Lecture 11 Newton's method

Alg: $x^+ = x - t(\nabla^2 f(x))^{-1}\nabla f(x)$

f is L-smooth, m-strongly convex and M-Hessian Lipschitz.

α, β are parameters of line search $0<\alpha\le\frac{1}{2}$, $0<\beta<1$

Two phases

① $f(x^{(k)}) - f^* \le f(x^{(0)}) - f(x^*) - \gamma k$. $k\le k_0$

② $f(x^{(k)}) - f^* \le \frac{2m^3}{M^2}\left(\frac{1}{2}\right)^{2^{k-k_0+1}}$ for $k>k_0$

k_0 is the # of steps until $\|\nabla f(x^{k_0})\|_2 \le \eta$

$\gamma = \alpha\beta^2\eta^2 m/L^2$, $\eta = \min\{1, 3(1-2\alpha)\}\, m^2/M$

Phase 1: ensured by backtracking line search.

$f(x^{(k+1)}) - f(x^{(k)}) \le -\gamma$ (exercise!)

Phase 2: $\|\nabla f(x^{(k_0)})\|_2 \le \eta \le m^2/M$, backtracking will always choose $t=1$

Fact: m-strong convex $\Rightarrow$ Polyak-Lojasiewicz Condition

$\forall x \quad f(x) - f^* \le \frac{1}{2m}\|\nabla f(x)\|_2^2$

$$\boxed{\text{Lemma: } \frac{M}{2m^2}\|\nabla f(x^+)\|_2 \le \left(\frac{M}{2m^2}\|\nabla f(x)\|_2\right)^2}$$

Note $v = -(\nabla^2 f(x))^{-1}\nabla f(x)$

$-\nabla f(x) = \nabla^2 f(x)\cdot v$

Proof: $\|\nabla f(x^+)\|_2 = \|\nabla f(x+v)\|_2$

$= \|\nabla f(x^+) - \nabla f(x) - \nabla^2 f(x) v\|_2$

$\overset{\text{Taylor's thm}}{=} \left\|\int_0^1 \nabla^2 f(x+tv)\cdot v\,dt - \nabla^2 f(x)v\right\|_2$

$= \left\|\int_0^1 (\nabla^2 f(x+tv) - \nabla^2 f(x))v\,dt\right\|_2$

$$\left\| \int_0^1 \left(\nabla^2 f(x+tv) - \nabla^2 f(x)\right) v \, dt \right\|_2$$

$$\leq \int_0^1 \underbrace{\left\| \left(\nabla^2 f(x+tv) - \nabla^2 f(x)\right) v \right\|_2}_{(*)} dt$$

$$\leq M \|v\|_2^2 \int_0^1 t \, dt = M\|v\|_2^2 \left[\frac{t^2}{2}\right]_0^1 = \frac{M}{2}\|v\|^2$$

$$= \frac{M}{2} \|\nabla^2 f(x)^{-1} \nabla f(x)\|_2^2 \leq \frac{M}{2} \|\nabla^2 f(x)^{-1}\|_{op}^2 \|\nabla f(x)\|_2^2$$

$$\leq \frac{M}{2m^2} \|\nabla f(x)\|_2^2$$

$(*) \leq \|\nabla^2 f(x+tv) - \nabla^2 f(x)\|_{op} \|v\|_2$

$\leq M\|tv\|_2 \cdot \|v\|_2 = Mt\|v\|_2^2$

recall $v = -(\nabla^2 f(x))^{-1} \nabla f(x)$

$(\nabla^2 f(x))^{-1} \preceq \frac{1}{m} I$

$$\underbrace{\frac{M}{2m^2} \|\nabla f(x^+)\|_2}_{\alpha_k} \leq \underbrace{\left(\frac{M}{2m^2}\right)^2 \|\nabla f(x)\|_2^2}_{(\alpha_{k-1})^2} \qquad \square$$

we want $f(x^{(k)}) - f^* \leq \frac{2m^3}{M^2} \left(\frac{1}{2}\right)^{2^{k-k_0}}$.

$\alpha_k \leq (\alpha_{k-1})^2 \leq (\alpha_{k-2})^4 \leq (\alpha_{k-3})^8 \leq \cdots \leq \alpha_{k_0}^{2^{k-k_0}} = \left(\frac{1}{2}\right)^{2^{k-k_0}}$, finally substitute in to the PL condition

$\alpha^{k_0} := \frac{M}{2m^2} \|\nabla f(x^{(k_0)})\|_2 \leq \eta \leq \frac{m^2}{M} \cdot \frac{M}{2m^2} = \frac{1}{2}$