Local and Superlinear Convergence

for

Truncated Iterated Projections Methods (1)

by

Trond Steihaug (2)

Rice MASC TR81-2

(1) A part of this paper was presented at the Third International Workshop on Numerical Analysis, IIMAS, University of Mexico, January 1981 and ORSA-TIMS National Meeting, Houston, October 1981.

(2) Department of Mathematical Sciences, Rice University, P.O. Box 1892, Houston, Texas 77001
Abstract: Least change secant updates can be obtained as the limit of iterated projections based on other secant updates. We show that these iterated projections can be terminated or truncated after any positive number of iterations and the local and the superlinear rate of convergence are still maintained. The truncated iterated projections method is used to find sparse and symmetric updates that are locally and superlinearly convergent.

Keywords: Symmetric and sparse secant methods, unconstrained optimization

Running Title: Truncated Iterated Projections
Introduction

Consider a system of \( n \) nonlinear equations in \( n \) unknowns

\[
F(x) = 0
\]  

(1.1)

where \( F: \mathbb{R}^n \rightarrow \mathbb{R}^n \) is continuously differentiable in a neighborhood of a solution \( x^* \). Quasi-Newton methods have been shown to be robust and efficient methods for solving these problems. Specifically quasi-Newton methods approximate the solution \( x^* \) by generating a sequence of iterates \( \{x_k\} \) as follows:

Given \( x_0 \) and \( B_0 \)

FOR \( k=0 \) STEP 1 UNTIL Convergence DO

Solve \( B_k s_k = -F(x_k) \)  

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

Update to obtain \( B_{k+1} \)

where \( \{B_k\} \) is a sequence of \( n \) by \( n \) nonsingular matrices. Derivative information of \( F \) is incorporated in the Jacobian approximations \( \{B_k\} \) by requiring that

\[
B_{k+1} s_k = F(x_k + s_k) - F(x_k).
\]  

(1.3)

Updates that satisfy (1.3) are called secant updates and (1.3) is called the secant equation. For given vectors \( s, y \in \mathbb{R}^n \), \( s \neq 0 \) let

\[
\mathcal{L}(y,s) = \{ B \in \mathbb{R}^{n \times n} : Bs = y \}.
\]  

(1.4)

The Jacobian approximation \( B_{k+1} \) is said to be a least change secant update if \( B_{k+1} \) solves

\[
\min[\|B-B_k\| : B \in \mathcal{L}(y_k,s_k)].
\]  

(1.5)

where
\[ y_k = F(x_k + s_k) - F(x_k) \]  \hspace{1cm} (1.6)

The norms that have been found to be effective in (1.5) are the inner product norms on \( \mathbb{R}^{n \times n} \), i.e., the Frobenius matrix norm

\[
\| M \|_F = \left( \sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij}^2 \right)^{1/2} = \left( \text{trace } M^T M \right)^{1/2}
\]

and the weighted Frobenius matrix norm

\[
\| M \|_W = \| W^{-1/2} M W^{-1/2} \|_F
\]

where \( W \) is a symmetric and positive definite \( n \) by \( n \) matrix. Unless otherwise indicated, the norm in this note will be the Frobenius matrix norm or the weighted Frobenius matrix norm.

If the Jacobian matrix of \( F \) has special structure such as symmetry then the matrices \( \{ B_k \} \) can be required to belong to the set \( \mathcal{C} \) of matrices which have this structure. The least change secant update with this structure is then the Jacobian approximation \( B_{k+1} \) which solves

\[
\min_{B} \left( \| B - B_k \| : B \in \mathcal{C} \cap \mathbb{R}^{n \times n} \right)
\]

with \( y_k \) given in (1.6). The method of iterated projections is a technique to derive a more complex update based on a basic update. Specifically, we construct a sequence of matrices \( \{ B_{(j)}^{(k+1)} \} \) that will converge to the update that solves (1.7), but each \( B_{(j)}^{(k+1)} \) involves solving (1.5).

For a given affine subspace \( \mathcal{B} \) of \( \mathbb{R}^{n \times n} \), let \( P_{\mathcal{B}} \) denote the projection operator onto \( \mathcal{B} \). Let \( \mathcal{C} \) be the subspace of \( \mathbb{R}^{n \times n} \) that have the desired structure of the Jacobian matrix of \( F \) and let \( B_{(0)} = B_k \in \mathcal{C} \). A method of iterated projections is
\[ B_{k+1}^{(j)} = \frac{P \cap 2(y_k, s_k)}{\omega} B_{k+1}^{(j)}, \quad j \geq 0 \]

or

\[ B_{k+1}^{(j+1)} = \left( \frac{P \cap 2(y_k, s_k)}{\omega} \right)^{j+1} B_k, \quad j \geq 0. \] (1.8)

From Dennis and Schnabel [8] we have that if \( \omega \cap 2(y_k, s_k) \neq \emptyset \) then \( \{B_{k+1}^{(j)}\} \) converges to \( B_{k+1} \), the solution of (1.7) or equivalently

\[ B_{k+1} = \frac{P \cap 2(y_k, s_k)}{\omega} B_k. \]

The key question of our paper is must \( j \) become infinite in order that the proper convergence results hold?

In Section 2, we will discuss least change updates and prove some basic properties that will be used to answer our key question for sparse and symmetric updates.

In Section 3, we will discuss the least change sparse and symmetric secant update which arises when we use the Frobenius norm. We will show that we can terminate the iterated projection method after any number of iterated projections, i.e. for \( j \) in (1.8) we need \( j \geq 0 \) to have a sparse and symmetric update that is locally and superlinearly convergent. We will discuss the relationship to the results of Steihaug [16].
Least Change Updates

Let \( \{x_k\} \) be a sequence of iterates found by a quasi-Newton method and assume that \( \{x_k\} \) converges to a solution \( x^* \) of

\[
F(x) = 0
\]

at which the Jacobian matrix \( F'(x^*) \) is nonsingular. From Dennis and Moré [6] we have \( \{x_k\} \) converges to \( x^* \) superlinearly if and only if

\[
\frac{\|y_k - B_k s_k\|}{\|s_k\|} \to 0 \quad \text{as} \quad k \to \infty
\]

where \( y_k \) is given by (1.6).

Let \( \|\cdot\|_2 \) be the Euclidean vector norm, and assume that the matrix \( B_{k+1} \) satisfies the secant equation (1.3). Then we have

\[
\frac{\|y_k - B_k s_k\|_2}{\|s_k\|_2} = \frac{\|(B_{k+1} - B_k)s_k\|_2}{\|s_k\|_2} \leq \|B_{k+1} - B_k\|_F
\]

where we used the fact that the Frobenius matrix norm is consistent with the Euclidean vector norm. It is thus reasonable to make \( \|B_{k+1} - B_k\|_F \) small.

We denote the projection operator in the Frobenius norm onto the affine space \( \mathcal{A} \) of \( \mathbb{R}^{n \times n} \) by \( P_{\mathcal{A}} \) and the projection operator onto \( \mathcal{F} \) in the corresponding weighted norm by \( P_{\mathcal{F},\mathcal{W}} \).

---

(2) We say that the sequence \( \{x_k\} \) converges superlinearly to \( x^* \) if

\[
\lim_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0
\]
Let $\mathcal{C}$ be any affine space of $\mathbb{R}^{n \times n}$, then for $M \in \mathcal{C}$

\[
\|P_{\mathcal{C},W}B - M\|_W^2 + \|P_{\mathcal{C},W}B - B\|_W^2 = \|B - M\|_W^2
\]  

(2.1)

If

\[
\mathcal{C} = \mathcal{C} \cap \mathcal{L}(y,s) \neq \emptyset
\]  

(2.2)

where $\mathcal{C}$ is an affine space of $\mathbb{R}^{n \times n}$ then

\[
\|P_{\mathcal{C},W}B - M\|_W^2 \leq \|B - M\|_W^2 - \left[\frac{\|W^{-1/2}(y - Bs)\|_2^2}{\|W^{1/2}s\|_2} \right]^2
\]  

(2.3)

**Proof** The first property (2.1) follows directly from the fact that $P_{\mathcal{C},W}$ is a projection operator and $\mathcal{C}$ is an affine space.

To establish the second property (2.3), we have

\[
\|W^{-1/2}(y - Bs)\|_2 = \|W^{-1/2}(P_{\mathcal{C},W}B - B)s\|_2
\]

\[
= \|W^{-1/2}(P_{\mathcal{C},W}B - B)W^{-1/2}(W^{1/2}s)\|_2
\]

\[
\leq \|W^{-1/2}(P_{\mathcal{C},W}B - B)W^{-1/2}\|_F\|W^{1/2}s\|_2
\]

since $P_{\mathcal{C},W}B \in \mathcal{L}(y,s)$ and the Frobenius matrix norm is consistent with the Euclidean vector norm. Hence from (2.1) we have

\[
\|P_{\mathcal{C},W}B - M\|_W^2 \leq \|B - M\|_W^2 - \left[\frac{\|W^{-1/2}(y - Bs)\|_2^2}{\|W^{1/2}s\|_2} \right]^2
\]

which is the desired result (2.3).

If we want to add a new restriction on the update, i.e., we want the update also to be in the affine subspace $\mathcal{C}$ of $\mathbb{R}^{n \times n}$, we may use iterated projections to accomplish this.
Assume that we may easily find $P_{B, W}^B$ and $P_{C, W}^B$ for all $B \in \mathbb{R}^{n \times n}$, where $B$ satisfies (2.2). We may now use iterated projections as follows

$$B(j) = P_{C, W}^B P_{B, W}^B (j-1) = (P_{C, W}^B P_{B, W}^B)^j B(0)$$ \hspace{1cm} (2.4)

where $B(0) = B$.

**Lemma 2.2**: Let $C \cap A \cap 2(y, x) \neq \emptyset$ and $M \in C \cap A \cap 2(y, s)$. Then

$$\left\| (P_{C, W}^B P_{B, W}^B)^j B - M \right\|_W^2 \leq \left\| B - M \right\|_W^2 - \left[ \frac{\|W^{-1/2} (y - Bs)\|_2^2}{\|W^{1/2} s\|_2} \right]^2$$

for all $j \geq 1$.

**Proof**: We will first show that

$$\left\| (P_{C, W}^B P_{B, W}^B)^j B - M \right\|_W \leq \left\| B - M \right\|_W, \quad j \geq 1 \hspace{1cm} (2.5)$$

Consider $B(j)$ defined in (2.4) for $j \geq 2$. Using (2.1) twice we have

$$\left\| B(j) - M \right\|_W^2 = \left\| P_{C, W}^B P_{B, W}^B (j-1) - M \right\|_W^2$$

$$= \left\| P_{B, W}^B (j-1) - M \right\|_W^2 - \left\| P_{C, W}^B P_{B, W}^B (j-1) - P_{B, W}^B (j-1) \right\|_W^2$$

$$\leq \left\| P_{B, W}^B (j-1) - M \right\|_W^2$$

$$= \left\| B(j-1) - M \right\|_W^2 - \left\| P_{B, W}^B (j-1) - B(j-1) \right\|_W^2$$

$$\leq \left\| B(j-1) - M \right\|_W^2$$

Hence, by induction we have

$$\left\| B(j) - M \right\|_W^2 \leq \left\| B(1) - M \right\|_W^2, \quad j \geq 1 \hspace{1cm} (2.6)$$
By (2.1) we have

\[ ||B^{(1)} - M||^2 = ||P_CWP_B WP_B^{(0)} - M||^2 \]

\[ = ||P_B WP_B^{(0)} - M||^2 - ||P_CWP_B WP_B^{(0)} - P_B WP_B^{(0)}||^2 \]

\[ \leq ||P_B WP_B^{(0)} - M||^2 \]

which combined with (2.6) is the desired result (2.5), since \( B^{(0)} = B \).

From (2.5) using (2.3) we have

\[ ||(P_CWP_B WP_B^{(0)})^j B - M||^2 \leq ||P_B WP_B^{(0)} - M||^2 \]

\[ \leq ||B - M||^2 - \left( \frac{||W^{-1/2}(y-Bs)||_2}{||W^{1/2}s||_2} \right)^2 \]

We note that the matrix \( B^{(j)} \) has the property that \( B^{(j)} \in \mathbb{C} \), but in general \( B^{(j)} \notin \mathbb{Q} \). However, if \( P_C \) and \( P_Q \) commute, i.e.,

\[ P_C P_Q = P_Q P_C \quad (2.7) \]

then \( B^{(j)} \in \mathbb{C} \cap \mathbb{Q} \). To see this, use \( A = \mathbb{Q} \cap \mathbb{Q}_2(y,s) \neq \emptyset \) and (2.7).

\[ B^{(j)} = P_C WP_B WP_B^{(j-1)} = P_C WP_B WP_B^{(j-1)} \]

\[ = P_C WP_B WP_B^{(j-1)} \]

\[ = P_C WP_B WP_B^{(j-1)} \]

\[ = P_C WP_B WP_B^{(j-1)} \]

Hence \( B^{(j)} \in \mathbb{Q} \) and we have \( B^{(j)} \in \mathbb{Q} \cap \mathbb{C} \).

We may interchange the order in which we do the projections, but if \( B \in \mathbb{Q} \cap \mathbb{C} \) then it is natural to require that \( B^{(j)} \) should have the structural properties.
We believe that the structural properties in this context are more important than the secant equation, i.e., we do not require that $B^{(j)} \in \Delta(y,s)$.

In the next section, we will use these lemmas to generate symmetric and sparse updates and we will see that $B^{(j)} \in \mathcal{A} \cap \mathcal{C}$ but in general $B^{(j)} \notin \Delta(y,s)$.

In contrast to Dennis and Schnabel [8] and Dennis and Walker [9], we make the simplifying assumption throughout the paper that

$$\mathcal{C} \cap \mathcal{A} \cap \Delta(y,s) \neq \emptyset.$$
Sparse and Symmetric Updates

In this section, we will consider solving \( n \) nonlinear equations in \( n \) unknowns

\[
F(x) = 0
\]

where \( F: \mathbb{R}^n \rightarrow \mathbb{R}^n \) satisfies the following conditions.

1. There exists \( x^* \) so that \( F(x^*) = 0 \)
2. \( F \) is continuously differentiable in an open neighborhood \( \Omega \) of \( x^* \)
3. \( F' \) is Hölder continuous at \( x^* \) with exponent \( 0 < p \leq 1 \), i.e. there exists \( L \geq 0 \) so that
   \[
   \|F'(x) - F'(x^*)\|_p \leq L\|x - x^*\|^p
   \]  
   (3.1)
4. \( F'(x^*) \) is nonsingular
5. \( F' \) is symmetric.

We will also assume that \( F' \) is sparse, i.e., there exists a set of indices \( K \) so that if \( (i,j) \notin K \), then \( [F'(x)]_{ij} = 0 \) \( \forall x \in \Omega \). Since \( F' \) is symmetric, we will assume that \( K \) preserves the symmetry, i.e., if \( (i,j) \in K \) then \( (j,i) \in K \). Finally, we note that if \( x^* \) is a local minimizer of a function \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) with gradient \( F \), then \( F'(x^*) \) is positive definite by assumption (5) and we know that \( F'(x^*)_{ii} \neq 0 \) \( i = 1,2,\ldots,n \) and \( (i,i) \in K \), \( i = 1,2,\ldots,n \). In the following, we assume that the set of indices \( K \) preserves the symmetry and contains the diagonal elements.

Let \( \mathcal{S} \) be the set of sparse matrices

\[
\mathcal{S} = \{ B \in \mathbb{R}^{n \times n} : B_{ij} = 0, (i,j) \notin K \}
\]  
(3.2)

and let \( \mathcal{T} \) be the set of symmetric matrices.

\[
\mathcal{T} = \{ B \in \mathbb{R}^{n \times n} : B = B^T \}.
\]  
(3.3)

We want to find

\[
B_+ = \left[ P_{\mathcal{T}\setminus\mathcal{S}}(y,s) \right]^B
\]  
(3.4)
However, to find the sparse and symmetric update (3.4) we have to solve a linear system of equations

$$Gu = y - Bs$$  \hspace{1cm} (3.5)

where $G$ is a $n \times n$ symmetric and positive definite matrix$^{(3)}$ [20], the matrix $G$ is sparse and has the same sparsity structure as the update, i.e. $G \in \mathcal{J} \cap \mathcal{J}$. In large scale optimization, solving the linear system (3.5) might be expensive. Instead, we will use iterated projection to include either sparsity, symmetry or both based on the symmetric secant update (the Powell Symmetric Broyden Update [7,13]), the sparse secant update (the Schubert update [2,15]) or the secant update (or Broyden update [1,7]) respectively. Let $P_\mathcal{G}$ be the projection operator onto the affine space $\mathcal{G}$ of $\mathbb{R}^{n \times n}$ in Frobenius norm.

Marwil [10] derived the sparse and symmetric secant update (3.4) from the iteration

$$B^{(j)} = \left[ P_{\mathcal{J}} P_{\mathcal{G}}(y,s) \right] B^{(j-1)} \hspace{1cm} (3.6)$$

based on the sparse secant update [2,15]

$$[P_{\mathcal{G}}(y,s) B]_{ij} = B_{ij} + \sigma^+_i (y-Bs)_{is} s_{js} \hspace{1cm} (i,j) \in K \hspace{1cm} (3.7)$$

where

$$\sigma^+_i = \sum_{j:(i,j) \in K} s_j^2 \hspace{1cm} i=1,\ldots,n \hspace{1cm} (3.8)$$

and $\sigma^+_i$ is the pseudo inverse

$$\sigma^+_i = \begin{cases} 0 & \text{if } \sigma_i = 0 \\ 1/\sigma_i & \text{otherwise.} \end{cases}$$

$^{(3)}$The matrix $G$ may be indefinite, but by eliminating all zero columns and rows the reduced matrix is positive definite.
Powell [14] describes an alternative set of projections

$$B^{(j)} = [P_pP_{\mathcal{T}\cap 2}(y,s)]^jB$$  \hspace{1cm} (3.9)

based on the symmetric secant update [7,13]

$$P_{\mathcal{T}\cap 2}(y,s)B = B + \frac{1}{s_s}[(y-Bs)s^T + s(y-Bs)^T] - \frac{(y-Bs)^T s}{(s^T s)^2}ss^T.$$  \hspace{1cm} (3.10)

The final iterated projections method is

$$B^{(j)} = [P_pP_2(y,s)]^jB$$  \hspace{1cm} (3.11)

based on the secant update [1,7]

$$P_2(y,s)B = B + \frac{1}{s_s}(y-Bs)s^T.$$  \hspace{1cm} (3.12)

It is easily seen that \[8\]

$$P_pP_{\mathcal{T}} = P_{\mathcal{T}p} = P_{\mathcal{T}\cap \mathcal{Q}}$$  \hspace{1cm} (3.13)

hence (3.6), (3.9) and (3.11) are the only iterated projections we can construct so that

$$B^{(j)} \in \mathcal{T} \cap \mathcal{Q}.$$  

For $x, x_+ \in \Omega$ let

$$y = F(x+s) - F(x)$$  \hspace{1cm} (3.14)

where $s = x_+ - x$. 
Lemma 3.1:

Let $x, x_+ \in \Omega$ and $y$ given by (3.14). Then

$$T \cap \mathcal{F} \cap \mathcal{A}(y, s) \neq \emptyset.$$  

Proof. The result follows from Ortega and Rheinboldt [12, 3.2.7] since the $n \times n$ matrix $M$ where element $(i,j)$ is given by

$$M_{ij} = \int_0^1 F'(x + \tau(x_+ - x))_{ij} \, d\tau$$

satisfies

$$Ms = y, \ M \in T \cap \mathcal{F}.$$ 

We will now consider the iterated projections methods (3.6), (3.9) and (3.11). Let $m$ be the number of iterated projections and

$$\sigma(x, z) = \max\{\|x - x^*\|_P, \|z - x^*\|_P\}$$

(3.16)

where $p$ is given in (3.1).

Lemma 3.2: Let $x, x_+ \in \Omega$ and let the number of iterations in (3.6), (3.9), or (3.11) be $m \geq 1$, and set $B_+ = B(m)$. Then

$$\|B_+ - F'(x^*)\|_F \leq \|B - F'(x^*)\|_F + 2L\sigma(x, x_+)$$

(3.17)

Let $\|x - x^*\| \leq \varepsilon$ and $\|x_+ - x^*\| \leq \varepsilon$. Then for

$$\beta = 2L(2\|B - F'(x^*)\|_F + 3\varepsilon^2L)$$

we have

$$\|B_+ - F'(x^*)\|_F^2 \leq \|B - F'(x^*)\|_F^2 - \left[\frac{\|y - Bs\|_2}{\|s\|_2}\right]^2 + \beta \sigma(x, x_+).$$

(3.18)
Proof. Let $M$ be given by (3.15). Then from Lemma 2.2 and 3.1 we have

$$||B_+ - M||_F \leq ||B - M||_F.$$  \hspace{1cm} (3.19)

Hence

$$||B_+ - F'(x^*)||_F \leq ||B_+ - M||_F + ||M - F'(x^*)||_F$$

$$\leq ||B - M||_F + ||M - F'(x^*)||_F$$

$$\leq ||B - F'(x^*)||_F + 2||M - F'(x^*)||_F$$

From Ortega and Rheinboldt [12, 3.2.11] we have

$$||M - F'(x^*)||_F \leq \int_0^1 ||F'(x + \tau s) - F'(x^*)||_F d\tau$$

$$\leq L \sigma (x, x_+)$$ \hspace{1cm} (3.20)

and we have shown (3.17).

From Lemma 2.2, we have

$$||B_+ - F'(x^*)||_F^2 \leq (||B_+ - M||_F + ||M - F'(x^*)||_F)^2$$

$$\leq ||B_+ - M||_F^2 + (2L||B_+ - M||_F + L^2eP)\sigma (x, x_+)$$

$$\leq ||B - M||_F^2 - \left[\frac{||y - Bs||}{||s||^2} \right]^2 + L(2||B - M||_F + L^2eP)\sigma (x, x_+)$$

$$\leq ||B - F'(x^*)||_F^2 - \left[\frac{||y - Bs||}{||s||^2} \right]^2 + \beta \sigma (x, x_+)$$

where

$$2L(2||B - M||_F + LeP) \leq 2L(2||B - F'(x^*)||_F + 2L\sigma (x, x_+) + LeP) \leq \beta .$$

We will now study the effect of terminating the iterated projections and using the corresponding update in a quasi-Newton method. Since the update is sparse, we will use the Inexact Quasi-Newton method [16].
Given $x_0$ and $B_0$

FOR $k=0$ STEP 1 UNTIL Convergence DO

Find some $s_k$ so that

$$\frac{||B_k s_k + F(x_k)||}{||F(x_k)||} \leq \eta_k$$ (3.21)

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

Update $B_{k+1}$ by one or more iterated projections (3.6), (3.9) or (3.11).

If $B_k = F'(x_k)$ then we have the Inexact Newton method [4].

Theorem 3.3: Let $\eta_k < \eta < \theta < 1$ and $m_k \geq 1$.

There exist $\varepsilon > 0$ and $\delta > 0$ so that if

$$||x_0 - x^*|| \leq \varepsilon \quad \text{and} \quad ||B_0 - F'(x^*)|| \leq \delta$$

then $\{x_k\}$ converges to $x^*$. Moreover the convergence is linear in the sense that

$$||x_{k+1} - x^*|| \leq \theta ||x_k - x^*|| \quad k \geq 0$$ (3.22)

where $||z||_\theta = ||F'(x^*)z||$, and the sequences $||B_k||$ and $||B_k^{-1}||$ are bounded.

Proof (3.17) is a bounded deterioration condition. If $\eta = 0$, then the convergence results follow from Broyden-Dennis and Moré [3]. The general case is proved in Steihaug [16].

The next lemma characterizes superlinear convergence and can be found in [16].

Lemma 3.4: Assume $\{x_k\}$ converges to $x^*$ and

$$\frac{||y_k - B_k s_k||}{||s_k||} \to 0 \quad \text{as} \quad k \to \infty$$ (3.23)
Then the sequence \( \{x_k\} \) converges superlinearly if and only if

\[
\frac{||B_k s_k + F(x_k)||}{||F(x_k)||} \to 0 \quad \text{as} \quad k \to \infty
\]  

(3.24)

**Theorem 3.5:** Assume that the conditions in Theorem 3.3 hold. If \( \eta_k \to 0 \) as \( k \to \infty \), then \( \{x_k\} \) converges superlinearly.

**Proof** From (3.21) and \( \eta_k \to 0 \) as \( k \to \infty \) we have (3.24). Hence we have super-linear convergence if (3.23) holds. We now show that (3.23) holds.

Since \( \{||B_k||\} \) is bounded, it follows that \( \{||B_k - F'(x*)||_F\} \) is bounded, say by \( \rho \). Let \( \beta = 2L(2\rho + 3\epsilon^* L) \), then from (3.18) we have \( \{B_k\} \) satisfies

\[
||B_{k+1} - F'(x^*)||_F^2 \leq ||B_k - F'(x^*)||_F^2 - \frac{||y_k - B_k s_k||_2^2}{||s_k||_2^2} + \beta \sigma (x_k, x_{k+1})
\]

for all \( k \geq 0 \). Thus

\[
\sum_{i=1}^{k} \left[ \frac{||y_i - B_i s_i||_2^2}{||s_i||_2^2} \right] \leq ||B_0 - F'(x^*)||_F^2 - ||B_{k+1} - F'(x^*)||_F^2 + \beta \sum_{i=1}^{k} \sigma (x_i, x_{i+1})
\]

From (3.22) we have that

\[
\sum_{i \geq 0} ||x_{i+1} - x^*||_P^p \leq \frac{1}{1-tp} ||x_0 - x^*||_P^p
\]

hence

\[
\sum_{i \geq 0} \sigma (x_i, x_{i+1}) < +\infty
\]

where we have used the equivalence of norms. Hence we have

\[
\sum_{i \geq 0} \left[ \frac{||y_i - B_i s_i||}{||s_i||} \right]^2 < \infty
\]
and we have shown that (3.23) holds.

Broyden, Dennis and Moré [3] show that one step of iterated projections

\[ B_+ = F \Pi_2(y, s)^B \]

gives a locally and superlinearly convergent quasi-Newton method. Dennis [5] has pointed out that

\[ B_+ = F \Pi_2(y, s)^B \]

gives rise to a locally and superlinearly convergent quasi-Newton method. Moré [11] has pointed out that

\[ B_+ = F \Pi_2(y, s)^B \]

has the same properties. However, the analysis for these cases were restricted to one step of the iterated projections method.

The results in this section may be extended to global convergence results using a line search approach like the backtracking strategy [18] or using trust regions [17].

Consider the sparsity operator \( P_\cdot \) and symmetry operator \( P_\tau \). From [8] we have
Let \( P_j B \) be defined as:

\[
P_j B = \begin{cases} B_{ij} & \text{if } (i,j) \in K \\ 0 & \text{otherwise} \end{cases}
\]  

(3.25)

and

\[
P_j B = \frac{1}{2} (B + B^T).
\]  

(3.26)

Consider the symmetric secant update (3.10). Let \( C = \frac{1}{s^T s} (I - \frac{1}{2s^T s} s s^T) \).

From the Sherman-Morris formula \([7]\) we have

\[
(s^T s I + s s^T)^{-1} = \frac{1}{s^T s} (I - \frac{1}{2s^T s} s s^T)
\]

and (3.10) may be written as

\[
P_j \cap L(y, s) B = B + s [C(y - Bs)]^T + C(y - Bs) s^T.
\]

Let

\[ d = C(y - Bs) \]

then from (3.25) the iterated projections method (3.9) is

\[
B^{(t+1)} = \begin{cases} B^{(t)} + d^{(t)} s_j + s_i d^{(t)} & (i,j) \in K \\ 0 & \text{otherwise} \end{cases}
\]

(3.27)

Similarly it follows from (3.13) and (3.26) that the iterated projections methods (3.6) and (3.11) may be written in the form (3.27) where \( d^{(t)} \in \mathbb{R}^n \) is given by

\[
d^{(t)} = C(y - B^{(t)} s)
\]

(3.28)
for some n by n matrix C that depends on the method used. We note that if we use (3.5) then from (3.7) and (3.26), C is a diagonal matrix with diagonal elements

$$C_{ii} = \frac{1}{2} \sigma_{i}^{+}$$

where $\sigma_{i}^{+}$ is the pseudo inverse of $\sigma_{i}$ where $\sigma_{i}$ is given in (3.8). The method (3.11) gives a diagonal matrix

$$C = \frac{1}{2s^T s} I .$$

The matrix G in (3.5) is a n by n matrix given by

$$G_{ij} = \begin{cases} \delta_{ij} \sigma_{i} + s_i s_j & (i,j) \in K \\ 0 & \text{otherwise} \end{cases}$$

(3.29)

where $\delta_{ij}$ is

$$\delta_{ij} = \begin{cases} 1 & i=j \\ 0 & \text{otherwise} \end{cases} .$$

We now show that the methods of iterated projections are equivalent to one step stationary iterative methods for solving (3.5).

**Theorem 3.6:** The methods of iterated projections (3.6), (3.9) and (3.11) generate matrices $B^{(\ell)}$ given by

$$B^{(\ell)}_{ij} = \begin{cases} B_{ij} + u_{i}^{(\ell)} s_i + s_i u_{j}^{(\ell)} & (i,j) \in K \\ 0 & \text{otherwise} \end{cases}$$

(3.30)
where \( u^{(\ell)} \) is determined by

\[
u^{(0)} = 0
\]

\[
u^{(\ell)} = u^{(\ell-1)} - C \left[ G u^{(\ell-1)} - (y - Bs) \right]
\]

where

\[
C = \begin{cases}
\frac{1}{2} \text{diag} (\sigma_1^+, \ldots, \sigma_n^+) & \text{if we use (3.6)} \\
\frac{1}{T} \left( I - \frac{1}{T} s s^T \right) & \text{if we use (3.9)} \\
\frac{1}{2s^T s} I & \text{if we use (3.11)}
\end{cases}
\]

**Proof.** Consider \( B^{(\ell+1)} \) given in (3.27) and \( c^{(\ell)} \) given in (3.28).

Define

\[
u^{(m+1)} = \sum_{k=0}^{m} d^{(k)}
\]

Then from (3.27)

\[
B_{ij}^{(\ell)} = B_{ij}^{(0)} + s_j \sum_{k=0}^{\ell-1} d^{(k)}_{i} + s_i \sum_{k=0}^{\ell-1} d^{(k)}_{j}
\]

\[
= B_{ij} + u_i^{(\ell)} s_j + s_i u_j^{(\ell)}
\]

using (3.32), and \( B^{(0)} = B \) and we have shown (3.30). Consider

\[
(y - B^{(\ell-1)} s)_i = y_i - \sum_{j: (i,j) \in K} (B_{ij} s_j + s_j u_i^{(\ell-1)} + s_j s u_j^{(\ell-1)})
\]

\[
= (y - Bs)_i - (G u^{(\ell-1)})_i
\]

using (3.33) and (3.29). Since

\[
u^{(\ell)} = u^{(\ell-1)} + d^{(\ell-1)}
\]

where \( d^{(\ell)} \) is given by (3.28) we have from (3.34) that
\[ u^{(\ell)} = u^{(\ell-1)} + d^{(\ell-1)} \]
\[ = u^{(\ell-1)} + C(y - B^{(\ell-1)}s) \]
\[ = u^{(\ell-1)} - C(Gu^{(\ell-1)} - (y-Bs)) \]

and we have shown (3.31).

Since the method of iterated projections is converging [8] we have that the sparse and symmetric least change update is given by

\[
B^{(\omega)}_{ij} = \begin{cases} 
B_{ij} + u_{s_{i}} + s_{i}u_{j} & (i,j) \in K \\
0 & \text{otherwise}
\end{cases}
\]

for some \( u \). And it follows that the matrix \( G \) in (3.5) is given in (3.29).

Hence we have shown that a truncated iterated projections method is equivalent to terminating a specific one step stationary iterative method for solving (3.5).

It follows from the eigenvalue analysis in [16] that for each of the iterated projections method there exists \( \alpha < 1 \) so that

\[ \|B^{(\ell+1)} - B^{(\omega)}\|_F \leq \alpha \|B^{(\ell)} - B^{(\omega)}\|_F. \]

Of special interest is (3.6) for which

\[ \alpha = 1 - \frac{1}{m} \]

where \( m \) is the maximum number of nonzero elements in any row of \( B \).

Steinhaug [16, 19] discusses updates where we only partially solve (3.5) and gives a general theory for how accurate a solution of (3.5) is needed to achieve the proper convergence results.
Let 
\[ b = y - Bs \] 
and
\[ q(u) = \frac{1}{2} u^T Gu - b^T u \]
where \( G \) is given in (3.5). For \( u \in \mathbb{R}^n \) let
\[
B(u)_{ij} = \begin{cases} 
B_{ij} + u_i s_{ij} + s_{ij} & (i,j) \in K \\
0 & \text{otherwise}
\end{cases}
\]
In \([16, 19]\) it is shown that if the new update \( B_{k+1} \) is chosen as
\[ B_{k+1} = B_k(u_k) \]
where \( u_k \) satisfies
\[
q_k(u_k) \leq -\beta \left( \frac{\|b_k\|}{\|s_k\|} \right)^2
\]
for \( 0 < \beta < 1 \) then Theorem 3.3 holds. In \([6]\) it is also shown that the iterative methods (3.31) with \( \ell = 1 \) satisfies (3.35) with \( \beta = \frac{1}{4} \).

Let \( C \) and \( Q \) be subspaces of \( \mathbb{R}^{n \times n} \) and assume \( F'(x) \in C, \quad F'(x) \in Q \)
for all \( x \in \Omega \). Then for \( x, x_+ \in \Omega \) and \( y \) given by (3.14)
\[ C \cap Q \cap \mathcal{A} (y,s) \neq \emptyset . \]
Let \( \beta \) be given in (2.2). Hence by Lemma 2.2 we can generalize Theorem 3.3 and conclude that the iterated projections (2.4) can be terminated after any positive number of iterations while maintaining the conclusions of the theorem. Also note that the results are valid for fixed scale least-change updates.
Acknowledgments: The author acknowledges a pleasant evening at Ninfa's in November 1980, where the ideas of this paper were discussed with E. Marwil and G. Tapia. Discussions with R. Tapia have greatly benefited the results presented.
References


linear convergence of quasi-Newton methods", *Journal of the Institute
of Mathematics and its Applications* 12(1973)223-245.


convergence and its application to quasi-Newton methods", *Mathematics


[9] J.E. Dennis, Jr. and H.F. Walker, "Convergence theorems for least-
change secant update methods", *SIAM Journal of Numerical Analysis*


in J.B. Rosen, O.L. Mangasarian, and K. Ritter, eds., *Nonlinear


equations with a sparse Jacobian", *Mathematics of Computation* 24


[19] T. Steihaug, "On the sparse and symmetric least-change secant update", Technical Report MASC TR 82-4, Department of Mathematical Sciences, Rice University (Houston, TX, 1982).