

# L10

## Newton's Method

CS 169 / CS 268  
Introduction to Optimization

Ioannis Panageas

Fall 2026

# Gradient descent and curvature

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

The derivative gives a direction. The second derivative describes how quickly the derivative changes.

Newton's method uses this **curvature** to choose the step.

We use exact derivatives in this lecture.

# Taylor's theorem: second order

Near a reference point  $x_0$ , the second-order approximation is

$$f(x_0 + h) \approx f(x_0) + f'(x_0)h + \frac{1}{2}f''(x_0)h^2.$$

Equivalently, with  $h = x - x_0$ ,

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2.$$

# Minimizing the approximation

At  $x_k$ , suppose  $f''(x_k) > 0$ . We choose  $h$  to minimize

$$f(x_k) + f'(x_k)h + \frac{1}{2}f''(x_k)h^2.$$

Its derivative with respect to  $h$  is zero when

$$f'(x_k) + f''(x_k)h = 0, \quad h = -\frac{f'(x_k)}{f''(x_k)}.$$

# Newton's update

For  $f''(x_k) > 0$ , the minimizing displacement satisfies

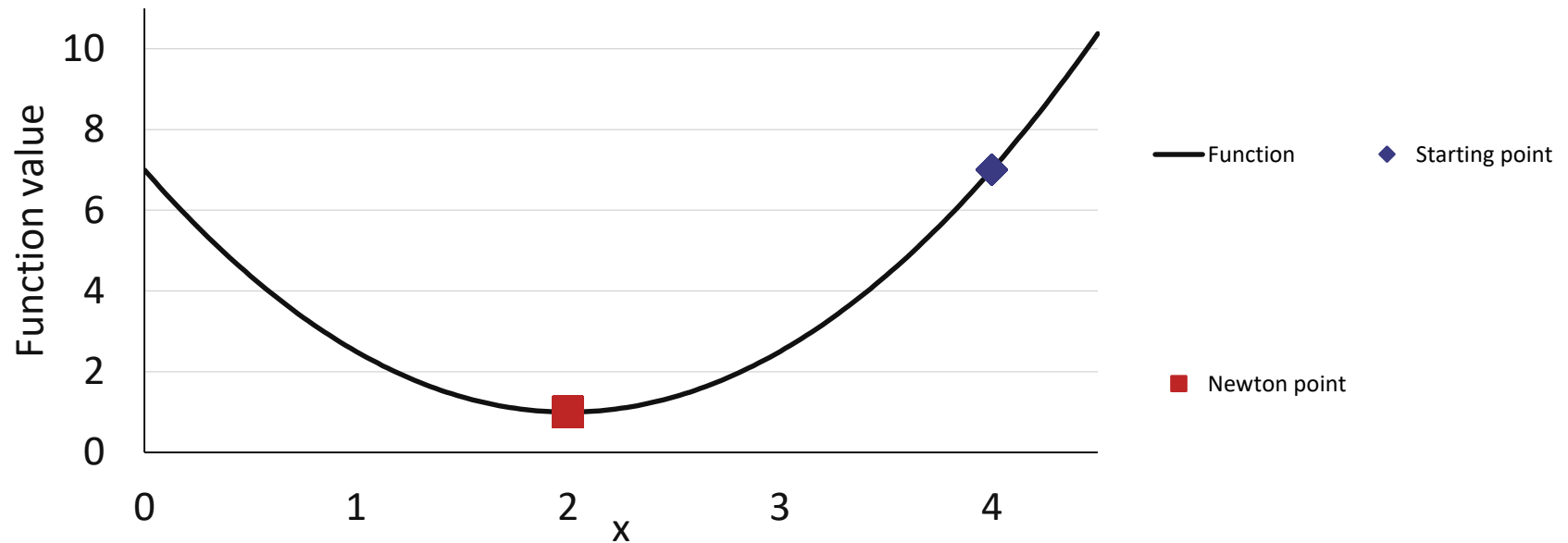
$$f'(x_k) + f''(x_k)h_k = 0 \quad \Longrightarrow \quad h_k = -\frac{f'(x_k)}{f''(x_k)}.$$

The approximation is for  $f(x_k + h)$ , so we choose  $x_{k+1} = x_k + h_k$ .

$$x_{k+1} = x_k + h_k = x_k - \frac{f'(x_k)}{f''(x_k)}.$$

# One step for a quadratic

$$f(x) = \frac{3}{2}(x - 2)^2 + 1, \quad x_{k+1} = x_k - \frac{3(x_k - 2)}{3} = 2.$$



The second-order approximation equals the function everywhere.

# A nonlinear example

$$f(x) = x - \log x - 1, \quad x > 0.$$

$$f'(x) = 1 - \frac{1}{x}, \quad f''(x) = \frac{1}{x^2} > 0.$$

$$x^* = 1, \quad f(x^*) = 0.$$

The curvature changes with  $x$ .

# Newton iterations in the example

$$x_{k+1} = x_k - \frac{1 - 1/x_k}{1/x_k^2} = 2x_k - x_k^2.$$

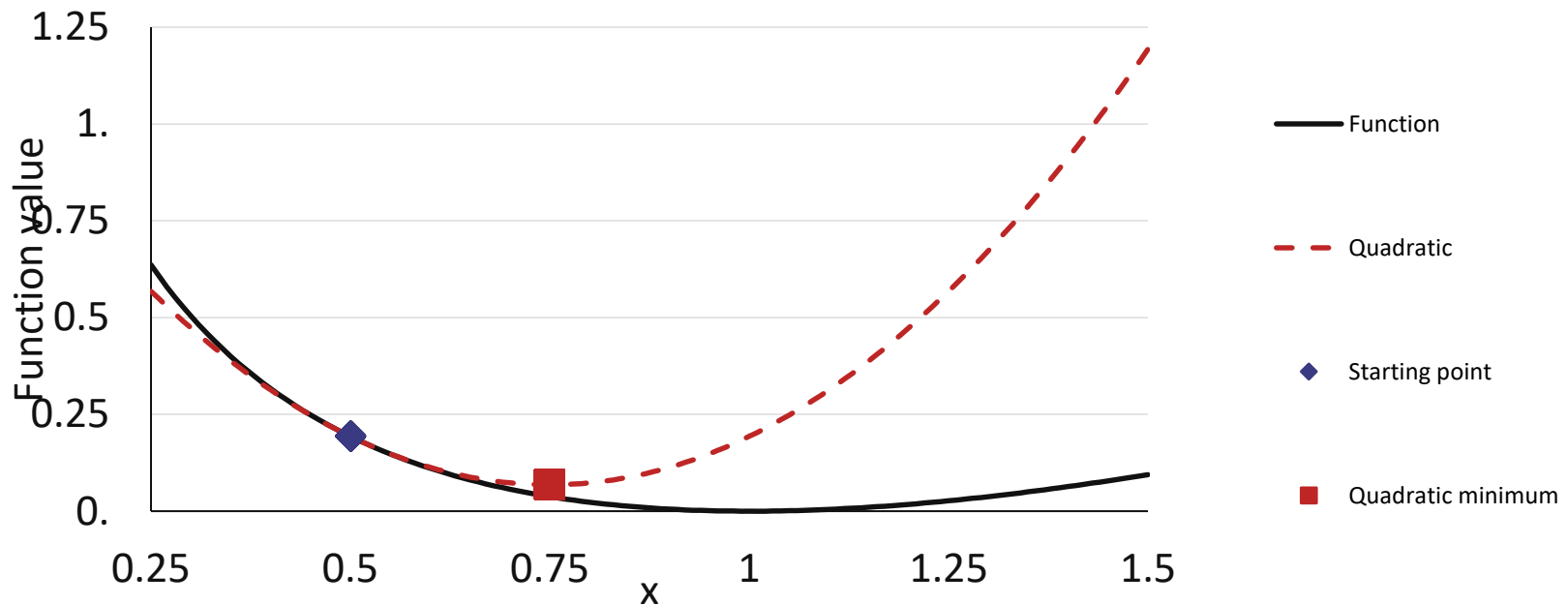
Starting at  $x_0 = 1/2$  gives

| $k$   | 0             | 1             | 2               | 3                 | 4                     |
|-------|---------------|---------------|-----------------|-------------------|-----------------------|
| $x_k$ | $\frac{1}{2}$ | $\frac{3}{4}$ | $\frac{15}{16}$ | $\frac{255}{256}$ | $\frac{65535}{65536}$ |

The iterates quickly approach  $x^* = 1$ .

# The quadratic approximation

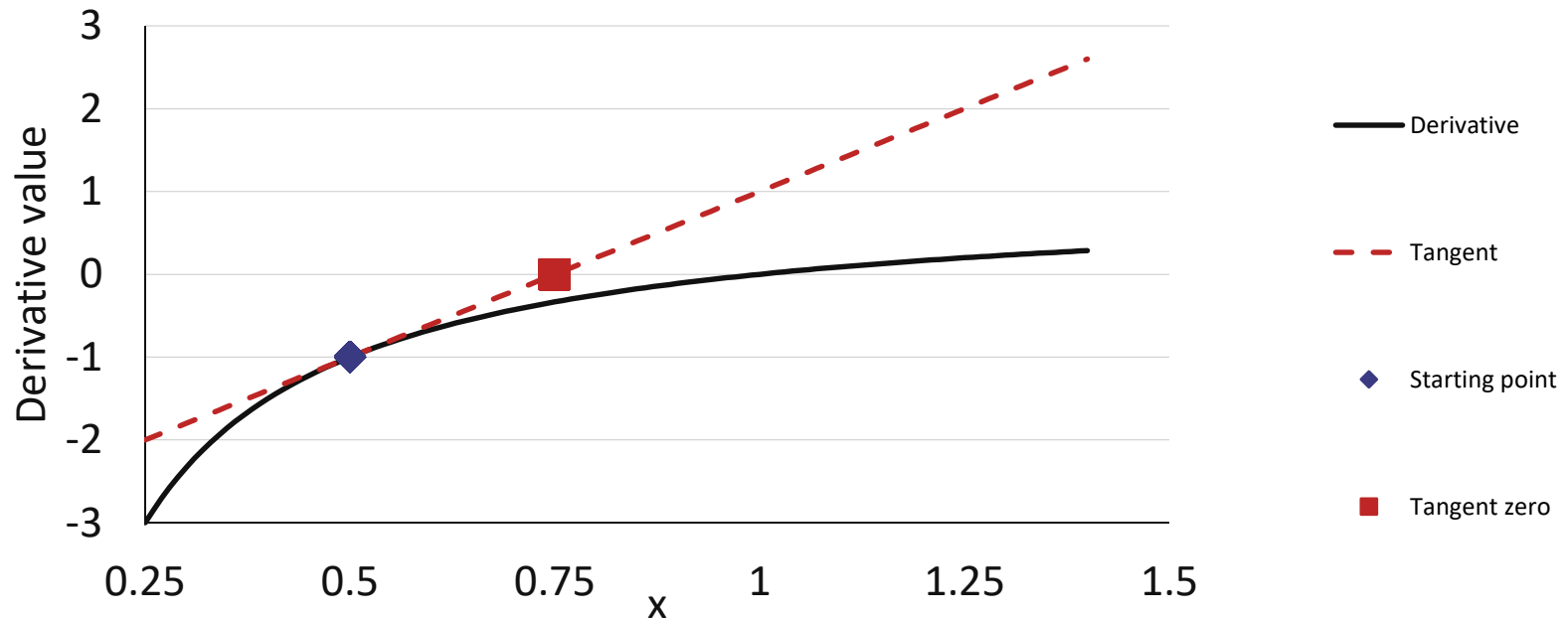
$$x_0 = \frac{1}{2}, \quad f'(x_0) = -1, \quad f''(x_0) = 4, \quad x_1 = \frac{3}{4}.$$



Newton minimizes the local quadratic, then rebuilds it at the new point.

# Newton as a root-finding method

$$f'(x_k) + f''(x_k)(x_{k+1} - x_k) = 0.$$



We use the zero of the tangent line to the derivative to estimate a stationary point.

# Quadratic convergence in the example

Let  $e_k = x_k - x^* = x_k - 1$ . Then

$$\begin{aligned} e_{k+1} &= 2x_k - x_k^2 - 1 \\ &= -(x_k - 1)^2 = -e_k^2. \end{aligned}$$

$$|e_{k+1}| = |e_k|^2.$$

For example, an error of  $10^{-2}$  becomes  $10^{-4}$ , then  $10^{-8}$ .

# Local quadratic convergence

Suppose  $f$  has three continuous derivatives near  $x^*$  and

$$f'(x^*) = 0, \quad f''(x) \geq \mu > 0, \quad |f'''(x)| \leq M.$$

For a sufficiently close starting point, Newton satisfies

$$|x_{k+1} - x^*| \leq \frac{M}{2\mu} |x_k - x^*|^2.$$

# Analysis: Taylor's theorem

Let  $e_k = x_k - x^*$ . Apply Taylor's theorem to  $f'$  at  $x_k$ , using  $f'(x^*) = 0$ .

$$0 = f'(x_k) - f''(x_k)e_k + \frac{1}{2}f'''(\xi_k)e_k^2.$$

Here  $\xi_k$  lies between  $x_k$  and  $x^*$ . Divide by  $f''(x_k) > 0$  and rearrange:

$$\frac{f'(x_k)}{f''(x_k)} = e_k - \frac{f'''(\xi_k)}{2f''(x_k)}e_k^2.$$

# Analysis: the Newton update

Substitute the preceding identity into the error after one Newton step:

$$\begin{aligned} e_{k+1} &= e_k - \frac{f'(x_k)}{f''(x_k)} \\ &= e_k - \left( e_k - \frac{f'''(\xi_k)}{2f''(x_k)} e_k^2 \right) \\ &= \frac{f'''(\xi_k)}{2f''(x_k)} e_k^2. \end{aligned}$$

The linear error cancels. Now use  $|f'''(\xi_k)| \leq M$  and  $f''(x_k) \geq \mu$  to bound the coefficient.

## Why the starting point must be close

$$|e_{k+1}| \leq \frac{M}{2\mu} |e_k|^2 = C |e_k|^2, \quad C = \frac{M}{2\mu}.$$

Let the bounds hold for  $|x - x^*| \leq r$ . Choose the starting point so that

$$|e_0| \leq r, \quad C |e_0| \leq \frac{1}{2}.$$

$$|e_{k+1}| \leq C |e_k|^2 \leq \frac{1}{2} |e_k|.$$

The iterates stay in the interval and converge to  $x^*$ . The squared-error bound gives quadratic convergence.

# The starting point matters

$$f(x) = x - \log x - 1, \quad x > 0.$$

$$x_0 = 3, \quad x_1 = 2(3) - 3^2 = -3.$$

The new point lies outside the domain, even though the function is strictly convex.

The local quadratic approximation can propose a step that is too large.

# Damped Newton's method

Damping controls the step length in **unconstrained optimization**.

$$p_k = -\frac{f'(x_k)}{f''(x_k)}, \quad x_{k+1} = x_k + \alpha_k p_k, \quad 0 < \alpha_k \leq 1.$$

Full Newton steps ( $\alpha_k = 1$ ) can fail to converge. Damping shortens steps when the local approximation is inaccurate.

Backtracking chooses  $\alpha_k$  for sufficient decrease. A domain check ensures that  $f$  can be evaluated.

# A descent direction

If  $f''(x_k) > 0$  and  $f'(x_k) \neq 0$ , then

$$f'(x_k)p_k = -\frac{[f'(x_k)]^2}{f''(x_k)} < 0.$$

$$f(x_k + \alpha p_k) = f(x_k) + \alpha f'(x_k)p_k + o(\alpha).$$

Every sufficiently small positive  $\alpha$  gives a decrease.

# Backtracking

We try  $\alpha = 1, 1/2, 1/4, \dots$  until the trial point is in the domain and

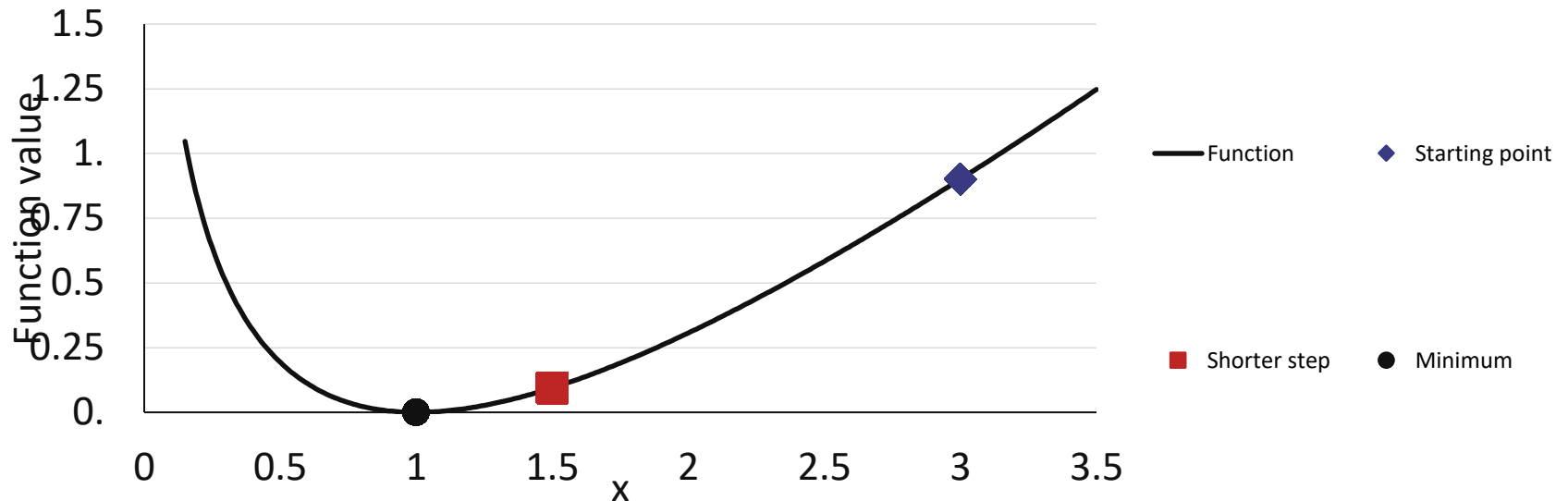
$$f(x_k + \alpha p_k) \leq f(x_k) + \frac{1}{4} \alpha f'(x_k) p_k.$$

The right side is below  $f(x_k)$  when  $p_k$  is a descent direction.

We accept the first trial that passes this **sufficient-decrease** test.

# Backtracking in the example

$$x_0 = 3, \quad p_0 = -6, \quad x_{\text{trial}} = x_0 + \alpha p_0 = 3 - 6\alpha.$$



$\alpha = 1$  gives  $-3$ , and  $\alpha = 1/2$  gives  $0$ . Both trials are outside  $x > 0$ : **reject**.

With  $\alpha = 1/4$ , the test passes:  $x_1 = 3/2$ .

# Several variables: Taylor approximation

For  $x_k, h \in \mathbb{R}^n$ , the second-order approximation is

$$f(x_k + h) \approx f(x_k) + \nabla f(x_k)^T h + \frac{1}{2} h^T \nabla^2 f(x_k) h.$$

$\nabla^2 f(x_k)$  is the **Hessian**, the matrix of second partial derivatives.

A positive definite Hessian makes this quadratic strictly convex.

# Newton's multivariate update

Let  $H_k = \nabla^2 f(x_k)$  and  $g_k = \nabla f(x_k)$ . If  $H_k$  is **invertible**,

$$H_k p_k = -g_k \quad \implies \quad p_k = -H_k^{-1} g_k.$$

$$x_{k+1} = x_k - \alpha_k [\nabla^2 f(x_k)]^{-1} \nabla f(x_k).$$

In computation, solve  $H_k p_k = -g_k$ . Positive definiteness guarantees descent for  $g_k \neq 0$ .

# A Newton step in two variables

$$f(x, y) = 2(x - 1)^2 + (x - 1)(y + 1) + (y + 1)^2.$$

$$x_0 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad g_0 = \begin{pmatrix} 6 \\ 5 \end{pmatrix}, \quad H_0 = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}.$$

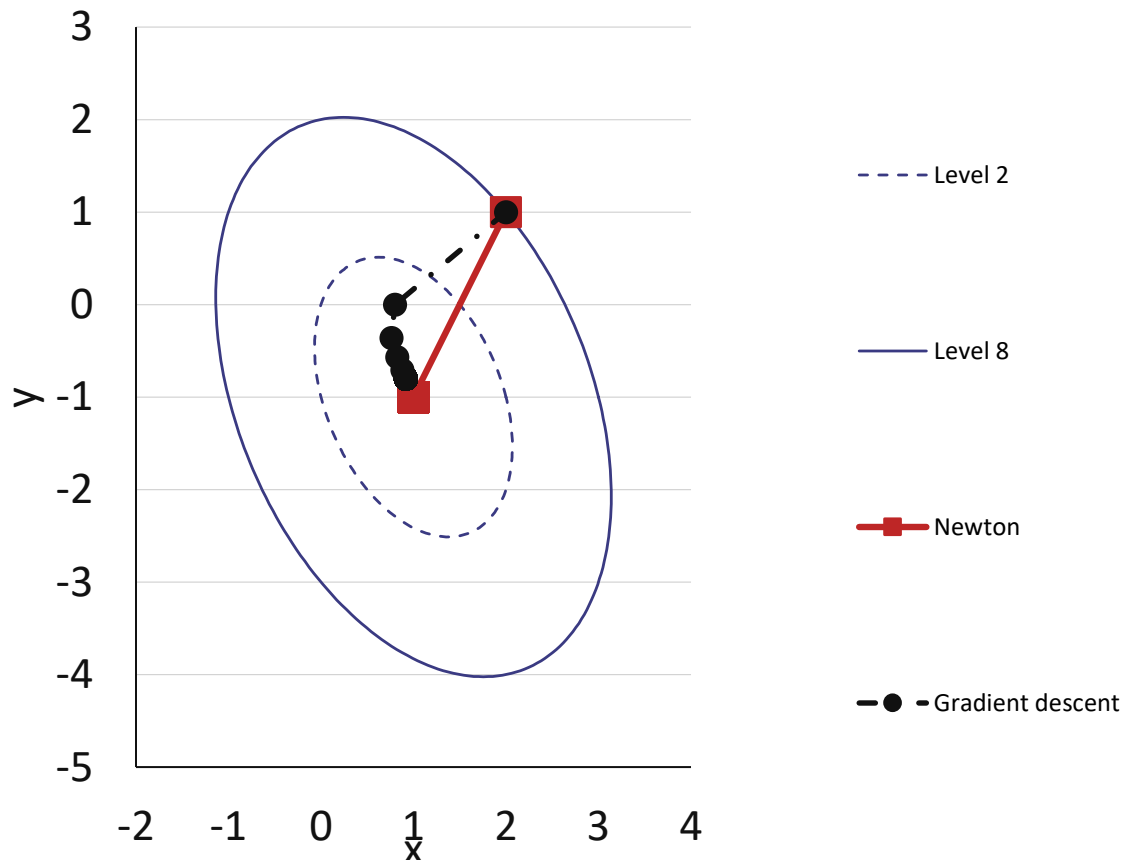
$$H_0 p_0 = -g_0 \quad \Longrightarrow \quad p_0 = \begin{pmatrix} -1 \\ -2 \end{pmatrix}.$$

For the full Newton step,  $\alpha_0 = 1$ :

$$x_1 = x_0 + p_0 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

# Newton and gradient descent

Same starting point  $(2, 1)$ . Gradient descent uses  $\eta = 1/5$ .



# Convergence and computational cost

Near a minimizer with positive curvature, Newton can converge quadratically.

In several variables, each step requires second derivatives and a linear-system solve.

Gradient descent has cheaper steps. Newton can need far fewer steps when curvature information is affordable.

# Reading and next lecture

Chong, Lu, and Zak, An Introduction to Optimization, fifth edition.

**Sections 9.1–9.2:** Newton's method and its analysis.

Review Section 5.6 for Taylor expansions. Section 7.5 gives the root-finding viewpoint.

**Lecture 11:** Diagnostic Midterm 2 solutions.

# Conceptual questions

1. Why can Newton's direction be a descent direction while the full step is unsafe? How does backtracking help?
2. Why does Newton solve a positive definite quadratic in one step, while its fast-convergence guarantee for a general function requires a nearby starting point?