SUPERLINEAR CONVERGENCE OF A CLASS OF
\( \theta \)-BOUNDED RANK-ONE UPDATE METHODS

by

Ai-Ping Liao

\( ^1 \)
SUPERLINEAR CONVERGENCE OF A CLASS OF 
θ-BOUNDED RANK-ONE UPDATE METHODS

Ai-Ping Liao

Abstract

In this paper, we propose a class of single-rank Quasi-Newton methods, called the class $\mathcal{A}$ of $\theta$-bounded rank-one updates, which are affine combinations of Pearson's method and a modified McCormick method, and prove the local superlinear convergence of this class of methods. Some numerical results are given to test these methods.

Key words. Nonlinear equations, quasi-Newton updates, $\theta$-bounded class, superlinear convergence.

---

1School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853.

Research supported in part by NSF, AFORS and ONR through NSF grant DMS-8920550.
1 Introduction

Quasi-Newton methods do not need higher order derivative information yet possess quite satisfactory convergence rates compared with Newton's method for zero-finding or function minimization. Todd [9] derives well-known matrix updates from an abstract vector space setting. Todd also proposed a single-rank update which is an affine combination of Pearson's update and McCormick's update. In this note we are interested in a class of methods which are affine combinations of Pearson's method and a modified McCormick update method. McCormick's update is

$$B_+ = B + \frac{(y - Bs)(B^T s)^T}{(B^T s)^T s},$$

(1)

and the modified McCormick update is given by:

$$B_+ = B + \frac{(y - Bs)(B s)^T}{s^T B s},$$

(2)

which can be obtained by replacing the term $B^T s$ in (1) with $Bs$. We do not know yet what the geometric meaning of this update is.

As we know the convergence results for Pearson's method and McCormick's method were proved using bounded deterioration methods in the $B$-form as above and the $H$-form, where $H$ is the inverse of $B$, respectively; see Broyden, Dennis and Moré [1]. The standard method in [1] first proving linear convergence and then superlinear convergence, cannot be used directly for the affine combination of these methods. However, an affine combination of Pearson's method and the modified McCormick method (2) seems more consistent than that of Pearson's method and McCormick's method, and we can combine these two arguments into one as Stachurski [8] does for Broyden's bounded $\theta$-class of methods and Ip and Todd [5] do for a different class of rank-one methods.
2 The class $A$ of rank-one update methods

Let $R^n$ denote the real $n$-dimensional space. We are interested in finding a zero $x_*$ of a function $F : R^n \to R^n$ whose domain $D$ and range $F(D)$ lie in $R^n$.

**Algorithm**

0. Choose a nonsingular matrix $B_0$ and $x_0 \in D$. Set $k = 0$.

1. Compute

$$s_k = -B_k^{-1}F(x_k),$$
$$x_{k+1} = x_k + s_k,$$
$$y_k = F(x_{k+1}) - F(x_k).$$

2. Update $B_k$ so that $B_{k+1}s_k = y_k$. Set $k = k + 1$ and go to step 1.

In this note the class $A$ of quasi-Newton methods will be considered.

**Definition 1** The class $A$ of quasi-Newton methods is a subclass of single-rank update methods in which

$$B_{k+1} = \theta_k(B_k + \frac{(y_k - B_k s_k)y_k^T}{y_k^T s_k}) + (1 - \theta_k)\left(B_k + \frac{(y_k - B_k s_k)(B_k s_k)^T}{(B_k s_k)^T s_k}\right),$$

where $\{\theta_k\}$ is a bounded sequence of scalar parameters.

In the next section, we will prove the Q-superlinear convergence of this class of methods.
3 Superlinear convergence

In this section we discuss the local properties of the class $A$ of methods for the problem of finding a solution to the system of $n$ equations in $n$ unknowns given by

$$f_i(x_1, \ldots, x_n) = 0, \quad 1 \leq i \leq n,$$

where $f_1, \ldots, f_n$ are the component functions of $F$. We assume that the function $F$ is continuously differentiable in an open convex set $D$ and there is an $x_*$ in $D$ such that $F(x_*) = 0$ and $J_* := J(x_*)$ is nonsingular. The notation $J(x)$ denotes the Jacobian matrix $(\partial f_i(x))$ evaluated at $x$. Thus $x_*$ is a locally unique solution to the equation $F(x) = 0$. We also assume that there is a constant $\kappa_0$ such that

$$\|J(x) - J(x_*)\| \leq \kappa_0 \|x - x_*\|, \quad x \in D. \quad (4)$$

Our final assumption is that $J_*$ is symmetric and positive definite. (This is the case when we use these methods for the problem of function minimization.) Pearson's method and McCormick's method are proved to be Q-superlinearly convergent in Broyden, Dennis and Moré [1] under this assumption too.

We first need some lemmas.

Lemma 3.1 (Lemma 3.1 of [1]) Let $x_*$ be given as above. Then for all $v, u \in D$

$$\|F(v) - F(u) - J(x_*)(v - u)\| \leq \kappa_0 \max\{\|v - x_*\|, \|u - x_*\|\}\|v - u\|. \quad (5)$$

Lemma 3.2 Assume that $J_*$ is positive definite. Then there exist positive scalars $\varepsilon, \Delta, \mu_l$ and $\mu_u$ such that, if $\|x - x_*\| \leq \varepsilon$, and $\|B - J_*\| \leq 2\Delta$, then $\|Bs\| \leq \mu_u \|s\|$, $\|J_*s\| \leq \mu_u \|s\|$, $\|y\| \leq \mu_u \|s\|$, $\mu_l \|s\|^2 \leq y^T s \leq \mu_u \|s\|^2$, and $\mu_l \|s\|^2 \leq s^T Bs \leq \mu_u \|s\|^2$. 
Proof. Note that \( \|B\| = \|B - J_* + J_*\| \leq 2\Delta + \|J_*\| \), thus \( \|Bs\| \leq \mu_u \|s\| \) if we take \( \mu_u > 3\Delta + \|J_*\| \). Obviously, \( \|J_*s\| \leq \mu_u \|s\| \). By Taylor’s theory, if \( \varepsilon \) is sufficiently small, \( \|y\| \leq \mu_u \|s\| \), which also implies \( y^T s \leq \mu_u \|s\|^2 \). Define

\[
\bar{J} = \int_0^1 J(x + \tau s) d\tau.
\]

Then we have \( y = \bar{J}s \), so that (further restricting \( \varepsilon \), if necessary)

\[
y^T s = s^T \bar{J} s = s^T J_* s + s^T (\bar{J} - J_*) s \\
\geq s^T J_* s - O(\|s\|^3) \geq \mu_l \|s\|^2
\]

for some \( \mu_l > 0 \) such that \( \mu_l \) is less than the smallest eigenvalue of \( J_* \). Note that \( s^T Bs = s^T J_* s + s^T (B - J_*) s \); thus if \( \Delta \) is sufficiently small we have

\[
\mu_l \|s\|^2 \leq s^T Bs \leq \mu_u \|s\|^2.
\]

\( \square \)

From the above lemma, the class \( \mathcal{A} \) of rank-one methods is locally well-defined.

Lemma 3.3 The update formula (3) can be rewritten as:

\[
B_{k+1} = B_{k+1}^{Pearson} + (1 - \theta_k)\delta B_k, \quad (6)
\]

where

\[
B_{k+1}^{Pearson} = B_k + \frac{(y_k - B_k s_k) y_k^T}{y_k^T s_k} \quad (7)
\]

and

\[
\delta B_k = -(y_k - B_k s_k) \left[ \frac{s_k^T B_k s_k \cdot y_k^T - y_k^T s_k \cdot (B_k s_k)^T}{y_k^T s_k \cdot s_k^T B_k s_k} \right] . \quad (8)
\]

\( \square \)

This lemma is easy to verify. The following result can be proved using the results in Broyden, Dennis and Moré [1].
Lemma 3.4 Let $M := J_{**}^\frac{1}{2}$. Suppose $\| x_k - x_* \| \leq \varepsilon$, $\| x_{k+1} - x_* \| \leq \varepsilon$ and $\| B_k - J_* \|_M \leq 2\Delta$, where $\varepsilon$ and $\Delta$ are sufficiently small. Here $\| \cdot \|_M$ is defined by means of $\| Q \|_M = \| MQM \|$. Then there are positive constants $\alpha_0, \alpha_1$ and $\alpha_2$ such that

$$\| B_{k+1}^{Pearson} - J_* \|_M \leq [1 - \alpha_0 \Theta_k^2 + \alpha_1 \sigma_k] \| B_k - J_* \|_M + \alpha_2 \sigma_k,$$

where

$$\Theta_k := \begin{cases} \frac{\| (B_k - J_*) s_k \|}{\| B_k - J_* \|_M \| s_k \|} & \text{for } B_k \neq J_*, \\ 0 & \text{otherwise,} \end{cases}$$

and $\sigma_k := \max\{\| x_{k+1} - x_* \|, \| x_k - x_* \|\}$.

From Corollary 3.3 of Broyden, Dennis and Moré [1], if some subsequence of $\{ \| B_k - J_* \| \}$ converges to zero, then $\{ x_k \}$ converges $Q$-superlinearly at $x_*$. Thus we assume that $B_k \neq J_*$ for all $k$.

Corollary 3.5 Assume that $\| B_0 - J_* \|_M \leq \Delta$, and $\| B_k - J_* \|_M \leq 2\Delta$ for $k = 0, 1, \ldots, m - 1$. Then

$$\| B_{k+1}^{Pearson} - J_* \|_M \leq \| B_k - J_* \|_M - \frac{\alpha_0 \| (B_k - J_*) s_k \|^2}{2\Delta \| s_k \|^2} + W \sigma_k,$$

where $W = 2\Delta \alpha_1 + \alpha_2$.

Proof. Using Lemma 3.4 directly. \qed

Lemma 3.6 Assume that $\| B_0 - J_* \|_M \leq \Delta$ and $\| x_0 - x_* \| \leq \varepsilon$, and for $k = 0, 1, \ldots, m - 1$,

$$\| x_{k+1} - x_* \| \leq r \| x_k - x_* \|, \quad \| B_k - J_* \|_M \leq 2\Delta < 1,$$
where \(0 < r < 1\). Then there is a constant \(\kappa > 0\), depending only on \(\kappa_0, \mu_1, \mu_u\) and \(\|M\|\), such that
\[
\|\delta B_k\|_M \leq \kappa \frac{\|(B_k - J_\ast)s_k\|^2}{\|s_k\|^2} + \kappa \sigma_k. \tag{11}
\]

**Proof.** By Lemma 3.1 and Lemma 3.2,
\[
y_k^T s_k^T s_k B_k s_k \geq \mu_1^2\|s_k\|^4; \tag{12}
\]
\[
\|y_k - B_k s_k\| \leq \|y_k - J_\ast s_k\| + \|(B_k - J_\ast)s_k\|
\leq \kappa_0 \sigma_k\|s_k\| + \|(B_k - J_\ast)s_k\|.
\]

Also,
\[
\|s_k^T B_k s_k y_k^T - y_k^T s_k (B_k s_k)^T\|
\leq \|s_k^T B_k s_k y_k^T - s_k^T B_k s_k (J_\ast s_k)^T\| + \|s_k^T B_k s_k (J_\ast s_k)^T - y_k^T s_k (B_k s_k)^T\|
\leq \mu_u \kappa_0 \sigma_k\|s_k\|^3 + \mu_u |s_k^T B_k s_k - y_k^T s_k|\|s_k\||
+ \|y_k^T s_k (J_\ast s_k)^T - y_k^T s_k (B_k s_k)^T\|
\leq \mu_u \kappa_0 \sigma_k\|s_k\|^3 + \mu_u |s_k^T B_k s_k - y_k^T s_k|\|s_k\|
+ \|y_k^T s_k (J_\ast s_k)^T - y_k^T s_k (B_k s_k)^T\|
\leq \mu_u \kappa_0 \sigma_k\|s_k\|^3 + \mu_u |s_k^T (B_k s_k - J_\ast s_k) + s_k^T (J_\ast s_k - y_k)|\|s_k\|
+ \|y_k^T s_k (J_\ast s_k)^T - y_k^T s_k (B_k s_k)^T\|
\leq \mu'_u \sigma_k\|s_k\|^3 + \mu_u \|s_k\|^2\|(B_k - J_\ast)s_k\| + \|y_k^T s_k (J_\ast s_k)^T - y_k^T s_k (B_k s_k)^T\|
\leq \mu'_u \sigma_k\|s_k\|^3 + \mu_u \|s_k\|^2\|(B_k - J_\ast)s_k\| + \mu_u \|s_k\|^2\|(B_k - J_\ast)s_k\|
= \mu'_u \sigma_k\|s_k\|^3 + 2\mu_u \|s_k\|^2\|(B_k - J_\ast)s_k\|.
\]

7
Thus,

\[ \| \delta B_k \| \leq \frac{1}{y_k^T s_k B_k s_k} \| y_k - B_k s_k \| \| s_k B_k s_k y_k^T - y_k^T s_k (B_k s_k)^T \| \]

\[ \leq \frac{1}{\mu_k^2 \| s_k \|^2} (\kappa_0 \sigma_k \| s_k \| + \| (B_k - J_*) s_k \|)(\mu'_u \sigma_k \| s_k \|^3 + 2 \mu_u \| s_k \|^2 \| (B_k - J_*) s_k \|)
\]

\[ \leq \frac{1}{\mu_k^2} (\kappa_0 \sigma_k + \| (B_k - J_*) s_k \|)(\mu'_u \sigma_k + 2 \mu_u \| (B_k - J_*) s_k \|)
\]

\[ \leq \frac{1}{\mu_k^2} (\kappa_0 \mu'_u \sigma_k^2 + (\mu'_u + \mu'_{u'}) \sigma_k \| (B_k - J_*) s_k \| \| s_k \| + 2 \mu_u \| (B_k - J_*) s_k \|)(\| s_k \|^2)
\]

\[ \leq 2 \mu'_u \| (B_k - J_*) s_k \|^2 + \mu_k^{-2} (\kappa_0 \mu'_u \sigma_k + 4 \mu_u \mu'_{u'}) \sigma_k
\]

\[ \leq \kappa \| (B_k - J_*) s_k \|^2 + \kappa' \sigma_k,
\]

where \( \kappa' = \max\{2 \mu'_u \mu_k, \mu_k^{-2} (\kappa_0 \mu'_u + 4 \mu_u \mu'_{u'})\} \) (we assume that \( \sigma_k \leq 1 \)). Noting that

\[ \| Q \|_M \leq \| M \|^2 \| Q \|, \]

we thus have

\[ \| \delta B_k \|_M \leq \kappa \| (B_k - J_*) s_k \|^2 \| s_k \|^2 + \kappa \sigma_k,
\]

where \( \kappa = \| M \|^2 \kappa' \).

Now we prove our main theorem.

**Theorem 3.7** For each \( r \in (0, 1) \), there are positive constants \( \epsilon \) and \( \Delta \) such that for all \( x_0, B_0 \) satisfying the conditions:

\[ \| x_0 - x_* \| \leq \epsilon, \quad \| B_0 - J_* \|_M \leq \Delta,
\]

then the sequence

\[ x_{k+1} = x_k - B_k^{-1} F(x_k)
\]

defined by any method from the class \( \mathcal{A} \) of rank-one methods with \( |1 - \theta_k| \leq \tilde{\theta} \) for each \( k \), where \( \tilde{\theta} \) is a positive constant scalar, is well-defined and converges to \( x_* \) at a \( Q \)-superlinear rate.
Proof. First we note that there is a constant $\eta > 0$ such that $\| A \| \leq \eta \| A \|_M$ for any $n \times n$ matrix. Let $r \in (0, 1)$ be given and set $\gamma \geq \| J_* \|$. We first choose $\Delta > 0$ such that

$$\frac{\alpha_0}{2\Delta} - \kappa \tilde{\theta} > 1$$

and then choose $\varepsilon > 0$ and further restrict $\Delta$, if necessary, so that

$$(2\alpha_1 \Delta + \alpha_2 + \tilde{\theta} \kappa) \frac{\varepsilon}{1 - r} + \kappa \tilde{\theta} \frac{W_{\frac{\varepsilon}{1 - r}}}{\frac{\alpha_0}{2\Delta} - \kappa \tilde{\theta}} \leq \Delta,$$

$$\gamma (1 + r) [ \kappa_0 \varepsilon + 2\eta \Delta ] \leq r,$$

and the hypotheses of Lemma 3.4 hold, and $x \in D$ if $\| x - x_* \| < \varepsilon$. Now suppose that $\| B_0 - J_* \|_M < \Delta$ and $\| x_0 - x_* \| < \varepsilon$. Then $\| B_0 - J_* \| < \eta \Delta < 2\eta \Delta$, and since (15) yields

$$2\gamma (1 + r) \eta \Delta \leq r,$$

the Banach Perturbation Lemma (Ortega and Rheinboldt [6]) gives

$$\| B_0^{-1} \| \leq (1 + r) \gamma.$$

Lemma 3.1 now implies that

$$\| x_1 - x_* \| \leq \| B_0^{-1} \| [ \| F(x_0) - F(x_*) - J_*(x_0 - x_*) \| + \| B_0 - J_* \| \| x_0 - x_* \| ]$$

$$\leq \gamma (1 + r) [ \kappa_0 \varepsilon + 2\eta \Delta ] \| x_0 - x_* \|,$$

and by (15), $\| x_1 - x_* \| \leq r \| x_0 - x_* \|$. Hence $\| x_1 - x_* \| < \varepsilon$ and thus $x_1 \in D$.

Assume that for $k = 0, 1, \ldots, m - 1$,

$$\| B_k - J_* \|_M \leq 2\Delta, \quad \| x_{k+1} - x_* \| \leq r \| x_k - x_* \|.$$

By Lemma 3.4 and (6)

$$\| B_{k+1} - J_* \|_M - \| B_k - J_* \|_M \leq 2\alpha_1 \Delta \varepsilon r^k + \alpha_2 \varepsilon r^k + | 1 - \theta_k | \| \delta B_k \|_M.$$
So,
\[ \|B_m - J_*\|_M - \|B_0 - J_*\|_M \leq (2\alpha_1 + \alpha_2) \frac{\varepsilon}{1 - r} + \bar{\theta} \sum_{k=0}^{m-1} \|\delta B_k\|_M. \] (16)

By Corollary 3.5
\[ \|B_{k+1}^{Pearson} - J_*\|_M - \|B_k - J_*\|_M \leq -\frac{\alpha_0}{2\Delta} \sum_{k=0}^{m-1} \frac{\|(B_k - J_*)s_k\|^2}{\|s_k\|^2} + W\sigma_k. \]

Thus, by (6),
\[ \|B_m^{Pearson} - J_*\|_M - \|B_0 - J_*\|_M \leq -\frac{\alpha_0}{2\Delta} \sum_{k=0}^{m-1} \frac{\|(B_k - J_*)s_k\|^2}{\|s_k\|^2} + W\frac{\varepsilon}{1 - r} + \bar{\theta} \sum_{k=0}^{m-1} |1 - \theta_k| \|\delta B_k\|_M. \]

Therefore,
\[ \alpha_0 \sum_{k=0}^{m-1} \frac{\|(B_k - J_*)s_k\|^2}{\|s_k\|^2} \leq W \frac{\varepsilon}{1 - r} + \bar{\theta} \sum_{k=0}^{m-1} \|\delta B_k\|_M + \|B_0 - J_*\|_M - \|B_m^{Pearson} - J_*\|_M \]
\[ \leq (W \frac{\varepsilon}{1 - r} + \Delta) + \bar{\theta} \sum_{k=0}^{m-1} \|\delta B_k\|_M. \] (17)

From Lemma 3.6,
\[ \sum_{k=0}^{m-1} \|\delta B_k\|_M \leq \kappa \sum_{k=0}^{m-1} \frac{\|(B_k - J_*)s_k\|^2}{\|s_k\|^2} + \kappa \frac{\varepsilon}{1 - r}. \] (18)

Thus, by (17) and (18), we have
\[ \sum_{k=0}^{m-1} \frac{\|(B_k - J_*)s_k\|^2}{\|s_k\|^2} \leq \frac{W \frac{\varepsilon}{1 - r} + \Delta + \bar{\theta} \kappa \frac{\varepsilon}{1 - r}}{\frac{\alpha_0}{2\Delta} - \kappa \bar{\theta}} =: q. \] (19)

By (16), (18) and (19)
\[ \|B_m - J_*\|_M \leq \Delta + (2\alpha_1 + \alpha_2) \frac{\varepsilon}{1 - r} + \bar{\theta} \sum_{k=0}^{m-1} \|\delta B_k\|_M \]
\[ \leq \Delta + (2\alpha_1 + \alpha_2) \frac{\varepsilon}{1 - r} + \bar{\theta} q + \bar{\theta} \kappa \frac{\varepsilon}{1 - r} \]
\[ \leq \Delta + (2\alpha_1 + \alpha_2 + \bar{\theta} \kappa) \frac{\varepsilon}{1 - r} + \bar{\theta} \frac{W \frac{\varepsilon}{1 - r} + \Delta + \bar{\theta} \kappa \frac{\varepsilon}{1 - r}}{\frac{\alpha_0}{2\Delta} - \kappa \bar{\theta}}. \]
By (14), we thus have \( \|B_m - J_*\|_M \leq 2\Delta \). Hence \( \|B_m - J_*\| \leq 2\eta \Delta \), from which using the Banach Perturbation Lemma we have \( \|B_m^{-1}\| \leq (1 + r)\gamma \), and then as in the case of \( m = 1 \),

\[
\|x_{m+1} - x_*\| \leq r\|x_m - x_*\|.
\]

Therefore, for all \( k \)

\[
\|B_k - J_*\|_M \leq 2\Delta, \quad \|x_{k+1} - x_*\| \leq r\|x_k - x_*\|,
\]

i.e., \( \{x_k\} \) converges to \( x_* \). From (19),

\[
\left\{ \sum_{k=0}^{m-1} \frac{\|(B_k - J_*)s_k\|^2}{\|s_k\|^2} \right\}_{m=0}^{\infty}
\]

is uniformly bounded. Thus

\[
\frac{\|(B_k - J_*)s_k\|}{\|s_k\|} \to 0
\]

which implies that \( \{x_k\} \) converges to \( x_* \) Q-superlinearly to \( x_* \) by Theorem 2.2 of Dennis and Moré [2].

We have thus proved Q-superlinear convergence of this class of methods. Clearly, if we take \( \theta_k = 1 \), it gives Pearson’s method and if we take \( \theta_k = 0 \) it corresponds to the modified McCormick method (2). In particular, if we take

\[
\theta_k = \frac{y_k^T s_k}{(y_k + B_k s_k)^T s_k},
\]

then the corresponding update is

\[
B_{k+1} = B + \frac{(y - Bs)(y + Bs)^T}{(y + Bs)^T s}
\]

and it is easy to prove that this update provides a local Q-superlinear convergence rate. We note that (21) is similar to Todd’s update [9]

\[
B_{k+1} = B + \frac{(y - Bs)(y + B^T s)^T}{(y + Bs)^T s}
\]
One advantage of Todd’s update is that it preserves positive definiteness. Unfortunately, this property does not hold for update (21). We also note that both methods with updates (21) and (22) solve an $n \times n$ nonsingular system of linear equations in at most $2n$ steps by the result of Gay [3].

4 Numerical results

Since rank-one updates are not symmetric, so for any update formula there are two ways to implement the resulting quasi-Newton method for unconstrained minimization:

**version a:** solve $B_k s_k = -F_k$;

**version b:** solve $B_k^T s_k = -F_k$.

Our description of the methods in Section 2 and analysis in Section 3 only apply to version a. However, sometimes version b has some nice properties, as we know that, with version b, McCormick’s and Pearson’s methods have the quadratic termination property under certain conditions (see, for example, Broyden, Dennis and Moré [1]). It is thus appropriate to test version b as well.

We will compare Pearson’s method, McCormick’s method, the modified McCormick method, Todd’s method and method (21) (we call it the L method below), in both versions, and two famous rank-two methods: BFGS and DFP. The test problems are some function minimization problems.

**Problem 1.** Rosenbrock’s banana-shaped valley function:

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2,$$
with starting point $x_0 = (-1.2, 1.0)^T$.

**Problem 2.** Wood’s function:

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$
$$+ 90(x_4 - x_3^2)^2 + (1 - x_3)^2$$
$$+ 10.1((x_2 - 1)^2 + (x_4 - 1)^2)$$
$$+ 19.8(x_2 - 1)(x_4 - 1),$$

with $x_0 = (-3, -1, -3, -1)^T$.

**Problem 3.** Powell’s singular function:

$$f(x) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2$$
$$+ (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4,$$

with $x_0 = (3, -1, 0, 1)^T$.

**Problem 4.** The trigonometric function:

$$f(x) = \sum_{i=1}^n f_i(x)^2$$

with

$$f_i(x) = n - \sum_{j=1}^n (\cos x_j + i(1 - \cos x_i) - \sin x_i),$$

$i = 1, \ldots, n$. $x_0 = (\frac{1}{n}, \ldots, \frac{1}{n})^T$.

All methods begin with $B_0 = I$, and with a stopping test $\|\nabla f(x_k)\| < 10^{-5}$.

We also use a simple binary line search. Table 1 gives the results of these methods with the number of iterations for Problem 1. Similarly, Tables 2–4 give the results for Problems 2–4, where $n = 10$ in Problem 4. The runs were performed using MATLAB on a Sun 4 (SPARC station). (Numbers in parentheses are the results obtained by the Optimization toolbox [4].)
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>version a</th>
<th>version b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>89</td>
<td>65</td>
</tr>
<tr>
<td>McCormick</td>
<td>$B \to$ singular</td>
<td>30</td>
</tr>
<tr>
<td>M-McCormick</td>
<td>47</td>
<td>$B \to$ singular</td>
</tr>
<tr>
<td>Todd</td>
<td>52</td>
<td>37</td>
</tr>
<tr>
<td>L</td>
<td>51</td>
<td>$&gt;$ 100</td>
</tr>
<tr>
<td>BFGS</td>
<td>34 (88)</td>
<td>34 (88)</td>
</tr>
<tr>
<td>DFP</td>
<td>48 (92)</td>
<td>48 (92)</td>
</tr>
</tbody>
</table>

Table 1: Numerical results for Problem 1.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>version a</th>
<th>version b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>$&gt;$ 100</td>
<td>73</td>
</tr>
<tr>
<td>McCormick</td>
<td>$B \to$ singular</td>
<td>29</td>
</tr>
<tr>
<td>M-McCormick</td>
<td>$&gt;$ 100</td>
<td>$B \to$ singular</td>
</tr>
<tr>
<td>Todd</td>
<td>27</td>
<td>35</td>
</tr>
<tr>
<td>L</td>
<td>34</td>
<td>$&gt;$ 100</td>
</tr>
<tr>
<td>BFGS</td>
<td>27 (Failed)</td>
<td>27 (Failed)</td>
</tr>
<tr>
<td>DFP</td>
<td>$&gt;$ 100 (Failed)</td>
<td>$&gt;$ 100 (Failed)</td>
</tr>
</tbody>
</table>

Table 2: Numerical results for Problem 2.
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>version a</th>
<th>version b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>$&gt; 100$</td>
<td>47</td>
</tr>
<tr>
<td>McCormick</td>
<td>$B \rightarrow\text{singular}$</td>
<td>44</td>
</tr>
<tr>
<td>M-McCormick</td>
<td>$&gt; 100$</td>
<td>$B \rightarrow\text{singular}$</td>
</tr>
<tr>
<td>Todd</td>
<td>26</td>
<td>46</td>
</tr>
<tr>
<td>L</td>
<td>27</td>
<td>$&gt; 100$</td>
</tr>
<tr>
<td>BFGS</td>
<td>24 (74)</td>
<td>24 (74)</td>
</tr>
<tr>
<td>DFP</td>
<td>82 (62)</td>
<td>82 (62)</td>
</tr>
</tbody>
</table>

Table 3: Numerical results for Problem 3.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>version a</th>
<th>version b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>McCormick</td>
<td>$B \rightarrow\text{singular}$</td>
<td>22</td>
</tr>
<tr>
<td>M-McCormick</td>
<td>34</td>
<td>$B \rightarrow\text{singular}$</td>
</tr>
<tr>
<td>Todd</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>L</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>BFGS</td>
<td>19 (57)</td>
<td>19 (57)</td>
</tr>
<tr>
<td>DFP</td>
<td>19 (59)</td>
<td>19 (59)</td>
</tr>
</tbody>
</table>

Table 4: Numerical results for Problem 4($n = 10$).
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>n = 20</th>
<th>n = 50</th>
<th>n = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Todd</td>
<td>35</td>
<td>74</td>
<td>108</td>
</tr>
<tr>
<td>L</td>
<td>34</td>
<td>74</td>
<td>100</td>
</tr>
<tr>
<td>BFGS</td>
<td>28 (84)</td>
<td>73 (131)</td>
<td>107 (102)</td>
</tr>
</tbody>
</table>

Table 5: Numerical results for Problem 5(n = 20, 50, 100).

The numerical results show that, in version a, Todd’s method and the L method are better than the other rank-one methods. The modified McCormick method simply does not work in version b. Among all the methods we test here the BFGS method is the best; however, the other rank-two method, DFP, does not perform as well. We also note that Todd’s method (in both versions), the L method (in version a) and McCormick’s method (in version b) give very similar results. Finally, we compare Todd’s method (version a), the L method (version a) and the BFGS method with Problem 4 with n = 20, n = 50 and n = 100. The test results are in Table 5. Interestingly, the numbers of iterations for these three methods are almost the same.

5 Conclusions

We propose a class of rank-one update methods and prove each method in this class possesses locally a superlinearly convergent rate, provided that |θ_k| is bounded. In particular, if we choose

$$θ_k = \frac{y_k^T s_k}{(y_k + B_k s_k)^T s_k},$$

then we have the L method which is also locally superlinearly convergent. From these very limited numerical tests it seems that the L method and Todd’s method,
which are single-rank updates, are comparable to the BFGS and DFP methods. By performing the symmetrization procedure of Powell [7] from the L update (21) or Todd's update (22) we have the following rank-two update [9]

\[
B_+ = B + \frac{(y - Bs)(y + Bs)^T + (y + Bs)(y - Bs)^T}{(y + Bs)^T s} - \frac{s^T(y - Bs)(y + Bs)(y + Bs)^T}{((y + Bs)^T s)^2}
\]

(23)

which is an affine combination of the BFGS and the DFP methods as shown by Todd [9]. By Stachurski's result [8], the method (23) is locally superlinearly convergent. We can also symmetrize the class \( \mathcal{A} \) of updates as follows: given a symmetric matrix \( B \), set \( B^{(0)} = B \) and update \( B^{(0)} \) by (3) to obtain \( \tilde{B}^{(1)} \) which is not necessarily symmetric, so we set \( B_+ = \frac{1}{2} (\tilde{B}^{(1)} + (\tilde{B}^{(1)})^T) \). It can be proven that this update provides a locally superlinearly convergent method. Actually, since

\[
\| \frac{1}{2} (\tilde{B} + \tilde{B}^T) - J_* \|_M \leq \frac{1}{2} (\| \tilde{B} - J_* \|_M + \| \tilde{B}^T - J_* \|_M) = \| \tilde{B} - J_* \|,
\]

(24)

(16) holds for the symmetrized \( B_k \). If we keep using inequality (24) in the proof of Theorem 3.7, (20) can be obtained eventually. However the secant equation does not hold now. Actually, the condition (20) implies that the secant equation holds asymptotically. In particular the symmetrical update corresponding to the L method, as well as Todd's method, has the form

\[
B_+ = B + \frac{yy^T - (Bs)(Bs)^T}{(y + Bs)^T s},
\]

(25)

which is locally superlinearly convergent and preserves positive definiteness. Hence, for minimization problems, \( s_k := B_k^{-1} \nabla f(x_k) \) is a descent direction. However, it is not a secant method. Its numerical performance is between those of the L method (Todd's method) and the BFGS method. Finally, we note that the rank-1 update matrices in (25) are the same as those in the BFGS method but with different scalars.
Acknowledgment: The author would like to thank Professor Michael J. Todd for a careful reading of this paper and his many valuable comments and suggestions.
References


