Local Analysis of Inexact Quasi-Newton Methods+

by

Stanley C. Eisenstat* and Trond Steihaug#

MASC TR 82-7

* Department of Computer Science
Yale University
New Haven, Connecticut

Department of Mathematical Sciences, Rice University
Houston, Texas
and
Statoil, Stavanger, Norway

+This research was supported by the Norwegian Research Council for Science and the Humanities and DoE Contract #EY-76-S-05-5046.

Abstract

Quasi-Newton methods are well known iterative methods for solving nonlinear problems. At each stage, a system of linear equations has to be solved. However, for large scale problems, solving the linear system of equations can be expensive and may not be justified when the iterate is far from the solution or when the matrix is an approximation to the Jacobian or Hessian matrix. Instead we consider a class of Inexact Quasi-Newton methods which solves the linear system only approximately. We derive conditions for local and superlinear rate of convergence in terms of a relative residual.

1. Introduction

Consider the system of n nonlinear equations in n unknowns

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

where the mapping $F:\mathbf{R}^n \to \mathbf{R}^n$ has the following properties

A.1: There exists x^* so that $F(x^*)=0$;

A.2: F is differentiable in an open neighborhood Ω of x^*;

A.3: $F'(x^*)$ is nonsingular.

Let $\|\ \|$ be a norm on $\mathbf{R}^n$ and its induced matrix norm. The modulus of continuity ω of F' at x^* is

$$\omega(\delta) \equiv \sup \{ \|F'(x) - F'(x^*)\| : \|x - x^*\| \le \delta \} \tag{1.2}$$

Assume that

A.4: There exists $\varepsilon > 0$ so that the modulus of continuity ω of F' at x^* satisfies

$$\int_0^{\varepsilon} \frac{\omega(\delta)}{\delta}\, d\delta < +\infty. \tag{1.3}$$

Quasi-Newton methods have proved themselves in dealing with the problem of finding a solution x^* of (1.1). These methods approximate the solution x^* by generating a sequence of iterates $\{ x_k \}$ where the correction of an iterate is found by solving a linear system of equations. Let B_k, $k=0,1,\cdots$ be an n by n nonsingular matrix that approximates the Jacobian matrix of F at x_k. The quasi-Newton method is:

Given x_0, and B_0

FOR $k=0$ STEP 1 UNTIL Convergence DO

Solve $B_k s_k = -F(x_k)$ (1.4)

Set $x_{k+1} = x_k + s_k$

Update to obtain B_{k+1}. (1.5)

For each iteration, a linear system in n equations and n unknowns (1.4) has to be solved. For large problems, solving the linear system (1.4) may not be justified when x_k is far from the solution x^* or when B_k is an approximation to the Jacobian matrix of F at x_k. Instead Steihaug [7] introduces the inexact quasi-Newton framework which accepts an approximate solution s_k of (1.4) if a relative residual is less than a tolerance θ_k that may depend on x_k. The inexact quasi-Newton framework is as follows

Given x_0, and B_0

FOR k=0 STEP 1 UNTIL Convergence DO

Find *some* s_k so that for $r_k = B_k s_k + F(x_k)$, then

$$\frac{\|r_k\|}{\|F(x_k)\|} \leq \theta_k \tag{1.6}$$

Set $x_{k+1} = x_k + s_k$ (1.7)

Update to obtain B_{k+1}. (1.8)

The non negative *forcing sequence* $\{\theta_k\}$ is used to control the level of accuracy. If $B_k = F'(x_k)$ then we have the Inexact Newton method [2].

The update function U introduced by Broyden, Dennis, and More [1] and generalized by Dennis and Walker [4], is a set valued approximation rule defined in a neighborhood of the solution of (1.1) and B_* where B_* is an approximation to the Jacobian matrix of F at x^*. The update B_{k+1} in (1.5) and (1.8) is chosen so that $B_{k+1} \in \mathrm{U}(x_k, x_{k+1}, B_k)$.This paper analyses the local behavior of methods in the Inexact quasi-Newton framework where we restrict the update function to be of bounded deterioration. In the case when $B_* = F'(x^*)$ we show in Section 2 that the methods are locally convergent if the forcing term is uniformly less than one. In Section 3, we characterize superlinear rate of convergence, and indicate how to choose the forcing sequence which yields superlinear convergence for secant methods. In Section 4 we discuss fixed scale least change secant update methods and in Section 5 we discuss iteratively rescaled least change secant update methods.

2. Local convergence

In this section we will prove local convergence theorem for the class of inexact methods using a wide class of update functions. Let B_* be a nonsingular matrix and define the vector norm $\|y\|_* = \|B_* y\|$ and let $\|\cdot\|_{M_0}$ be a matrix norm. From assumption A.4 we know there exists $\varepsilon > 0$ so that the modulus of continuity of F' at x^* satisfies

$$\omega(\delta) \equiv \sup \{ \|F'(x) - F'(x^*)\|_{M_0} : \|x - x^*\|_* \leqq \delta \} \tag{2.1}$$

using the the matrix $\| \ \|_{M_0}$ norm and the vector $\| \ \|_*$ norm so that

$$\int_0^{\varepsilon} \frac{\omega(\delta)}{\delta} \, d\delta < +\infty. \tag{2.2}$$

Let $P\{R^{n\times n}\}$ be the collection of all nonempty subsets of $R^{n\times n}$, i.e. the power set of $R^{n\times n}$. Further, let B_* be an approximation to the Jacobian matrix $F'(x^*)$. The update function U, is a set valued approximation rule for B_* with respect to F' at x^* defined in open neighborhood N of (x^*, x^*, B_*) in $R^n \times R^n \times R^{n\times n}$, $U: N \to P\{R^{n\times n}\}$. We say that the update function is of bounded deterioration for B_* with respect to F' at x^* if there exists $\alpha \geqq 0$ so that for $(x, \bar{x}, B) \in N$ and

$$B^+ \in U(x, \bar{x}, B) \tag{2.3}$$

there exist norms $\|\cdot\|_{M_+}$ and $\|\cdot\|_M$ so that

$$\|B^+ - B_*\|_{M_+} \leqq [1 + m(x,\bar{x})] \, \|B - B_*\|_M + m(x,\bar{x}) \tag{2.4}$$

where

$$m(x,z) = \alpha \max\{ \omega(\|x - x^*\|_*), \omega(\|z - x^*\|_*) \}, \tag{2.5}$$

and the norms which may depend on x, $\bar{x}$, and N satisfy

$$\| A \|_M \leqq [1 + m(x,\bar{x})] \, \| A \|_{M_+} \tag{2.6}$$

for all $A \in R^{n\times n}$. Before stating our main results we need a technical lemma.

Lemma 2.1: Let α, $\varepsilon > 0$ and $0 < t < 1$ be given. Assume $0 \leqq \rho_0 \leqq \varepsilon$ and

$$0 \leqq \rho_{i+1} \leqq [1 + \alpha\, \omega(t^i \varepsilon)] \, \rho_i + \alpha\, \omega(t^i \varepsilon), \quad 0 \leqq i < k, \tag{2.7}$$

where $\omega(\delta)$ satisfies

(i) ω is continuous and monotone increasing for $\delta > 0$;

(ii) $\omega(0) = 0$;

$$(iii) \int_0^{\varepsilon/t} \frac{\omega(\delta)}{\delta}\, d\delta < +\infty. \tag{2.8}$$

Then[1]

$$\rho_k \leqq (1+\varepsilon) \exp \left(\frac{\alpha}{1-t} \int_0^{\varepsilon/t} \frac{\omega(\delta)}{\delta}\, d\delta \right) - 1 \equiv W(\varepsilon). \tag{2.9}$$

Proof: Let $\omega_i = \omega(t^i\varepsilon)$. By an easy induction argument[2],

$$\rho_k \leqq \rho_0 \prod_{i=0}^{k-1}(1+\alpha\omega_i) + \sum_{i=0}^{k-1}\alpha\omega_i \prod_{j=i+1}^{k-1}(1+\alpha\omega_j). \tag{2.10}$$

Since $1+\alpha\omega \leqq \exp(\alpha\omega)$ for $\alpha\omega \geqq 0$, we have from (2.10)

$$\begin{aligned} \rho_k &\leqq \rho_0 \prod_{i=0}^{k-1}\exp(\alpha\omega_i) + \sum_{i=0}^{k-1}[\exp(\alpha\omega_i) - 1] \prod_{j=i+1}^{k-1}\exp(\alpha\omega_j). \\ &= \exp(\alpha\sum_{i=0}^{k-1}\omega_i)\,\rho_0 + \sum_{i=0}^{k-1}\{\exp(\alpha\sum_{j=i}^{k-1}\omega_j) - \exp(\alpha\sum_{j=i+1}^{k-1}\omega_j)\} \\ &= \exp(\alpha\sum_{i=0}^{k-1}\omega_i)\,\rho_0 + \{\exp(\alpha\sum_{i=0}^{k-1}\omega_i) - 1\}. \end{aligned} \tag{2.11}$$

But

$$\sum_{i=0}^{k-1}\omega(t^i\varepsilon) = \frac{1}{1-t}\sum_{i=0}^{k-1}\frac{\omega(t^i\varepsilon)}{t^{i-1}\varepsilon}(t^{i-1}\varepsilon - t^i\varepsilon) \leqq \frac{1}{1-t}\int_{t^{k-1}\varepsilon}^{\varepsilon/t}\frac{\omega(\delta)}{\delta}\, d\delta, \tag{2.12}$$

using that the sum is a lower Rieman sum for the integral. Further, since the integrand is non-negative we have

$$\sum_{i=0}^{k-1}\omega(t^i\varepsilon) \leqq \frac{1}{1-t}\int_0^{\varepsilon/t}\frac{\omega(\delta)}{\delta}\, d\delta,$$

using (2.12) and (2.8). From (2.11) we have

$$\rho_k \leqq (1+\rho_0)\exp\left(\alpha\sum_{i=0}^{k-1}\omega_i\right) - 1 \leqq W(\varepsilon)$$

since $\rho_0 \leqq \varepsilon$ and we have shown (2.9).

Q.E.D.

Theorem 2.2: Let U be of bounded deterioration for B, with respect to F' at x^* where the

[1] Let $z \in \mathbf{R}$ then $\exp(z) = e^z$.

[2] The convention $\prod_{j=k}^{k-1} = 1$ is understood.

matrix B_* is nonsingular and satisfies

$$||I - F'(x^*)B_*^{-1}|| \leq r^* < 1. \tag{2.13}$$

Consider the iteration (1.7) where the residual satisfies (1.6). Let

$$\theta_k \leq \theta < \frac{1 - r^*}{1 + r^*}$$

and choose $B_{k+1} \in \mathrm{U}(x_k, x_{k+1}, B_k)$ where the sequence of norms $\{||\cdot||_{M_k}\}$ satisfies

$$\frac{1}{1 + m(x_k, x_{k+1})} ||\cdot||_{M_k} \leq ||\cdot||_{M_{k+1}}.$$

For any t which satisfies $\theta(1+r^*) + r^* < t < 1$ there exists $\varepsilon > 0$ so that if

$$||x_0 - x^*||_* \leq \varepsilon, \quad ||B_0 - B_*||_{M_0} \leq \varepsilon$$

then $\{x_k\}$ is defined for any method in the inexact quasi-Newton framework, $x_k \to x^*$ as $k \to \infty$, and

$$||x_{k+1} - x^*||_* \leq t\, ||x_k - x^*||_* . \tag{2.14}$$

Moreover $\{||B_k||\}$ and $\{||B_k^{-1}||\}$ are bounded.

Proof: Let $\mu \equiv \max \{ ||F'(x^*)||, ||B_*||, ||B_*^{-1}|| \}$ then

$$\frac{1}{\mu} ||y|| \leq ||y||_* \leq \mu\, ||y|| \quad \forall\, y \in \mathbf{R}^n . \tag{2.15}$$

Let $\beta > 0$ be such that

$$||B|| \leq \beta\, ||B||_{M_0} \quad \forall\, B \in \mathbf{R}^{n \times n} \tag{2.16}$$

and define for $\varepsilon \geq 0$

$$E(\varepsilon) \equiv \exp\left(\frac{\alpha}{1-t} \int_0^{\varepsilon/t} \frac{\omega(\delta)}{\delta}\, d\delta\right) \tag{2.17}$$

and

$$W(\varepsilon) \equiv (1+\varepsilon)\, E(\varepsilon) - 1 . \tag{2.18}$$

Note from the assumption (1.3) that F' is continuous at x^* and W is continuous and monotone increasing with $W(0)=0$ and $E(0)=1$. Define

$$R(\varepsilon) \equiv \frac{\mu^2 \beta E(\varepsilon) W(\varepsilon)}{1 - \mu\beta E(\varepsilon) W(\varepsilon)} \tag{2.19}$$

and

$$T(\varepsilon) \equiv r^* + 2\,\mu\, R(\varepsilon) + [1+2R(\varepsilon)]\,\{\,\beta\mu\omega(\varepsilon)+\theta[1+r^*+\mu\beta\omega(\varepsilon)]\} \tag{2.20}$$

Choose $\varepsilon > 0$ sufficiently small that

$$W(\varepsilon) < \frac{1}{\mu\beta E(\varepsilon)} \quad \text{and} \quad T(\varepsilon) < t.$$

This is possible since W, R, and T are continuous and $W(0)=0<1/(\mu\beta)$, $R(0)=0$, $T(0)=r^*+\theta(1+r^*) < t$. If necessary further restrict ε so that $(x,\overline{x},B) \in \mathbf{N}$ whenever $\|B-B_*\| \leqq \beta E(\varepsilon)W(\varepsilon)$, $\|x-x^*\|_* \leqq \varepsilon$, and $\|\overline{x}-x^*\|_* \leqq \varepsilon$.

The proof is by induction. Assume that

$$\|x_i-x^*\|_* \leqq t^i\|x_0-x^*\|_* \leqq t^i\varepsilon, \quad 0 \leqq i \leqq k. \tag{2.21}$$

Then, from (2.5)

$$m(x_i,x_{i+1}) \leqq \alpha \max\{\,\omega(t^i\varepsilon),\, \omega(t^{i+1}\varepsilon)\,\} \leqq \alpha\, \omega(t^i\varepsilon), \quad 0 \leqq i < k$$

since ω is monotone increasing. By the choice of B_{i+1},

$$\|B_{i+1}-B_*\|_{M_{i+1}} \leqq [1 + \alpha\, \omega(t^i\varepsilon)]\, \|B_i-B_*\|_{M_i} + \alpha\, \omega(t^i\varepsilon) \quad 0 \leqq i < k.$$

Let $\rho_i = \|B_i-B_*\|_{M_i}$. Then from Lemma 2.1 and (2.18)

$$\|B_k-B_*\|_{M_k} \leqq W(\varepsilon) \quad \text{for } k \geqq 0.$$

The above inequality holds for $k=0$ since $W(\varepsilon)>\varepsilon$. From (2.6) and by (2.16) we have

$$\begin{aligned}
\|B_k-B_*\| &\leqq \beta\, \|B_k-B_*\|_{M_0} \\
&\leqq \beta \prod_{i=0}^{k-1}(1+\alpha\omega(t^i\varepsilon)\,)\, \|B_k-B_*\|_{M_k} \\
&\leqq \beta\, E(\varepsilon)W(\varepsilon)\,.
\end{aligned} \tag{2.22}$$

By the Banach Perturbation Lemma (c.f. Ortega and Rheinboldt [6, 2.3.3]), B_k is invertible and

$$\begin{aligned}
\|B_k^{-1} - B_*^{-1}\| &\leqq \frac{\|B_*^{-1}\|^2\, \|B_k-B_*\|}{1 - \|B_*^{-1}\|\, \|B_k-B_*\|} \\
&\leqq \frac{\mu^2\beta E(\varepsilon)W(\varepsilon)}{1-\mu\beta E(\varepsilon)W(\varepsilon)} = R(\varepsilon),
\end{aligned} \tag{2.23}$$

using (2.22) and (2.19). Therefore it is possible to find some s_k in (1.6). Further

$$\|I - F'(x^*)B_k^{-1}\| \leqq \|I - F'(x^*)B_*^{-1}\| + \|F'(x^*)\|\, \|B_k^{-1} - B_*^{-1}\|$$

$$\leq r^* + \mu \, ||B_k^{-1} - B_*^{-1}||$$

$$\leq r^* + \mu \, R(\varepsilon), \tag{2.24}$$

using (2.13) and (2.23). Since

$$B_*(x_{k+1}-x^*) = [I - F'(x^*)B_*^{-1} + B_*(B_*^{-1}-B_k^{-1})\, F'(x^*)B_*^{-1}]B_*(x_k-x^*)$$
$$+ [I - B_*(B_*^{-1}-B_k^{-1})] \, \{ [B_k s_k+F(x_k)] - [F(x_k)-F(x^*)-F'(x^*)(x_k-x^*)] \} .$$

Taking norms,

$$||x_{k+1}-x^*||_* \leq ||I - F'(x^*)B_k^{-1}|| \, ||x_k-x^*||_*$$
$$+ \; ||B_*|| \, (||I-F'(x^*)B_*^{-1}||+1) \, ||B_*^{-1}-B_k^{-1}|| \, ||x_k-x^*||_*$$
$$+ \; [1+||B_*|| \, ||B_*^{-1}-B_k^{-1}||] \, \{ ||F(x_k)-F(x^*)-F'(x^*)(x_k-x^*)||+||B_k s_k+F(x_k)|| \} . \tag{2.25}$$

By Ortega and Rheinboldt [6, 3.2.5],

$$||F(x_k)-F(x^*)-F'(x^*)(x_k-x^*)|| \leq \sup_{0\leq\tau\leq 1} ||F'(x^*+\tau(x_k-x^*))-F'(x^*)|| \, ||x_k-x^*||$$
$$\leq \beta \, \omega(||x_k-x^*||_*) \, ||x_k-x^*||$$
$$\leq \beta \, \mu \, \omega(\varepsilon) \, ||x_k-x^*||_* \tag{2.26}$$

using (2.16), (2.1) and (2.15). By the choice of s_k

$$||B_k s_k+F(x_k)|| \leq \theta_k ||F(x_k)|| \leq \theta \, ||F(x_k)|| \tag{2.27}$$

Since

$$F(x_k) = B_*(x_k-x^*) - [I-F'(x^*)B_*^{-1}]B_*(x_k-x^*) + [F(x_k)-F(x^*)-F'(x^*)(x_k-x^*)].$$

Taking norms

$$||F(x_k)|| \leq ||x_k-x^*||_* + ||I-F'(x^*)B_*^{-1}|| \, ||x_k-x^*||_* + ||F(x_k)-F(x^*)-F'(x^*)(x_k-x^*)||$$
$$\leq [1+r^*+ \beta \, \mu\omega(\varepsilon)] \, ||x_k-x^*||_* . \tag{2.28}$$

using (2.13) and (2.26). Therefore, by (2.25)

$$||x_{k+1}-x^*||_* \leq (r^*+\mu R(\varepsilon))||x_k-x^*||_* + \mu \, 2 \, R(\varepsilon)||x_k-x^*||_*$$
$$+ (1+\mu R(\varepsilon))[\beta\mu\omega(\varepsilon) + \theta(1+r^*+\beta \, \mu \, \omega(\varepsilon))] \, ||x_k-x^*||_*$$
$$\leq T(\varepsilon) \, ||x_k-x^*||_* $$

using (2.24), (2.13), (2.23), (2.26), (2.27), (2.28) and (2.20). By the choice of t and we have (2.14). Further

$$\|x_{k+1}-x^*\|_* \leq t\ \|x_k-x^*\|_* \leq t^{k+1}\ \|x_0-x^*\|_*$$

and we have shown (2.21) for $i=k+1$. The uniform boundedness of $\|B_k\|$, and $\|B_k^{-1}\|$ are immediate.

Q.E.D.

We say that F' is Holder continuous with exponent p $(0<p\leq 1)$ at x^* if there exists $L\geq 0$ such that

$$\|F'(y) - F'(x^*)\| \leq L\ \|y-x^*\|^p$$

for $\|y-x^*\|$ sufficiently small. It follows directly that if F' is Holder continuous with exponent p $(0<p\leq 1)$ at x^* then the modulus of continuity ω satisfies $\omega(\delta) \leq L\ \delta^p$ so

$$\int_0^{\varepsilon} \frac{\omega(\delta)}{\delta}\, d\delta \leq \frac{L}{p}\ \varepsilon^p$$

and (1.3) holds. In the following discussion, we will assume that F' is Holder continuous. In the following, we also assume that the update function is of bounded deterioration using a fixed norm in (2.4). Broyden Dennis and More [1] showed that the quasi-Newton method is locally Q-linearly convergent when $B_*=F'(x^*)$. This follows from Theorem 2.2 by choosing $\theta=r^*=0$. Dennis and Walker [4] showed local Q-linear convergence of $\{x_k\}$ of the quasi-Newton method. This follows from Theorem 2.2 by choosing $\theta=0$. Steihaug [7] showed that the inexact quasi-Newton method is locally and Q-linearly convergent when $B_*=F'(x^*)$ and $\theta<1$.

3. Characterization of Superlinear Convergence

In this section, we characterize superlinear convergence of methods in the inexact quasi-Newton framework in terms of the relative residual. We will in this section also use that[3] $x_k\to x^*$ as $k\to\infty$. We need the following lemma [2].

Lemma 3.1: There is an $\varepsilon>0$ and a $\rho\geq 1$ such that for $\|x-x^*\|\leq\varepsilon$ then

$$\frac{1}{\rho}\ \|x - x^*\| \leq \|F(x)\| \leq \rho\ \|x - x^*\|. \tag{3.1}$$

□

[3]For the analysis in this section, F' continuous at x^* would suffice in A.4.

We say that the sequence $\{x_k\}$ is superlinearly convergent to x^* if

$$\lim_{k\to\infty} \frac{||x_{k+1}-x^*||}{||x_k-x^*||} = 0. \tag{3.2}$$

Let $s_k = x_{k+1}-x_k$. The residual is $r_k = B_k s_k + F(x_k)$ The next condition states that the relative residual goes to 0 as $k\to\infty$.

$$\lim_{k\to\infty} \frac{||B_k s_k + F(x_k)||}{||F(x_k)||} = 0. \tag{3.3}$$

If $\theta_k \to 0$ in (1.6) as $k\to\infty$ then we have (3.3) If B_k is the Jacobian matrix of F at x_k then $B_k s_k + F(x_k) = F(x_k+s_k) + o(||s_k||)$. Condition (3.4) states that this should hold for approximations of the Jacobian matrix and it implies that the residual approximates the function at the new point.

$$\lim_{k\to\infty} \frac{||B_k s_k + F(x_k) - F(x_k+s_k)||}{||s_k||} = 0. \tag{3.4}$$

Theorem 3.2: Let $x_k \to x^*$ as $k\to\infty$. Any pair of the conditions (3.1), (3.2), and (3.3) implies the third.

Proof: Let ε be as in Lemma 3.1 and choose index k_0 so that for $k \geqq k_0$ then $||x_k - x^*|| \leqq \varepsilon$. Put

$$\gamma_k = \frac{||B_k s_k + F(x_k) - F(x_{k+1})||}{||s_k||}.$$

Consider

$$B_k s_k + F(x_k) = [B_k s_k + F(x_k) - F(x_{k+1})] + F(x_{k+1}).$$

Taking norm

$$\frac{||B_k s_k + F(x_k)||}{||F(x_k)||} \leqq \gamma_k \frac{||x_{k+1}-x_k||}{||F(x_k)||} + \frac{||F(x_{k+1})||}{||F(x_k)||}$$

$$\leqq \gamma_k \rho \left[\frac{||x_{k+1}-x^*||}{||x_k-x^*||} + 1 \right] + \rho^2 \frac{||x_{k+1}-x^*||}{||x_k-x^*||},$$

using the triangle inequality and (3.1). Assume (3.2) and (3.4), then (3.3) holds.

Assume (3.3) and (3.4) and consider

$$\frac{||x_{k+1}-x^*||}{||x_k-x^*||} \leqq \rho \frac{||F(x_{k+1})||}{||x_k-x^*||}$$

$$\leq \rho \left[\frac{\|B_k s_k + F(x_k)\|}{\|F(x_k)\|} \frac{\|F(x_k)\|}{\|x_k - x^*\|} + \gamma_k \frac{\|s_k\|}{\|x_k - x^*\|} \right]$$

$$\leq \rho \left[\rho \frac{\|B_k s_k + F(x_k)\|}{\|F(x_k)\|} + \gamma_k \left(\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} + 1 \right) \right]$$

Hence from (3.3) and (3.4) we have (3.2).

Assume (3.2) and (3.3). Then

$$\frac{\|B_k s_k + F(x_k) - F(x_k + s_k)\|}{\|s_k\|} \leq \left[\frac{\|B_k s_k + F(x_k)\|}{\|F(x_k)\|} + \frac{\|F(x_k + s_k)\|}{\|F(x_k)\|} \right] \frac{\|F(x_k)\|}{\|s_k\|}. \tag{3.5}$$

From Lemma 3.1 we have

$$\frac{\|F(x_k + s_k)\|}{\|F(x_k)\|} \leq \rho^2 \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|}.$$

Choose index $k_1 \geq k_0$ so that

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq \frac{1}{2} \text{ for } k \geq k_1.$$

Then from Lemma 3.1

$$\frac{\|F(x_k)\|}{\|s_k\|} \leq \rho \frac{\|x_k - x^*\|}{\|x_{k+1} - x_k\|} \leq \rho \frac{\|x_k - x^*\|}{\|x_k - x^*\| - \|x_{k+1} - x^*\|}$$

$$\leq \rho \frac{1}{1 - \dfrac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|}} \leq 2\rho.$$

From (3.5) we have

$$\frac{\|B_k s_k + F(x_k) - F(x_k + s_k)\|}{\|s_k\|} \leq 2\rho \left[\frac{\|B_k s_k + F(x_k)\|}{\|F(x_k)\|} + \rho^2 \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \right].$$

so from (3.2) and (3.3) we have (3.4).

Q.E.D.

Theorem 3.2 summerizes the following known results.

Corollary 3.3: (Dennis and More [3])Let $x_k \to x^*$ as $k \to \infty$ and consider the iteration $x_{k+1} = x_k + s_k$ where s_k satisfies $B_k s_k + F(x_k) = 0$ Then $x_k \to x^*$ superlinearly if and only if

$$\lim_{k \to \infty} \frac{\|(B_k - F'(x^*)) s_k\|}{\|s_k\|} = 0. \tag{3.6}$$

Proof From Ortega and Rheinboldt [6, 3.2.5] we have

$$\lim_{k\to\infty} \frac{\|F(x_{k+1}) - F(x_k) - F'(x^*)s_k\|}{\|s_k\|} = 0. \tag{3.7}$$

Hence, (3.4) and (3.6) are equivalent. From the choice of s_k, (3.3) holds.

Q.E.D.

Corollary 3.4: (Dembo, Eisenstat and Steihaug [2]) If $B_k = F'(x_k)$ then (3.2) and (3.3) are equivalent.

Proof From Ortega and Rheinboldt [6, 3.2.5] we have that (3.4) holds. Hence Theorem 3.2 shows that (3.2) and (3.3) are equivalent.

Q.E.D.

Corollary 3.5: (Steihaug [7]) If (3.4) holds then (3.1') and (3.2) are equivalent.

□

4. Applications: Least Change Secant Updates.

In the last years there have been a growing interest in generating sparse matrix updates. Most of these methods are based on the least change technique. In this section, we will show that these updates satisfy the conditions for local and superlinear rate of convergence. In addition to the assumptions A.1 to A.4 we will assume that

A.5: F' is continuous in Ω.

To incorporate Jacobian information in the new update B_{k+1}, it is common to require that the update satisfies the secant equation

$$B_+ s = F(x+s) - F(x) \equiv y. \tag{4.1}$$

Let the set of matrices that satisfies (4.1) be

$$\mathbf{Q}(y,s) = \{ B \in \mathbf{R}^{n\times n} : Bs = y \}. \tag{4.2}$$

If the Jacobian matrix is sparse we will also require that the update is sparse, and if the Jacobian has other structural properties like symmetry then the updates will be required to have these properties. The set of symmetric matrices, the set of sparse matrices, etc., are subspaces of $\mathbf{R}^{n\times n}$. Let $\mathbf{A}$ be a subspace of $\mathbf{R}^{n\times n}$ and

$$\mathbf{V} = \mathbf{A} \cap \mathbf{Q}(y,s). \tag{4.3}$$

For given $B\in \mathbf{R}^{n\times n}$, $s\in \mathbf{R}^n$ and y in (4.1) the least change secant update in Frobenius norm is the minimizer B_+ of

$$\min \{ \|\tilde{B}_+ - B\|_F : \tilde{B}_+ \in \mathbf{V} \}. \tag{4.4}$$

Lemma 4.1: Let $J \in \mathbf{V}$ for $\mathbf{V}$ in (4.3) and let B_+ be the least change secant update. Then

$$\|B_+ - J\|_F^2 \le \|B - J\|_F^2 - \left[\frac{\|Bs - y\|_2}{\|s\|_2} \right]^2. \tag{4.5}$$

Proof: Let $C, D \in \mathbf{V}$ then $C - D \in \mathbf{A}$ so $\mathbf{V}$ is an affine space or linear manifold [5]. Further,

$$\|B_+ - J\|_F^2 + \|B_+ - B\|_F^2 = \|B - J\|_F^2 \tag{4.6}$$

since the Frobenius norm is the Euclidean norm on $\mathbf{R}^{n \times n}$. The Frobenius norm is consistent with the Euclidean norm[4] and $B_+ \in \mathbf{Q}(y,s)$ so

$$\|Bs - y\|_2 = \|Bs - B_+ s\|_2 \le \|B_+ - B\|_F \|s\|_2 .$$

From (4.6) we have the desired result (4.5).

Q.E.D.

Let

$$\omega(\delta) \equiv \sup \{ \|F'(x) - F'(x^*)\|_F : \|x - x^*\|_* \le \delta \}$$

where $\|y\|_* = \|F'(x^*)y\|$.

Lemma 4.2: Assume that $F'(x) \in \mathbf{A}$ for all $x \in \Omega$. There exists $\varepsilon > 0$ so that for $\|x - x^*\|_* \le \varepsilon$, $\|x + s - x^*\|_* \le \varepsilon$, and $y = F'(x+s) - F(x)$ we have $J \in \mathbf{A} \cap \mathbf{Q}(y,s)$ where

$$J_{ij} = \int_0^1 F'(x + \tau s)_{ij} \, d\tau \quad 1 \le i,j \le n. \tag{4.7}$$

and

$$\|J - F'(x^*)\|_F \le \max \{ \omega(\|x+s-x^*\|_*), \omega(\|x - x^*\|_*) \} \tag{4.8}$$

Proof: The existence of J follows from Ortega and Rheinboldt [6, 3.2.7]. From [6,3.2.11] we have

[4] We say that the matrix norm $\| \ \|$ and the vector norm $| \ |$ is consistent if for all matrices A and vectors x we have $|Ax| \le \|A\| \, |x|$.

$$\begin{aligned}
\|\int_0^1 F'(x+\tau s)d\tau - F'(x^*)\|_F &\leq \int_0^1 \|F'(x+\tau s) - F'(x^*)\|_F \, d\tau \\
&\leq \sup_{0\leq\tau\leq 1} \|F'(x+\tau s) - F'(x^*)\|_F \\
&\leq \sup_{0\leq\tau\leq 1} \omega(\|x + \tau s - x^*\|_*) \\
&\leq \omega(\max \{ \|x+s-x^*\|_*, \|x-x^*\|_* \}) \\
&\leq \max \{ \omega(\|x+s-x^*\|_*), \omega(\|x-x^*\|_*) \},
\end{aligned}$$

using (1.2), the convexity of the norm, and the monotonicity of ω.

Q.E.D.

We now combine Lemma 4.1 and 4.2 to show that the least change secant update is of bounded deterioration for $F'(x^*)$ at x^*.

Lemma 4.3: Assume that $F'(x) \in \mathbf{A}$ for all $x \in \Omega$. There exists $\varepsilon > 0$ so that for $\|x-x^*\|_* \leq \varepsilon$, $\|\bar{x}-x^*\|_* \leq \varepsilon$, and $\mathbf{V}$ in(4.3) the least change update B_+ in (4.4) satisfies

$$\|B_+ - F'(x^*)\|_F \leq \|B - F'(x^*)\|_F + m(x,\bar{x})$$

where $s = \bar{x}-x$, $y = F(\bar{x})-F(x)$, and

$$m(x,z) = 2 \max \{ \omega(\|x-x^*\|_*), \omega(\|z-x^*\|_*) \}.$$

Proof: From Lemma 4.2 we have $J \in \mathbf{Q}(y,s) \cap \mathbf{A}$. Consider

$$\begin{aligned}
\|B_+ - F'(x^*)\|_F &\leq \|B_+ - J\|_F + \|J - F'(x^*)\|_F \\
&\leq \|B - J\|_F + \|J - F'(x^*)\|_F \\
&\leq \|B - F'(x^*)\|_F + 2\, \|J - F'(x^*)\|_F \\
&\leq \|B - F'(x^*)\|_F + m(x, \bar{x})
\end{aligned}$$

using the triangle inequality, (4.5), and (4.8).

Q.E.D.

For x_k and s_k let

$$y_k = F(x_k+s_k) - F(x_k), \tag{4.9}$$

and

$$J_k = \int_0^1 F'(x_k+\tau(x_{k+1}-x_k)) \, d\tau. \tag{4.10}$$

Theorem 4.4: Assume that $F'(x) \in \mathbf{A}$ for all $x \in \Omega$ and assume $B_{k+1} = (B_k)_+$ satisfies (4.5) using (4.9) and (4.10). Let $\theta_k \leqq \theta < t < 1$. There exists $\varepsilon > 0$ so that if

$$\|x_0 - x^*\|_* \leqq \varepsilon \text{ , and } \|B_0 - F'(x^*)\|_F \leqq \varepsilon \text{ ,}$$

then $\{x_k\}$ is defined for any method in the inexact quasi-Newton framework, $x_k \to x^*$ as $k \to \infty$, and

$$\|x_{k+1} - x^*\|_* \leqq t \, \|x_k - x^*\|_*, \tag{4.11}$$

where $\|y\|_* = \|F'(x^*)y\|$ and

$$\lim_{k \to \infty} \frac{\|y_k - B_k s_k\|}{\|s_k\|} = 0 \, . \tag{4.12}$$

If $\theta_k \to 0$ as $k \to \infty$ then $\{x_k\}$ converges Q-superlinearly.

Proof: The convergence of $\{x_k\}$ and (4.11) follows from Lemma 4.3 and Theorem 2.2 with $B_* = F'(x^*)$, $r^* = 0$ and $\|\cdot\|_{M_k} = \|\cdot\|_F$.

We will now show that (4.12) holds. Let

$$\rho_k = \|B_k - F'(x^*)\|_F.$$

then from the proof of Theorem 2.2 $\rho_k \leqq W(\varepsilon)$. From (4.11) we have $\omega(x_{k+1}, x_k) \leqq w(t^k \varepsilon) \leqq \omega(\varepsilon)$ using the monotonicity of ω. Consider

$$\|B_k - J_k\|_F \leqq \rho_k + \|J_k - F'(x^*)\|_F \leqq W(\varepsilon) + m(x_{k+1}, x_k) \leqq W(\varepsilon) + \omega(\varepsilon) \tag{4.13}$$

using Lemma 2.1 and (4.8). Further let

$$\gamma_k = \frac{\|y_k - B_k s_k\|_2}{\|s_k\|_2} \, .$$

Then

$$\begin{aligned}
\rho_{k+1}^2 &\leqq (\|B_{k+1} - J_k\|_F + \|J_k - F'(x^*)\|_F)^2 \\
&\leqq \|B_{k+1} - J_k\|_F^2 + 2 \, \|B_{k+1} - J_k\|_F \, \omega(t^k \varepsilon) + \omega(\varepsilon) \, \omega(t^k \varepsilon) \\
&\leqq \|B_{k+1} - J_k\|_F^2 + 2 \, (W(\varepsilon) + \omega(\varepsilon)) \, \omega(t^k \varepsilon) + \omega(\varepsilon) \, \omega(t^k \varepsilon) \\
&\leqq \|B_k - J_k\|_F^2 - \gamma_k^2 + (2 \, W(\varepsilon) + 3\omega(\varepsilon)) \, \omega(t^k \varepsilon) \\
&\leqq \rho_k^2 - \gamma_k^2 + 4 \, (W(\varepsilon) + \omega(\varepsilon)) \, \omega(t^k \varepsilon)
\end{aligned}$$

Hence

$$\sum_{i=0}^{k} \gamma_i^2 \leqq \varepsilon^2 - \rho_{k+1}^2 + 4 \, (W(\varepsilon) + \omega(\varepsilon)) \, W(\varepsilon)$$

using (2.12) and $\rho_0 \leqq \varepsilon$. From (4.13) we have $\{\rho_j\}$ is bounded, hence

$$\sum_{i=0}^{\infty}\gamma_i^2 < +\infty ,$$

and (4.12) follows directly using that all norms on $\mathbf{R}^n$ are equivalent. From Theorem 3.2 and (4.9) we have that (3.3) and (3.4) holds, hence the sequence $\{x_k\}$ converges Q-superlinearly to x^*.

Q.E.D.

5. Applications: Iteratively Rescaled Least Change Updates.

An important nonlinear problem is the unconstrained optimization problem. Namely the problem of finding a local minimizer x^* of a function $f:\mathbf{R}^n \to \mathbf{R}$

$$f(x^*) \leq f(x) \text{ for all } x: \|x - x^*\| < \varepsilon.$$

for some $\varepsilon > 0$. If f is continuously differentiable in an open neighborhood Ω of x^* then x^* satisfies (1.1) where $F \equiv \nabla f$. And if f is twice continously differentiable and strictly convex in Ω, then the Jacobian matrix of F (the Hessian matrix of f) is symmetric and positive definite. In this section, we assume that the function f has the following properties. The Hessian matrix of f is denoted H, i.e. $H \equiv \nabla^2 f$.

A.1: f is twice continuously differentiable in an open neighborhood Ω of x^*;

A.2: x^* satisfies $\nabla f(x^*) = 0$;

A.3: $H(x^*)$ is positive definite;

Let W be a positive definite matrix and consider the matrix norm

$$\|\cdot\|_{W,F} = \|W^{-\frac{1}{2}} \cdot W^{-\frac{1}{2}}\|_F , \tag{5.1}$$

and the vector norm $\|y\|_* = \|H(x^*)y\|$. Let

$$\omega(\delta) \equiv \sup \{\|H(x) - H(x^*)\|_{H(x^*),F} : \|x - x^*\|_* \leq \delta \}. \tag{5.2}$$

A.4: There exists $\varepsilon > 0$ so that the modulus of continuity ω of H at x^* satisfies

$$\int_0^{\varepsilon} \frac{\omega(\delta)}{\delta}\, d\delta \;<\; +\infty. \tag{5.3}$$

Consider the scaled function

$$\overline{f}(\overline{x}) \;=\; f(W^{-1/2}\,\overline{x})$$

induced by the scaling $\overline{x} = W^{1/2}x$ which makes

$$\nabla\overline{f}(\overline{x}) \;=\; W^{-1/2}\nabla f(W^{-1/2}\,\overline{x})$$

and

$$\overline{H}(\overline{x}) \;=\; W^{-1/2}\,H(W^{-1/2}\,\overline{x})\,W^{-1/2}.$$

The natural candidates for the scaling matrices are those that make the condition number of the scaled problem small, ie. we want $W \simeq H$. If we do a least change secant update in Frobenius norm (4.4) in the transformed variables, then in the untransformed variables this is equivalent to

$$\min\,\{\|\widetilde{B}_+ - B\|_{W,F} : \widetilde{B}_+ \in \mathbf{V}\,\} \tag{5.4}$$

where

$$\mathbf{V} \;=\; \mathbf{A} \cap \mathbf{Q}(y,s). \tag{5.5}$$

for $\mathbf{Q}(y,s)$ defined in (4.2) with

$$y \;\equiv\; \nabla f(x+s) - \nabla f(x)\,. \tag{5.6}$$

The corresponding result of Lemma 4.1 in the weighted Frobenius norm is:

Lemma 5.1: Let $J \in \mathbf{V}$ for $\mathbf{V}$ in (5.5) and let B_+ be the least change update in the $\|\cdot\|_{W,F}$ norm. Then

$$\|B_+ - J\|^2_{W,F} \;\leq\; \|B - J\|^2_{W,F} \;-\; \left[\frac{\|W^{-1/2}(Bs - y)\|_2}{\|W^{1/2}s\|_2}\right]^2. \tag{5.7}$$

Proof: Just note that[5]

$$\|W^{-1/2}(Bs - y)\|_2 \;=\; \|W^{-1/2}\,(B - B_+)\,W^{-1/2}\,(W^{1/2}s)\|_2 \;\leq\; \|B_+ - B\|_{W,F}\,\|W^{1/2}s\|_2\,.$$

Q.E.D.

We now relate the matrix norms $\|\cdot\|_{W_+,F}$ and $\|\cdot\|_{W,F}$ where W_+ and W are positive definite matrices. First we need a technical lemma.

---

[5] This is the only modification of the proof of Lemma 4.1

Lemma 5.2: Let C be a symmetric n by n matrix and $\|C\|_2 < 1$. Then

$$\| (I + C)^{-1/2} - I \|_2 \leqq \frac{\|C\|_2}{1 - \|C\|_2}$$

Proof: Let $\|\cdot\|$ be the Euclidean norm. Note that I+C is nonsingular and

$$I - (I + C)^{-1} = C(I + C)^{-1}$$

so

$$\| I - (I + C)^{-1} \| \leqq \|C\| \, \|(I + C)^{-1}\| \leqq \frac{\|C\|}{1 - \|C\|}$$

Further, since $I+C$ is symmetric and positive definite, $(I+C)^{-1/2}$ is positive definite and we have

$$\| [I + (I + C)^{-1/2}]^{-1} \| \leqq 1$$

Hence

$$\begin{aligned} \|(I + C)^{-1/2} - I\| &\leqq \|(I + C)^{-1} - I\| \, \| [I + (I + C)^{-1/2}]^{-1}\| \\ &\leqq \frac{\|C\|}{1 - \|C\|} \end{aligned}$$

Q.E.D.

We will use Lemma 5.2 to show that for positive definite matrices W and W_+ that are close, then for any matrix A $\|A\|_{W,F}$ and $\|A\|_{W_+,F}$ are close.

Lemma 5.3: Let W and W_+ be symmetric and positive definite n by n matrices and put

$$\rho = \max \{ \|W_+^{-1}\|_2 , \|W^{-1}\|_2 \} \, \|W_+ - W\|_2 . \tag{5.8}$$

If $\rho \leqq 1/2$ then for any real n by n matrix A

$$\frac{1}{1 + 6\rho} \|A\|_{W,F} \leqq \|A\|_{W_+,F} \leqq (1 + 6\rho) \|A\|_{W,F} . \tag{5.9}$$

Proof: In the proof, we will frequently use the following inequality. Let A and B be real n by n matrices. Then

$$\| A B \|_F \leqq \min \{ \|A\|_2 \|B\|_F , \|A\|_F \|B\|_2 \} . \tag{5.10}$$

Let $\bar{A} = W^{-1/2} A W^{-1/2}$ and consider

$$\begin{aligned} &\|W_+^{-1/2} A W_+^{-1/2} - W^{-1/2} A W^{-1/2}\|_F \\ &\qquad \leqq \|W_+^{-1/2} A (W_+^{-1/2} - W^{-1/2})\|_F + \|(W_+^{-1/2} - W^{-1/2}) A W^{-1/2}\|_F \end{aligned}$$

$$= \|W_+^{-1/2}W^{1/2}\bar{A}(W^{1/2}W_+^{-1/2}-I)\|_F + \|(W_+^{-1/2}W^{1/2}-I)\bar{A}\|_F \tag{5.11}$$

using the triangle inequality. Consider

$$\|(W_+^{-1/2}W^{1/2}-I)\bar{A}\|_F \leqq \|\bar{A}\|_F \|W_+^{-1/2}W^{1/2}-I\|_2 = \|\bar{A}\|_F \|I-W^{1/2}W_+^{-1/2}\|_2 \tag{5.12}$$

using (5.10) and $\|B^T\|_2 = \|B\|_2$ for any matrix B. Further

$$\|W_+^{-1/2}W^{1/2}\bar{A}(W^{1/2}W_+^{-1/2}-I)\|_F \leqq$$

$$\|(W_+^{-1/2}W^{1/2}-I)\bar{A}\|_F \|(W^{1/2}W_+^{-1/2}-I)\|_F + \|\bar{A}(W^{1/2}W_+^{-1/2}-I)\|_F$$

$$\leqq \|\bar{A}\|_F (1 + \|W^{1/2}W_+^{-1/2}-I\|_2)\|W^{1/2}W_+^{-1/2}-I\|_2$$

using (5.10) and (5.12) Let $W_+=W+E$. Then

$$W_+ = W^{1/2}(I + W^{-1/2}E\, W^{-1/2})W^{1/2} \text{ and } W_+^{-1/2} = W^{-1/4}(I + W^{-1/2}E\, W^{-1/2})^{-1/2}W^{-1/4}$$

Let $C = W^{-1/2}E\, W^{-1/2}$. Then

$$\|C\|_2 = \|W^{-1/2}E\, W^{-1/2}\|_2 \leqq \|W^{-1}\|_2 \|W_+-W\|_2 \leqq \rho < 1$$

and

$$\|W^{1/2}W_+^{-1/2}-I\|_2 = \|W^{1/4}W_+^{-1/2}W^{1/4} - I\|_2$$

$$= \|(I+C)^{-1/2} - I\|_2 \leqq \frac{\|C\|_2}{1-\|C\|_2} \leqq \frac{\rho}{1-\rho}$$

using Lemma 5.2. Hence from (2.11) we have

$$\|W_+^{-1/2}AW_+^{-1/2} - W^{-1/2}AW^{-1/2}\|_F \leqq \rho\,\frac{2-\rho}{(1-\rho)^2}\,\|\bar{A}\|_F \leqq 6\,\rho\|\bar{A}\|_F$$

for $0 \leqq \rho \leqq 1/2$. Since $\|\bar{A}\|_F = \|A\|_{W,F}$ we have

$$\|A\|_{W_+,F} \leqq \|A\|_{W,F} + \|W_+^{-1/2}AW_+^{-1/2} - W^{-1/2}AW^{-1/2}\|_F \leqq (1 + 6\,\rho)\|A\|_{W,F}\,.$$

The left inequality follows by interchanging W_+ and W in the proof.

Q.E.D.

Since $H(x)$ is positive definite in a neighborhhod of x^*, Lemma 5.4 shows that the norms $\|\cdot\|_{H(x),F}$ and $\|\cdot\|_{H(x_+)}$ will satisfy (2.6).

Lemma 5.4: There exists $\varepsilon > 0$ so that for $\|x-x^*\|_* \leqq \varepsilon$ and $\|x_+-x^*\|_* \leqq \varepsilon$ then for any real n by n matrix A

$$\|A\|_{H(x),F} \leqq (1 + 12\,\mu^2 m(x,x_+))\|A\|_{H(x_+),F}\,.$$

where

$$m(x,z) \;=\; \max\{\, \omega(\|x-x^*\|_*)\,,\, \omega(\|z-x^*\|_*) \,\}$$

and

$$\mu \;=\; \max\{\, \|H(x^*)\|_2,\ \|H(x^*)^{-1}\|_2 \,\}\,.$$

Proof: Choose $\varepsilon > 0$ so that the following hold

$$2\,\mu^2\,\omega(\varepsilon) \leqq \frac{1}{2}$$

$$\max\{\, \|H(x)\|_2,\ \|H(x)^{-1}\|_2 \,\} \leqq 2\,\mu \quad \text{for } \|x-x^*\|_* \leqq \varepsilon\,.$$

This can be done in view of the continuity of H. Let $\|x-x^*\| \leqq \varepsilon$, and $\|x_+-x^*\| \leqq \varepsilon$. Then from Lemma 5.3

$$\begin{aligned}
\rho \;&=\; \max\{\, \|H(x)^{-1}\|_2, \|H(x_+)\|_2 \,\}\, \|H(x)-H(x_+)\|_2 \;\leqq\; 2\,\mu\, \|H(x)-H(x_+)\|_2 \\
&\leqq\; 2\,\mu^2\, \|H(x)-H(x_+)\|_{H(x^*),F} \\
&\leqq\; 2\,\mu^2 \max\{\, \omega(\|x-x^*\|_*)\,,\, \omega(\|x_+-x^*\|_*) \,\} \\
&\leqq\; 2\,\mu^2\omega(\varepsilon) \;\leqq\; \frac{1}{2}\,.
\end{aligned}$$

Hence

$$1 + 6\,\rho \;\leqq\; 1 + 12\,\mu^2 m(x,x_+) \;\leqq\; 4$$

and the result follows from (5.9).

Q.E.D.

We now combine Lemma 5.1 and 5.4 to show that the least change secant updates in the weighted Frobenius norms $\|\cdot\|_{H(x_+),F}$ and $\|\cdot\|_{J,F}$ where

$$J_{ij} \;=\; \int_0^1 H(x+\tau s)_{ij}\, d\tau \qquad 1 \leqq i,j \leqq n.$$

are of bounded deterioration at for $H(x^*)$ at x^*.

Lemma 5.5: Assume $H(x) \in \mathbf{A}$ for all $x \in \Omega$. There exists $\varepsilon > 0$ so that for $\|x-x^*\|_* \leqq \varepsilon$ and $\|x_+-x^*\|_* \leqq \varepsilon$ the least change secant update B_+ (5.4) in the norm $\|\cdot\|_{H(x_+),F}$ satisfies

$$\|B_+-H(x^*)\|_{H(x_+),F} \leqq (1+12\mu^2 m(x,x_+))\|B-H(x^*)\|_{H(x),F} + 8m(x,x_+)\,.$$

The least change secant update C_+ (5.4) in the norm $\|\cdot\|_{J,F}$ satisfies

$$\|C_+-H(x^*)\|_{H(x_+),F} \leqq (1+60\mu^2 m(x,x_+))\|C-H(x^*)\|_{H(x),F} + 32\, m(x,x_+)\,.$$

Proof: From Lemma 4.2 we know that $J \in \mathbf{A} \cap \mathbf{Q}(y,s)$. From Lemma 5.1 we have

$$\|B_+ - J\|_{H(x_+),F} \leqq \|B - J\|_{H(x_+),F} .$$

Choose $\varepsilon > 0$ so that Lemma 5.4 holds. The norms $\|\cdot\|_{H(x_+),F}$ and $\|\cdot\|_{H(x),F}$ satisfy

$$\|\cdot\|_{H(x),F} \leqq (1 + 12\,\mu^2 m(x,x_+))\|\cdot\|_{H(x_+),F}$$

Hence

$$\begin{aligned}\|B_+ - H(x^*)\|_{H(x_+),F} &\leqq \|B - H(x^*)\|_{H(x_+),F} + 2\,\|H(x^*) - J\|_{H(x_+),F} \\ &\leqq (1 + 12\,\mu^2 m(x,x_+))\,\|B - H(x^*)\|_{H(x),F} + 8\,\|H(x^*) - J\|_{H(x^*),F} \\ &\leqq (1+12\mu^2 m(x,x_+))\|B-H(x^*)\|_{H(x),F} + 8m(x,x_+)\end{aligned}$$

By similar arguments as in Lemma 4.2 and Lemma 5.4, we have

$$\frac{1}{1 + 12\,\mu^2 m(x,x_+)}\,\|A\|_{J,F} \leqq \|A\|_{H(x),F} ,$$

and

$$\|A\|_{H(x_+),F} \leqq (1 + 12\,\mu^2 m(x,x_+))\,\|A\|_{J,F} .$$

From Lemma 5.1 we have

$$\|B_+ - J\|_{J,F} \leqq \|B - J\|_{J,F} .$$

Hence we have

$$\begin{aligned}\|B_+ - H(x^*)\|_{J,F} &\leqq \|B - H(x^*)\|_{J,F} + 2\,\|H(x^*) - J\|_{J,F} \\ &\leqq (1 + 12\,\mu^2 m(x,x_+))\,\|B - H(x^*)\|_{H(x),F} + 8\,\|H(x^*) - J\|_{H(x^*),F} \\ &\leqq (1+12\mu^2 m(x,x_+))\|B-H(x^*)\|_{H(x),F} + 8m(x,x_+)\end{aligned}$$

Since

$$\|B-H(x^*)\|_{H(x_+),F} \leqq (1+12\mu^2 m(x,x_+))\|B-H(x^*)\|_{J,F} ,$$

we have

$$\|B-H(x^*)\|_{H(x_+),F} \leqq$$

$$(1+12\mu^2 m(x,x_+))^2\|B-H(x^*)\|_{H(x),F} + 8(1+12\mu^2 m(x,x_+))m(x,x_+)$$

From the choice of ε, $12\mu^2 m(x,x_+) \leqq 3$ and

$$(1 + 12\,\mu^2 m(x,x_+))^2 \leqq 1 + 60\,\mu^2 m(x,x_+)$$

and we have the desired result.

Q.E.D.

For x_k and s_k let $y_k \in \mathbf{R}^n$

$$y_k = \nabla f(x_k+s_k) - \nabla f(x_k) \tag{5.13}$$

and define the matrix $J_k \in \mathbf{R}^{n\times n}$

$$J_k \equiv \int_0^1 H(x_k+\tau s_k)d\tau \tag{5.14}$$

Theorem 5.6 Assume $H(x)\in$ **A** for all $x\in \Omega$. Let B_{k+1} be a least change secant update either in the norm $\|\cdot\|_{H(x_{k+1}),F}$ or $\|\cdot\|_{J_k,F}$ using (5.13) where J_k is given in (5.14). Let $\theta_k \leqq \theta \leqq t < 1$. There exists $\varepsilon>0$ so that if

$$\|x_0-x^*\|_* \leqq \varepsilon \text{ , and } \|B_0-H(x^*)\|_{H(x^*),F} \leqq \varepsilon \text{ ,}$$

then $\{x_k\}$ is defined for any method in the inexact quasi-Newton framework, $x_k \to x^*$ as $k \to \infty$, and

$$\|x_{k+1}-x^*\|_* \leqq t\, \|x_k-x^*\|_*,$$

where $\|y\|_*=\|H(x^*)y\|$ and

$$\lim_{k\to\infty} \frac{\|y_k-B_k s_k\|}{\|s_k\|} = 0 .$$

If $\theta_k\to 0$ as $k\to\infty$ then $\{x_k\}$ converges Q-superlinearly.

References

[1] C. G. Broyden, J. E. Dennis Jr. and J. J. More, On the local and superlinear convergence of quasi-Newton methods, *J. Inst. Math. Appl.*, 12 (1973) 223-245.

[2] R. S. Dembo, S. C. Eisenstat, and T. Steihaug, Inexact Newton methods, *SIAM J. Numer. Anal.*, 19 (1982) 400-408.

[3] J.E.Dennis Jr. and J.J.More, A characterization of superlinear convergence and its application to quasi-Newton methods, *Math. Comp.*, 28 (1974) 549-560.

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

[5] A. A. Goldstein, Constructive Real Analysis, Harper & Row, Publishers, New York, 1967

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

[7] T. Steihaug, *Quasi-Newton Methods for Large Scale Nonlinear Problems*, Ph.D Thesis, SOM Technical Report #49, Yale University, 1980.