$Q^{1/2}$ $\qquad$ $Q^{1/2}$ is sym. and $Q^{1/2}Q^{1/2} = Q$

$$\left\| \begin{pmatrix} \frac{Q^{1/2}x}{\sqrt{2}} \\ \frac{1+t}{2} \end{pmatrix} \right\|_2 \le \frac{1-t}{2}$$

$$\frac{x^T Q^{1/2} Q^{1/2} x}{2} + \cancel{\frac{1}{4}} + \cancel{\frac{t^2}{4}} + \frac{t}{2} \le \cancel{\frac{1}{4}} + \cancel{\frac{t^2}{4}} - \frac{t}{2}$$

$$\frac{1}{2} x^T Q x \le -t$$

(Replace t with $-t$)

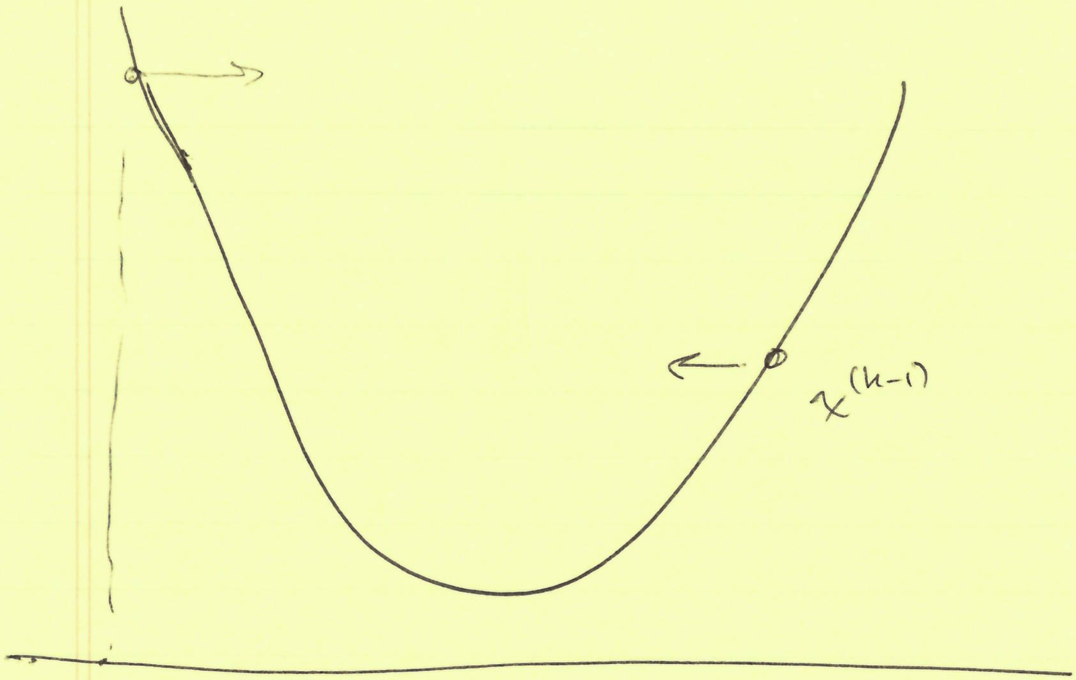
x^+ $\quad$ x

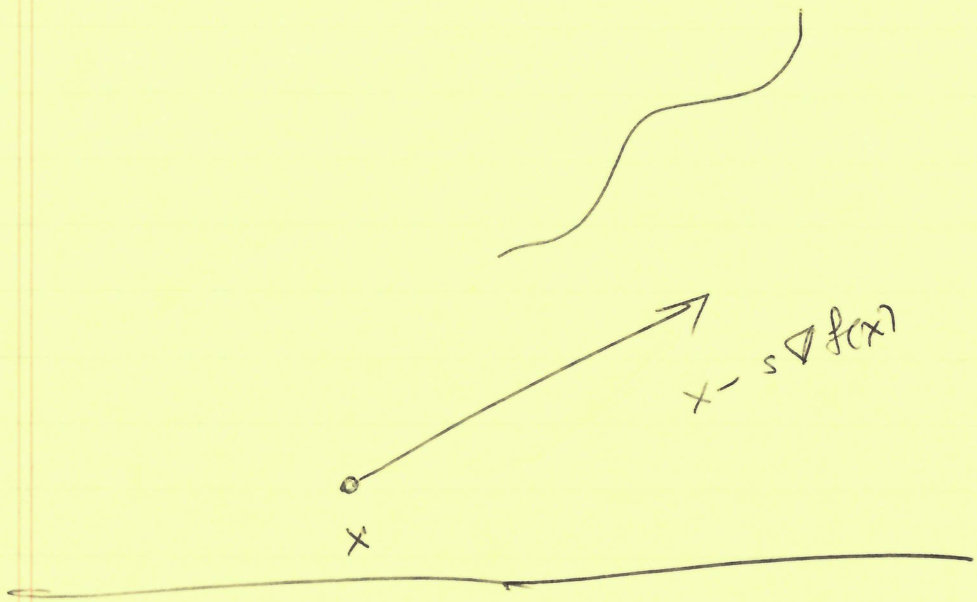
$$f(x) + \nabla f(x)^T (y-x) + \underbrace{\frac{1}{2t}\|y-x\|_2^2}$$

2nd order Taylor replaces this with $\frac{1}{2}(y-x)^T \nabla^2 f(x)(y-x)$

chosen to min. this quadratic approximation (over y)

$$x^+ = x - t\nabla f(x)$$





$$\frac{\|x^{(0)} - x^\star\|_2^2}{2tk} = \varepsilon$$

$$k = \cancel{\frac{2t}{\|x^{(0)} - x^\star\|_2^2}} \cdot \frac{\|x^{(0)} - x^\star\|_2^2}{2t} \cdot \frac{1}{\varepsilon}$$

∇f Lipschitz, constant L

$$\Longrightarrow \quad f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{L}{2} \|y-x\|_2^2 \quad \text{all } x, y$$

Suppose we are at x in gradient desc iterations, go to $x^+ = x - t\nabla f(x)$

use above ineq. with $y = x^+$:

$$\begin{aligned} f(x^+) &\le f(x) + \nabla f(x)^T \cdot (-t\nabla f(x)) + \frac{L}{2} \| -t \cdot \nabla f(x) \|_2^2 \\ &= f(x) - t \|\nabla f(x)\|_2^2 + \frac{Lt^2}{2} \cdot \|\nabla f(x)\|_2^2 \\ &= f(x) - \left(1 - \frac{Lt}{2}\right) \cdot t \cdot \|\nabla f(x)\|_2^2 \end{aligned}$$

if $t \le \frac{1}{L}$, note we get

$$f(x^+) \le f(x) - \frac{t}{2} \cdot \|\nabla f(x)\|_2^2$$

[important aside: get descent! $f(x^+) < f(x)$]

use convexity of f:

$$f(x^*) \geq f(x) + \nabla f(x)^T (x^* - x)$$

OR $\quad f(x) \le f(x^*) + \nabla f(x)^T (x - x^*)$

hence

$$f(x^+) \le f(x^*) + \nabla f(x)^T (x - x^*) - \frac{t}{2} \|\nabla f(x)\|_2^2$$

$$\begin{aligned} f(x^+) - f(x^*) &\le \nabla f(x)^T (x - x^*) - \frac{t}{2} \|\nabla f(x)\|_2^2 \\ &= \frac{1}{2t}\left(\|x - x^*\|_2^2 - \|x^+ - x^*\|_2^2 \right) \end{aligned}$$

check : just expand.

$$\frac{1}{2t}\left(\|x - x^\star\|_2^2 - \|x - t\nabla f(x) - x^\star\|_2^2 \right)$$

$$= \frac{1}{2t}\Big(\cancel{\|x - x^\star\|_2^2} - \cancel{\|x - x^\star\|_2^2} - t^2 \|\nabla f(x)\|_2^2$$

$$2 \cdot t \nabla f(x)^T (x - x^\star) \Big)$$

$$= - \frac{t}{2} \|\nabla f(x)\|_2^2 + \nabla f(x)^T (x - x^\star) \checkmark$$

Back up, state in terms of iteration i

$$f(x^{(i)}) - f(x^\star) \le \frac{1}{2t}\left(\|x^{(i-1)} - x^\star\|_2^2 - \|x^{(i)} - x^\star\|_2^2 \right)$$

$$\sum_{i=1}^{k} \left[f(x^{(i)}) - f(x^\star) \right] \le \frac{1}{2t}\left(\|x^{(0)} - x^\star\|_2^2 - \|x^{(k)} - x^\star\|_2^2 \right)$$

$$\le \frac{1}{2t} \|x^{(0)} - x^\star\|_2^2$$

Because we've proved that $f(x^{(1)}) \geq f(x^{(2)}) \geq f(x^{(3)}) \geq \ldots$

$$f(x^{(k)}) - f(x^\star) \le \frac{1}{k} \sum_{i=1}^{k} \left[f(x^{(i)}) - f(x^\star) \right]$$

$$\le \underbrace{\|x^{(0)} - x^\star\|_2^2}_{2t} \cdot \frac{1}{k}$$
