CONVERGENCE OF QUASI-NEWTON METHODS

David F. Gleich

April 25, 2023

In this lecture note, we'll show that the BFGS method is globally convergent for the problem:

$$\text{minimize} \quad f(\mathbf{x}) \quad .$$

The material here is from Section 6.4 in Nocedal and Wright.

1 THE ALGORITHM

We begin by restarting the BFGS Quasi-Newton method using an approximation of the inverse Hessian

$$\boldsymbol{H}(\mathbf{x}_k)^{-1} \approx \boldsymbol{T}_k.$$

```
Given a starting point $\mathbf{x}_0$ and tolerance $\varepsilon$,
  and inverse Hessian $\boldsymbol{T}_0$
Let $k$ start at 0.
While $\|\mathbf{g}_k\| \ge \varepsilon$
  Compute search direction $\mathbf{p}_k = -\boldsymbol{T}_k \mathbf{g}_k$
    Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha \mathbf{p}_k$ based
      on a strong Wolfe line search.
    Define $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$
    Define $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$
    Define $\rho_k = 1/\mathbf{y}_k^T \mathbf{s}_k$
    Set $\boldsymbol{T}_{k+1} = (\boldsymbol{I} - \rho_k \mathbf{s}_k \mathbf{y}_k^T)\boldsymbol{T}_k(\boldsymbol{I} - \rho_k \mathbf{y}_k \mathbf{s}_k) + \rho_k \mathbf{s}_k \mathbf{s}_k^T$
```

2 THE ASSUMPTIONS!

f is twice continuously differentiable.

We need this assumption otherwise the Hessian isn't defined.

the level set below the function $f(\mathbf{x}_0)$ is convex, i.e. the set $L = \{\mathbf{x} : f(\mathbf{x}) \le f(\mathbf{x}_0)\}$ is convex.

In the proof, we'll need to study the behavior of the function along the lines $\mathbf{x}_k \to \mathbf{x}_{k+1}$. Without this assumption, this step could become complicated.

on the set L, the Hessian $\boldsymbol{H}(\mathbf{x})$ satisfies:

$$m\|\mathbf{z}\|^2 \le \mathbf{z}^T \boldsymbol{H}(\mathbf{x})\mathbf{z} \le M\|\mathbf{z}\|^2$$

for all $\mathbf{z} \in \mathbb{R}$ at any $\mathbf{x} \in L$.

We need this to be able to control the behavior of the true Hessian and make sure our approximation is well posed.

Although we did not assume that the minimizer is unique, these assumptions imply that as well! This is because we are looking at an everywhere positive definite Hessian in a convex set.

3 THE RESULT

THEOREM 1 (Nocedal & Wright 6.5) *Let $\boldsymbol{B}_0$ be any symmetric, positive definite approximation of the Hessian $\boldsymbol{H}_0$ and let $\mathbf{x}_0$ be a starting point for which the previous assumptions are satisfied. Then the sequence $\mathbf{x}_k$ generated by the BFGS Quasi-Newton method converges to the minimizer $\mathbf{x}^*$ of f.*

4 THE PROOF

ARCHITECTURE

In this proof, we actually work with the BFGS approximation of the Hessian, instead of the inverse Hessian. Thus, the search direction is the solution of the linear system:

$$\boldsymbol{B}_k \mathbf{p}_k = -\mathbf{g}_k$$

and the BFGS update (when stated about the Hessian) is:

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

Our goal in the proof is to apply Zoutendijk's conditions to show that the gradient must converge to zero. As there is only one minimizer in the region, this result implies that the quasi-Newton method must find it.

Let $\cos\theta_j$ be the angle between the search direction $\mathbf{p}_k$ and the gradient descent direction $-\mathbf{g}_k$:

$$\cos\theta_j = \frac{-\mathbf{g}_k^T \mathbf{p}_k}{\|\mathbf{g}_k\| \, \|\mathbf{p}_k\|}.$$

If $\cos\theta_j \to 0$, then the algorithm may not converge. The proof we present shows this fact by way of contradiction. That is, we show that if $\cos\theta_j$ goes to zero, we contradict another fact we know is true.

The first part of this proof is just getting our expressions all lined up so that we can put the pieces together to get the result. Some of this work will seem out-of-place until you see how it's used.

4.1 STEP 1

We first show that:

$$\cos\theta_j = \frac{\mathbf{s}_k^T \boldsymbol{B}_k \mathbf{s}_k}{\|\mathbf{s}_k\| \, \|\boldsymbol{B}_k \mathbf{s}_k\|}.$$

This is important because it translates the problem back into something we have direct control over: the matrix $\boldsymbol{B}_k$. In a future step of the proof, we'll try and translate the upper-and-lower-bounds on the positive definiteness of the Hessian matrix into a bound on the properties of $\boldsymbol{B}_k$.

Proving this step is just algebra. Try and work it out. If not, see the cheat-sheet at the bottom of this paper.

Note that $\mathbf{p}_k^T / \|\mathbf{p}_k\| = \mathbf{s}_k^T / \|\mathbf{s}_k\|$. Also note that $-\mathbf{g}_k = \boldsymbol{B}_k \mathbf{p}_k = \boldsymbol{B}_k \frac{\mathbf{x}_{k+1} - \mathbf{x}_k}{\alpha} = \alpha^{-1} \boldsymbol{B}_k \mathbf{s}_k$. Combining these two substitutions provides the result.

4.2 STEP 2

This comes out of the blue, but bear with us for a second. Consider the function

$$\phi(\boldsymbol{A}) = \operatorname{trace}(\boldsymbol{A}) - \log\det(\boldsymbol{A}).$$

This function measures the sum of the eigenvalue and subtracts the log of their product:

$$\phi(\boldsymbol{A}) = \sum \lambda_i - \sum \log \lambda_i = \sum (\lambda_i - \log \lambda_i).$$

Note that the scalar function $x - \log x \geq 1$ for all $x > 0$, thus:

$$\phi(\boldsymbol{A}) \geq n > 0$$

for any positive definite matrix $\boldsymbol{A}$. In particular, $\phi(\boldsymbol{B}_k) > 0$ because $\boldsymbol{B}_k$ is positive at each step (by assumption).

The idea with this function is that we want to get some control over the matrix $\boldsymbol{B}_k$.

4.3 STEP 3

In particular, we'll show shortly that:

$$\phi(\boldsymbol{B}_{k+1}) \le \phi(\boldsymbol{B}_0) + c(k+1) + \sum_{j=0}^{k} \log\cos^2\theta_j.$$

Here c is just some positive constant (it'll depend on m and M, actually). But before we show that, let's see how we'll be done once we do it.

Suppose that $\cos\theta_j \to 0$ then $\cos^2\theta_j \to 0$ as well.[1] We'll show that, if this is true, then $\phi(\boldsymbol{B}_{k+1})$ must become negative at some point. But this contradicts our construction that $\boldsymbol{B}_k$ is positive definite at each step (that's how we chose it!) and hence, that can't be.

How will this become negative? Look at the expression, we have a $\log\cos^2\theta_j$ in the sum. If $\cos\theta_j \to 0$, then $\log\cos^2\theta_j$ will take a value below any negative number. Let t be the step when $\log\cos^2\theta_j < -2c$.

For $k > t$,

$$0 < \phi(\boldsymbol{B}_0) + c(k+1) + \sum_{j=0}^{k} \log\cos^2\theta_j \le \phi(\boldsymbol{B}_0) + c(k+1) + \sum_{j=0}^{t} \log\cos^2\theta_j + (-2c)(k-t)$$

Put another way, we have:

$$0 < \text{constant} + c(k+1) + \text{constant} - 2c(k-t) = \text{constant} + 2ct + c - ck.$$

Once $k > t$, we are *losing* a factor of c in this expression for each additional term! Thus, we can drive this down to something below zero and reach our contradiction.

Hence, we have that $\cos\theta_j$ cannot go to zero. It's still possible that it visits zero periodically, but check the Zoutendijk condition on a strongly convex objective. This result implies that we'll converge to the minimizer.

4.4 STEP 4

What remains is to show that:

$$\phi(\boldsymbol{B}_{k+1}) \le \phi(\boldsymbol{B}_0) + c(k+1) + \sum_{j=0}^{k} \log\cos^2\theta_j.$$

We'll do this in a few steps. First, we want to show that

$$\phi(\boldsymbol{B}_{k+1}) = \phi(\boldsymbol{B}_k) + c_k - d_k + \log\cos^2\theta_k,$$

where c_k is arbitrary and d_k is positive term. Then, if we can show that $c_k < c$ for all k, we will have our result:

$$\begin{aligned} 0 < \phi(\boldsymbol{B}_{k+1}) &\le \phi(\boldsymbol{B}_k) + c + \log\cos^2\theta_k \\ &\le \phi(\boldsymbol{B}_{k-1} + c + \log\cos^2\theta_{k-1} + c + \log\cos^2\theta_k \\ &\le \phi(\boldsymbol{B}_0)c(k+1) + \sum_{j=0}^{k} \log\cos^2\theta_k \end{aligned}$$

[1] This piece is the key part of the proof by contradiction.

4.5 STEP 5

Let's prove the bound on each term:

$$\phi(\boldsymbol{B}_{k+1}) = \operatorname{trace}(\boldsymbol{B}_{k+1}) - \log\det(\boldsymbol{B}_{k+1}) = \phi(\boldsymbol{B}_k) + c_k - d_k + \log\cos^2\theta_k,$$

where c_k is arbitrary and d_k is positive term.

Recall

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

Let's work on the terms $\operatorname{trace}(\boldsymbol{B}_{k+1})$ and $\det(\boldsymbol{B}_{k+1})$ separately:

$$\operatorname{trace}(\boldsymbol{B}_{k+1}) = \operatorname{trace}(\boldsymbol{B}_k) - \operatorname{trace}\left(\frac{\boldsymbol{B}_k \mathbf{s}_k \mathbf{s}_k^T \boldsymbol{B}_k}{\mathbf{s}_k^T \boldsymbol{B}_k \mathbf{s}_k}\right) + \operatorname{trace}\left(\frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}\right) = \operatorname{trace}(\boldsymbol{B}_k) - \frac{\|\boldsymbol{B}_k \mathbf{s}_k\|^2}{\mathbf{s}_k^T \boldsymbol{B}_k \mathbf{s}_k} + \frac{\|\mathbf{y}_k\|^2}{\mathbf{y}_k^T \mathbf{s}_k}.$$

This derivation holds because the trace of two vectors is their inner-product: $\operatorname{trace}(\mathbf{f}\mathbf{g}^T) = \mathbf{f}^T\mathbf{g}$, which holds right from the definition of trace as the sum of the diagonal.

The determinant will be harder to handle because $\det(\boldsymbol{A} + \boldsymbol{B}) \neq \det(\boldsymbol{A}) + \det(\boldsymbol{B})$ as was true for the trace. What does hold for the determinant is $\det(\boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$. We'll use this property:

$$\det(\boldsymbol{B}_{k+1}) = \det\left(\boldsymbol{B}_k\left(\boldsymbol{I} - \frac{\mathbf{s}_k \mathbf{s}_k^T \boldsymbol{B}_k}{\mathbf{s}_k^T \boldsymbol{B}_k \mathbf{s}_k} + \frac{\boldsymbol{B}_k^{-1} \mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k}\right)\right) = \det(\boldsymbol{B}_k)\det(\text{the rest}).$$

To deal with "the rest", we'll first need to one small results:[2]

$$\det(\boldsymbol{I} + \mathbf{x}\mathbf{y}^T + \mathbf{u}\mathbf{v}^T) = (1 + \mathbf{y}^T\mathbf{x})(1 + \mathbf{v}^T\mathbf{u}) - \mathbf{x}^T\mathbf{v}\mathbf{y}^T\mathbf{u}.$$

If we apply this result, then

$$\det(\text{the rest}) = \frac{\mathbf{y}_k^T \mathbf{s}_k}{\mathbf{s}_k^T \boldsymbol{B}_k \mathbf{s}_k}.$$

At this point, we'll introduce a few terms that will help us simplify these expressions:

$$m_k = \frac{\mathbf{y}_k^T \mathbf{s}_k}{\|\mathbf{s}_k\|^2} \quad M_k = \frac{\|\mathbf{y}_k\|^2}{\mathbf{s}_k^T \mathbf{y}_k} \quad q_k = \frac{\mathbf{s}_k^T \boldsymbol{B}_k \mathbf{s}_k}{\mathbf{s}_k^T \mathbf{s}_k} \quad \cos\theta_j = \frac{\mathbf{s}_k^T \boldsymbol{B}_k \mathbf{s}_k}{\|\mathbf{s}_k\|\,\|\boldsymbol{B}_k \mathbf{s}_k\|}$$

The last expression was what we worked out in Step 1. We can insert this into the expression for the determinant:

$$\det(\boldsymbol{B}_{k+1}) = \det(\boldsymbol{B}_k) m_k / q_k$$

Using the same expressions for the trace, we have:

$$\operatorname{trace}(\boldsymbol{B}_{k+1}) = \operatorname{trace}(\boldsymbol{B}_k) + M_k - q_k / \cos^2\theta_j.$$

Hence,

$$\begin{aligned}\phi(\boldsymbol{B}_{k+1}) &= \operatorname{trace}(\boldsymbol{B}_k) + M_k - \frac{q_k}{\cos^2\theta_j} - \log\det(\boldsymbol{B}_k) - \log m_k + \log q_k \\ &= \phi(\boldsymbol{B}_k) + M_k - \log m_k - 1 + \left(1 - \frac{q_k}{\cos^2\theta_j} - \log\cos^2\theta_k + \log q_k\right) + \log\cos^2\theta_k\end{aligned}$$

Recall that $1 - t + \log(t) < 0$ for all $t > 0$. From this, we conclude that

$$1 - \frac{q_k}{\cos^2\theta_j} - \log\cos^2\theta_k + \log q_k = 1 - \frac{q_k}{\cos^2\theta_j} + \log(q_k/\cos^2\theta_k) < 0.$$

And we've finished this step!

$$\phi(\boldsymbol{B}_{k+1}) = \phi(\boldsymbol{B}_k) + \underbrace{(M_k - \log m_k - 1)}_{c_k} + \underbrace{(1 - \frac{q_k}{\cos^2\theta_j} + \log(q_k/\cos^2\theta_k))}_{-d_k} + \log\cos^2\theta_k.$$

[2] Which is worked out in more detail in Exercise 6.10 in the book starting with the simple case: $\det(\boldsymbol{I} + \mathbf{x}\mathbf{y}^T) = 1 + \mathbf{y}^T\mathbf{x}$.

4.6 STEP 6

In the final step, we just need to show that $c_k < c$. Each term $c_k = M_k - \log m_k - 1$. We now show that the assumptions we have imply that $M_k < M$ and $m_k > m$ for the values of M and m given in the assumptions.[3]

For this task, we first use a nice property of Quasi-Newton methods that relates to the *average Hessian between the time-steps:*

$$\begin{aligned}
\mathbf{y}_k &= \mathbf{g}_{k+1} - \mathbf{g}_k = \mathbf{g}(\mathbf{x}_k + \alpha\mathbf{p}_k) - \mathbf{g}(\mathbf{x}_k) \\
&= \underbrace{\int_0^1 \boldsymbol{H}(\mathbf{x} + \alpha\tau\mathbf{p})\alpha\mathbf{p}\, d\tau}_{\text{This is Taylor's theorem again!}} \\
&= \int_0^1 \boldsymbol{H}(\mathbf{x} + \alpha\tau\mathbf{p})\mathbf{s}_k\, d\tau = \underbrace{\left[\int_0^1 \boldsymbol{H}(\mathbf{x} + \alpha\tau\mathbf{p})\, d\tau\right]}_{\equiv \bar{\boldsymbol{H}}_k} \mathbf{s}_k \\
&= \bar{\boldsymbol{H}}_k \mathbf{s}_k.
\end{aligned}$$

Thus:

$$\frac{\mathbf{y}_k^T \mathbf{s}_k}{\mathbf{s}_k^T \mathbf{s}_k} = \frac{\mathbf{s}_k^T \bar{\boldsymbol{H}}_k \mathbf{s}_k}{\mathbf{s}_k^T \mathbf{s}_k} \geq m,$$

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

where $\mathbf{z}_k = \bar{\boldsymbol{H}}_k^{1/2} \mathbf{s}_k$.

Consequently, we have $c_k \leq M - \log m - 1$. If $M - \log m - 1$ happens to be negative, then we can just increase M until it becomes positive and the rest of the theorem falls into place!

4.7 STEP 7

Give yourself a pat on the back! That was a lot of work!

4.8 CONVERGENCE RATE

In Theorem 6.6, Nocedal and Wright utilize a characterization of super-linear convergence from Theorem 3.6. At a high level, the idea is to transform the iterates to look at them in the space of the Hessian nearby the solution. This involves looking at very similar quantities, but *transformed* by $\boldsymbol{H}_*^{-1/2}$ where $\boldsymbol{H}_*$ is the Hessian at the solution.

In a nut-shell, Theorem 3.6 states: *if the Hessian and the approximate Hessian behave the same on the step* $\mathbf{s}$*, then we'll get super-linear convergence.* Formally, the statement we need to show about Quasi-Newton is that:

$$\lim_{k \to \infty} \frac{\|(\boldsymbol{B}_k - \boldsymbol{H}_*)\mathbf{s}\|}{\|\mathbf{s}\|} \to 0.$$

Checkout Theorem 6.6 to see how this is done!

[3] These were "on the set L, the Hessian $\boldsymbol{H}(\mathbf{x})$ satisfies:

$$m\|\mathbf{z}\|^T \leq \mathbf{z}^T \boldsymbol{H}(\mathbf{x})\mathbf{z} \leq M\|\mathbf{z}\|^2$$

for all $\mathbf{z} \in \mathbb{R}$ at any $\mathbf{x} \in L$."