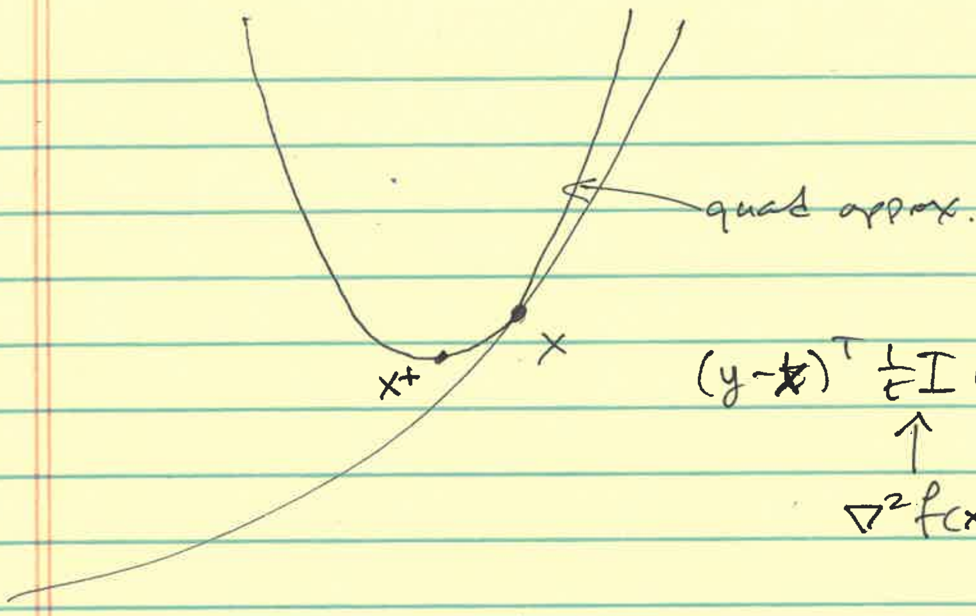


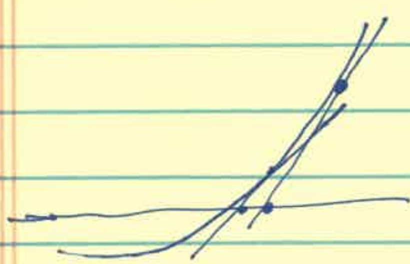
$$(y-x)^T \underbrace{\tfrac{1}{t} I}_{\nabla^2 f(x)} (y-x) = \tfrac{1}{t}\|y-x\|_2^2$$

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

$$0 = \nabla Q(y) = \nabla f(x) + \nabla^2 f(x)\cdot(y-x).$$

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

$F(x) = 0.$ $\quad$ F nonlinear function.
(for us $F(x) = \nabla f(x)$.)



$$F(y) \approx \underbrace{F(x) + F'(x)(y-x)}_{\text{solve for its root:}}$$

$$y = x - (F'(x))^{-1} F(x)$$

Applied to $F(x) = \nabla f(x)$:

$$y = x - (\nabla^2 f(x))^{-1}\nabla f(x).$$

---

f. $\quad$ nonsingular square matrix A. $\quad$ $g(y) = f(Ay)$

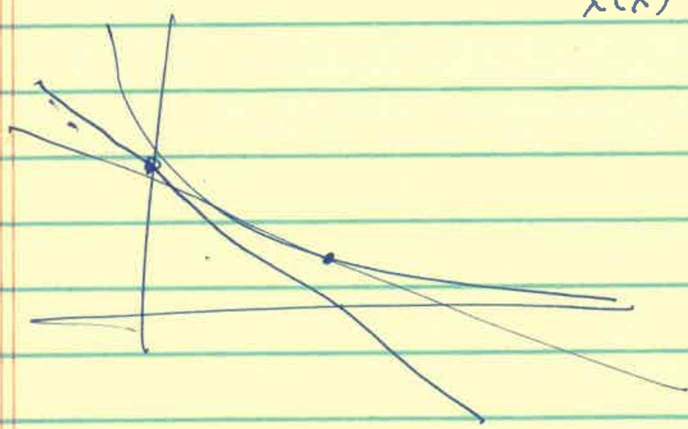
Newton on f $\quad = \quad$ Newton on g

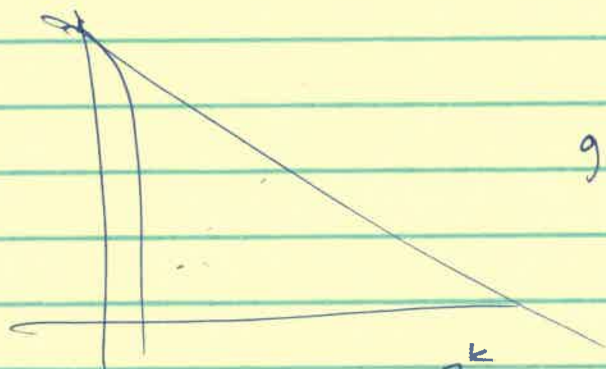
$$(AB)^{-1} = B^{-1}A^{-1}$$

$$x^+ = x - tv. \qquad v = (\nabla^2 f(x))^{-1} \nabla f(x)$$

$$\|x\|_A^2 = x^T A x.$$

$$\|v\|^2_{\nabla^2 f(x)} = \nabla f(x)^T (\nabla^2 f(x))^{-1} \cancel{\nabla^2 f(x) (\nabla^2 f(x))^{-1}} \nabla f(x)$$
$$= \lambda(x)^2.$$





grad descent converging at c^k. for some $c<1$.

$$\underbrace{\left(\tfrac{1}{2}\right)^{2^k}.}$$

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_{op.} \leq M \|x-y\|_2$$

Assume that we're in pure phase and backtracking gives $t=1$.

Fact 1. f is m-strongly convex

$$\Rightarrow \quad f(x) - f^\star \le \frac{1}{2m} \|\nabla f(x)\|_2^2 .$$

Fact 2. $$\frac{M}{2m^2} \|\nabla f(x^+)\|_2 \le \left(\frac{M}{2m^2} \|\nabla f(x)\|_2\right)^2$$

Fact 3. $$f(x^{(k)}) - f^\star \le \frac{2m^3}{M^2} \underline{\left(\frac{1}{2}\right)^{2^{k-k_0}}}$$

Proof Fact 1.

$$f(y) \ge f(x) + \nabla f(x)^T (y-x) + \frac{m}{2} \|y-x\|_2^2 .$$

minimize both sides over y.

$$f^\star \ge \min_y \left(f(x) + \nabla f(x)^T (y-x) + \frac{m}{2} \|y-x\|_2^2 \right)$$

$$0 = \nabla f(x) + m(y-x)$$

$$y = -\frac{1}{m} \nabla f(x) + x.$$

plug in:

$$= f(x) - \frac{1}{m} \|\nabla f(x)\|_2^2 + \frac{1}{2m} \|\nabla f(x)\|_2^2$$

$$= f(x) - \frac{1}{2m} \|\nabla f(x)\|_2^2 .$$

rearrange:

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

Proof of Fact 2.

$$\|\nabla f(x^+)\|_2 = \|\nabla f(x+v)\|_2 \quad \text{where } v = -(\nabla^2 f(x))^{-1} \nabla f(x)$$

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

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

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

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

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

$$\leq M t \|v\|_2^2$$

$$\leq M \|v\|_2^2 \cdot \int_0^1 t \, dt$$

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

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

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

multiply both sides by $M/2m^2$:

$$\frac{M}{2m^2} \| \nabla f(x^+) \|_2 \leq \left(\frac{M}{2m^2} \| \nabla f(x) \|_2 \right)^2 . \checkmark$$

Pf: Fact 3.

we've established $\underbrace{\frac{M}{2m^2} \| \nabla f(x^{(k)}) \|_2}_{a_k} \leq \left(\frac{M}{2m^2} \| \nabla f(x^{(k-1)}) \|_2 \right)^2$

$\leq a_{k-1}^2 .$

$$a_k \leq a_{k-2}^4$$
$$\vdots$$
$$\leq a_{k_0}^{2^{k-k_0}} .$$

$$\frac{M}{2m^2} \| \nabla f(x^{(k)}) \|_2 \leq \left(\frac{M}{2m^2} \| \nabla f(x^{(k_0)}) \|_2 \right)^{2^{k-k_0}}$$

~~But~~ But at k_0 we know $\|\nabla f(x^{(k_0)})\|_2 < \eta \le \frac{m^2}{M}$

so $\frac{M}{2m^2}\|\nabla f(x^{(k)})\|_2 \le \left(\frac{1}{2}\right)^{2^{k-k_0}}$.

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

$$\le \frac{1}{2m}\left(\frac{2m^2}{M}\right)^2\left(\frac{1}{2}\right)^{2^{k-k_0+1}}$$

$$= \frac{2m^3}{M^2}\left(\frac{1}{2}\right)^{2^{k-k_0+1}}. \checkmark$$

set $c\cdot\left(\frac{1}{2}\right)^{2^{k-k_0+1}} = \varepsilon$.

$k - k_0 = \log\log(\varepsilon_0/\varepsilon)$.

$\log\log(1/\varepsilon) = 6$.