

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 2018

Outline

- 1 Convergence Analysis — BFGS (& SR1)
 - Global Convergence of BFGS
 - Superlinear Convergence of BFGS

Convergence Analysis of the BFGS and SR1 Methods

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.

Quasi-Newton Methods — What We *Cannot* Prove...

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. [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.

Global Convergence of BFGS: Assumptions

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

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

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

The second assumption implies that the Hessian is positive definite on Ω and that f has a unique minimizer $\bar{\mathbf{x}}^* \in \Omega$.

Global Convergence of BFGS: Building Blocks

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

$$G_k = \int_0^1 \nabla^2 f(\bar{\mathbf{x}}_k + \tau \alpha_k \bar{\mathbf{p}}_k) d\tau \quad \text{to get} \quad \bar{\mathbf{y}}_k = \mathbf{G}_k \alpha_k \bar{\mathbf{p}}_k = \mathbf{G}_k \bar{\mathbf{s}}_k.$$

In combination with the second assumption, we get

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

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

$$\frac{\bar{\mathbf{y}}_k^T \bar{\mathbf{y}}_k}{\bar{\mathbf{y}}_k^T \bar{\mathbf{s}}_k} = \frac{\bar{\mathbf{s}}_k^T \mathbf{G}_k^2 \bar{\mathbf{s}}_k}{\bar{\mathbf{s}}_k^T \mathbf{G}_k \bar{\mathbf{s}}_k} = \frac{\bar{\mathbf{z}}_k^T \mathbf{G}_k \bar{\mathbf{z}}_k}{\bar{\mathbf{z}}_k^T \bar{\mathbf{z}}_k} \leq M$$

Square Roots of SPD Matrices

[SUPPLEMENTAL]

- A positive semi-definite matrix, M has a unique positive semi-definite square root, $R = M^{1/2}$.
- When $M = X\Lambda X^{-1} \stackrel{\text{SPD}}{=} Q\Lambda Q^T$, let $R = QSQ^T$, and

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

showing that

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

- \exists 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**, $\det(A)$ 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 (BFGS Global Convergence)

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

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

this (eventually) implies **super-linear** convergence.

Global Convergence of BFGS: Proof

1 of 6

First, we define

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

where m and M are the constants in the assumption.

We compute the trace and determinant of the BFGS-update

$$\begin{cases} \text{trace}(B_{k+1}) &= \text{trace}(B_k) - \frac{\|\mathbf{B}_k \bar{\mathbf{s}}_k\|^2}{\bar{\mathbf{s}}_k^T \mathbf{B}_k \bar{\mathbf{s}}_k} + \frac{\|\bar{\mathbf{y}}_k\|^2}{\bar{\mathbf{y}}_k^T \bar{\mathbf{s}}_k} \\ \det(B_{k+1}), &= \det(B_k) \frac{\bar{\mathbf{y}}_k^T \bar{\mathbf{s}}_k}{\bar{\mathbf{s}}_k^T \mathbf{B}_k \bar{\mathbf{s}}_k}. \end{cases}$$

Global Convergence of BFGS: Proof

2 of 6

Further we define the angle between $\bar{\mathbf{s}}_k$ and $B_k \bar{\mathbf{s}}_k$, θ_k , and the **Rayleigh quotient**, q_k

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

From this we can obtain

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

We can also write

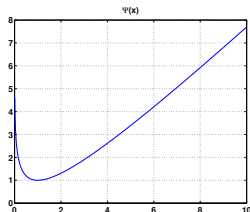
$$\det(B_{k+1}) = \det(B_k) \frac{m_k}{q_k}.$$

Global Convergence of BFGS: Proof

3 of 6

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 $\Psi(\cdot)$

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

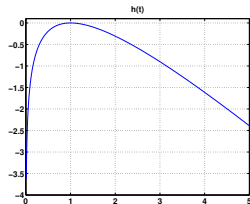
Since the function $h(t) = 1 - t + \ln t \leq 0$ is non-positive for all $t > 0$ the term inside the square bracket is non-positive...

Global Convergence of BFGS: Proof

4 of 6

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 $\mathbf{c} = (\mathbf{M} - \ln \mathbf{m} - 1)$ is positive.

For quasi-Newton methods $\bar{\mathbf{s}}_k = -\alpha_k B_k^{-1} \nabla f(\bar{\mathbf{x}}_k)$, so that θ_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 $\|\nabla f(\bar{\mathbf{x}}_k)\|$ is bounded away from zero (non-convergence) only if $\cos \theta_j \rightarrow 0$.

Global Convergence of BFGS: Proof

We prove convergence by contradiction: let us assume that $\cos \theta_j \rightarrow 0$. 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 $k > k_1$

$$\begin{aligned} 0 &< \Psi(B_1) + ck + \sum_{j=1}^{k_1} \ln \cos^2 \theta_j + \sum_{j=k_1+1}^k (-2c) \\ &= \Psi(B_1) + \sum_{j=1}^{k_1} \ln \cos^2 \theta_j + 2ck_1 - \text{ck} \end{aligned}$$

However, for large enough k , the right-hand-side will be negative, resulting in a contradiction. Therefore, $\cos \theta_j \not\rightarrow 0$, and $\|\nabla f(\bar{x}_k)\| \rightsquigarrow 0$ (Zoutendijk), and $x_k \rightarrow x^*$ (convexity).

Global Convergence of BFGS: Proof

6 of 6

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

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

the limit implies that $\liminf \|\nabla f(\bar{\mathbf{x}}_k)\| \rightarrow 0$. Since the problem is strongly convex, this shows that $\bar{\mathbf{x}}_k \rightarrow \bar{\mathbf{x}}^*$. \square

Superlinear Convergence of BFGS

We need one further assumption in order to show that if

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

holds, then the rate of convergence is super-linear

Assumption

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

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

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

Superlinear Convergence of BFGS: Pre-Proof

1 of 4

We introduce a coordinate transformation

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

In these coordinates, the terminal Hessian is the identity matrix.

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

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

and we let

$$\tilde{M}_k = \frac{\|\tilde{\mathbf{y}}_k\|^2}{\tilde{\mathbf{y}}_k^T \tilde{\mathbf{s}}_k}, \quad \tilde{m}_k = \frac{\tilde{\mathbf{y}}_k^T \tilde{\mathbf{s}}_k}{\|\tilde{\mathbf{s}}_k\|^2}$$

Superlinear Convergence of BFGS: Pre-Proof

Now, if we pre- and post-multiply the BFGS update formula

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

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

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

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

$$\begin{aligned} \Psi(\tilde{B}_{k+1}) &= \Psi(\tilde{B}_k) + (\tilde{M}_k - \ln(\tilde{m}_k) - 1) \\ &\quad + \left[1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \left(\frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} \right) \right] + \ln(\cos^2 \tilde{\theta}_k) \end{aligned}$$

Superlinear Convergence of BFGS: Pre-Proof

3 of 4

We recall

$$\bar{\mathbf{y}}_k = G_k \alpha_k \bar{\mathbf{p}}_k = G_k \bar{\mathbf{s}}_k,$$

so that we can write

$$\bar{\mathbf{y}}_k - G_* \bar{\mathbf{s}}_k = (G_k - G_*) \bar{\mathbf{s}}_k, \quad \text{where } G_* = \nabla^2 f(\bar{\mathbf{x}}^*),$$

and pre-multiplying by $G_*^{-1/2}$ gives

$$\tilde{\mathbf{y}}_k - \tilde{\mathbf{s}}_k = G_*^{-1/2} (G_k - G_*) G_*^{-1/2} \tilde{\mathbf{s}}_k.$$

By the (assumed) Lipschitz continuity of the Hessian we have

$$\|\tilde{\mathbf{y}}_k - \tilde{\mathbf{s}}_k\| \leq \|G_*^{-1/2}\|^2 \|\tilde{\mathbf{s}}_k\| \|G_k - G_*\| \leq \|G_*^{-1/2}\|^2 \|\tilde{\mathbf{s}}_k\| L \epsilon_k,$$

where

$$\epsilon_k = \max\{\|\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}^*\|, \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}^*\|\}.$$

Superlinear Convergence of BFGS: Pre-Proof

It now follows that

$$\frac{\|\tilde{\mathbf{y}}_k - \tilde{\mathbf{s}}_k\|}{\|\tilde{\mathbf{s}}_k\|} \leq \bar{c}\epsilon_k$$

for some positive constant \bar{c} .

This results, together with

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

will be used to show super-linear convergence for BFGS.

Superlinear Convergence of BFGS: Theorem

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 $\bar{\mathbf{x}}^$ at which point the Hessian is Lipschitz continuous. Suppose also that*

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

holds. Then $\bar{\mathbf{x}}$ converges to $\bar{\mathbf{x}}^$ at a superlinear rate.*

Superlinear Convergence of BFGS: Proof

1 of 7

We use the result from the pre-proof:

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

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

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

Now,

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

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

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

We use the **left half** of the inequality above to obtain...

Superlinear Convergence of BFGS: Proof

2 of 7

$$(1 - \bar{c}\epsilon_k)^2 \|\tilde{\mathbf{s}}_k\|^2 + \|\tilde{\mathbf{s}}_k\|^2 - 2\tilde{\mathbf{y}}_k^T \tilde{\mathbf{s}}_k \leq \|\tilde{\mathbf{y}}_k\|^2 + \|\tilde{\mathbf{s}}_k\|^2 - 2\tilde{\mathbf{y}}_k^T \tilde{\mathbf{s}}_k \leq \bar{c}^2 \epsilon_k^2 \|\tilde{\mathbf{s}}_k\|^2,$$

and therefore

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

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

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

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

$$\tilde{M}_k \leq 1 + c\epsilon_k.$$



Superlinear Convergence of BFGS: Proof

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 $\bar{c}\epsilon_k < 1/2$, so that

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

Together with

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

we have

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

Superlinear Convergence of BFGS: Proof

Putting together

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

and

$$\begin{aligned} \Psi(\tilde{B}_{k+1}) &= \Psi(\tilde{B}_k) + (\tilde{M}_k - \ln(\tilde{m}_k) - 1) \\ &\quad + \left[1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \left(\frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} \right) \right] + \ln(\cos^2 \tilde{\theta}_k) \end{aligned}$$

gives

$$\begin{aligned} 0 < \Psi(\tilde{B}_{k+1}) &\leq \Psi(\tilde{B}_k) + 3c\epsilon_k + \ln(\cos^2 \tilde{\theta}_k) \\ &\quad + \left[1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \left(\frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} \right) \right] \end{aligned}$$

We sum this expression over k ...

Superlinear Convergence of BFGS: Proof

5 of 7

...and make use of

$$\sum_{k=1}^{\infty} \underbrace{\|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}^*\|}_{\sim \epsilon_k} < \infty$$

to get

$$\begin{aligned} & \sum_{k=0}^{\infty} \left(\underbrace{\ln \left(\frac{1}{\cos^2 \tilde{\theta}_k} \right)}_{\geq 0} - \underbrace{\left[1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \left(\frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} \right) \right]}_{\leq 0} \right) \\ & \leq \psi(\tilde{B}_0) + 3c \sum_{k=0}^{\infty} \epsilon_k < +\infty \end{aligned}$$

We must have

$$\lim_{k \rightarrow \infty} \ln \left(\frac{1}{\cos^2 \tilde{\theta}_k} \right) = 0, \quad \lim_{k \rightarrow \infty} \left[1 - \frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} + \ln \left(\frac{\tilde{q}_k}{\cos^2 \tilde{\theta}_k} \right) \right] = 0$$

Superlinear Convergence of BFGS: Proof

In order for this to be true we must have

$$\lim_{k \rightarrow \infty} \cos \tilde{\theta}_k = 1, \quad \lim_{k \rightarrow \infty} \tilde{q}_k = 1$$

We are now (not so obviously) done!

In the previous proof we had the relation

$$\begin{aligned} \frac{\|G_*^{-1/2}(B_k - G_*)\tilde{\mathbf{s}}_k\|^2}{\|G_*^{1/2}\tilde{\mathbf{s}}_k\|^2} &= \frac{\|(\tilde{B}_k - I)\tilde{\mathbf{s}}_k\|^2}{\|\tilde{\mathbf{s}}_k\|^2} \\ &= \frac{\|\tilde{B}_k\tilde{\mathbf{s}}_k\|^2 - 2\tilde{\mathbf{s}}_k^T \tilde{B}_k \tilde{\mathbf{s}}_k + \tilde{\mathbf{s}}_k^T \tilde{\mathbf{s}}_k}{\tilde{\mathbf{s}}_k^T \tilde{\mathbf{s}}_k} \\ &= \frac{\tilde{q}_k^2}{\cos^2 \tilde{\theta}_k} - 2\tilde{q}_k + 1 \end{aligned}$$

The right-hand-side converges to zero...

Superlinear Convergence of BFGS: Proof

7 of 7

...and we can conclude that (recall $G_* = \nabla^2 f(\bar{\mathbf{x}}^*)$)

$$\lim_{k \rightarrow \infty} \frac{\|(B_k - \nabla^2 f(\bar{\mathbf{x}}^*))\bar{\mathbf{s}}_k\|}{\|\bar{\mathbf{s}}_k\|} = 0$$

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

$$\lim_{k \rightarrow \infty} \frac{\|(B_k - \nabla^2 f(\bar{\mathbf{x}}^*))\bar{\mathbf{p}}_k\|}{\|\bar{\mathbf{p}}_k\|} = 0$$

which establishes **superlinear convergence**. \square

Phew!!! (a.k.a. "That was trivially obvious!")



Index

assumptions

 Global Convergence of BFGS, 5

 Superlinear Convergence of BFGS, 16

Rayleigh quotient, 11

theorem

 BFGS global convergence, 9

trace of a matrix, $\text{trace}(A)$, 8