

# L07

## Gradient Descent

### Convergence and Constraints

CS 169 / CS 268  
Introduction to Optimization

Ioannis Panageas

Fall 2026

# Unconstrained optimization

$$\min_{x \in \mathbb{R}} f(x)$$

Assume that  $f$  is differentiable. We can evaluate  $f(x)$  and its derivative  $f'(x)$ .

Start from  $x_0$ . Use local derivative information to construct points  $x_1, x_2, \dots$  with smaller function values.

Our main question: **which direction, and how far?**

# Gradient descent in one variable

$$x_{k+1} = x_k - \eta_k f'(x_k), \quad \eta_k > 0.$$

$x_k$ : current point.       $f'(x_k)$ : current derivative.

$\eta_k$ : **step size**. A fixed step size uses  $\eta_k = \eta$  for every iteration.

$f'(x_k) > 0$  : move left,       $f'(x_k) < 0$  : move right.

# Why the negative derivative?

Taylor's first-order expression predicts the change along a small displacement  $h$ :

$$f(x + h) \approx f(x) + f'(x)h.$$

$$h = -\eta f'(x) \implies f'(x)h = -\eta [f'(x)]^2.$$

This is negative when  $f'(x) \neq 0$ . We still need to control the **Taylor error**.

# A one-variable quadratic

$$f(x) = \frac{1}{2}x^2, \quad f'(x) = x, \quad x_0 = 4.$$

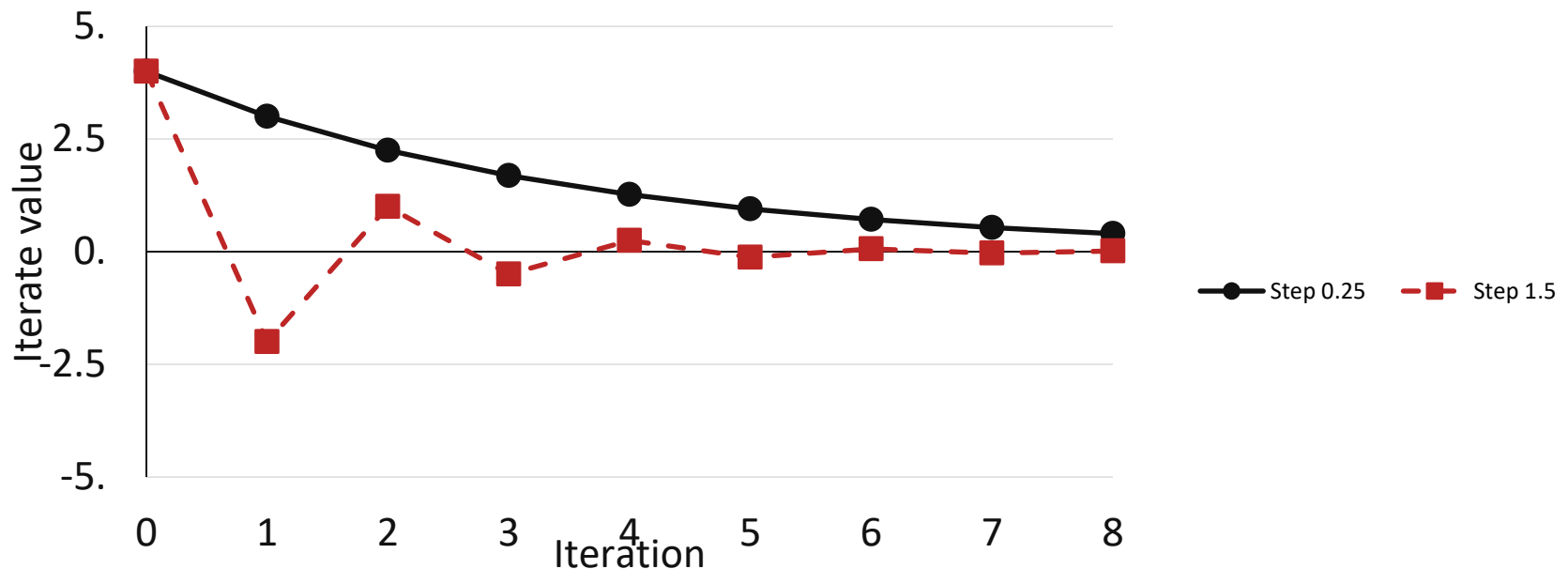
$$x_{k+1} = x_k - \eta x_k = (1 - \eta)x_k.$$

$$x_k = (1 - \eta)^k x_0, \quad f(x_k) = \frac{1}{2}(1 - \eta)^{2k} x_0^2.$$

Convergence depends on the multiplier  $1 - \eta$ .

# Two convergent step sizes

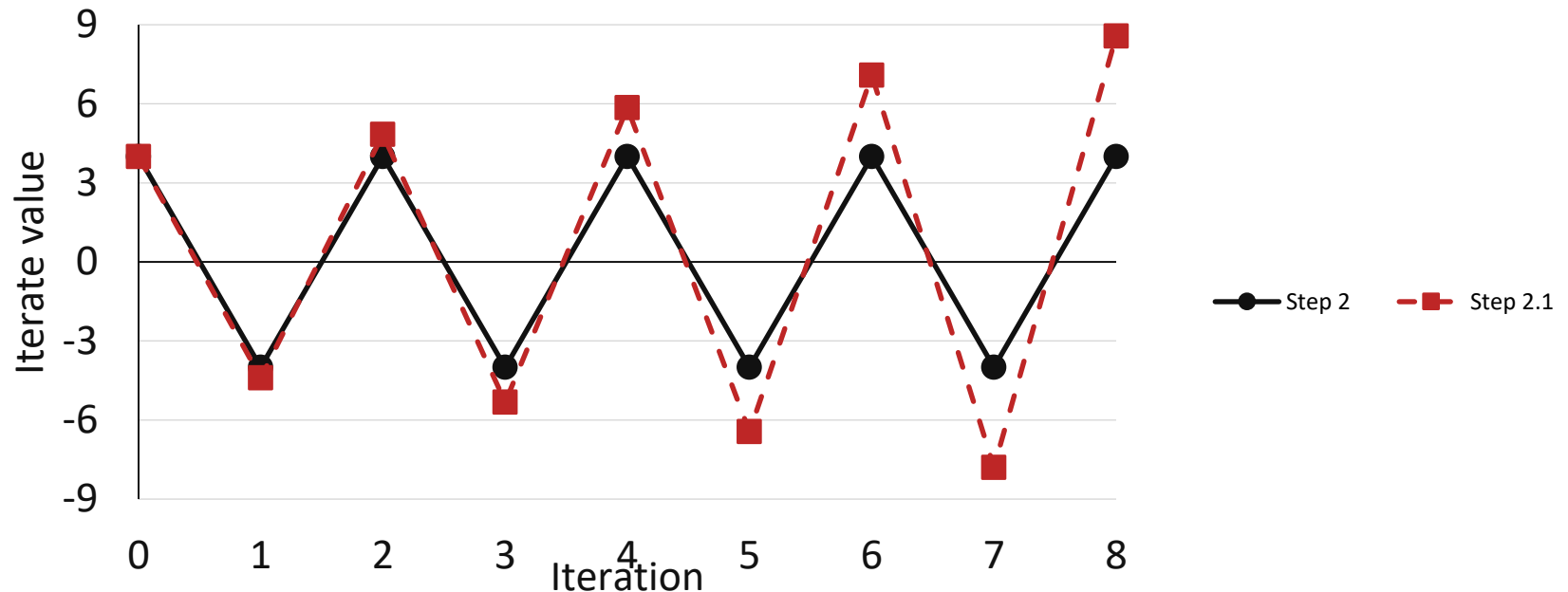
$$\eta = \frac{1}{4} : x_{k+1} = \frac{3}{4}x_k, \quad \eta = \frac{3}{2} : x_{k+1} = -\frac{1}{2}x_k.$$



The iterates may cross the minimizer while **the function value keeps decreasing.**

# When the step is too large

$$\eta = 2 : x_{k+1} = -x_k, \quad \eta = 2.1 : x_{k+1} = -1.1x_k.$$



At  $\eta = 2$ , the objective is constant. At  $\eta = 2.1$ , it grows even though we move opposite the gradient.

# The exact step-size condition

$$x_0 \neq 0 : \quad x_k \longrightarrow 0 \quad \Longleftrightarrow \quad |1 - \eta| < 1 \quad \Longleftrightarrow \quad 0 < \eta < 2.$$

At  $\eta = 1$ , this particular quadratic reaches the minimum in one step.

$$f(x) = \frac{a}{2}x^2, \quad a > 0 : \quad x_{k+1} = (1 - a\eta)x_k.$$

The condition becomes  $0 < \eta < 2/a$ . **More curvature requires a smaller step.**

# Smoothness in one variable

Assume  $f : \mathbb{R} \rightarrow \mathbb{R}$  is twice continuously differentiable. It is  **$L$ -smooth** if  $L > 0$  and

$$|f'(x) - f'(y)| \leq L|x - y| \quad \text{for all } x, y.$$

The constant  $L$  bounds how quickly the derivative can change as we move along the line.

$$|f''(x)| \leq L \quad \text{for all } x \quad \text{is sufficient.}$$

Example:  $f(x) = x^2/2$  has  $f''(x) = 1$ , so we can take  $L = 1$ .

# Taylor's theorem gives an upper bound

For a displacement  $h$ , Taylor's theorem gives

$$f(x + h) = f(x) + f'(x)h + \frac{1}{2}f''(\xi)h^2,$$

Here  $\xi$  lies between  $x$  and  $x + h$ . Smoothness gives  $f''(\xi) \leq L$ , so

$$f(x + h) \leq f(x) + f'(x)h + \frac{L}{2}h^2.$$

# Why gradient descent decreases the function

For  $x_{k+1} = x_k - \eta f'(x_k)$ , substitute  $h = -\eta f'(x_k)$  into the upper bound.

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) - \eta [f'(x_k)]^2 + \frac{L\eta^2}{2} [f'(x_k)]^2 \\ &= f(x_k) - \eta \left(1 - \frac{L\eta}{2}\right) [f'(x_k)]^2. \end{aligned}$$

If  $0 < \eta < 2/L$  and  $f'(x_k) \neq 0$ , then  $f$  strictly decreases.  
A convenient choice is  $\eta = 1/L$ :

$$f(x_{k+1}) \leq f(x_k) - \frac{[f'(x_k)]^2}{2L}.$$

# Squared distance after a gradient step

For  $\eta > 0$ , subtract  $x^\star$  from the gradient update:

$$x_{k+1} - x^\star = (x_k - x^\star) - \eta f'(x_k).$$

Square both sides and expand:

$$\begin{aligned}(x_{k+1} - x^\star)^2 &= [(x_k - x^\star) - \eta f'(x_k)]^2 \\&= (x_k - x^\star)^2 - 2\eta f'(x_k)(x_k - x^\star) \\&\quad + \eta^2 [f'(x_k)]^2.\end{aligned}$$

# Convexity relates values to distance

Move the cross term to the left, then divide by  $2\eta$ :

$$\begin{aligned} 2\eta f'(x_k)(x_k - x^*) &= (x_k - x^*)^2 - (x_{k+1} - x^*)^2 \\ &\quad + \eta^2 [f'(x_k)]^2. \\ f'(x_k)(x_k - x^*) &= \frac{(x_k - x^*)^2 - (x_{k+1} - x^*)^2}{2\eta} \\ &\quad + \frac{\eta}{2} [f'(x_k)]^2. \end{aligned}$$

Now assume  $f$  is convex and  $x^*$  is a minimizer. First-order convexity gives

$$f(x_k) - f(x^*) \leq f'(x_k)(x_k - x^*).$$

# Convex functions: one-step progress

Set  $\eta = 1/L$  in the distance identity and apply convexity:

$$f(x_k) - f(x^*) \leq \frac{L}{2} [(x_k - x^*)^2 - (x_{k+1} - x^*)^2] \\ + \frac{[f'(x_k)]^2}{2L}.$$

The descent bound cancels the derivative term:

$$f(x_{k+1}) - f(x^*) \leq f(x_k) - f(x^*) - \frac{[f'(x_k)]^2}{2L} \\ \leq \frac{L}{2} [(x_k - x^*)^2 - (x_{k+1} - x^*)^2].$$

# Convex functions: the last-iterate bound

Summing over  $k = 0, \dots, K - 1$  ( $K \geq 1$ ) gives

$$\begin{aligned}\sum_{k=1}^K [f(x_k) - f(x^*)] &\leq \frac{L}{2} [(x_0 - x^*)^2 - (x_K - x^*)^2] \\ &\leq \frac{L}{2} (x_0 - x^*)^2.\end{aligned}$$

The function values decrease, so every term in the sum is at least  $f(x_K) - f(x^*)$ . Therefore

$$f(x_K) - f(x^*) \leq \frac{L(x_0 - x^*)^2}{2K}.$$

# Nonconvex functions: summing the decrease

Assume  $f(x) \geq f_{\inf}$  for every  $x$ , where  $f_{\inf}$  is a finite lower bound. Use  $\eta = 1/L$ .

$$\frac{[f'(x_k)]^2}{2L} \leq f(x_k) - f(x_{k+1}).$$

Sum for  $k = 0, \dots, K-1$ . The intermediate function values cancel:

$$\begin{aligned} \frac{1}{2L} \sum_{k=0}^{K-1} [f'(x_k)]^2 &\leq f(x_0) - f(x_K) \\ &\leq f(x_0) - f_{\inf}. \end{aligned}$$

# Nonconvex functions: a small derivative

$$\min_{0 \leq k < K} |f'(x_k)|^2 \leq \frac{2L(f(x_0) - f_{\inf})}{K}.$$

For some iterate to satisfy  $|f'(x_k)| \leq \varepsilon$ , it suffices to take  $K \geq 1$  with

$$K \geq \frac{2L(f(x_0) - f_{\inf})}{\varepsilon^2}.$$

The smallest absolute derivative is  $O(1/\sqrt{K})$ . This gives **approximate stationarity**.

# What the two guarantees measure

**With convexity:** if a minimizer  $x^*$  exists, then

$$f(x_K) - f(x^*) \leq \frac{L(x_0 - x^*)^2}{2K} = O\left(\frac{1}{K}\right).$$

**Without convexity:** smoothness and a finite lower bound.

$$\min_{0 \leq k < K} |f'(x_k)| \leq \sqrt{\frac{2L(f(x_0) - f_{\inf})}{K}} = O\left(\frac{1}{\sqrt{K}}\right).$$

Both use  $\eta = 1/L$ . The bounds measure different quantities.

# Stationary need not mean minimum

$$f(x) = \cos x, \quad f'(x) = -\sin x.$$

$$f'(0) = 0, \quad f''(0) = -1.$$

Starting from  $x_0 = 0$ , gradient descent remains at a **local maximum**.

The nonconvex guarantee controls the derivative. It does not certify a global minimum.

# Adding an interval constraint

$$\min_{x \in [a,b]} f(x), \quad a < b.$$

A gradient step may leave the feasible interval. We need an update that keeps every iterate feasible.

$$f(x) = \frac{1}{2}(x - 2)^2, \quad x \in [0, 1].$$

The unconstrained minimum is at 2. The constrained minimum is at **1**.

# Projection onto an interval

The **projection** of  $z$  is the feasible point closest to  $z$ .

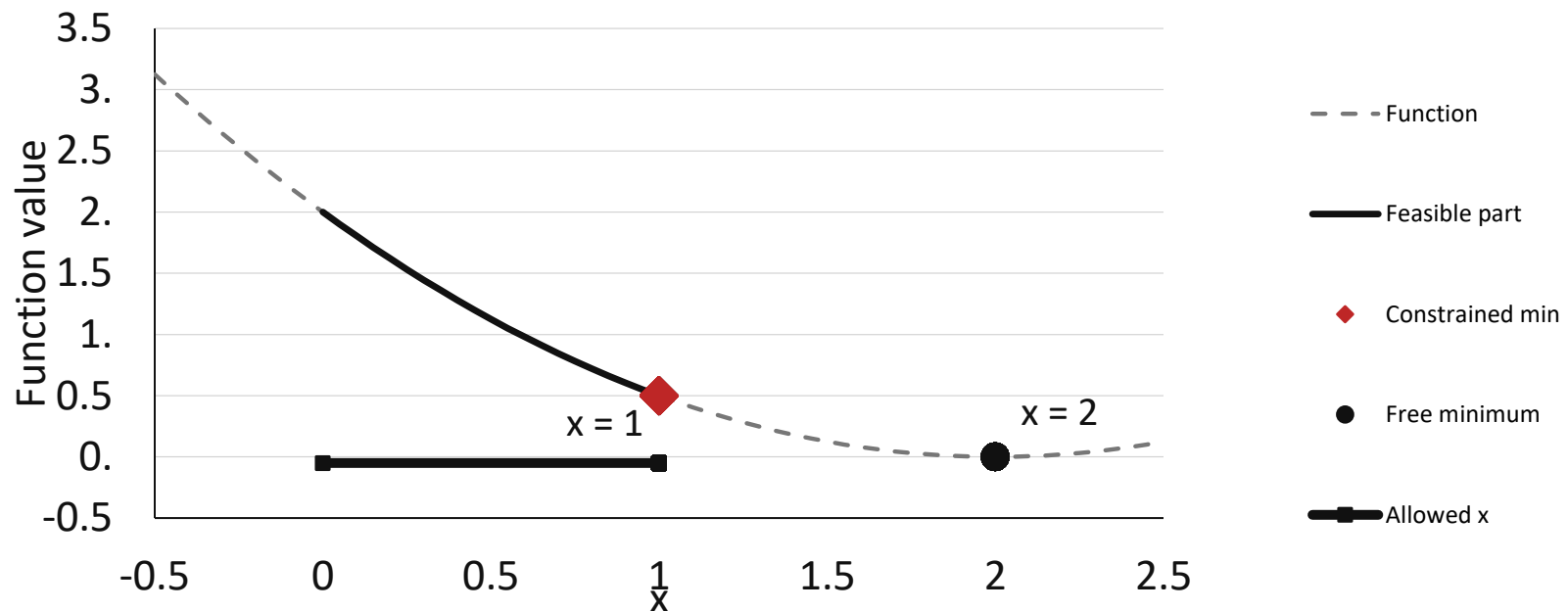
$$\Pi_{[a,b]}(z) = \begin{cases} a, & z < a, \\ z, & a \leq z \leq b, \\ b, & z > b. \end{cases}$$

$$x_{k+1} = \Pi_{[a,b]}(x_k - \eta f'(x_k)).$$

In one variable, projection is simply **clipping to the interval**.

# A projected step

$$x_0 = \frac{1}{2}, \quad \eta = 1 : \quad x_0 - f'(x_0) = 2, \quad x_1 = \Pi_{[0,1]}(2) = 1.$$



The projection makes the step feasible and reaches the constrained minimum.

# Optimality at the boundary

$$f'(1) = -1 \neq 0, \quad \Pi_{[0,1]}(1 - \eta f'(1)) = 1 \quad (\eta > 0).$$

The unconstrained step points outside the interval.  
Projection leaves the point unchanged. Necessary conditions at a local minimum:

$$\begin{aligned} a < x^* < b : \quad & f'(x^*) = 0, \\ x^* = a : \quad & f'(a) \geq 0, \\ x^* = b : \quad & f'(b) \leq 0. \end{aligned}$$

## Several variables: unconstrained

$$x_{k+1} = x_k - \eta \nabla f(x_k), \quad x_k \in \mathbb{R}^n.$$

Assume  $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$ . With  $\eta = 1/L$ , use the same proofs and assumptions.

$$\text{Convex: } f(x_K) - f(x^*) \leq \frac{L\|x_0 - x^*\|^2}{2K},$$

$$\text{Nonconvex: } \min_{0 \leq k < K} \|\nabla f(x_k)\|^2 \leq \frac{2L(f(x_0) - f_{\inf})}{K}.$$

# Projection in several variables

$$\Pi_C(z) = \arg \min_{y \in C} \|y - z\|^2,$$

For a nonempty closed convex set  $C$ , the projection exists and is unique.

$$x_{k+1} = \Pi_C(x_k - \eta \nabla f(x_k)).$$

For a box, clip each coordinate to its interval. For a general set, computing the projection can itself require optimization.

# Reading

Chong, Lu, and Zak, 5th edition

Chapter 8: Gradient Methods. Sections 8.1–8.3 support the algorithm and its analysis. Review Section 5.6 for Taylor expansions.

For the introduction to constraints: Section 24.2 (Projections) and Section 24.3 (Projected Gradient Methods with Linear Constraints).

# Conceptual questions

1. For unconstrained optimization, why does a zero derivative certify a global minimum of a differentiable convex function, but not of a nonconvex function?
2. At a constrained minimum, why might the derivative be nonzero even though the projected gradient step is zero?