such a reduction has been observed by many authors as we indicated above.

As is standard, we update the approximate Jacobian only if

$$|s^T(y - B_c s)| \geq \sigma \|s\| \|y - B_c s\| \tag{1.5}$$

for some $\sigma \in (0, 1)$ fixed. In many recent papers [5], [15], [3], on SR1, (1.5), with an arbitrary choice of σ, is one of the assumptions used to prove convergence of B_n to $F'(x^*)$. The numerical results presented in [5], [15], and [3] use very small values of σ. In this paper, as was done in [14], we use a larger value of σ than in other treatments of the SR1 iteration as a way to improve stability. The estimates in § 2 can be used as a rough guide in selection of an appropriate value of σ. The numerical results in § 5 illustrate the benefits of varying σ.

In this paper we show that if the initial approximations to the solution and Jacobian are sufficiently good and if the sequence of steps that satisfy (1.5) also satisfy a uniform linear independence condition [5], then the iteration will be locally linearly convergent and the sequence $\{B_k\}$ will remain near to $F'(x^*)$. A stronger uniform linear independence condition implies k-step superlinear convergence for some integer k. Thus, our goal is different from that in [5], [15], and [3], where convergence of the iteration to the solution was an assumption and conditions were given that guaranteed convergence of the approximate Jacobians to $F'(x^*)$.

While the uniform linear independence condition may seem strong, it is a reasonable condition for very small problems. Such problems arise as part of the pointwise SR1 method proposed in [14] for certain optimal control problems and the results in this paper give insight that we use to improve the performance of the pointwise SR1 method. In § 2 we state and prove our basic convergence results. We consider structured updates in § 3 and the application to pointwise updates in § 4. Finally we report numerical results in § 5.

2. Basic Lemmas and Convergence Results. As we deal with local convergence only in this paper, we take full steps ($\lambda = 1$) and use the fact that

$$y - B_c s_c = F(x_+) \tag{2.1}$$

which is a direct consequence of the definition $y = F(x_+) - F(x_c)$, $\lambda = 1$, and the equation for the quasi-Newton step $B_c s_c = -F(x_c)$. Hence the SR1 update can be written

$$B_+ = B_c + \frac{F(x_+)F(x_+)^T}{s_c^T F(x_+)^T}. \tag{2.2}$$

We use the notation

$$E = B - F'(x^*) \text{ and } e = x - x^* \tag{2.3}$$

for errors in Jacobian and solution approximations.

It may happen that many iterations take place between updates of B. We must introduce notation to keep track of those iterations that result in a change in B. We say that s_n is a **major step**, x_{n+1} is a **major iteration**, and B_{n+1} a **major update** if $s = s_n$ and $y - B_n s_n = F(x_{n+1})$ satisfy (1.5) with $\sigma \in (0, 1)$. In this case $B_{n+1} \neq B_n$. A step is a **minor step** if it is not a major step (and hence no update of B takes place and $B_{n+1} = B_n$). A local convergence theory must show that the new approximation B_{n+1} to the Jacobian at the solution is nonsingular. We do this by proving an analog of the "bounded deterioration" results used in the study of other quasi-Newton methods. However, an inherent instability must be kept in check and this is where the uniform linear independence becomes important.

2.1. Stability. We base our main result on several lemmas. The first simply summarizes some well known results in nonlinear equations [6], [13], [16], and has nothing to do with an assumption of symmetry of $F'(x^*)$ or any particular quasi-Newton method.

LEMMA 2.1. *Let Assumption 1.1 hold and let $\rho \in (0, 1)$ be given. Then there are ϵ_0 and δ_0 such that if x_c and B_c satisfy*

$$\|e_c\| \leq \epsilon_0 \textit{ and } \|E_c\| \leq \delta_0 \tag{2.4}$$

then B_c is nonsingular and $x_+ = x_c - B^{-1}F(x_c)$ satisfies

$$\|e_+\| \leq \rho \|e_c\|. \tag{2.5}$$

Moreover

$$\|F(x_+) + E_c s_c\| \leq \gamma \|s_c\|^2. \tag{2.6}$$

Lemma 2.2. *Let the hypotheses of Lemma 2.1 hold. Then there is C_1 such that if $\|e_c\| \leq \epsilon_0$, $\|E_c\| \leq \delta_0$, and s_c is a major step, then*

$$\|E_+\| \leq (1+\sigma^{-1})\|E_c\| + C_1\|s_c\| \tag{2.7}$$

and

$$\|E_+ s_c\| \leq C_1 \|s_c\|^2. \tag{2.8}$$

Proof. By (2.2)

$$E_+ = E_c + \frac{F(x_+)F(x_+)^T}{s_c^T F(x_+)^T} \tag{2.9}$$

and hence, using (2.6),

$$\begin{aligned}\|E_+\| &\leq \|E_c\| + \sigma^{-1}\frac{\|F(x_+)\|}{\|s_c\|} \leq \|E_c\| + \sigma^{-1}\frac{\|E_c s_c\| + \gamma\|s_c\|^2}{\|s_c\|} \\ &\leq (1+\sigma^{-1})\|E_c\| + \sigma^{-1}\gamma\|s_c\|.\end{aligned}$$

This is (2.7) with $C_1 = \gamma\sigma^{-1}$.

We apply (2.9) again to obtain

$$E_+ s = E_c s + F(x_+)$$

and hence (2.8) follows from (2.6) and the the fact that $C_1 > \gamma$. □

At this point we need to consider a sequence of major steps. Several minor steps in which (1.5) fails could lie between any two major steps, but they have no effect on the estimates of the approximate Jacobians. This is also the first place where symmetry of E (and hence of $F'(x^*)$) plays an important role. We remark that (2.11) and (2.12) differ from estimates of $\|E_k\|$, $\|y - B_k s_k\|$, and $\|E_k s_k\|$, used in other recent papers [5], [15] in that the assumption of good approximations to x^* and $F'(x^*)$ is used in a crucial way. The next lemma uses the general approach of [5] and the observation from [5] and [15] that only the major steps need be considered in the estimates.

Lemma 2.3. *Let Assumption 1.1 hold, let $\rho \in (0,1)$, and let ϵ_0 and δ_0 be such that the conclusions of Lemma 2.1 hold. Let $\bar{k} \geq 0$, x_0, and B_0 be such that $F'(x^*) - B_0$ is symmetric and*

$$\frac{(1+C_1)(1+\rho)}{1-\rho}(1+2\sigma^{-1})^{\bar{k}}(\|E_0\| + \|e_0\|) \leq \delta_0. \tag{2.10}$$

Then at least $\bar{k}$ major steps can be taken with B_{k+1} nonsingular for all $0 \le k \le \bar{k}-1$. Moreover, there is $C_2 > 0$ such that if $\{s_j\}_{k=0}^{k-1}$, $\{x_{j+1}\}_{j=0}^{k}$, and $\{B_{j+1}\}_{j=0}^{k-1}$ are the sequence of the first $k \le \bar{k}$ major steps, iterations, and updates, then for any $1 \le k \le \bar{k}$,

$$\|E_k\| \le C_2(1+2\sigma^{-1})^k(\|E_0\| + \|e_0\|) \le \delta_0. \tag{2.11}$$

Also, for any $0 \le j < k$

$$\|E_k s_j\| \le C_2(1+2\sigma^{-1})^{k-j-1}\|s_j\|^2. \tag{2.12}$$

Proof. We set

$$C_2 = \frac{(1+C_1)(1+\rho)}{1-\rho}. \tag{2.13}$$

We note that $\|e_{k+1}\| \le \rho\|e_k\|$ implies that

$$\|s_k\| \le (1+\rho)\|e_k\| \le (1+\rho)\rho^k\|e_0\| \tag{2.14}$$

We prove the lemma by induction on k. For $k = 1$, (2.11) follows from (2.7), (2.10), and (2.14). We obtain (2.12) from (2.8) and (2.13).

If (2.11) holds for all $k < K \le \bar{k}$, then from (2.7) and (2.14),

$$\begin{aligned}\|E_K\| &\le (1+\sigma^{-1})\|E_{K-1}\| + C_1\|s_K\| \\ &\le (1+\sigma^{-1})\|E_{K-1}\| + C_1(1+\rho)\rho^K\|e_0\|.\end{aligned}$$

We use the induction hypothesis and $C_1(1+\rho) < C_2$ to obtain

$$\begin{aligned}\|E_K\| &\le (1+\sigma^{-1})C_2(1+2\sigma^{-1})^{K-1}(\|E_0\| + \|e_0\|) + C_2\rho^K\|e_0\| \\ &\le C_2(1+2\sigma^{-1})^K\|E_0\| + C_2(\rho^K + (1+\sigma^{-1})(1+2\sigma^{-1})^{K-1})\|e_0\|.\end{aligned}$$

Since $\rho < 1 < \sigma^{-1}$ we have

$$C_2(\rho^K + (1+\sigma^{-1})(1+2\sigma^{-1})^{K-1}) \le C_2(1+2\sigma^{-1})^K$$

which proves the first inequality in (2.11). The second inequality $\|E_K\| \le \delta_0$ follows from (2.10).

Since $\|e_k\| \le \epsilon_0$, we may apply Lemma 2.1 to conclude that at least $\bar{k}$ major steps can be taken and $\|e_k\| \le \rho^k\|e_0\|$ for all $0 \le k \le \bar{k}$. Hence, for $1 \le l < k \le \bar{k}$,

$$\|s_k\| \le \frac{\rho^{k-l}(1+\rho)}{(1-\rho)}\|s_l\|.$$

Assuming that (2.12) holds for $1 \le k < K$, we note that (2.12) is a consequence of Lemma 2.2 if $k = K$ and $j = K-1$. For $j < K-1$ we have

$$\|E_K s_j\| \le \|E_{K-1}s_j\| + \frac{|F(x_K)^T s_j|}{\sigma\|s_{K-1}\|}. \tag{2.15}$$

Every E_k is symmetric because E_0 is. Hence, if we write

$$F(x_K) = -E_{K-1}s_{K-1} + \Delta_{K-1}$$

then

$$\begin{aligned} F(x_K)^T s_j &= -(E_{K-1}s_{K-1})^T s_j + \Delta_{K-1}^T s_j \\ &= -s_{K-1}^T(E_{K-1}^T s_j) + \Delta_{K-1}^T s_j \\ &= -s_{K-1}^T(E_{K-1}s_j) + s_j^T \Delta_{K-1}. \end{aligned} \tag{2.16}$$

Combining (2.16) and (2.6) yields

$$|F(x_K)^T s_j| \le \|s_{K-1}\|(\|E_{K-1}s_j\| + \gamma\|s_j\|\|s_{K-1}\|).$$

Hence, by (2.15),

$$\begin{aligned} \|E_K s_j\| &\le (1+\sigma^{-1})\|E_{K-1}s_j\| + C_1\|s_j\|\|s_{K-1}\| \\ &\le (1+\sigma^{-1})\|E_{K-1}s_j\| + C_1(1+\rho)\rho^{K-j-1}\|s_j\|^2/(1-\rho). \end{aligned} \tag{2.17}$$

As in the proof of (2.11) we have

$$C_1\rho^{K-1-j}(1+\rho)/(1-\rho) \le C_2\rho^{K-1-j} \tag{2.18}$$

and

$$(1+\sigma^{-1})(1+2\sigma^{-1})^{K-2-j} + \rho^{K-1-j} \le (1+2\sigma^{-1})^{K-1-j}. \tag{2.19}$$

We apply (2.18), (2.19), and the induction hypothesis to (2.17) to get

$$\begin{aligned} \frac{\|E_K s_j\|}{\|s_j\|^2} &\le (1+\sigma^{-1})(C_2(1+2\sigma^{-1})^{K-2-j}) + C_2\rho^{K-1-j} \\ &\le C_2(1+2\sigma^{-1})^{K-1-j} \end{aligned} \tag{2.20}$$

verifying (2.12). $\square$

The estimates in Lemma 2.3 are analogs to the bounded deterioration results common in the quasi-Newton method literature. However, in this case the deviation of B_k from $F'(x^*)$ is exponentially increasing and, at least according to the bounds in Lemma 2.3, can eventually become so large that convergence may be lost. The SR1 update has, however, a self-correcting property in the linear case [9] that has been exploited in much of the recent work on the method [15], [3], [5]. This self-correction property overcomes the instability indicated in the estimates (2.11) and (2.12).

2.2. Uniform Linear Independence. Our uniform linear independence assumption differs slightly from that in [5]. We only consider major steps and are not concerned at this point with the total number (major + minor) steps required to form the sequence of linearly independent major steps.

ASSUMPTION 2.1. *There are $c_{min} > 0$ and $\bar{k} \geq n$ such that the hypotheses of Lemma 2.3 hold. Moreover from each of the sets of columns of normalized major steps*

$$\{v_k\}_{k=(p-1)\bar{k}+1}^{p\bar{k}} \equiv \{s_k/\|s_k\|\}_{k=(p-1)\bar{k}+1}^{p\bar{k}}$$

a subsequence $\{v_j^p\}_{j=1}^N$ can be extracted such that the matrix $\mathcal{S}_p$ with columns $\{v_j^p\}_{j=1}^N$ has minimum singular value at least c_{min}.

2.3. Convergence Results. Using the assumptions above we can prove q-linear convergence directly using Lemma 2.3 and the methods for exploitation of uniform linear independence from [5].

THEOREM 2.4. *Let Assumptions 2.1 and the hypotheses of Lemma 2.3 hold. Let $\rho \in (0,1)$ be given. Then if $\|e_0\|$ and $\|E_0\|$ are sufficiently small the SR1 iterates converge q-linearly to x^* with q-factor at most ρ.*

Proof. The proof is based on the simple observation that Lemma 2.3 states that the iteration may proceed through $\bar{k}$ major steps ($p = 1$) and that then Assumption 2.1 implies that the iteration may continue.

Let $p = 1$. Lemma 2.3 implies that $\|E_k\| \leq \delta_0$ for $0 \leq k \leq \bar{k}$ and that by (2.14)

$$\begin{aligned} \frac{\|E_{\bar{k}} s_k\|}{\|s_k\|} &\leq C_2(1 + 2\sigma^{-1})^{\bar{k}-k-1}\|s_k\| \\ &\leq C_2(1 + 2\sigma^{-1})^{\bar{k}-k-1}(1+\rho)\rho^k\|e_0\| \\ &\leq C_2(1 + 2\sigma^{-1})^{\bar{k}}(1+\rho)\|e_0\| \end{aligned} \tag{2.21}$$

Note that

$$\|E_{\bar{k}}\| \leq c_{min}^{-1}\|E_{\bar{k}}\mathcal{S}_0\| \leq c_{min}^{-1}\sqrt{N}C_2(1 + 2\sigma^{-1})^{\bar{k}}(1+\rho)\|e_0\|.$$

Now $\|e_{\bar{k}}\| \leq \rho^{\bar{k}}\|e_0\|$. Hence, if $\|e_0\|$ is small enough so that

$$C_2(1 + 2\sigma^{-1})^{\bar{k}}(C_2(1 + 2\sigma^{-1})^{\bar{k}}\sqrt{N}c_{min}^{-1} + \rho^{\bar{k}+1})\|e_0\| < \delta_0 \tag{2.22}$$

then (2.10) holds if E_0 and e_0 are replaced by $E_{\bar{k}}$ and $e_{\bar{k}}$. Hence, we may continue the iteration by Lemmas 2.1 and 2.3, obtaining q-linear convergence with q-factor at most ρ. ∎

3. Structured Updates. In order to apply the convergence results to optimal control problems in the context of pointwise updates, as a next step, we extend the statements from the previous sections to the case of structured updates. We use the notational conventions of [7] in the formulation.

Suppose that the Jacobian F' of F can be split into a Lipschitz continuous computable part $C(x)$ and a part $A(x)$ which will be approximated by a SR1 update:

$$F'(x) = A(x) + C(x).$$

We define the SR1 update by

$$B_+ = B_c + \frac{(y^\# - B_c s_c)(y^\# - B_c s_c)^T}{(y^\# - B_c s_c)^T s_c}, \tag{3.1}$$

where the step is computed by solving

$$(B_c + C(x_c))s_c = -F(x_c). \tag{3.2}$$

If we choose

$$y^\# = F(x_+) - F(x_c) - C(x_+)s_c, \tag{3.3}$$

the secant condition $(B_+ + C(x_+))s_c = F(x_+) - F(x_c)$ holds and we obtain from (3.2) and (3.3)

$$y^\# - B_c s_c = F(x_+) + (C(x_c) - C(x_+))s_c = \tilde{F}_+ \tag{3.4}$$

Hence the SR1 update can be written with a perturbation $\tilde{F}_+$ of $F(x_+)$

$$B_+ = B_c + \frac{\tilde{F}_+ \tilde{F}_+^T}{s_c^T \tilde{F}_+} \tag{3.5}$$

We use the notation

$$E = B + C(x) - F'(x^*) = B - A(x_*) + C(x) - C(x_*) \tag{3.6}$$

for errors in the Jacobian.

Next we apply Lemma 2.1 to obtain a similar estimate

LEMMA 3.1. *Let Assumption 1.1 hold and let $\rho \in (0,1)$ be given. Then there are ϵ_0 and $\tilde{\delta}_0$ such that if x_c and B_c for a structured SR1 update satisfy*

$$\|e_c\| \le \epsilon_0 \text{ and } \|B_c - A(x_*)\| \le \tilde{\delta}_0 \tag{3.7}$$

then $x_+ = x_c - (B_c + C(x_c))^{-1}F(x_c)$ is well defined and satisfies

$$\|e_+\| \le \rho\|e_c\|. \tag{3.8}$$

Moreover

$$\|F(x_+) + E_c s_c\| \le \gamma\|s_c\|^2 \quad \text{and} \quad \|\tilde{F}_+ + E_c s_c\| \le \gamma_1\|s_c\|^2. \tag{3.9}$$

Proof. Note that

$$\|E_c\| \leq \|B_c - A(x_*)\| + \|C(x_c) - C(x_*)\| \leq \tilde{\delta}_0 + \gamma_C \epsilon_0$$

with a Lipschitz constant γ_C for C. Then $\|E_c\| \leq \delta_0$ holds and since Lemma 2.1 holds for arbitrary approximations of the Jacobian, x_+ exists and (3.8) holds. Observing that

$$\|(C(x_c) - C(x_+))s_c\| \leq \gamma_C \|s_c\|^2$$

proves (3.9) and completes the proof. □

The next lemma is an extension of Lemma 2.2. Before we can state the lemma we define when an update is skipped. We update B only if

$$|s_c^T(y^\# - B_c s_c)| \geq \sigma \|s_c\| \|y^\# - B_c s_c\| \tag{3.10}$$

for some $\sigma \in (0,1)$ fixed. The definition of minor and major steps is the same as that in § 2 with (3.10) playing the role of (1.5).

LEMMA 3.2. *Let the hypotheses of Lemma 3.1 hold. Then there is C_1 such that if $\|e_c\| \leq \epsilon_0$, $\|E_c\| \leq \delta_0$, and s_c is a major step, then*

$$\|E_+\| \leq (1 + \sigma^{-1})\|E_c\| + C_1 \|s_c\| \tag{3.11}$$

and

$$\|E_+ s_c\| \leq C_1 \|s_c\|^2. \tag{3.12}$$

Proof. For the structured updates we have

$$E_+ = E_c + \frac{\tilde{F}_+ \tilde{F}_+^T}{s^T \tilde{F}_+} + C(x_+) - C(x_c). \tag{3.13}$$

If we note that

$$\|C(x_+) - C(x_c)\| \leq \gamma_C \|s_c\|$$

and (3.10) has been changed accordingly then the proof is the same as for Lemma 2.2 with

$$C_1 = \sigma^{-1}(\gamma + \gamma_C)$$

and $F(x)$ replaced by $\tilde{F}$. □

In the next lemma symmetry is assumed which causes some additional changes for the proof.

LEMMA 3.3. *Let Assumption 1.1 and the hypotheses of Lemma 3.1 hold. Let $\bar{k} \geq 0$, x_0, and B_0 be such that $B_0 - A(x^*)$ is symmetric and*

$$\frac{(1 + C_1)(1 + \rho)}{1 - \rho}(1 + 2\sigma^{-1})^{\bar{k}}(\|E_0\| + \|e_0\|) \leq \delta_0. \tag{3.14}$$

Then at least $\bar{k}$ major steps can be taken with B_{k+1} nonsingular for all $0 \le k \le \bar{k}-1$. Moreover, there is $C_2 > 0$ such that if $\{s_j\}_{k=0}^{k-1}$, $\{x_{j+1}\}_{j=0}^{k}$, and $\{B_{j+1}\}_{j=0}^{k-1}$ are the sequence of the first $k \le \bar{k}$ major steps, iterations, and updates, then for any $1 \le k \le \bar{k}$,

$$\|E_k\| \le C_2(1+2\sigma^{-1})^k(\|E_0\| + \|e_0\|) \le \delta_0. \tag{3.15}$$

Also, for any $0 \le j < k$

$$\|E_k s_j\| \le C_2(1+2\sigma^{-1})^{k-j-1}\|s_j\|^2. \tag{3.16}$$

Proof. The first part of the induction proof for (3.15) is identical to the one for Lemma 2.3 and therefore omitted.

Assuming that (3.16) holds for $1 \le k < K$, we note that (3.16) is a consequence of Lemma 3.2 if $k = K$ and $j = K-1$. For $j < K-1$ we have from (3.13)

$$\|E_K s_j\| \le \|E_{K-1}s_j\| + \frac{|\tilde{F}_K^T s_j|}{\sigma\|s_{K-1}\|} + \gamma_C\|s_{K-1}\|\|s_j\|. \tag{3.17}$$

If we write

$$\tilde{F}_K = -E_{K-1}s_{K-1} + \Delta_{K-1}$$

then by (3.9)

$$\|\Delta_{K-1}\| \le \gamma\|s_{K-1}\|^2$$

and

$$\begin{aligned} \tilde{F}_K^T s_j &= -(E_{K-1}s_{K-1})^T s_j + \Delta_{K-1}^T s_j \\ &= -s_{K-1}^T(E_{K-1}^T s_j) + \Delta_{K-1}^T s_j \end{aligned} \tag{3.18}$$

Combining (3.18) and (3.9) yields

$$|\tilde{F}_K^T s_j| \le \|s_{K-1}\|(\|E_{K-1}^T s_j\| + \gamma\|s_j\|\|s_{K-1}\|).$$

Note that by assumption $B - A(x_*)$ is symmetric and hence by definition (3.6)

$$\|E - E^T\| = \|C(x) - C(x_*) - (C(x) - C(x_*))^T\| \le 2\gamma_C\|e\|.$$

Hence, by (3.17),

$$\begin{aligned} \|E_K s_j\| &\le \|E_{K-1}s_j\| + \sigma^{-1}(\|E_{K-1}^T s_j\| + \gamma\|s_j\|\|s_{K-1}\|) + \gamma_C\|s_{K-1}\|\|s_j\| \\ &\le (1+\sigma^{-1})\|E_{K-1}s_j\| + 2\gamma_C\sigma^{-1}\|e_{K-1}\|\|s_j\| + (\gamma\sigma^{-1} + \gamma_C)\|s_{K-1}\|\|s_j\| \\ &\le (1+\sigma^{-1})\|E_{K-1}s_j\| + (2\gamma_C(1-\rho)^{-1}\sigma^{-1} + \gamma\sigma^{-1} + \gamma_C)\|s_{K-1}\|\|s_j\| \\ &\le (1+\sigma^{-1})\|E_{K-1}s_j\| + C_2\rho^{K-j-1}\|s_j\|^2, \end{aligned} \tag{3.19}$$

where, as in § 2

$$C_2 = C_2 = \frac{(1+C_1)(1+\rho)}{1-\rho}.$$

As in the proof of (2.11) we have

$$C_1\rho^{K-1-j}(1+\rho)/(1-\rho) \le C_2\rho^{K-1-j} \tag{3.20}$$

and

$$(1+\sigma^{-1})(1+2\sigma^{-1})^{K-2-j} + \rho^{K-1-j} \le (1+2\sigma^{-1})^{K-1-j}. \tag{3.21}$$

We apply (3.20), (3.21), and the induction hypothesis to (3.19) to get

$$\begin{aligned} \frac{\|E_K s_j\|}{\|s_j\|^2} &\le (1+\sigma^{-1})(C_2(1+2\sigma^{-1})^{K-2-j}) + C_2\rho^{K-1-j} \\ &\le C_2(1+2\sigma^{-1})^{K-1-j} \end{aligned} \tag{3.22}$$

verifying (3.16). □

Using the definition of uniform linear independence from § 2.2 we may state the structured analog of Theorem 2.4. The proof is essentially identical to that of Theorem 2.4.

THEOREM 3.4. *Let Assumptions 2.1 and the hypotheses of Lemma 3.3 hold. Let $\rho \in (0,1)$ be given. Then if $\|e_0\|$ and $\|E_0\|$ are sufficiently small the SR1 iterates converge q-linearly to x^* with q-factor at most ρ.*

4. Pointwise Structured Updates for Optimal Control. In order to apply the convergence results to optimal control problems in the context of pointwise updates, as a next step, we extend the statements from the previous sections to the case of pointwise structured updates.

Our nonlinear equations represent the necessary conditions for the optimal control problem

$$\text{minimize} \int_0^T L(x(t),\, u(t),\, t)dt$$

over $u \in L^\infty_M[0,T]$ such that $x \in C_N[0,T]$ solves

$$\dot{x} = f(x,u,t) \quad , \quad x(0) = x_0 \quad , \quad t \in (0,T).$$

Here $f : I\!R^N \times I\!R^M \times [0,T] \to I\!R^N$ and $L : I\!R^N \times I\!R^M \times [0,T] \to I\!R$.

We set

$$H(x,u,t) = f^T(x,u,t)\,p(t) + L(x,u,t),$$

where $p \in C_N[0,T]$ solves the adjoint equation

$$-\dot{p} = f_x^T p + L_x^T \quad , \quad p(T) = 0.$$

We seek to solve the nonlinear system

$$F(z) = F(p,x,u) \;=\; \begin{pmatrix} \dot{x} - f(x,u,t) \\ \dot{p} + H_x(x,u,t) \\ H_u(x,u,t) \end{pmatrix} \;=\; \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{4.1}$$

for $z = (p,x,u)^T \in X^\infty$ which satisfy the boundary conditions $x(0) = x_0$ and $p(T) = 0$ and $F : X^\infty \to Y^\infty$. Here, for $1 \le r \le \infty$,

$$\begin{aligned} X^r &= W_N^{1,r}[0,T] \times W_N^{1,r}[0,T] \times L_M^r[0,T] \\ Y^r &= L_N^r[0,T] \times L_N^r[0,T] \times L_M^r[0,T]. \end{aligned}$$

We use the following assumption

ASSUMPTION 4.1. *f, L and their first and second partial derivatives with respect to x and u are continuous on $\mathbb{R}_N \times \mathbb{R}_M \times [0,T]$.*

Under the above assumption, F is Fréchet -differentiable and the Fréchet -derivative is given by

$$F'(p,x,u) \begin{pmatrix} \pi \\ \xi \\ \nu \end{pmatrix} \;=\; \begin{pmatrix} 0 & D - f_x & -f_u \\ D + f_x^T & H_{xx} & H_{xu} \\ f_u^T & H_{ux} & H_{uu} \end{pmatrix} \begin{pmatrix} \pi \\ \xi \\ \nu \end{pmatrix} \tag{4.2}$$

with $D = \frac{d}{dt}$ and all other components as multiplication operators.

$F'(p,x,u)$ is split into two parts $F'(z) = A(z) + C(z)$ with $C(z)$ containing all information from first derivatives

$$C(z) \;=\; \begin{pmatrix} 0 & D - f_x & -f_u \\ D + f_x^T & 0 & 0 \\ f_u^T & 0 & 0 \end{pmatrix}$$

and $A(z)$ consisting of second order derivatives

$$A(z) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & H_{xx}(z) & H_{xu}(z) \\ 0 & H_{ux}(z) & H_{uu}(z) \end{pmatrix},$$

All entries depend on time t so that $A(z)$ is typically approximated by a family of quasi Newton updates depending also on time t

$$B \;=\; \begin{pmatrix} 0 & 0 & 0 \\ 0 & B_{11} & B_{12} \\ 0 & B_{21} & B_{22} \end{pmatrix} \tag{4.3}$$

with

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \;\in\; L^\infty_{(M+N)\times(M+N)}\,[0,T].$$

Set

$$y^{\#} = F(z_+) - F(z_c) - C(z_+)s_c. \tag{4.4}$$

We use a pointwise analog of (3.10)

$$|s_c^T(t)(y^{\#} - B_c s_c)(t)| \geq \sigma \|s_c(t)\| \|(y^{\#} - B_c s_c)(t)\|. \tag{4.5}$$

We update B at each $t \in [0,T]$ by a structured SR1 update

$$B_+(t) = \begin{cases} B_c(t) + \dfrac{(y^{\#} - B_c s_c)(t)(y^{\#} - B_c s_c)^T(t)}{(y^{\#} - B_c s_c)^T(t) s_c(t)} & \text{if (4.5) holds} \\ B_c(t) & \text{otherwise.} \end{cases} \tag{4.6}$$

In order to justify the use of pointwise updates, we state the next lemma.

LEMMA 4.1. *If B_0 is given of the form* (4.3), *then all B_k defined by* (4.6) *are also multiplication operators with*

$$\begin{pmatrix} B_{k\ 11} & B_{k\ 12} \\ B_{k\ 21} & B_{k\ 22} \end{pmatrix} \in L^\infty_{(M+N)\times(M+N)}[0,T].$$

Proof. The proof is via induction and we show the step from B_c to B_+. We write $s_c = (s_{pc}, s_{xc}, s_{uc})^T$. Note that by (4.4)

$$y^{\#} - B_c s_c = F(z_+) - F(z_c) - C(z_+)s_c - B_c s_c.$$

Since the differentiation operator D appears linearly in F and in C it cancels in

$$F(z_+) - F(z_c) - C(z_+)s_c = \begin{pmatrix} -f(x_+,u_+) + f(x_c,u_c) + f_x(x_+,u_+)s_{xc} + f_u(x_+,u_+)s_{uc} \\ H_x(x_+,u_+) - H_x(x_c,u_c) - f_x(x_+,u_+)^T s_{pc} \\ H_u(x_+,u_+) - H_u(x_c,u_c) - f_u(x,u_+)^T s_{uc} \end{pmatrix}$$

This implies with the differentiability assumption on the data that pointwise

$$\|F(z_+)(t) - F(z_c)(t) - (C(z_+)s_c)(t)\| \leq \phi(t)\|s_c(t)\|$$

holds for some $\phi \in L^\infty[0,T]$. Since B_c is also in L^∞ we have

$$\|y^{\#}(t) - B_c(t)s_c(t)\| \leq (\phi(t) + \|B_c(t)\|)\|s_c(t)\|. \tag{4.7}$$

The decision in the update formula (4.6) shows that the components of B_+ are measurable functions. They are also essentially bounded because either

$$\|B_+(t)\| = \|B_c(t)\|$$

or using the choice of σ in (4.6) and (4.7)

$$\begin{aligned}\|B_+(t)\| &\le \|B_c(t)\| + \|y^\#(t) - B_c(t)s_c(t)\|^2 |(y^\#(t) - B_c(t)s_c(t))^T s_c(t)|^{-1} \\ &\le \|B_c(t)\| + \sigma^{-1}\|y^\#(t) - B_c(t)s_c(t)\| \|s_c(t)\|^{-1} \\ &\le \|B_c(t)\| + \sigma^{-1}(\phi(t) + \|B_c(t)\|)\end{aligned}$$

This completes the proof. □

The next Lemma gives pointwise estimates on the error in the Jacobian in the context of pointwise updates which will be used later for uniform estimates.

LEMMA 4.2. *Assume that for some* $t \in [0,T]$ $E_c(t)$ *is a multiplication operator and let* $\|E_c(t)\| \le \delta_0$ *and* $\|e_c(t)\| \le \epsilon_0$. *If* $s_c(t)$ *is a major step, then* $E_+(t)$ *is also a multiplation operator and the following estimates hold:*

$$\|E_+(t)\| \le (1+\sigma^{-1})\|E_c(t)\| + (1+\sigma^{-1})\gamma_C\|s_c(t)\| + \sigma^{-1}\gamma_F(\|e_+(t)\| + \|e_c(t)\|) \tag{4.8}$$

$$\|E_+ s_c(t)\| \le (\gamma_F(\|e_c(t)\| + \|e_+(t)\|)\|s_c(t)\| \tag{4.9}$$

and

$$\begin{aligned}\|y^\#(t) - B_c(t)s_c(t) & + E_+(t)s_c(t)\| \\ & \le (\gamma_F(\|e_+(t)\| + \|e_c(t)\|) + \gamma_C\|s_c(t)\|)\|s_c(t)\|.\end{aligned} \tag{4.10}$$

Proof. Observe that we can rewrite (4.4) pointwise

$$y^\# - B_c s_c = (F(z_+) - F(z_c) - F'(z_*)s_c) + (C(z_c) - C(z_+))s_c - E_c s_c. \tag{4.11}$$

By assumption, the last term is a multiplication operator and therefore does not contain the differentiation operator D. In first and the second term in parentheses the D cancels, so that we can estimate pointwise under the given smoothness assumptions on the data:

$$\begin{aligned}\|(y^\# - B_c s_c)(t)\| &\le \gamma_F(\|e_+(t)\| + \|e_c(t)\|)\|s_c(t)\| \\ &\quad + \gamma_C\|s_c(t)\|^2 + \|E_c(t)\|\|s_c(t)\|\end{aligned} \tag{4.12}$$

Hence we obtain from (3.13) using (4.12) that pointwise

$$\begin{aligned}\|E_+(t)\| \le \|E_c(t)\| & + \sigma^{-1}\|y^\#(t) - B_c(t)s_c(t)\|\|s_c(t)\|^{-1} \\ & + \|C(z_+)(t) - C(z_c)(t)\| \\ \le \|E_c(t)\| & + \sigma^{-1}(\gamma_F(\|e_+(t)\| + \|e_c(t)\|) \\ & + \gamma_C\|s_c(t)\| + \|E_c(t)\|) + \gamma_C\|s_c(t)\|\end{aligned}$$

To estimate $\|E_+ s_c\|$ recall that from the secant condition

$$E_+ s_c = (B_+ + A_+)s_c - F'(z_*)s_c = F(z_+) - F(z_c) - F'(z_*)s_c$$

and therefore (4.9) holds. Furthermore, from (4.11)

$$y^\# - B_c s_c + E_+ s_c = (F(z_+) - F(z_c) - F'(z_*)s_c) + (C(z_c) - C(z_+))s_c$$

and

$$\|y^\#(t) - B_c(t)s_c(t) + E_+(t)s_c(t)\| \leq (\gamma_F(\|e_+(t)\| + \|e_c(t)\|)\|s_c(t)\| + \gamma_C\|s_c\|^2$$

□

The next Lemma describes a linear rate estimate in a uniform norm.

LEMMA 4.3. *For $\rho \in (0,1)$ there are ϵ_0 and δ_0 such that if s_c is a major step for all t and B_c for the pointwise structured SR1 update satisfies*

$$\|e_c\|_\infty \leq \epsilon_0 \text{ and } \|B_c - A(z_*)\|_\infty \leq \delta$$

then $z_+ = z_c - (B_c + C(z_c))^{-1}F(z_c)$ is well defined and satisfies

$$\|e_+\|_\infty \leq \rho\|e_c\|_\infty. \tag{4.13}$$

Moreover for some γ, C_1, C_2

$$\|E_+\|_\infty \leq (1 + \sigma^{-1})\|E_c\|_\infty + C_1\|s_c\|_\infty \tag{4.14}$$

$$\|E_+ s_c\|_\infty \leq C_1\|s_c\|_\infty^2 \tag{4.15}$$

and

$$\|y^\# - B_c s_c + E_+ s_c\|_\infty \leq \gamma\|s_c\|^2 \tag{4.16}$$

Proof. The assumptions imply that

$$\|E_c\|_\infty \leq \|B_c - A(z_*)\|_\infty + \|C(z_c) - C(z_*)\|_\infty \leq \delta_0 + \gamma_C\epsilon_0$$

is small so that Lemma 2.1 can be invoked to yield (4.13). From this we deduce $\|e_c\|_\infty \leq (1-\rho)^{-1}\|s_c\|_\infty$. (4.8) gives

$$\begin{aligned}\|E_+\|_\infty &\leq (1 + \sigma^{-1})\|E_c\|_\infty + (1 + \sigma^{-1})\gamma_C\|s_c\|_\infty + \sigma^{-1}\gamma_F(\|e_+\|_\infty + \|e_c\|_\infty) \\ &\leq (1 + \sigma^{-1})\|E_c\|_\infty + ((1 + \sigma^{-1})\gamma_C + \sigma^{-1}\gamma_F(1+\rho)(1-\rho)^{-1})\|s_c\|_\infty\end{aligned}$$

which is (4.14). In the same way we use (4.9) and (4.10) to obtain (4.15) and (4.16), resp. □

5. Numerical Results. We present some numerical results which illustrate the observations from the previous sections. Let us consider the following class of examples:

Let $N = 2, M = 1$ and

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2, u) \\ \dot{x}_2 &= x_1 \end{aligned}$$

First we set

$$\begin{aligned} f_1(x_1, x_2, u) &= (1 - x_2^2)x_1 - x_2 + u \\ L(x_1, x_2, u) &= (x_1^2 + x_2^2 + u^2)/2 \end{aligned}$$

Furthermore, we set $x_0 = (0, 3)^T, \quad T = 3$. The initial data are given by

$$B_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in I\!R^{3\times 3} \text{ and } z_0 = (0, 0, 0)^T.$$

We did update if

$$|(y^{\#}(t) - B(t)\hat{s}(t))^T \hat{s}(t)| \geq \sigma \|(y^{\#}(t) - B(t)\hat{s}(t))\|_2 \|\hat{s}(t)\|_2$$

is true. In [14] we used

$$\sigma = h^4.$$

The discretization parameter $h = 1/120$ comes from the discretization of the two-point boundary value problem by the trapezoid finite difference scheme used with a Richardson extrapolation to achieve 4th order accurate results, see e.g. [12]. This indicates that the termination criterion should be

$$\|F'(z_k)\|_2 \leq (2h)^4 \approx 0.6\ 10^{-5}.$$

We approximate the norm in the discrete case also with accuracy of order 4.

The numbers in column [No Upd %] give the percentage of the 121 3×3-matrices which have not been updated at iteration k. The difference in the matrices is computed as follows:

$$\|B - A^*\| = \max_{0 \leq i \leq 120} \|B(t_i) - A^*(t_i)\|_F,$$

where $\|B(t)\|_F$ denotes the Frobenius-Norm in $R^{3\times 3}$.

In Table 5.1 we report on the results for the SR-1 update where the choice of σ is based on the truncation error of the discretization scheme ($h^4 \approx 0.6 \times 10^{-5}$). Tables 5.2 and 5.3 show the effects of more conservative updating strategies. The analysis in the preceding sections indicates that a larger value of σ will keep $\|E_k\|$ smaller and might thereby allow for a more monotone iteration. While an increase by a factor of 100 did reduce the size of $\|E_k\|$, it did not lead to any improvement in overall performance (see Table 5.2). Increasing σ by a factor of 1000 (see Table 5.3) led to a performance improvement of about 20%.

In Table 5.2 we used a less stringent requirement for not updating the Hessians. In this table σ was increased by a factor of 10^2.

Note the reduced error in the Hessian updates in the early phase of the algorithm.

TABLE 5.1
Pointwise SR-1 update, $\sigma = 0.6\ 10^{-5}$

k	$\|F(z_k)\|_2$	$\|F(z_k)\|_2/\|F(z_{k-1})\|_2$	No Upd %	$\|B - A^*\|$
1	0.42870D+01	4.287	0.8	13870.320
2	0.20575D+01	0.480	0.8	13877.437
3	0.14663D+01	0.713	0.8	13870.092
4	0.26374D+00	0.180	0.8	13307.899
5	0.28341D+00	1.075	1.7	13719.676
6	0.62710D-01	0.221	1.7	6218.692
7	0.51961D-01	0.829	1.7	4083.302
8	0.42872D-01	0.825	1.7	2949.300
9	0.72865D-04	0.002	1.7	332.568
10	0.22618D-04	0.310	1.7	52.916
11	0.21262D-05	0.094	1.7	15.803
12	0.92609D-06	0.436	0.8	129.835

TABLE 5.2
Pointwise SR-1 update, $\sigma = 0.6\ 10^{-3}$

k	$\|F(z_k)\|_2$	$\|F(z_k)\|_2/\|F(z_{k-1})\|_2$	No Upd %	$\|B - A^*\|$
1	0.42870D+01	4.287	1.7	152.417
2	0.20581D+01	0.480	0.8	296.288
3	0.14651D+01	0.712	1.7	252.101
4	0.25807D+00	0.176	1.7	155.621
5	0.29463D+00	1.142	5.0	155.621
6	0.43605D-01	0.148	2.5	155.621
7	0.75704D-03	0.017	1.7	155.621
8	0.44425D-04	0.059	3.3	120.137
9	0.10275D-04	0.231	2.5	383.280
10	0.23941D-04	2.330	3.3	120.137
11	0.10516D-05	0.044	6.6	120.137
12	0.19081D-08	0.002	1.7	0.074

TABLE 5.3
Pointwise SR-1 update, $\sigma = 0.6\ 10^{-2}$

k	$\|F(z_k)\|_2$	$\|F(z_k)\|_2/\|F(z_{k-1})\|_2$	No Upd %	$\|B - A^*\|$
1	0.42870D+01	4.287	1.7	65.960
2	0.20581D+01	0.480	1.7	112.648
3	0.14648D+01	0.712	5.8	126.006
4	0.25750D+00	0.176	5.0	69.245
5	0.29470D+00	1.144	12.4	69.245
6	0.37371D-01	0.127	12.4	69.245
7	0.55627D-03	0.015	9.9	69.245
8	0.38187D-04	0.069	8.3	69.245
9	0.13565D-05	0.036	6.6	62.154
10	0.20748D-06	0.153	5.8	62.154

REFERENCES

[1] C. G. Broyden, *Quasi-Newton methods and their application to function minimization*, Math. Comp., 21 (1967), pp. 368–381.

[2] C. G. Broyden, J. E. Dennis, and J. J. Moré, *On the local and superlinear convergence of quasi-Newton methods*, J. Inst. Maths. Applics., 12 (1973), pp. 223–246.

[3] R. H. Byrd, H. F. Khalfan, and R. B. Schnabel, *Analysis of a symmetric rank-one trust region method*, Tech. Report CU-CS-657-93, University of Colorado at Boulder, Department of Computer Science, July 1993.

[4] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Testing a class of methods for solving minimization problems with simple bounds on the variables*, Math. Comp., 50 (1988), pp. 399–430.

[5] ——, *Convergence of quasi-Newton matrices generated by the symmetric rank one update*, Math. Programming A, 50 (1991), pp. 177–195.

[6] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Nonlinear Equations and Unconstrained Optimization*, Prentice-Hall, Englewood Cliffs, N.J, 1983.

[7] J. E. Dennis and H. F. Walker, *Convergence theorems for least change secant update methods*, SIAM J. Numer. Anal., 18 (1981), pp. 949–987.

[8] P. Deuflhard, R. W. Freund, and A. Walter, *Fast secant methods for the iterative solution of large nonsymmetric linear systems*, Impact of Computing in Science and Engineering, 2 (1990), pp. 244–276.

[9] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming*, John Wiley and Sons, New York, 1968.

[10] P. Gilmore, *An Algorithm for Optimizing Functions with Multiple Minima*, PhD thesis, North Carolina State University, 1993.

[11] ——, *IFFCO: Implicit Filtering for Constrained Optimization*, Tech. Report CRSC-TR93-7, Center for Research in Scientific Computation, North Carolina State University, May 1993. available by anonymous ftp from math.ncsu.edu in pub/kelley/iffco/ug.ps.

[12] H. B. Keller, *Numerical Solution of Two Point Boundary Value Problems*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1982.

[13] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, no. 16 in SIAM Frontiers in Applied Mathematics, SIAM, Philadelphia, 1995. To appear: preliminary version available by anonymous ftp from math.ncsu.edu in pub/kelley/notes/iterative.PS.Z.

[14] C. T. Kelley, E. W. Sachs, and B. Watson, *A pointwise quasi-Newton method for unconstrained optimal control problems, II*, J. Optim. Theory Appl., 71 (1991), pp. 535–547.

[15] H. Khalfan, R. H. Byrd, and R. B. Schnabel, *A theoretical and experimental study of the symmetric rank one update*, SIAM J. Optimization, 3 (1993), pp. 1–24.

[16] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[17] M. R. Osborne and L. P. Sun, *A new approach to the symmetric rank-one updating algorithm*, tech. report, Department of Statistics, IAS, Australian National University, 1988.

[18] D. Stoneking, G. Bilbro, R. Trew, P. Gilmore, and C. T. Kelley, *Yield optimization using a GaAs process simulator coupled to a physical device model*, IEEE Transactions on Microwave Theory and Techniques, 40 (1992), pp. 1353–1363.

[19] P. L. Toint, *On large scale nonlinear least squares calculations*, SIAM J. Scientific and Statistical Computing, 8 (1987), pp. 416–435.

[20] B. Watson, *Quasi-Newton Methoden für Minimierungsprobleme mit strukturierter Hesse-Matrix*, Diploma Thesis, Universität Trier, 1990.

[21] S. J. Wright, *Compact storage of Broyden-class quasi-Newton matrices.* Argonne National Laboratory, preprint, 1994.