Mathematics of Image and Data Analysis
Math 5467

Newton's Method

Instructor: Jeff Calder
Email: jcalder@umn.edu

http://www-users.math.umn.edu/~jwcalder/5467

Announcements

- HW2 due Feb 25

Last time

- Gradient Descent

Today

- Newton's Method

Gradient Descent

Recall gradient descent

$$x_{k+1} = x_k - \alpha \nabla f(x_k) \tag{1}$$

has the minimizing movements interpretation:

Exercise 1. Fix x_k and define

$$T(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2\alpha} \|x - x_k\|^2. \tag{2}$$

If we define x_{k+1} as the minimizer of T, show that

$$x_{k+1} = x_k - \alpha \nabla f(x_k).$$

△

Second order Taylor expansion

We say $\nabla^2 f$ is L-Lipschitz if

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\|,$$

where $\nabla^2 f$ is the Hessian of f.

Theorem 2 (Second Order Taylor Expansion)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *and assume* $\nabla^2 f$ *is* L*-Lipschitz. Then*

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x) + R \tag{3}$$

where

$$|R| \leq \frac{L}{6}\|x - y\|^3.$$

$$(D^2 f)_{ij} = \frac{\partial^2 f}{\partial x(i)\,\partial x(j)} \qquad n\times n \text{ matrix}$$

Norm: For an $n\times n$ matrix A

$$\|A\| = \max\{\|Ax\| : x\in\mathbb{R}^n,\ \|x\|=1\}.$$

Operator norm

Note $\|Ax\| = \underbrace{\left\|A\frac{x}{\|x\|}\right\|}_{\le \|A\|}\cdot\|x\| \le \|A\|\,\|x\|$

Assume A is symmetric. Then

$$\|Ax\| = \sqrt{\|Ax\|^2} = \sqrt{(Ax)^T(Ax)}$$

$$= \sqrt{x^T A^T A x}$$

$$= \sqrt{x^T A^2 x}$$

$$\max_{\|x\|=1} \|Ax\| = \max_{\|x\|=1} \sqrt{x^T A^2 x}$$

$$= \sqrt{\lambda_{max}(A^2)}$$

$\lambda_{max}(A^2)$ = largest eigenvalue A^2.

$\|A\| = \max_{\|x\|=1} \|Ax\|$ = Largest magnitude eigenvalue of A

(Spectral norm)

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(x)(y-x) + R$$

$$|R| \leq \frac{L}{6}\|x-y\|^3.$$

Define $g(t) = f(x + t(y-x))$

$$g(1) = g(0) + \int_0^1 g'(t)\, dt$$

$$= g(0) + \int_0^1 g'(0) + \int_0^t g''(s)\,ds\, dt$$

$$= g(0) + g'(0) + \int_0^1 \int_0^t g''(s)\,ds\,dt$$

$$= g(0) + g'(0) + \int_0^1 \int_0^t g''(0)\,ds\,dt + R$$

where $R = \int_0^1 \int_0^t g''(s) - g''(0)\, ds\, dt$

$$\int_0^1 \int_0^t ds\,dt = \int_0^1 t\,dt = \frac{1}{2} t^2 \Big|_0^1 = \frac{1}{2}$$

$$g(1) = g(0) + g'(0) + \frac{1}{2} g''(0) + R$$

$g(1) \leftarrow f(y)$, $g(0) \leftarrow f(x)$, $g'(0) \leftarrow \nabla f(x)^T (y-x)$

$g(t) = f(x + t(y-x))$

$\frac{1}{2} g''(0) = \frac{1}{2} (y-x) \nabla^2 f (y-x)$

$$g'(t) = \nabla f(x + t(y-x))^T (y-x)$$

$$g''(t) = (y-x)^T \nabla^2 f(x + t(y-x)) (y-x)$$

↑ <u>exercise</u> Chapter 2

$$|R| \leq \int_0^1 \int_0^t |g''(s) - g''(0)| \, ds \, dt$$

$$= \int_0^1 \int_0^t \left| (y-x)^T \nabla^2 f(x+s(y-x))(y-x) - (y-x)^T \nabla^2 f(x)(y-x) \right| ds dt$$

$$= \int_0^1 \int_0^t \left| (y-x)^T \left(\nabla^2 f(x+s(y-x))(y-x) - \nabla^2 f(x)(y-x) \right) \right| ds dt$$

Cauchy-Schwarz

$$\leq \int_0^1 \int_0^t \|y-x\| \, \|\nabla^2 f(x+s(y-x))(y-x) - \nabla^2 f(x)(y-x)\| \, ds dt$$

$$= \int_0^1 \int_0^t \|y-x\| \, \left\| \left(\nabla^2 f(x+s(y-x)) - \nabla^2 f(x) \right)(y-x) \right\| ds dt$$

Spectral norm

$$\leq \int_0^1 \int_0^t \|y-x\|^2 \, \|\nabla^2 f(x+s(y-x)) - \nabla^2 f(x)\| \, ds\, dt$$

L-Lipshitz

$$\leq \int_0^1 \int_0^t \|y-x\|^2 L \, \|\cancel{x}+s(y-x)-\cancel{x}\| \, ds\, dt$$

$$= L\|x-y\|^3 \int_0^1 \int_0^t s \, ds\, dt = \frac{L}{6}\|x-y\|^3$$

$\blacksquare$

$\|x\| = \|-x\|$

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x)(y - x) + R$$

Newton's Method

Newton's method is based on minimizing a better (second order) approximation of f: We define

$$L(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)$$

and choose x_{k+1} to minimize L. This yields

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k). \tag{4}$$

$$\nabla L(x) = \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) = 0$$

$$\nabla^2 f(x_k)(x - x_k) = -\nabla f(x_k)$$

$$x_{k+1} - x_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

Convergence of Newton's Method

We assume that f is μ-strongly convex for $\mu > 0$. This implies that

$$\|\nabla^2 f(x)^{-1} y\| \leq \frac{1}{\mu}\|y\|, \tag{5}$$

for any $x, y \in \mathbb{R}^n$.

$\|\nabla^2 f(x)^{-1}\| \leq \frac{1}{\mu}$

f μ-strongly convex

(notes) $\iff$ Smallest eigenvalue of $\nabla^2 f$ is at least $\mu > 0$

$\iff$ Largest eigenvalue of $\nabla^2 f^{-1}$

is less than $\frac{1}{\mu}$.

$$\Longleftrightarrow \quad \|\nabla^2 f^{-1}\| \le \frac{1}{\mu}$$

Second order Taylor expansion

(*) $$\nabla f(y) = \nabla f(x) + \nabla^2 f(x)(y-x) + R$$

$$|R| \le \frac{L}{2}\|x-y\|^2$$

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \le L\|x-y\|$$

Compare to

$$f(y) = f(x) + \nabla f(x)^T(y-x) + R$$

Convergence of Newton's Method

Theorem 3. *Let $f : \mathbb{R}^n \to \mathbb{R}$. Assume that f is μ-strongly convex, $\nabla^2 f$ is L-Lipschitz and*

$$e(0) = \beta := \frac{L}{2\mu^2} \|\nabla f(x_0)\| < 1.$$

Then Newton's method converges as $k \to \infty$ to the unique minimizer of f, and furthermore for any $k \geq 0$ we have

$$\|\nabla f(x_k)\| \leq \frac{2\mu^2}{L} \beta^{2^k}. \tag{6}$$

Proof: Define $e(k) = \frac{L}{2\mu^2} \|\nabla f(x_k)\|$

Claim: $e(k+1) \leq e(k)^2$ (Quadratic conv.)

Newton

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k).$$

$$\nabla^2 f(x_k)^{-1} \nabla f(x_k) = x_k - x_{k+1}$$

$$\nabla f(x_k) = \nabla^2 f(x_k)(x_k - x_{k+1})$$

$$\boxed{\nabla f(x_k) + \nabla^2 f(x_k)(x_{k+1} - x_k) = 0}$$

$$e(k+1) = \frac{L}{2\mu^2} \| \nabla f(x_{k+1}) \|$$

$$= \frac{L}{2\mu^2} \| \underbrace{\nabla f(x_{k+1}) - \nabla f(x_k) - \nabla^2 f(x_k)(x_{k+1} - x_k)}_{R} \|$$

$$\nabla f(y) = \nabla f(x) + \nabla^2 f(x)(y-x) + R$$

$y = x_{k+1}, \; x = x_k$

$$\|R\| \le \frac{L}{2}\|x-y\|^2$$

<u>Taylor Expansion</u>

$$e(k+1) \le \frac{L}{2\mu^2} \cdot \frac{L}{2}\|x_{k+1} - x_k\|^2$$

Newton's method.

$$= \frac{L^2}{4\mu^2}\|\nabla^2 f(x_k)^{-1}\nabla f(x_k)\|^2$$

μ-strong convexity

$\|\nabla^2 f^{-1}\| \le \frac{1}{\mu}$

$$\le \frac{L^2}{4\mu^2} \cdot \left(\frac{1}{\mu}\|\nabla f(x_k)\|\right)^2$$

$$= \frac{L^2}{4\mu^4} \|\nabla f(x_k)\|^2 = e(k)^2 \quad \blacksquare$$

Newton Notebook (.ipynb)