4/10/20

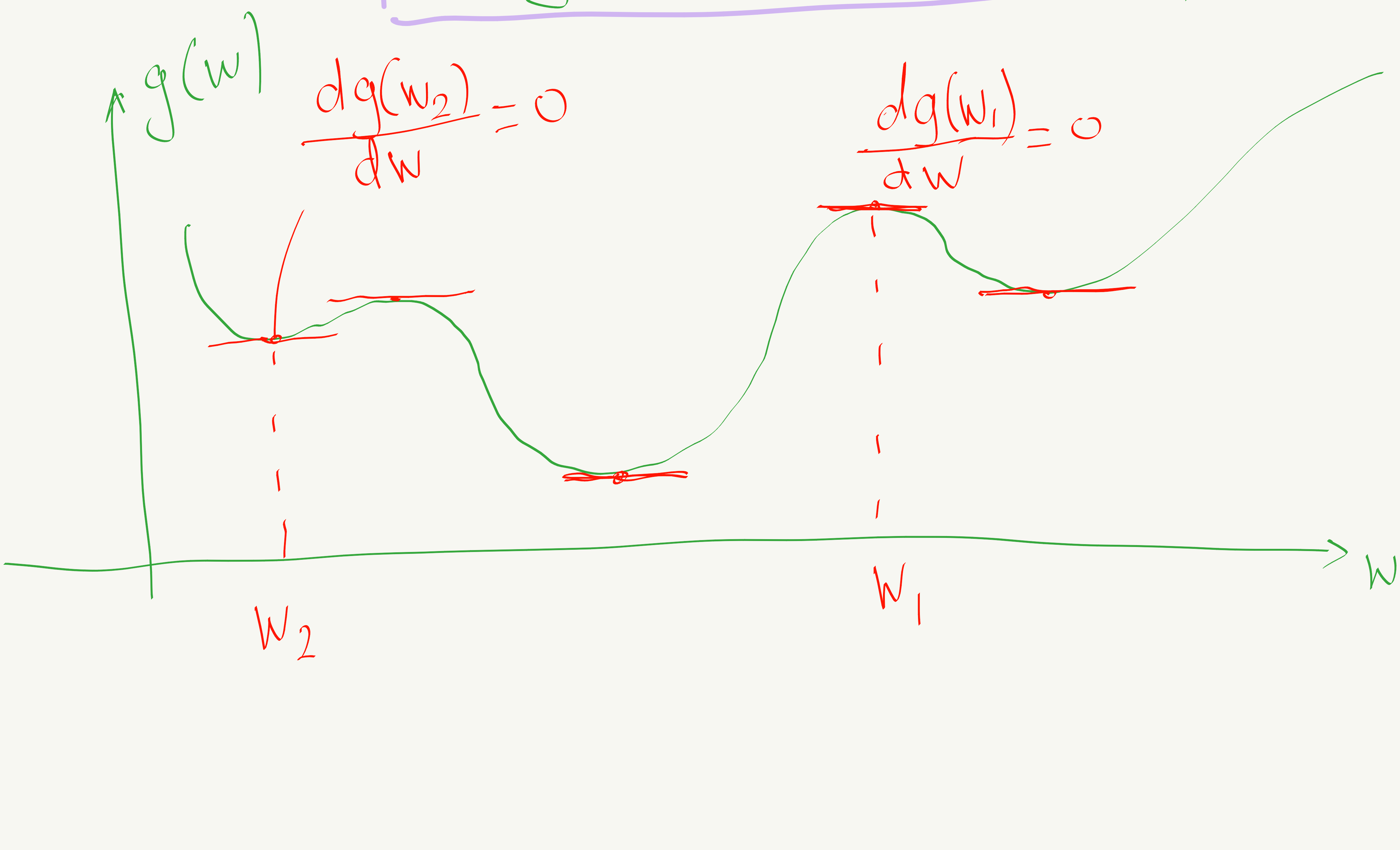
Numerical Optimization

Cost function $g(\bar{w})$ — g scalar, $\bar{w} \in \mathbb{R}^N$

"Optimal" points of $g(\bar{w})$

First order optimality condition

$$\boxed{\nabla g(\bar{w}) = 0 \Rightarrow \text{stationary points}}$$



L3_041020

Example 3.1 in textbook

saddle point

$\sin(2w)$ w^3 $\sin(3w) + 0.3w^2$

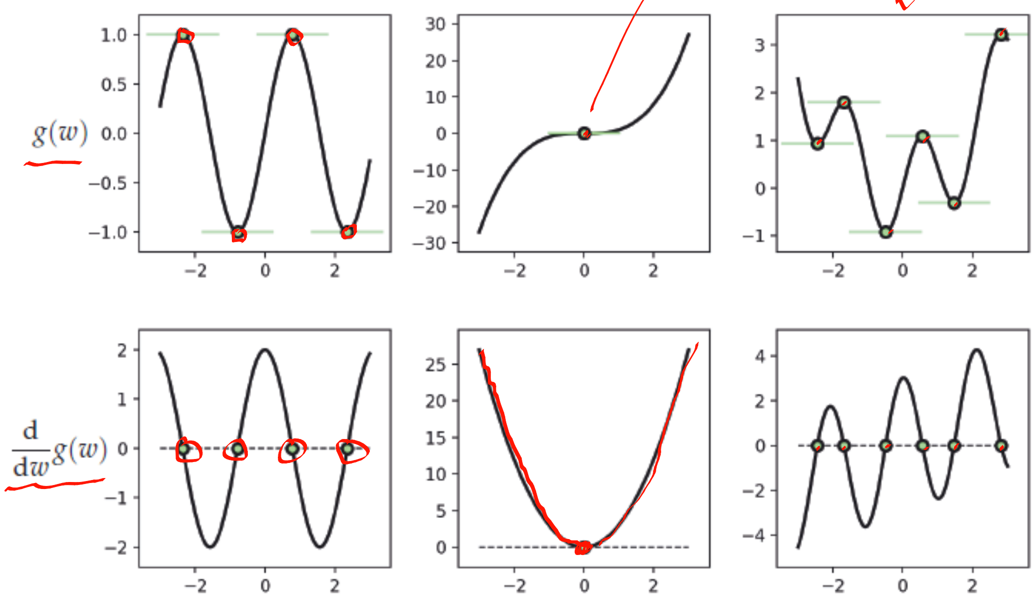


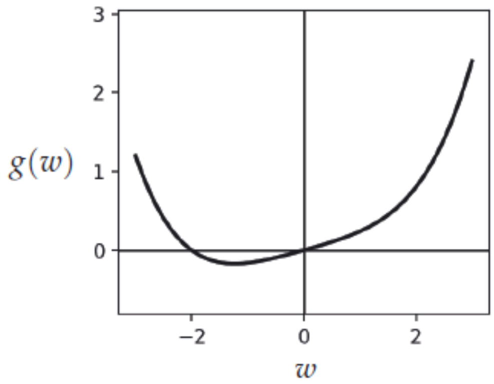
Figure 3.2 Figure associated with Example 3.1. From left to right in the top row, the functions $g(w) = \sin(2w)$, w^3, and $\sin(3w) + 0.3w^2$ are plotted along with their derivatives in the bottom row. See text for further details.

$$\frac{dw^3}{dw} = 3w^2 = 0 \rightarrow w = 0$$

L3_041020

Example 3.3 in textbook

$$g(w) = \frac{1}{50}\left(w^4 + w^2 + 10w\right)$$



$$w = \frac{\sqrt[3]{\sqrt{2031} - 45}}{\sqrt[3]{36}} - \frac{1}{\sqrt[3]{6\left(\sqrt{2031} - 45\right)}}$$

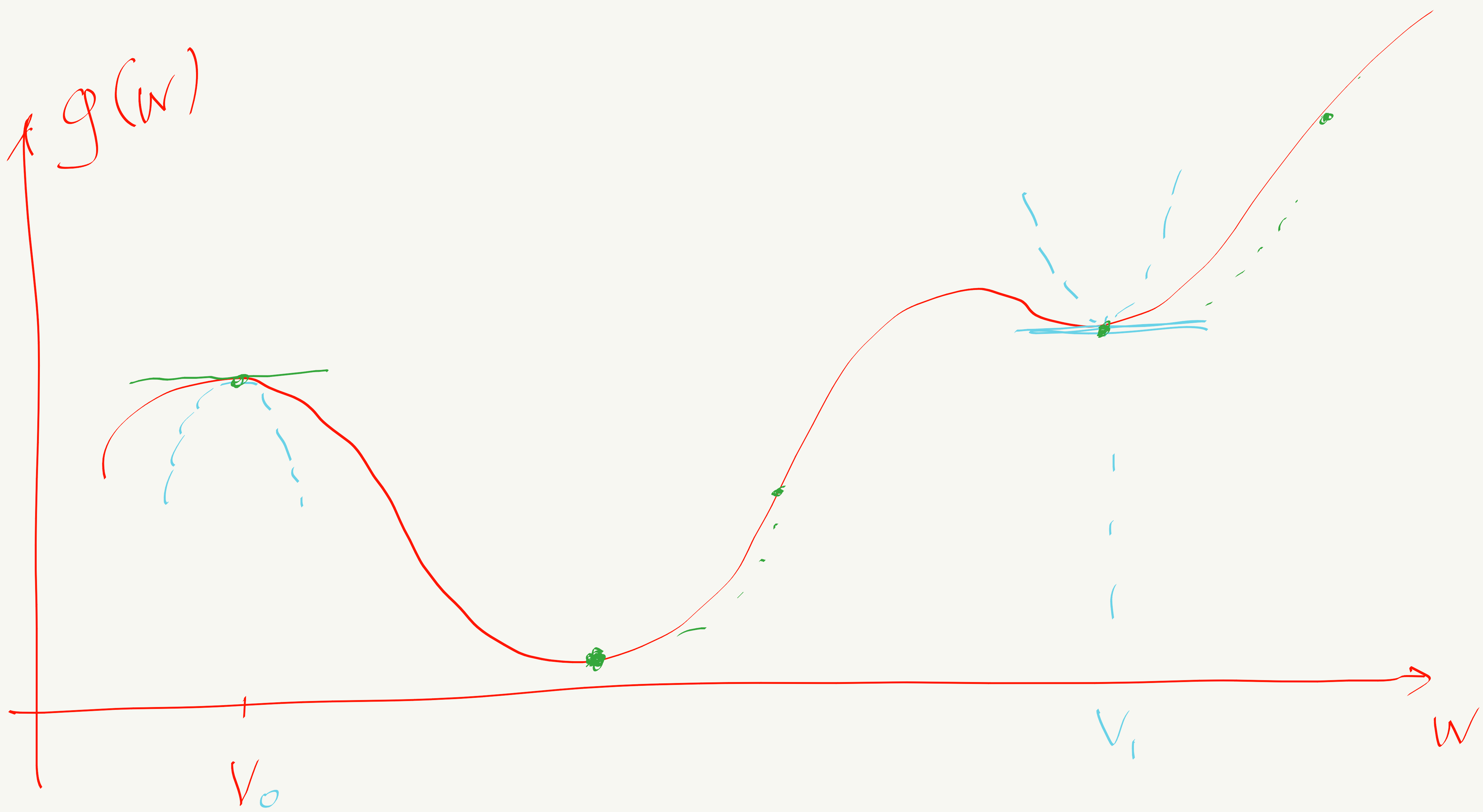
Convexity at a point

Curvature information of $g(w)$ $\in \mathbb{R}$ is contained in the 2^{nd} derivative

$\underline{N=1}$

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

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



$g(v)$ is concave at v_0

$g(v)$ is convex at v_1

N-dim space $g(\bar{w}) \in R^N$

$$\nabla^2_{\bar{w}}\, \underbrace{g(\bar{v})}_{N \times N} = \nabla^2_{\bar{w}}\, g(\bar{w})\Big|_{\bar{w}=\bar{v}}$$

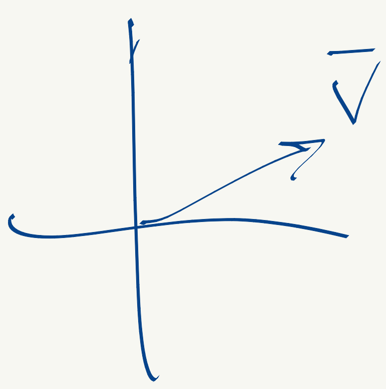
eigenvalues of the Hessian will determine the curvature

---

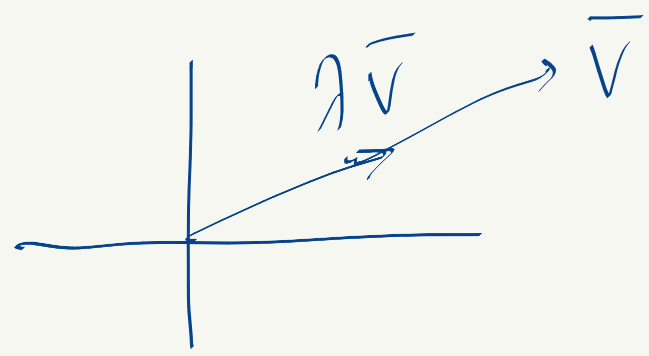
matrix X

if $X\bar{v} = \lambda \bar{v} \Rightarrow \bar{v}$ is an eigen-vector & λ an eigen-value

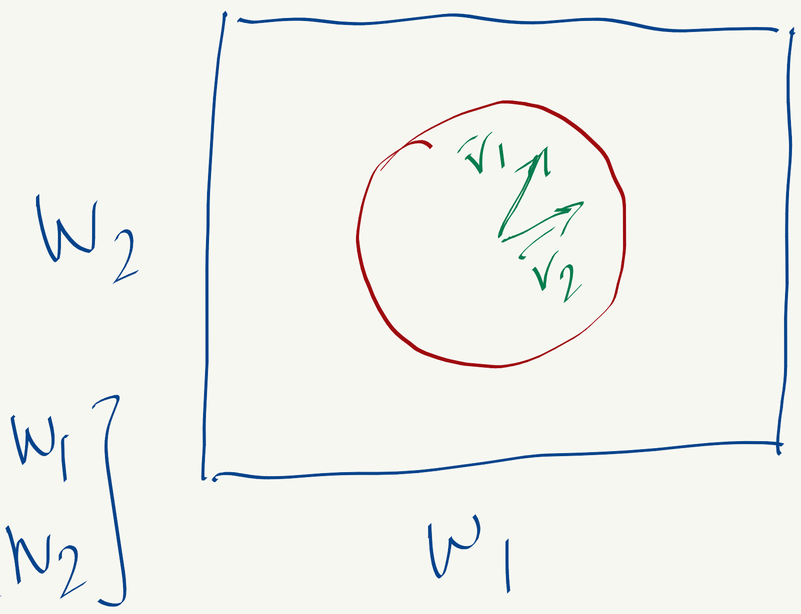
$$\bar{w} \longrightarrow \boxed{X} \longrightarrow X\bar{w}$$



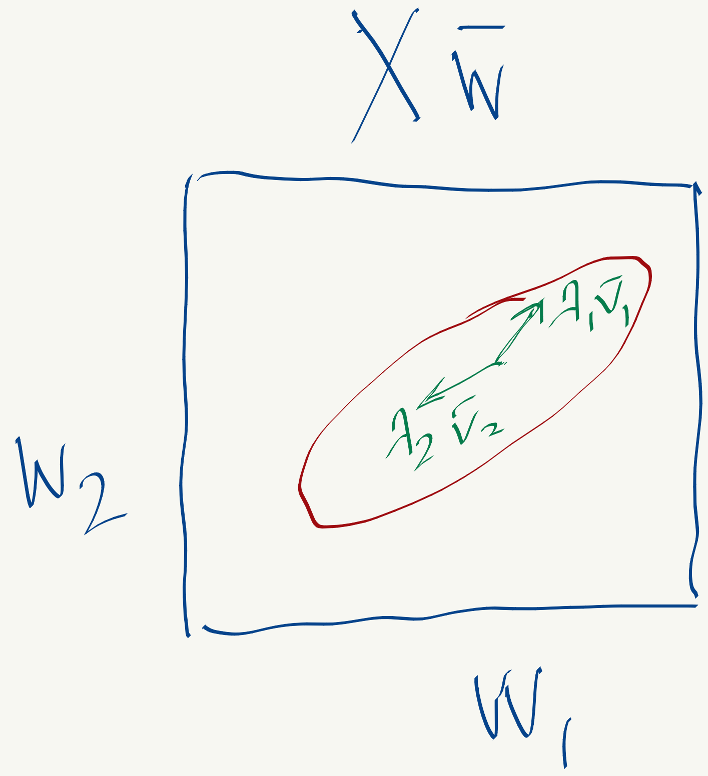
$$\bar{v} \longrightarrow \boxed{X} \longrightarrow \lambda \bar{v}$$



appendix C



$\bar{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$



if $X_{N \times N}$ is symmetric $(X = X^T)$

is Hessian symmetric? Yes!

$\longrightarrow X\bar{v} = \lambda \bar{v}$, N eigen-values
N eigen-vectors

$V = \begin{bmatrix} | & | & & | \\ \bar{v}_1 & \bar{v}_2 & \cdots & \bar{v}_N \\ | & | & & | \end{bmatrix}$

$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{bmatrix}$

$\boxed{XV = VD}$ $\left\{ \begin{array}{l} X\bar{v}_1 = \lambda_1 \bar{v}_1 \\ \vdots \\ X\bar{v}_N = \lambda_N \bar{v}_N \end{array} \right.$

assume $\|\bar{v}_i\|_2 = 1$ $\left\{ \underbrace{\bar{v}_i' = \frac{\bar{v}_i}{\|\bar{v}_i\|_2}} \rightarrow \|\bar{v}_i'\|_2 = 1 \right.$

$\Rightarrow V^T V = V V^T = I$

$\Rightarrow \boxed{X = V D V^T}$

N-dim space

if all eigen-values of $\nabla^2 g(\bar{v})$
are non-negative

$\Longrightarrow g(\bar{w})$ is convex at $\bar{v}$

$$\nabla^2 g(\bar{v}) \succeq 0$$

$$\uparrow \succeq$$

---

if all e-values of $\nabla^2 g(\bar{v})$
are non-positive

$\Longrightarrow g(\bar{w})$ is concave at $\bar{v}$

$$\nabla^2 g(\bar{v}) \preceq 0$$

L3_041020

one min　　　　many minima　　　　Saddle point

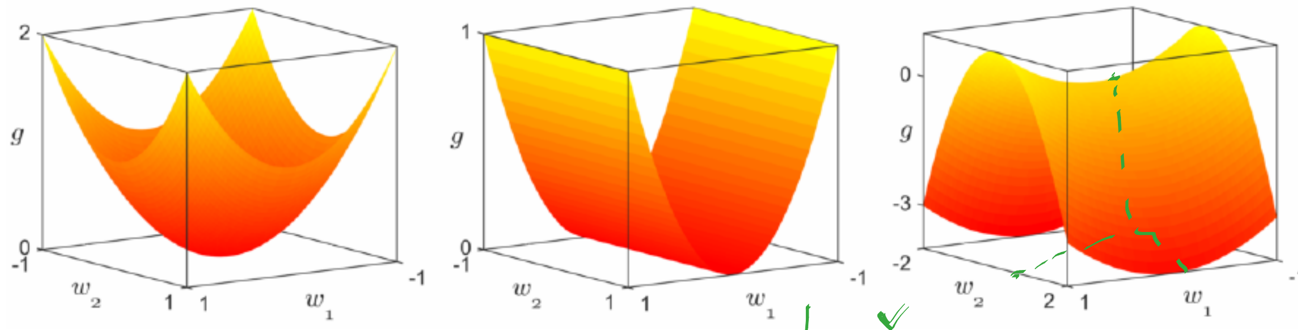


Figure 2.5. *Three quadratic cost function* $\boxed{g(\mathbf{w}) = \mathbf{w}^T\mathbf{A}\mathbf{w} + \mathbf{b}^T\mathbf{w} + c}$ *generated by different instances of matrix* $\mathbf{A}$. *In all three cases* $\mathbf{b}$ *and* c *are set to zero. (left)* $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ *generates a (strictly) convex upward facing cup. (middle)* $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ *generates a long narrow half-pipe with infinitely many global minima along the bottom. (right)* $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ *produces the upward and downward curving quadratic surface with a saddle point at* $\mathbf{w} = \mathbf{0}$.

$$\nabla g(\bar{w}) = \nabla\left(\bar{w}^T A \bar{w} + \bar{b}^T \bar{w} + c\right)$$

$$= \underbrace{\nabla\left(\bar{w}^T A \bar{w}\right)}_{= 2A\bar{w}} + \underbrace{\nabla\left(\bar{b}^T \bar{w}\right)}_{= \bar{b}} + \cancel{\nabla(c)}^{\,0}$$

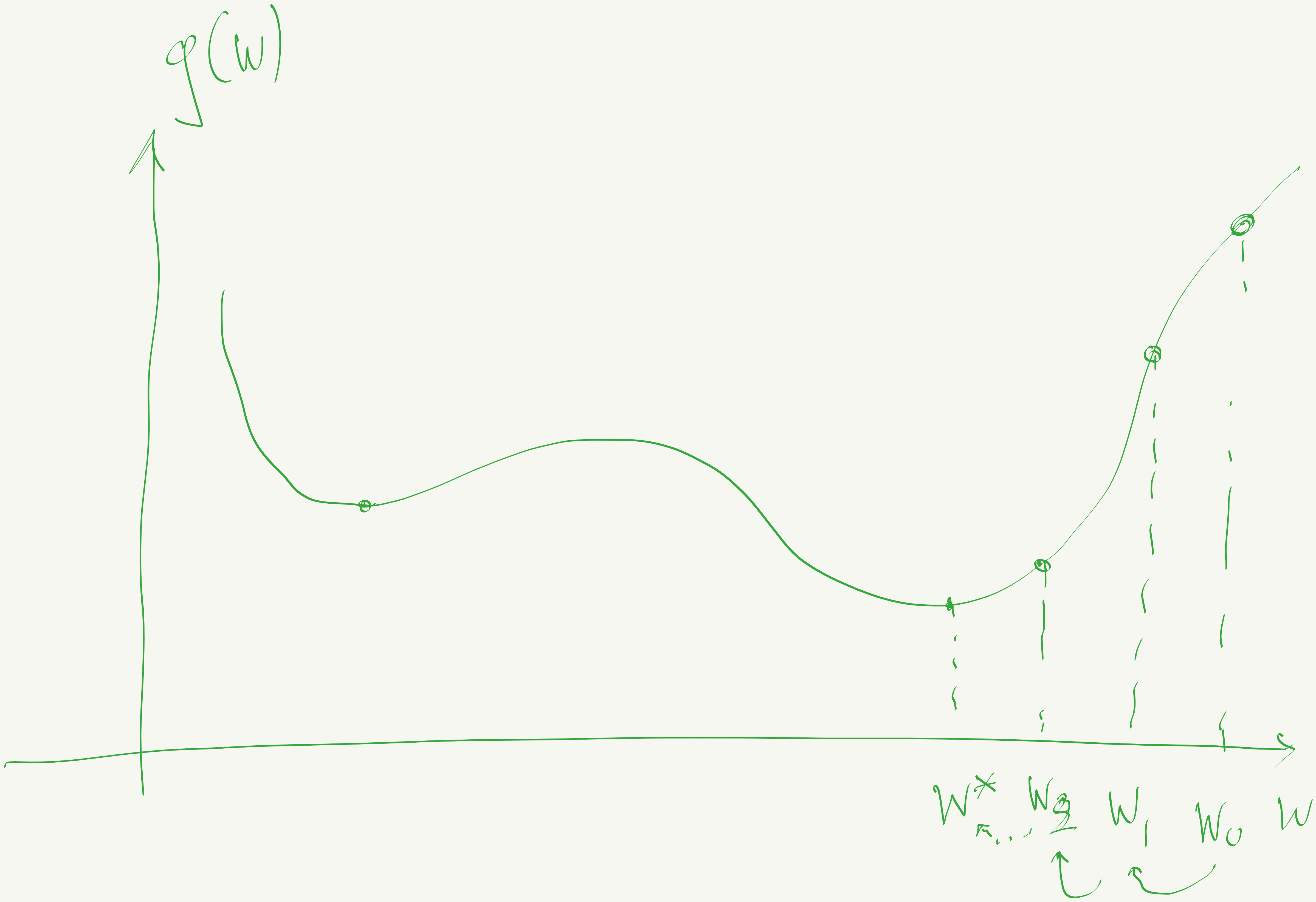
A symmetric

$$\boxed{\begin{aligned} \nabla\left(\bar{w}^T \bar{b}\right) &= \bar{b} \\ \bar{w}^T \bar{b} &= \bar{b}^T \bar{w} \end{aligned}}$$

$$\nabla\left(\bar{w}^T A \bar{w}\right) = \left(A + A^T\right)\bar{w}$$

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

$$\nabla^2 g(\bar{w}) = \nabla\left(\nabla g(\bar{w})\right) = \nabla\left(2A\bar{w} + \bar{b}\right) = 2A$$



$\longrightarrow$ • gradient descent method

$\longrightarrow$ • Newton's method