Justification of the Gradient Direction (Orthogonality to Level Sets)

Group 2 – Ferschl Martin, Reiter Roman, Zenkic Mirza

December 12, 2025

1 Task specification

We consider a scalar function
$$f : \mathbb{R}^2 \to \mathbb{R}, \qquad f(x, y),$$
which is sufficiently smooth (at least continuously differentiable). For a fixed value $c \in \mathbb{R}$, the equation
$$f(x, y) = c$$
defines a curve in the plane (a *level set* or *contour line* of f). Under suitable regularity assumptions (e.g. f_x and f_y not both zero on the curve), this level set can locally be parameterized as
$$\boldsymbol{x}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}, \qquad t \in I \subset \mathbb{R},$$
so that
$$f(x(t), y(t)) = c \quad \text{for all } t \in I.$$

The vector
$$\boldsymbol{x}'(t) = \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix}$$
is the tangent (velocity) vector to the curve at the point $\boldsymbol{x}(t) = (x(t), y(t))^\top$. The claim is that the gradient
$$\nabla f(x, y) = \begin{pmatrix} f_x(x, y) \\ f_y(x, y) \end{pmatrix}$$
is orthogonal to this tangent vector at each point on the level curve.

We prove this using the chain rule.

2 Chain rule argument

Let $f(x, y)$ be C^1 (continuously differentiable), and let $\boldsymbol{x}(t) = (x(t), y(t))^\top$ be a differentiable parameterization of the level set $f(x, y) = c$. By definition of the level set, we have
$$f(x(t), y(t)) = c \quad \text{for all } t.$$

We now differentiate this identity with respect to t.

Step 1: Differentiate the composition

Consider the composite function

$$g(t) := f(x(t), y(t)).$$

We know that $g(t) \equiv c$ is constant, hence

$$g'(t) = 0 \quad \text{for all } t.$$

On the other hand, by the (one-dimensional) chain rule applied to the composition $f \circ \boldsymbol{x}$, we have

$$g'(t) = \frac{d}{dt} f(x(t), y(t)) = f_x(x(t), y(t))\, x'(t) + f_y(x(t), y(t))\, y'(t),$$

where f_x and f_y denote the partial derivatives of f with respect to its first and second argument.

Combining both expressions for $g'(t)$, we obtain the scalar equation

$$f_x(x(t), y(t))\, x'(t) + f_y(x(t), y(t))\, y'(t) = 0.$$

Step 2: Interpret as a dot product

We can rewrite this as a dot product of two vectors in $\mathbb{R}^2$:

$$\nabla f(x(t), y(t)) \cdot \boldsymbol{x}'(t) = \begin{pmatrix} f_x(x(t), y(t)) \\ f_y(x(t), y(t)) \end{pmatrix} \cdot \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix} = 0.$$

Thus, at every point on the level curve where $\boldsymbol{x}'(t) \neq \mathbf{0}$, we have

$$\nabla f(x(t), y(t)) \cdot \boldsymbol{x}'(t) = 0,$$

i.e. the gradient vector is orthogonal to the tangent vector.

3 Geometric interpretation and connection to gradient descent

The above calculation shows that at any point (x, y) on the level set $f(x, y) = c$, the gradient $\nabla f(x, y)$ is perpendicular to the level curve passing through that point. Equivalently:

- $\nabla f(x, y)$ points in the direction of *steepest increase* of f.
- $-\nabla f(x, y)$ points in the direction of *steepest decrease* of f.
- The level curve $f(x, y) = c$ is locally "flat" in the sense that moving along the curve (i.e. along $\boldsymbol{x}'(t)$) leaves the value of f unchanged; hence there is no component of ∇f in the tangential direction, only in the normal direction.

This orthogonality property justifies the gradient descent step used in Task 4:

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

At each iteration, we move from the current point in the direction $-\nabla f$, i.e. orthogonally to the contours of f, in the direction of steepest decrease of the function value. For convex functions, this iteration can be shown (under suitable step-size conditions) to converge to a global minimizer.