Gradient Descent for a Simple Convex Function

Group 2 – Ferschl Martin, Reiter Roman, Zenkic Mirza

December 12, 2025

1 Task specification

We are given a (convex) function $f(x, y)$ of two real variables. A simple model example is

$$f(x, y) = x^2 + y^2,$$

which has a unique minimum at $(x, y) = (0, 0)$.

The gradient

$$\nabla f(x, y) = \big(f_x(x, y), f_y(x, y)\big)$$

points in the direction of steepest ascent of f, while $-\nabla f(x, y)$ points in the direction of steepest descent.

1. For $f(x, y) = x^2 + y^2$, we are asked to sketch the contour lines $x^2 + y^2 = \text{const}$ and verify that the gradient ∇f is orthogonal to the contour lines at each point (geometric interpretation).
2. We then implement and test a simple gradient descent strategy for the model problem $f(x, y) = x^2 + y^2$, starting from a given initial point (x_0, y_0). The iteration is

   $$(x_{k+1}, y_{k+1}) = (x_k, y_k) - \gamma \, \nabla f(x_k, y_k),$$

   where $\gamma > 0$ is a step size that must be chosen appropriately.

2 Contour lines and orthogonality of the gradient

For the specific function

$$f(x, y) = x^2 + y^2,$$

the level sets (contour lines) are defined by

$$f(x, y) = c \quad \Longleftrightarrow \quad x^2 + y^2 = c,$$

for constants $c \geq 0$. These are circles centred at the origin with radius $\sqrt{c}$.

The partial derivatives are

$$f_x(x, y) = \frac{\partial}{\partial x}(x^2 + y^2) = 2x, \qquad f_y(x, y) = \frac{\partial}{\partial y}(x^2 + y^2) = 2y,$$

so the gradient is

$$\nabla f(x, y) = (2x, 2y).$$

Geometrically, at a point (x, y) on the circle $x^2 + y^2 = c$, the vector (x, y) points radially outwards from the origin. Hence $\nabla f(x, y) = 2(x, y)$ also points radially outwards. A tangent vector to the circle at (x, y) is, for example,

$$\boldsymbol{t} = (-y, x),$$

which is orthogonal to the radial direction. Indeed,

$$\nabla f(x, y) \cdot \boldsymbol{t} = (2x, 2y) \cdot (-y, x) = 2x(-y) + 2yx = -2xy + 2xy = 0.$$

Thus the gradient ∇f is orthogonal to the contour line $x^2 + y^2 = c$ at each point (x, y) on that circle.

In a sketch of the level sets (concentric circles around the origin), the gradient vectors are drawn as arrows pointing radially outward, perpendicular to the circles. The direction of steepest descent is given by $-\nabla f$, i.e. radially inward, pointing directly toward the minimum at $(0, 0)$. This orthogonality of the gradient to the contour lines is the geometric basis of the gradient descent method.

3 Gradient descent for $f(x, y) = x^2 + y^2$

3.1 Iteration formula and convergence condition

We consider the gradient descent iteration

$$(x_{k+1}, y_{k+1}) = (x_k, y_k) - \gamma \nabla f(x_k, y_k),$$

with step size $\gamma > 0$.

For our model function, the gradient is

$$\nabla f(x, y) = (2x, 2y),$$

so the iteration becomes

$$(x_{k+1}, y_{k+1}) = (x_k, y_k) - \gamma(2x_k, 2y_k) = \big((1 - 2\gamma)x_k,\ (1 - 2\gamma)y_k\big).$$

Hence each component evolves independently according to

$$x_{k+1} = (1 - 2\gamma)x_k, \qquad y_{k+1} = (1 - 2\gamma)y_k.$$

By induction we obtain

$$x_k = (1 - 2\gamma)^k x_0, \qquad y_k = (1 - 2\gamma)^k y_0.$$

Thus the iterates (x_k, y_k) converge to the minimum $(0, 0)$ if and only if

$$|1 - 2\gamma| < 1.$$

This inequality is equivalent to

$$-1 < 1 - 2\gamma < 1 \quad \Longleftrightarrow \quad 0 < \gamma < 1.$$

- If $0 < \gamma < \frac{1}{2}$, then $1 - 2\gamma \in (0, 1)$. The iterates move toward the origin monotonically without overshooting.

- If $\frac{1}{2} < \gamma < 1$, then $1 - 2\gamma \in (-1, 0)$. The iterates converge to the origin but alternate signs (they overshoot and "zigzag" around the minimum).
- If $\gamma = 1$, then $1 - 2\gamma = -1$, so x_k and y_k alternate between (x_0, y_0) and $(-x_0, -y_0)$ and never converge: this is an example of overshooting with no damping.
- If $\gamma > 1$, then $|1 - 2\gamma| > 1$ and the iterates diverge.

This exactly matches the intuition from the exercise text: choosing $\gamma = 1$ can cause the iteration to "overshoot". In that case, one can reduce γ, e.g. to $\gamma = \frac{1}{2}$, which yields $1 - 2\gamma = 0$ and the minimum is reached in a single step.

3.2 Python implementation and test

We now implement gradient descent for $f(x, y) = x^2 + y^2$ in Python and test different values of γ. The code prints a few iterations and the final distance to the minimum.

```
import numpy as np

def f(x, y):
    """Objective function f(x, y) = x^2 + y^2."""
    return x**2 + y**2

def grad_f(x, y):
    """Gradient of f: f(x, y) = (2x, 2y)."""
    return np.array([2*x, 2*y], dtype=float)

def gradient_descent(x0, y0, gamma, n_steps):
    """Run n_steps of gradient descent for f starting at (x0, y0)."""
    x, y = float(x0), float(y0)
    trajectory = [(x, y)]
    for k in range(n_steps):
        g = grad_f(x, y)
        x -= gamma * g[0]
        y -= gamma * g[1]
        trajectory.append((x, y))
    return trajectory

# Parameters
x0, y0 = 1.0, -1.0    # initial point
gamma_good = 0.25     # step size (convergent without oscillation)
gamma_overshoot = 1.0   # step size that leads to oscillation

# Run gradient descent with a good step size
traj_good = gradient_descent(x0, y0, gamma_good, n_steps=10)
print("Gradient descent with gamma =", gamma_good)
for k, (xk, yk) in enumerate(traj_good):
    print(f"k={k:2d}: (x, y) = ({xk: .6f}, {yk: .6f}), f = {f(xk, yk): .6e}")

# Run gradient descent with gamma = 1 (overshooting)
traj_over = gradient_descent(x0, y0, gamma_overshoot, n_steps=6)
print("\nGradient descent with gamma =", gamma_overshoot, "(overshoot)")
for k, (xk, yk) in enumerate(traj_over):
    print(f"k={k:2d}: (x, y) = ({xk: .6f}, {yk: .6f}), f = {f(xk, yk): .6e}")
```

A typical output (for $\gamma = 0.25$) shows the iterates moving steadily towards $(0, 0)$, with the function values $f(x_k, y_k)$ decreasing monotonically. For $\gamma = 1$, the iterates jump back and forth between two points of equal distance from the origin, illustrating overshooting and the importance of choosing an appropriate step size.

4 Conclusion

For the simple convex function $f(x, y) = x^2 + y^2$, we have:

- The contour lines $f(x, y) = \text{const}$ are circles centred at the origin.
- The gradient $\nabla f(x, y) = (2x, 2y)$ is orthogonal to these contour lines at every point and points in the direction of steepest ascent.
- Gradient descent with iteration $(x_{k+1}, y_{k+1}) = (x_k, y_k) - \gamma \nabla f(x_k, y_k)$ converges to the minimum $(0, 0)$ if and only if $0 < \gamma < 1$. Larger step sizes lead to oscillations or divergence.

This model problem illustrates the basic ideas behind gradient descent that are used, in more complex forms, throughout optimization and machine learning.