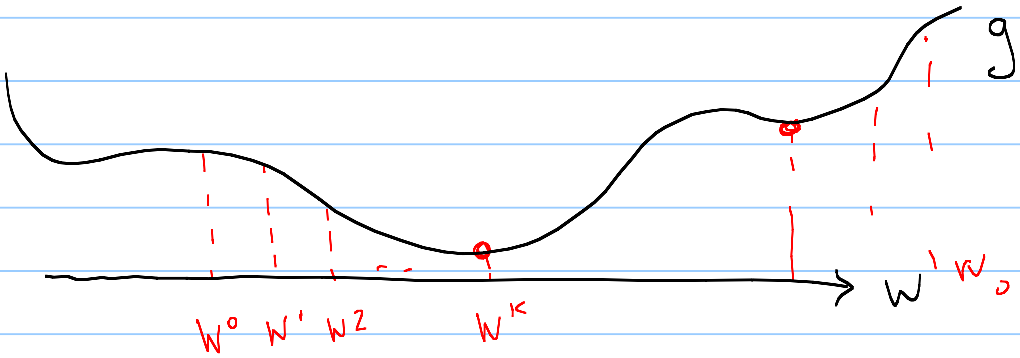
4/9/13

Note Title 1/4/2000

Numerical Optimization

$\min \; g(\bar{w})$

$\boxed{\nabla g(\bar{w}) = 0}$ stationary point



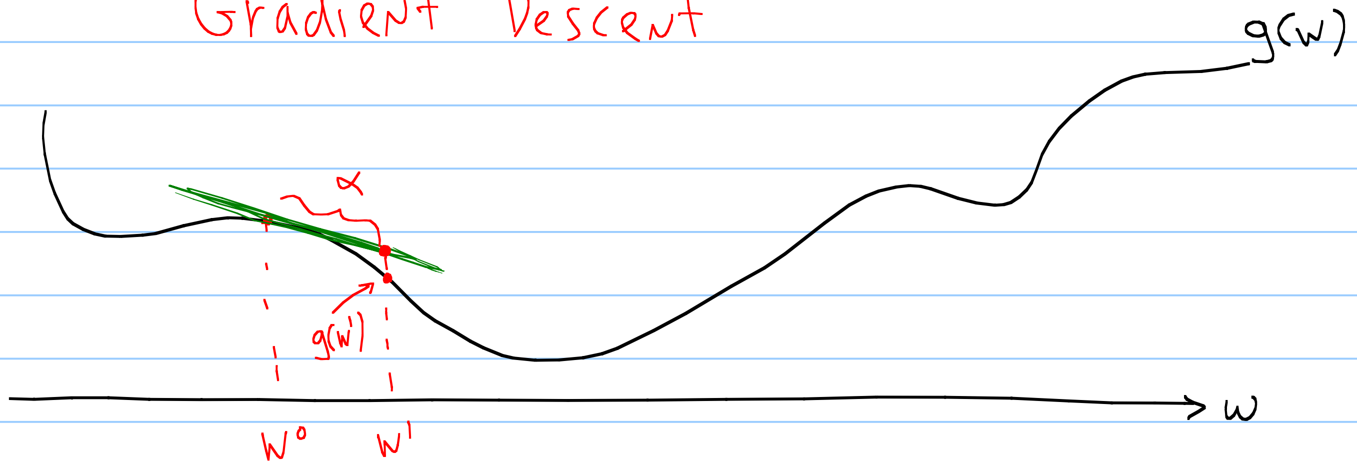
How do you stop the iteration?

- $\| \nabla g(\bar{w}) \|_2^2 < \varepsilon$

ℓ_2 norm
Euclidean norm
$\|x\|_2^2 = \sum_i x_i^2$

- $\dfrac{\|w^k - w^{k-1}\|}{\|w^k\|} < \delta$

Gradient Descent



- Linear approximation at $\bar{w}^{0}$ is given by the 1st order Taylor approximation

$$h(\bar{w}) = g(\bar{w}^{0}) + \nabla g(\bar{w}^{0})^{T}(\bar{w} - \bar{w}^{0})$$

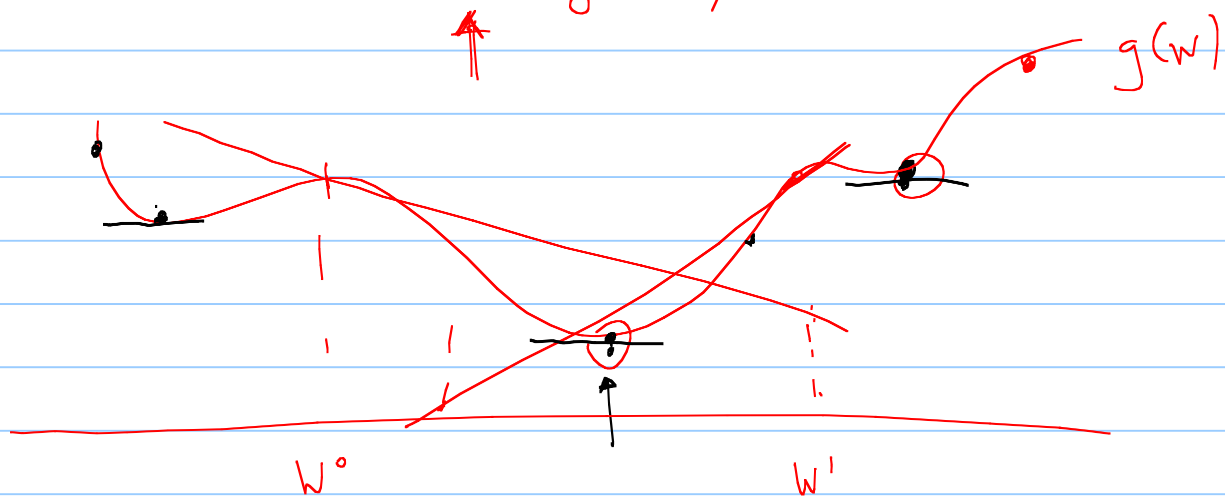
- Downhill direction is given by $-\nabla g(\bar{w}^{0})$ (footnote @ Ch 2)
- Start @ $\bar{w}^{0}$

$$\bar{w}^{1} = \bar{w}^{0} - \alpha \nabla g(\bar{w}^{0})$$

step size or learning rate

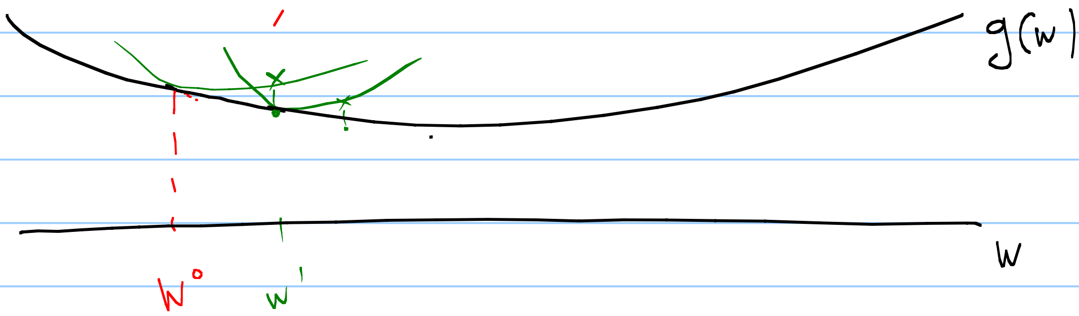
general form

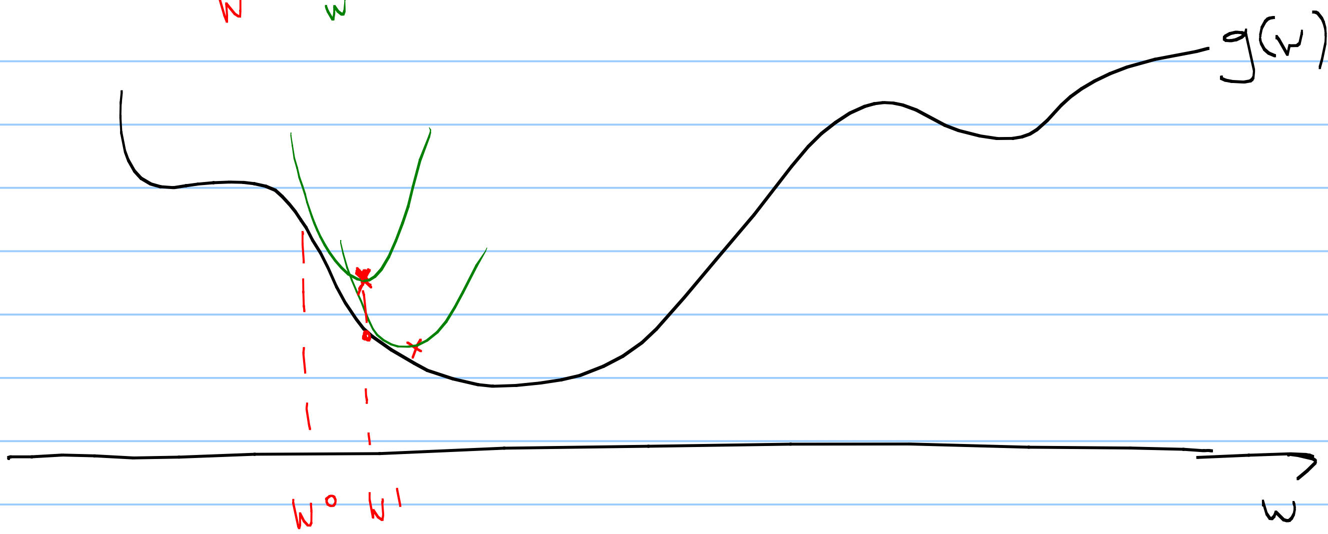
$$\bar{w}^{k} = \bar{w}^{k-1} - \alpha \nabla g(\bar{w}^{k-1})$$



Newton's Method

Use quadratic approximation at each step





2nd order Taylor series approximation

$$h(\bar{w}) = g(\bar{w}^0) + \underbrace{\nabla^T g(\bar{w}^0)\cdot(\bar{w}-\bar{w}^0)} + \frac{1}{2}(\bar{w}-\bar{w}^0)^T \nabla^2 g(\bar{w}^0)\cdot(\bar{w}-\bar{w}^0)$$

Find stationary point of $h(\bar{w})$

$$\nabla h(\bar{w}) = 0$$

$$\cancel{\nabla_{\bar{w}} g(\bar{w}^0)}^{\,0} + \nabla_{\bar{w}}\left(\nabla^T g(\bar{w}^0)\cdot\bar{w}\right) - \cancel{\nabla_{\bar{w}}\left(\nabla^T g(\bar{w}^0)\,\bar{w}^0\right)}^{\,0}$$

$$+ \frac{1}{2}\nabla_{\bar{w}}\left(\underbrace{(\bar{w}-\bar{w}^0)^T \nabla^2 g(\bar{w}^0)(\bar{w}-\bar{w}^0)}\right) = 0$$

$$\rightarrow \nabla g(\bar{w}^0) + \nabla^2 g(\bar{w}^0)\,(\underset{\uparrow\,\bar{w}^1}{\bar{w}} - \bar{w}^0) = 0$$

$\nabla_x(\bar{a}^T\bar{x})$

$= \nabla_{\bar{x}}(\bar{x}^T\bar{a}) = \bar{a}$

$\nabla(x^T A x) = 2Ax$

$$\nabla g(\bar{w}^0) + \nabla^2 g(\bar{w}^0) w_1 - \nabla^2 g(\bar{w}^0)\bar{w}^0 = 0$$

$$\Rightarrow \nabla^2 g(\bar{w}^0)\bar{w}_1 = \nabla^2 g(\bar{w}^0)\bar{w}^0 - \nabla g(\bar{w}^0)$$

if $\nabla^2 g(\bar{w}^0)$ is invertible

$$\Rightarrow \bar{w}^1 = \bar{w}^0 - \left(\nabla^2 g(\bar{w}^0)\right)^{-1} \nabla g(\bar{w}^0)$$

In general

$$\boxed{\bar{w}^k = \bar{w}^{k-1} - \underbrace{\left(\nabla^2 g(\bar{w}^{k-1})\right)^{-1}} \nabla g(\bar{w}^{k-1})}$$