Note Title 3/31/17 2/9/2000

Chapter 2 : Numerical Optimization

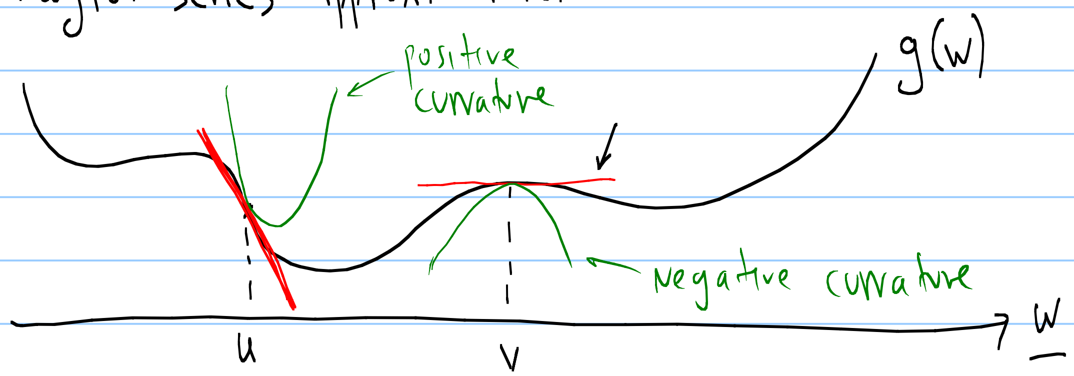
2.1. Calculus defined optimality

$$g(\bar{w}), \quad \bar{w} = [w_1 \; w_2 \; \dots \; w_N]^T$$

scalar

many times differentiable

- Taylor series approximation



→ (N=1) linear approximation to g(w) at w=v

$$\boxed{h(w) = g(v) + g'(v)(w-v)}$$

1st order Taylor series approximation

$$h(v) = g(v) + \cancelto{0}{g'(v)(v-v)} = g(v)$$

$$h'(v) = \frac{\partial}{\partial w}\left[g(v) + g'(v)(w-v)\right] = g'(v)$$

- for general N-dim vector $\bar{w}$, 1st order Taylor approximation

$$\boxed{h(\bar{w}) = g(\bar{v}) + \underbrace{\nabla g(v)^T(\bar{w}-\bar{v})}_{\text{Inner product}}}$$

$$x^T y = [x_1 \cdots x_N]\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \sum_{i=1}^{N} x_i y_i$$

$$\underline{\nabla g(\bar{v})} = \begin{bmatrix} \frac{\partial g(\bar{v})}{\partial v_1} \\ \frac{\partial g(\bar{v})}{\partial v_2} \\ \vdots \\ \frac{\partial g(\bar{v})}{\partial v_N} \end{bmatrix}_{N \times 1}$$

$$\nabla g(v)^T = \begin{bmatrix} \frac{\partial g(\bar{v})}{\partial v_1} & \frac{\partial g(\bar{v})}{\partial v_2} & \cdots & \frac{\partial g(\bar{v})}{\partial v_N} \end{bmatrix}_{1 \times N}$$

- Quadratic approximation – 2^{nd} order Taylor series of $g(\bar{w})$ at $\bar{v}$

$\underline{\underline{N=1}}$

$$\boxed{h(w) = g(v) + g'(v)(w-v) + \frac{1}{2} g''(v)(w-v)^2}$$

Notice: this approximation is tangent at v and contains 1^{st} and 2^{nd} order derivatives of $g(v)$

$$h(v) = g(v)$$
$$h'(v) = g'(v)$$
$$h''(v) = g''(v)$$

<u>Examples</u>

1) $g(w) = \log(1 + e^{w^2})$

$$g'(w) = \frac{1}{1+e^{w^2}} e^{w^2} \cdot 2w = \frac{2we^{w^2}}{1+e^{w^2}}$$

$$h(w) = \log(1+e^{v^2}) + \frac{2ve^{v^2}}{1+e^{v^2}}(w-v)$$

2) $g(\bar{w}) = \frac{1}{2}\bar{w}^T A \bar{w} + \bar{b}^T \bar{w} + c$

$\left(\underline{\bar{b}^T \bar{w} = \bar{w}^T \bar{b}}\right)$

$$\nabla_{\bar{w}} g(\bar{w}) = \frac{1}{2}(A^T \bar{w} + A\bar{w}) + \bar{b}$$

A symmetric $= \frac{1}{2}(A\bar{w} + A\bar{w}) + \bar{b} = A\bar{w} + \bar{b}$

$$\nabla\left(\|\bar{w}\|^2\right)$$

$$\|\bar{w}\|^2 = \bar{w}^T\bar{w} = \sum_i w_i^2$$

$$\frac{\partial}{\partial x} x^2 = 2x$$

$$\nabla_{\bar{w}}\left(\bar{w}^T\bar{w}\right) = \nabla\left(\begin{bmatrix} w_1 & w_2 & \cdots & w_N \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}\right)$$

$$= \nabla \sum_i w_i^2 = \begin{bmatrix} \frac{\partial}{\partial w_1} \sum_i w_i^2 \\ \frac{\partial}{\partial w_2} \sum_i w_i^2 \\ \vdots \\ \frac{\partial}{\partial w_N} \sum_i w_i^2 \end{bmatrix} = \begin{bmatrix} 2w_1 \\ 2w_2 \\ \vdots \\ 2w_N \end{bmatrix} = 2w$$

$$\boxed{\nabla\left(\bar{w}^T A \bar{w}\right) = 2Aw \qquad A \text{ symmetric}}$$

$$\nabla\left(\bar{w}^T\bar{b}\right) = \nabla\left(\begin{bmatrix} w_1 & \cdots & w_N \end{bmatrix}\begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix}\right) = \nabla \sum w_i b_i = \bar{b}$$

Taylor Series for a general N-dim vector

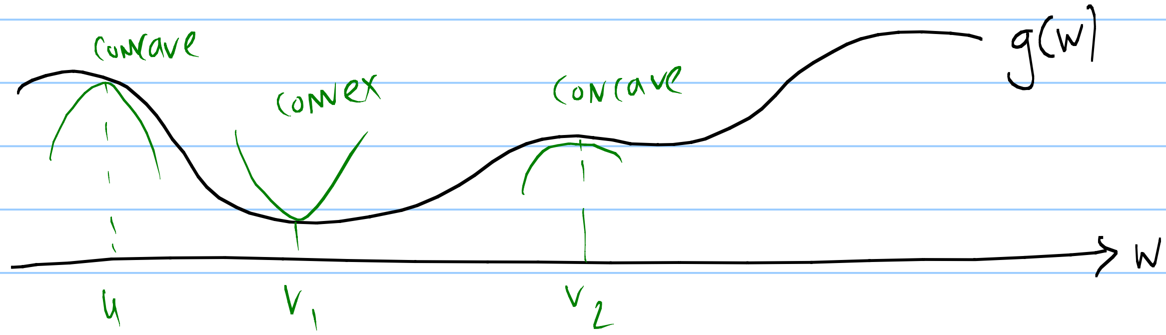
$$h(\bar{w}) = g(\bar{v}) + \nabla g(\bar{v})^T(\bar{w}-\bar{v}) + \frac{1}{2}(\bar{w}-\bar{v})^T \underbrace{\nabla^2 g(\bar{v})}_{\text{Hessian of } g(v)} (\bar{w}-\bar{v})$$

Convexity at a point

- Curvature information of g is contained in its 2^{nd} derivative

$g''(v) \geq 0 \longrightarrow g$ is convex at v (facing upwards)

$g''(v) \leq 0 \longrightarrow g$ is concave at v (" downwards)

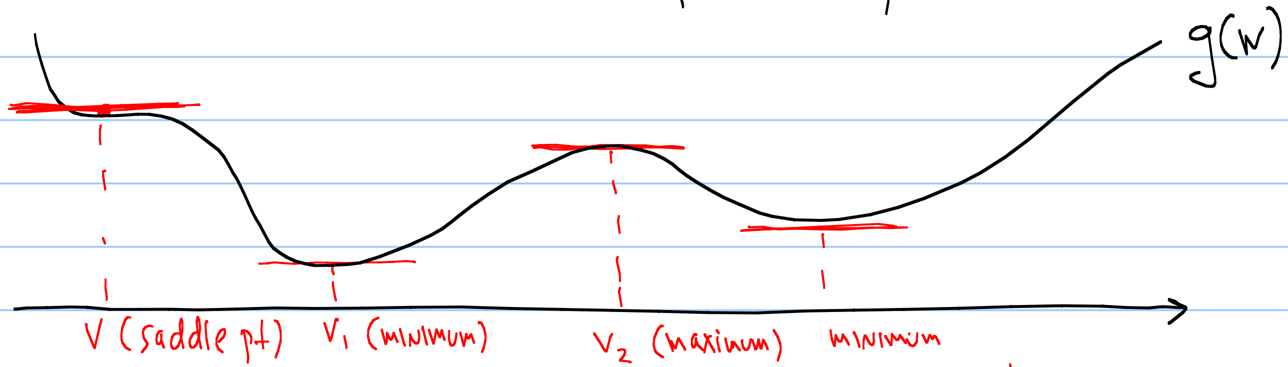


- analogous expressions for general N

$\nabla^2 g(\bar{v})$ if all eigenvalues are non negative $\longleftrightarrow$ g is convex at v

$\nabla^2 g(\bar{v})$ if all eigenvalues are nonpositive $\longleftrightarrow$ g is concave at v

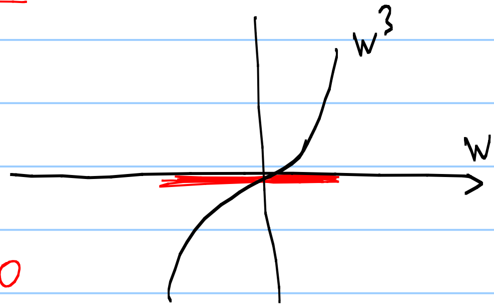
1^{st} order condition for optimality



Each point v where $g'(v) = 0$ or $\nabla g(\bar{v}) = 0$ } 1^{st} order condition of optimality

is called a stationary point

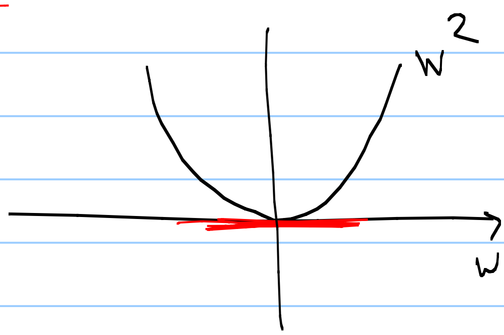
Examples



1) $g(w) = w^3$

$g'(w) = 3w^2 = 0$ for $w = 0$ stationary pt

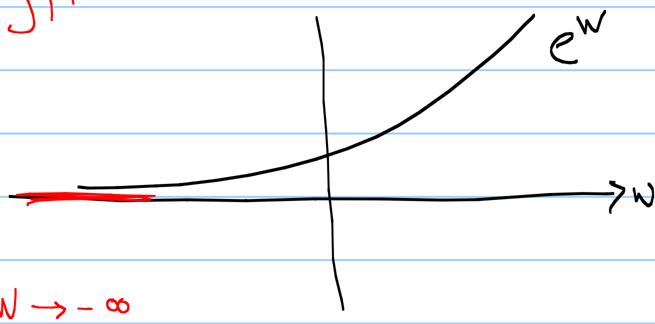
$g''(w) = 6w$



2) $g(w) = w^2$

$g'(w) = 2w = 0 \rightarrow$ for $w = 0$ stationary pt

$g''(w) = 2$



3) $g(w) = e^w$

$g'(w) = e^w = 0 \rightarrow$ for $w \rightarrow -\infty$

$g''(w) = e^w > 0$

4) $g(\bar{w}) = \frac{1}{2}\bar{w}^T A \bar{w} + \bar{b}^T \bar{w} + c$

$\nabla g(\bar{w}) = A\bar{w} + \bar{b} = 0 \Rightarrow A\bar{w} = -\bar{b}$

$\Rightarrow \bar{w} = A^{-1}\bar{b}$

if A invertible