# Computational Foundations for ML

## 10-607

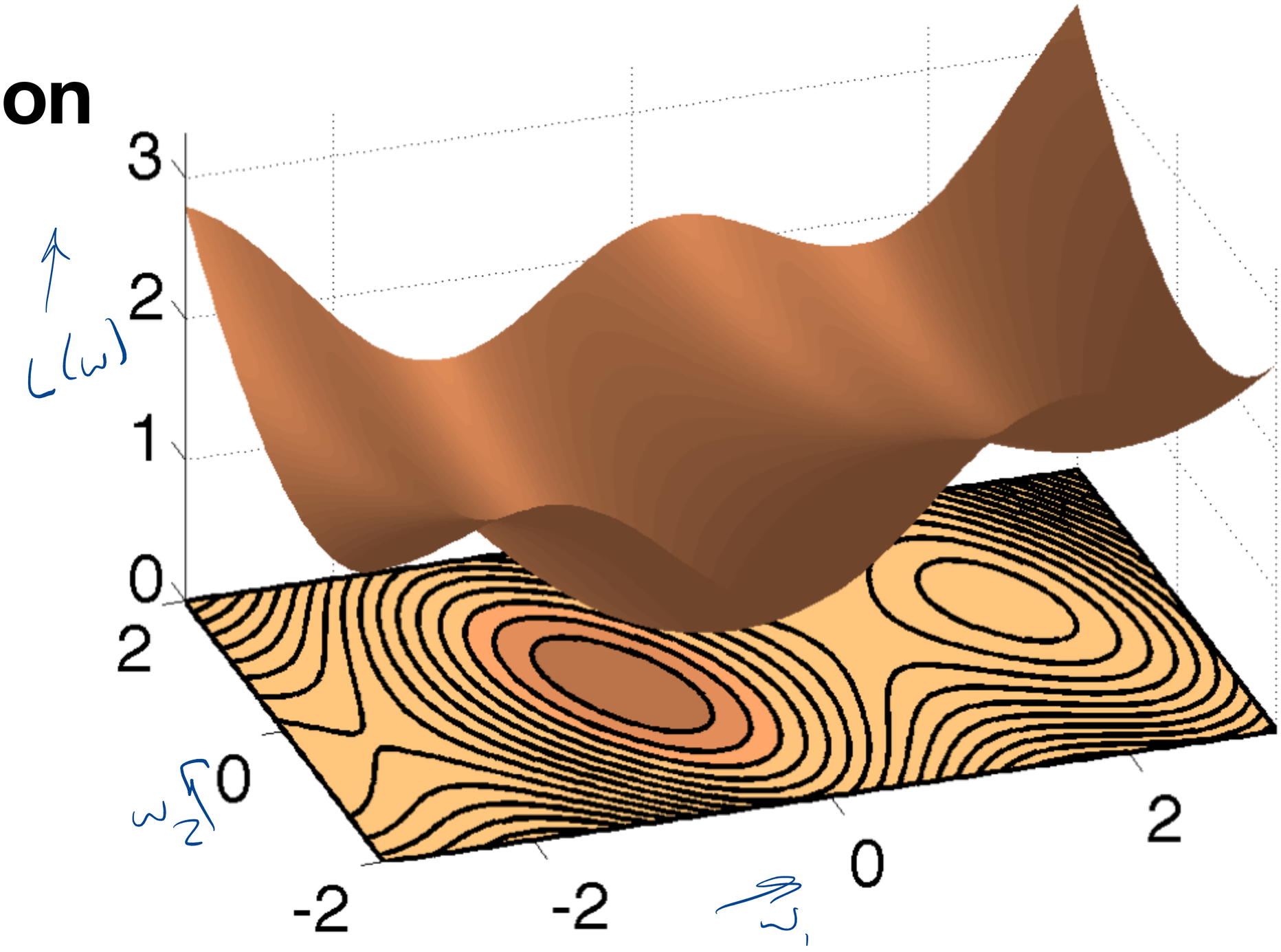
Geoff Gordon

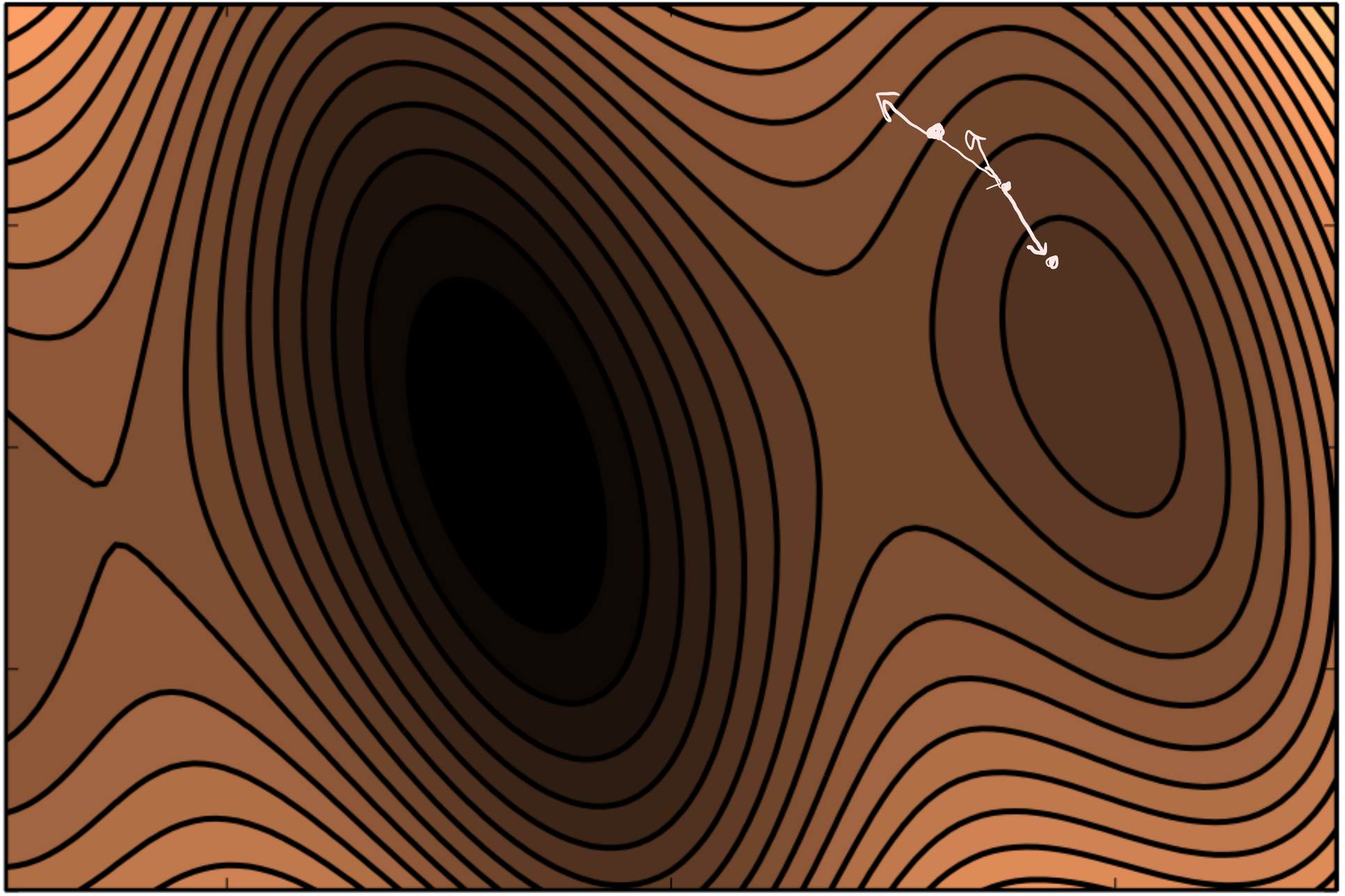
# Notes and reminders

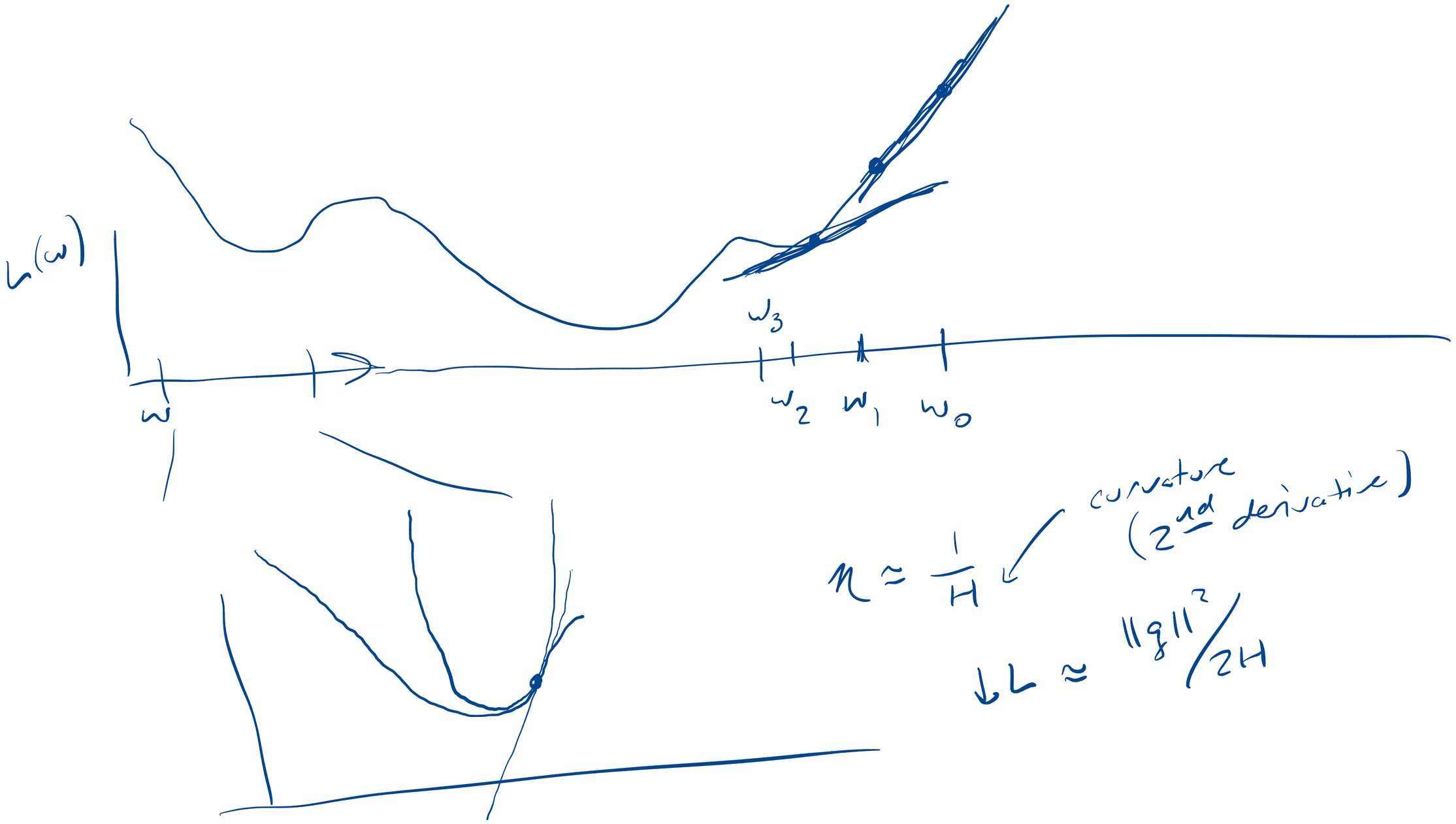
- Please submit repl.it ID through Canvas

# Optimization

$\min_w L(w)$

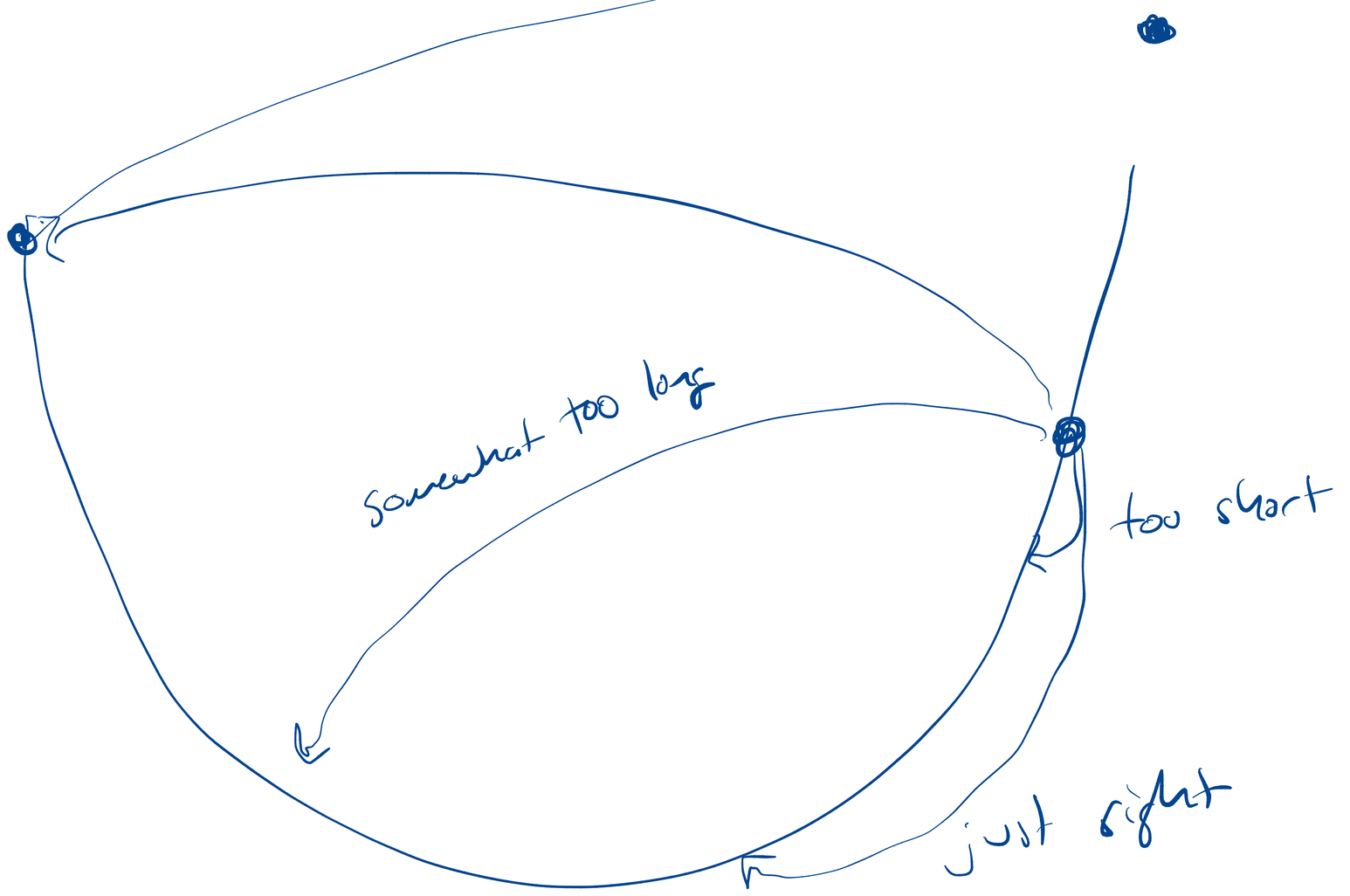






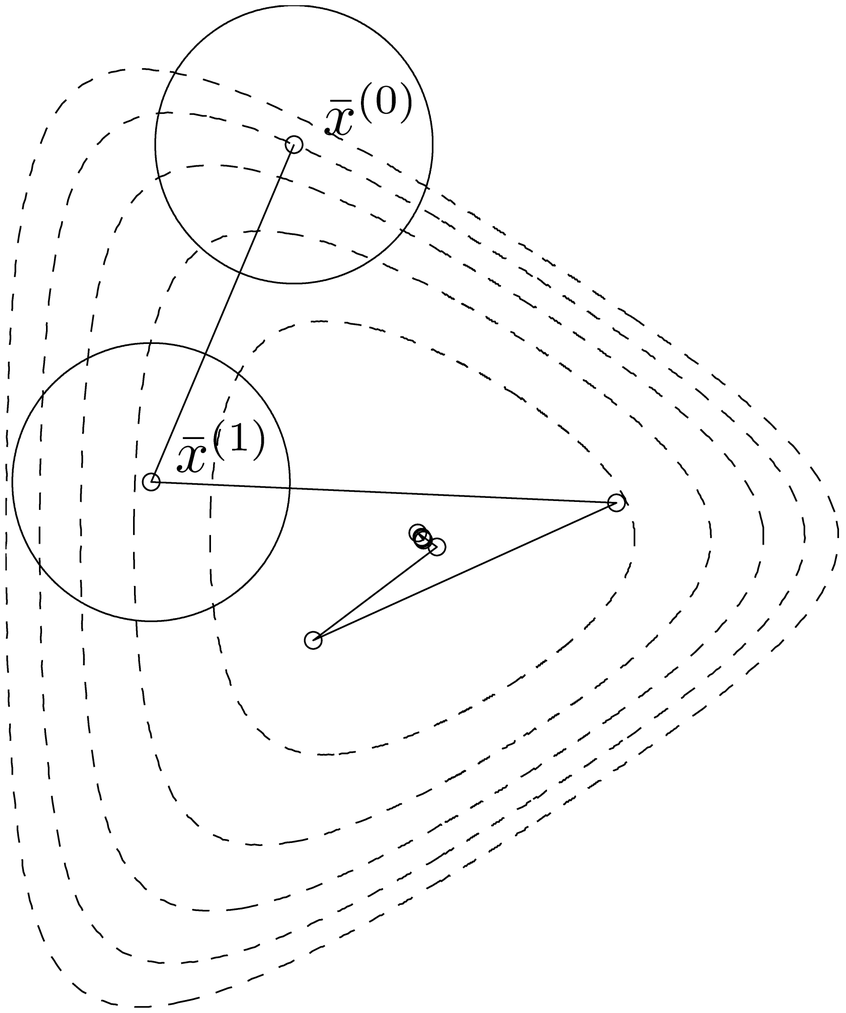

$\eta \simeq \frac{1}{H}$ ← curvature (2nd derivative)

$\downarrow L \simeq \|g\|^2 / 2H$


somewhat too long
too short
just right

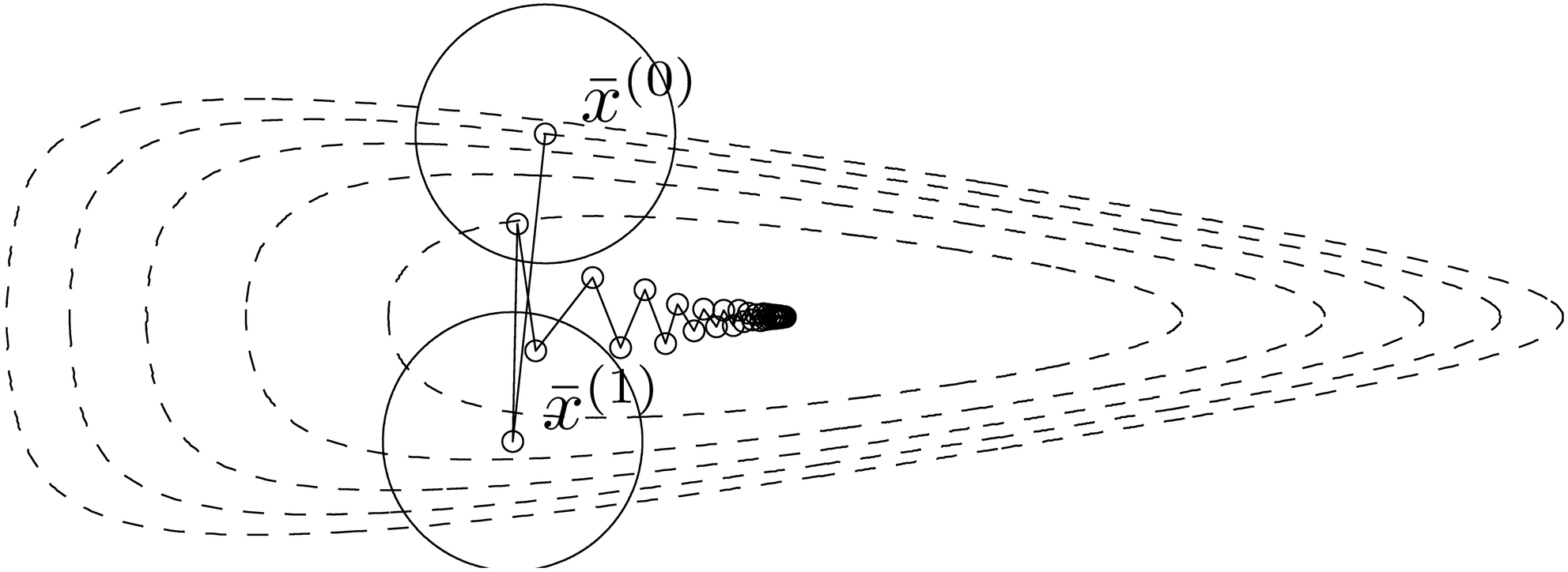
# Conditioning

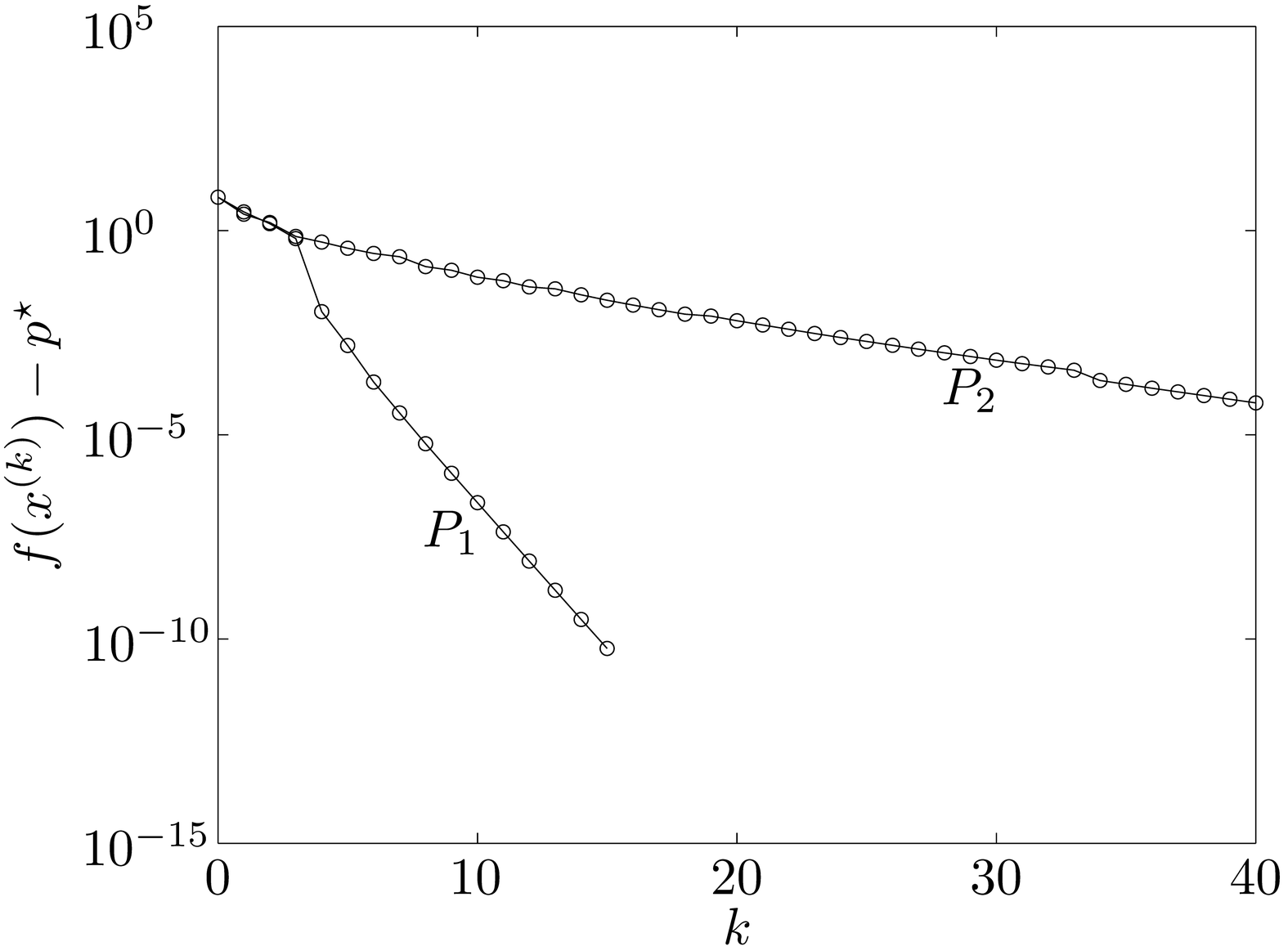
## low (good) condition number



# Conditioning

## high (bad) condition number




$10^5$
$10^0$
$10^{-5}$
$10^{-10}$
$10^{-15}$
$f(x^{(k)}) - p^\star$
0
10
20
30
40
$k$
$P_1$
$P_2$

$$H = \frac{d^2}{dw^2} L(w)$$

$w \in \mathbb{R}^d$ $\quad g \in \mathbb{R}^d$

$H \in \mathbb{R}^{d \times d}$

$$H = U \Sigma U^T$$

orthonormal ← diagonal +ve

$$\kappa = \Sigma_{11} / \Sigma_{dd}$$

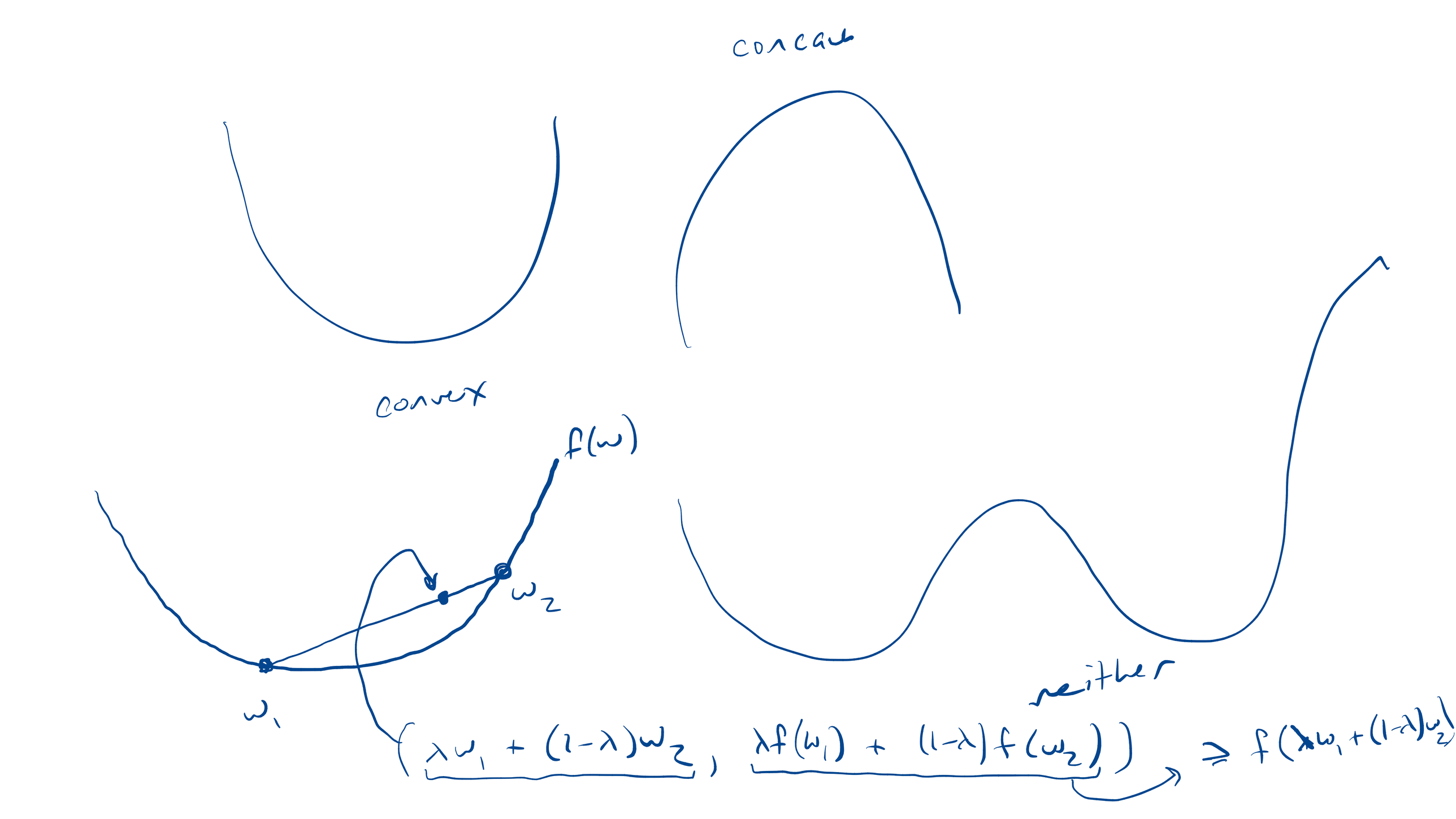
Preconditioning

$$\min_w L(w) \longrightarrow \min_v L(f(v))$$

↑ we pick

$$w = f(v) = \begin{pmatrix} 2 & 0 \\ 0 & .5 \end{pmatrix} v$$

$$v = f^{-1}(w)$$


concave
convex
f(w)
$w_2$
$w_1$
neither
$(\lambda w_1 + (1-\lambda) w_2,\ \lambda f(w_1) + (1-\lambda) f(w_2)) \geq f(\lambda w_1 + (1-\lambda) w_2)$

$w \leftarrow w_0 \qquad \beta \in (0,1)$

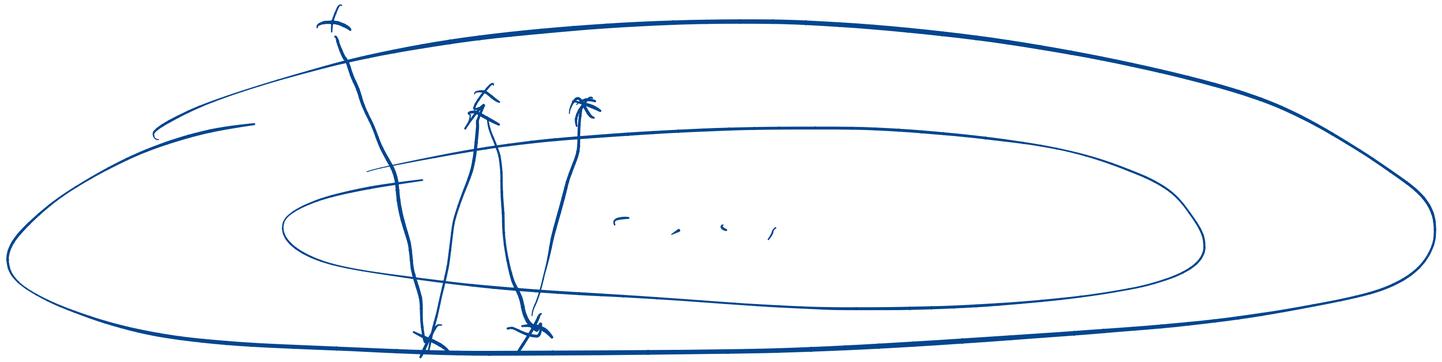
$m_0 \leftarrow 0$

for $i = 1, 2, \ldots$

$$g_t = \frac{dL}{dw}\Big|_{w_t}$$

$$m_t = (1-\beta) g_t + \beta m_{t-1}$$

$$w_{t+1} = w_t - \frac{\eta m_t}{1-\beta^t}$$



$$= (1-\beta) g_t + \beta (1-\beta) g_{t-1} + \beta^2 m_{t-2}$$