Numerical Optimization
Lecture Notes #20
Quasi-Newton Methods — Convergence Analysis

Peter Blomgren,
 ⟨blomgren.peter@gmail.com⟩

Department of Mathematics and Statistics
Dynamical Systems Group
Computational Sciences Research Center
San Diego State University
San Diego, CA 92182-7720

http://terminus.sdsu.edu/

Fall 2017
Outline

1. Convergence Analysis — BFGS (& SR1)
   - Global Convergence of BFGS
   - Superlinear Convergence of BFGS
We look at some local and global convergence results for the BFGS method.

The BFGS results are more general than the results for the SR1 method, hence we omit the SR1 discussion.

Since the Hessian approximations evolve by updating formulas, the analysis of quasi-Newton methods is much more complex than the corresponding analysis for steepest descent and Newton methods.
We cannot prove the following (desirable) result:

- The iterates of a quasi-Newton method approach a stationary point from any starting point and any suitable initial Hessian approximation. \([\text{GLOBAL CONVERGENCE}]\)

In the analysis we must either assume that the objective function is convex, or we must impose restrictions on the iterates.

Under reasonable assumptions we will be able to show local, superlinear convergence results.
In our analysis of global convergence for BFGS in the context of an inexact line search, we need the following assumptions:

**Assumptions**

1. The objective function $f$ is twice continuously differentiable.

2. The level set $\Omega = \{ \bar{x} \in \mathbb{R}^n : f(\bar{x}) \leq f(\bar{x}_0) \}$ is convex, and there exist positive constants $m$ and $M$ such that

   $$m\|\bar{z}\|^2 \leq \bar{z}^T \nabla^2 f(\bar{x}) \bar{z} \leq M\|\bar{z}\|^2$$

   for all $\bar{z} \in \mathbb{R}^n$, and $\bar{x} \in \Omega$.

The second assumption implies that the Hessian is positive definite on $\Omega$ and that $f$ has a unique minimizer $\bar{x}^* \in \Omega$. 
Global Convergence of BFGS: Building Blocks

Using Taylor’s theorem we can express a relation between the quantities $\bar{s}_k = \bar{x}_{k+1} - \bar{x}_k \equiv \alpha_k \bar{p}_k$, and $\bar{y}_k = \nabla f(\bar{x}_{k+1}) - \nabla f(\bar{x}_k)$ using the average Hessian $G_k$ over the step

$$G_k = \int_0^1 \nabla^2 f(\bar{x}_k + \tau \alpha_k \bar{p}_k) \, d\tau$$

to get $\bar{y}_k = G_k \alpha_k \bar{p}_k = G_k \bar{s}_k$.

In combination with the second assumption, we get

$$\frac{\bar{y}_k^T \bar{s}_k}{\bar{s}_k^T \bar{s}_k} = \frac{\bar{s}_k^T G_k \bar{s}_k}{\bar{s}_k^T \bar{s}_k} \geq m$$

Further, since $G_k$ is SPD, its square-root is well defined and we can set $\bar{z}_k = G_k^{1/2} \bar{s}_k$, and we have

$$\frac{\bar{y}_k^T \bar{y}_k}{\bar{y}_k^T \bar{s}_k} = \frac{\bar{s}_k^T G_k^2 \bar{s}_k}{\bar{s}_k^T \bar{s}_k} = \frac{\bar{z}_k^T G_k \bar{z}_k}{\bar{z}_k^T \bar{z}_k} \leq M$$
A positive semi-definite matrix, $M$ has a unique positive semi-definite square root, $R = M^{1/2}$.

When $M = X\Lambda X^{-1} \triangleq Q\Lambda Q^T$, let $R = QSQ^T$, and

$$R^2 = (QSQ^T)^2 = QSQ^T QSQ^T = QSSQ^T = QS^2Q^T = M,$$

showing that

$$S = \Lambda^{1/2}, \text{ and therefore } R = Q\Lambda^{1/2}Q^T$$

∃ other approaches.
Global Convergence of BFGS: Tools

In order to estimate the size of the largest and smallest eigenvalues of the generated Hessian approximations $B_k$ we need two linear algebra tools, the determinant and the trace.

**Definition (Trace)**

The trace of an $n \times n$ matrix $A$ is

$$\text{trace}(A) = \sum_{i=1}^{n} a_{ii}$$

Both the determinant and the trace are invariant under the operations performed by Gaussian elimination (LU-factorization), hence

- The trace is the sum of the eigenvalues of $A$.
- The determinant is the product of the eigenvalues of $A$. 
Global Convergence of BFGS: Theorem

Theorem

Let $B_0$ be any symmetric positive definite initial matrix, and let $\bar{x}_0$ be a starting point for which the stated assumptions are satisfied. Then the sequence $\{\bar{x}_k\}$ generated by the BFGS method converges to the minimizer $\bar{x}^*$ of $f$.

This theorem can be generalized to the entire restricted Broyden class, except for the DFP method.

The analysis can also be extended to show that convergence is rapid enough that

$$
\sum_{k=1}^{\infty} \|\bar{x}_k - \bar{x}^*\| < \infty,
$$

this (eventually) implies \textbf{super-linear} convergence.
First, we define

\[
m_k = \frac{\bar{y}_k^T \bar{s}_k}{\|\bar{s}_k\|^2} \geq m, \quad M_k = \frac{\|\bar{y}_k\|^2}{\bar{y}_k^T \bar{s}_k} \leq M,
\]

where \(m\) and \(M\) are the constants in the assumption.

We compute the trace and determinant of the BFGS-update

\[
\begin{align*}
\text{trace}(B_{k+1}) &= \text{trace}(B_k) - \frac{\|B_k \bar{s}_k\|^2}{\bar{s}_k^T B_k \bar{s}_k} + \frac{\|\bar{y}_k\|^2}{\bar{y}_k^T \bar{s}_k} \\
\text{det}(B_{k+1}) &= \text{det}(B_k) \frac{\bar{y}_k^T \bar{s}_k}{\bar{s}_k^T B_k \bar{s}_k}.
\end{align*}
\]
Further we define the angle between $\bar{s}_k$ and $B_k\bar{s}_k$, $\theta_k$, and the Rayleigh quotient, $q_k$

$$\cos(\theta_k) = \frac{\bar{s}_k^T B_k \bar{s}_k}{\|\bar{s}_k\| \|B_k\bar{s}_k\|}, \quad q_k = \frac{\bar{s}_k^T B_k \bar{s}_k}{\|\bar{s}_k\|^2}.$$ 

From this we can obtain

$$\frac{\|B_k\bar{s}_k\|^2}{\bar{s}_k^T B_k \bar{s}_k} = \frac{\|B_k\bar{s}_k\|^2 \|\bar{s}_k\|^2}{(\bar{s}_k^T B_k \bar{s}_k)^2} \frac{\bar{s}_k^T B_k \bar{s}_k}{\|\bar{s}_k\|^2} = \frac{q_k}{\cos^2(\theta_k)}.$$ 

We can also write

$$\det(B_{k+1}) = \det(B_k) \frac{m_k}{q_k}.$$
We introduce the following positive function, valid for positive definite matrices $B$

$$\Psi(B) = \text{trace}(B) - \ln(\det(B)).$$

Combining our previous results, we get an update formula for $\Psi(\cdot)$

$$\Psi(B_{k+1}) = \Psi(B_k) + (M_k - \ln(m_k) - 1) + \left[ 1 - \frac{q_k}{\cos^2 \theta_k} + \ln \left( \frac{q_k}{\cos^2 \theta_k} \right) \right] + \ln(\cos^2 \theta_k).$$

Since the function $h(t) = 1 - t + \ln t \leq 0$ is non-positive for all $t > 0$ the term inside the square bracket is non-positive...
Using the non-positiveness of \( h(t) \) we get

\[
0 < \Psi(B_{k+1}) \leq \Psi(B_1) + ck + \sum_{j=1}^{k} \ln \cos^2 \theta_j,
\]

where we can assume that the constant \( c = (M - \ln m - 1) \) is positive.

For quasi-Newton methods \( \bar{s}_k = -\alpha_k B_k^{-1} \nabla f(\bar{x}_k) \), so that \( \theta_k \) is the angle between the steepest descent direction and the search direction.

From earlier discussion (on the convergence of line-search methods) we know that \( \| \nabla f(\bar{x}_k) \| \) is bounded away from zero (non-convergence) only if \( \cos \theta_j \to 0 \).
We prove convergence by contradiction: let us assume that $\cos \theta_j \to 0$. Then there exists $k_1 > 0$ such that $\forall j > k_1$ we have

$$\ln \cos^2 \theta_j < -2c,$$

where $c = (M - \ln m - 1)$ (as above). Plugging this into the update-inequality for $\Psi()$ gives, for $k > k_1$

$$0 < \Psi(B_1) + ck + \sum_{j=1}^{k_1} \ln \cos^2 \theta_j + \sum_{j=k_1+1}^{k} (-2c)$$

$$= \Psi(B_1) + \sum_{j=1}^{k_1} \ln \cos^2 \theta_j + 2ck_1 - ck.$$

However, for large enough $k$, the right-hand-side will be negative, resulting in a contradiction. Therefore, $\cos \theta_j \not\to 0$, and $\|\nabla f(\bar{x}_k)\| \sim 0$ (Zoutendijk), and $x_k \to x^*$ (convexity).
We can conclude that there exists a subsequence of indices \( \{ j_k \} \) such that \( \{ \cos \theta_{j_k} \} \geq \delta > 0 \). By Zoutendijk’s result

\[
\sum_{k=0}^{\infty} \cos^2 \theta_k \| \nabla f(\bar{x}_k) \|^2 < \infty
\]

the limit implies that \( \lim \inf \| \nabla f(\bar{x}_k) \| \to 0 \). Since the problem is strongly convex, this shows that \( \bar{x}_k \to \bar{x}^* \). \( \square \)
Superlinear Convergence of BFGS

We need one further assumption in order to show that if

$$\sum_{k=1}^{\infty} \|\bar{x}_k - \bar{x}^*\| < \infty$$

holds, then the rate of convergence is super-linear

**Assumption**

The Hessian matrix $\nabla^2 f(\bar{x})$ is Lipschitz continuous at $\bar{x}^*$, that is

$$\|\nabla^2 f(\bar{x}) - \nabla^2 f(\bar{x}^*)\| \leq L \|\bar{x} - \bar{x}^*\|$$

for all $\bar{x}$ near $\bar{x}^*$, where $L$ is a positive constant.
We introduce

$$\tilde{s}_k = [\nabla^2 f(\bar{x}^*)]^{1/2} s_k, \quad \tilde{y}_k = [\nabla^2 f(\bar{x}^*)]^{-1/2} y_k, \quad \tilde{B}_k = [\nabla^2 f(\bar{x}^*)]^{-1/2} B_k [\nabla^2 f(\bar{x}^*)]^{-1/2}$$

Further we define the angle between $\tilde{s}_k$ and $B_k \tilde{s}_k$, $\tilde{\theta}_k$, and the Rayleigh quotient $q_k$

$$\cos(\tilde{\theta}_k) = \frac{\tilde{s}_k^T B_k \tilde{s}_k}{\|\tilde{s}_k\| \|B_k \tilde{s}_k\|}, \quad q_k = \frac{\tilde{s}_k^T B_k \tilde{s}_k}{\|\tilde{s}_k\|^2}$$

and we let

$$\tilde{M}_k = \frac{\|\tilde{y}_k\|^2}{\tilde{y}_k^T \tilde{s}_k}, \quad \tilde{m}_k = \frac{\tilde{y}_k^T \tilde{s}_k}{\|\tilde{s}_k\|^2}$$
Now, if we pre- and post-multiply the BFGS update formula

\[
B_{k+1} = B_k - \frac{B_k \bar{s}_k \bar{s}_k^T B_k}{\bar{s}_k^T B_k \bar{s}_k} + \frac{\bar{y}_k \bar{y}_k^T}{\bar{y}_k^T \bar{s}_k}
\]

by \([\nabla^2 f(\bar{x}^*)]^{-1/2}\), we obtain

\[
\tilde{B}_{k+1} = \tilde{B}_k - \frac{\tilde{B}_k \tilde{s}_k \tilde{s}_k^T \tilde{B}_k}{\tilde{s}_k^T \tilde{B}_k \tilde{s}_k} + \frac{\tilde{y}_k \tilde{y}_k^T}{\tilde{y}_k^T \tilde{s}_k}
\]

Since this has exactly the same form as the BFGS update, it follows from the argument in the previous proof that

\[
\psi(\tilde{B}_{k+1}) = \psi(\tilde{B}_k) + (\tilde{M}_k - \ln(\tilde{m}_k) - 1)
\]

\[
+ \left[ 1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \left( \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} \right) \right] + \ln(\cos^2 \tilde{\theta}_k)
\]
We recall
\[ \tilde{y}_k = G_k \alpha_k \tilde{p}_k = G_k \tilde{s}_k, \]
so that we can write
\[ \tilde{y}_k - G_* \tilde{s}_k = (G_k - G_*) \tilde{s}_k, \quad \text{where} \ G_* = \nabla^2 f(\bar{x}^*), \]
and pre-multiplying by \( G_*^{-1/2} \) gives
\[ \tilde{y}_k - \tilde{s}_k = G_*^{-1/2} (G_k - G_*) G_*^{-1/2} \tilde{s}_k. \]

By the (assumed) Lipschitz continuity of the Hessian we have
\[ \|\tilde{y}_k - \tilde{s}_k\| \leq \|G_*^{-1/2}\|^2 \|\tilde{s}_k\| \|G_k - G_*\| \leq \|G_*^{-1/2}\|^2 \|\tilde{s}_k\| L \epsilon_k, \]
where
\[ \epsilon_k = \max\{\|\bar{x}_{k+1} - \bar{x}^*\|, \|\bar{x}_k - \bar{x}^*\|\}. \]
It now follows that

\[ \frac{\|\tilde{y}_k - \tilde{s}_k\|}{\|\tilde{s}_k\|} \leq \overline{c}\epsilon_k \]

for some positive constant \( \overline{c} \).

This results, together with

\[ \sum_{k=1}^{\infty} \|\bar{x}_k - \bar{x}^*\| < \infty \]

will be used to show super-linear convergence for BFGS.
We now show the following theorem

**Theorem**

Suppose that $f$ is twice continuously differentiable and that the iterates generated by the BFGS algorithm converge to a minimizer $\bar{x}^*$ at which point the Hessian is Lipschitz continuous. Suppose also that

$$\sum_{k=1}^{\infty} \| \bar{x}_k - \bar{x}^* \| < \infty$$

holds. Then $\bar{x}$ converges to $\bar{x}^*$ at a superlinear rate.
We use the result from the pre-proof:

\[
\frac{\|\mathbf{y}_k - \mathbf{s}_k\|}{\|\mathbf{s}_k\|} \leq \bar{c}\epsilon_k, \quad \Leftrightarrow \quad \|\mathbf{y}_k - \mathbf{s}_k\| \leq \bar{c}\epsilon_k \|\mathbf{s}_k\|,
\]

and the triangle-inequality (\(\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|\)) to obtain

\[
\|\mathbf{y}_k\| - \|\mathbf{s}_k\| \leq \bar{c}\epsilon_k \|\mathbf{s}_k\|, \quad \|\mathbf{s}_k\| - \|\mathbf{y}_k\| \leq \bar{c}\epsilon_k \|\mathbf{s}_k\|.
\]

Now,

\[
(1 - \bar{c}\epsilon_k)\|\mathbf{s}_k\| \leq \|\mathbf{y}_k\| \leq (1 + \bar{c}\epsilon_k)\|\mathbf{s}_k\|.
\]

If we square the result at the top right, we get

\[
\|\mathbf{y}_k\|^2 + \|\mathbf{s}_k\|^2 - 2\mathbf{y}_k^T\mathbf{s}_k \leq \bar{c}^2\epsilon_k^2 \|\mathbf{s}_k\|^2.
\]

We use the left half of the inequality above to obtain...
(1 − \bar{c}\epsilon_k)^2 \|\tilde{s}_k\|^2 + \|\tilde{s}_k\|^2 - 2\tilde{y}_k^T\tilde{s}_k \leq \|\tilde{y}_k\|^2 + \|\tilde{s}_k\|^2 - 2\tilde{y}_k^T\tilde{s}_k \leq \bar{c}^2 \epsilon_k^2 \|\tilde{s}_k\|^2,

and therefore

$$2\tilde{y}_k^T\tilde{s}_k \geq (1 - 2\bar{c}\epsilon_k + \bar{c}^2 \epsilon_k^2 + 1 - \bar{c}^2 \epsilon_k^2)\|\tilde{s}_k\|^2 = 2(1 - \bar{c}\epsilon_k)\|\tilde{s}_k\|^2.$$  

Finally we can say something useful [Above + Triangle Ineq.]

$$\tilde{m}_k = \frac{\tilde{y}_k^T\tilde{s}_k}{\|\tilde{s}_k\|^2} \geq (1 - \bar{c}\epsilon_k), \quad \tilde{M}_k = \frac{\|\tilde{y}_k\|^2}{\tilde{y}_k^T\tilde{s}_k} \leq \frac{1 + \bar{c}\epsilon_k}{1 - \bar{c}\epsilon_k}$$

Since \(\bar{x}_k \rightarrow \bar{x}^*\), we have that \(\epsilon_k \rightarrow 0\). There must exists a \(K\) and \(c > \bar{c}\) so that for all \(k > K\), we have

$$\tilde{M}_k \leq 1 + c\epsilon_k.$$
We use the non-positiveness of the function $h(t) = 1 - t + \ln t$:

$$\frac{-x}{1 - x} - \ln(1 - x) = h\left(\frac{1}{1 - x}\right) \leq 0$$

for large enough $k$ we can assume $\overline{c} \epsilon_k < 1/2$, so that

$$\ln(1 - \overline{c} \epsilon_k) \geq \frac{-\overline{c} \epsilon_k}{1 - \overline{c} \epsilon_k} \geq -2\overline{c} \epsilon_k.$$

Together with

$$\tilde{m}_k = \frac{\tilde{y}_k^T \tilde{s}_k}{\|\tilde{s}_k\|^2} \geq (1 - \overline{c} \epsilon_k),$$

we have

$$\ln \tilde{m}_k \geq \ln(1 - \overline{c} \epsilon_k) \geq \frac{-\overline{c} \epsilon_k}{1 - \overline{c} \epsilon_k} \geq -2\overline{c} \epsilon_k.$$
Putting together

\[ \ln \tilde{m}_k \geq -2c \epsilon_k, \quad \tilde{M}_k \leq 1 + c \epsilon_k, \]

and

\[
\psi(\tilde{B}_{k+1}) = \psi(\tilde{B}_k) + (\tilde{M}_k - \ln(\tilde{m}_k) - 1)
\]

\[
+ \left[ 1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \left( \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} \right) \right] + \ln(\cos^2 \tilde{\theta}_k)
\]

gives

\[
0 < \psi(\tilde{B}_{k+1}) \leq \psi(\tilde{B}_k) + 3c \epsilon_k + \ln(\cos^2 \tilde{\theta}_k)
\]

\[
+ \left[ 1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \left( \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} \right) \right]
\]

We sum this expression over \( k \).
...and make use of
\[ \sum_{k=1}^{\infty} \| \bar{x}_k - \bar{x}^* \| \sim \epsilon_k \]
to get
\[ \sum_{k=0}^{\infty} \left( \ln \left( \frac{1}{\cos^2 \tilde{\theta}_k} \right) - \left[ 1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \left( \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} \right) \right] \right) \geq 0 \]
\[ \leq \psi(\tilde{B}_0) + 3c \sum_{k=0}^{\infty} \epsilon_k < +\infty \]

We must have
\[ \lim_{k \to \infty} \ln \left( \frac{1}{\cos^2 \tilde{\theta}_k} \right) = 0, \quad \lim_{k \to \infty} \left[ 1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \left( \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} \right) \right] = 0 \]
In order for this to be true we must have

\[
\lim_{k \to \infty} \cos \bar{\theta}_k = 1, \quad \lim_{k \to \infty} \bar{q}_k = 1
\]

We are now (not so obviously) done!

In the previous proof we had the relation

\[
\frac{\| G_*^{-1/2} (B_k - G_*) \tilde{s}_k \|^2}{\| G_*^{1/2} \tilde{s}_k \|^2} = \frac{\| (\bar{B}_k - I) \tilde{s}_k \|^2}{\| \tilde{s}_k \|^2}
\]

\[
= \frac{\| \bar{B}_k \tilde{s}_k \|^2 - 2 \tilde{s}_k^T \bar{B}_k \tilde{s}_k + \tilde{s}_k^T \bar{B}_k \tilde{s}_k}{\tilde{s}_k^T \tilde{s}_k}
\]

\[
= \frac{\bar{q}_k^2}{\cos \bar{\theta}_k^2} - 2\bar{q}_k + 1
\]

The right-hand-side converges to zero...
...and we can conclude that

$$\lim_{k \to \infty} \frac{\|(B_k - G_*)\bar{s}_k\|}{\|\bar{s}_k\|} = 0$$

close to the solution, the step $\alpha_k = 1$ will be accepted, so that $\bar{s}_k \equiv \bar{p}_k$, therefore

$$\lim_{k \to \infty} \frac{\|(B_k - G_*)\bar{p}_k\|}{\|\bar{p}_k\|} = 0$$

which establishes \textbf{superlinear convergence}.  \qed

Phew!!!
Convergence Analysis — BFGS (& SR1)

Global Convergence of BFGS

Superlinear Convergence of BFGS

Index