Single-Step Comparison — Gradient Descent vs Newton’s Method

**Given**

[ f(x\_1,x\_2)=x\_1^2 + 2x\_22,x{(0)} = ]

## 1. Gradient Descent (one step — direction)

### (a) Gradient (f(x\_1,x\_2))

Compute partial derivatives:

* (=2x\_1)
* (=4x\_2)

So [ f(x\_1,x\_2)=]

Evaluate at (x^{(0)}=(2,1)): [ f(2,1)=]

### (b) Direction of steepest descent at (x^{(0)})

The steepest descent direction is the **negative gradient**: [ d\_{} = -f(2,1)=]

(If you want a unit direction, divide by its norm ( |f|===4); the unit descent direction would be ([-1/( ), -1/( )]), but the algorithm usually uses (-f) scaled by a step size ().)

## 2. Newton’s Method (one step — Newton direction)

### (a) Hessian (H=^2 f(x))

Second derivatives:

* (=2)
* (=4)
* mixed partials are 0

So the Hessian (constant) is [ H=^2 f(x)=]

### (b) Inverse Hessian (H^{-1})

Since (H) is diagonal: [ H^{-1}=]

### (c) Newton direction

Formula: [ d\_{} = -H{-1}f(x{(0)}).]

Plug in values: [d\_{} = -](1/4)\end{bmatrix} = -

=

. ]

## 3. Comparison

* Gradient at ((2,1)): (f(2,1)=
* ).
* Gradient descent (steepest descent) direction: (;d\_{}=
* .)
* Hessian: (H=
* ,; H^{-1}=
* .)
* Newton direction: (d\_{}=
* .)

**Interpretation:** Newton’s direction scales each gradient component by the inverse curvature in that coordinate (here dividing the first component by 2 and the second by 4), so it accounts for curvature and points to a more balanced step. In fact, with a unit Newton step ((x{(1)}=x{(0)}+d\_{})) we get (x^{(1)} = [2,1] + [-2,-1] = [0,0]), which is the exact minimizer for this quadratic. Gradient descent would need a step size chosen appropriately (here (/2) would give the same update: (x^{(1)} = [2,1] + [-4,-4] = [0, -1]) — not the optimum).