Directional Derivatives and the Gradient

In this lecture we specialize to the case where \( f : \mathbb{R}^n \to \mathbb{R} \) is a real-valued function of several variables. For such a function the differential \( Df \) reduces to an \( 1 \times n \) matrix, or equivalently an \( n \)-dimensional vector. In fact we have

\[
Df = \begin{pmatrix}
\frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n}
\end{pmatrix}
= \begin{pmatrix}
\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}
\end{pmatrix} \equiv \nabla f
\]

so \( Df \) can be identified with the gradient of \( f \).

We'll come back to the gradient in a minute. But first let me introduce the notion of directional derivatives.

**Definition 8.1.** Let \( f \) be a function from \( \mathbb{R}^n \) to \( \mathbb{R} \), and let \( u \) be a unit vector in \( \mathbb{R}^n \) (i.e., a vector of length 1). Then the directional derivative of \( f \) in the direction \( u \) at the point \( x \) is the limit

\[
D_uf(x) \equiv \left. \frac{df}{dt} (x + tu) \right|_{t=0} \equiv \lim_{t \to 0} \frac{f(x+tu) - f(x)}{t}
\]

The directional derivative of \( f : \mathbb{R}^n \to \mathbb{R} \) along the direction \( u \) at the point \( x \) is interpretable as the rate of change in \( f \) as one moves away from the point \( x \) in the direction of \( u \).

**Remark 8.2.** We restrict \( u \) to be a unit vector because most often we're interested only in how a function changes when we move in different directions. Since, we care only about the direction of \( u \) and not its magnitude; we simply fix its magnitude to be 1.

**Example 8.3.** Compute the rate of change of \( f : (x, y, z) \mapsto x^2yz \) in the direction \( u = \left( \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}} \right) \) at the point \((1,1,0)\).
We need to compute

\[ \mathbf{D}_u f(x) = \left. \frac{d}{dt} \left[ f(1,1,0) + t \left( \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}} \right) \right] \right|_{t=0} \]

\[ = \left. \frac{d}{dt} \left( 1 + \frac{t}{\sqrt{3}}, 1 + \frac{t}{\sqrt{3}}, 0 - \frac{t}{\sqrt{3}} \right) \right|_{t=0} \]

\[ = \left. \frac{d}{dt} \left( \left(1 + \frac{t}{\sqrt{3}}\right)^2 \left( 1 + \frac{t}{\sqrt{3}}, -\frac{t}{\sqrt{3}} \right) \right) \right|_{t=0} \]

\[ = \left. \left(2 \left(1 + \frac{t}{\sqrt{3}}\right) \left(1 + \frac{t}{\sqrt{3}}, -\frac{t}{\sqrt{3}}\right) \left(1 + \frac{t}{\sqrt{3}}\right) \right) \right|_{t=0} \]

\[ + \left. \left(\left(1 + \frac{t}{\sqrt{3}}\right)^2 \left( \frac{1}{\sqrt{3}}, -\frac{t}{\sqrt{3}} \right) \right) \right|_{t=0} \]

\[ + \left. \left(\left(1 + \frac{t}{\sqrt{3}}\right)^2 \left(1 + \frac{t}{\sqrt{3}}, -\frac{t}{\sqrt{3}}\right) \right) \right|_{t=0} \]

\[ = (2) \left(\frac{1}{\sqrt{3}}\right) (1)(0) \]

\[ + (1)^2 \left(1\right) (0) \]

\[ + (1)^2 (1) \left(\frac{1}{\sqrt{3}}\right) \]

\[ = -\frac{1}{\sqrt{3}} \]

Below we give a theorem that makes computations such as the one above a lot simpler.

**Theorem 8.4.** If \( f : \mathbb{R}^n \to \mathbb{R} \) is differentiable then all directional derivatives exist and, moreover, the directional derivative of \( f \) in the direction \( \mathbf{u} \) at the point \( x \) is given by

\[ \nabla f(x) \cdot \mathbf{u} \]

**Proof.** Let \( \gamma : \mathbb{R} \to \mathbb{R}^n \) be the function

\[ \gamma(t) = x + t\mathbf{u} \]

so that

\[ \gamma_1(t) = x_1 + tu_1 \]

\[ \gamma_2(t) = x_2 + tu_2 \]

\[ \vdots \]

\[ \gamma_n(t) = x_n + tu_n \]

and

\[ f(x + t\mathbf{u}) = f(\gamma(t)) \]
By the chain rule we have

\[ D_u f(x) = \frac{d}{dt} f(x + tu) \bigg|_{t=0} = \frac{d}{dt} (f \circ \gamma) \bigg|_{t=0} = D(f \circ \gamma) \bigg|_{t=0} = Df(\gamma(0)) D\gamma(0) \]

\[ = \left( \frac{\partial f}{\partial x_1}(\gamma(0)) \ \frac{\partial f}{\partial x_2}(\gamma(0)) \ \cdots \ \frac{\partial f}{\partial x_n}(\gamma(0)) \right) \left( \begin{array}{c} u_1 \\ u_2 \\ \vdots \\ u_n \end{array} \right) \]

\[ = \left( \frac{\partial f}{\partial x_1}(x) \ \frac{\partial f}{\partial x_2}(x) \ \cdots \ \frac{\partial f}{\partial x_n}(x) \right) \left( \begin{array}{c} u_1 \\ u_2 \\ \vdots \\ u_n \end{array} \right) = \nabla f(x) \cdot u \]

**Example 8.5.** Let’s return to the preceding example and use our spanking new formula to compute the directional derivative of \( f(x,y,z) = x^2yz \) along the direction \( u = \left( \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}} \right) \) at the point \((1,1,0)\).

\[ \nabla f(1,1,0) = \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right) \bigg|_{(1,1,0)} = (2yz, x^2, x^2) \bigg|_{(1,1,0)} = (0,0,1) \]

So

\[ \frac{d}{dt} f((1,1,0) + tu) \bigg|_{t=0} = \nabla f(1,1,0) \cdot u = (0,0,1) \cdot \left( \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}} \right) = -\frac{1}{\sqrt{3}} \]

The gradient \( \nabla f \) not only makes the computation of directional derivatives easier, it also makes it easy to identify the direction in which a function increases most rapidly.

**Theorem 8.6.** Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) be a differentiable function and assume that \( \nabla f(x) \neq 0 \). Then the direction of \( \nabla f(x) \) coincides with the direction in which \( f(x) \) is increasing most rapidly.

**Proof.** We want to determine the direction \( u \) in which a directional derivative

\[ D_u f(x) = \frac{d}{dt} f(x + tu) \bigg|_{t=0} \]
is maximized. Using the preceding theorem we have
\[ D_u f(x) = \nabla f(x) \cdot u \]
where \( u \) is a unit vector and so
\[ \frac{d}{dt} f(x + tu) \bigg|_{t=0} = \nabla f(x) \cos(\theta) \]
The right-hand side is obviously maximized when \( \theta = 0 \); i.e. when \( u \) points in the same direction as \( \nabla f(x) \).

Remark 8.7. Another way of phrasing the result of this theorem is that, when one imagines the graph of \( f \) as a surface with hilltops and valleys, the direction of the \( \nabla f(x) \) corresponds to the direction uphill at the point \( x \).

Here is another application of the gradient.

Theorem 8.8. Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a differentiable function and let \( x_0 \) be a point on the level surface
\[ S = \{ x \in \mathbb{R}^n \mid f(x) = k \} \]
Then \( \nabla f(x_0) \) is normal to the surface \( S \) at the point \( x_0 \) in the following sense: if \( v \) is the tangent vector at \( t = 0 \) to any curve \( \gamma(t) \) that lies within \( S \) and satisfies \( \gamma(t) = 0 \), then \( v \cdot \nabla f(x_0) = 0 \).

Proof. Let \( \gamma(t) \) be such a curve. Since \( \gamma(t) \) lies in \( S \) for all \( t \) we must have
\[ f(\gamma(t)) = k \]
Therefore,
\[ 0 = \frac{d}{dt}(f \circ \gamma) \bigg|_{t=0} = D f(\gamma(0)) D \gamma(0) \]
\[ = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x_0) & \frac{\partial f}{\partial x_2}(x_0) & \cdots & \frac{\partial f}{\partial x_n}(x_0) \end{pmatrix} \begin{pmatrix} \frac{dx_1}{dt}(0) \\ \frac{dx_2}{dt}(0) \\ \vdots \\ \frac{dx_n}{dt}(0) \end{pmatrix} \]
\[ = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x_0) & \frac{\partial f}{\partial x_2}(x_0) & \cdots & \frac{\partial f}{\partial x_n}(x_0) \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \]
\[ = \nabla f(x_0) \cdot v \]
Because the gradient of \( f \) at the point \( x_0 \) is perpendicular to the tangent vector at \( x_0 \) to any curve \( \gamma(t) \) that lives in a level surface \( S = \{ x \in \mathbb{R}^n \mid f(x) = k \} \) it is reasonable to define the plane tangent to the surface \( S \) at the point \( x_0 \) in terms of the gradient.

Definition 8.9. Let \( f : \mathbb{R}^n \to \mathbb{R} \) be a differentiable function and let \( S \) be a surface in \( \mathbb{R}^n \) of the form \( S = \{ x \in \mathbb{R}^n \mid f(x) = k \} \), the tangent plane to \( S \) at the point \( x_0 \) is defined by the equation
\[ \nabla f(x_0) \cdot (x - x_0) = 0 \]