ED5340 - Data Science: Theory and Practise

L14 - Optimization

Ramanathan Muthuganapathy (https://ed.iitm.ac.in/~raman)
Course web page: https://ed.iitm.ac.in/~raman/datascience.html
Moodle page: Available at https://courses.iitm.ac.in/

Why optimization

- Fundamental to machine and deep learning
- Cost function solving needs optimization (or solve using direct methods)
- Basic differential calculus / linear algebra

Optimization

- Unconstrained (e.g. min $J(w)$, e.g $J(w) = w^2, J(w) = w^3,$ $J(w) = w^2 + 54/w$)
- constrained optimization (e.g. $min\, J(w), w > 0$)

Unconstrained optimization

- Single variable (e.g. min $J(w)$, e.g $J(w) = w^2$, $J(w) = w^3$, $J(w) = w^2 + 54/w$)
- multivariable (e.g. $min\, J(w_0, w_1) = (w_0 - 2)^2 + (w_1 - 2)^2$)

Ramanathan Muthuganapathy, Department of Engineering Design, IIT Madras

Optimality criteria - single variable

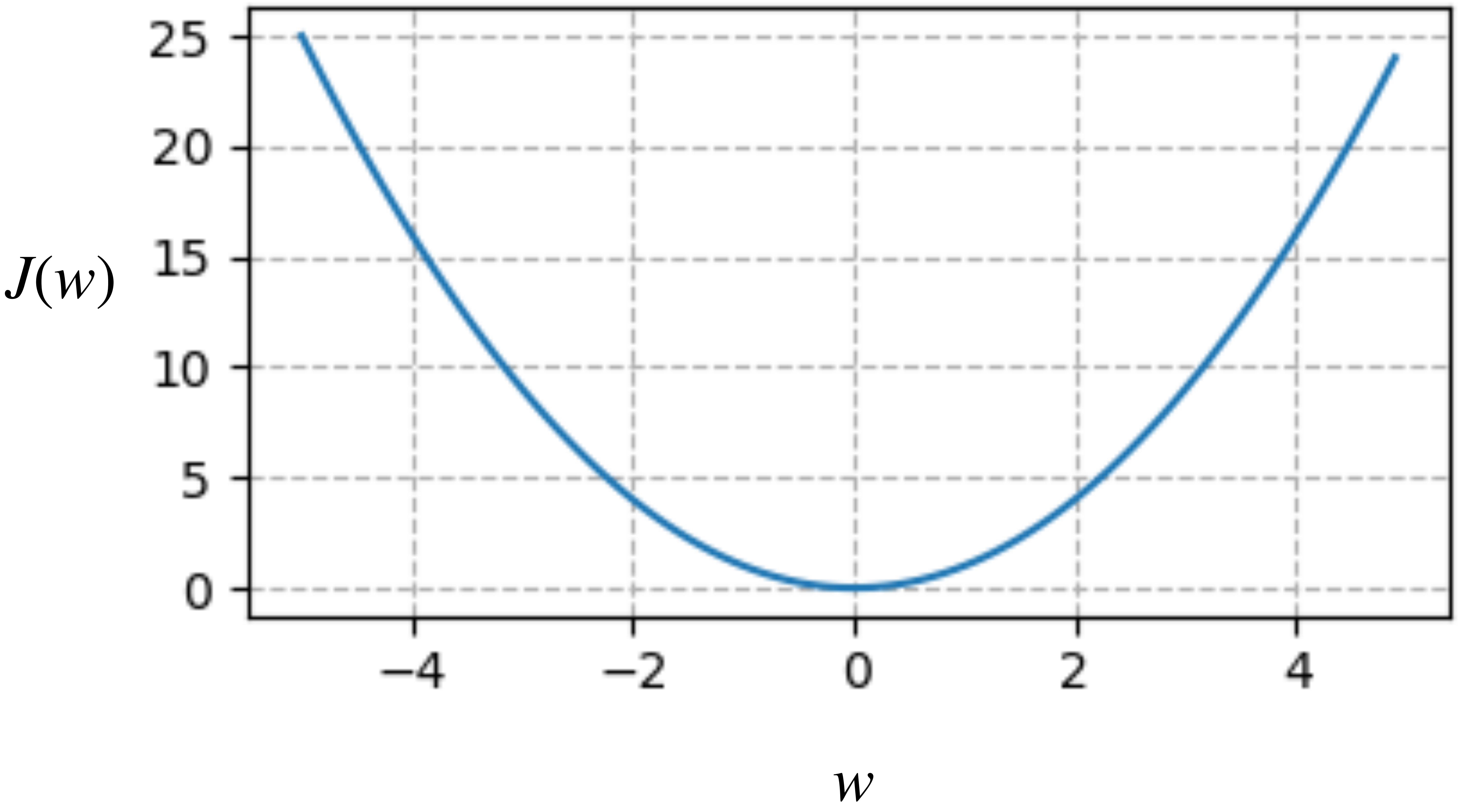
- Single variable (e.g $J(w) = w^2, J(w) = w^3,\ J(w) = w^4$)
- min $J(w)$
  - The value of w for which the function $J(w)$ has the least (minimum) value
  - Unimodal function
  - Local minimum (in this case, this is also global minimum)

Demo of various power functions

Optimality criteria - single variable

$J(w) = w^2$

- $J(w) = w^2$
- $J^{'}(w) = dJ(w)/dw$
- $J^{''}(w) = \dfrac{d^2J(w)}{dw^2} = \dfrac{d}{dw}\left(\dfrac{dJ}{dw}\right)$

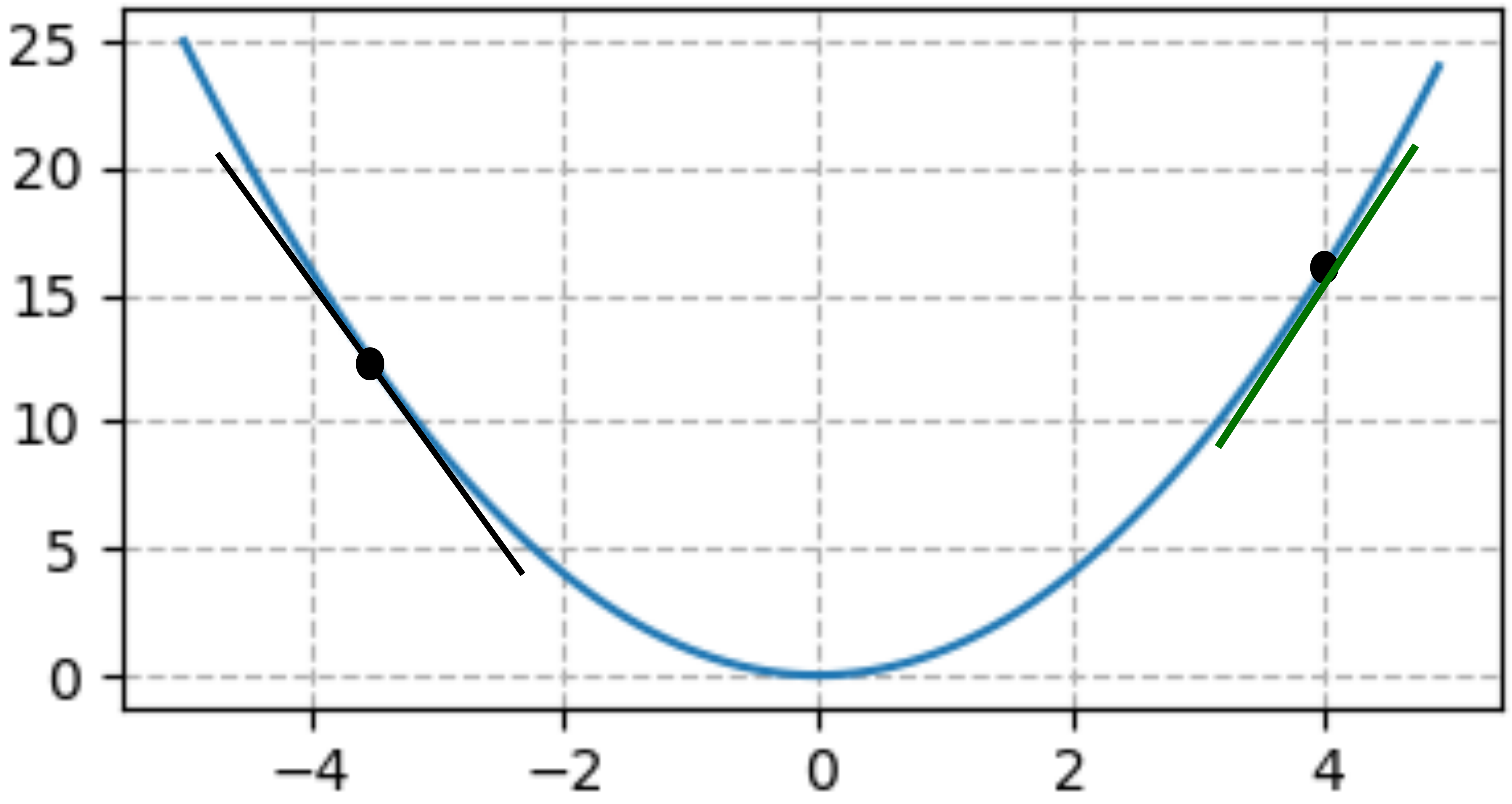


Optimality criteria - single variable

$J^{'}(w) = dJ(w)/dw$

- $J^{'}(w) = dJ(w)/dw$
    - slope / tangent at a point on the curve.
    - Continuous curve

$J(w)$

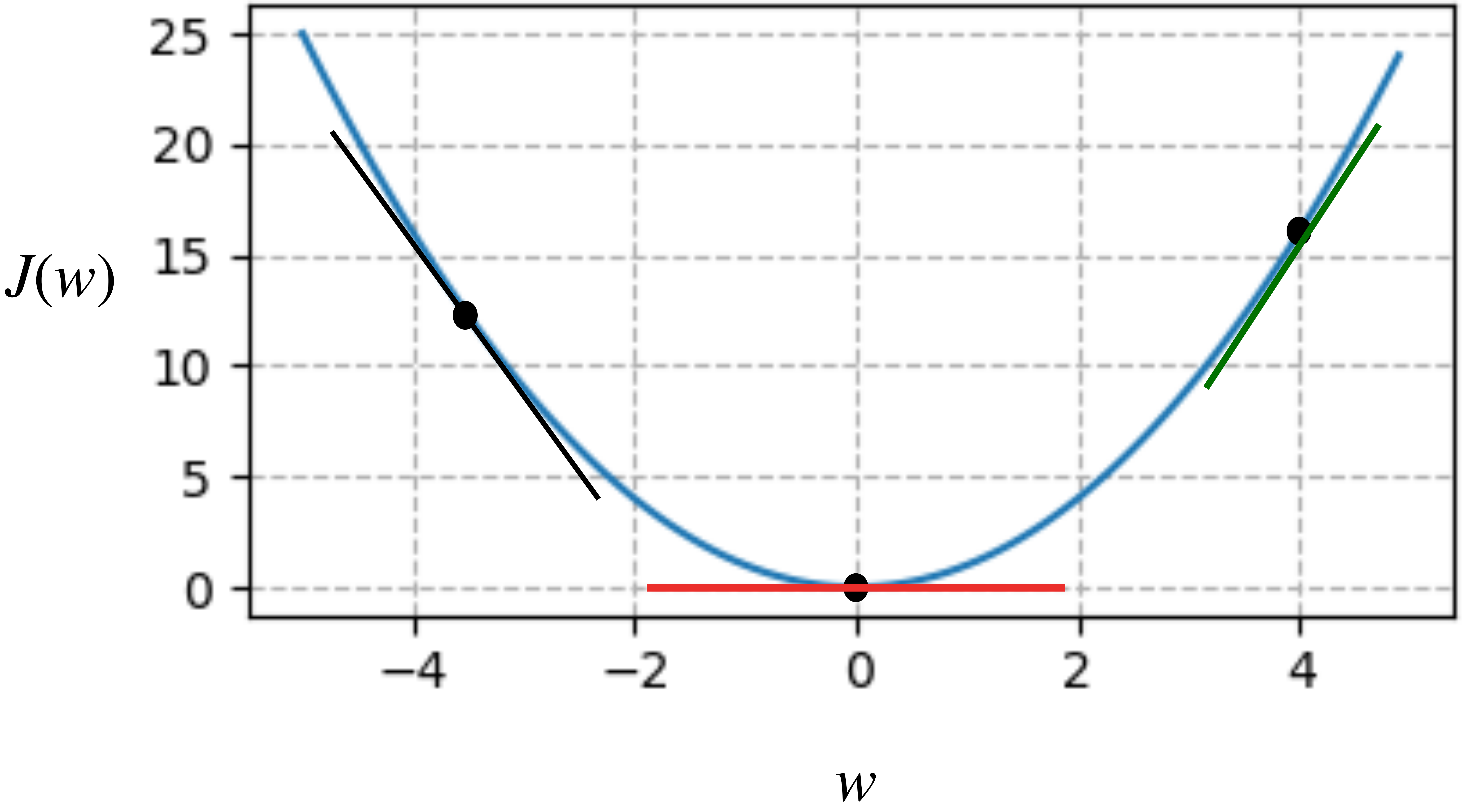


w

Optimality criteria - single variable

$J'(w) = dJ(w)/dw$

- $J'(w) = dJ(w)/dw$
  - slope / tangent at a point on the curve.
- At minimum function value, $J'(w) = 0$
- The corresponding $w = \overline{w}$

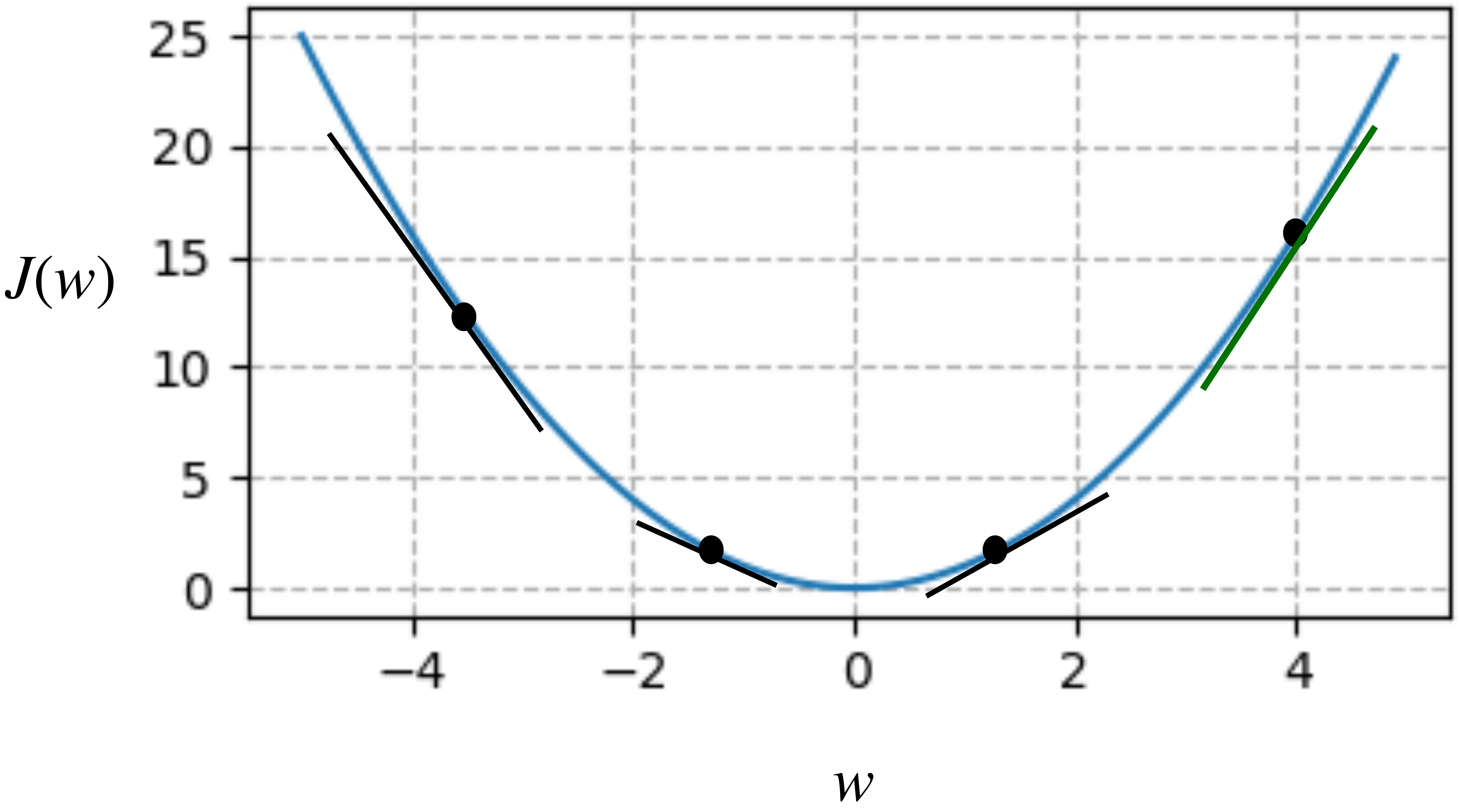


Ramanathan Muthuganapathy, Department of Engineering Design, IIT Madras

Optimality criteria - single variable

$J''(w) = \frac{d^2J(w)}{dw^2} = \frac{d}{dw}\left(\frac{dJ}{dw}\right)$

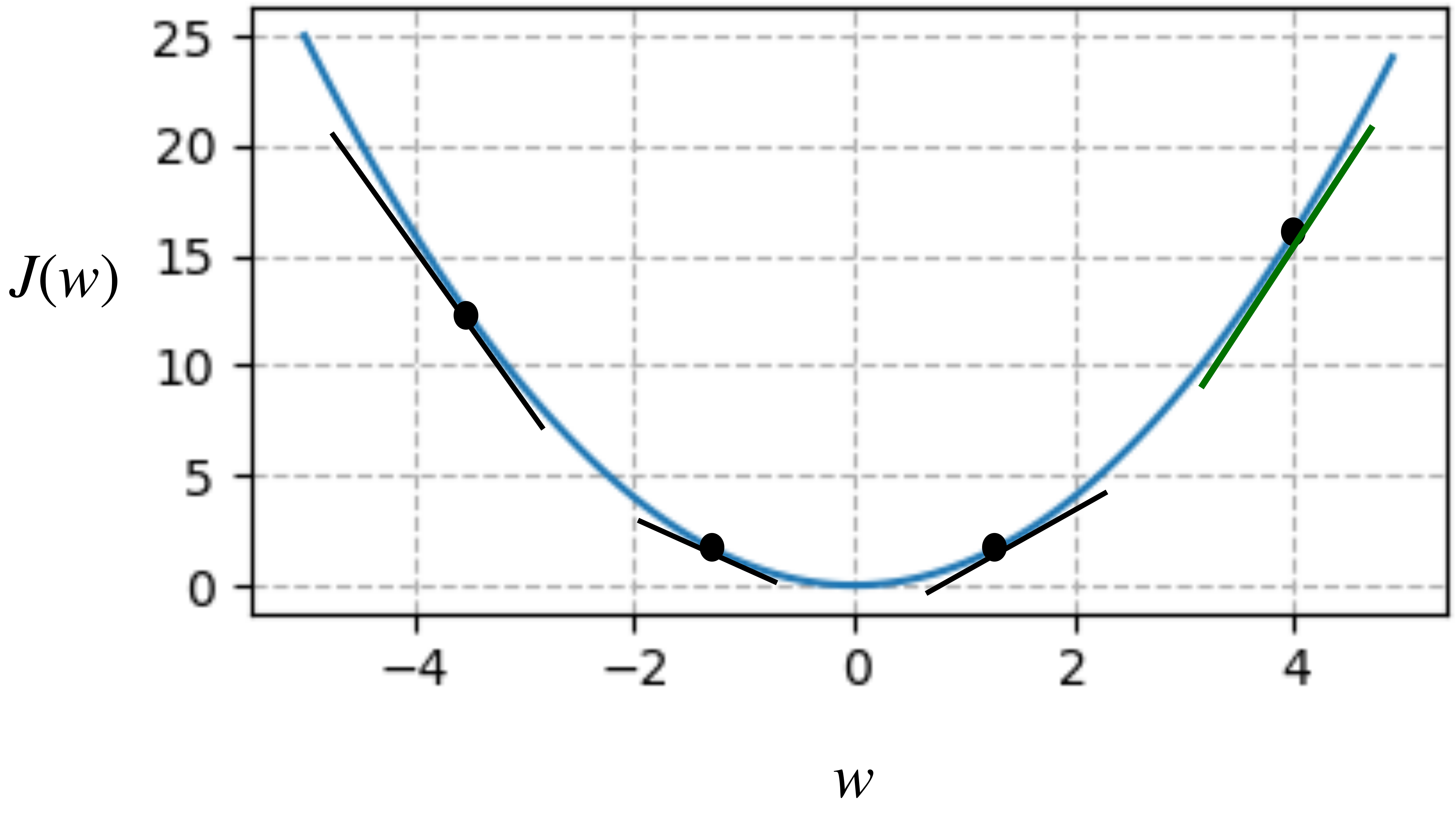
- $J''(w) = \frac{d^2J(w)}{dw^2} = \frac{d}{dw}\left(\frac{dJ}{dw}\right)$
  - rate of change of slope / tangent at a point on the curve.



Optimality criteria - single variable

$J^{''}(w) = \frac{d^2J(w)}{dw^2} = \frac{d}{dw}\left(\frac{dJ}{dw}\right)$

- $J^{''}(w) = \frac{d^2J(w)}{dw^2} = \frac{d}{dw}\left(\frac{dJ}{dw}\right)$
    - > 0 in the nbghd of minimum point.



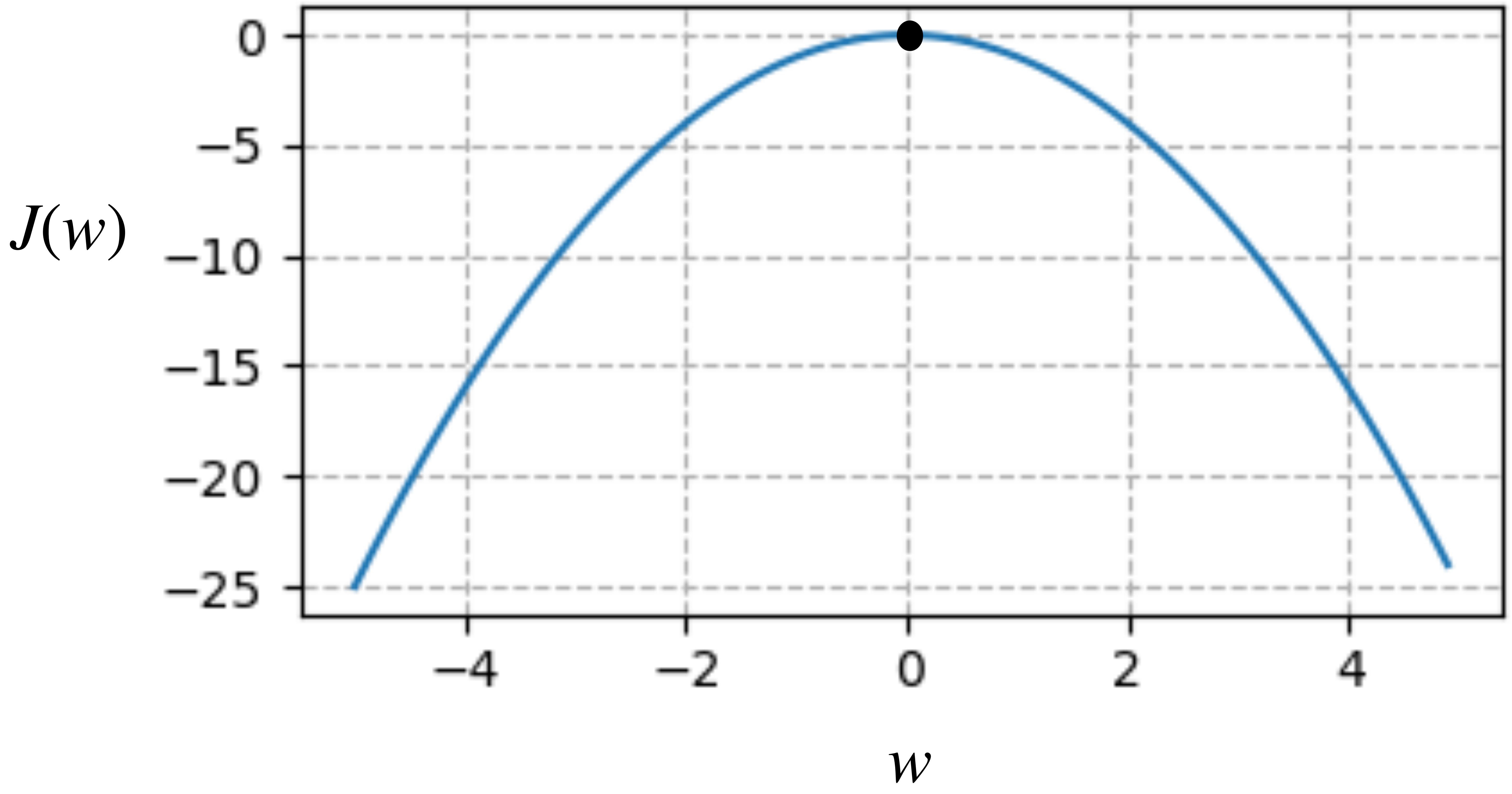
Critical Points

- Minimum
- Maximum
- Inflection

Optimality criteria - Maximum

$J(w) = -w^2$

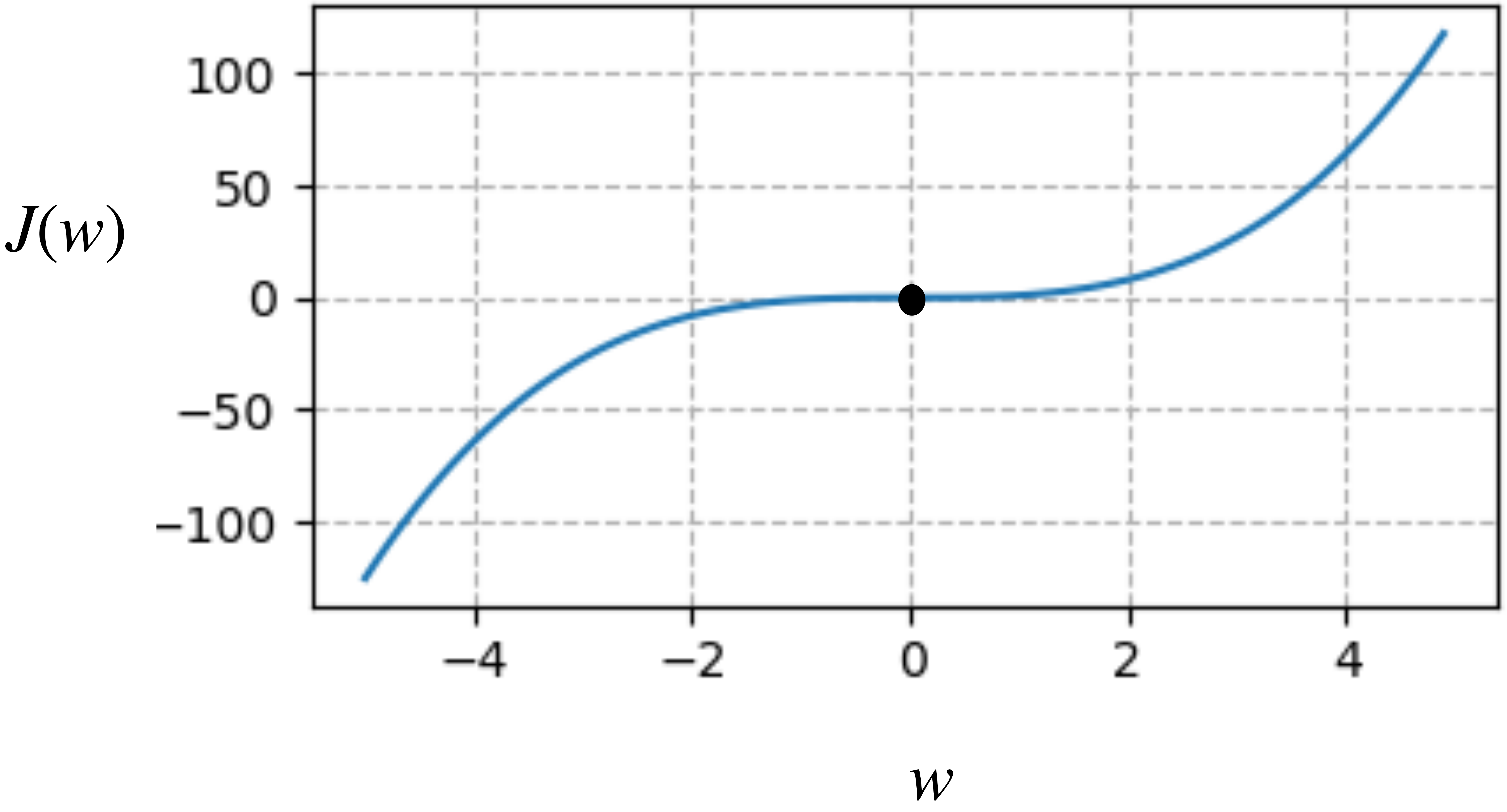
- $J(w) = -w^2$
- $J^{'}(w) = 0, at\, w = \overline{w},$
- At $\overline{w}, J^{''}(\overline{w}) < 0$



Optimality criteria - Inflection

$J(w) = w^3$

- $J(w) = w^3$
- $J^{'}(w) = 0, at\, w = \overline{w},$
- At $\overline{w}, J^{''}(\overline{w})$ is ?
- At $\overline{w}, J^{'''}(\overline{w})$ is ?

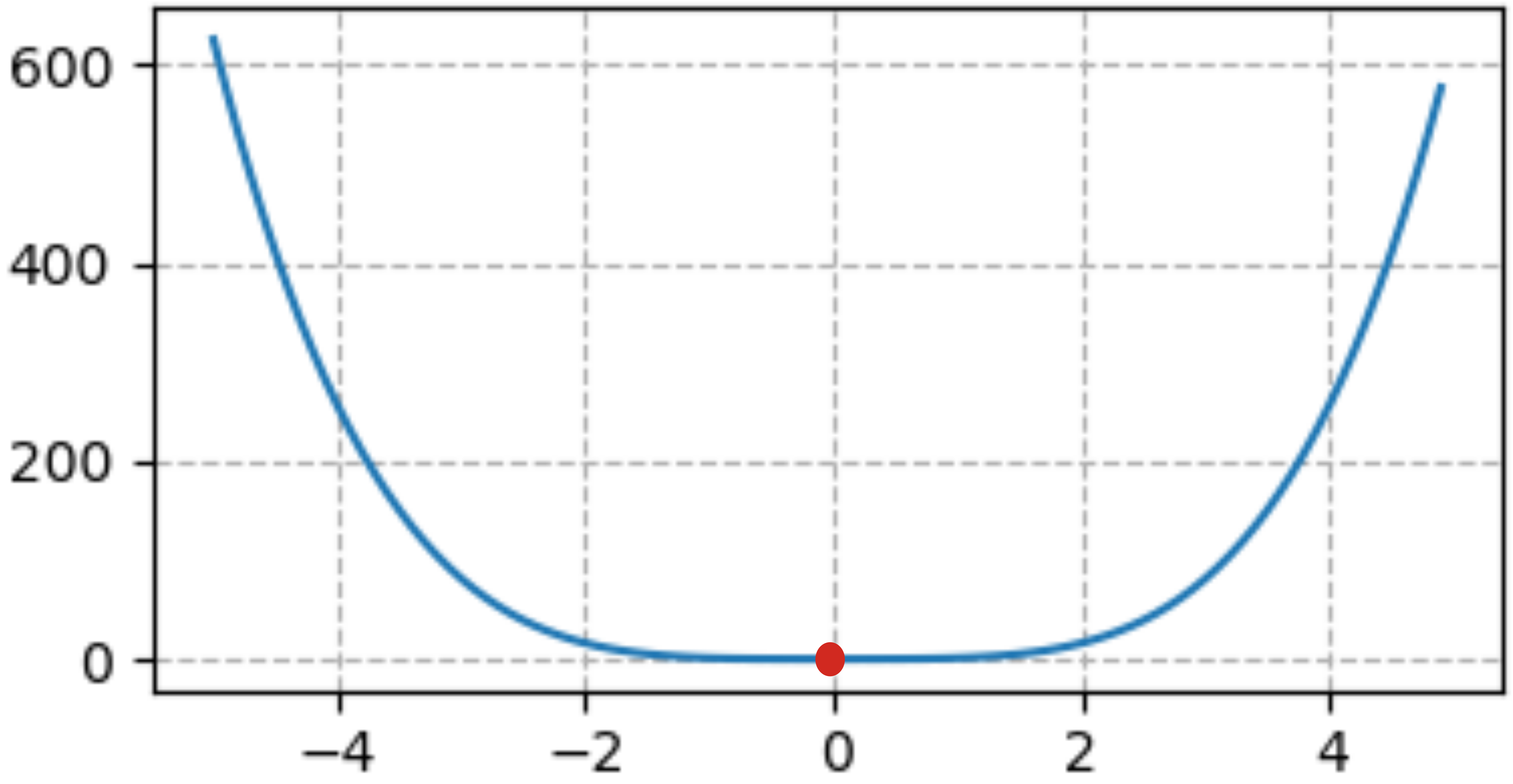


Optimality criteria ?

$J(w) = w^4$

- $J(w) = w^4$
- $J^{'}(w) = 0, at\, w = \overline{w},$
- At $\overline{w}, J^{''}(\overline{w})$ is ?

$J(w)$

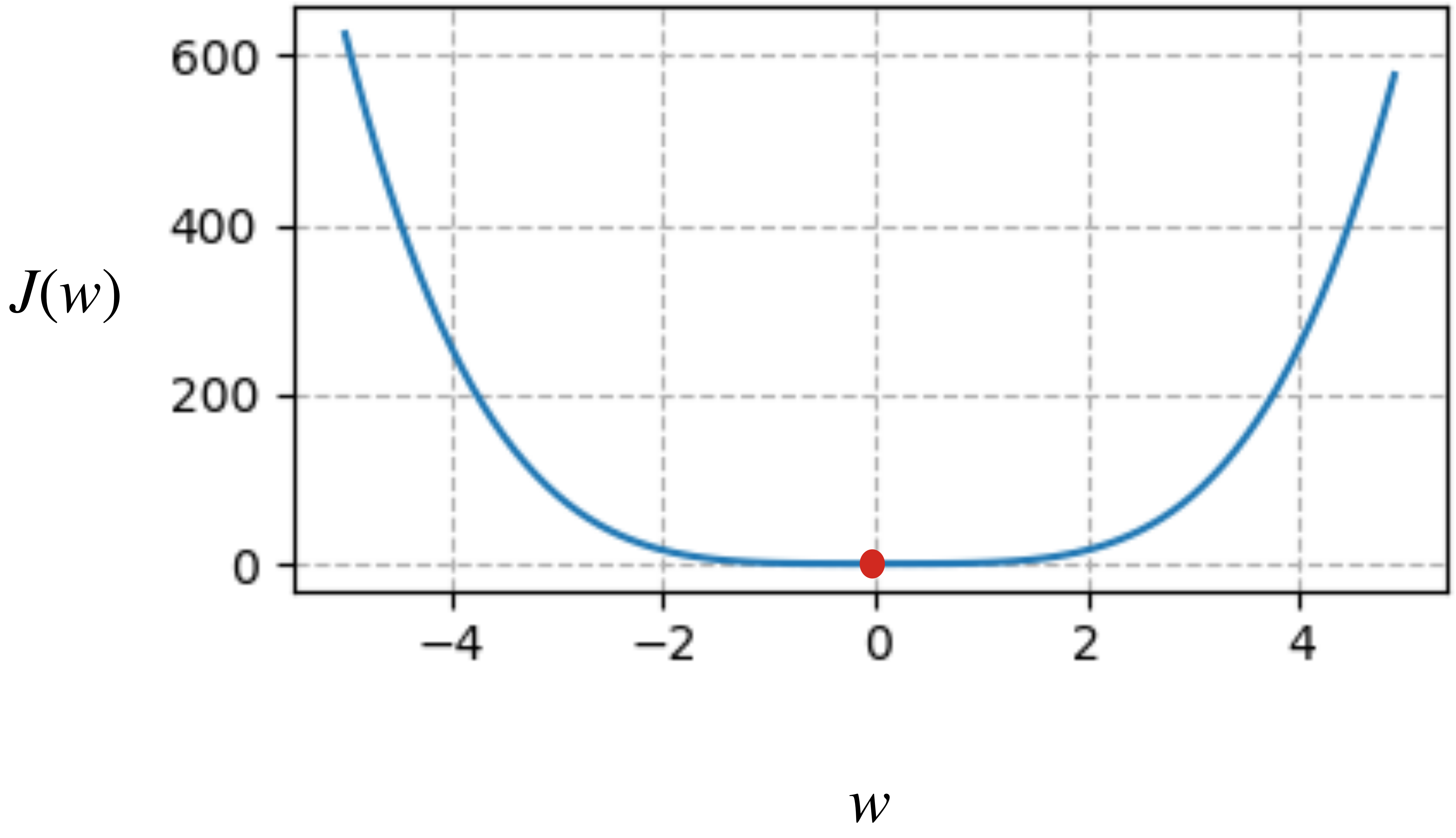


w

Optimality criteria ?

$J(w) = w^4$

- $J(w) = w^4$
- $J^{'}(w) = 0, at\, w = \overline{w},$
- At $\overline{w}, J^{''}(\overline{w})$ is ?
- At $\overline{w}, J^{'''}(\overline{w})$ is ?
- At $\overline{w}, J^{''''}(\overline{w})$ is ?



Optimality criteria - Generalization

From Kalyanmoy Deb

- Suppose at point $\overline{w}$ the first derivative is zero and the first nonzero higher order derivative is denoted by n; then
  - If n is odd, $\overline{w}$ an inflection point
  - If n is even, $\overline{w}$ is a local optimum.
    - (i) If the derivative is positive, $\overline{w}$ is a local minimum.
    - (ii) If the derivative is negative, $\overline{w}$ is a local maximum.

Optimality criteria - How to use

- Given a point on the curve, whether it belongs to any of the optimal ones
- The more pressing one - Given a function $J(w)$, how to find the optimal points (in our case, mostly ‘min’)

Methods to find local minimum

Unimodal functions

- Given a point on the curve, whether it belongs to any of the optimal ones
- The more pressing one - Given a function $J(w)$, how to find the optimal points (in our case, mostly 'min')

Methods to find local minimum

Unimodal functions

- First, a crude approach to find bounds
- Use a more sophisticated method to find the min
- In general, any method takes the following pattern:
    - Identify initial guess and and their function values
    - Make appropriate changes in the next values for w's
    - Continue the procedure till the termination is reached.

Methods (iterative) to find local minimum

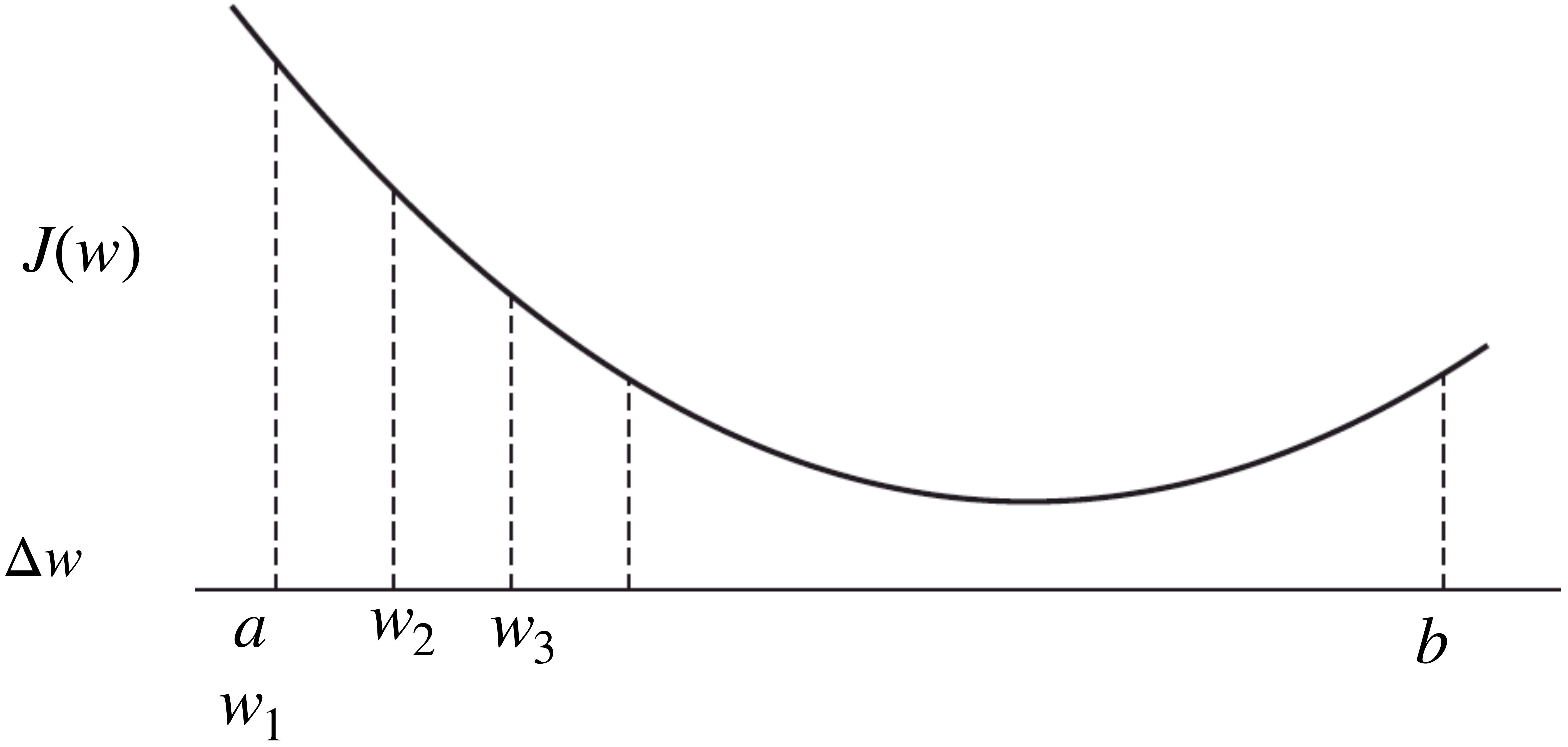
Unimodal functions

- Bracketing methods
    - Exhaustive search
    - Bounding phase
- Region elimination approaches
    - Interval halving
    - Fibonacci search
    - Golden section search
- Gradient-based ones
    - Newton-Raphson
    - Bisection
    - Secant

Bracketing - Exhaustive search method

Unimodal functions

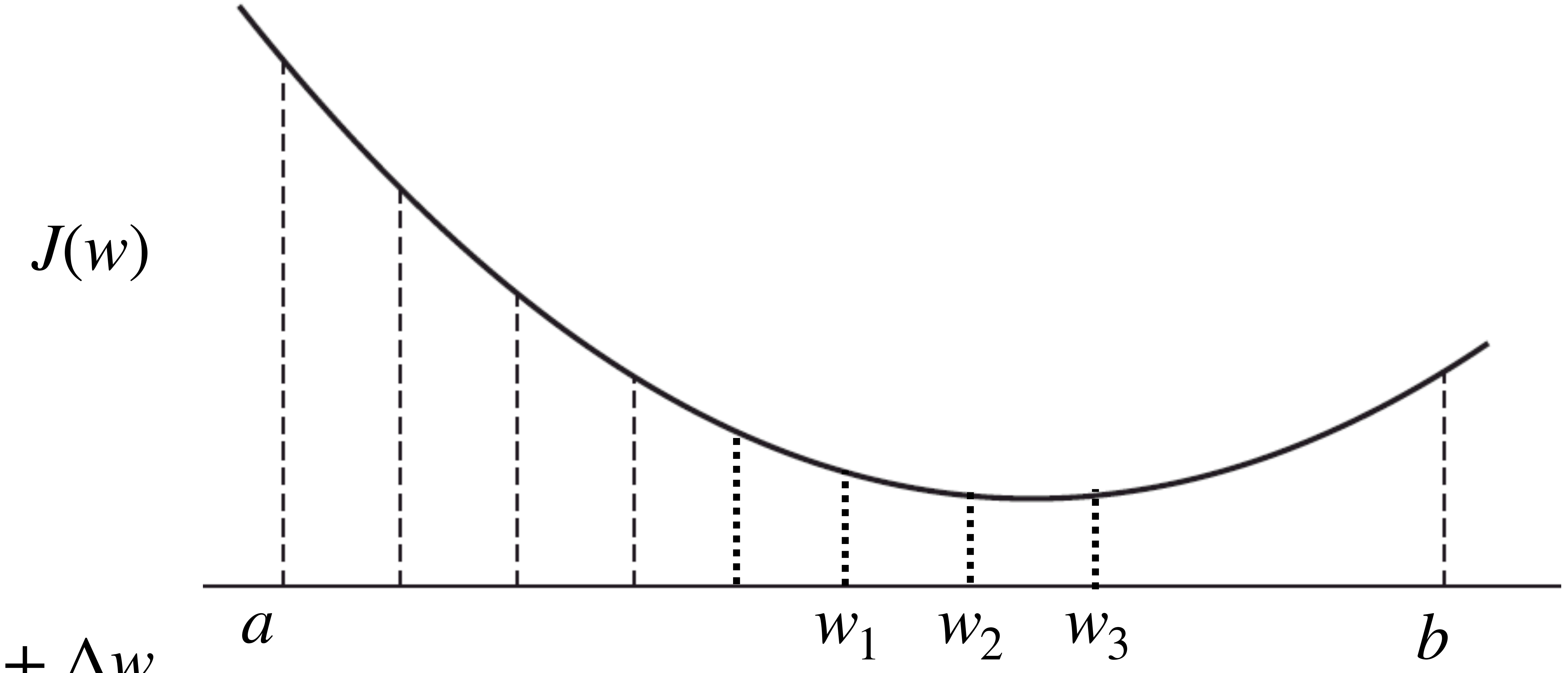
- Let n be the number of intermediate points .
- Step 1:
  - $\Delta w = (b - a)/n$
  - $w_1 = a, w_2 = w_1 + \Delta w, w_3 = w_2 + \Delta w$



Bracketing - Exhaustive search method

Unimodal functions

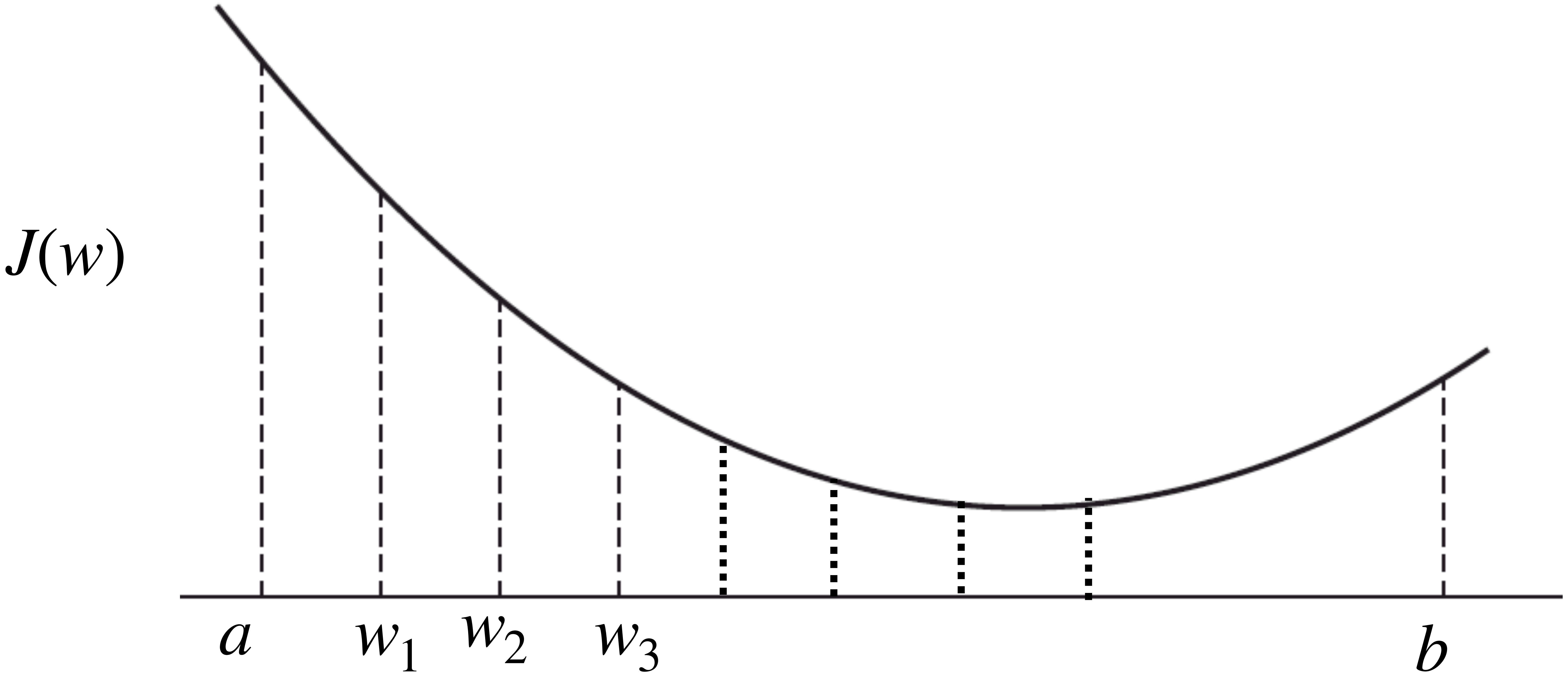
- Step 2:
  - If $J(w_1) \geq J(w_2) \leq J(w_3)$
    - then min lies between (w_1, w_3)
  - Else
    - $w_1 = w_2, w_2 = w_3, w_3 = w_2 + \Delta w$
    - Go to Step 3



Bracketing - Exhaustive search method

Unimodal functions

- $w_1 = w_2, w_2 = w_3, w_3 = w_2 + \Delta w$
- Step 3:
  - Is $w_3 \leq b$, the go to Step 2
  - Otherwise, no min exists between (a, b).
  - Min could be one of the bdry points.



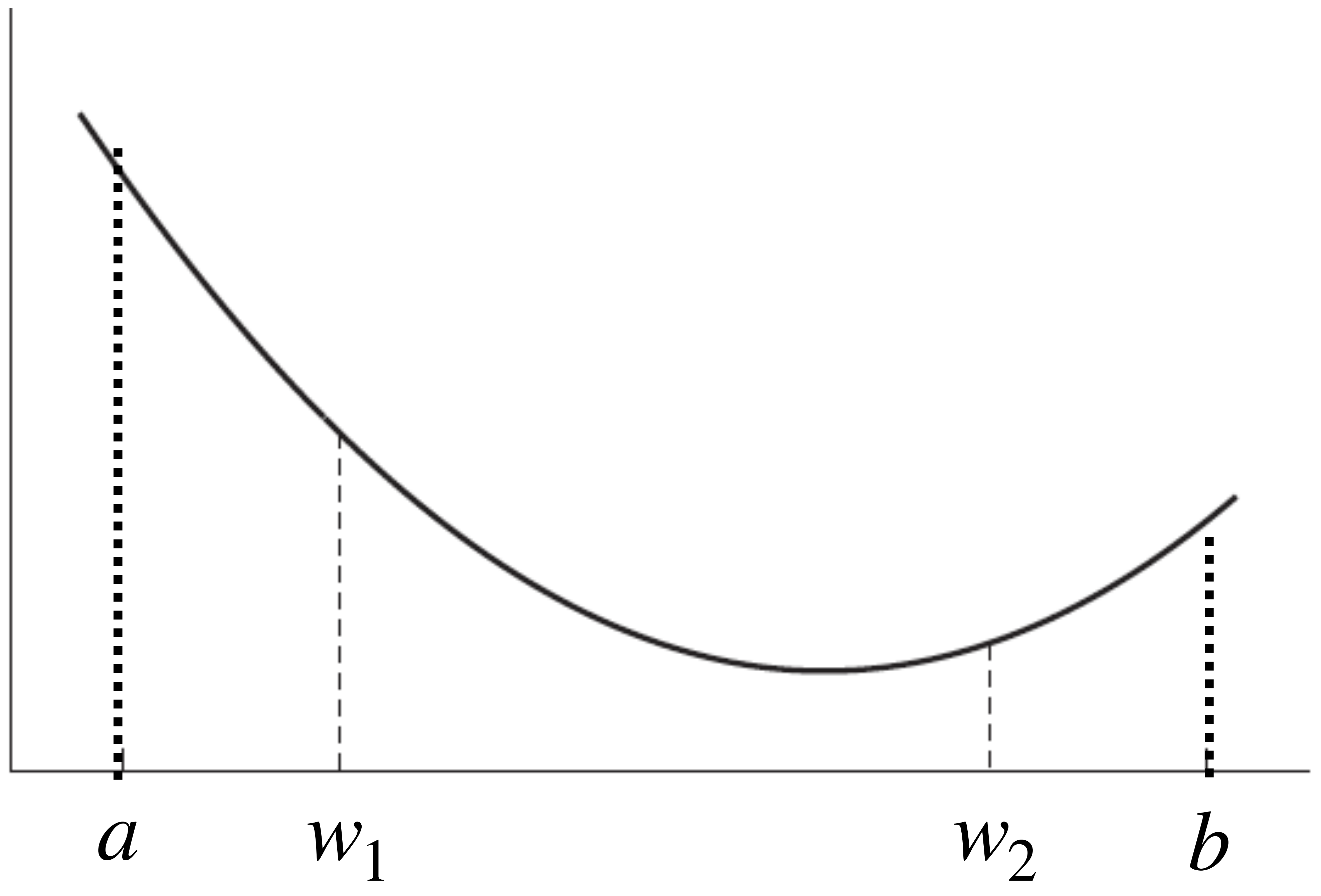
Ramanathan Muthuganapathy, Department of Engineering Design, IIT Madras

Region elimination method

Overall idea

- $w_1 \leq w_2, if\, J(w_1) \geq J(w_2)$
- Region (a, w_1) can be eliminated

$J(w)$

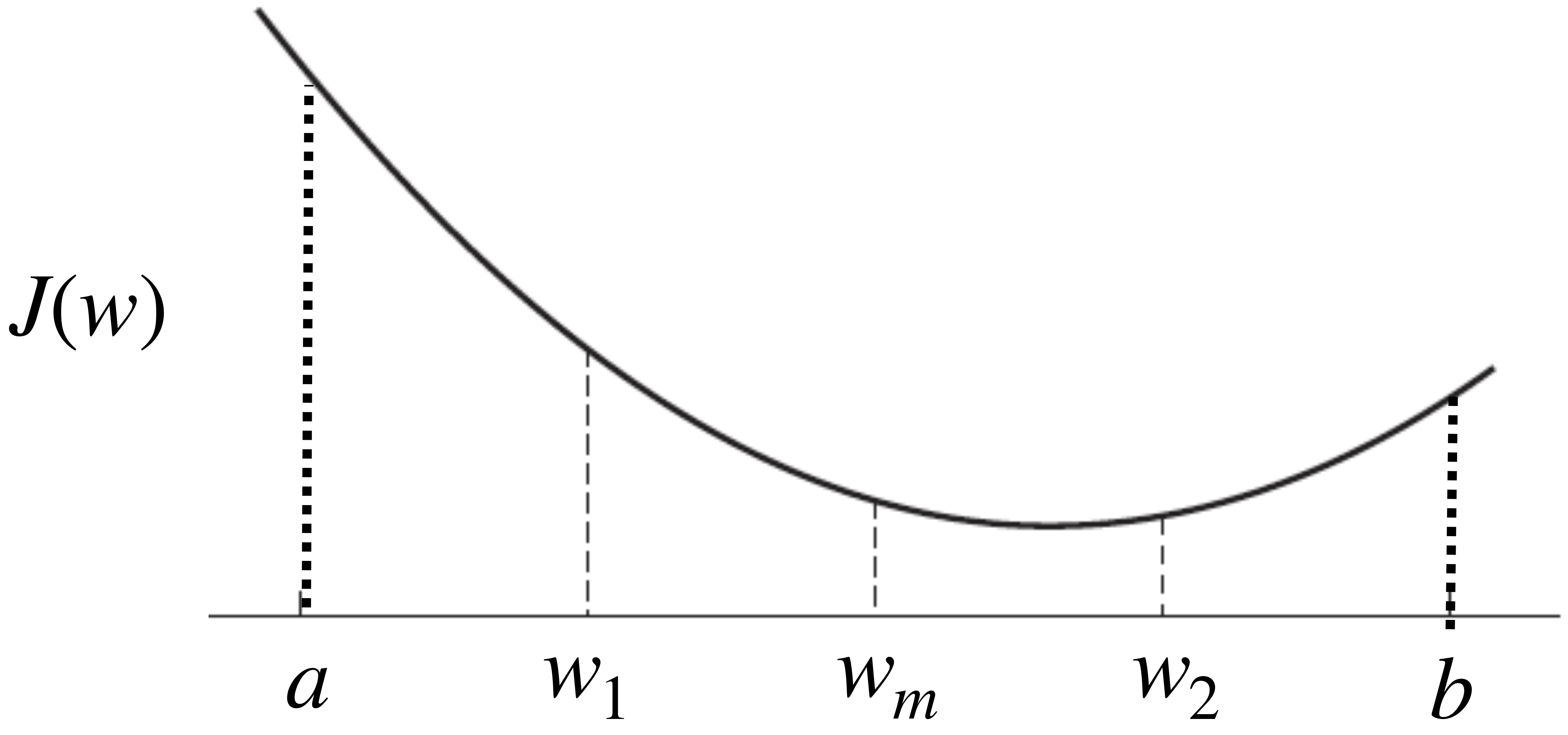


w_3

Region elimination method

Steps

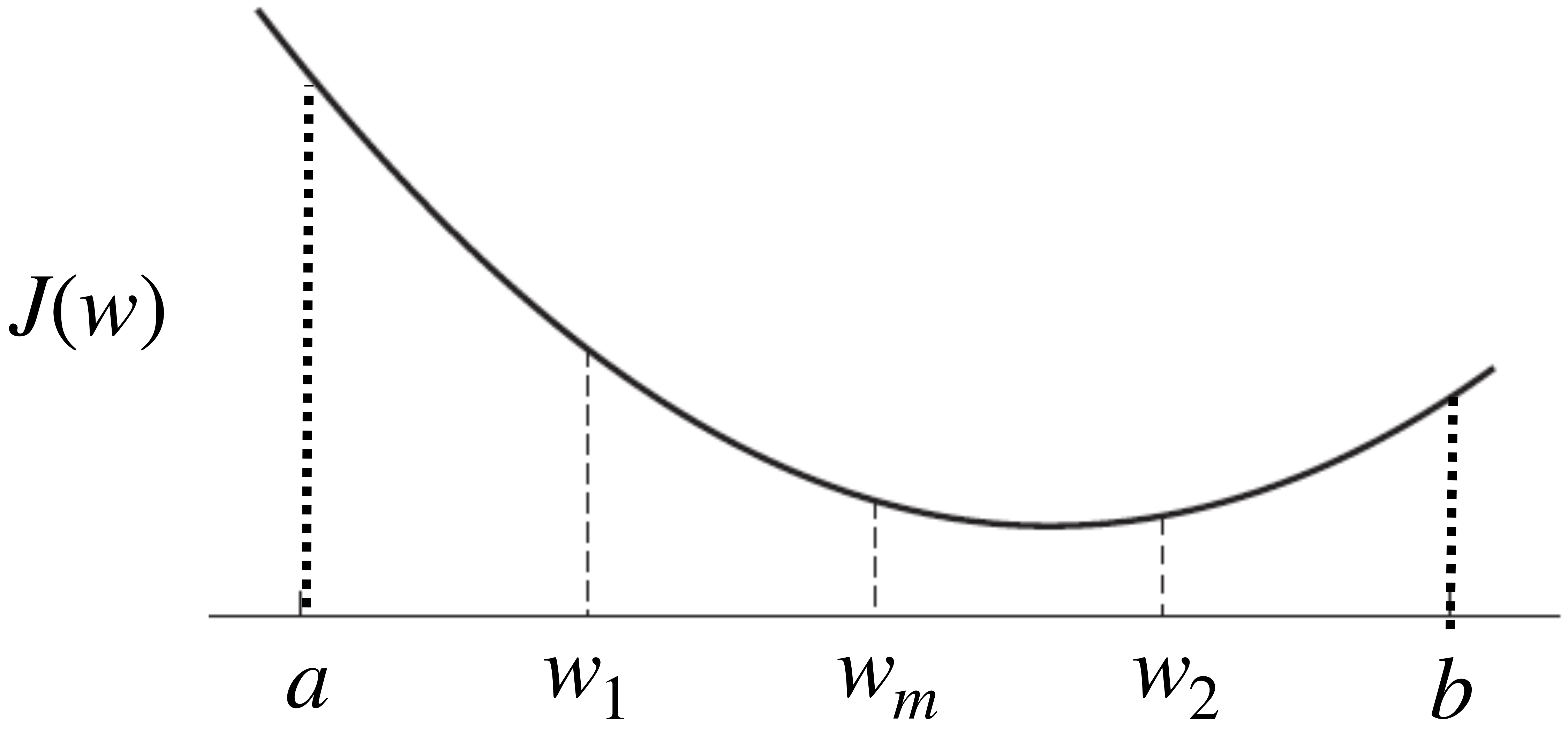
- Step 1
  - Choose $a, b, \epsilon, w_m = (a + b)/2, L = (b - a)$
  - Compute $J(w_m)$



Region elimination method

Steps

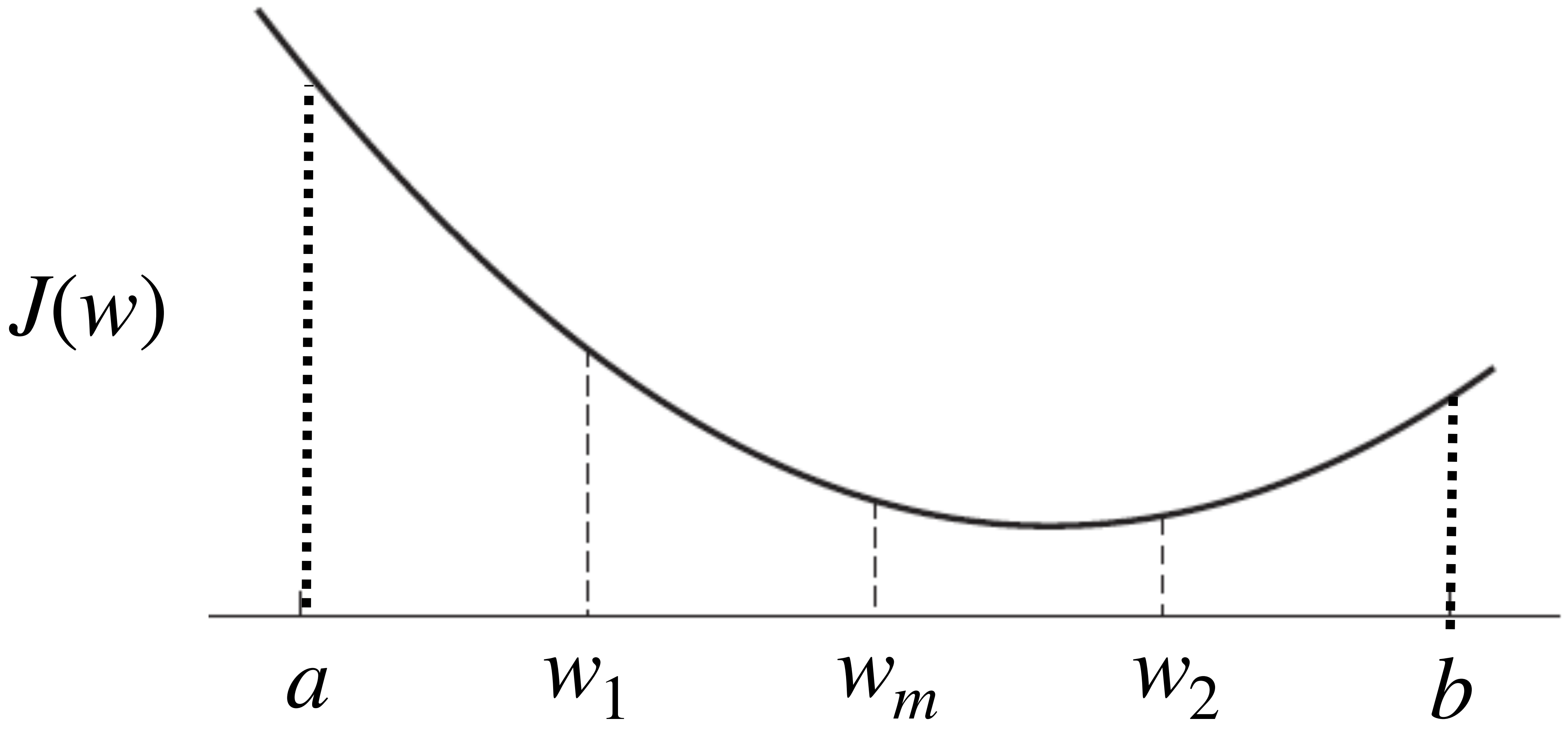
- Step 2
  - Set
    $w_1 = a + L/4, w_2 = b - L/4$
  - Compute $J(w_1), J(w_2)$



Region elimination method

Steps

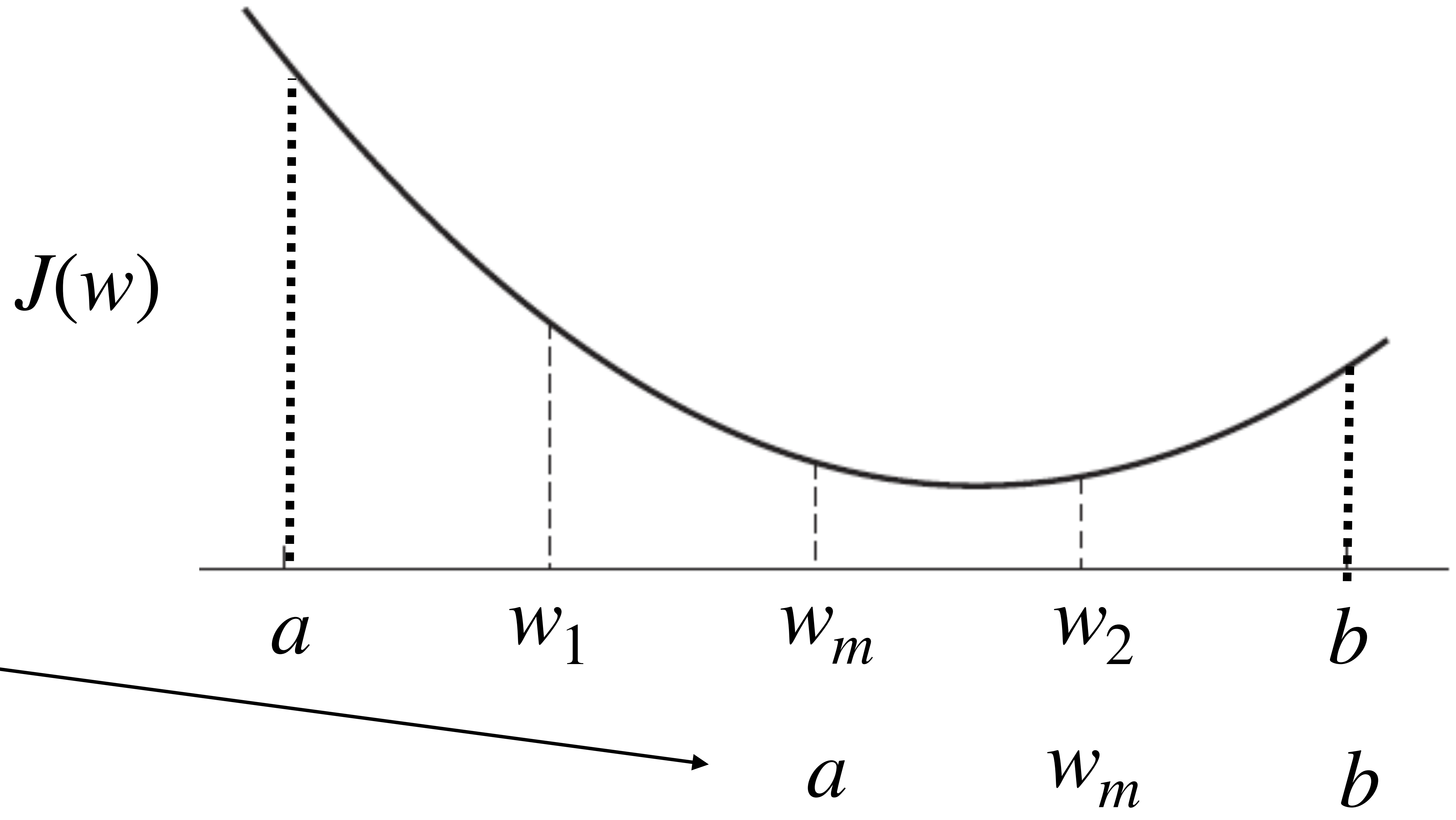
- Step 3
  - If $J(w_1) < J(w_m)$
    - set $b = w_m, w_m = w_1$, go to Step 5
  - Else
    - go to Step 4



Region elimination method

Steps

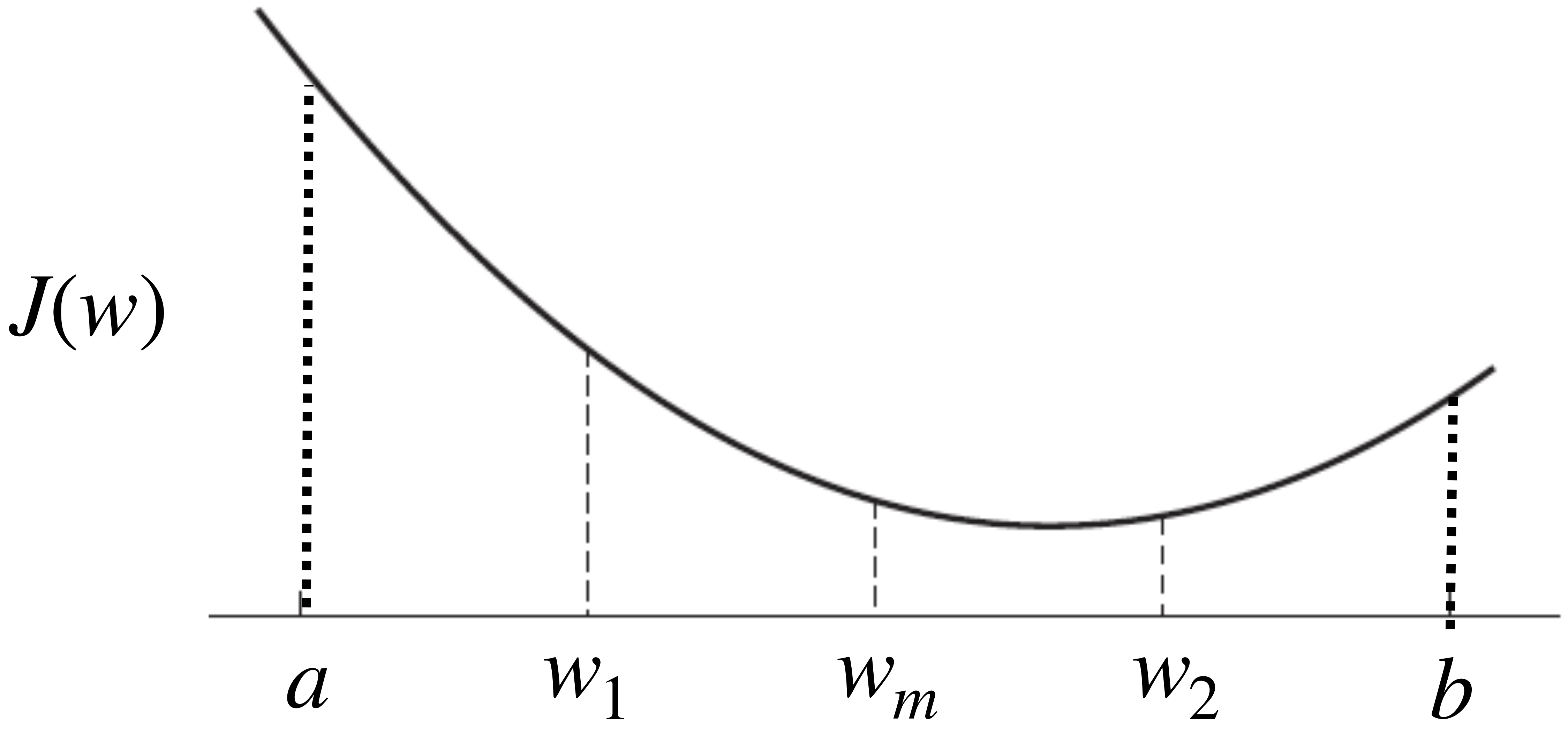
- Step 4
  - If $J(w_2) < J(w_m)$
    - set $a = w_m, w_m = w_2$; go to Step 5
  - Else
    - $a = w_1, b = w_2$; go to Step 5



Region elimination method

Steps

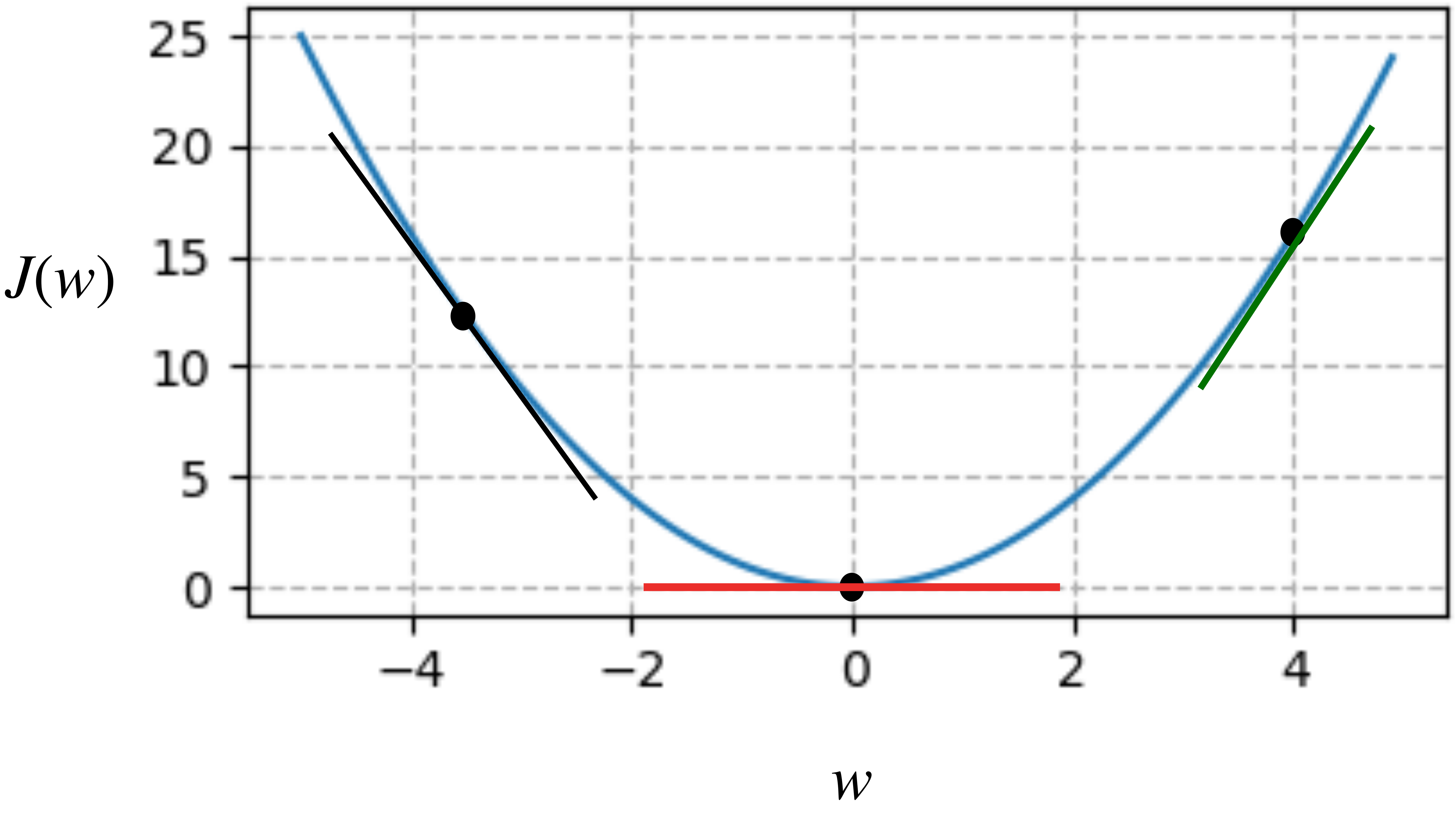
- Step 5
  - Calculate $L = b - a$
  - If $|L| < \epsilon$
    - Terminate
  - Else
    - go to Step 2



Gradient-based approaches

$J^{'}(w) = dJ(w)/dw$

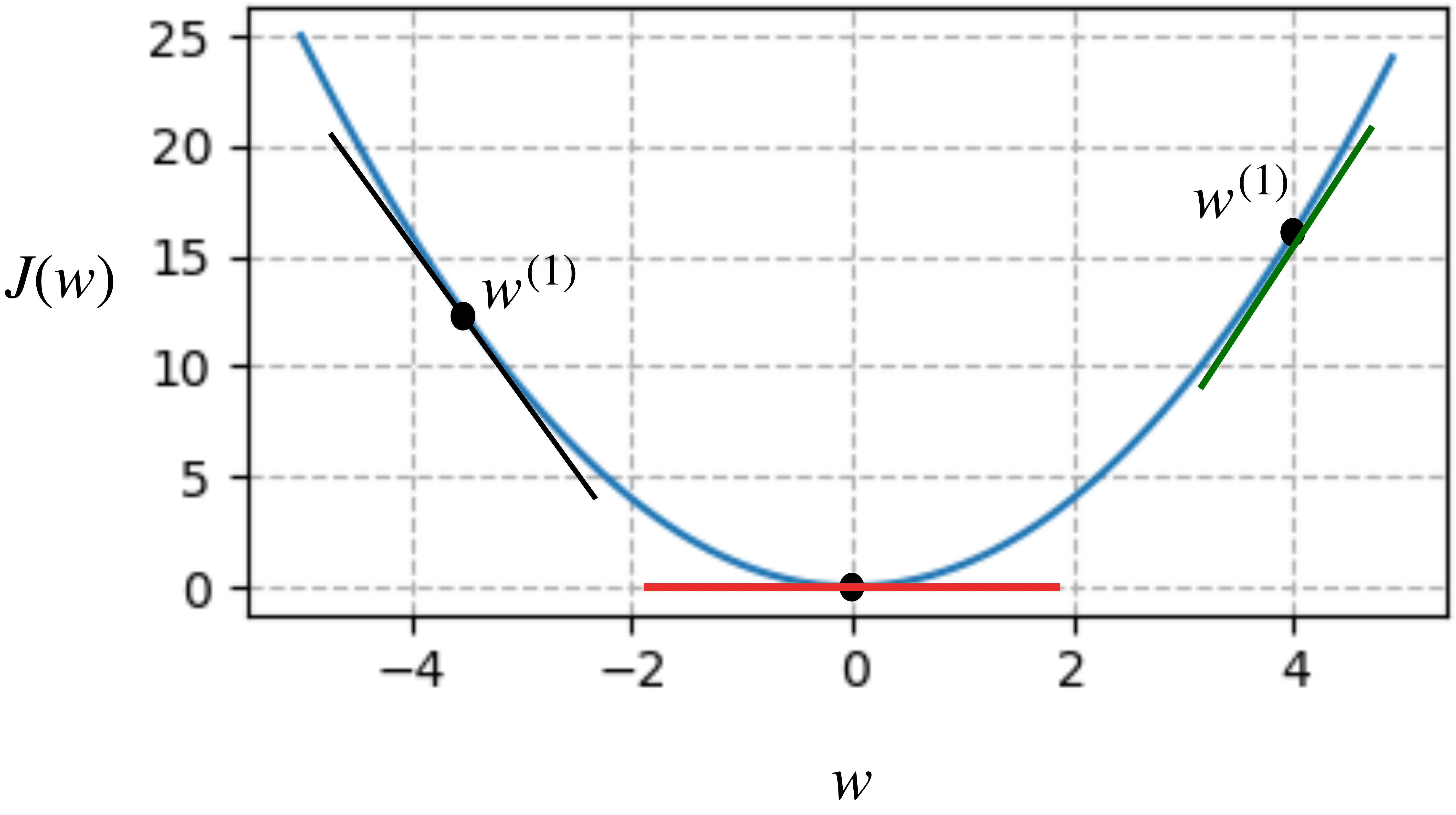
- uses $J^{'}(w) = dJ(w)/dw$ and other higher order derivatives.
- Exact / Numerical approach for derivative
- Mim —-> point where $J^{'}(w) \approx 0$



Newton-Raphson

$J^{'}(w) = dJ(w)/dw$

- uses $w^{(k)}$ to compute $w^{(k+1)}$
- Using Taylor's appox,
- $w^{(k+1)} = w^{(k)} - \dfrac{J^{'}(w)}{J^{''}(w)}$
- k = 1 to N (num of iterations)
- Initial guess $w^{(1)}$ is needed.



Newton-Raphson

Steps

- Step 1: Choose $w^{(1)}$, ϵ, set $k = 1$. Compute $J^{'}(w^{(1)})$
- Step 2: Compute $J^{''}(w^{(k)})$
- Step 3: Calculate $w^{(k+1)} = w^{(k)} - \dfrac{J^{'}(w)}{J^{''}(w)}$. Compute $J^{'}(w^{(k+1)})$
- Step 4: If $|J^{'}(w^{(k+1)}) < \epsilon|$, Terminate. Else set $k = k + 1$ go to Step 2