4/8/20

Numerical Optimization

$g(\mathbf{w})$ — many times differentiable

g: scalar

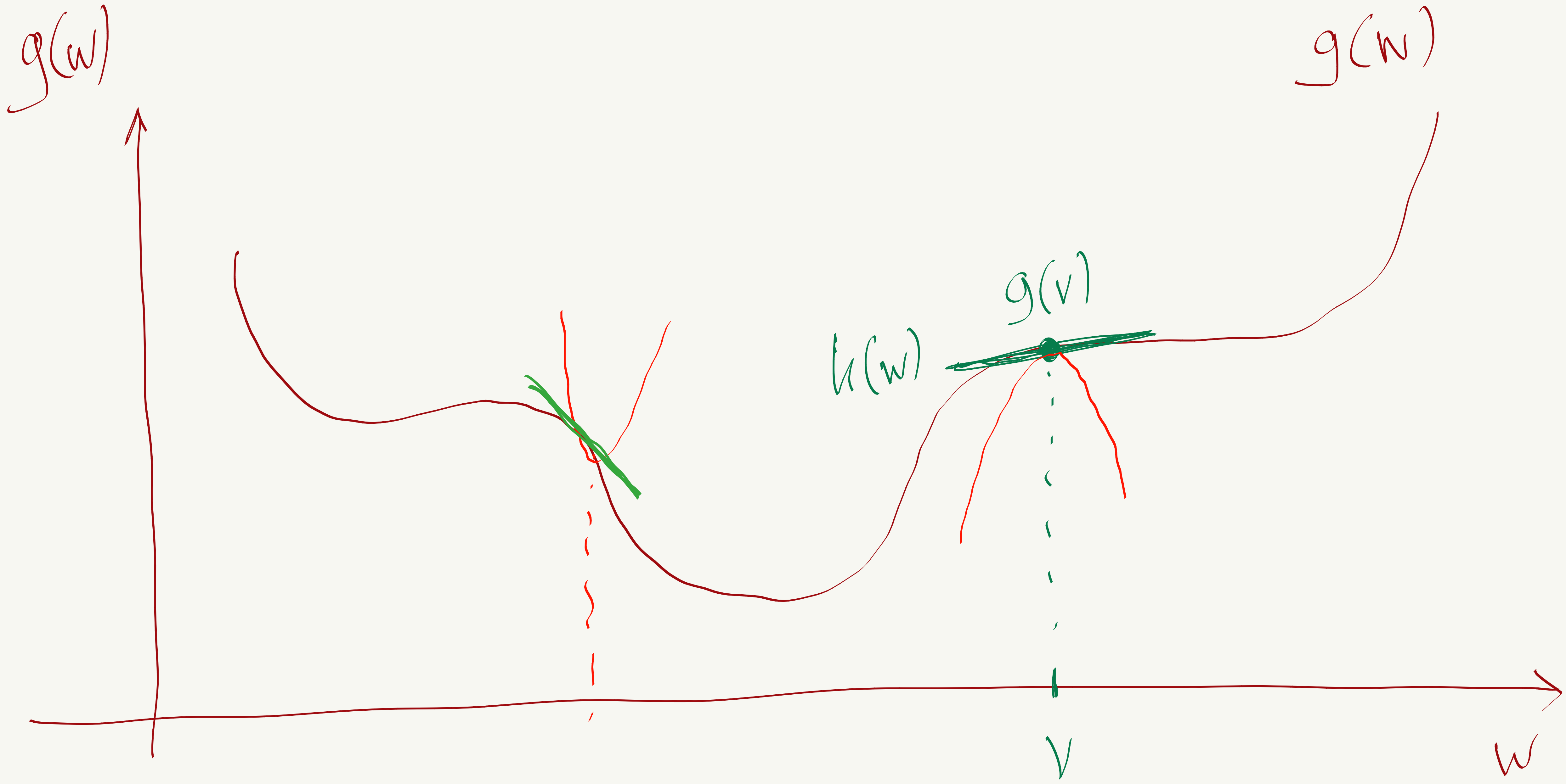
$\mathbf{w} \in \underline{\underline{R}}^N$ N-dimensional

$$g(\mathbf{w}) = \|\mathbf{w}\|_2^2 = \mathbf{w}^T\mathbf{w} = \sum_{i=1}^{N} w_i^2$$

ℓ_2 - Norm

or Euclidean Norm

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}_{N \times 1}$$



Taylor series approximation

1st order approximation

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

h(w) is tangent to g(w) at v

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

$$= \cancel{\frac{d}{dw} g(v)}^{0} + \frac{d}{dw}\left[g'(v)(w-v)\right]$$

$$= 0 + g'(v) \cdot 1 = g'(v)$$

$\underline{N}$ - dimensional space

1st order Taylor series approximation

$\underline{\text{Vector}}$ ~~w~~ , $\bar{w}$

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

gradient vector

Gradient of $g(\bar{w})$: $\nabla g(\bar{w})_{N \times 1}$

scalar $\in \mathbb{R}^N$

$$\nabla g(\bar{w}) = \begin{bmatrix} \frac{\partial g(\bar{w})}{\partial w_1} \\ \frac{\partial g(\bar{w})}{\partial w_2} \\ \vdots \\ \frac{\partial g(\bar{w})}{\partial w_N} \end{bmatrix} = \begin{bmatrix} \frac{\partial g(\bar{w})}{\partial w_1} & \cdots & \frac{\partial g(\bar{w})}{\partial w_N} \end{bmatrix}^T$$

Example

Inner product

$$g(\bar{w}) = \bar{w}^T \bar{v} = \sum_{i=1}^{N} w_i v_i = \underline{w_1 v_1} + w_2 v_2 + \cdots + w_N v_N$$

$$\nabla_{\bar{w}}\, g(\bar{w}) = \begin{bmatrix} \frac{\partial g(\bar{w})}{\partial w_1} \\ \vdots \\ \frac{\partial g(\bar{w})}{\partial w_N} \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} = \bar{v}$$

$$g(\bar{w}) = \|\bar{w}\|_2^2 = \underline{w_1^2} + w_2^2 + \cdots + w_N^2$$

$$\nabla g(\bar{w}) = 2\bar{w}$$

2^{nd} order Taylor series approximation

$N=1$

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

Note: this approximation is tangent at v & contains 1st & 2nd order derivatives

$$\left.\begin{aligned} h(v) &= g(v) \\ h'(v) &= g'(v) \\ h''(v) &= g''(v) \end{aligned}\right\} \text{check}$$

N-dim space

$$h(\bar{w}) = g(\bar{v}) + \nabla^T g(\bar{v})(\bar{w}-\bar{v}) + \frac{1}{2}(\bar{w}-\bar{v})^T_{1\times N} \cdot \underbrace{\nabla^2 g(\bar{v})_{N\times N}}_{\text{matrix}} (\bar{w}-\bar{v})_{N\times 1}$$

Hessian of $g(\bar{w})$
(2nd derivative)

Hessian

$$\nabla^2 g(\bar{w}) = \begin{bmatrix} \frac{\partial^2}{\partial w_1 \partial w_1} g(\bar{w}) & \frac{\partial^2}{\partial w_1 \partial w_2} g(\bar{w}) & \cdots & \frac{\partial^2}{\partial w_1 \partial w_N} g(\bar{w}) \\ \frac{\partial^2}{\partial w_2 \partial w_1} g(\bar{w}) & \cdots & & \frac{\partial^2}{\partial w_2 \partial w_N} g(\bar{w}) \\ \vdots & & & \vdots \\ \frac{\partial^2}{\partial w_N \partial w_1} g(\bar{w}) & \cdots & & \frac{\partial^2}{\partial w_N \partial w_N} g(\bar{w}) \end{bmatrix}$$