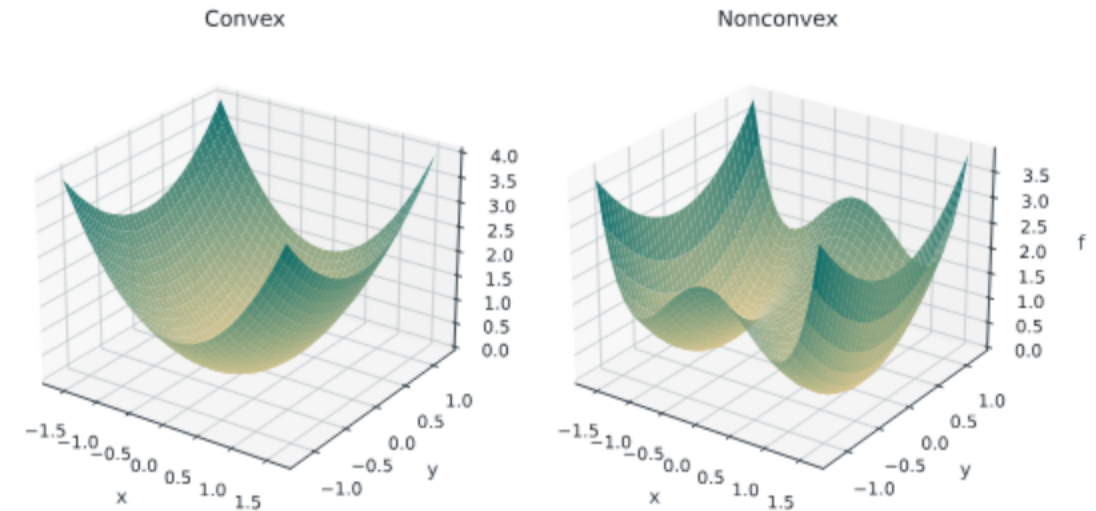


Convexity and unconstrained optimization

NLP session 1

2105623 Optimization of Chemical Processes

Geometry, derivatives, and numerical steps



Learning goals

Classify convex sets and convex functions.

Compute gradient descent and Newton steps.

Check a numerical result using derivatives and geometry.

The three-hour session

60 min: convexity, examples, and short exercises

10 min break, then 75 min: unconstrained methods

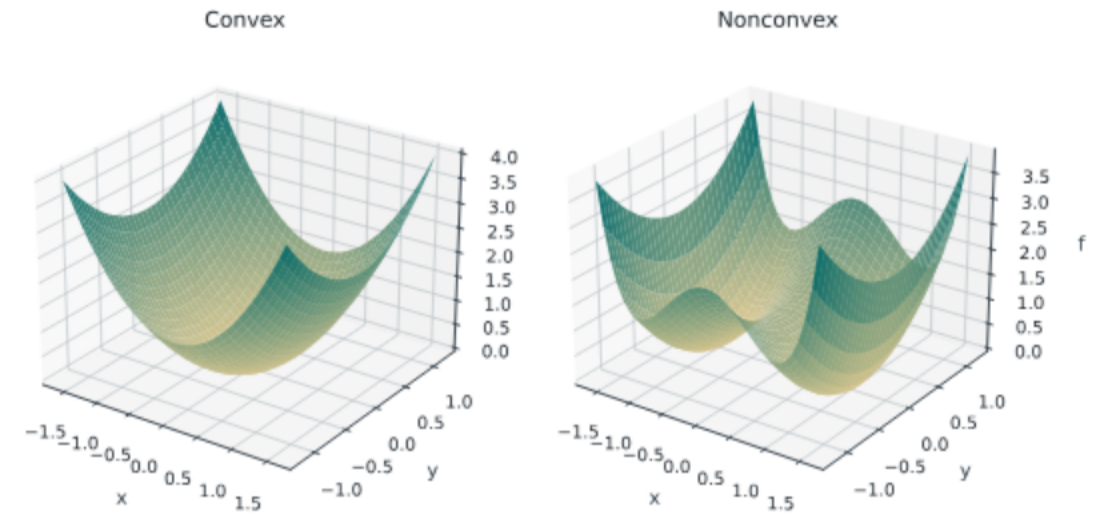
35 min: Python lab, practice, and closing questions

The NLP problem

Minimize a scalar function $f(x)$ over its domain.

Unconstrained: no additional equality or inequality constraints.

Convexity describes geometry, independent of the variable count.

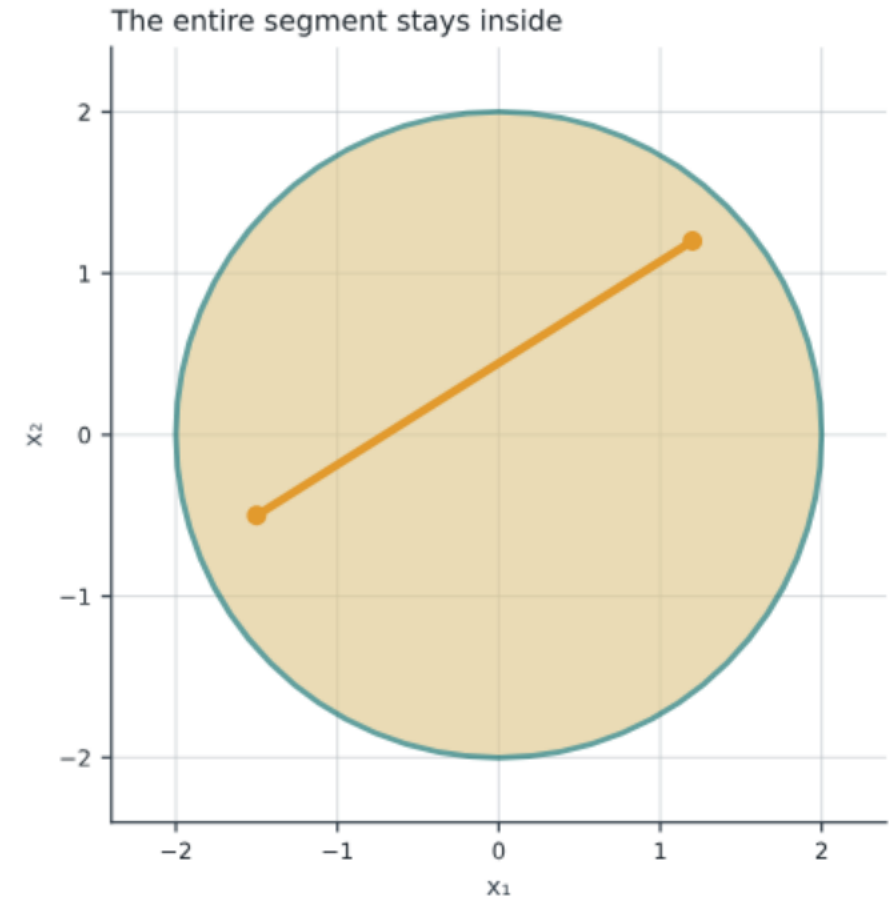


Convex sets

Every segment between two feasible points stays in the set.

$$z = \theta x + (1-\theta)y, \text{ with } 0 \leq \theta \leq 1.$$

Half-spaces, balls, and their intersections are convex.

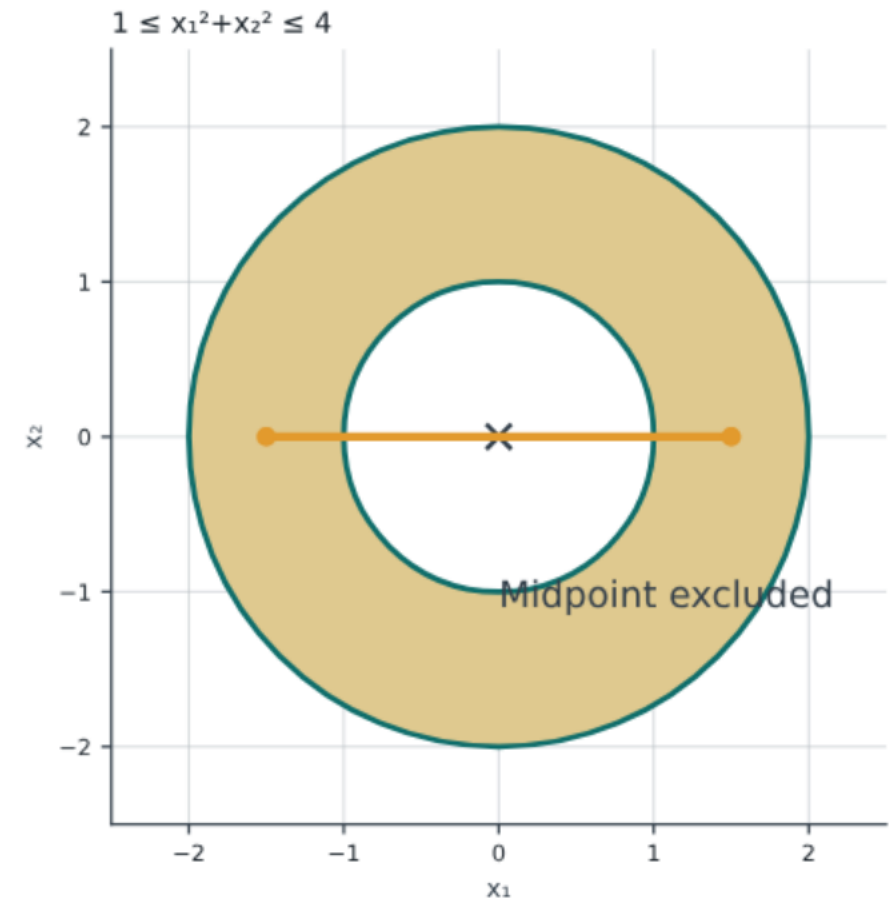


A nonconvex feasible set

Annulus: $1 \leq x_1^2 + x_2^2 \leq 4$.

The points $(-1.5, 0)$ and $(1.5, 0)$ are feasible.

Their midpoint $(0, 0)$ lies in the hole.

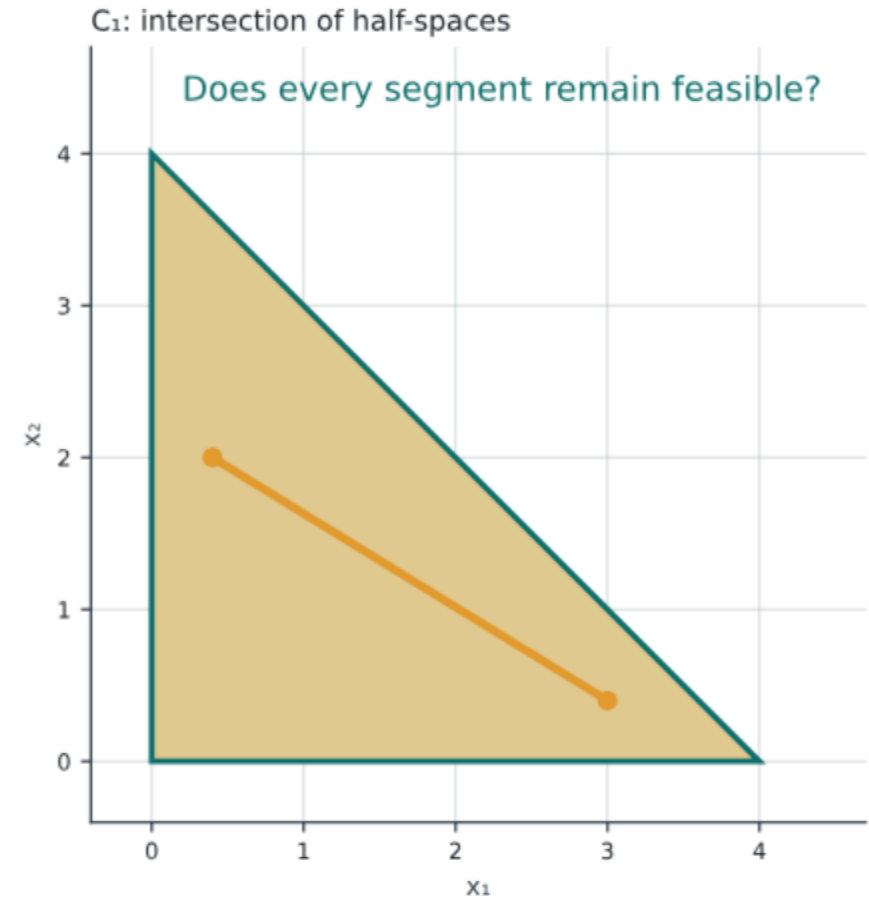


Set exercise

$$C_1: x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 4.$$

$$C_2: x_1^2 + x_2^2 \geq 1.$$

Which set is convex? Give a reason or counterexample.

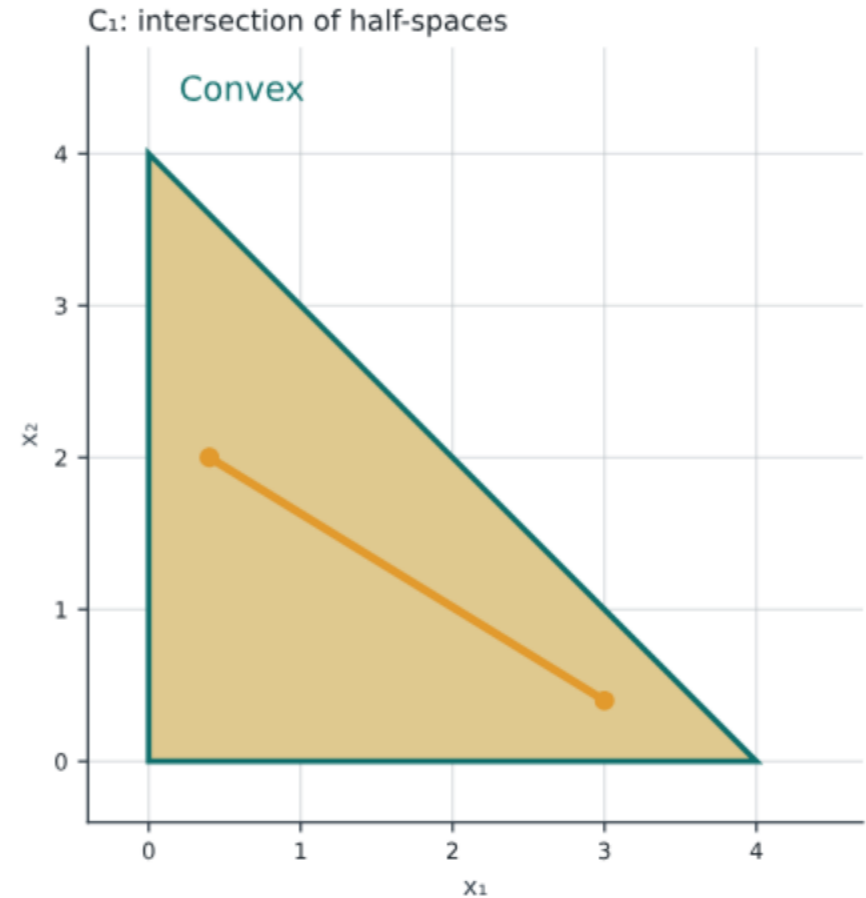


Set exercise: solution

C_1 is the intersection of three half-spaces.

For C_2 , use $p=(-1,0)$ and $q=(1,0)$.

Both pass, but $(p+q)/2=(0,0)$ violates $x_1^2+x_2^2 \geq 1$.

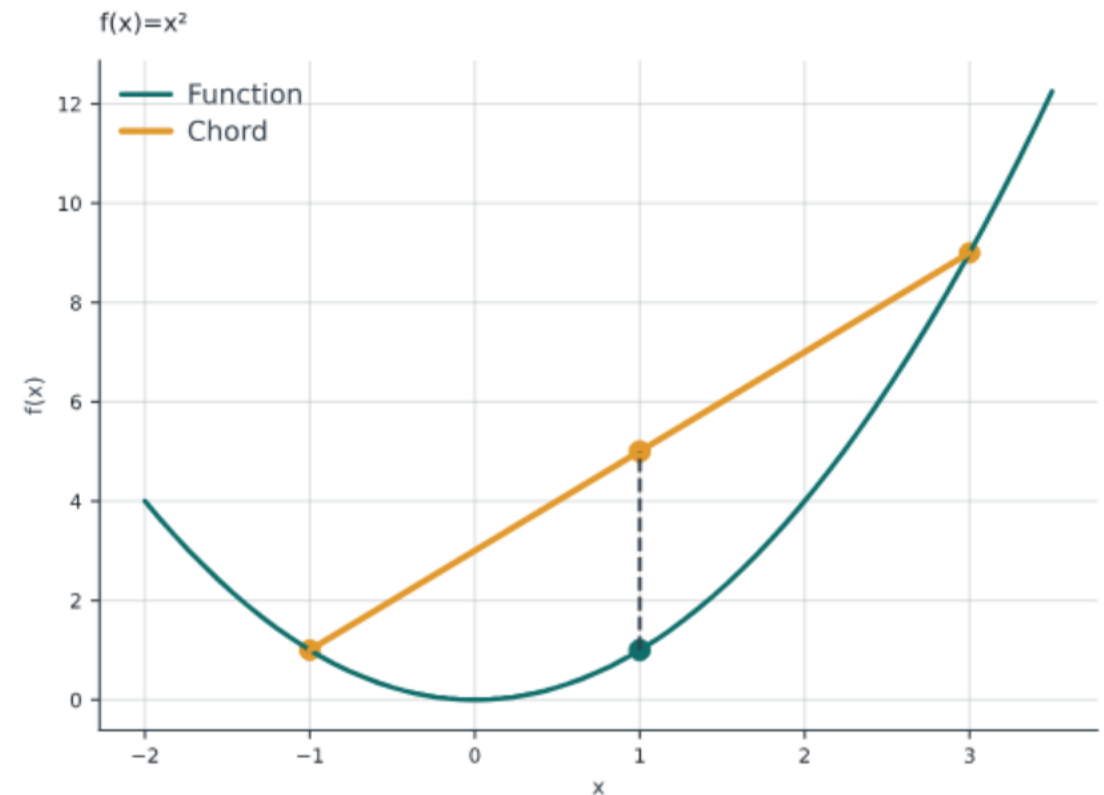


A convex function and its chord

The graph lies below every chord joining two graph points.

The domain must also be convex.

Convexity compares function values along line segments.

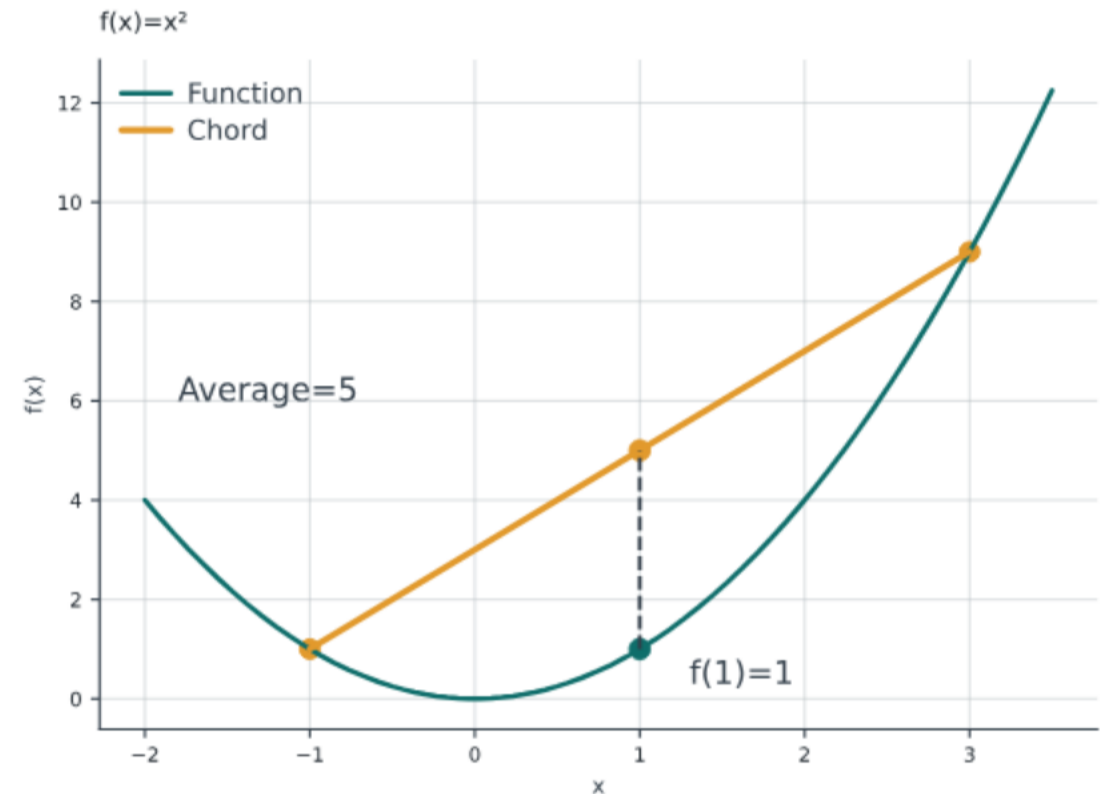


The convexity inequality

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$$

The left side evaluates f at an interpolated input.

The right side interpolates the two endpoint values.

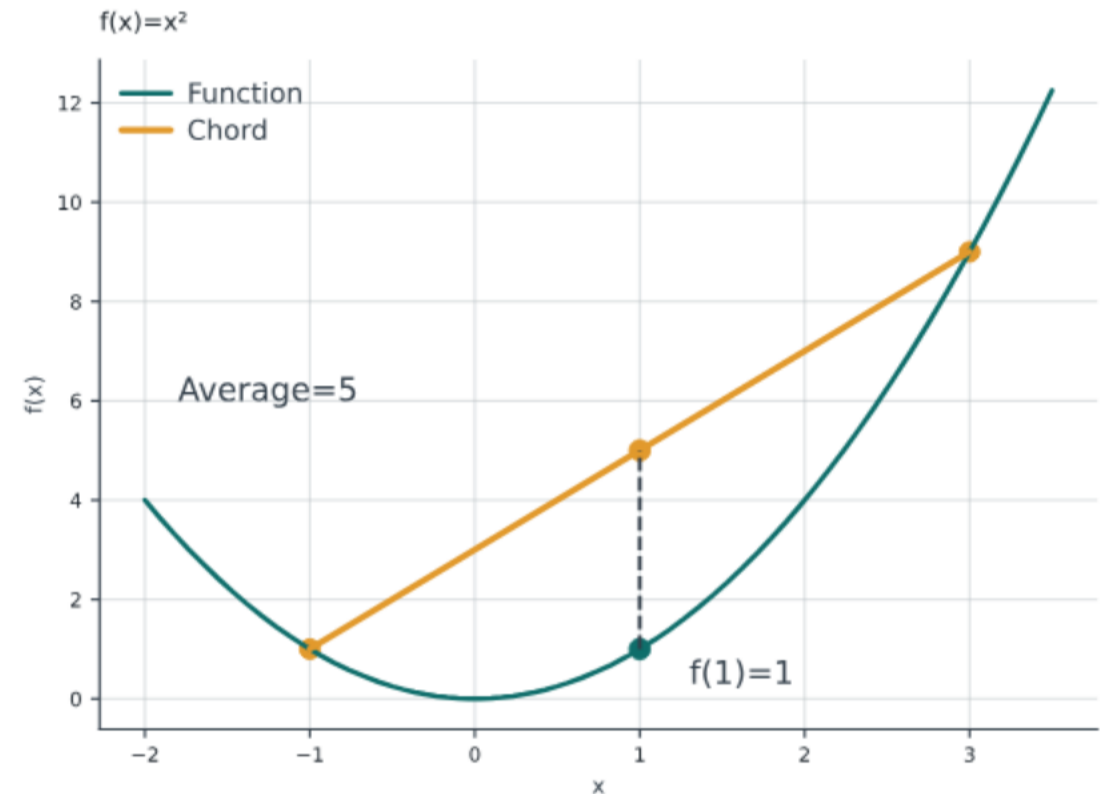


Example: $f(x)=x^2$

Choose $x=-1$, $y=3$, and $\theta=1/2$.

Midpoint value: $f(1)=1$.

Average endpoint value: $[f(-1)+f(3)]/2=5$. Thus $1 \leq 5$.

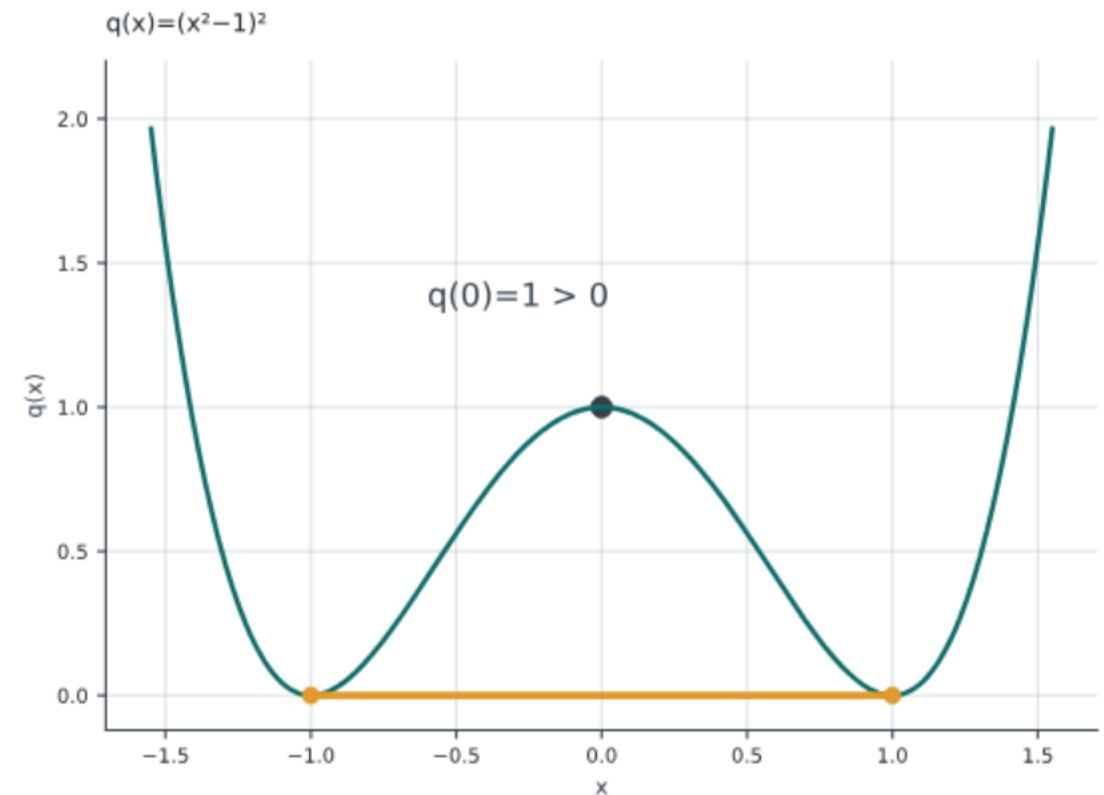


A counterexample for nonconvexity

$q(x)=(x^2-1)^2$ on the real line.

$q(-1)=q(1)=0$, but $q(0)=1$.

The midpoint inequality would require $1 \leq 0$, which is false.

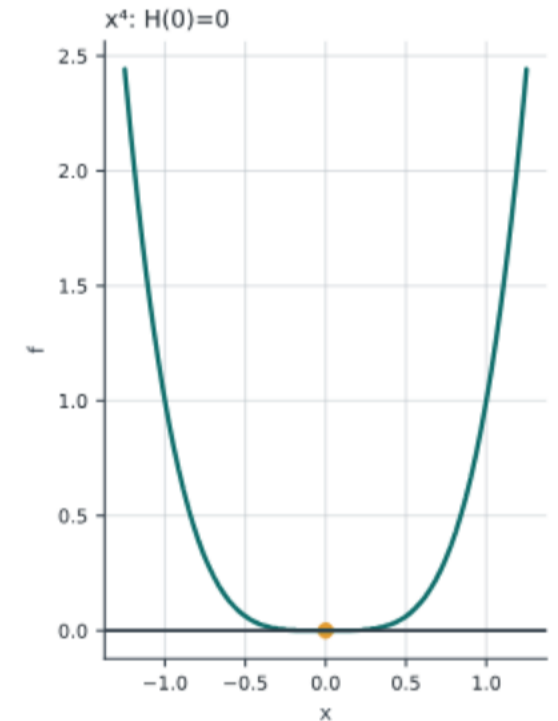
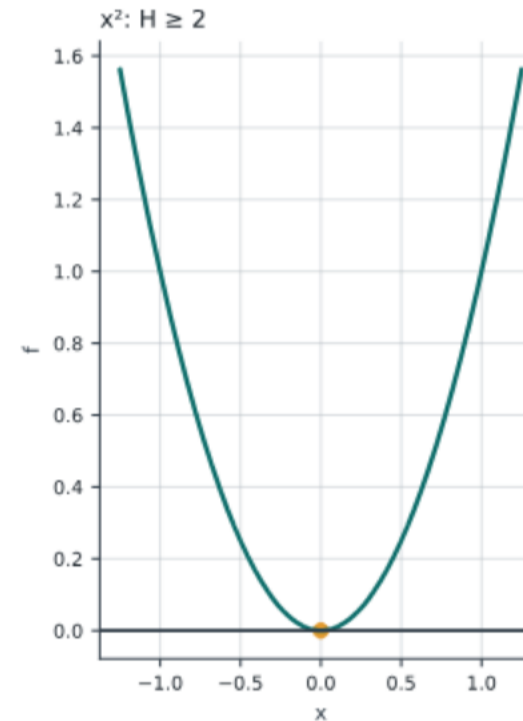


Convex, strictly convex, and strongly convex

Convex: the chord inequality may allow equality.

Strictly convex: strict inequality for distinct endpoints and $0 < \theta < 1$.

For C^2 f : strongly convex if $H(x) \geq m$ everywhere, for a fixed $m > 0$.

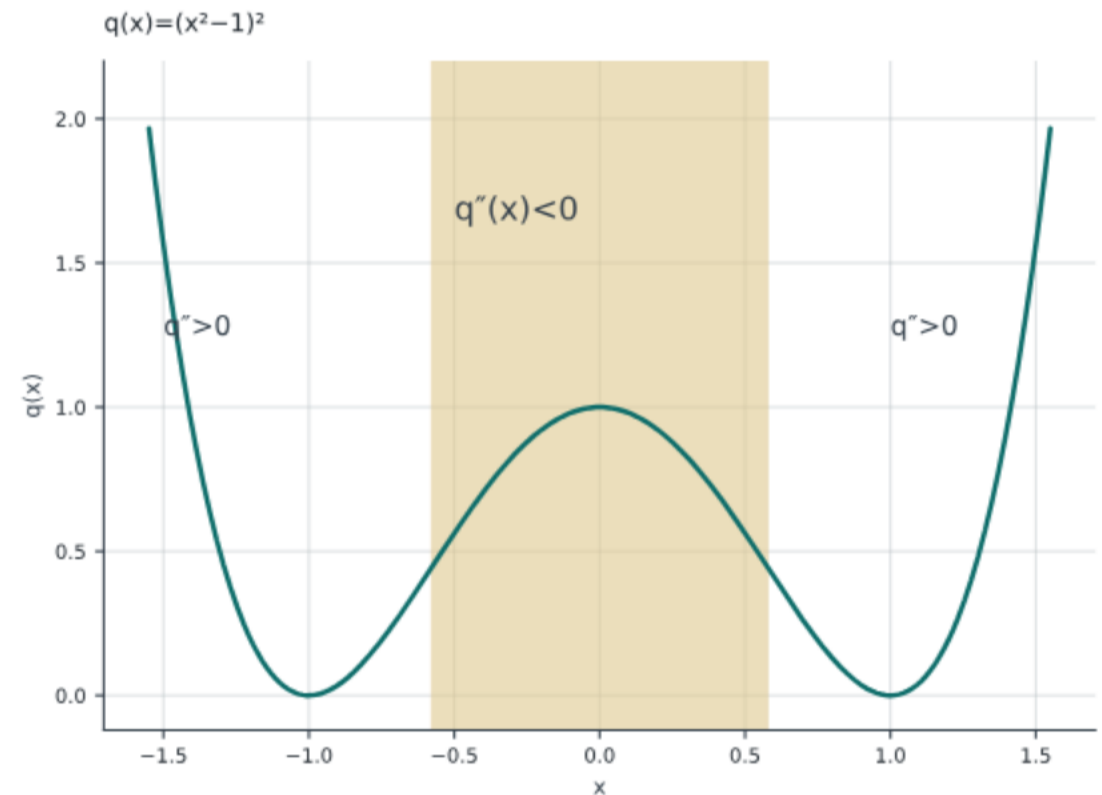


The Hessian test

Assume f is twice continuously differentiable on an open convex domain.

f is convex exactly when $H(x) \geq 0$ at every point in that domain.

Positive curvature at one point alone does not prove global convexity.

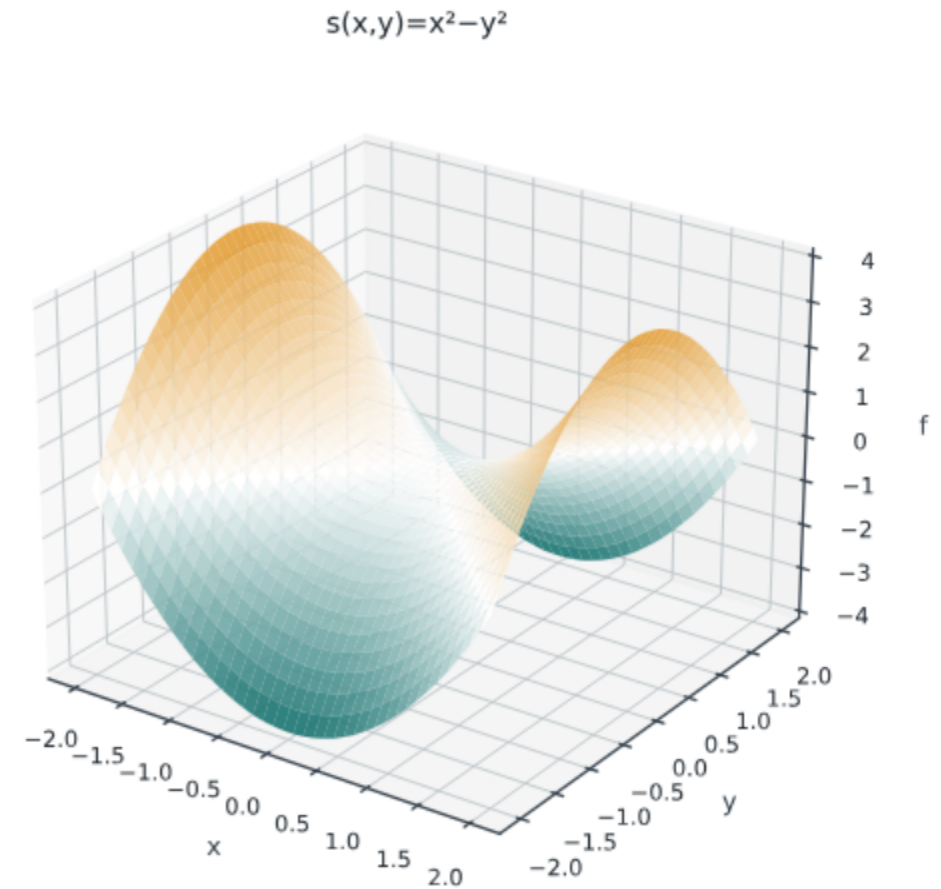


Reading a symmetric Hessian

$H \succeq 0$: $d^T H d \geq 0$ for every direction d .

Equivalently, every eigenvalue is nonnegative.

Mixed eigenvalue signs mean opposite curvature directions.

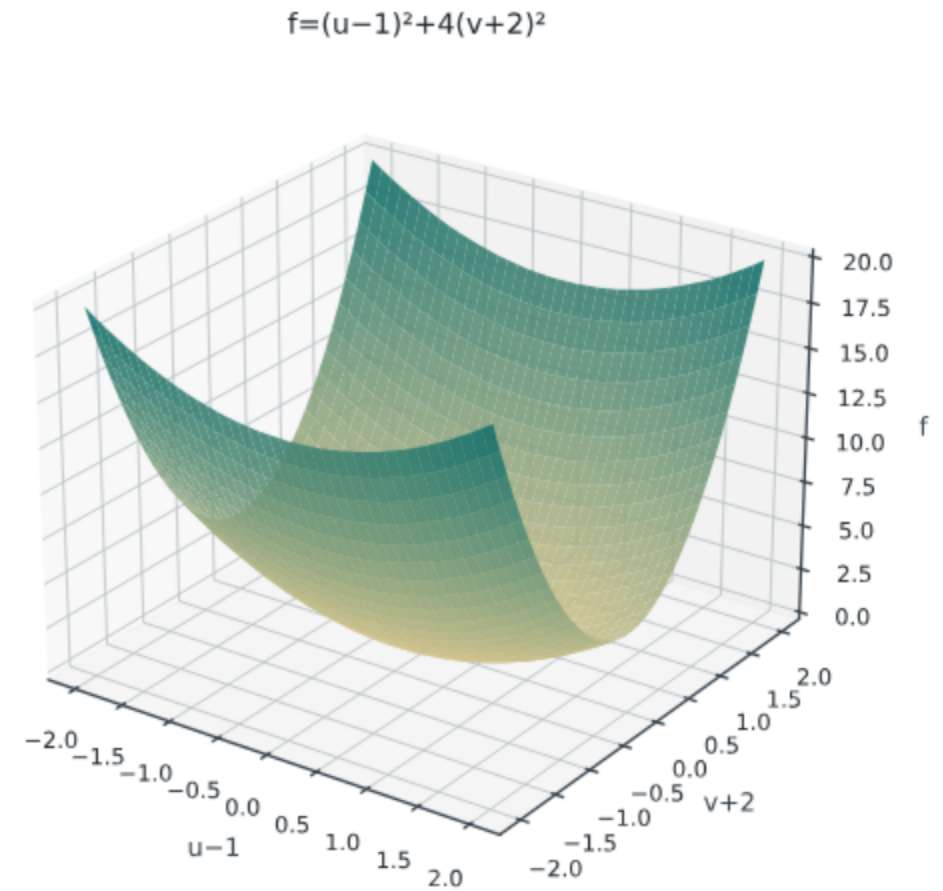


Example: a convex quadratic

$$f(u,v)=(u-1)^2+4(v+2)^2$$

H is constant, with eigenvalues 2 and 8.

Both are positive, so f is strongly convex on $\mathbb{R}^2$.

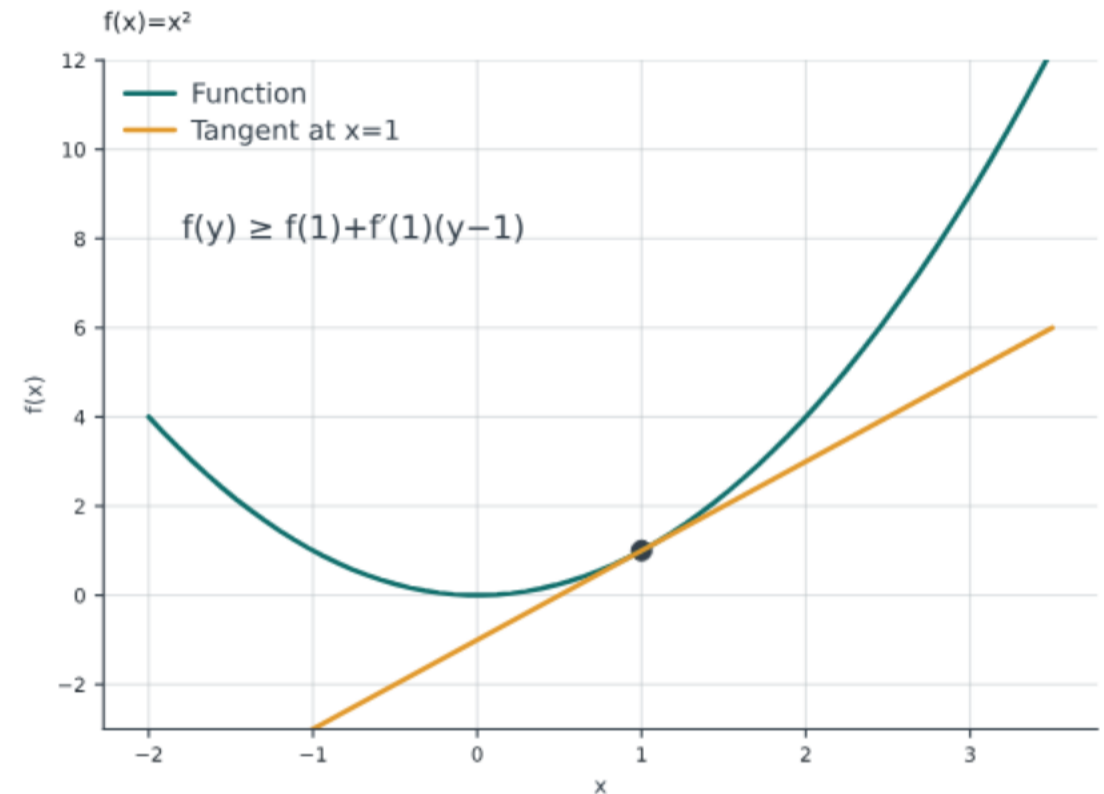


A global certificate from the gradient

For differentiable convex f : $f(y) \geq f(x) + \nabla f(x)^\top (y - x)$.

If $\nabla f(x^*) = 0$, then $f(y) \geq f(x^*)$ for every y in the domain.

Strict convexity gives at most one minimizer.

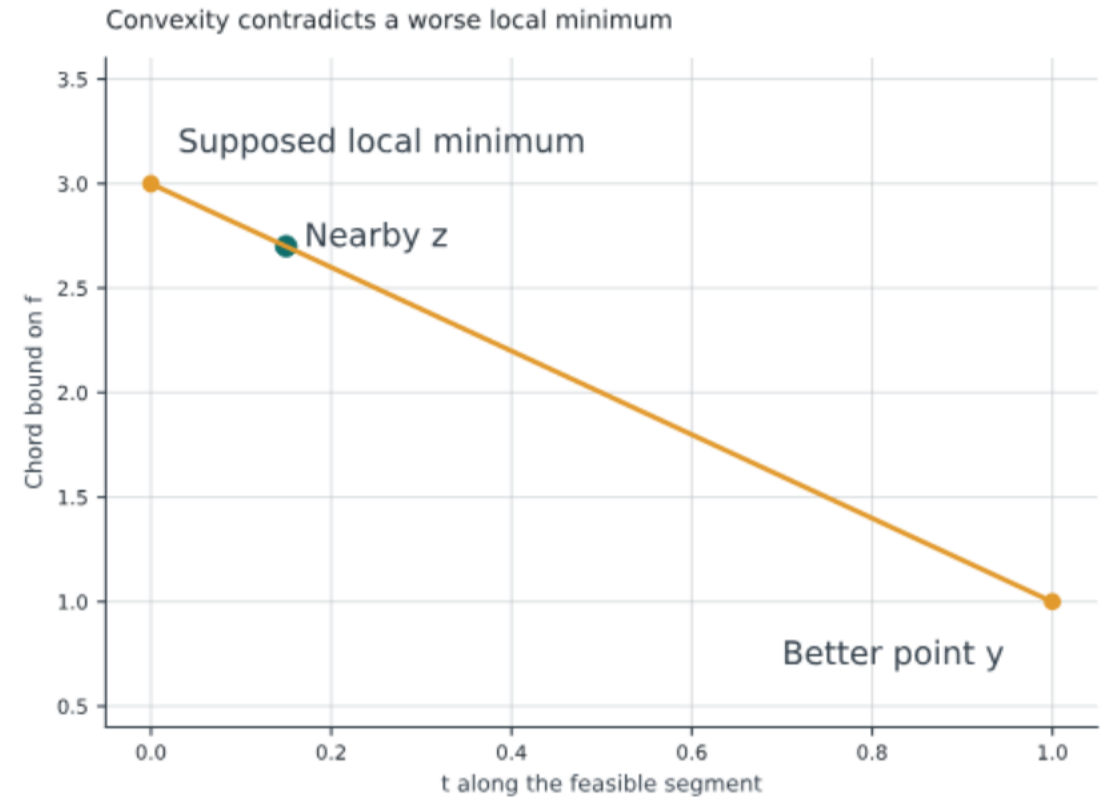


Why local minima are global in a convex problem

Suppose a feasible y has $f(y) < f(x^*)$.

Points $z = (1-t)x^* + ty$ remain feasible and can lie arbitrarily close to x^* .

Convexity gives $f(z) < f(x^*)$, contradicting local minimality.

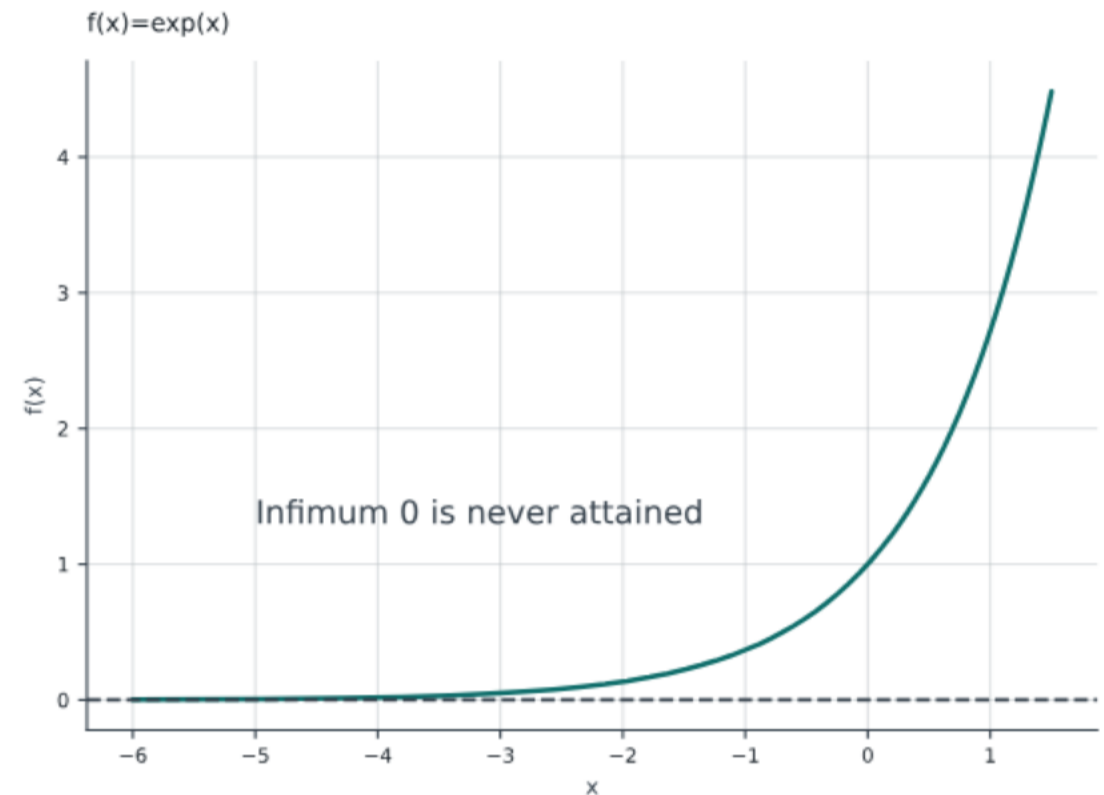


Existence and uniqueness are separate

Strict convexity means at most one minimizer, if one exists.

$f(x)=\exp(x)$ is strictly convex on $\mathbb{R}$.

Its infimum is 0, but no finite x attains 0.



Convex optimization problems

Minimize a convex objective over a convex feasible set.

Standard form: convex inequalities $g_i(x) \leq 0$ and affine equalities.

For maximization, a concave objective over a convex set is the matching case.

Objective

Convex $f(x)$

Inequalities

Convex $g_i(x) \leq 0$

Equalities

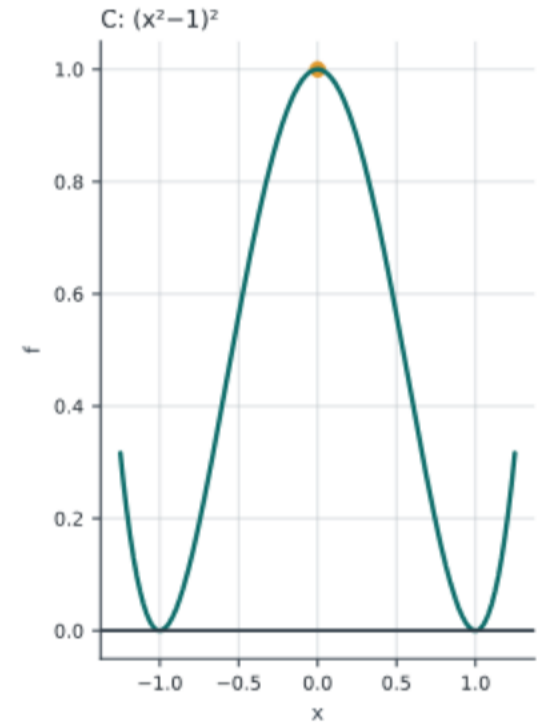
Affine $Ax = b$

Classification exercise

A: minimize x^4 on $\mathbb{R}$. B: minimize $-x^2$ on $\mathbb{R}$.

C: minimize $(x^2-1)^2$ on $\mathbb{R}$.

Classify convexity, existence, and uniqueness of each minimizer.

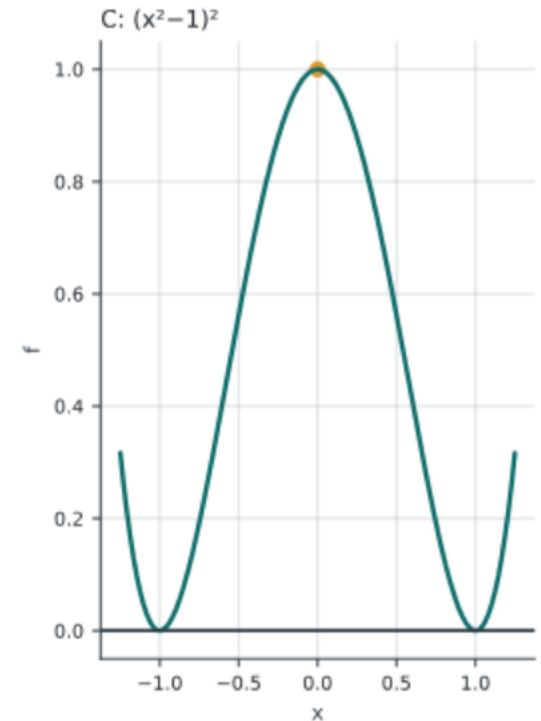


Classification exercise: solution

A: strictly convex, unique global minimum $x=0$, value 0.

B: concave and unbounded below. $x=0$ is a maximum.

C: nonconvex, two global minima $x=-1$ and $x=1$, value 0.

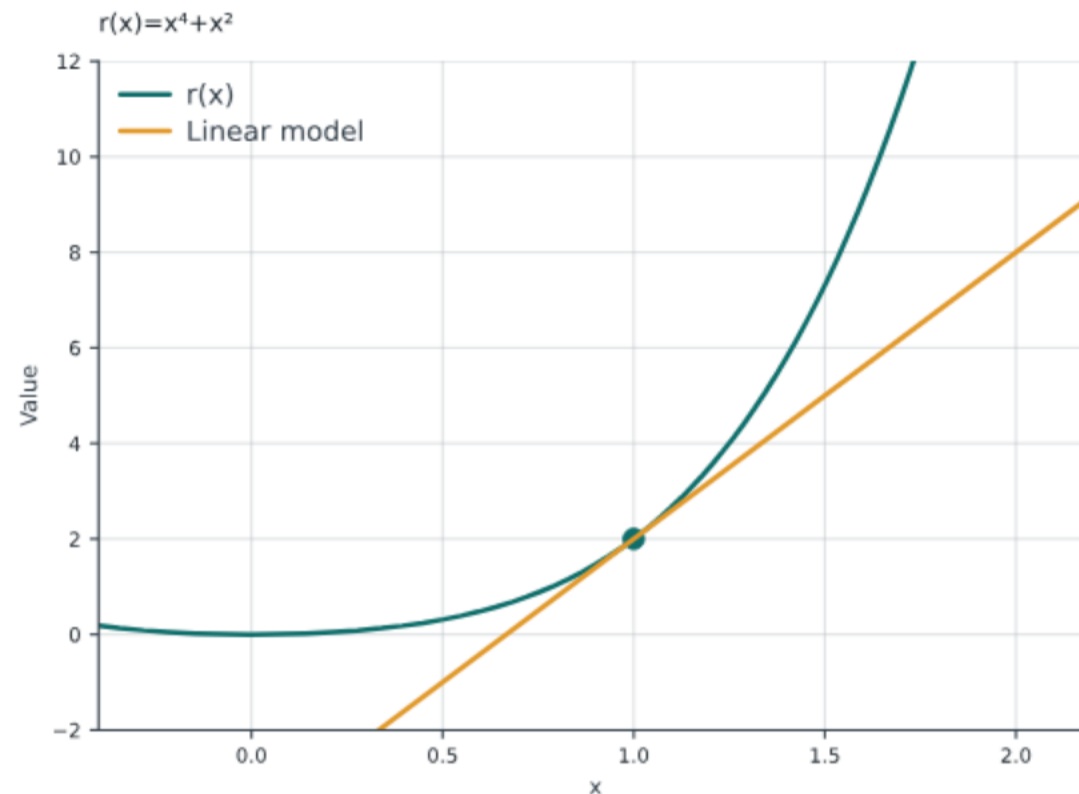


Derivatives and local change

Gradient $g = \nabla f$ collects first derivatives.

$f(x+p) \approx f(x) + g^T p$ for small p .

The Hessian adds the quadratic term $\frac{1}{2}p^T H p$.



Stationary points and curvature

At an interior local minimum of differentiable f , $\nabla f=0$.

At a C^2 local minimum, $H \geq 0$ is necessary.

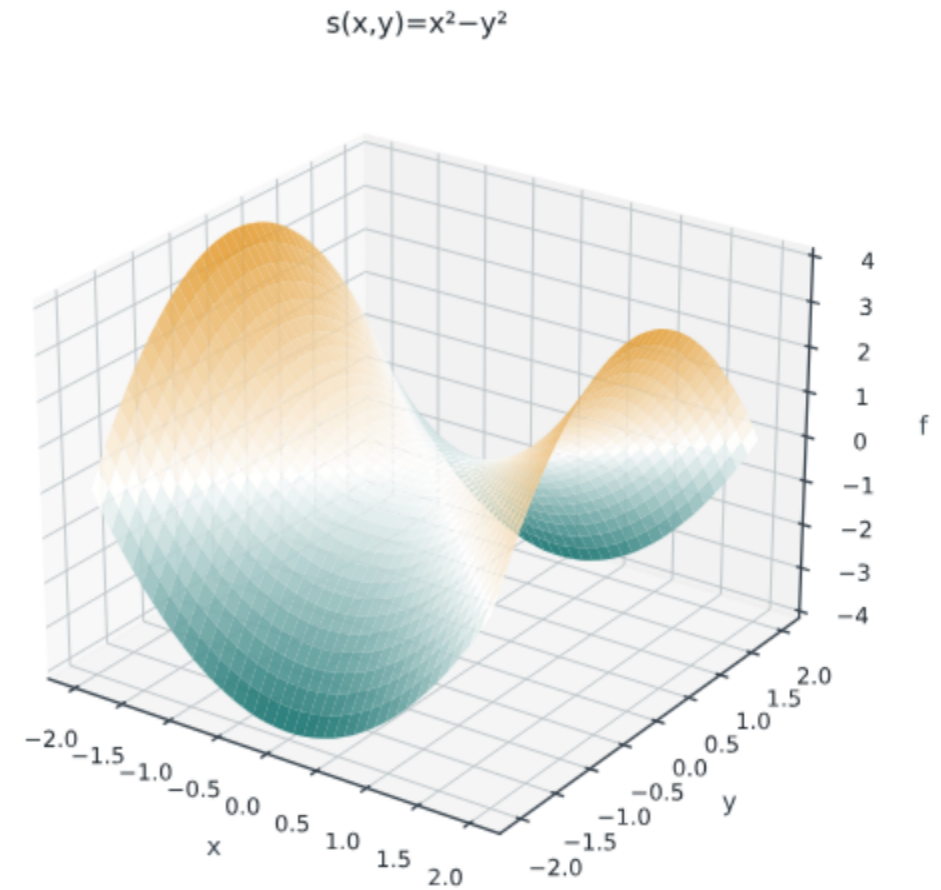
At a stationary point, $H > 0$ is sufficient for a strict local minimum.

A stationary saddle point

$s(x,y)=x^2-y^2$, with $\nabla s(0,0)=0$.

Along $y=0$: $s=x^2>0$ for $x\neq 0$.

Along $x=0$: $s=-y^2<0$ for $y\neq 0$.

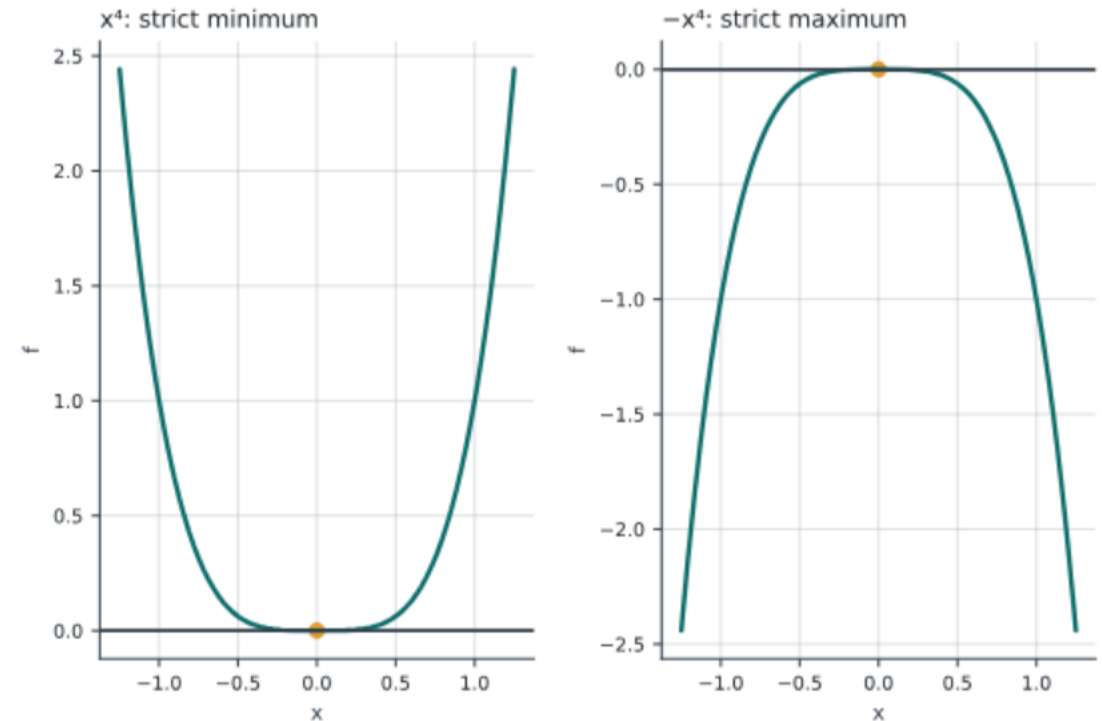


A zero Hessian can be inconclusive

At $x=0$, x^4 and $-x^4$ both have $f'=0$ and $f''=0$.

x^4 has a strict minimum at 0.

$-x^4$ has a strict maximum at 0. Higher-order behavior matters.



The shared quadratic example

$$f(u,v)=(u-1)^2+4(v+2)^2$$

$$\nabla f=(2(u-1), 8(v+2))^T$$

$H=\text{diag}(2,8)$. Initial point $x_0=(3,0)^T$.

$$H = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$$

$$\lambda_1 = 2 > 0 \quad \lambda_2 = 8 > 0$$

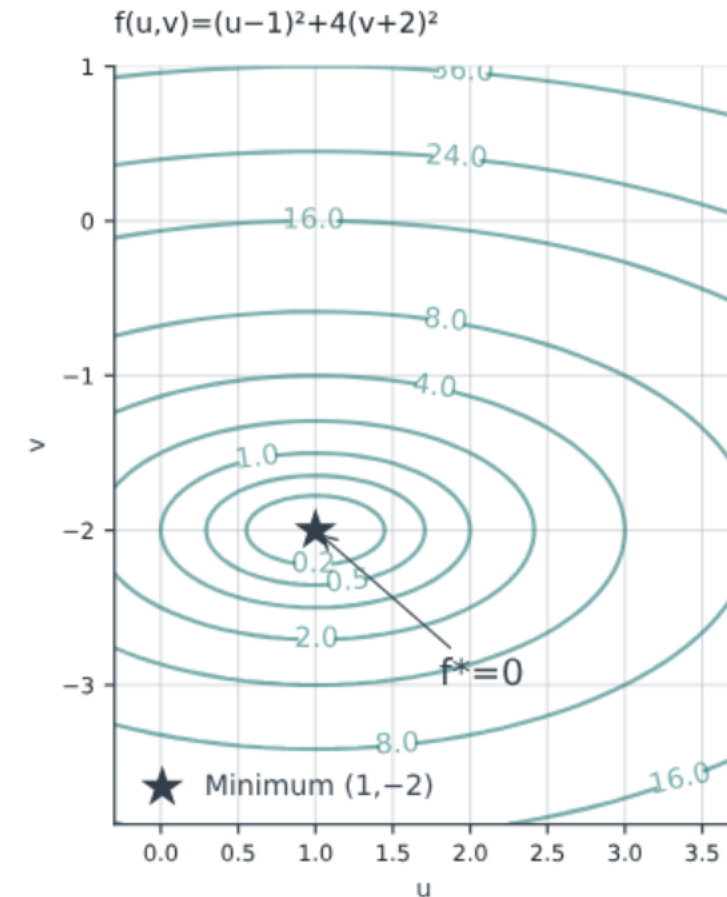
Positive definite
Unique quadratic minimum

The exact solution as a reference

Set $2(u-1)=0$ and $8(v+2)=0$.

The unique solution is $x^*=(1,-2)^T$, with $f(x^*)=0$.

Both squared terms are nonnegative, giving an independent certificate.

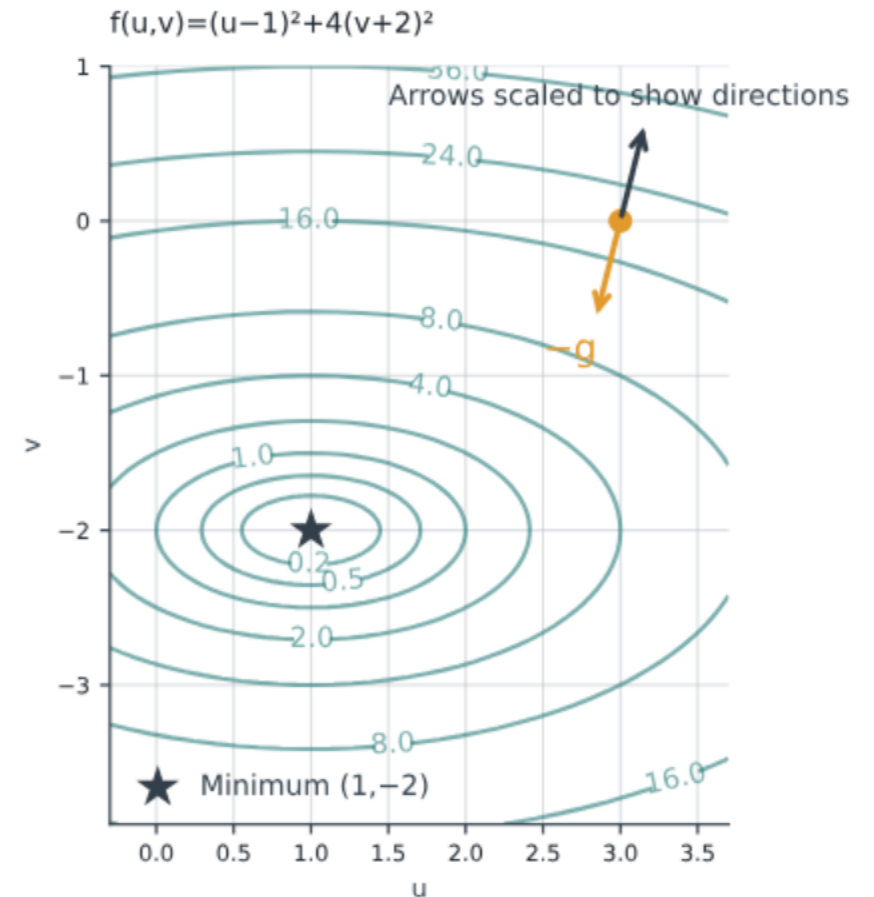


The negative gradient direction

At $x_0=(3,0)$, $g_0=(4,16)^T$.

Choose $p_0=-g_0=(-4,-16)^T$.

The direction is perpendicular to the local contour.

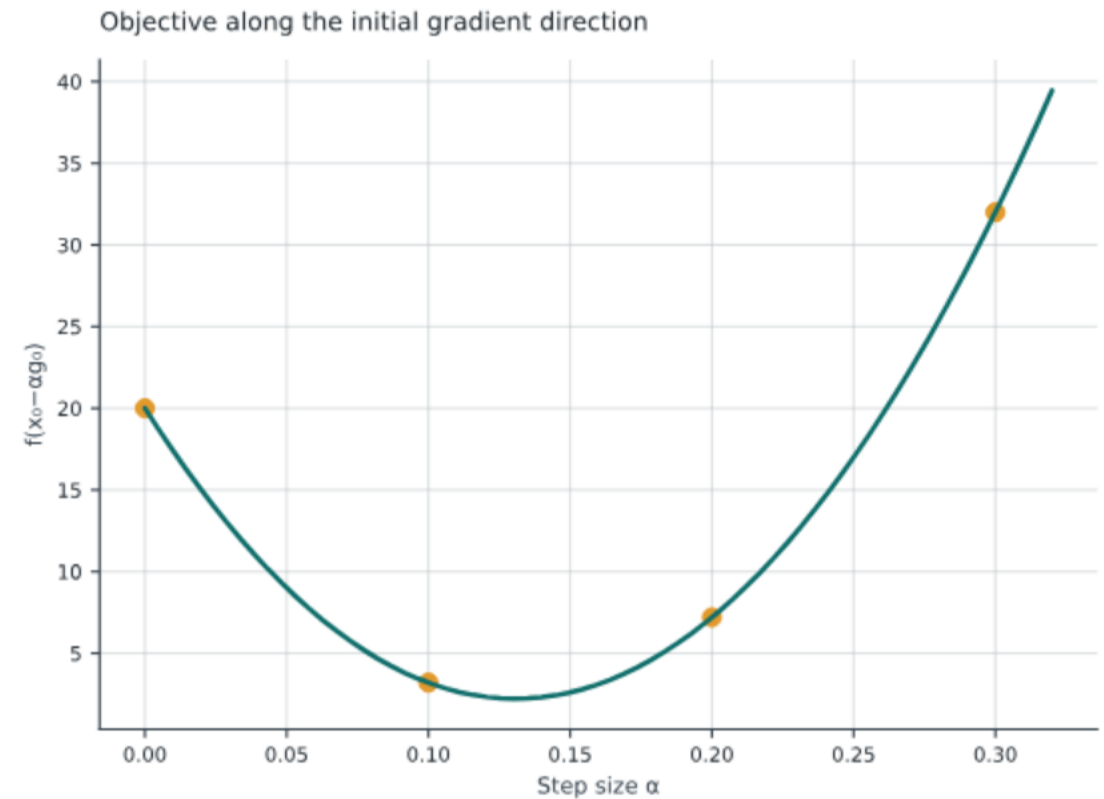


Why $-g$ is a descent direction

Directional derivative: $g^T p$.

For $p = -g \neq 0$, $g^T p = -\|g\|^2 < 0$.

Here $g_0^T p_0 = -4^2 - 16^2 = -272$. Small positive steps reduce f .

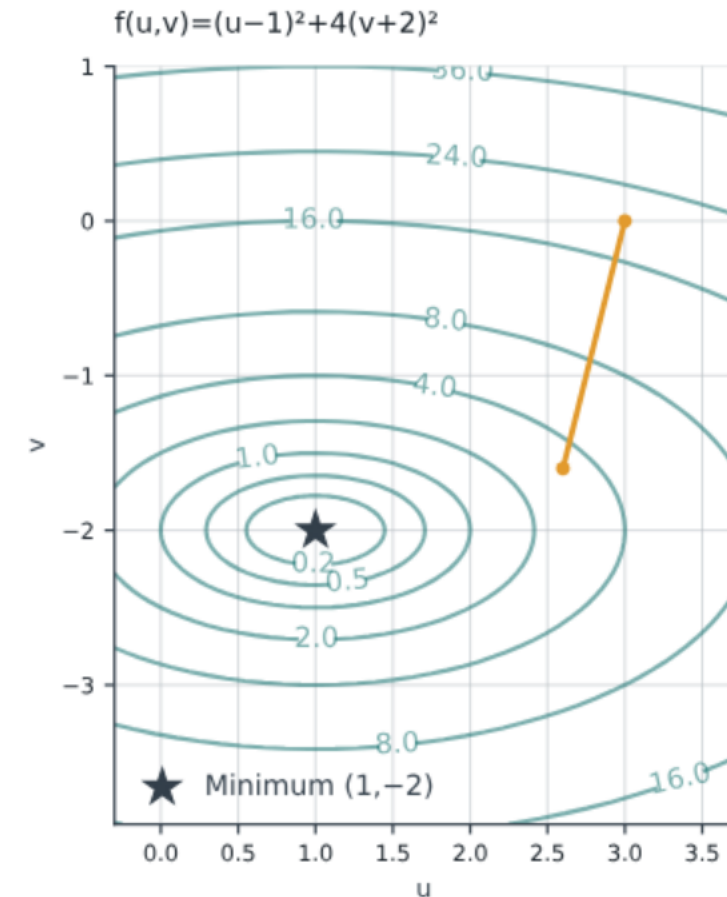


The iterative update

$$x_{k+1} = x_k + \alpha_k p_k, \text{ with } \alpha_k > 0.$$

The direction p_k and the step size α_k play different roles.

Today: fixed-step gradient descent and full Newton steps on selected examples.

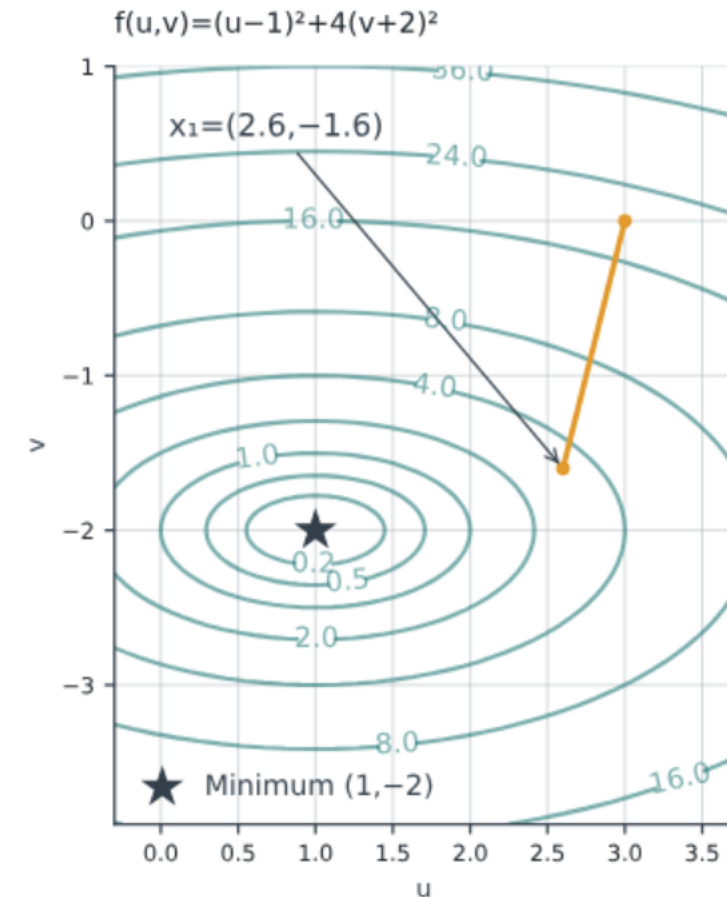


Gradient descent: first step

At $x_0=(3,0)$, $f_0=4+16=20$ and $g_0=(4,16)$.

With $\alpha=0.1$: $x_1=(3,0)-0.1(4,16)=(2.6,-1.6)$.

$f_1=1.6^2+4(0.4^2)=2.56+0.64=3.20$.

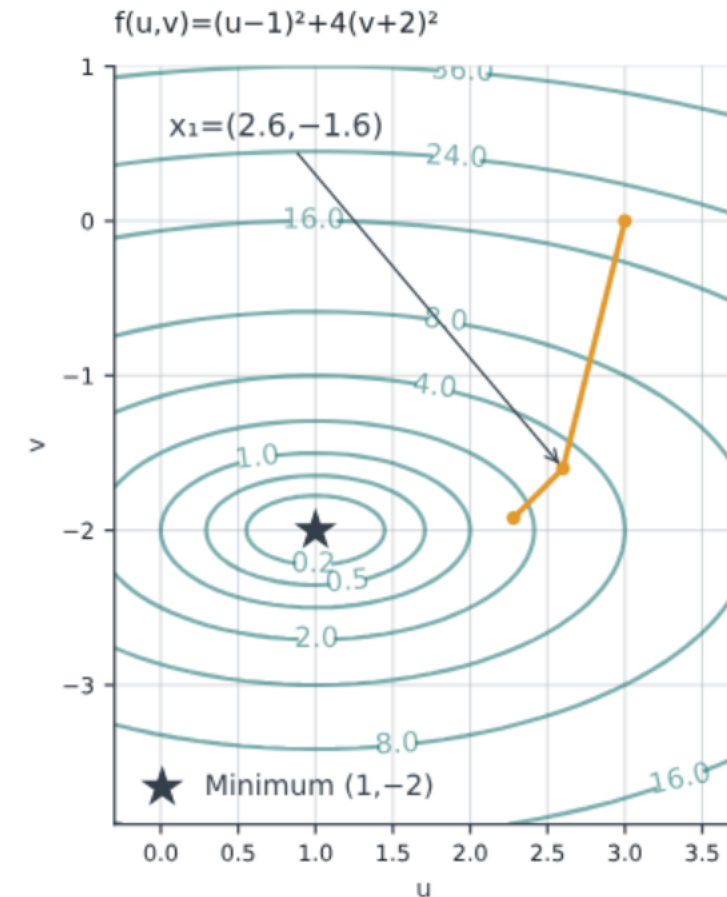


Gradient descent: second step

At $x_1=(2.6,-1.6)$, $g_1=(3.2,3.2)$.

$x_2=(2.6,-1.6)-0.1(3.2,3.2)=(2.28,-1.92)$.

$f_2=1.28^2+4(0.08^2)=1.6384+0.0256=1.6640$.

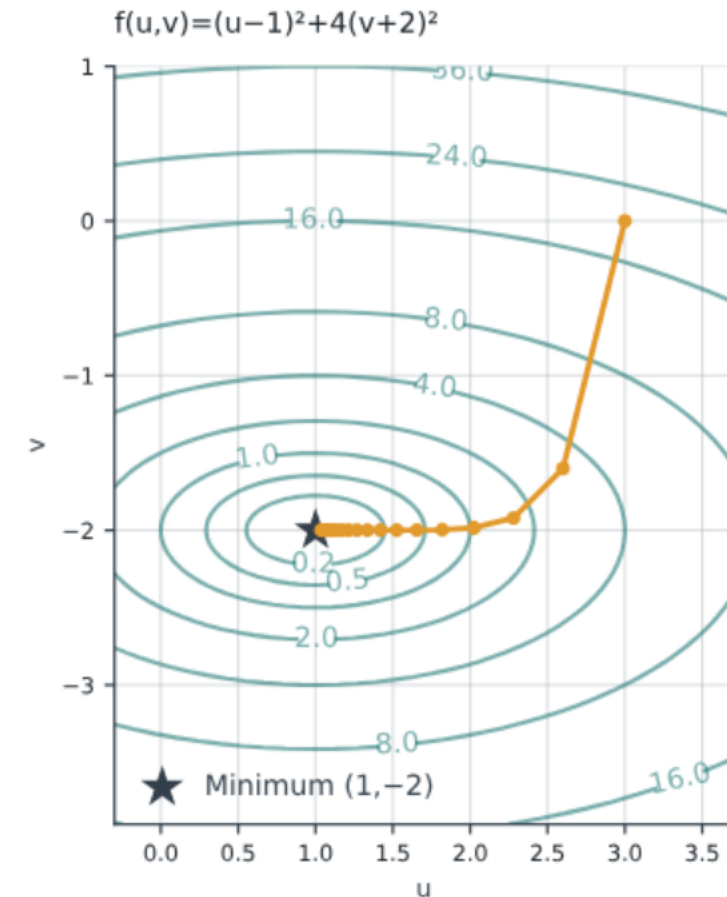


The gradient descent path

With $\alpha=0.1$, the v error contracts faster than the u error.

The path approaches $(1, -2)$, while the gradient approaches zero.

A decreasing objective alone does not identify the type of stationary point.

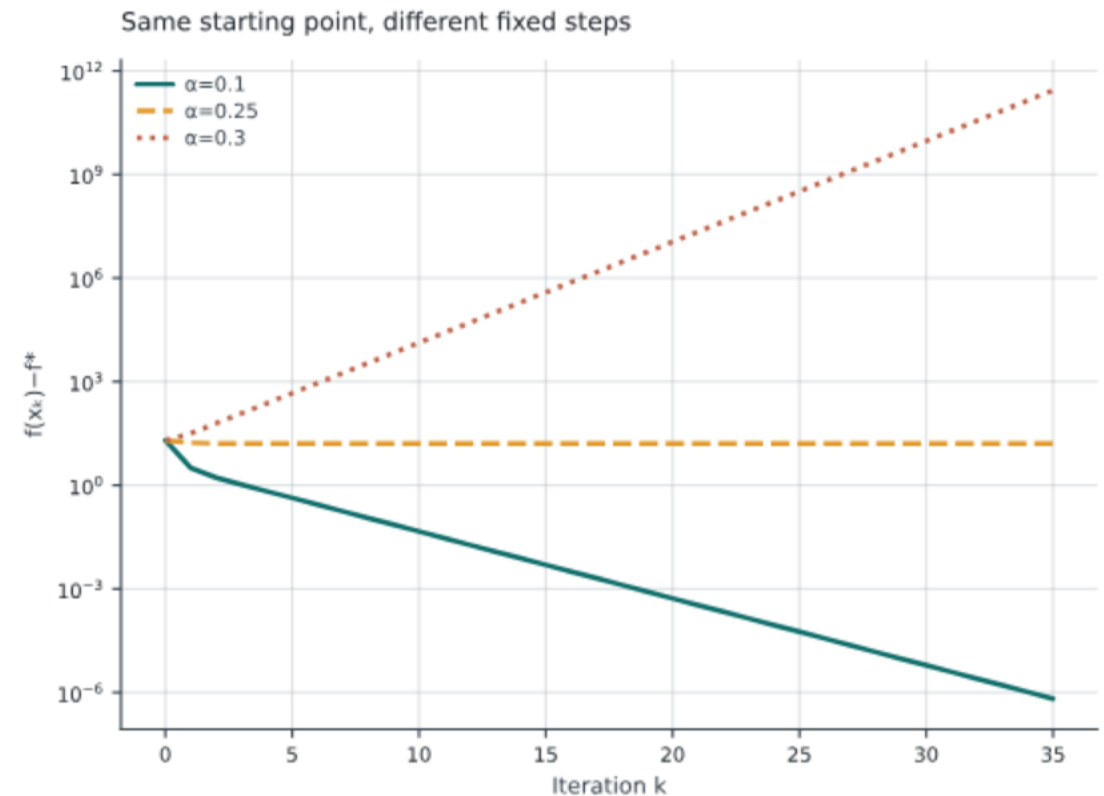


Step size changes the behavior

For this quadratic, convergence from every start requires $0 < \alpha < 0.25$.

At $\alpha=0.25$, the v error changes sign without shrinking.

At $\alpha=0.30$, a nonzero v error grows in magnitude.

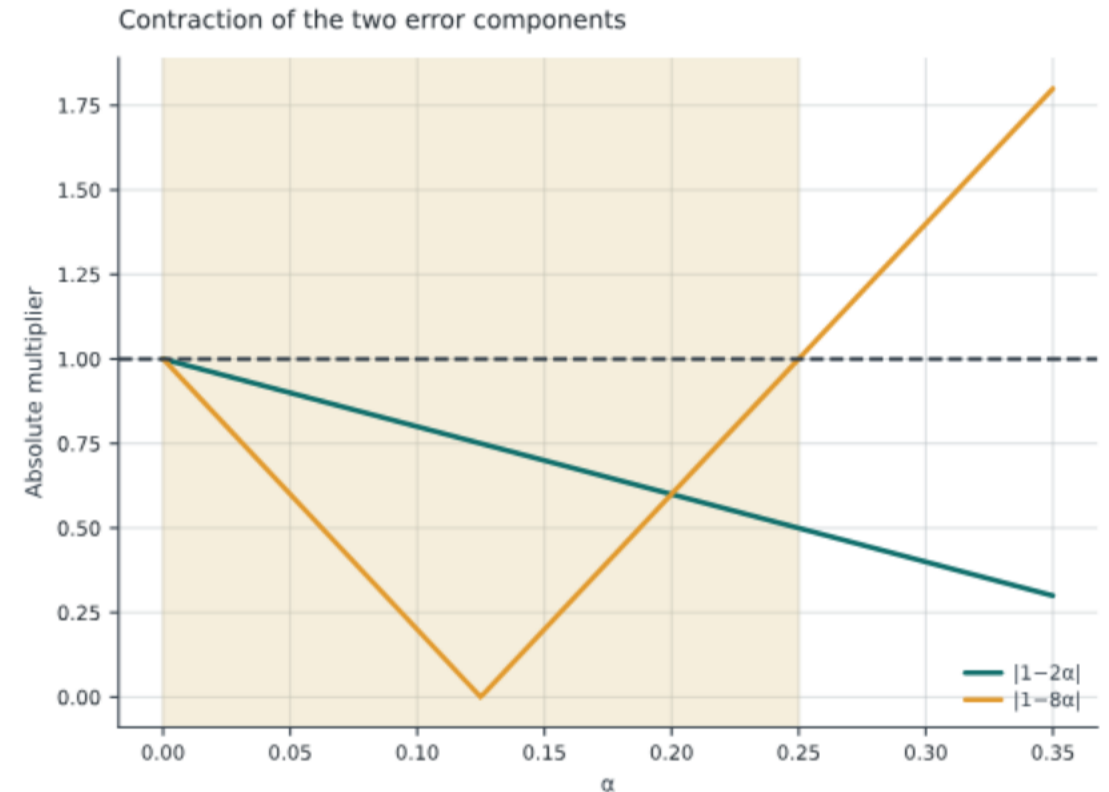


Derivation of the step-size interval

Let $e_u = u - 1$ and $e_v = v + 2$.

$e_{u,\text{new}} = (1 - 2\alpha)e_u$ and $e_{v,\text{new}} = (1 - 8\alpha)e_v$.

Both errors shrink when $|1 - 2\alpha| < 1$ and $|1 - 8\alpha| < 1$, giving $0 < \alpha < 1/4$.

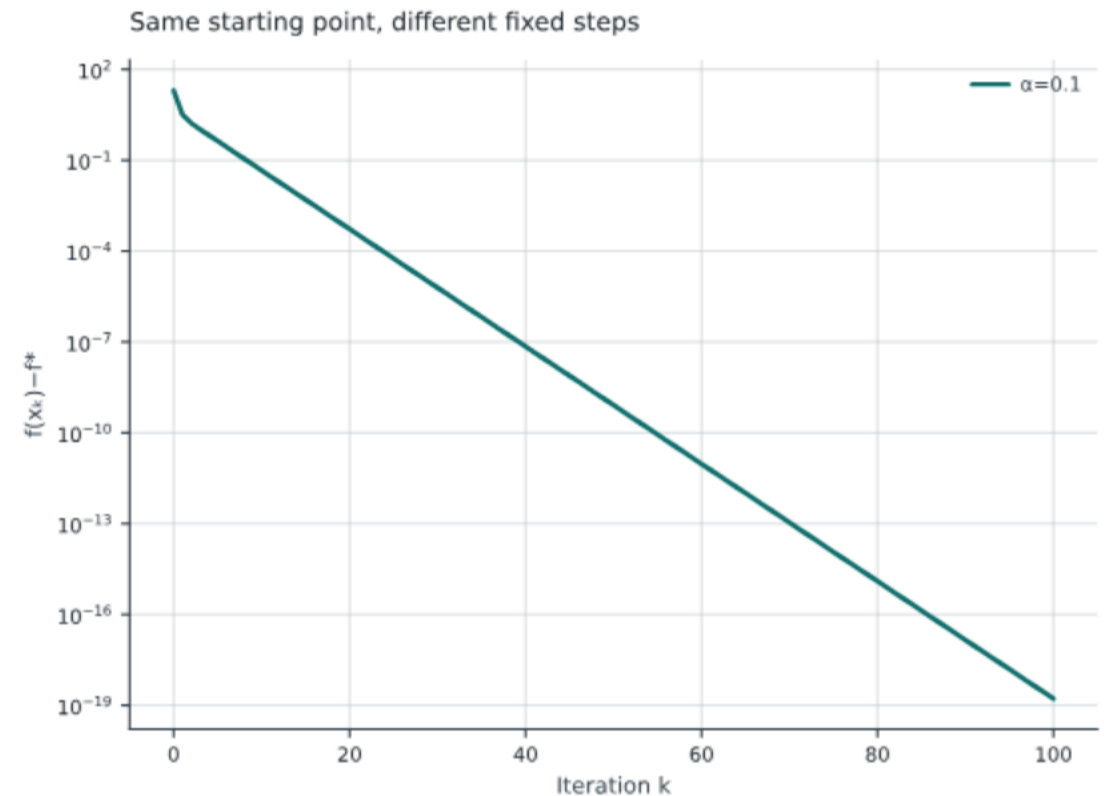


Stopping and checking a result

Use a gradient tolerance, an iteration limit, and finite-value checks.

Small steps can also reflect poor scaling or stagnation.

Report x , $f(x)$, $\|\nabla f(x)\|$, termination reason, and curvature.

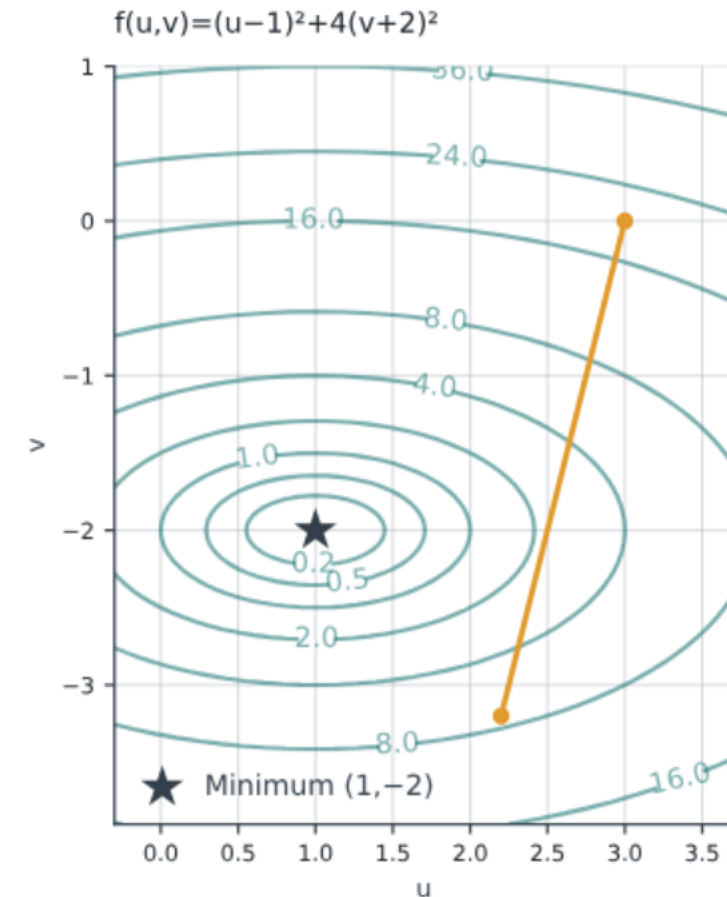


Practice: a larger gradient step

Use the same $x_0=(3,0)$ and $g_0=(4,16)$.

Take one step with $\alpha=0.2$.

Calculate x_1 and $f(x_1)$. Did either coordinate cross the minimizer?

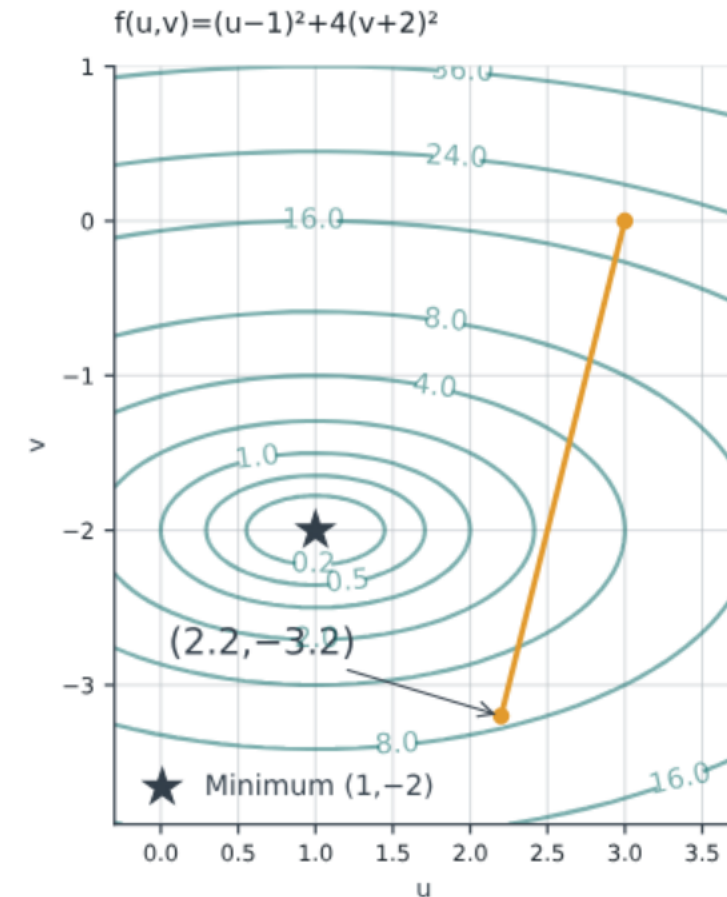


Practice solution: $\alpha=0.2$

$$x_1 = (3, 0) - 0.2(4, 16) = (2.2, -3.2).$$

$$f(x_1) = 1.2^2 + 4(-1.2)^2 = 1.44 + 5.76 = 7.20.$$

v crosses -2 , but f falls from 20 to 7.2.

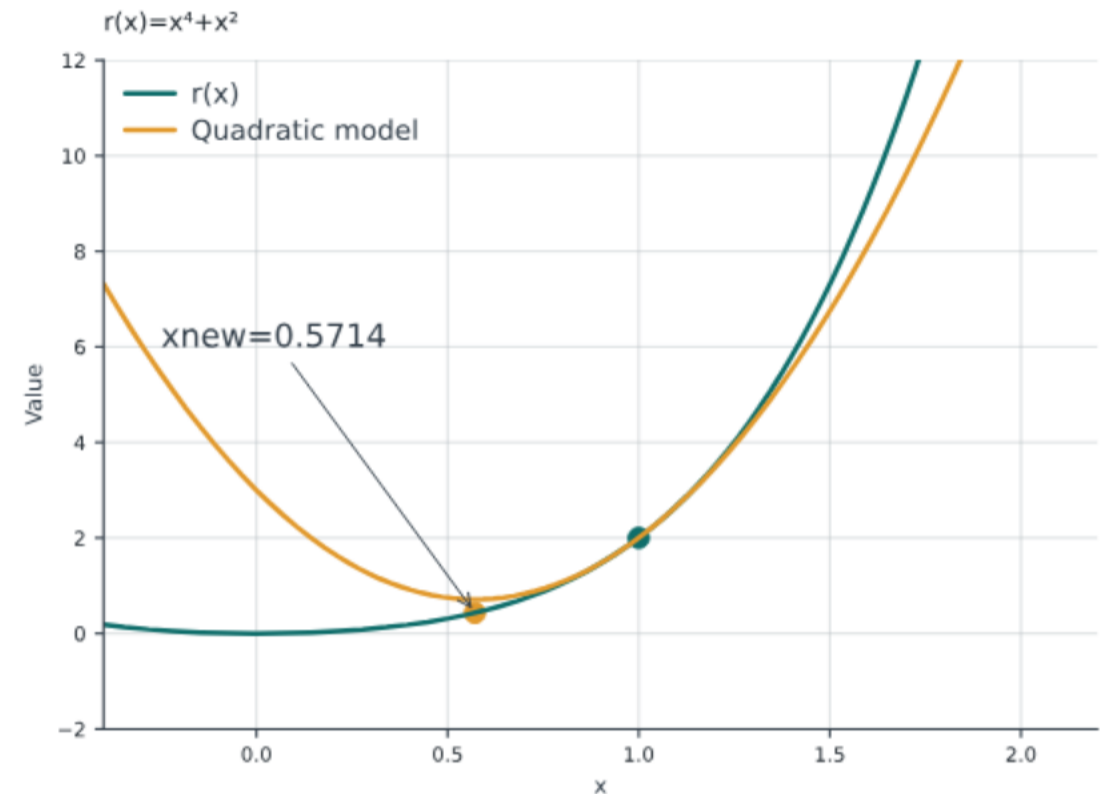


The Newton quadratic model

$$m(p) = f(x) + g^T p + \frac{1}{2} p^T H p$$

Stationarity of m gives $g + H p = 0$.

Solve $H p = -g$ for p , then form $x_{\text{new}} = x + p$.



Newton direction at the initial point

At $x_0=(3,0)$, $H=\text{diag}(2,8)$ and $g_0=(4,16)$.

The two equations are $2p_u=-4$ and $8p_v=-16$.

$$H = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$$

$$\lambda_1 = 2 > 0 \quad \lambda_2 = 8 > 0$$

Therefore $p=(-2,-2)$. Solve the system instead of forming H^{-1} .

$$2p_1 = -4 \Rightarrow p_1 = -2$$

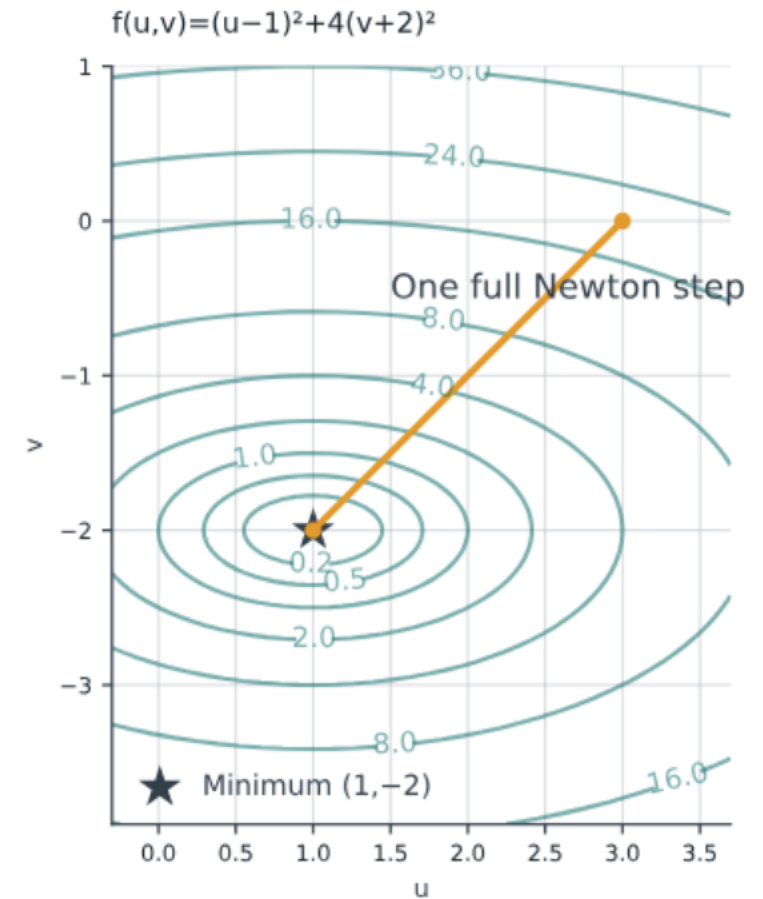
$$8p_2 = -16 \Rightarrow p_2 = -2$$

Newton update and verification

$$x_1 = (3, 0) + (-2, -2) = (1, -2).$$

$$f(x_1) = 0 \text{ and } \nabla f(x_1) = (0, 0).$$

This positive-definite quadratic reaches its exact minimizer in one full step.

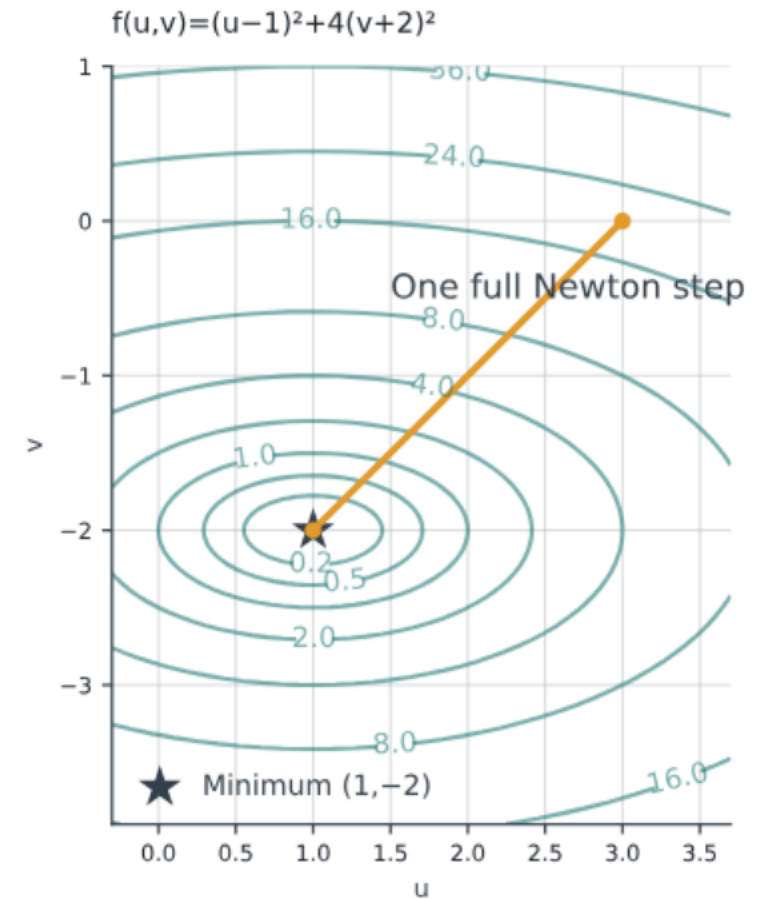


When the one-step result applies

For $f(x) = \frac{1}{2}x^T H x + c^T x + r$ with constant $H > 0$:

$H p = -(H x + c)$ implies $x + p = -H^{-1}c = x^*$.

Nonquadratic functions have higher-order terms, so further iterations are usually needed.

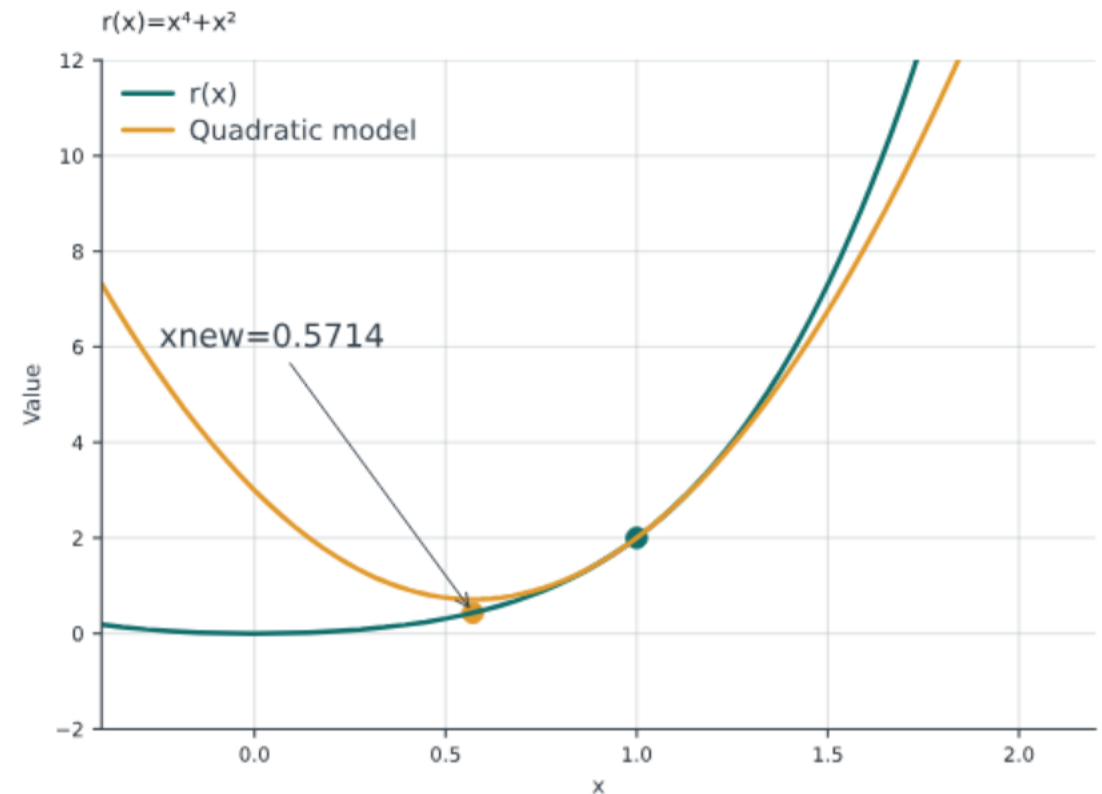


Newton on a nonquadratic function

$r(x)=x^4+x^2$. At $x_0=1$: $r'=6$ and $r''=14$.

$p=-6/14=-3/7$, so $x_1=4/7$.

$r(x_1)=1040/2401\approx0.433153$, down from 2. The minimum is still at 0.

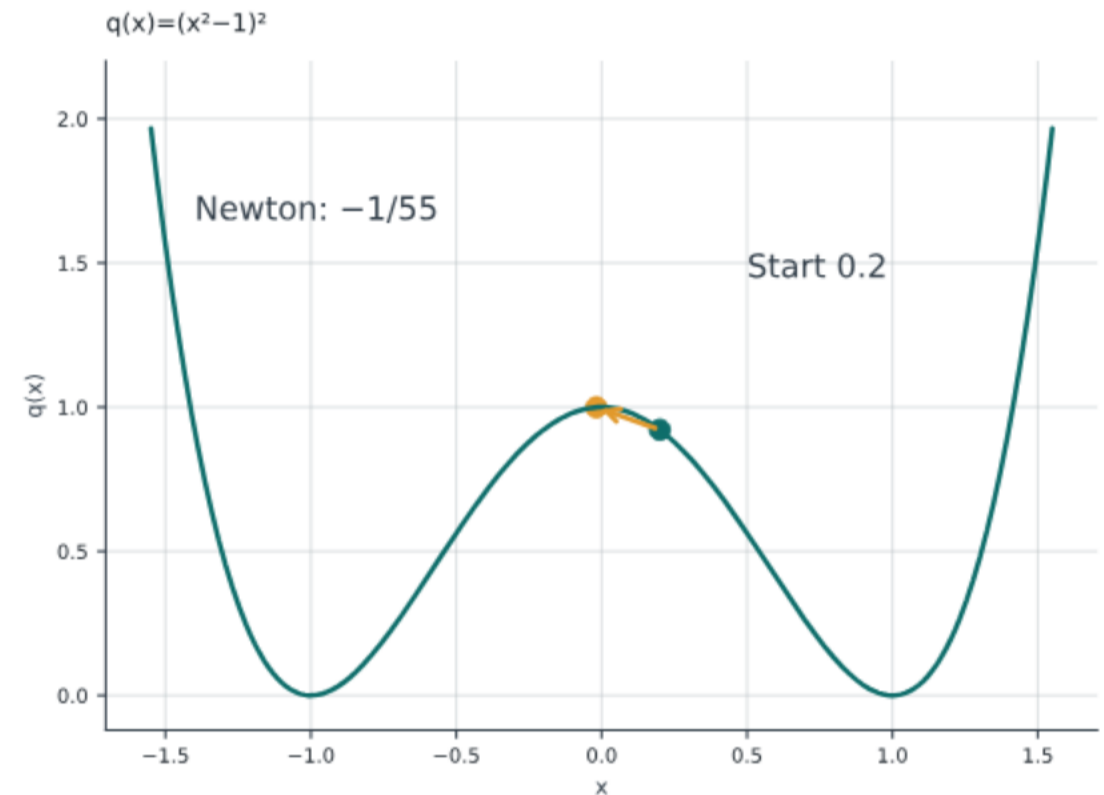


Newton can move uphill in a nonconvex region

$q(x)=(x^2-1)^2$. At $x_0=0.2$: $g=-0.768$ and $H=-3.52$.

Newton gives $p \approx -0.218182$ and $x_1 = -1/55 \approx -0.018182$.

q rises from 0.9216 to about 0.999339. Here $g_p > 0$.

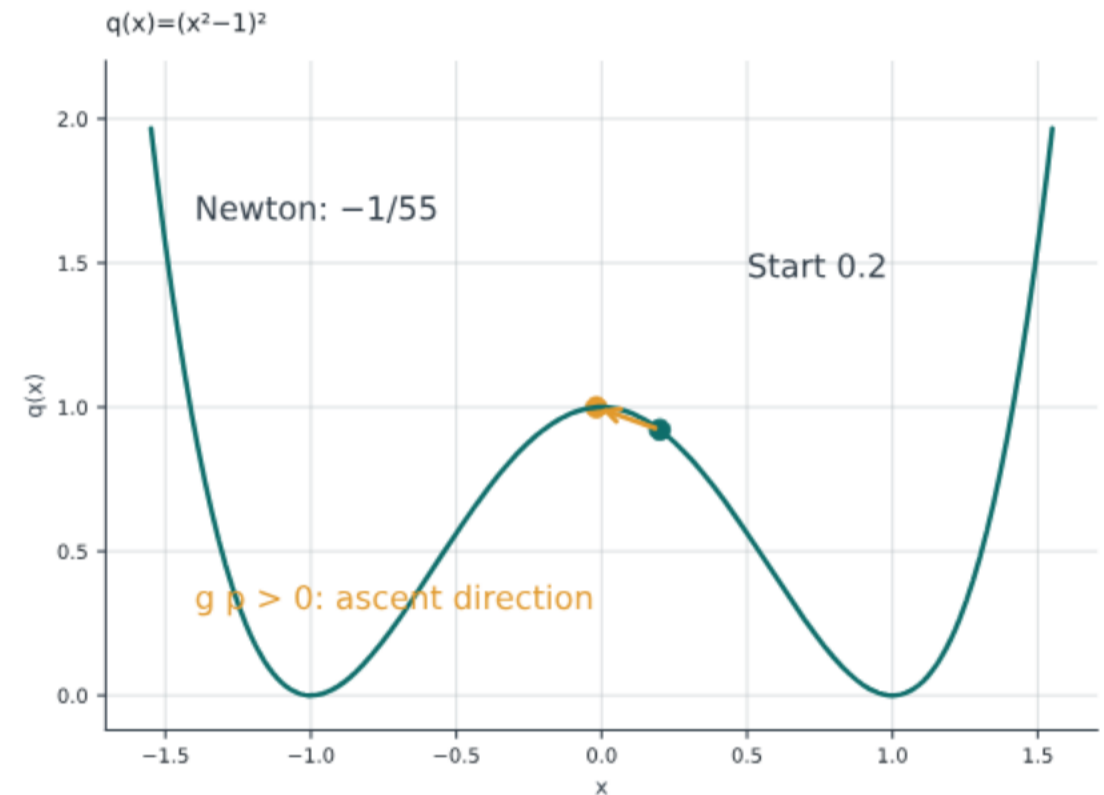


A direction needs a descent check

Before stepping, inspect $g^T p$. A negative value indicates local descent.

If $H > 0$ and $g \neq 0$, Newton gives $g^T p = -g^T H^{-1} g < 0$.

An indefinite Hessian may require a modified direction or a trust-region method.

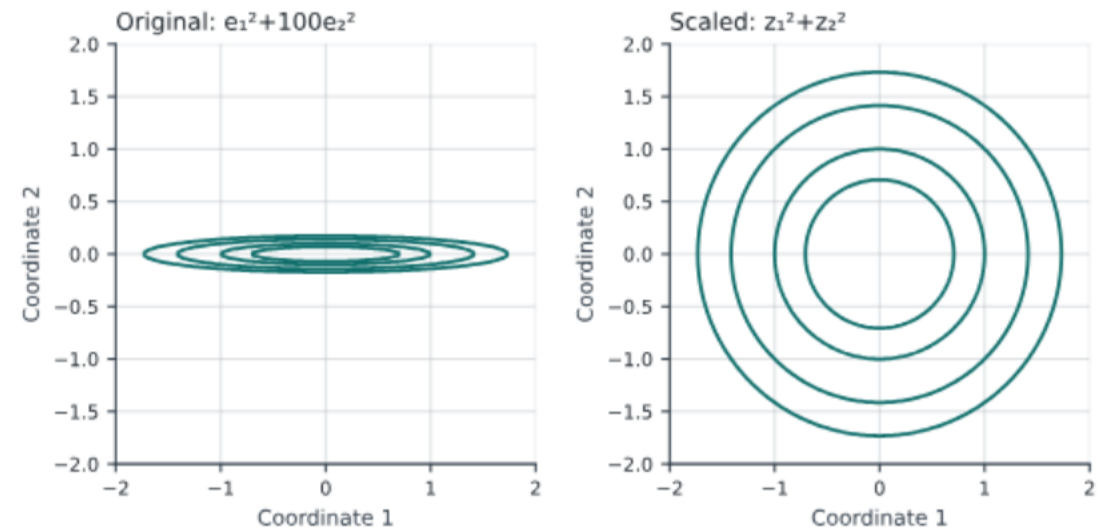


Scaling and elongated contours

Example: $F=e_1^2+100e_2^2$ has Hessian eigenvalues 2 and 200.

The curvature ratio is 100. A single fixed step must respect the steep direction.

With $z_1=e_1$ and $z_2=10e_2$, $F=z_1^2+z_2^2$.



Gradient descent and Newton

Gradient descent uses first derivatives and a chosen step size.

Newton also uses curvature and solves a linear system.

Both require stopping checks. Nonconvex problems need additional interpretation.

Gradient descent

$$p = -g$$

Simple steps; scale matters

Newton

Solve $Hp = -g$

Uses local curvature

Both methods

Check descent and stopping

Python lab: workflow

1. Define f , its gradient, and its Hessian.
2. Trace gradient descent and compare fixed step sizes.
3. Verify Newton, then classify nonconvex stationary points.

Lab 1: functions and derivative checks

Use NumPy arrays in the fixed order $[u,v]$.

At $[3,0]$, expect $f=20$, $\text{gradient}=[4,16]$, and $\text{Hessian}=\text{diag}(2,8)$.

Compare analytic derivatives against central finite differences.

Objective

Return a scalar

Gradient

Return a vector of length 2

Hessian

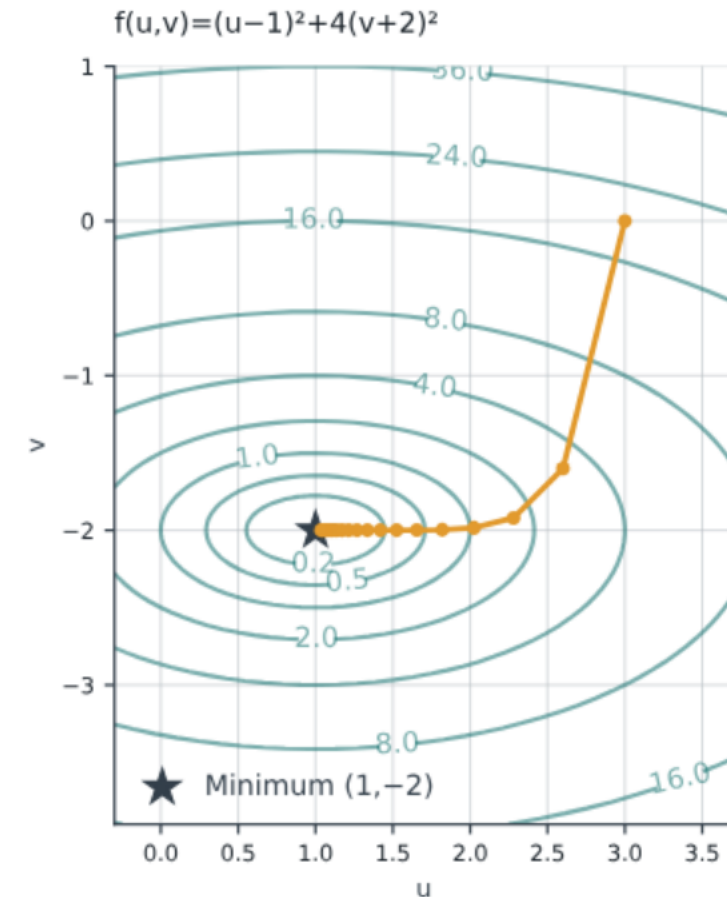
Return a symmetric 2×2 array

Lab 2: a gradient descent trace

Start at $[3,0]$, with $\alpha=0.1$.

Expected first values: 20.0000, 3.2000, 1.6640.

Store each iterate and objective before checking convergence.

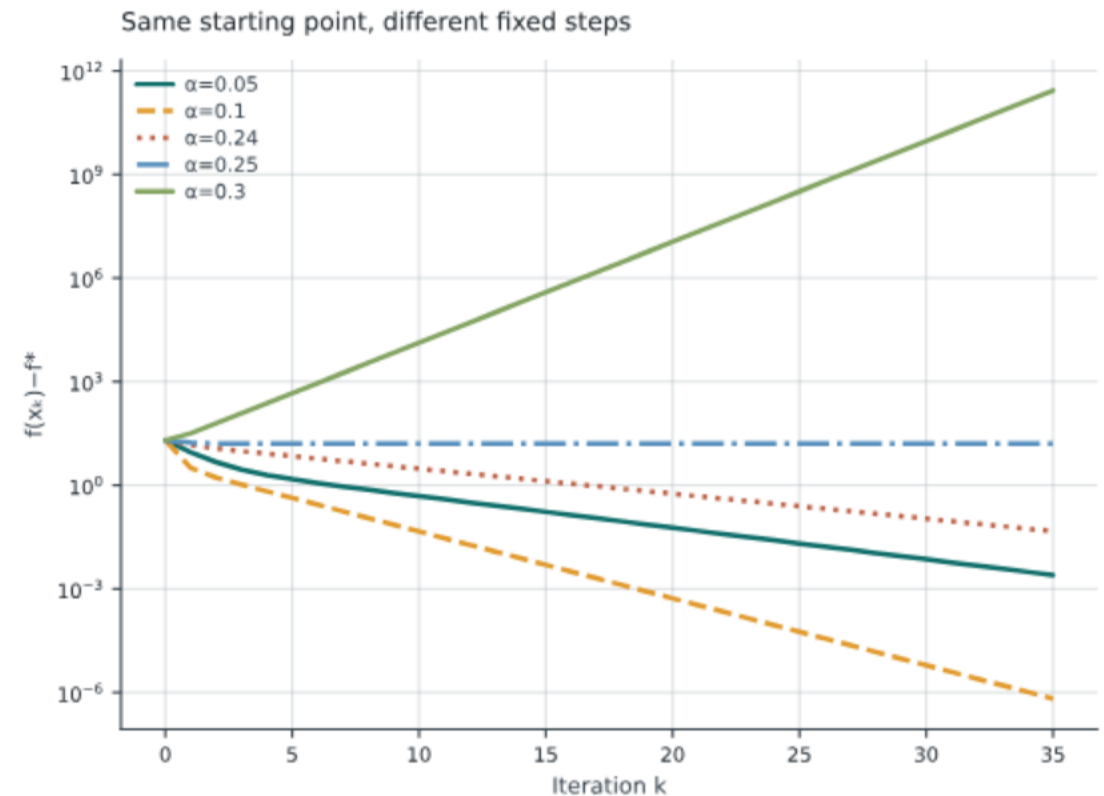


Lab 3: step-size experiments

Run $\alpha=0.05$, 0.10 , 0.24 , 0.25 , and 0.30 .

Keep x_0 , tolerance, and iteration cap fixed across runs.

Report convergence, oscillation, or growth using the trace.

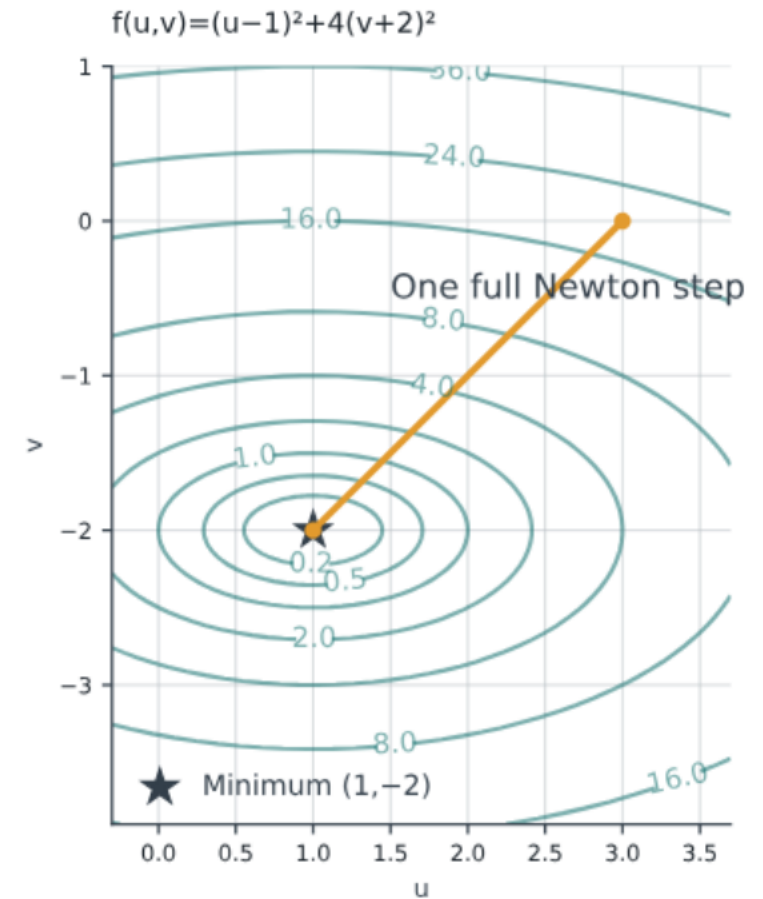


Lab 4: Newton and a reference solver

Compute `p=np.linalg.solve(H(x), -grad(x))`.

One full quadratic step must give $[1, -2]$, $f=0$.

Optional: compare with `scipy.optimize.minimize` using analytic derivatives.

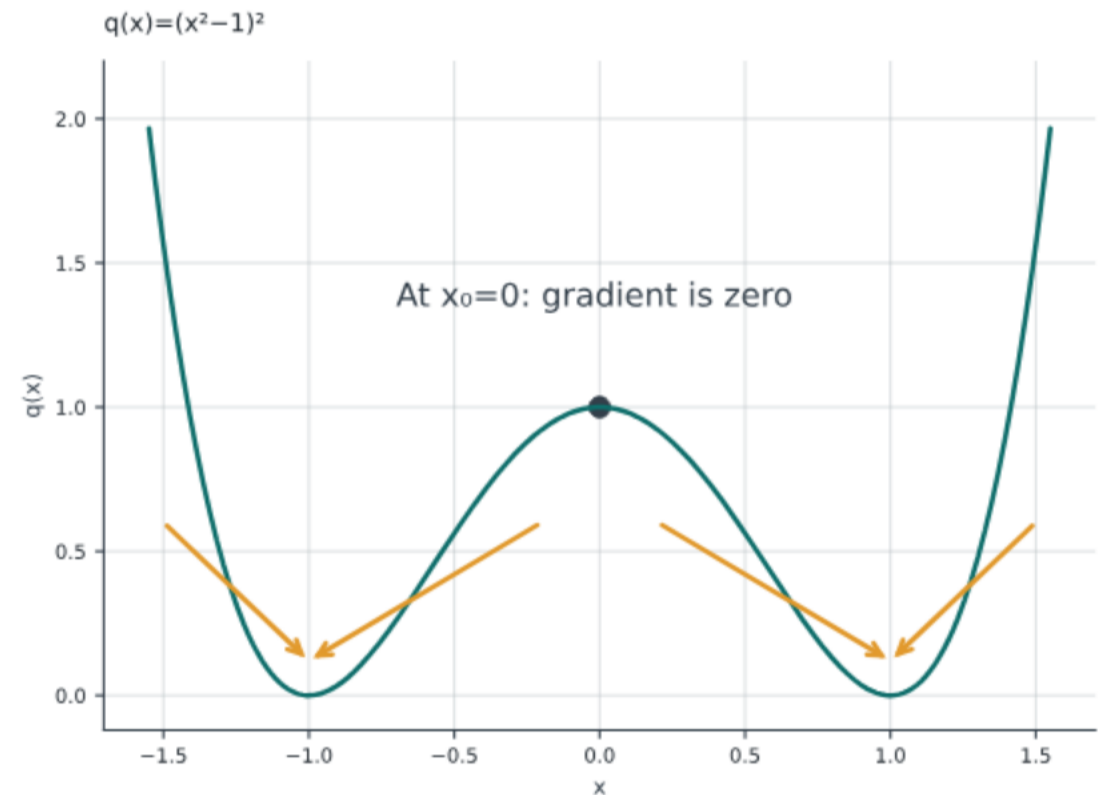


Lab 5: nonconvex starting points

For $q(x)=(x^2-1)^2$, compare starts -1.5 , -0.2 , 0 , 0.2 , and 1.5 .

Use gradient descent with $\alpha=0.05$, then inspect q' and q'' .

The start $x_0=0$ can stop immediately at a local maximum.



Independent practice

A. Minimize $h=x^2+2xy+3y^2-4x-8y+7$. Find H , eigenvalues, and x^* .

B. Take one Newton step for $r=x^4+x^2$ from $x_0=2$.

C. Explain why $f'(0)=f''(0)=0$ cannot identify a minimum.

Practice A: coupled quadratic solution

H is shown at right; its eigenvalues are $4 \pm 2\sqrt{2} > 0$.

Stationarity gives $x+y=2$ and $x+3y=4$, hence $(x^*, y^*) = (1, 1)$.

$h(1, 1) = 1$. Positive-definite H certifies the unique global minimum.

$$H = \begin{bmatrix} 2 & 2 \\ 2 & 6 \end{bmatrix}$$

$$\lambda = 4 \pm 2\sqrt{2} > 0$$

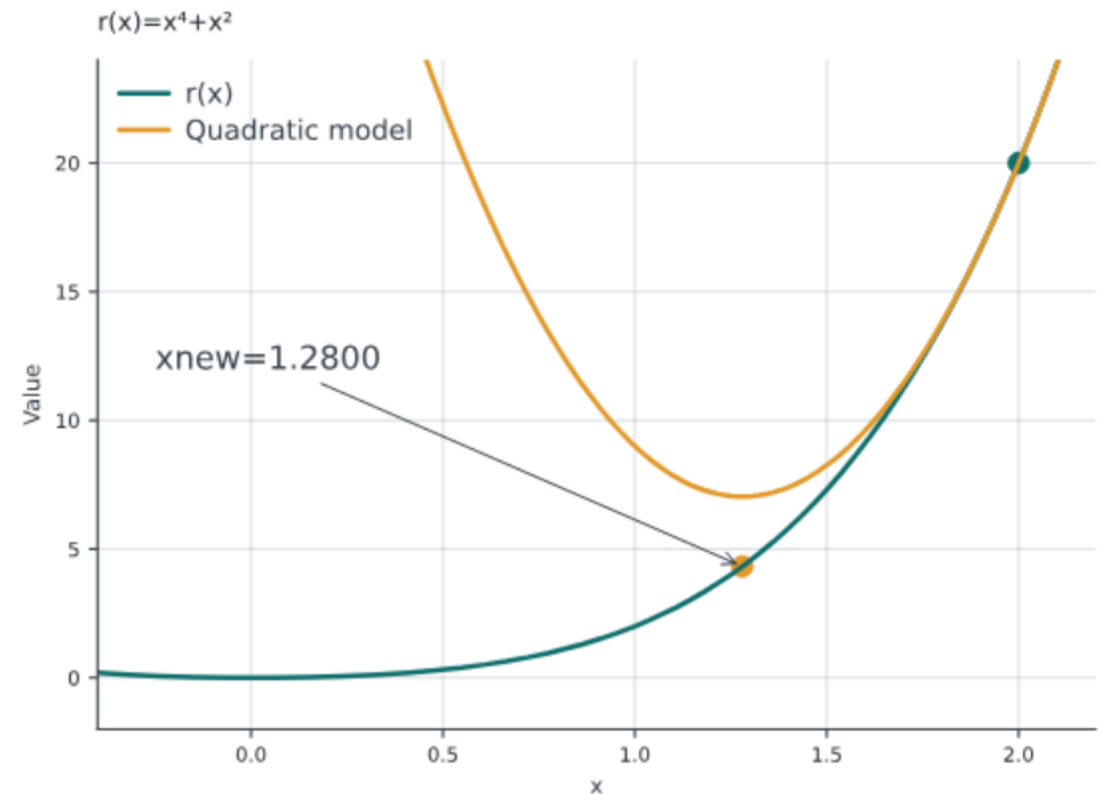
Positive definite
Unique quadratic minimum

Practice B and C: worked answers

B. At $x=2$: $r'=36$, $r''=50$, and $p=-0.72$.

$x_1=1.28$, with $r(x_1)=4.32275456$, down from 20.

C. x^4 and $-x^4$ share zero first and second derivatives at 0, but have opposite extrema.



Closing questions

1. Does a positive Hessian at one point prove global convexity?
2. What happens at $\alpha=0.25$ for the shared quadratic from [3,0]?
3. Why can Newton increase q at $x=0.2$?

Closing answers

1. No. The convexity Hessian test requires the whole open convex domain.
2. The v error alternates without shrinking, so the method does not reach x^* .
3. $q''(0.2) < 0$ gives an ascent Newton direction: $g_p > 0$.

Next session: choosing the step size

Today: geometry, gradient descent, Newton, and result checks.

Next: backtracking, Armijo, BFGS, and practical stopping rules.

Keep the notebook and the questions raised by the failed steps.

Geometry

Convex or nonconvex?

Derivatives

Stationary point; curvature

Algorithm

Direction + step size

Verification

Value, gradient, and meaning