DSA3102: Superlinear Convergence of Newton's Method Near the Optimum

Vincent Y. F. Tan

November 4, 2023

In this document, we provided an analysis of Newton's method near the optimum. We show that the convergence is quadratic under some conditions. The material here is based on the book by Bertsekas [Ber99].

Newton's method consists of the iterations.

$$x^{(k+1)} = x^{(k)} - t_k(\nabla^2 f(x^{(k)}))^{-1}\nabla f(x^{(k)}). \tag{1}$$

assuming the Newton direction

$$\Delta x^{(k)} := -(\nabla^2 f(x^{(k)}))^{-1}\nabla f(x^{(k)}) \tag{2}$$

is defined and is a direction of descent, i.e.,

$$(\Delta x^{(k)})^T \nabla f(x^{(k)}) < 0. \tag{3}$$

When the backtracking line search is used with initial stepsize 1, no reduction in the stepsize is necessary near a non-singular minimum (positive definite Hessian). Thus, near convergence, the method takes the form

$$x^{(k+1)} = x^{(k)} - (\nabla^2 f(x^{(k)}))^{-1}\nabla f(x^{(k)}). \tag{4}$$

In other words, we take $t_k = 1$ for all k. Here, we only discuss the situation when the iterates are near the minimum, known as *local convergence analysis*.

Newton's method can be seen as a solver for the following system of n equations with n unknowns:

$$g(x) = 0 \tag{5}$$

where $g : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable function. This method has the form

$$x^{(k+1)} = x^{(k)} - (\nabla g(x^{(k)}))^{-1} g(x^{(k)}) \tag{6}$$

and for the special case $g(x) = \nabla f(x)$, we recover Newton's method.

There is a simple but informal argument that shows the fast convergence of Newton's method. Suppose the method generates a sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ that converges to a vector x^* satisfying $g(x^*) = 0$ and $\nabla g(x^*)$ is invertible. Then by Taylor's theorem,

$$0 = g(x^*) = g(x^{(k)}) + (x^{(k)} - x^*)^T \nabla g(x^{(k)}) + o(\|x^{(k)} - x^*\|). \tag{7}$$

By multiplying this relation with $(\nabla g(x^{(k)}))^{-1}$, we obtain

$$x^{(k)} - x^* - (\nabla g(x^{(k)}))^{-1} g(x^{(k)}) = o(\|x^{(k)} - x^*\|). \tag{8}$$

so for the pure Newton iterations in (6), we obtain

$$x^{(k)} - x^* = o(\|x^{(k)} - x^*\|). \tag{9}$$

or

$$\lim_{k\to\infty} \frac{\|x^{(k+1)} - x^*\|}{\|x^{(k)} - x^*\|} = 0. \tag{10}$$

Hence, the convergence is *superlinear*. This is faster than gradient descent which recall, typically guarantees that

$$\|x^{(k+1)} - x^*\| \le c\|x^{(k)} - x^*\| \tag{11}$$

for some c that is related to the condition number of $\nabla^2 f(x)$. Hence, we can only say that

$$\limsup_{k\to\infty} \frac{\|x^{(k+1)} - x^*\|}{\|x^{(k)} - x^*\|^1} \le c, \tag{12}$$

which is called *linear convergence*.

More formally, we state and prove the following theorem, which shows that Newton's method guarantees quadratic convergence locally and under the right conditions.

Theorem 1. *Consider a function $g : \mathbb{R}^n \to \mathbb{R}^n$ and a vector x^* such that $g(x^*) = 0$. For $\delta > 0$, let $S_\delta := \{x : \|x - x^*\| \le \delta\}$. Assume g is continuously differentiable within some sphere S_δ and that $\nabla g(x^*)$ is invertible.*

1. *There exists some $\delta > 0$ such that if $x^{(0} \in S_\delta$, the sequence $\{x^{(k)}\}$ generated by the iterates in* (6) *converges to x^*. Furthermore, $\{\|x^{(k)} - x^*\|\}$ converges superlinearly.*
2. *If for some $L > 0$, $M > 0$ and $\delta > 0$ and for all $x, y \in S_\delta$,*

$$\|\nabla g(x) - \nabla g(y)\| \le L\|x - y\| \tag{13}$$

$$\|(\nabla g(x))^{-1}\| \le M \tag{14}$$

then if $x^{(0)} \in S_\delta$, we have

$$\|x^{(k+1)} - x^*\| \le \frac{LM}{2}\|x^{(k)} - x^*\|^2, \qquad \forall\, k \in \mathbb{N} \cup \{0\} \tag{15}$$

so if $LM\delta/2 < 1$, $\{\|x^{(k)} - x^*\|\}$ converges superlinearly with order at least 2, i.e.,

$$\limsup_{k\to\infty} \frac{\|x^{(k+1)} - x^*\|}{\|x^{(k)} - x^*\|^2} \le \frac{LM}{2}. \tag{16}$$

It is instructive to carefully compare (12) to (16). The proof is based on some basic calculus and careful use of the conditions of the theorem.

Proof. Choose $\delta > 0$ such that $(\nabla g(x))^{-1}$ exists for all $x \in S_\delta$. Let $M > 0$ be chosen to satisfy

$$\|(\nabla g(x))^{-1}\| \le M, \qquad \forall\, x \in S_\delta. \tag{17}$$

Assume that $x \in S_\delta$. Now we claim that the following relation holds

$$g(x^{(k)}) = \int_0^1 \nabla g(x^* + t(x^{(k)} - x^*))^T (x^{(k)} - x^*)\, \mathrm{d}t. \tag{18}$$

This relation can be verified as follows. Let $\tilde{g}(t) := g(x^* + t(x^{(k)} - x^*))$. Then it is true that

$$\tilde{g}(1) - \tilde{g}(0) = \int_0^1 \frac{\mathrm{d}\tilde{g}}{\mathrm{d}t}\, \mathrm{d}t \tag{19}$$

However, by the chain rule,

$$\frac{\mathrm{d}\tilde{g}}{\mathrm{d}t} = (\nabla g(x^* + t(x^{(k)} - x^*))^T (x^{(k)} - x^*). \tag{20}$$

So plugging this into the integral and noting that $\tilde{g}(0) = g(x^*) = 0$, we recover (18). Now, we estimate $\|x^{(k+1)} - x^*\|$ as follows:

$$\begin{aligned}
\|x^{(k+1)} - x^*\| &= \left\| x^{(k)} - x^* - (\nabla g(x^{(k)}))^{-1} g(x^{(k)}) \right\| && (21)\\
&= \left\| (\nabla g(x^{(k)}))^{-1} \left((\nabla g(x^{(k)}))(x^{(k)} - x^*) - g(x^{(k)}) \right) \right\| && (22)\\
&= \left\| (\nabla g(x^{(k)}))^{-1} \left(\nabla g(x^{(k)}) - \int_0^1 \nabla g(x^* + t(x^{(k)} - x^*))^T \,\mathrm{d}t \right) (x^{(k)} - x^*) \right\| && (23)\\
&= \left\| (\nabla g(x^{(k)}))^{-1} \left(\int_0^1 \nabla g(x^{(k)}) - \nabla g(x^* + t(x^{(k)} - x^*))^T \,\mathrm{d}t \right) (x^{(k)} - x^*) \right\| && (24)\\
&\le M \left(\int_0^1 \left\| \nabla g(x^{(k)}) - \nabla g(x^* + t(x^{(k)} - x^*)^T \right\| \,\mathrm{d}t \right) \|x^{(k)} - x^*\| && (25)
\end{aligned}$$

where (23) follows from (18). By continuity of ∇g, we can take δ sufficiently small to ensure that the term under the integral sign is arbitrarily small. The convergence of $x^{(k)}$ to x^* and the superlinear convergence of $\{\|x^{(k)} - x^*\|\}$ are established.

Now, for part 2, the preceding relations simplify to

$$\|x^{(k+1)} - x^*\| \le M \left(\int_0^1 Lt \|x^{(k)} - x^*\| \,\mathrm{d}t \right) \|x^{(k)} - x^*\| = \frac{LM}{2} \|x^{(k)} - x^*\|^2 \tag{26}$$

and we are done. □

References

[Ber99] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 1999.