Optimization

Last episode

Mei-Chen Yeh

Numerical Methods (Fall 2022)

14 programming practices!

Solving equations

- Nested multiplication
- Bisection / Fixed-point iteration
- Newton / Secant's method

Interpolating data

- Lagrange interpolation
- Newton's divided difference

Optimization

- Gradient descent

Solving systems

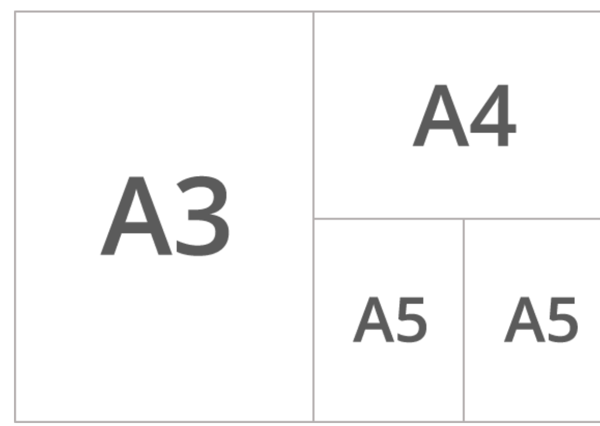
- Gaussian elimination
- Partial pivoting
- Jacobi / Gauss-Seidel / SOR
- Multivariate Newton

Least squares

- Normal equations
- QR factorization
- Gauss-Newton method

Announcements

- We will have the final next Friday (Dec. 23), starting at 14:20.
  - Programming problems
  - Written problems
- A cheat sheet (A5) is allowed.



Optimization

Some slides are modified from the lecture slides of Nicholas Ruozzi, University of Texas at Dallas

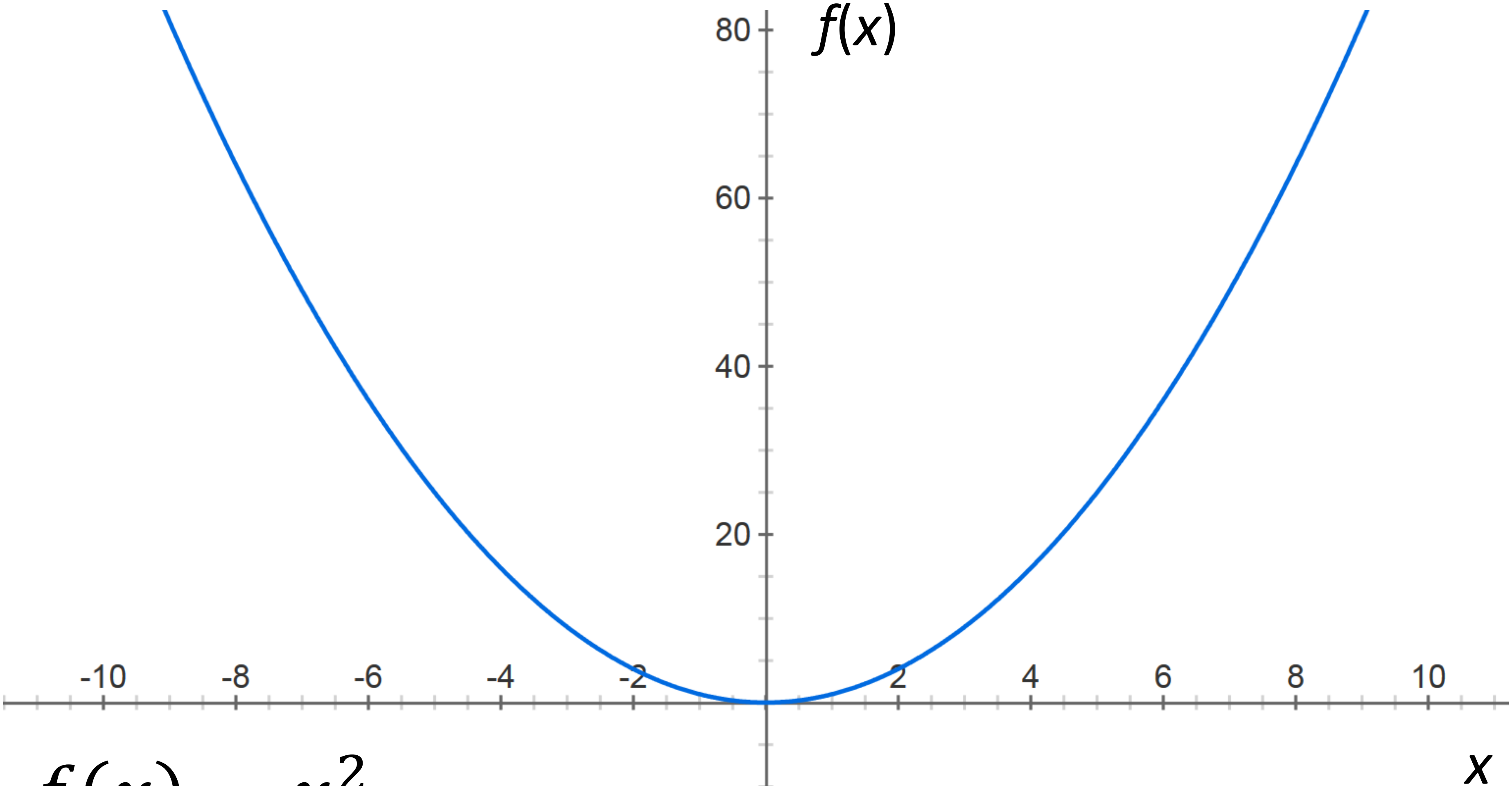
Optimization

- Finds the values of n-variables $(x_1, x_2, x_3, \ldots, x_n)$ that minimize (or maximize) some objective function $f(\underline{x})$

$$\underline{x}^* = \operatorname{argmin} f(\underline{x})$$

- Assumptions
  - We have a single objective function.
  - No constraint is applied.

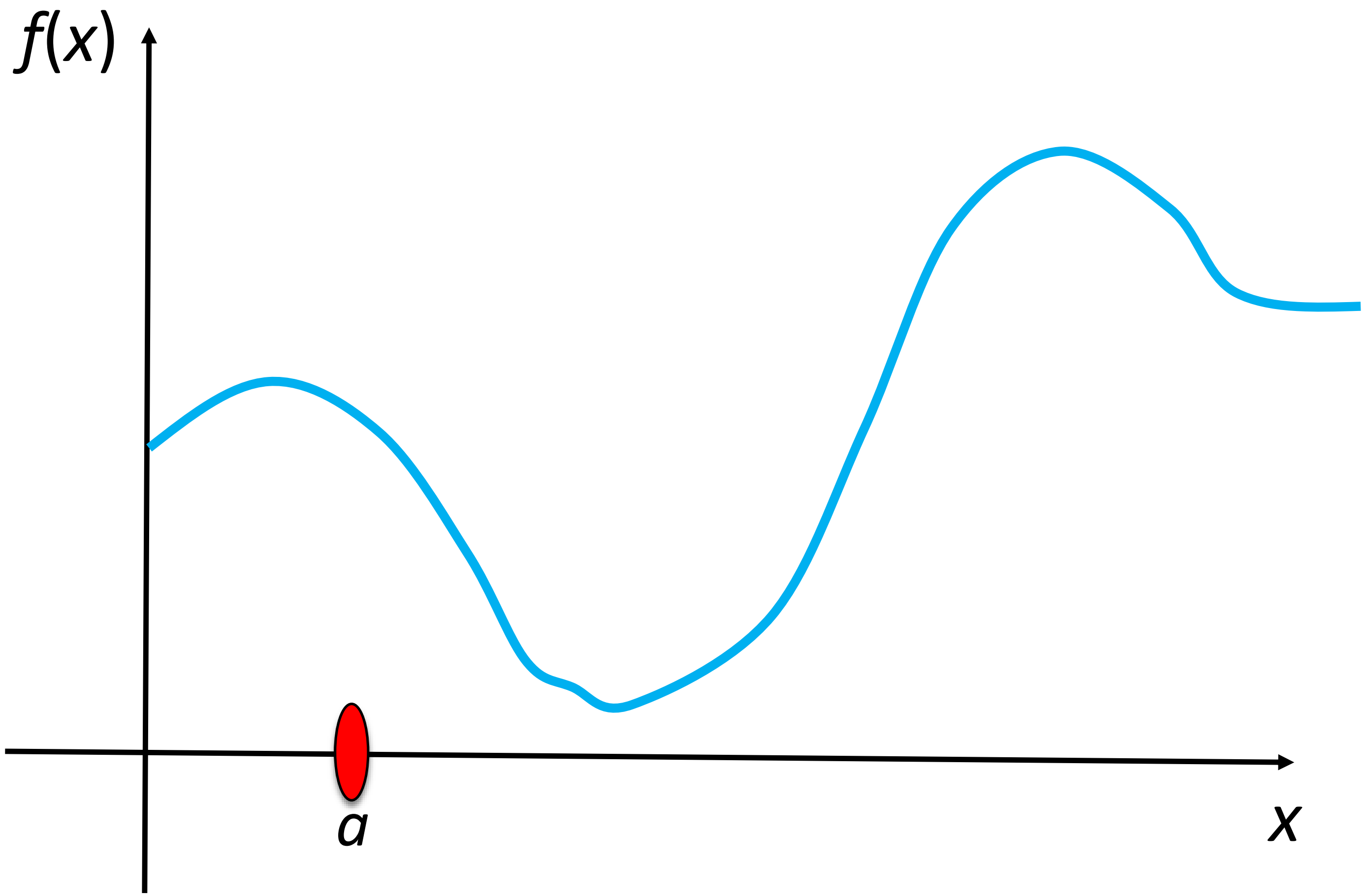
Example (one variable)



$$f(x) = x^2$$

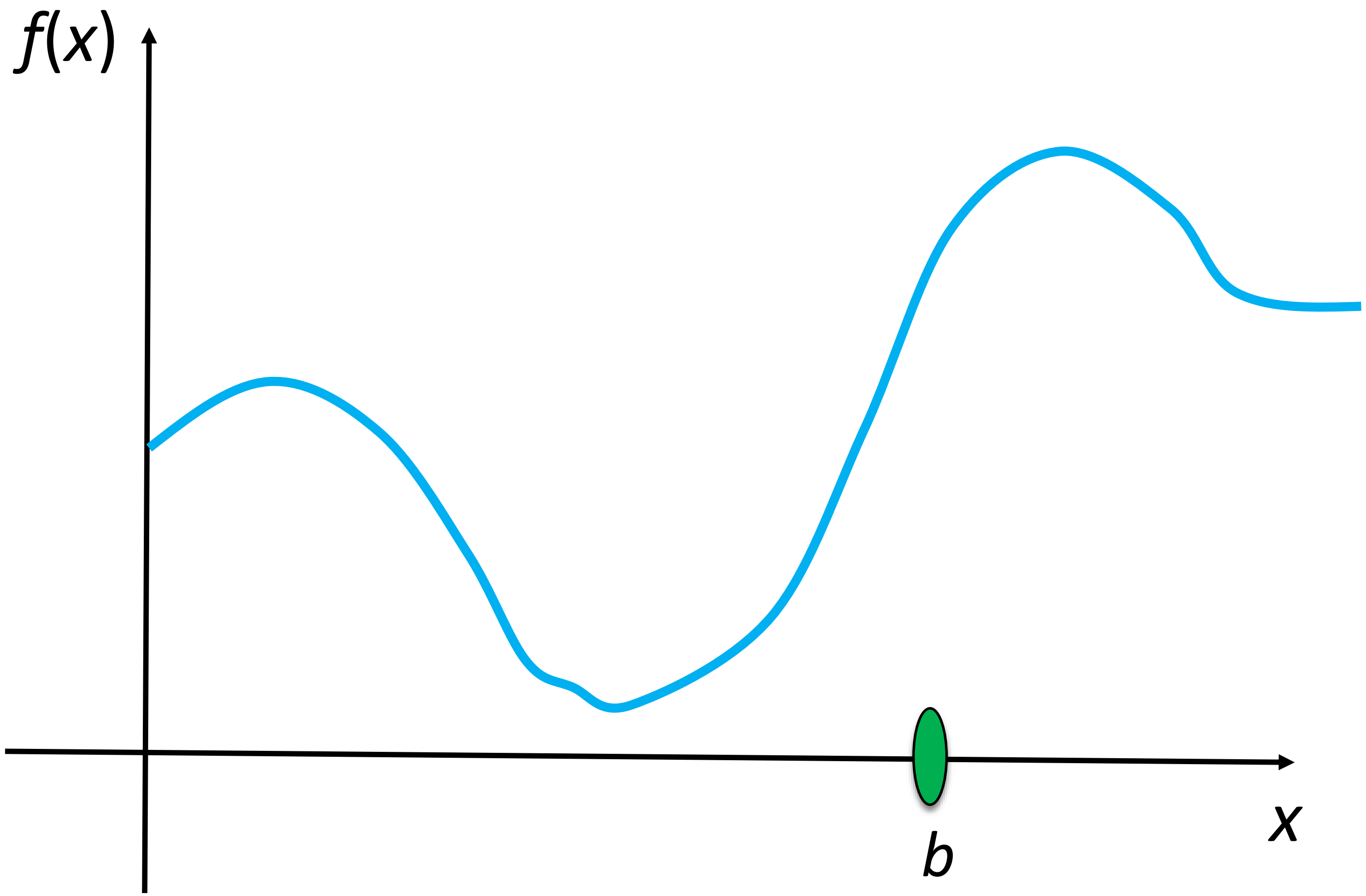
Gradient descent

How do I adjust a to lower $f(x)$?



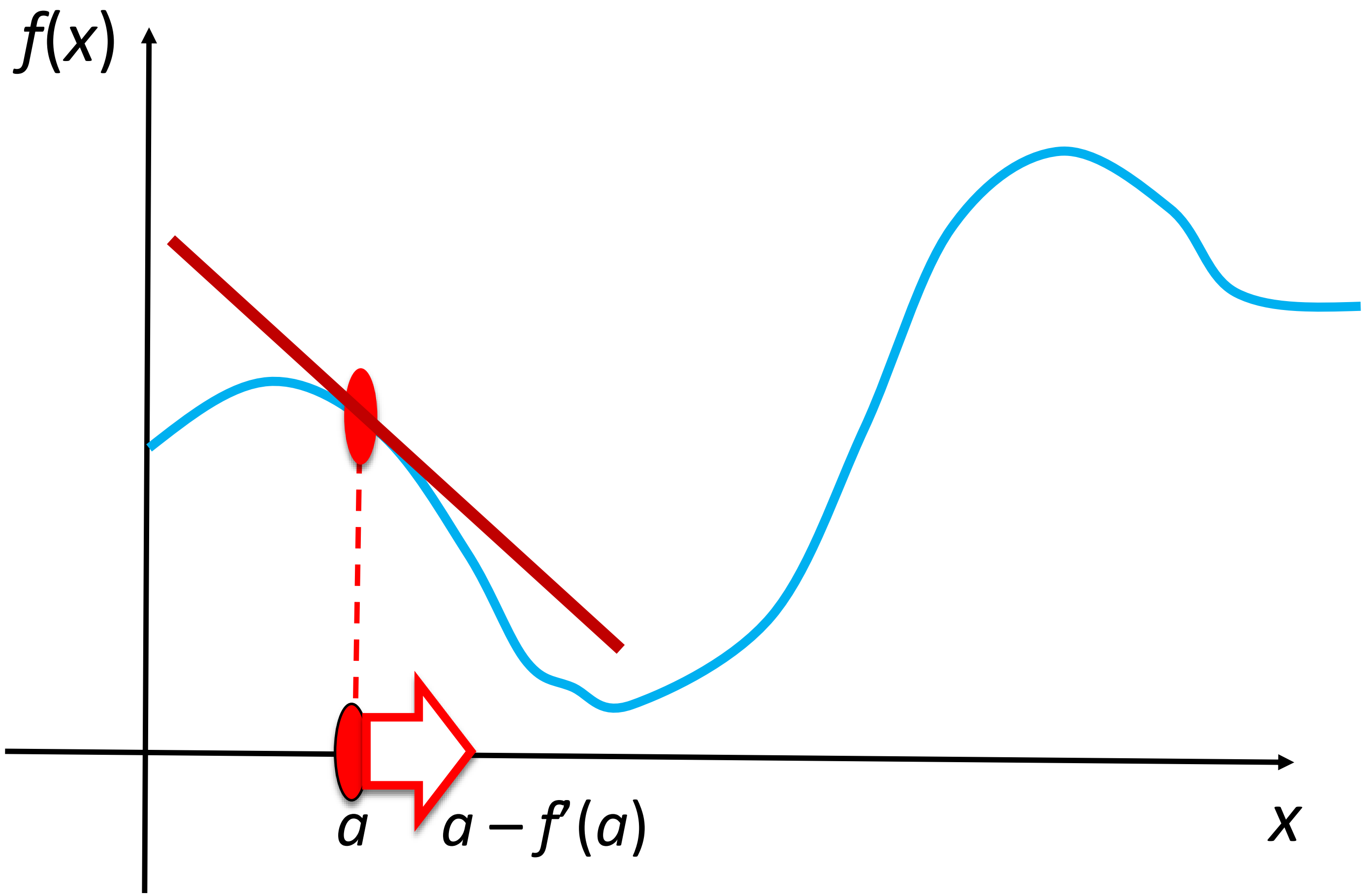
Gradient descent

How do I adjust *b* to lower *f*(*x*)?



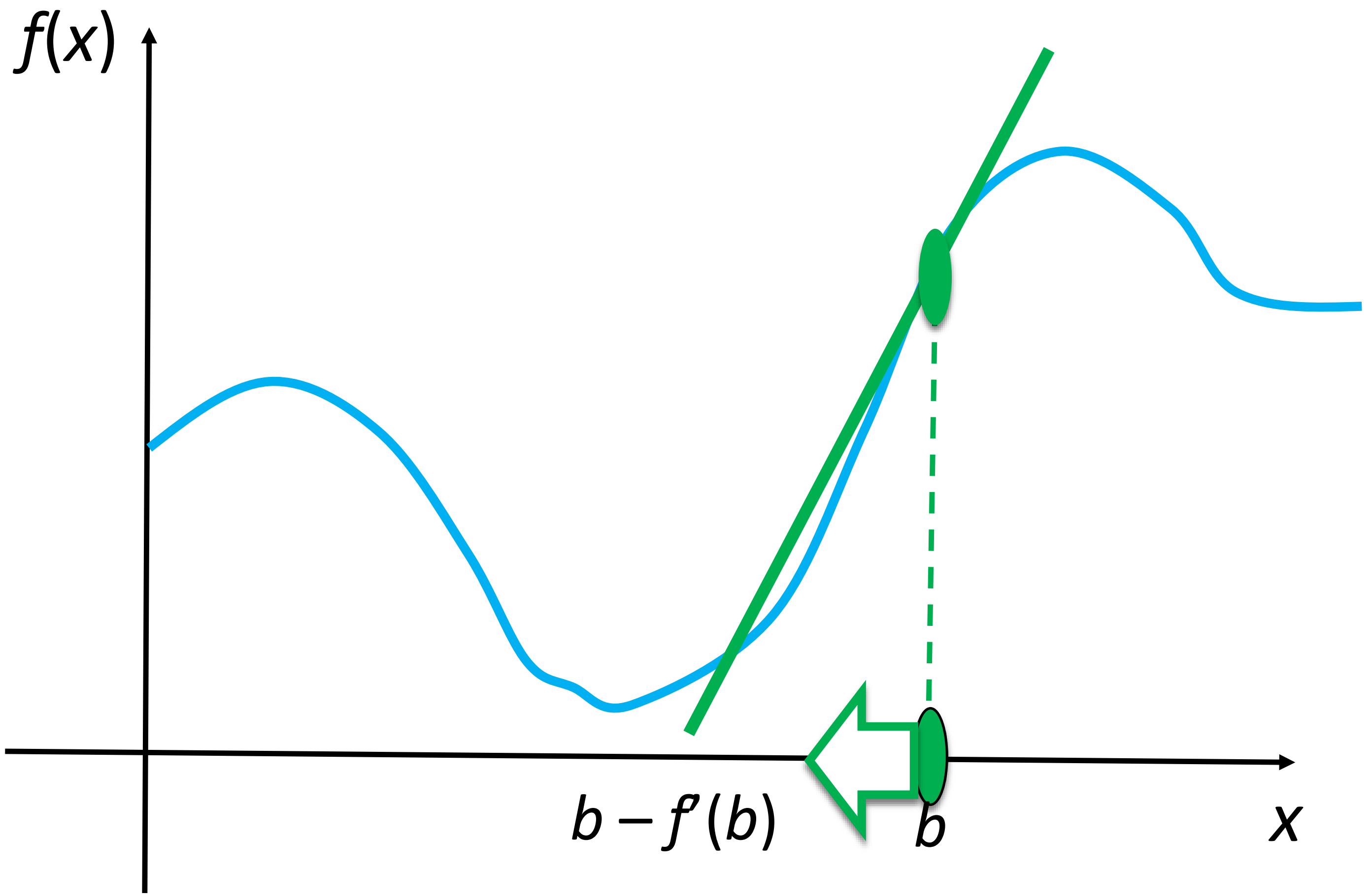
Gradient descent

How do I adjust *a* to lower *f*(*x*)?



Gradient descent

How do I adjust *b* to lower *f*(*x*)?



“Fine-tuning”

- $x_{\text{new}} = x_{\text{old}} - \eta\, f'(x_{\text{old}})$
- η: learning rate
- When do we stop?

Gradient descent

Algorithm:

- Pick an initial point x_0
- Iterate until convergence

$$x_{t+1} = x_t - \eta \nabla f(x_t)$$

where η is the step size (sometimes called learning rate)

When do we stop?

Gradient descent

Algorithm:

- Pick an initial point x_0
- Iterate until convergence

$$x_{t+1} = x_t - \eta \nabla f(x_t)$$

where η is the step size (sometimes called learning rate)

iterate until $\|\nabla f(x_t)\| \leq \epsilon$
for some $\epsilon > 0$

Multivariate

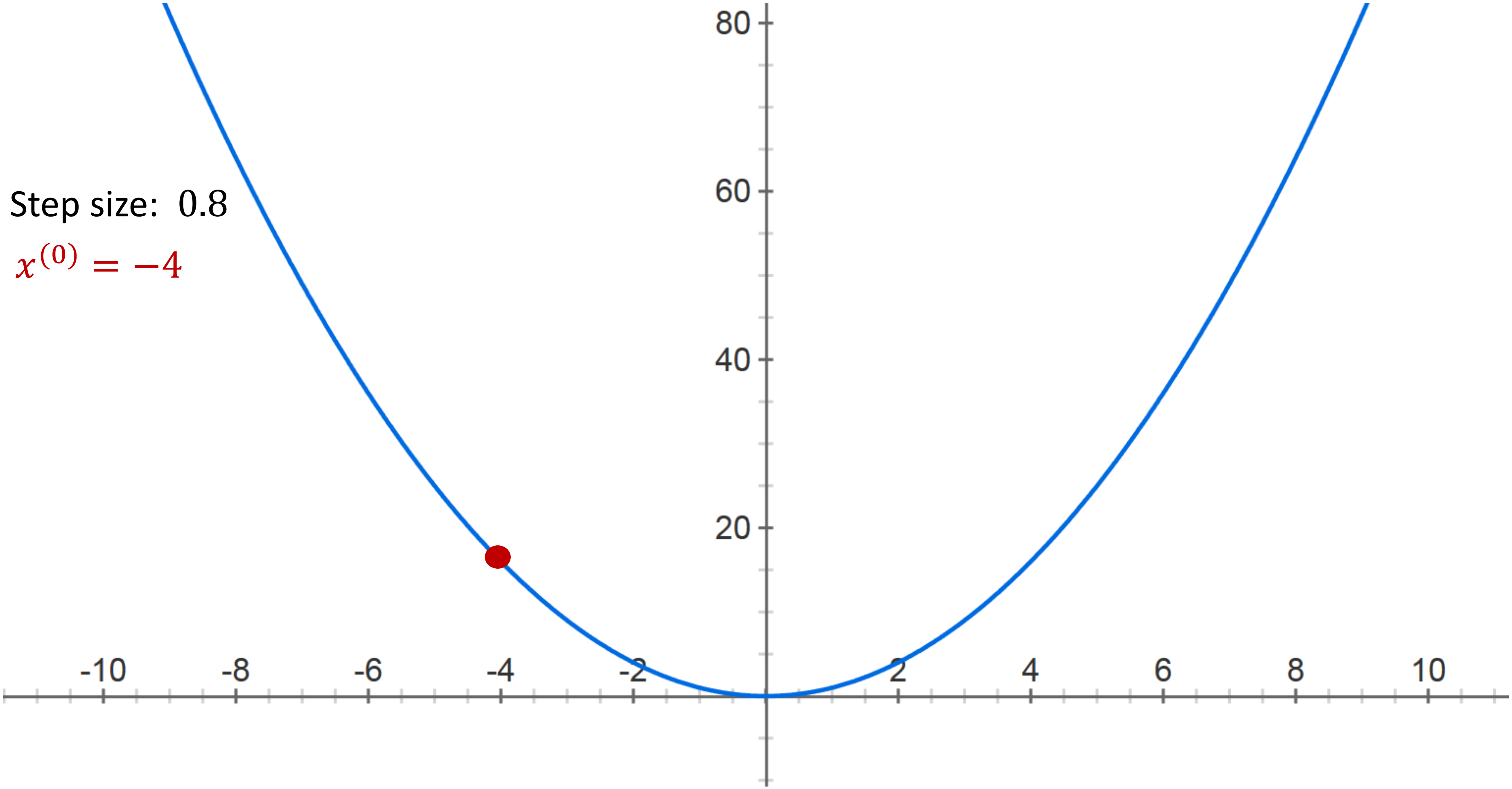
- x scalar → $\underline{x}$ vector
- Do partial derivative of $f(\underline{x})$ of each variable

Example

$$f(x) = x^2$$

Step size: 0.8

$x^{(0)} = -4$



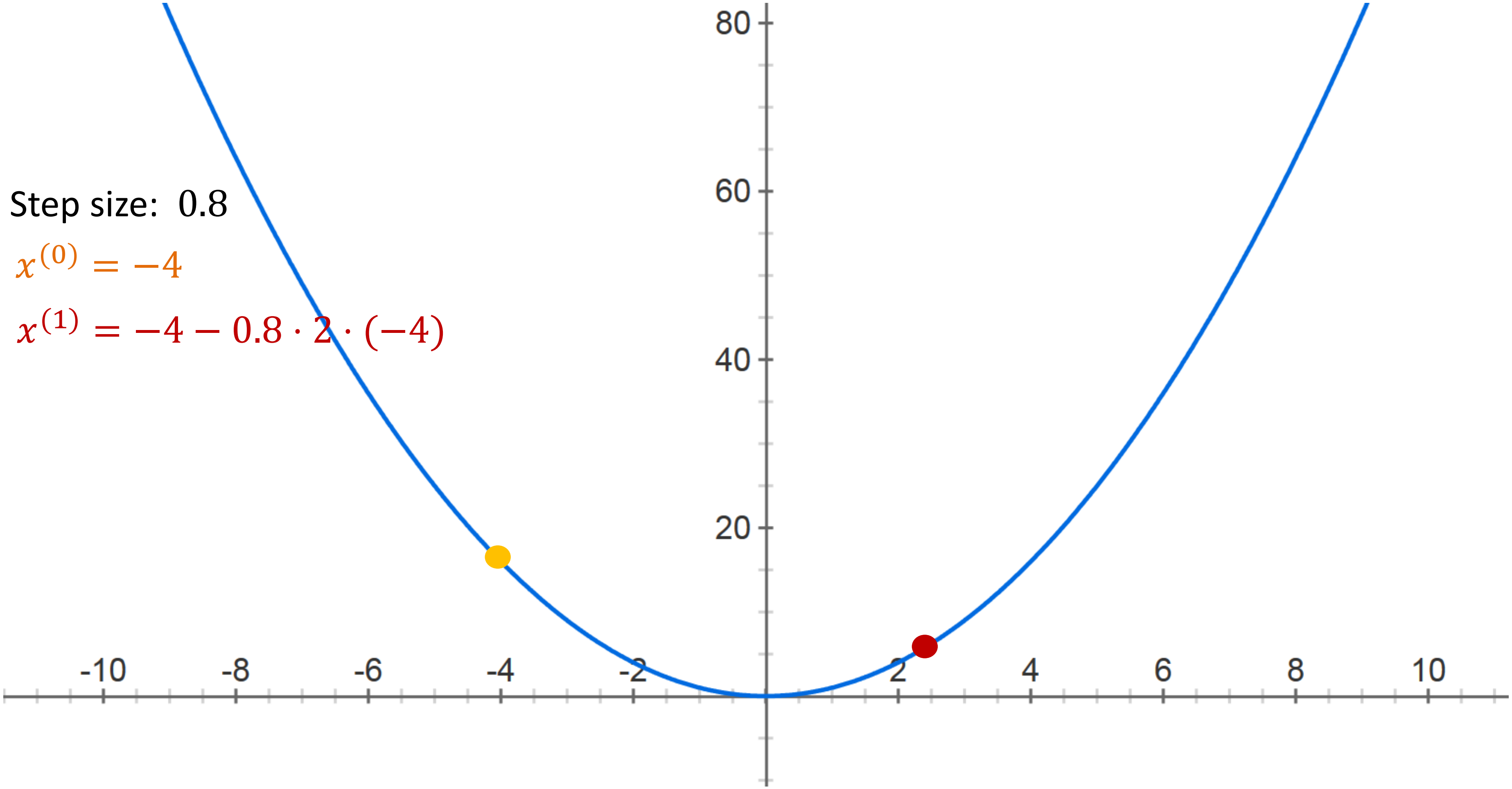
Example

$f(x) = x^2$

Step size: 0.8

$x^{(0)} = -4$

$x^{(1)} = -4 - 0.8 \cdot 2 \cdot (-4)$



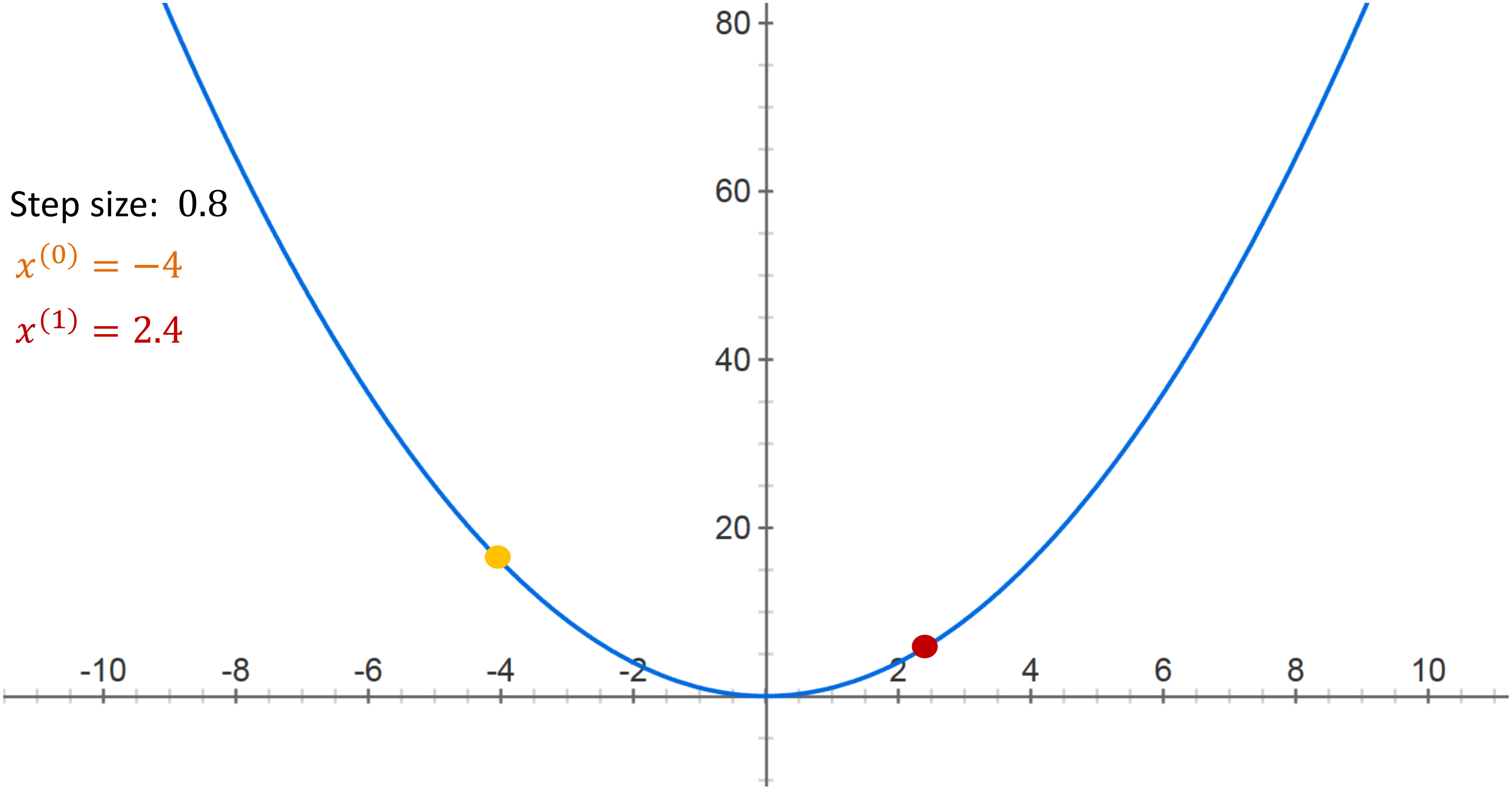
Example

$f(x) = x^2$

Step size: 0.8

$x^{(0)} = -4$

$x^{(1)} = 2.4$



Example

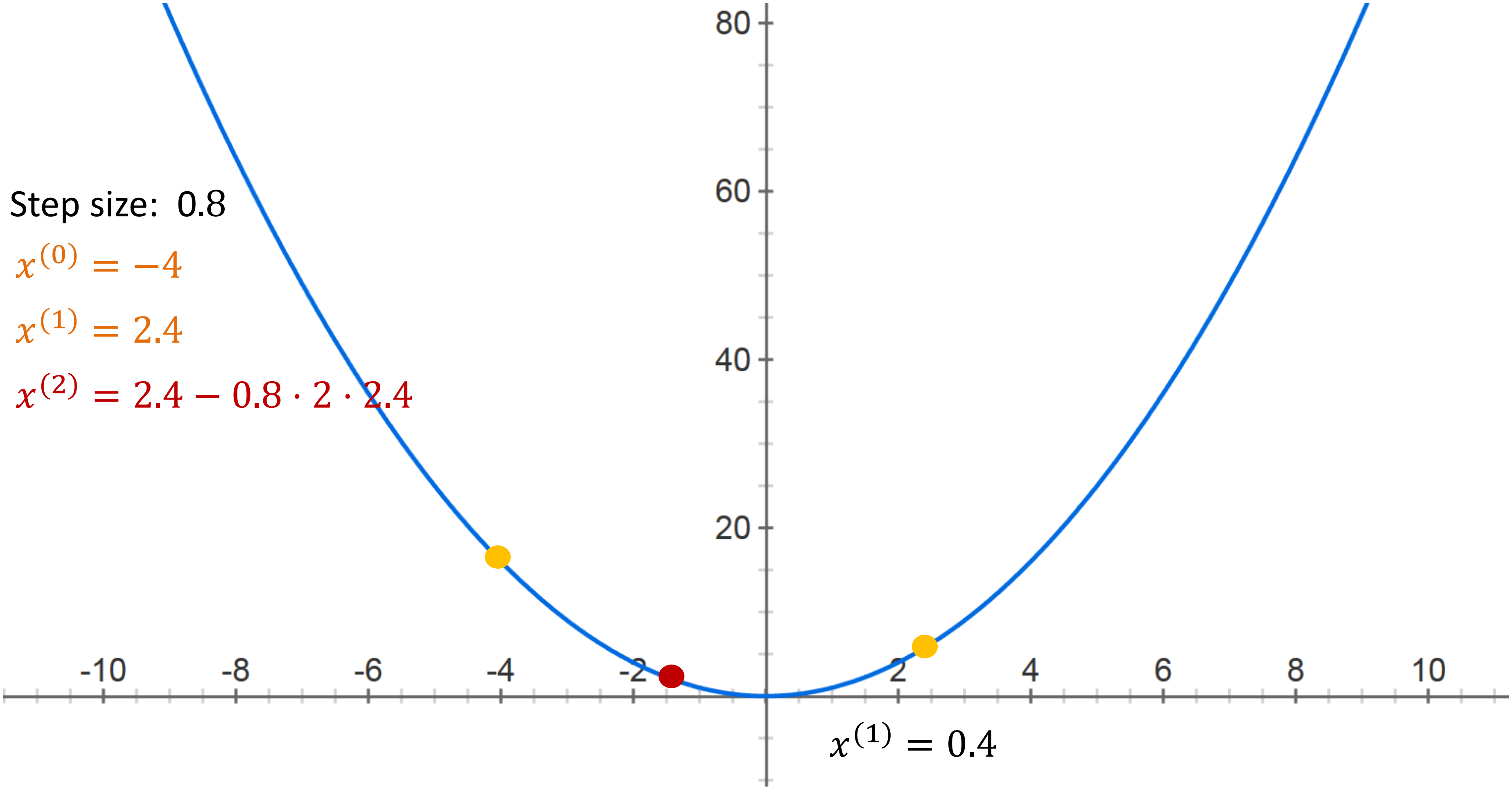
$f(x) = x^2$

Step size: 0.8

$x^{(0)} = -4$

$x^{(1)} = 2.4$

$x^{(2)} = 2.4 - 0.8 \cdot 2 \cdot 2.4$



$x^{(1)} = 0.4$

Example

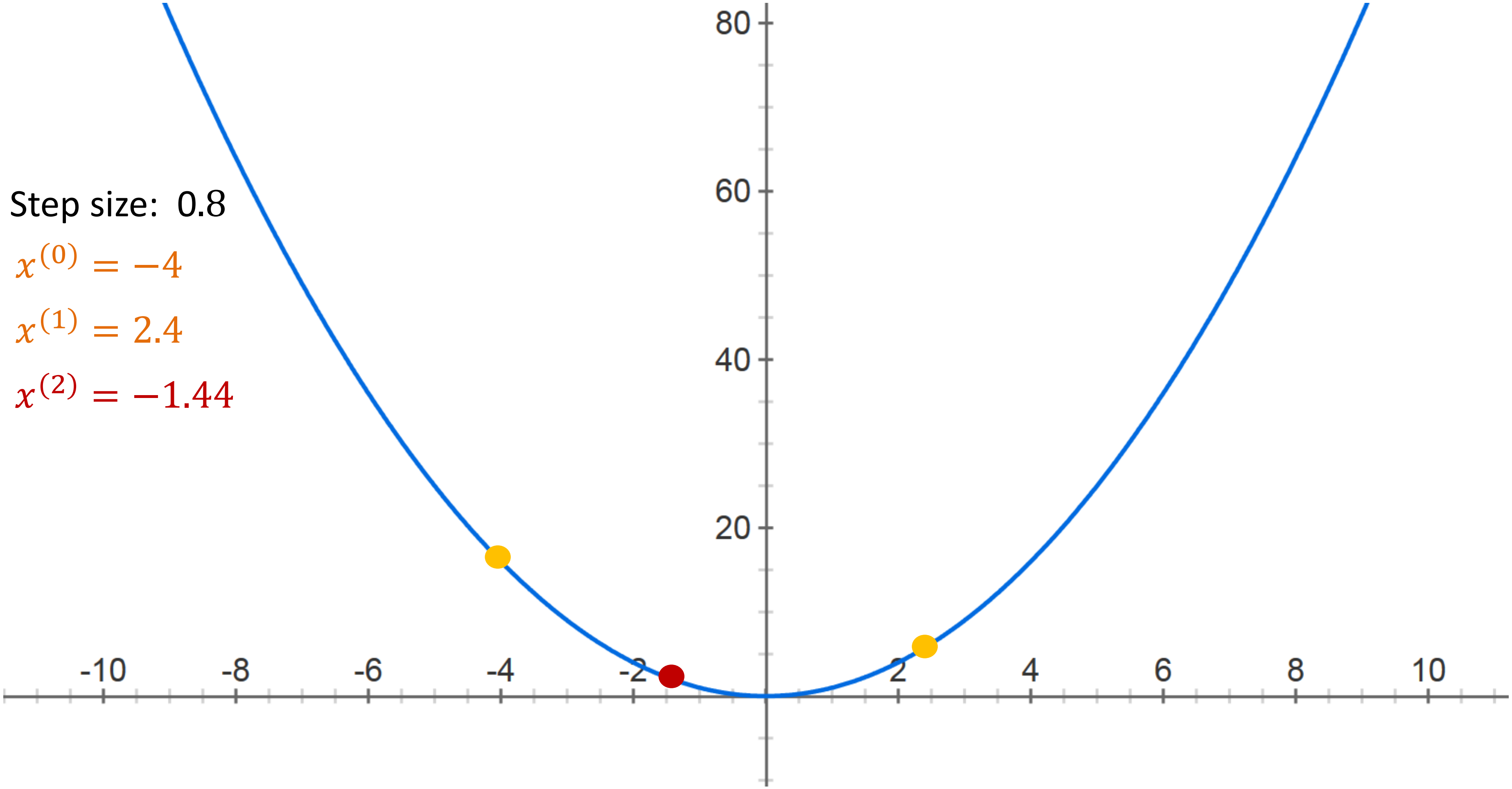
$f(x) = x^2$

Step size: 0.8

$x^{(0)} = -4$

$x^{(1)} = 2.4$

$x^{(2)} = -1.44$



Example

$f(x) = x^2$

Step size: 0.8

$x^{(0)} = -4$

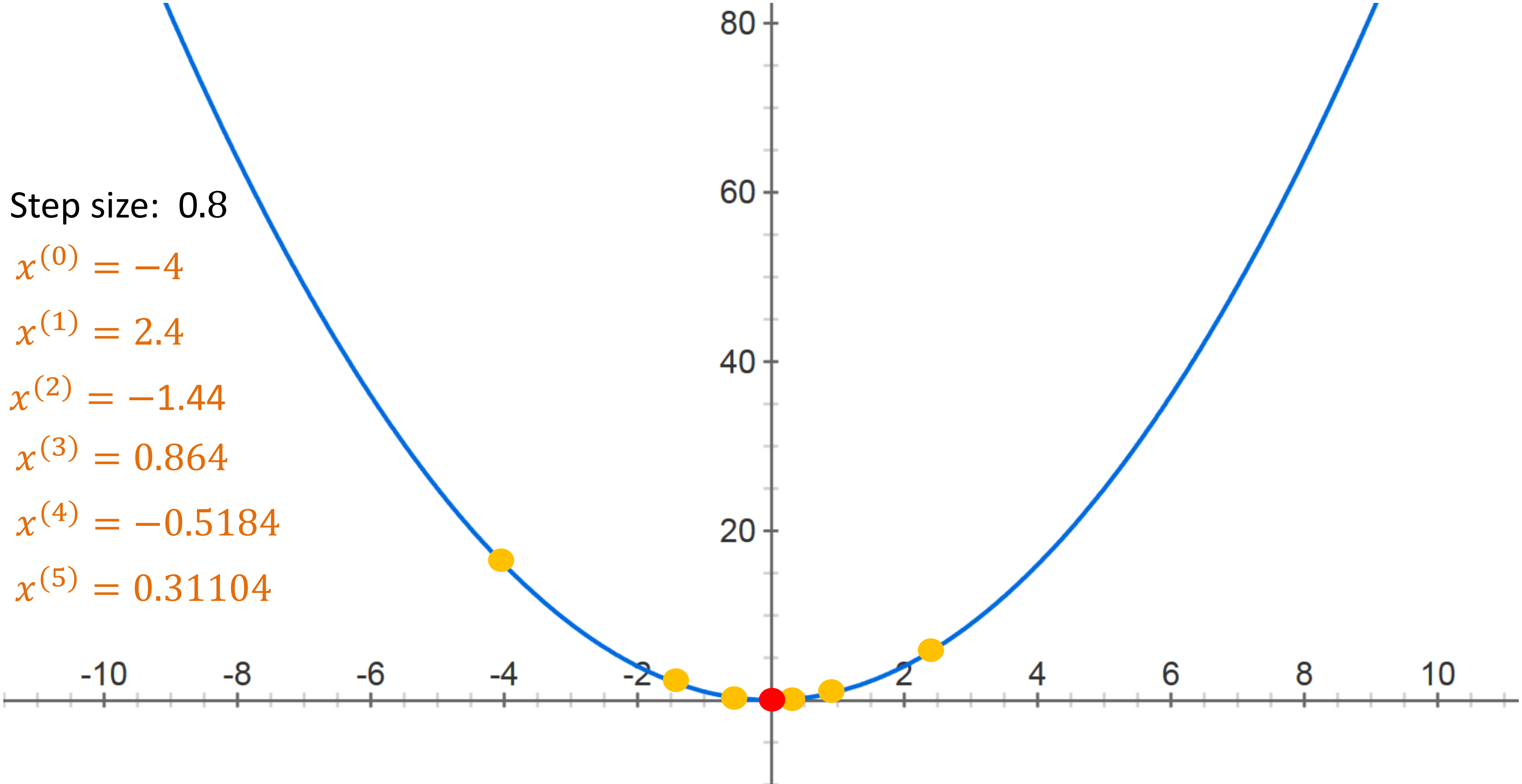
$x^{(1)} = 2.4$

$x^{(2)} = -1.44$

$x^{(3)} = 0.864$

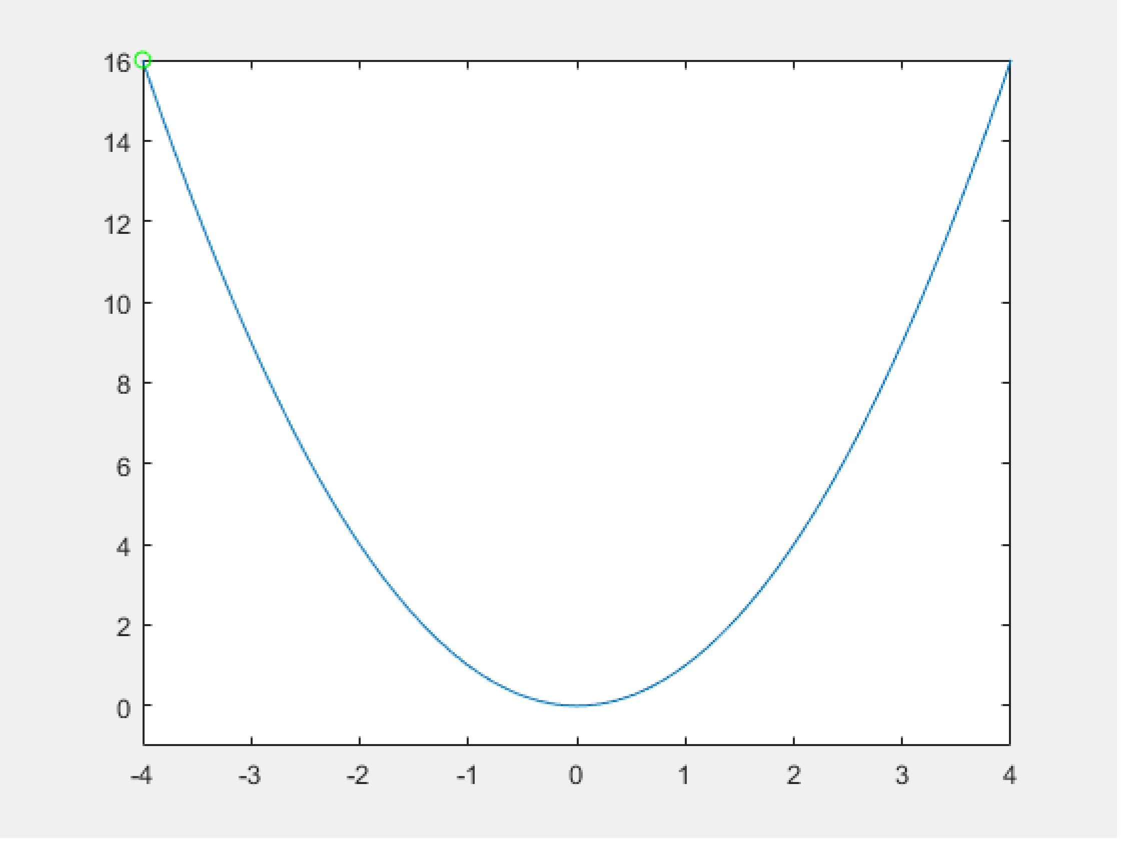
$x^{(4)} = -0.5184$

$x^{(5)} = 0.31104$



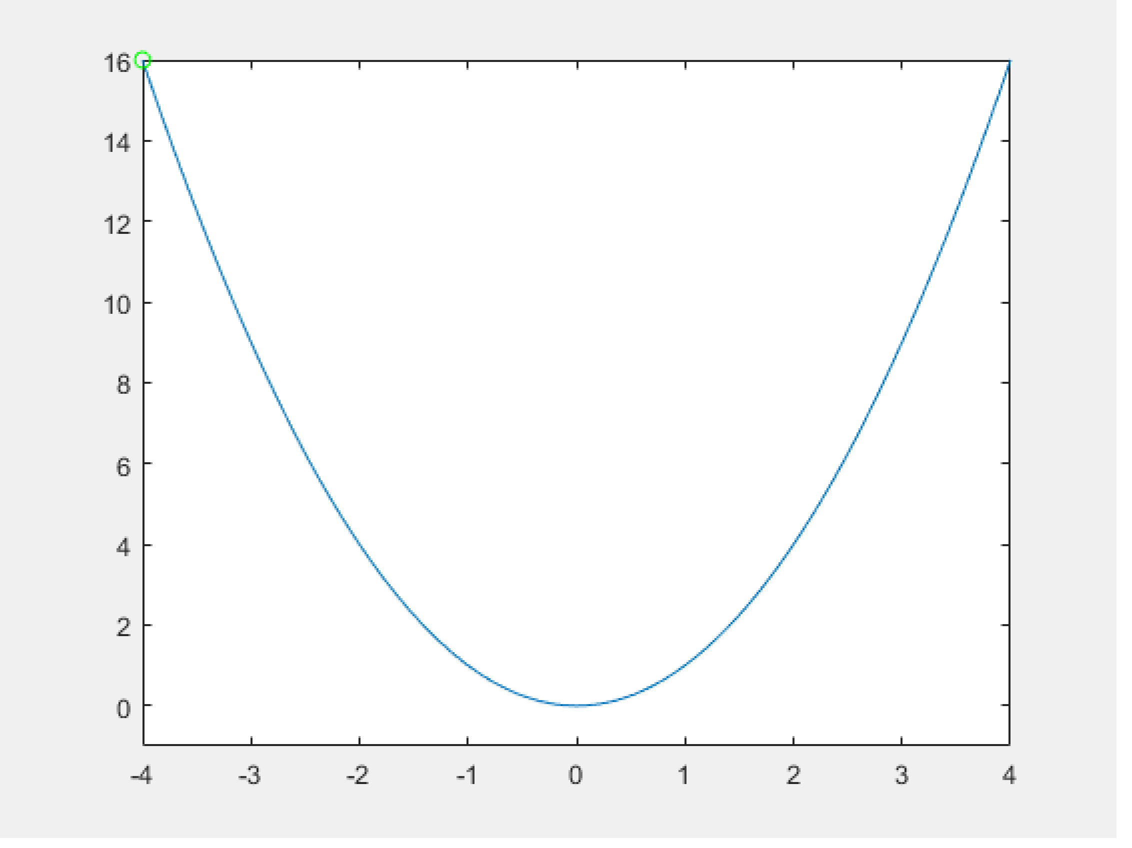
$x^{(30)} = -8.84296e - 07$

Step size matters



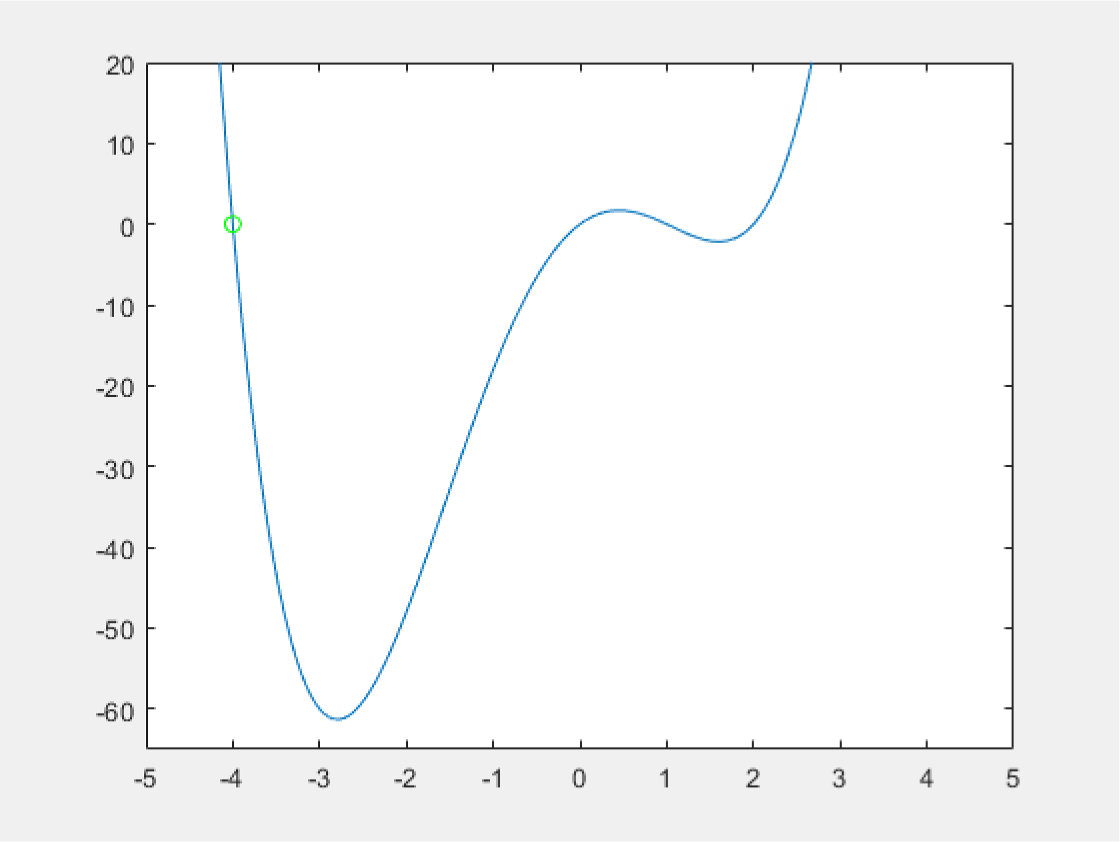
Step size: 0.9

Step size matters

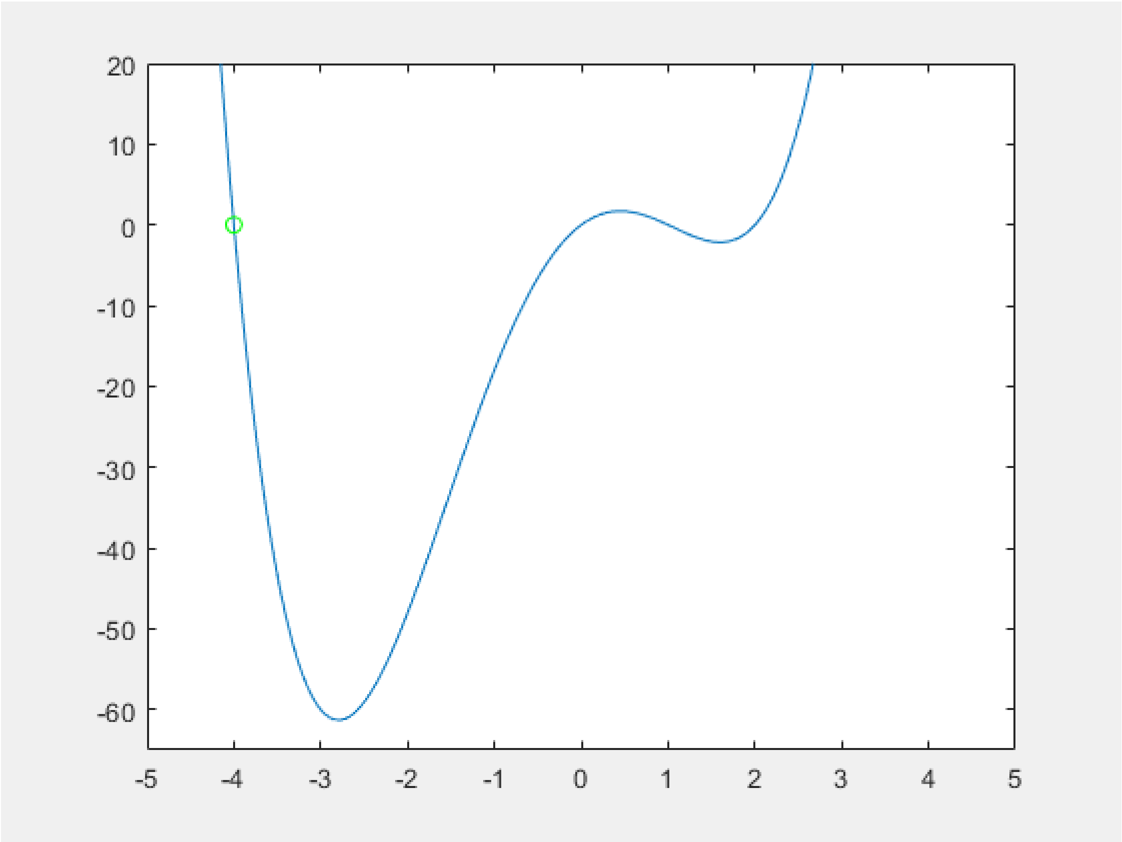


Step size: 0.2

Step size matters



Step size matters



Gradient descent is a greedy method!
It may return a local minimum!

程式練習（最終回）

And, please upload your program on moodle.

- Find *x* that minimizes

$$f(x) = x^4 - 2x^3 + 2$$

Closing remarks

- Thank you for your participation in this course.
- The final exam is scheduled on 14:20, Dec. 23. Please be prepared.