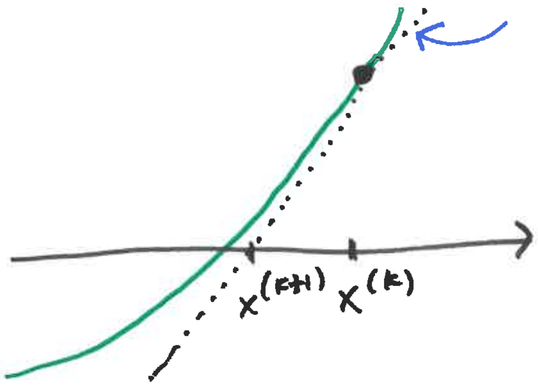
Newton's method	Quasi-Newton		Gradient Descent
$O(n^3)$	$O(n^2)$	$\uparrow$	$O(n)$
		Limited memory QN	

Secant equation

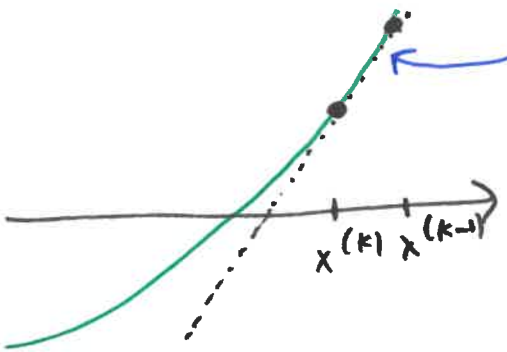
$$\nabla f(x^+) = \nabla f(x) + B(x^+ - x)$$

Newton' method (1D, for root finding)



$$\tilde{f}(x) = f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)})$$

Secant method



$$\tilde{f}(x) = f(x^{(k)}) + \underbrace{\frac{f(x^{(k)}) - f(x^{(k-1)})}{x^{(k)} - x^{(k-1)}}}_{B} (x - x^{(k)})$$

$$B^+ s = y$$

$$s = x^+ - x, \qquad y = \nabla f(x^+) - \nabla f(x)$$

SR1

$$B^+ = B + a u u^T$$

$$Bs + \underline{a}\, u\, \underbrace{u^T s} = y$$

$$\Rightarrow (a u^T s) u = y - Bs \quad \Rightarrow \quad u = y - Bs$$

$$a u^T s = 1 \Rightarrow a = \frac{1}{u^T s} = \frac{1}{(y - Bs)^T s}$$

$$B^+ = B + a u u^T \quad \Rightarrow \quad \tilde{u} = a u, \quad \tilde{v} = u$$

$$(B + \hat{u}\hat{v}^T)^{-1} = B^{-1} - \frac{B^{-1}\hat{u}\hat{v}^T B^{-1}}{1 + \hat{v}^T B^{-1}\hat{u}}$$

$$\underbrace{C^+ = C - \frac{C\hat{u}\hat{v}\,C}{1 + \hat{v}^T C\hat{u}}}_{O(n^2)}$$

$$C^+ \nabla f(x^{(k)}) \leftarrow O(n^2)$$

$$B^+ = B + a u u^T + b v v^T$$

$$\cancel{y - Bs}$$

$$\cancel{y - Bs = a u u^T s + b v v^T s}$$

$$\underline{y} - \underline{Bs} = (a u^T s)\underline{u} + (b v^T s)\underline{v}$$

$$a u^T s = 1 \Rightarrow a = \frac{1}{u^T s} = \frac{1}{y^T s}$$

$$b v^T s = -1 \Rightarrow b = \frac{-1}{v^T s} = \frac{-1}{s^T B s}$$

$$C^+ = \underbrace{\left(I - \frac{s y^T}{y^T s}\right) C \left(I - \frac{y s^T}{y^T s}\right)} + \frac{s s^T}{\underbrace{\boxed{y^T s}}_{>0}}$$

$$\left(I - \frac{sy^T}{y^Ts}\right) c \left(I - \frac{ys^T}{y^Ts}\right) g + \underbrace{\frac{ss^T}{y^Ts} g}_{\alpha}$$

$$= \left(I - \frac{sy^T}{y^Ts}\right) c \left(g - \alpha y\right) + \alpha s$$

Linear $\|x^{(k)} - x^*\|_2 \leq c\|x^{(k-1)} - x^*\|_2$

Quadratic $\|x^{(k)} - x^*\|_2 \leq c\|x^{(k-1)} - x^*\|_2^2$

Superlinear $\|x^{(k)} - x^*\|_2 \leq c_k\|x^{(k-1)} - x^*\|_2$

* $c_k \to 0$ as $k \to \infty$

<u>Explicit form</u>

$O(n^2)$ memory

~~$O(n^2)$ time per update~~

$\Rightarrow O(kn^2)$ time total

<u>Implicit form</u>

$O(kn)$ memory

$O(k^2 n)$ time

$\Downarrow$

Limited memory

<u>LBFGS</u>

$O(mn)$

$O(kmn)$ time

* same as gradient descent if m is