Newton's Method

Equation Solving

$\mathcal{E}$ **:** Given $g : \mathbf{E} \to \mathbf{E}$ continuously differentiable, find $x \in \mathbf{E}$ for which $g(x) = 0$.

Equation Solving

$\mathcal{E}$ **:** Given $g : \mathbf{E} \to \mathbf{E}$ continuously differentiable, find $x \in \mathbf{E}$ for which $g(x) = 0$.

Newton's Approach:
Given an approximate solution $x_0 \in \mathbf{E}$ to $\mathcal{E}$ and wishes to improve upon it. If $\overline{x}$ is an actual solution to $\mathcal{E}$, then

$$0 = g(\overline{x}) = g(x_0) + g'(x_0)(\overline{x} - x_0) + o\|\overline{x} - x_0\|.$$

Thus, if x_0 is "close" to $\overline{x}$, it is reasonable to suppose that the solution to the linearized system

$$0 = g(x_0) + g'(x_0)(x - x_0)$$

is even closer.

Equation Solving

$\mathcal{E}$ **:** Given $g : \mathbf{E} \to \mathbf{E}$ continuously differentiable, find $x \in \mathbf{E}$ for which $g(x) = 0$.

Newton's Approach:
Given an approximate solution $x_0 \in \mathbf{E}$ to $\mathcal{E}$ and wishes to improve upon it. If $\overline{x}$ is an actual solution to $\mathcal{E}$, then

$$0 = g(\overline{x}) = g(x_0) + g'(x_0)(\overline{x} - x_0) + o\|\overline{x} - x_0\|.$$

Thus, if x_0 is "close" to $\overline{x}$, it is reasonable to suppose that the solution to the linearized system

$$0 = g(x_0) + g'(x_0)(x - x_0)$$

is even closer.

To implement Newton's method we need conditions under which solving this equation is meaningful. In particular, we assume that $g'(x_0)$ is nonsingular.

Newton and Newton Like Iterations

The Newton iteration:

$$x_{k+1} := x_k - [g'(x_i)]^{-1} g(x_k).$$

The associated search direction

$$d := -[g'(x_k)]^{-1} g(x_k).$$

is called the Newton direction.

We analyze the convergence behavior of this scheme under the additional assumption that only an approximation to $g'(x_k)$ is available.

Denote the approximation to $g'(x_k)$ by J_k. The Newton-Like iteration scheme:

$$x_{k+1} := x_k - J_k^{-1} g(x_k).$$

Convergence of Newton's Method

Theorem: Let $g : \mathbf{R}^n \to \mathbf{R}^n$ be differentiable, $x_0 \in \mathbf{R}^n$, and $J_0 \in \mathbf{R}^{n \times n}$. Suppose that there exists $\bar{x},\ x_0 \in \mathbf{R}^n$, and $\epsilon > 0$ with $\|x_0 - \bar{x}\| < \epsilon$ such that

1. $g(\bar{x}) = 0$,
2. $g'(x)^{-1}$ exists for $x \in B(\overline{x}; \epsilon)$ with

$$\sup\{\|g'(x)^{-1}\| : x \in B(\overline{x}; \epsilon)] \leq M_1$$

3. g' is Lipschitz continuous on $c\ell B(\overline{x}; \epsilon)$ with Lipschitz constant L, and
4. $\theta_0 := \frac{LM_1}{2}\|x_0 - \overline{x}\| + M_0 K < 1$ where $K \geq \|(g'(x_0)^{-1} - J_0^{-1})y^0\|$, $y^0 := g(x^0)/\|g(x^0)\|$, and $M_0 = \max\{\|g'(x)\| : x \in B(\overline{x}; \epsilon)\}$.

Convergence of Newton's Method

Further suppose that iteration $x_{k+1} := x_k - J_k^{-1} g(x_k)$ is initiated at x_0 where the J_k's are chosen to satisfy one of the following conditions;

(i) $\|(g'(x_k)^{-1} - J_k^{-1})y_k\| \leq K$,

(ii) $\|(g'(x_k)^{-1} - J_k^{-1})y_k\| \leq \theta_1^k K$ for some $\theta_1 \in (0, 1)$,

(iii) $\|(g'(x_k)^{-1} - J_k^{-1})y_k\| \leq \min\{M_2\|x_k - x_{k-1}\|, K\}$, for some $M_2 > 0$, or

(iv) $\|(g'(x_k)^{-1} - J_k^{-1})y_k\| \leq \min\{M_2\|g(x_k)\|, K\}$, for some $M_3 > 0$,

where for each $k = 1, 2, \ldots,$ $y_k := g(x_k)/\|g(x_k)\|$.

Convergence of Newton's Method

These hypotheses on the accuracy of the approximations J_k yield the following conclusions about the rate of convergence of the iterates x_k.

(a) If (i) holds, then $x_k \to \overline{x}$ linearly.

(b) If (ii) holds, then $x_k \to \overline{x}$ superlinearly.

(c) If (iii) holds, then $x_k \to \overline{x}$ two step quadratically.

(d) If (iv) holds, then $x_i \to \overline{x}$ quadratically.

Convergence of Newton's Method

Proof: We begin by establishing the basic inequalities

$$(\star) \quad \|x_{k+1} - \overline{x}\| \leq \frac{LM_1}{2}\|x_k - \overline{x}\|^2 + \|(g'(x_k)^{-1} - J_k^{-1})g(x_k)\|,$$

and

$$(\star\star) \quad \|x_{k+1} - \overline{x}\| \leq \theta_0 \|x_k - \overline{x}\|$$

and the inclusion $x_{k+1} \in B(\bar{x}; \epsilon)$ by induction on k. For $k = 0$ we have

$$\begin{aligned} x_1 - \overline{x} &= x_0 - \overline{x} - g'(x_0)^{-1}g(x_0) + [g'(x_0)^{-1} - J_0^{-1}]g(x_0) \\ &= g'(x_0)^{-1}[g(\overline{x}) - (g(x_0) + g'(x_0)(\overline{x} - x_0))] \\ &\qquad +[g'(x_0)^{-1} - J_0^{-1}]g(x_0), \end{aligned}$$

since $g'(x_0)^{-1}$ exists by the hypotheses. Consequently, the hypotheses (1)–(4) plus the quadratic bound lemma imply that

$$\begin{aligned} \|x_{k+1} - \overline{x}\| \leq \quad & \|g'(x_0)^{-1}\| \|g(\overline{x}) - (g(x_0) + g'(x_0)(\overline{x} - x_0)) \\ & \qquad + \|(g'(x_0)^{-1} - J_0^{-1})g(x_0)\|\| \end{aligned}$$

Convergence of Newton's Method

$$
\begin{aligned}
\|x_{k+1} - \overline{x}\| &\leq \|g'(x_0)^{-1}\| \|g(\overline{x}) - (g(x_0) + g'(x_0)(\overline{x} - x_0))\| \\
&\quad + \|(g'(x_0)^{-1} - J_0^{-1})g(x_0)\| \\
&\leq \frac{M_1 L}{2}\|x_0 - \overline{x}\|^2 + K\|g(x_0) - g(\overline{x})\| \\
&\leq \frac{M_1 L}{2}\|x_0 - \overline{x}\|^2 + M_0 K\|x_0 - \overline{x}\| \\
&\leq \theta_0\|x_0 - \overline{x}\| < \epsilon,
\end{aligned}
$$

whereby the induction is established for $k = 0$.

Convergence of Newton's Method

Next suppose that the induction hypothesis holds for $k = 0, 1, \ldots, s-1$. We show that it holds at $k = s$. Since $x_s \in B(\overline{x}, \epsilon)$, hypotheses (2)–(4) hold at x_s, one can proceed exactly as in the case $k = 0$ to obtain ($\star$). Now if any one of (i)–(iv) holds, then (i) holds. Thus, by ($\star$), we find that

$$\begin{aligned}
\|x_{s+1} - \overline{x}\| &\leq \tfrac{M_1 L}{2}\|x_s - \overline{x}\|^2 + \|(g'(x_s)^{-1} - J_s^{-1})g(x_s)\| \\
&\leq [\tfrac{M_1 L}{2}\theta_0^s\|x_0 - \overline{x}\| + M_0 K]\|x_s - \overline{x}\| \\
&\leq [\tfrac{M_1 L}{2}\|x_0 - \overline{x}\| + M_0 K]\|x_s - \overline{x}\| \\
&= \theta_0\|x_s - \overline{x}\|.
\end{aligned}$$

Hence $\|x_{s+1} - \overline{x}\| \leq \theta_0\|x_s - \overline{x}\| \leq \theta_0\epsilon < \epsilon$ and so $x_{s+1} \in B(\overline{x}, \epsilon)$.

Convergence of Newton's Method

(a) If $\|(g'(x_k)^{-1} - J_k^{-1})y_k\| \leq K$ holds, then $x_k \to \overline{x}$ linearly.

(b) If $\|(g'(x_k)^{-1} - J_k^{-1})y_k\| \leq \theta_1^k K$ for some $\theta_1 \in (0,1)$ holds, then $x_k \to \overline{x}$ superlinearly.

Proof:

$$\begin{aligned}
\|x_{k+1} - \overline{x}\| &\leq \frac{LM_1}{2}\|x_k - \overline{x}\|^2 + \|(g'(x_k)^{-1} - J_k^{-1})g(x_k)\| \\
&\leq \frac{LM_1}{2}\|x_k - \overline{x}\|^2 + \theta_1^k K\|g(x_k)\| \\
&\leq [\frac{LM_1}{2}\theta_0^k\|x_0 - \overline{x}\| + \theta_1^k M_0 K]\|x_k - \bar{x}\|
\end{aligned}$$

Hence $x_k \to \overline{x}$ superlinearly.

Convergence of Newton's Method

(d) If $\|(g'(x_k)^{-1} - J_k^{-1})y_k\| \leq \min\{M_2\|g(x_k)\|, K\}$, for some $M_3 > 0$ holds, then $x_k \to \overline{x}$ quadratically.

Proof: Again by $(\star)$ and the fact that $x_k \to \bar{x}$, we eventually have

$$\begin{aligned}
\|x_{k+1} - \overline{x}\| &\leq \frac{LM_1}{2}\|x_k - \overline{x}\|^2 + \|(g'(x_k)^{-1} - J_k^{-1})g(x_k)\| \\
&\leq \frac{LM_1}{2}\|x_k - \overline{x}\|^2 + M_2\,\|g(x_k)\|^2 \\
&\leq [\frac{LM_1}{2} + M_2 M_0^2]\|x_k - \overline{x}\|^2\ .
\end{aligned}$$

Newton's Method for Minimization

Theorem: Let $f : \mathbf{R}^n \to \mathbf{R}$ be twice differentiable, $x_0 \in \mathbf{R}^n$, and $H_0 \in \mathbf{R}^{n\times n}$. Suppose that

1. there exists $\overline{x} \in \mathbf{R}^n$ and $\epsilon > \|x_0 - \bar{x}\|$ such that $f(\overline{x}) \leq f(x)$ whenever $\|x - \bar{x}\| \leq \epsilon$,
2. there is a $\delta > 0$ such that $\delta \|z\|_2^2 \leq z^T \nabla^2 f(x) z$ for all $x \in B(\overline{x}, \epsilon)$,
3. $\nabla^2 f$ is Lipschitz continuous on $\mathrm{cl}\, B(\overline{x}; \epsilon)$ with Lipschitz constant L, and
4. $\theta_0 := \frac{L}{2\delta}\|x_0 - \overline{x}\| + M_0 K < 1$ where $M_0 > 0$ satisfies $z^T \nabla^2 f(x) z \leq M_0 \|z\|_2^2$ for all $x \in B(\overline{x}, \epsilon)$ and $K \geq \|(\nabla^2 f(x_0)^{-1} - H_0^{-1}) y^0\|$ with $y^0 = \nabla f(x^0) / \left\|\nabla f(x^0)\right\|$.

Newton's Method for Minimization

Further, suppose that the iteration

$$x_{k+1} := x_k - H_k^{-1} \nabla f(x_k)$$

is initiated at x_0 where the H_k's are chosen to satisfy one of the following conditions:

(i) $\|(\nabla^2 f(x_k)^{-1} - H_k^{-1}) y_k\| \le K$,

(ii) $\|(\nabla^2 f(x_k)^{-1} - H_k^{-1}) y_k\| \le \theta_1^k K$ for some $\theta_1 \in (0,1)$,

(iii) $\|(\nabla^2 f(x_k)^{-1} - H_k^{-1}) y_k\| \le \min\{M_2 \|x_k - x_{k-1}\|, K\}$, for some $M_2 > 0$, or

(iv) $\|(\nabla^2 f(x_k)^{-1} - H_k^{-1}) y_k\| \le \min\{M_2 \|\nabla f(x_k)\|, K\}$, for some $M_3 > 0$,

where for each $k = 1, 2, \ldots$ $y_k := \nabla f(x_k) / \|\nabla f(x_k)\|$.

Newton's Method for Minimization

These hypotheses on the accuracy of the approximations H_k yield the following conclusions about the rate of convergence of the iterates x_k.

(a) If (i) holds, then $x_k \to \overline{x}$ linearly.

(b) If (ii) holds, then $x_k \to \overline{x}$ superlinearly.

(c) If (iii) holds, then $x_\epsilon \to \overline{x}$ two step quadratically.

(d) If (iv) holds, then $x_k \to \overline{k}$ quadradically.

Newton's Method: Example

Let $f(x) = x^2 + e^x$. Then $f'(x) = 2x + e^x$, $f''(x) = 2 + e^x$, $f'''(x) = e^x$. Given $x_0 = 1$ we may take $L = 2$, $M_0 = 4$, and $M_1 = \frac{1}{2}$. Hence the pure Newton strategy should converge to $\overline{x} \approx -0.3517337$ with

$$\|x_k - \overline{x}\| \leq 2(.676)^{2^k}.$$

The actual interates are given in the following table.

x	$f'(x)$
1	4.7182818
0	1
$-1/3$	.0498646
$-.3516893$	.00012
$-.3517337$	.00000000064

Convergence of Steepest Descent with Backtracking

K	X	$f(x)$	$f'(x)$	s
0	1	.37182818	4.7182818	0
1	0	1	1	0
2	−.5	.8565307	−0.3934693	1
3	−.25	.8413008	0.2788008	2
4	−.375	.8279143	−.0627107	3
5	−.34075	.8273473	.0297367	5
6	−.356375	.8272131	−.01254	6
7	−.3485625	.8271976	.0085768	7
8	−.3524688	.8271848	−.001987	8
9	−.3514922	.8271841	.0006528	10
10	−.3517364	.827184	−.0000072	12