	First Order		Second Order
Unconstrained smooth	Gradient descent	Quasi-Newton	Newton's method
Nonsmooth "simple" i.e. w/ prox	Proximal gradient	Today	
Nonsmooth ~~complex~~ "general"	Subgradient descent *		Interior point

* not fast convergence

$$x^{(k+1)} = \mathrm{prox}_{t_k}\left(x^{(k)} - t_k \nabla g(x^{(k)})\right)$$

$$\mathrm{prox}_t(x) = \underset{z}{\mathrm{argmin}}\ \frac{1}{2t}\|x - z\|_2^2 + h(z)$$

Proximal gradient

$$x^+ = \underset{z}{\mathrm{argmin}}\ \underbrace{\nabla g(x)^T(z - x) + \frac{1}{2t}\|x - z\|^2}_{H = \frac{1}{t} I} + h(z)$$

$$v^+ = \underset{v}{\mathrm{argmin}}\ \nabla g(x)^T v + \frac{1}{2t}\|v\|^2 + h(x + v)$$

Proximal gradient

$$\Downarrow$$

Proximal Newton

$$v^{(k)} = \underset{v}{\mathrm{argmin}}\ \nabla g(x^{(k-1)})^T v + v^T H^{(k-1)} v + h(x^{(k-1)} + v)$$

$$x^{(k)} = x^{(k-1)} + t_k v^{(k)}$$

$$H^{(k-1)} = \nabla^2 g(x^{(k-1)})$$

$$\text{prox}_t(x) = \underset{z}{\text{argmin}} \; \frac{1}{2t}\|x-z\|_2^2 + h(z)$$

$$\text{prox}_H(x) = \underset{z}{\text{argmin}} \; \frac{1}{2}\|x-z\|_H^2 + h(z)$$

where $\quad \|x\|_H^2 = x^T H x$

prox Newton
$$\begin{cases} z^{(k)} = \text{prox}_{H^{(k-1)}}\left(x^{(k-1)} - {H^{(k-1)}}^{-1} \nabla g(x^{(k-1)})\right) \\ x^{(k)} = x^{(k-1)} + t_k\left(z^{(k)} - x^{(k-1)}\right) \end{cases}$$

(Local) Quadratic convergence

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

1) $t < \min\left\{1, \frac{2m}{M}(1-\alpha)\right\}$

2) For large k (or for $x^{(k)}$ close enough to x^*) $\Rightarrow t = 1$

3) $$\|x^+ - x^*\|_2 \le \frac{1}{\sqrt{m}} \|x^+ - x^*\|_H$$

$$\le \frac{1}{\sqrt{m}} \|\mathrm{prox}_H(x - H^{-1}\nabla g(x)) - \mathrm{prox}_H(x^* - H^{-1}\nabla g(x^*))\|_H$$

$$\le \frac{1}{\sqrt{m}} \|x - H^{-1}\nabla g(x) - x^* + H^{-1}\nabla g(x^*)\|_H$$

$$\le \frac{1}{m} \|H(x-x^*) - \nabla g(x) + \nabla g(x^*)\|_2$$

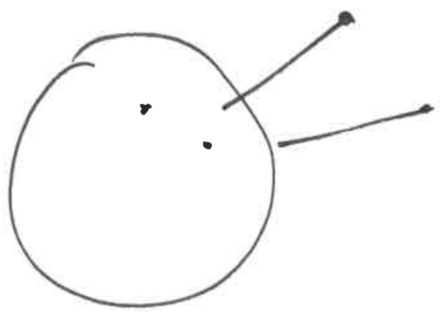
$$\le \frac{M}{2m} \|x - x^*\|_2^2$$

$\cancel{\nabla g(x^*) =}$

$$\|\nabla g(x) - \nabla g(x^*) + \underline{H}(x-x^*)\|_2 + O(\|x-x^*\|^2) \le M\|x-x\|_2^2$$

$$\begin{aligned}\frac{1}{\sqrt{m}}\|y\|_H &= \frac{1}{\sqrt{m}}\|H^{1/2}y\|_2 \\ &\le \frac{1}{m}\|H^{1/2}y\|_H \\ &= \frac{1}{m}\|Hy\|_2\end{aligned}$$

x, y $\quad \|\mathrm{prox}_H(x) - \mathrm{prox}_H(y)\|_H \le \|x-y\|_H$



$$\boxed{h(x) = \mathbb{1}_C(x)}$$

Proximal gradient /

$$\mathrm{prox}(x) = \underset{z}{\mathrm{argmin}}\ \|x - z\|_2^2 + h(x)$$

Projected gradient

$$= \underset{z \in C}{\mathrm{argmin}}\ \|x - z\|_2^2$$

Proximal Newton

$$z^+ = \underset{z \in C}{\mathrm{argmin}}\ \nabla g(x)^T (z - x) + \frac{1}{2}(z - x)^T H (z - x)$$

$$\min_x \ g(x)$$

$$\text{s.t.} \quad \ell \le x < u$$

$$\min_{x} \quad g(x)$$

$$\text{subject to} \quad \ell \le x \le u$$

$$B_{k-1} = \{i : x_i \le \ell_i + \varepsilon \text{ and } \nabla_i g(x^{(k-1)}) > 0\} \cup$$
$$\{i : x_i \ge u_i - \varepsilon \text{ and } \nabla_i g(x^{(k-1)}) < 0\}$$

$$F_{k-1} = \{1, \dots, n\} \setminus B_{k-1}$$

$$S^{(k-1)} = \left(\nabla^2 g(x^{(k-1)})_{F_{k-1}, F_{k-1}}\right)^{-1}$$

$$x^{(k)} = P_{[\ell,u]}\left(x^{(k-1)} - t_k \begin{bmatrix} S^{(k-1)} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \nabla_{F_{k-1}} g(x^{(k-1)}) \\ \nabla_{B_{k-1}} g(x^{(k-1)}) \end{bmatrix}\right.$$