Newton’s Method

Qi Gong
Associate Professor
Dept. of Applied Math & Stats.
Baskin School of Engineering
University of California
Santa Cruz, CA

Newton's Method – Motivation

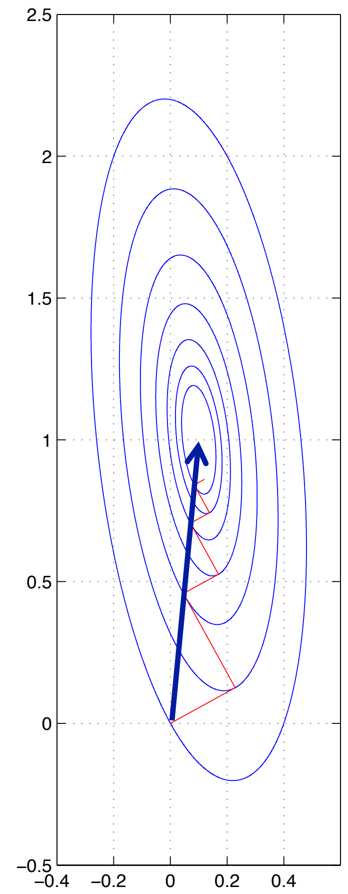
Quadratic Example:

$$\text{Min. } f(x) = \frac{1}{2}x^T Q x - b^T x$$

where $Q = Q^T$ is positive definite.

- At any current point x_k, the direction pointing to $x^* = Q^{-1}b$ is a perfect search direction, since exact line search along this direction will converge in just one step!

$$\begin{aligned} x^* &= x_k + \alpha p = Q^{-1}b \Longrightarrow \\ \alpha p &= Q^{-1}b - x_k = -Q^{-1}(Qx_k - b) \\ &= -\left[\nabla^2 f(x_k)\right]^{-1} \cdot \nabla f(x_k) \end{aligned}$$



Newton's Method – Motivation

Quadratic Example:

$$\text{Min. } f(x) = \frac{1}{2}x^T Q x - b^T x$$
$$\text{where } Q = Q^T \text{ is positive definite.}$$

- At any current point x_k, the direction pointing to $x^* = Q^{-1}b$ is a perfect search direction, since exact line search along this direction will converge in just one step!

$$\begin{aligned} x^* &= x_k + \alpha p = Q^{-1}b \Longrightarrow \\ \alpha p &= Q^{-1}b - x_k = -Q^{-1}(Qx_k - b) \\ &= \boxed{-\left[\nabla^2 f(x_k)\right]^{-1} \cdot \nabla f(x_k)} \end{aligned}$$

- Perfect search direction for quadratic and convex functions
- The step length is naturally $\alpha = 1$

Newton Iteration

Consider general nonlinear cost function f(x).

choose x_0 and tolerance $tol > 0$
$k \leftarrow 0$
evaluate $\nabla f(x_0)$ and $\nabla^2 f(x_0)$

while $\|\nabla f(x_k)\| > tol$
    Solve linear equation:
    $\nabla^2 f(x_k) p_k^N = -\nabla f(x_k)$ for p_k^N
    $x_{k+1} = x_k + p_k^N$
    $k = k + 1$
    evaluate gradient $\nabla f(x_k)$
    evaluate Hessian $\nabla^2 f(x_k)$
end

- Do not compute $[\nabla^2 f(x_k)]^{-1}$! Computing inverse of a matrix is very expensive!
- If the Hessian $\nabla^2 f(x_k)$ is positive definite, the Newton's direction p_k^N is a descent direction.
- If the Hessian $\nabla^2 f(x_k)$ is NOT positive definite, the Newton's direction p_k^N may not be a descent direction.
- Converge to minimum?

Newton Iteration – Example

Consider $\min_{x \in R} f(x) = \ln(1+x^2)$. $x^* = 0$ is the global minimizer.

$$f'(x) = \frac{2x}{1+x^2}, \qquad f''(x) = \frac{2(1-x^2)}{(1+x^2)^2}$$

$$\Longrightarrow \quad p_k^N = -\left[\nabla^2 f(x_k)\right]^{-1} \cdot \nabla f(x_k) = -\frac{x_k(1+x_k^2)}{1-x_k^2}$$

$$\Longrightarrow \quad x_{k+1} = x_k + p_k^N = \frac{2x_k^3}{x_k^2 - 1}$$

$\Longrightarrow$

- If $|x_0| > \frac{1}{\sqrt{3}}$, $\lim_{k\to\infty} |x_k| = \infty$;
- If $|x_0| < \frac{1}{\sqrt{3}}$, $\lim_{k\to\infty} x_k = 0$;
- If $|x_0| = \frac{1}{\sqrt{3}}$, $x_k = \{-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\}$.

- Unlike the steepest descent method, global convergence of Newton's method is NOT guaranteed.

Newton's Method – Convergence Results

Theorem 3.5.

Suppose that f is twice differentiable and that the Hessian $\nabla^2 f(x)$ is Lipschitz continuous (see (A.42)) in a neighborhood of a solution x^* at which the sufficient conditions (Theorem 2.4) are satisfied. Consider the iteration $x_{k+1} = x_k + p_k$, where p_k is given by (3.30). Then

(i) *if the starting point x_0 is sufficiently close to x^*, the sequence of iterates converges to x^*;*

(ii) *the rate of convergence of $\{x_k\}$ is quadratic; and*

(iii) *the sequence of gradient norms $\{\|\nabla f_k\|\}$ converges quadratically to zero.*

Let $\{x_k\}_{k=0}^{\infty}$ be a sequence in R^n that converges to x^*. The rate of convergence of $\{x_k\}$ is quadratic, if there is a constant $M > 0$ such that

$$\|x_{k+1} - x^*\| \leq M\|x_k - x^*\|^2$$

for all k sufficiently large.

Newton's Method – Convergence Analysis

Sketch of a proof:

1. Let $e_k = x_k - x^*$. Then

$$\begin{aligned} e_{k+1} &= x_k + p_k - x^* = x_k - x^* - \left[\nabla^2 f(x_k)\right]^{-1} \cdot \nabla f(x_k) \\ &= \left[\nabla^2 f(x_k)\right]^{-1} \left[\nabla^2 f(x_k)(x_k - x^*) - \nabla f(x_k)\right] \\ &= \left[\nabla^2 f(x_k)\right]^{-1} \left[\nabla^2 f(x_k)(x_k - x^*) - \boxed{(\nabla f(x_k) - \nabla f(x^*))}\right] \end{aligned}$$

Define $g(t) = \nabla f(x_k + t(x^* - x_k))$, $t \in [0,1]$. Then,

$$\begin{aligned} \nabla f(x_k) - \nabla f(x^*) &= g(0) - g(1) = -\int_0^1 g'(t)dt \\ &= \boxed{\int_0^1 \nabla^2 f(x_k + t(x^* - x_k))(x_k - x^*)dt} \end{aligned}$$

By Lipchitz continuity of $\nabla^2 f(x)$,

$$\|e_{k+1}\| \leq \frac{L}{2} \left\| \left[\nabla^2 f(x_k)\right]^{-1} \right\| \cdot \|e_k\|^2,$$

where L is the Lipchitz constant of $\nabla^2 f(x)$.

Newton's Method – Convergence Analysis

2. $\|e_{k+1}\| \;\leq\; \frac{L}{2}\boxed{\left\|\left[\nabla^2 f(x_k)\right]^{-1}\right\|}\cdot\|e_k\|^2$

Lemma. Let A and B be $n \times n$ matrices. Suppose A is invertible and $\|A - B\| \leq \frac{1}{\|A^{-1}\|}$. Then B is invertible. Moreover,

$$\|B^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \cdot \|A - B\|}.$$

If B is sufficiently close to an invertible matrix A, B is invertible with a bounded inverse. In particular, if $\|A - B\| \leq \frac{1}{2\|A^{-1}\|}$, B is invertible, and $\|B^{-1}\| \leq 2\|A^{-1}\|$.

Since $\nabla^2 f(x)$ is continuous, there exist a sufficient small r, such that

$$\forall x \text{ with } \|x - x^*\| \leq r,\; \|\nabla^2 f(x) - \nabla^2 f(x^*)\| \leq \frac{1}{2\|\nabla^2 f(x^*)\|}.$$

By the Lemma, $\nabla^2 f(x)$ is invertible and

$$\left\|\left[\nabla^2 f(x)\right]^{-1}\right\| \leq \boxed{2\left\|\left[\nabla^2 f(x^*)\right]^{-1}\right\|}.$$

Newton's Method – Convergence Analysis

$\Longrightarrow$ There exist a sufficient small r, such that $\forall x_k$ with $\|x_k - x^*\| \leq r$,

$$\boxed{\|e_{k+1}\| \;\leq\; \tilde{L}\|e_k\|^2.} \qquad (\tilde{L} = L\left\|\left[\nabla^2 f(x^*)\right]^{-1}\right\|)$$

3. Define a set $\mathbb{B} = \{x \in R^n \mid \|x - x^*\| < \min(r, \frac{1}{2\tilde{L}})\}$. For all $x_0 \in \mathbb{B}$,

$\Longrightarrow$
$$\begin{aligned} \|x_1 - x^*\| \;&\leq\; \tilde{L}\|x_0 - x^*\|^2 \;=\; \tilde{L}\|x_0 - x^*\|\|x_0 - x^*\| \\ &\leq\; \tilde{L}\frac{1}{2\tilde{L}}\|x_0 - x^*\| \;=\; \frac{1}{2}\|x_0 - x^*\| \end{aligned}$$

$\Longrightarrow$ $x_1 \in \mathbb{B}$. By induction, the sequence $\{x_k\}_{k=0}^{\infty} \in \mathbb{B}$, and

$$\|x_{k+1} - x^*\| \;\leq\; \frac{1}{2}\|x_k - x^*\|$$

$\Longrightarrow$ For all $x_0 \in \mathbb{B}$, $\lim_{k\to\infty} x_k = x^*$, and the rate of convergence is quadratic.

Steepest Descent / Newton's Method

	Steepest Descent	Newton's Method
Robust?	Yes (global convergence)	No (local convergence)
Efficient?	No (linear rate of convergence)	Yes (quadratic rate of convergence)
Computational Cost	Cheap	Expensive

Needs $f(x_k)$ and $\nabla f(x_k)$

Needs gradient $\nabla f(x_k)$, Hessian $\nabla^2 f(x_k)$, and solving $\nabla^2 f(x_k) p_k^N = -\nabla f(x_k)$

Steepest Descent / Newton's Method

	Steepest Descent	Newton's Method
Robust?	Yes (global convergence)	No (local convergence)
Efficient?	No (linear rate of convergence)	Yes (quadratic rate of convergence)
Computational Cost	Cheap	Expensive

Tradeoffs among these goals are essential issues in numerical optimization! ⇒

- Conjugate gradient methods
- Trust region methods
- Quasi-Newton methods
- …