THE CHOLESKY UPDATE FOR UNCONSTRAINED OPTIMIZATION

Saad Shakir Mahmood
College Of Education, Department Of Mathematics
Almustansiriya University
Email: saadshakirmahmood@yahoo.com

ABSTRACT

It is well known that the unconstrained Optimization often arises in economies, finance, trade, law, meteorology, medicine, biology, chemistry, engineering, physics, education, history, sociology, psychology, and so on. The classical Unconstrained Optimization is based on the Updating of Hessian matrix and computed of its inverse which make the solution is very expensive. In this work we will updating the LU factors of the Hessian matrix so we don't need to compute the inverse of Hessian matrix, so called the Cholesky Update for unconstrained optimization. We introduce the convergent of the update and report our findings on several standard problems, and make a comparison on its performance with the well-accepted BFGS update.

Keywords : Unconstrained Optimization, Cholesky Factorization, Convergence

ABSTRAK

Diketahui bahwa masalah optimasi tak berkendala mempunyai peran pada ilmu ekonomi, keuangan, Hukum Metrologi, Kedokteran, Biology, Kimia, Teknik, Fisika, pendidikan, Sejarah, Sosial, Psikologi , dan lain lainnya. Optimasi tak berkendala yang klasik selama ini menyelasaikan masalah berdasarkan penyesuaian matrik HJessian dan menghitung kebalikan dari matrik Hessian yang membuat beban komputasi sangat mahal. Pada naskah ini, kami akan menyesuaikan LU factor dari matrik Hessian Untuk itu kami tidak perlu menghitung kebalikan dari matrik Hessian dank arena itu penyesuaian ini kmami namakan Penyesuaian Cholesky untuk masalah optimasi tak berkendala. Kami memperlihatkan kekonvergenan pe.nyesuaian tersebut dan melaporkan penumuan kami pada beberapa masalah standar yang kami selesaikan dengan

menggunakan penyesuan tersebut, kemudian kami memandingkan hasil yang kami dapat dengan hasil dari penyesuaian BFGS yang terkenal pada masalah potimasi tak berkendala.

Kata Kunci : Unconstrained Optimization, Cholesky Factorization, Convergence

INTRODUCTION

Given $f : R^n \to R$, it is assumed that:

1. f is twice continuously differentiable.
2. f is uniformly convex, i.e., there are m_1 and $m_2 > 0$ such that $m_1 \|\underline{a}\|^2 \le \underline{a}^T \nabla^2 f \underline{a} \le m_2 \|\underline{a}\|^2$, were $\underline{a} \in R''$.

The method proposed by Hart (1990) is to update the Jacobian matrix of f on the basis of both L' unit lower triangular and U' upper triangular matrices. He takes $B = L'U'$ the current approximation of Jacobian matrix and he defines $B^+ = (L'+O)(U'+Q)$, with the updating matrices O and Q lower triangular and upper triangular, respectively. Then he determines O and Q from the following relations:

$$(L'+O)r = y$$
$$(U'+Q)s = r$$

and he used the following elementary lemma to solve the above system:

LEMMA

If $\beta \in R$ and $\underline{a} \in R''$ such that $\underline{a}^T \underline{a} \ne 0$, then the minimum value of $\underline{x}^T \underline{x} = \|\underline{x}\|_2$, such that $\underline{a}^T \underline{x} = \beta$ is $\underline{x} = \beta \frac{\underline{a}}{\underline{a}^T \underline{a}}$.

The update satisfies the Quasi-Newton condition but it's not symmetric.

THE CHOLESKY UPDATE

Given $f : R'' \to R$, suppose that B is the current approximation of the Hessian matrix, and B^+ is the next approximation of the Hessian matrix. Suppose also that B is positive definite. There exists matrices L' and $U' \in R^{n\times n}$, with L' unit lower triangular matrix and U' upper triangular matrix, such that

$$B = L'U' \qquad (1)$$

then L^+ and U^+ are to be updated, such that

$$L^+ = L'+O \qquad (2)$$

$$U^+ = U'+Q \qquad (3)$$

with O and Q are lower triangular and upper triangular matrices respectively, and

$$B^{+} = L^{+}U^{+} \qquad (4)$$

Since B is symmetric then there is a diagonal matrix D such that

$$U' = DL'^{T} \qquad (5)$$

and $B = L'DL'^{T}$

$$B = L'D^{\frac{1}{2}}D^{\frac{1}{2}}L'^{T}$$

with
$$D^{\frac{1}{2}} = \begin{bmatrix} \sqrt{d_{11}} & 0 & . & . & 0 \\ 0 & \sqrt{d_{22}} & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & 0 & . & . & \sqrt{d_{nn}} \end{bmatrix}$$

Hence

$$B = L\,L^{T} \qquad (6)$$

with $L = L'D^{\frac{1}{2}}$, and $O = Q^{T}$ so (4) becomes

$$B^{+} = L^{+}L^{+^{T}} \qquad (7)$$

by the Quasi-Newton condition it follows

$$L^{+}\underline{r} = \underline{y}$$

$$L^{+^{T}}\underline{s} = \underline{r} \qquad (8)$$

with $\underline{r} \in R^{n}$, from (8)

$$\underline{r}^{T}\underline{r} = \underline{s}^{T}L^{+}L^{+^{T}}\underline{s} = \underline{s}^{T}B^{+}\underline{s} = \underline{s}^{T}\underline{y} \qquad (9)$$

so the best choice of $\underline{r}$ which guarantees convergence is

$$\underline{r} = \sqrt{\frac{\underline{s}^{T}\underline{y}}{\underline{s}^{T}B\underline{s}}}L^{T}\underline{s}. \qquad (10)$$

The solution of system (8) is obtained by using the elementary lemma as follows:

$$l_{ij}^{+} = \frac{y_i r_j}{\sum_{k=i}^{n} r_k^2}, i = 1,\ldots n-1 \text{ and } j = 1,\ldots i \qquad (11)$$

$$l_{nj}^{+} = \frac{r_j - \sum_{k=j}^{n-1} l_{kj}^{*} s_k}{s_n}, j = 1,\ldots,n \qquad (12)$$

It is clear that the update is symmetric. In order to show that the update satisfies the Quasi-Newton condition, the following lemma is introduced:

Lemma 1

The Cholesky update given in (11) and (12) satisfies the Quasi-Newton condition.

Proof:

It is only needed to show that $\sum_{j=1}^{n} l_{nj}^{+} r_j = y_n$

.

$$\sum_{j=1}^{n} l_{nj}^{+} r_j = l_{n1}^{+} r_1 + l_{n2}^{+} r_2 + \ldots + l_{nn-1}^{+} r_{n-1} + l_{nn}^{+} r_n$$

$$= r_1\left[\frac{r_1 - \sum_{k=1}^{n-1} l_{k1}^{+} s_k}{s_n}\right] + r_2\left[\frac{r_2 - \sum_{k=2}^{n-1} l_{k1}^{+} s_k}{s_n}\right] + \ldots + r_{n-1}\left[\frac{r_{n-1} - l_{n-11}^{+} s_{n-1}}{s_n}\right] + \frac{r_n^2}{s_n}$$

$$= \frac{-y_1 s_1 - y_2 s_2 - \ldots - y_{n-1} s_{n-1} + \underline{y}^T \underline{s}}{s_n} = \frac{y_n s_n}{s_n} = y_n$$

THE CONVERGENCE OF THE UPDATE

Assumption I

(a) $\nabla f : R'' \rightarrow R''$ is continuously differentiable in an open convex set D,

(b) there is an $\underline{x}^*$ such that $\nabla f(\underline{x}^*) = \underline{0}$ and $\nabla^2 f(\underline{x}^*)$ is nonsingular,

(c) there is a constant K and $p > 0$ such that $\left\|\nabla^2 f(\underline{x}) - \nabla^2 f(\underline{x}^*)\right\| \le K \left\|\underline{x} - \underline{x}^*\right\|^p$,

In proving a convergence result of the same nature as for other Newton-like methods, a Lipschitz condition is assumed on the Hessian matrix. Using factorization and an open pivoting strategy, the factorizations must be shown to be continuous. This has been established in Dennis and Schnabel (1983).

Definition 1:

Let $M_L = \{M \in L(R^n) : m_{ij} = 0, 1 < j\}$ lower triangular matrices. Furthermore let the following mapping be defined:

$$L_L : L(R^n) \rightarrow M_L$$

such that $L_L(M) = M$.

Lemma 2 (Royden 1968)

i. $\|M \underline{x}\|_2 \le \|M\|_F \|\underline{x}\|_2$;

ii. $\|M\|_F \le c\|M\|_2$ for some $c > 0$;

iii. $\|M\|_F = \|L_L(M)\|_F$

Definition 2

Let $\phi : R^n \times M_L \rightarrow M_L$, and Let $(\underline{x}^*, L^*)$ be an element of $R'' \times M_L$. The update function ϕ has the bounded deterioration property at $(\underline{x}^*, L^*)$ if for some neighborhood $D_1 \times D_2$ of $(\underline{x}^*, L^*)$ there exist constants α_1 and α_2 such that if $x \in D_1, LL^T \underline{s} = -\nabla f(\underline{x})$ and $\underline{x}^+ = \underline{x} + \underline{s}$, then for any $L^+ \in \phi(x, L)$.

$$\|L^+ - L^*\| \le [1 + \alpha_1 \sigma(\underline{x}, \underline{x}^+)]\|L - L^*\| + \alpha_2 \sigma(\underline{x}, \underline{x}^+)$$

where $\sigma(\underline{x}, \underline{x}^+) = \max\left\{\|\underline{x} - \underline{x}^*\|^p, \|\underline{x}^+ - \underline{x}^*\|^p\right\}$

.

Assumption II

Let $\nabla^2 f(\underline{x}^*)$ be LL^T - factorable, i.e. $\nabla^2 f(\underline{x}^*) = L^* L^{*T}$, where $L^* \in M_L$, is invertible, and $\nabla f(\underline{x}^*) = 0$.

Lemma 3 (Technical Lemma, Hart, 1990)

Let $L \in M_L$, and $\delta, \gamma > 0$ satisfy

(i) $\|L - L^*\| \le 2\delta < 1$,

(ii) $\max\{\|L^*\|, \|L^{*-1}\|, \|\nabla f(\underline{x}^*)\|\} \le \gamma$.

Then each of the following holds:

(a) L is invertible and $\|L^{-1}\| \le \dfrac{\gamma}{1-2\gamma\delta}$,

(b) $\|L\| \le 2\delta + \gamma$,

(c) $\|L^{-1}L^T - L^{*-1}L^{*T}\| \le 2\delta\gamma \dfrac{1+\gamma^2}{1-2\delta\gamma}$,

(d) $\|LL^T - L^* L^{*T}\| \le 4\delta(2\delta + \gamma)$.

Note that, this technical lemma involves some useful norm inequalities. Since the lemma (and the proof) is independent of the way the lower triangular L is used, it is the choice to use L within the context of the Cholesky update. The inequality (c) is thus of no special interest. However, (d) is needed to represent the addition to the original lemma. The proof for (d) is straightforward.

Theorem 1

Let f satisfy assumption I, and assume that the update function ϕ satisfies the bounded deterioration at $(\underline{x}^*, L^*)$. If for any $p \in [0,1]$, there exists an $\varepsilon = \varepsilon(r)$ and $\delta = \delta(r)$ such that

(i) $\|\underline{x}_0 - \underline{x}^*\| < \varepsilon$,

(ii) $\|L_0 - L^*\| \le \delta$,

(iii) $L^+ \in \phi(x.L)$,

then the sequence $\{\underline{x}_k\}$, defined by $\underline{x}_{k+1} = \underline{x}_k + \underline{s}_k$, $LL^T\underline{s} = -\nabla f(\underline{x})$ satisfies for each $k \ge 0$, $\|\underline{x}_{k+1} - \underline{x}^*\| \le p \cdot \|\underline{x}_k - \underline{x}^*\|$; Moreover, each of $\|L\|, \|L^{-1}\|$ is uniformly bounded.

Proof

Let $p \in (0,1)$. Choose a neighborhood $D_1 \times D_2$ guaranteed by the bounded deterioration property, and restricted so that it contains only invertible matrices. Given c in lemma 2, and γ as in lemma 3, choose $\varepsilon(p)$ and $\delta(p)$ such that $(\underline{x}, L) \in D_1 \times D_2$ whenever $\|\underline{x}_0 - \underline{x}^*\| < \varepsilon$ and $\|L - L^*\| \le 2\delta$, such that

$$2\delta\gamma < 1$$

$$\frac{r^2}{(1-2\gamma\delta)^2}\left[K\varepsilon^p + 4\delta(\delta+\gamma)\right] \le p$$

$$(2\alpha_1\delta + \alpha_2)\frac{\varepsilon^p}{1-r^p} \le \delta \tag{13}$$

where α_1 and α_2 are determined by D_2 and the bounded deterioration property. First, it is to be noted that

$$\underline{x}_1 - \underline{x}^* = \underline{x}_0 - \underline{x}^* - L_0^{T^{-1}} L_0^{-1}\nabla f(\underline{x}_0)$$
$$= -L_0^{T^{-1}} L_0^{-1}[\nabla f(\underline{x}_0) - \nabla f(\underline{x}^*) -$$
$$\nabla^2 f(\underline{x}^*)(\underline{x}_0 - \underline{x}^*) -$$
$$(L_0 L_0^T - \nabla^2 f(\underline{x}^*))(\underline{x}_0 - \underline{x}^*)]$$

Consequently, by the above lemmas,

$$\|\underline{x}_1 - \underline{x}^*\| \le \|L_0^{T^{-1}} L_0^1\| [K\sigma(\underline{x}_0, \underline{x}^*) + 4\delta(2\delta + \alpha)] \|\underline{x}_0 - \underline{x}^*\|$$
$$\le \frac{r^2}{(1-2\gamma\delta)^2}[K\varepsilon^p + 4\delta(2\delta+\gamma)]\|\underline{x}_0 - \underline{x}^*\|$$
$$\le \rho\|\underline{x}_0 - \underline{x}^*\|$$

Thus $\|\underline{x}_1 - \underline{x}^*\| \le \varepsilon$ and $\underline{x}_1 \in D_1$. By way of induction. Assume that for $k = 0,1,2,\ldots,m-1$, then $\|L - L^*\| \le 2\delta$, and $\|\underline{x}_{k+1} - \underline{x}^*\| \le \rho\|\underline{x}_k - \underline{x}^*\|$.

Then by the bounded deterioration property and the fact that

$$\alpha(\underline{x}_{k+1}, \underline{x}_k) = \max\left\{\|\underline{x}_{k+1} - \underline{x}^*\|^p, \|\underline{x}^k - \underline{x}^*\|^p\right\} \le \rho^{pk}\varepsilon^p,$$

it following

$$\|L_{k+1} - L^*\| - \|L - L^*\| \le 2\alpha_1\delta\rho^{pk}\varepsilon^p + \alpha_2\rho^{pk}\varepsilon^p.$$

summing both sides from $k = 0$ to $m-1$, a telescoping sum on the left is obtained and thus

$$\|L_m - L^*\| \le \|L_0 - L^*\| + (2\alpha_1\delta + \alpha_2)\varepsilon^p \sum_{k=0}^{m-1} \rho^{pk} \le 2\delta$$

Hence, by the technical lemma, and $\|L_m\| \le 2\delta + \gamma$. It follows analogously to the case $m = 1$ that

$$\|\underline{x}_{m+1} - \underline{x}^*\| \le \rho\|\underline{x}_m - \underline{x}^*\|$$

Observe that for any $\varepsilon > 0$, there is a $\delta > 0$ such that if $\|\underline{x} - \underline{x}^*\| < \varepsilon$ and $\|L - L^*\| \le 2\delta$ then the computed iterate $\underline{x}^+ = \underline{x} - (LL^T)^{-1}\nabla f(\underline{x})$ must satisfy $\|\underline{x}^+ - \underline{x}^*\| \le \|\underline{x} - \underline{x}^*\|$.

Lemma 4 (Bartle 1975):

Let $\{x_n\}$ be a sequence of positive real numbers. Then $\sum_n x_n$ converges if and only if the sequence $S = \{s_n\}$ of partial sums is bounded.

Lemma 5 (Bartle 1975):

If $\sum_n x_n$ converges, then $\lim_{k\to\infty} x_k = 0$.

Now in this direction the following result will be needed. The straightforward proof is omitted.

Corollary 1 (Dennis and Schnabel, 1983)

If f satisfies the assumptions I and II then for any $\varepsilon > 0$, there is a neighborhood D_ε of $(\underline{x}^*, L^*)$ such that if $(\underline{x}, L) \in D_\varepsilon$, then $\sigma(\underline{x}, \underline{x}^+) < \varepsilon$.

Corollary 2

Define $\eta = \underline{y} - L\underline{r}$, with $\underline{r}$ computed from the Cholesky update which satisfies $L^+\underline{r} = \underline{y}$. Then $\eta = T\underline{y}$ with T lower triangular, bounded, and $T^* = (L^*)^{-1}$.

Proof

That $\eta = T\underline{y}$ and T lower triangular are obvious. The bounded- ness comes from the fact that it is invertible and L satisfies the hypothesis of theorem 1. The fact that $T^* = (L^*)^{-1}$ is obtained by direct substitution.

In order to obtain bounded deterioration property for the LL^T update, bound on $\|v_1\|, i = 1,2,3,\ldots,n, v_1 = [r_1 \quad r_2 \quad r_3 \quad . \quad . \quad . \quad r_n]^T$ is required. In particular, $\|v_1\|$ must approach to zero at essentially the same rate as $\|\underline{s}\|$.

Lemma 6 (Byrd and Nocedal, 1989).

Under assumption I, it follows

$$m\|\underline{s}\|^2 \le \underline{y}^T \underline{s} \le \|\underline{s}\|^2 M, \quad \frac{\|\underline{y}\|^2}{\underline{y}^T \underline{s}} \le M, \text{ for } m \text{ and } M \in R.$$

Lemma 7

Let f satisfies the assumption I. Let v_i defined as $v_i = [r_1 \quad r_2 \quad . \quad . \quad . \quad r_i]^T$ There exist numbers ε, ρ_1 and ρ_2 such that if $\sigma(\underline{x}, \underline{x}^+) < \varepsilon$, then

$$\rho_1 \le \frac{\|\underline{s}\|}{\|v_i\|} \le \rho_2, i = 1,2,3,\ldots,n.$$

Proof

By the Cholesky update

$$\underline{r}^T \underline{r} = r_1^2 + r_2^2 + \ldots + r_n^2 = \underline{y}^T \underline{s} \qquad (14)$$

and from (14)

$$\|v_i\| = \sqrt{\sum_{k=1}^{i} r_k^2} \le \sqrt{\sum_{k=}^{n} r_k^2} = \sqrt{\underline{y}^T \underline{s}},$$ thus by

Lemma 6, $\|v_i\| \le \sqrt{M}\|\underline{s}\| \le \sqrt{M}\varepsilon^{\frac{1}{p}}$

thus $\dfrac{1}{\sqrt{M}\varepsilon^{\frac{1}{p}}} \le \dfrac{1}{\|v_i\|}$, and so

$$\frac{\sqrt{s_n^2}}{M\varepsilon^{\frac{1}{p}}} \le \frac{\|\underline{s}\|}{M\varepsilon^{\frac{2}{p}}} \le \frac{\|\underline{s}\|}{\|v_i\|}.$$

If we set $\rho_1 = \dfrac{\sqrt{s_n^2}}{M\varepsilon^{\frac{1}{p}}}$ then

$$\rho_1 \le \frac{\|\underline{s}\|}{\|v_i\|} \qquad (15)$$

Since $\sqrt{r_1^2} \le \sqrt{\sum_{k=1}^{i} r_k^2}$, then

$\sqrt{r_1^2} = \left|\sqrt{\dfrac{\underline{y}^T \underline{s}}{\underline{s}^T B \underline{s}}} \sum_{k=1}^{n} l_{kl} s_k\right|$, by Theorem 1 $|L|$ is boded, and hypothesis $\|\underline{s}\| \le \varepsilon^{\frac{1}{p}}$, so there is a number $\xi_{ij} > 0$ such that $l_{ij} \le \xi_{ij}$ for all i and j, and such that $\left|\sqrt{\dfrac{\underline{y}^T \underline{s}}{\underline{s}^T B \underline{s}}} \sum_{i=1}^{n} \sum_{k=1}^{n} \xi_{ki} \varepsilon^{\frac{1}{p}}\right| \le \|v_i\|$

By lemma 6, and the boundedness property of $\|L\|$ and $\|\underline{s}\|$, exist a number $\zeta > 0$ such that $\dfrac{\sqrt{m}\|\underline{s}\|}{\xi} \sum_{i=1}^{n} \sum_{k=1}^{n} \xi_{ki} \varepsilon^{\frac{1}{p}} \le \|v_i\|$

Since $s_n \le \|\vec{\underline{s}}\|$, then it follows

$$\frac{\sqrt{m s_n^2}}{\xi} \sum_{i=1}^{n} \sum_{k=1}^{n} \xi_{ki} \varepsilon^{\frac{1}{p}} \le \|v_i\|$$

so, $\frac{\|\underline{s}\|}{\|v_i\|} \leq \frac{\zeta\varepsilon^{\frac{1}{p}}}{\sqrt{ms_n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}\xi_{ki}\varepsilon^{\frac{1}{p}}}$

The Lemma thus holds by letting

$$\rho_2 = \frac{\zeta\varepsilon^{\frac{1}{p}}}{\sqrt{ms_n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}\xi_{ki}\varepsilon^{\frac{1}{p}}}$$

Lemma 8

Let f satisfies the assumption I. Let v_i defined as $v_i = [r_1 \quad r_2 \quad . \quad . \quad . \quad r_i]^T$. There exist numbers $\varepsilon, \rho_1^{'}$ and $\rho_2^{'}$ such that if $\sigma(\underline{x}, \underline{x}^+) < \varepsilon$, then $\rho_1 \leq \frac{\|\underline{r}\|}{\|v_i\|} \leq \rho_2$.

Proof

By lemma 6, and lemma 7

$$\rho_1^{'} = \sqrt{m}\rho_1 \leq \frac{\sqrt{m}\|\underline{s}\|}{\|v_i\|} \leq \frac{\sqrt{\underline{y}^T\underline{s}}}{\|v_i\|} \leq \frac{\sqrt{M}\|\underline{s}\|}{\|v_i\|}$$
$$\leq \sqrt{M}\rho_2 = \rho_2^{'}$$

Now, in order to proceed with the main objective, the following abbreviation is introduced.

(a) $\underline{v}_i = [r_1 \quad r_2 \quad r_3 \quad . \quad . \quad . \quad r_i]$

(b) $C = L - L^*, C^+ = L^+ - L^*,$

(c) $\alpha = T\underline{y} - L^T\underline{s}, \alpha^* = T^*\underline{y} - L^T{}^*\underline{s},$

(d) $\beta = Cr,$

consequently,

(e) $C^+ = C - \left[L_L\left(\frac{\beta\underline{r}^T}{\|\underline{v}_i\|^2}\right) - L_L\left(\frac{\alpha^*\underline{r}^T}{\|\underline{v}_i\|^2}\right)\right]$

Lemma 9

Let f satisfies assumption I and II. Let α^* be as above. Then there exist numbers $K > 0$ and $\varepsilon > 0$ such that if $\sigma(\underline{x}, \underline{x}^+) < \varepsilon$ then

(a) $\frac{\|\alpha_i^*\|}{\|\underline{v}_i\|} \leq K'\sigma(\underline{x}, \underline{x}^+)$, for $i = 1, 2, \ldots, n$

(b) $\left\|L_L\left(\frac{\alpha^*\underline{r}^T}{\|\underline{v}_i\|^2}\right)\right\|_F \leq K''\sigma(\underline{x}, \underline{x}^+)$.

Proof

Since $\alpha^* = T^*\underline{y} - L^T{}^*\underline{s} = T^*(\underline{y} - \nabla^2 f(\underline{x}^*)\underline{s})$, if T_i^* is the i-th row of T^*, then $\alpha_i^* = T_i^*(\underline{y} - \nabla^2 f(\underline{x}^*)\underline{s})$, giving

$$|\alpha_i^*| \leq \|T_i^*\|\|\underline{y} - \nabla^2 f(\underline{x}^*)\underline{s}\|$$
$$\leq \|T_i^*\|K\sigma(\underline{x}, \underline{x}^+)\|\underline{s}\|$$

By lemma 7 there exist $\rho > 0$ and $\varepsilon > 0$ such that if $\sigma(\underline{x}, \underline{x}^+) < \varepsilon$ and

$$\frac{|\alpha_i^*|}{\|\underline{v}_i\|} \leq K\|T_i^*\|_F\,\sigma(\underline{x}, \underline{x}^+)\frac{\|\underline{s}\|}{\|\underline{v}_i\|} \leq K\rho\|T_i^*\|_F \cdot \sigma(\underline{x}, \underline{x}^+)$$

this proves (a) with $K' = K\rho\|T^*\|_F$. Now, to prove (b), let

$$A^* = L_L\left[\frac{\alpha_i^*\underline{r}^T}{\|\underline{v}_i\|^2}\right]$$

then if A_i^* denotes the i-th row of A^*

$$\|A_i *\|^2 = \sum_{j=1}^{i}\left(\frac{\alpha_i * r_j}{\|\underline{v}_i\|^2}\right)^2$$

$$\|A_i *\|^2 = \frac{(\alpha_i *)^2}{\|\underline{v}_i\|^4}\|\underline{v}_i\|^2$$

$$\|A_i *\|^2 = \frac{(\alpha_i *)^2}{\|\underline{v}_i\|^2}$$

$$\|A_i *\|_F^2 = \sum_{i=1}^{n}\left(\frac{\alpha_i * r_j}{\|\underline{v}_i\|^2}\right)^2$$

$$\leq \sum_{i=1}^{n} K'^2\left(\sigma(\underline{x}, \underline{x}^+)\right)^2$$

$$\|A*\|_F \leq \sqrt{n}K'\sigma(\underline{x}, \underline{x}^+)$$

this proves (b) with $K''=\sqrt{n}K'$.

Lemma 10

Let f satisfies the assumption I and II. Then the Cholesky update algorithm satisfies

$$(C^+)_F \leq \sqrt{(1-\theta)}\|C\|_F + K'\sigma(\underline{x}, \underline{x}^+),$$

where $\theta = \left[\frac{1}{\|C\|}\left\|\frac{\beta}{\|\underline{v}_i\|}\right\|\right]^2 \in (0,1).$

Proof

Let $\nabla = C^+ - L_L\left(\frac{\alpha * \underline{r}^T}{\|\underline{v}_i\|^2}\right)$. Then

$\nabla = C - L_L\left(\frac{\beta \underline{r}^T}{\|\underline{v}_i\|^2}\right)$. Thus,

if ∇_i denotes the ith row of ∇,

$$\|\nabla_i\|^2 = \sum_{j=1}^{i}\left[c_{ij} - \frac{\beta_i r_j}{\|\underline{v}_i\|^2}\right]^2$$

$$\|\nabla_i\|^2 = \sum_{j=1}^{i}\left[c_{ij}^2 - \frac{2c_{ij}\beta_i r_j}{\|\underline{v}_i\|^2} + \left(\frac{\beta_i r_j}{\|\underline{v}_i\|^2}\right)^2\right]$$

$$\|\nabla_i\|^2 = \sum c_{ij}^2 - \frac{2\beta_i}{\|\underline{v}_i\|^2}\sum_{j=1}^{i} c_{ij} r_j + \frac{\beta_i^2}{\|\underline{v}_i\|^4}\sum_{j=1}^{i} r_j^2$$

$$\|\nabla_i\|^2 = \sum c_{ij}^2 - \frac{2\beta_i^2}{\|\underline{v}_i\|^2} + \frac{\beta_i^2}{\|\underline{v}_i\|^4}\|\underline{v}_i\|^2$$

$$\|\nabla_i\|^2 = \sum c_{ij}^2 - \frac{\beta_i^2}{\|\underline{v}_i\|^2}$$

thus,

$$\|\nabla\|_F^2 = \sum_{i=1}^{n}\|\nabla_i\|^2 = \sum_{i=1}^{n}\sum_{j=1}^{i} c_{ij}^2 - \sum_{i=1}^{n}\frac{\beta_i^2}{\|\underline{v}_i\|^2}$$

$$0 < \|\nabla\|_F^2 = \|C\|_F^2 - \sum_{i=1}^{n}\frac{\beta_i^2}{\|\underline{v}_i\|^2}$$

observe that

$$0 < \sum_{i=1}^{n}\frac{\beta_i^2}{\|\underline{v}_i\|^2} < \|C\|_F^2$$

thus, setting

$$\theta = \frac{1}{\|C\|_F^2}\sum_{i=1}^{n}\frac{\beta_i^2}{\|\underline{v}_i\|^2}$$

gives $0 \leq \theta \leq 1$.

,and so by lemma 9,

$$\|C^+\|_F \leq \|\nabla\|_F \left\|L_L\left(\frac{\beta r^T}{\|\underline{v}_i\|^2}\right)\right\|$$

$$\|C^+\|_F \leq \sqrt{1-\theta}\|C\|_F + K'\sigma(\underline{x}, \underline{x}^+)$$

Theorem 2

Let f satisfies assumption I and II. Then there exists a neighbourhood $D_1 \times D_2$ of $(\underline{x}^*, L^*)$ such that in $D_1 \times D_2$ the LL^T update function $\phi(x, L)$ described above is well defined and the corresponding iteration

$\underline{x}^+ = \underline{x} - L^{T^{-1}} L^{-1} \nabla f(\underline{x})$,

where $L^+ = \phi(\underline{x}, L)$ is locally convergent at $\underline{x}^*$.

Proof

Set $\alpha_1 = \dfrac{\sqrt{1-\theta}}{\sigma(\underline{x}, \underline{x}^+)}$, and $\alpha_2 = K'$. Then by lemma 7

$$\|C^+\| \le \alpha_1 \sigma(\underline{x}, \underline{x}^+)\|C\| + \alpha_2 \sigma(\underline{x}, \underline{x}^+)$$
$$\|C^+\| \le [1 + \alpha_1 \sigma(\underline{x}, \underline{x}^+)]\|C\| + \alpha_2 \sigma(\underline{x}, \underline{x}^+)$$

So the Cholesky update function $\phi(x, L)$ satisfies the deterioration property and the convergence follows from theorem 1.

Corollary 3

Assume that the hypotheses of Theorem 2 hold. If some subsequence of $\{\|L_k - L^*\|_F\}$ converges to zero, then $\{\underline{x}_k\}$ converges q-superlinearly to $\underline{x}^*$.

Proof

It is necessary to show that

$\lim_{k \to \infty} \|L_k - L^*\|_F = 0$

From the proof of Theorem 1

$\|L_k - L^*\|_F \le 2\delta,$

δ is defined in Theorem 1, therefore $\|L_k - L^*\|_F$ is bounded above and by Lemma 4, converges, and by Lemma 5, $\lim_{k \to \infty} \|L_k - L^*\|_F = 0$. Now it is necessary to show that $\lim_{k \to \infty} \dfrac{\|\underline{x}_{k+1} - \underline{x}^*\|}{\|\underline{x}_k - \underline{x}^*\|} = 0.$

Let $h \in (0,1)$ be given. By Theorem 1 there are numbers $\varepsilon(h)$ and $\delta(h)$ such that $\|L_0 - L^*\|_F < \delta(h)$ and $\|\underline{x}_0 - \underline{x}^*\| < \varepsilon(h)$ which imply that $\|\underline{x}_{k+1} - \underline{x}^*\| \le h\|\underline{x}_k - \underline{x}^*\|$ for each $k \ge 0$. However, if $m > 0$ is chosen such that $\|L_m - L^*\|_F < \delta(h)$ and $\|\underline{x}_m - \underline{x}^*\| < \varepsilon(h)$ and thus, $\|\underline{x}_{k+1} - \underline{x}^*\| \le h\|\underline{x}_k - \underline{x}^*\|$ for each $k \ge m$. So $\dfrac{\|\underline{x}_{k+1} - \underline{x}^*\|}{\|\underline{x}_k - \underline{x}^*\|} \le h,$ and since $h \in (0,1)$ was arbitrary, then $\lim_{k \to \infty} \dfrac{\|\underline{x}_{k+1} - \underline{x}^*\|}{\|\underline{x}_k - \underline{x}^*\|} = 0.$

NUMERICAL RESULTS

In this section, we compare the numerical results of the Cholesky update with the BFGS update . The line search routine is inexact or satisfies the Wolfe conditions. In order to assess the value of this approach, numerical test were carried out on several unconstrained optimization problems. The test problems chosen were as follows:

1. $f(x) = 100(x_2 — x_1^2)^2 + (1 - x_1)^2$.
2. $f(x) = 100(x_2 — x_1^3)^2 + (1 - x_1)^2$.
3. $f(x)$ = 0.0001 ($x_1 - 3)^2 - (x_1 - x_2)$+ exp (20 $(x_1 - x_2)$).
4. $f(x) = (x_1 - 10^6)^2 + (x_2 - 2 \times$ *10−62+x1x2−22.*
5. $f(x) = (10^{-4}\, x_1 x_2 - 1)^2 +$ $[exp(-x_1) + exp(-x_2) - 1.0001]^2$.
6. $f(x) = (1 - x_1)^2 + (x_2^2 - x_3)^2$.
7. $f(x) = \sum_{i=1}^{99} \left\{exp\left[\frac{-(t_i - x_3)^{x_2}}{x_1}\right] - y_i\right\}^2$,

   $Where\ y_i = 0.1\, i \quad and \quad t_i = 25 + [-50 \ln y_i]^{\frac{2}{3}}$.
8. $f(x) = \sum_{i=1}^{n+1} f_i^2$,

   $where\ f_i = \sqrt{10^{-5}}\ (x_i - 1),\ 1 \le i \le 3\ and\ f_{n+1} = \sum_{j=1}^{n}\left(x_j^2\right) - \frac{1}{4}$.
9. $f(x) = \sum_{i=1}^{n}(x_i - 1)^2 +$ *f2+f 3, where*

   $$f = \sum_{j=1}^{n} j\left(x_j - 1\right).$$
10. $f(x) = \sum_{i=1}^{n}[n - \sum_{j=1}^{n} cosx_j +$ *i1−cosxi−sinxi+exi−12.*
11. $f(x) = \sum_{i=1}^{n}\left[i\left(\sum_{j=1}^{n} jx_j\right) - 1\right]^2$.

Tabel 1.

			BFGS			Cholesky		
No	n	Startingpoint	Itr	feval	f min	Itr	feval	f min
1	2	(-1.2,1)	30	188	4.86E-8	35	221	6.65E-9
1	10	(-1.2,1)	30	-428	4.86E-8	35	501	6.65E-9
1	20	(-12,1)	30	728	4.86E-8	35	851	6.65E-9
2	2	(-1.2,1)	31	195	7.11E-8	26	159	5.29E-6
2	2	(-3.635,5.621)	132	803	3.205E-10	35	229	2.03B-5
2	2	(639,-0.221	280	1818	339E4	44	263	6.13E-10
3	2	(0,0)	5	29	0.2008	5	29	0.2008
3	2	(9,12)	34	201	0.2053	14	79	0.2053

3	2	(30,30)	13	81	0.1998	3	17	0.2723
4	2	(1,1)	27	241	1.35E-6	24	201	4.87E40
4	2	(10,-0.5)	248	1280	3.69E-7	27	214	5.49E-11
4	2	(100,100)	392	1928	3.14E-8	32	260	1.96E-10
5	2	(0,1)	3	16	0.9999	3	16	0.9999
5	2	(0.01,5)	2	9	1	2	9	1
5	2	(3,-0.5)	3	16	-1	3	16	1
6	3	(2,4,2.5)	10	63	2.09E-4	10	68	2.23E-4
6	3	(0.6,6,3.5)	8	53	235E-6	14	91	3.06E-6
6	3	(0.7,5.5,3)	9	59	2.68E-5	12	87	1.03E-5
7	3	(250,03,0)	14	99	7.13E-4	13	91	7.92E-4
7	3	(25,-2,0)	1	4	2.8393	1	4	2.8393
7	3	(200,0.1,40)	17	123	3.20E-4	15	110	3.20E-4
8	10	(1,2,3,…)	5	70	7.60E-5	5	70	7.62E-5
8	20	(1,2,3,..)	17	414	9.47E-4	7	168	2.46E-4
8	30	(1,2,3,..)	53	1817	3.04E-4	265	9047	2.27E-4
8	10	-(1,2,3,..)	5	70	1.34E-4	5	70	'133E-4
8	20	-(1,2,3,..)	17	412	1.67E-4	8 -	199	1.68E-4
8	30	-(1,2,3,…)	59	2022	3.13E-4	188	6031	3.53E-4
9	20	1 – *iln*	25	604	2.20E-9	8	195	6.62E-9
9	30	1 – *iln*	91	3340	4.82E-10	9	309	132E-9
9	40	*1-i/n*	Filed	Filed	Filed	14	145	6.51E-9
9	20	1 - *i/2n*	11	274	3.50E-11	8	191	6.18E-9
9	30	1 - *il2n*	100	3360	1.63E-9	12	400	6.01E-10
9	40	1 - *il2n*	101	4550	2.08E-11	9	402	2.27E-10
10	10	(1,2,3,..)	90	1280	29.0596	1357	17143	9.2360
10	20	(1,2,3,..)	19	269	36.2260	1549	21175	4.7698
11	10	(1,2,3,..)	3	55	2.1429	2	25	2.1429
11	20	(1,2,3,..)	97	2089	4.6341	6	79	4.6341
11	30	(1,2,3,..)	113	3327	7.1311	3	107	7.1311
11	10	-(1,2,3,..)	2	25	2.1429	2	25	2.1429
11	20	-(1,2,3,..)	90	2019	4.6341	4	90	4.6341
11	30	-(1,2,3,..)	96	3003	7.1311	3	106	7.1311

CONCLUSION

Hart update is satisfying the Q-N condition but it's not symmetric, and since the Hessian matrix is symmetric so to preserve the symmetric property is very important. From equation 11 and equation 12, clear that the cholesky update is satisfying Q-N condition and preserve the symmetric property. Cholesky update does not compute the Hessian inverse matrix but this update compute the inverse of Cholesky factor with $2n^3$ (total operations to compute Hessian inverse matrix is $4n^3 - 2n^2$) operations were n represent the dimension of the objective function. From table 1, we can see that the Cholesky update with large dimension has a good results

REFERENCES

Byrd R. H. and Nocedal J. 1989. *A tool for the analysis of quasi-Newton methods with application to unconstrained minimization*. SIAM J. Number. Anal., 26: 727-739

Bartle R. G. 1975. *The elements of real analysis*. Wiley international edition. USA. 286-330

David G. Luenberger and Yinyu Ye. 2009. *linear and Nonlinear Programming.* Third Edition. Springer

Dennis J. E. and Schnabel R. B. 1983. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Hart, W. E. 1990. *Quasi-Newton methods for sparse Nonlinear system*. Memorandum CSM-151, Department of Computer Science, University of Essex.

Royden, H. L. 1968. *Real Analysis*. second edition, Macmillan Publishing co. INC. New York.

Saad S.. 2003. *Unconstrained Optimization Methods Based On Direct Updating of Hessian Factors*. Ph. D. Dissertation. Gadjah Mada University. Yogyakarta, Indonesia.

Steven C. Chapra. 2005."*Applied Numerical Methods With MATLAB for Engineers and Scientists*. McGRAW- HILL INTERNATIONAL EDITION.