

# CMO: Newton's method

Eklavya Sharma

By second-order Taylor series, we get

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

In Newton's method, we choose an update rule which minimizes the resulting quadratic function, assuming that  $H_f(x)$  is positive definite.

$$x^{i+1} = x^i - H_f(x^i)^{-1} \nabla f(x^i)$$

## 1 Matrix norm

**Definition 1** (Spectral Norm). *Let  $A$  be a  $d$  by  $d$  matrix. Then*

$$\|A\| = \max_{i=1}^d |\lambda_i(A)|$$

**Theorem 1** (Homework). *The spectral norm is a norm (i.e.  $\|A\| = 0 \iff A = 0$  and  $\|A\| \geq 0$ ).*

**Theorem 2** (Homework).  *$D(A, B) = \|A - B\| \implies D$  is a distance metric.*

*Proof hint for triangle inequality.* Get a [bound on eigenvalues of sum of matrices](#) using the facts that the sum of positive semidefinite matrices is also positive semidefinite and that if  $(\lambda, v)$  is an eigenpair of a matrix  $A$  then  $(\lambda - k, v)$  is an eigenpair of the matrix  $A - kI$ .  $\square$

## 2 Region of positive-definite hessian

Intuitively, Newton's method would work when the hessian is positive definite. Unfortunately, that need not be true for most real problems. However, at a local minimum, the hessian is guaranteed to be positive semidefinite. We hope that, due to  $f$  being in  $C^2$ , hessian would be positive semidefinite or definite *near* the local minimum too.

Therefore, if we somehow reach close to a local minimum, we can start using Newton's method. We'll now try to quantify how close do we need to get.

**Definition 2.** *Let  $\lambda(A)$  denote the set of eigenvalues of matrix  $A$ .*

**Definition 3.** For matrices  $A$  and  $B$ ,  $A \leq B$  iff  $B - A$  is positive semidefinite. For matrices  $A$  and  $B$ ,  $A < B$  iff  $B - A$  is positive definite. Similarly define  $\geq$  and  $>$ .

**Theorem 3** (Transitivity of  $\leq$  (Homework)).  $A \leq B \wedge B \leq C \implies A \leq C$

**Definition 4.**

$$f \in C_M^2 \iff (\forall x, y \in \mathbb{R}^d, \|H_f(y) - H_f(x)\| \leq M\|y - x\|)$$

**Theorem 4** (Homework).

$$f \in C_M^2 \implies (\forall x, y \in \mathbb{R}^d, H_f(x) - MrI \leq H_f(y) \leq H_f(x) + MrI)$$

where  $r = \|y - x\|$ .

**Theorem 5.** Let  $H_f(x^*) \geq aI$ , where  $a > 0$ . Then  $r = \|x - x^*\| < \frac{a}{M} \implies H_f(x)$  is positive definite.

*Proof.*

$$r < \frac{a}{M} \implies 0 < (a - Mr)I \leq H_f(x^*) - MrI$$

$$f \in C_M^2 \implies H_f(x) \geq H_f(x^*) - MrI > 0$$

□

We now have a region where  $H_f(x)$  is known to be positive definite. However, this is still not suitable for Newton's method, since the point in the next iteration may fall outside this region. We therefore impose another condition, that distance from  $x^*$  should reduce.

### 3 Newton's region and convergence

**Lemma 6** (Proof omitted (beyond the scope of course?)).  $\forall x, y \in \mathbb{R}^d$ ,

$$\nabla_f(y) - \nabla_f(x) = \int_0^1 H_f(x + \alpha(y - x))(y - x)d\alpha$$

**Lemma 7.** Let  $A$  be a symmetric matrix and  $u$  be a vector. Then  $\|Au\| \leq \|A\|\|u\|$ .

*Proof.*

$$\frac{\|Au\|^2}{\|u\|^2} = \frac{u^T A^2 u}{\|u\|^2} \in [0, \|A\|^2] \implies \frac{\|Au\|}{\|u\|} \leq \|A\|$$

□

**Theorem 8.** Let  $H_f(x^*) \geq a$ , where  $a > 0$ . Let  $r_k = \|x^k - x^*\|$ . Then  $r_k < \frac{2a}{3M} \implies r_{k+1} < r_k$ .

*Proof.* By theorem 6, we get

$$\nabla_f(x^k) = \int_0^1 \mathbf{H}_f(x^* + \alpha(x^k - x^*))(x^k - x^*) d\alpha$$

$$\begin{aligned} x^{k+1} - x^* &= (x^k - x^*) - \mathbf{H}_f^{-1}(x^k) \nabla_f(x^k) \\ &= \mathbf{H}_f^{-1}(x^k) (\mathbf{H}_f(x^k)(x^k - x^*) - \nabla_f(x^k)) \\ &= \mathbf{H}_f^{-1}(x^k) \int_0^1 (\mathbf{H}_f(x^k) - \mathbf{H}_f(x^k + \alpha(x^k - x^*))) (x^k - x^*) d\alpha \end{aligned}$$

$$\begin{aligned} r_{k+1} &= \left\| \mathbf{H}_f^{-1}(x^k) \int_0^1 (\mathbf{H}_f(x^k) - \mathbf{H}_f(x^k + \alpha(x^k - x^*))) (x^k - x^*) d\alpha \right\| \\ &\leq \|\mathbf{H}_f^{-1}(x^k)\| \left\| \int_0^1 (\mathbf{H}_f(x^k) - \mathbf{H}_f(x^k + \alpha(x^k - x^*))) (x^k - x^*) d\alpha \right\| && \text{(by lemma 7)} \\ &\leq \|\mathbf{H}_f^{-1}(x^k)\| \int_0^1 \|(\mathbf{H}_f(x^k) - \mathbf{H}_f(x^k + \alpha(x^k - x^*))) (x^k - x^*)\| d\alpha \\ &&& \text{(by triangle inequality)} \\ &\leq \|\mathbf{H}_f^{-1}(x^k)\| \int_0^1 \|\mathbf{H}_f(x^k) - \mathbf{H}_f(x^k + \alpha(x^k - x^*))\| \|x^k - x^*\| d\alpha && \text{(by lemma 7)} \\ &\leq \|\mathbf{H}_f^{-1}(x^k)\| \int_0^1 (M \|(1 - \alpha)(x^k - x^*)\|) \|x^k - x^*\| d\alpha && (f \in C_M^2) \\ &= \|\mathbf{H}_f^{-1}(x^k)\| \left( \frac{Mr_k^2}{2} \right) \end{aligned}$$

$$\begin{aligned} \mathbf{H}_f(x^k) \geq (a - Mr_k)I &\implies \mathbf{H}_f^{-1}(x^k) \leq \frac{1}{a - Mr_k}I \implies \|\mathbf{H}_f^{-1}(x^k)\| \leq \frac{1}{a - Mr_k} \\ \implies r_{k+1} &\leq \frac{Mr_k^2}{2(a - Mr_k)} \\ \frac{Mr_k}{2(a - Mr_k)} < 1 &\iff r_k < \frac{2a}{3M} \end{aligned}$$

Therefore,

$$r_k \leq \frac{2a}{3M} \implies \frac{r_{k+1}}{r_k} \leq \frac{Mr_k}{2(a - Mr_k)} < 1 \implies r_{k+1} < r_k$$

□

The  $\frac{2a}{3M}$ -neighborhood of  $x^*$  is called the Newton region. In this region, Newton's method will always be applicable. Furthermore,

$$r_k < \frac{2a}{3M} \implies \frac{a}{M} - r_k > \frac{a}{3M} \implies r_{k+1} = \frac{r_k^2}{2(\frac{a}{M} - r_k)} < \frac{3M}{2a} r_k^2$$

which shows that Newton's method gives us quadratic convergence.

## 4 Quadratic function

Let  $f(x) = \frac{1}{2}x^T Qx - b^T x$ , where  $Q$  is symmetric and positive definite.  $\nabla_f(x) = Qx - b = Q(x - x^*)$ .

$$x_1 = x_0 - H_f(x_0)^{-1} \nabla_f(x_0) = x_0 - Q^{-1}Q(x - x^*) = x^*$$

Therefore, Newton's method converges to the minimum in a single iteration.