LECTURE 8

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 $\mathbf{D}f$ reduces to an $1 \times n$ matrix, or equivalently an n-dimensional vector. In fact we have

$$\mathbf{D}f = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{pmatrix} = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n} \right) \equiv \nabla f$$

so $\mathbf{D}f$ 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 $\mathbf{u}$ be a unit vector in $\mathbb{R}^n$ (i.e, a vector of length 1). Then the* ***directional derivative*** *of f in the direction $\mathbf{u}$ at the point $\mathbf{x}$ is the limit*

$$\mathbf{D}_{\mathbf{u}} f(\mathbf{x}) \equiv \frac{d}{dt} f\left(\mathbf{x} + t\mathbf{u}\right)\bigg|_{t=0} \equiv \lim_{t \to 0} \frac{f\left(\mathbf{x} + t\mathbf{u}\right) - f\left(\mathbf{x}\right)}{t}$$

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

Remark 8.2. We restrict $\mathbf{u}$ 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 $\mathbf{u}$ and not its magnitude; we simply fix its magnitude to be 1.

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

We need to compute

$$
\begin{aligned}
\mathbf{D_u} f(\mathbf{x}) &= \frac{d}{dt} f\left((1,1,0) + t\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right)\right)\Bigg|_{t=0} \\
&= \frac{d}{dt} f\left(1 + \frac{1}{\sqrt{3}}t, 1 + \frac{1}{\sqrt{3}}t, 0 - \frac{1}{\sqrt{3}}t\right)\Bigg|_{t=0} \\
&= \frac{d}{dt}\left(\left(1+\frac{t}{\sqrt{3}}\right)^2 \left(y\backslash 1 + \frac{t}{\sqrt{3}}\right)\left(-\frac{t}{\sqrt{3}}\right)\right)\Bigg|_{t=0} \\
&= \left(2\left(1+\frac{t}{\sqrt{3}}\right)\left(\frac{1}{\sqrt{3}}\right)\left(1 + \frac{t}{\sqrt{3}}\right)\left(-\frac{t}{\sqrt{3}}\right)\right)\Bigg|_{t=0} \\
&+ \left(\left(1+\frac{t}{\sqrt{3}}\right)^2 \left(\frac{1}{\sqrt{3}}\right)\left(-\frac{t}{\sqrt{3}}\right)\right)\Bigg|_{t=0} \\
&+ \left(\left(1+\frac{t}{\sqrt{3}}\right)^2 \left(1 + \frac{t}{\sqrt{3}}\right)\left(-\frac{1}{\sqrt{3}}\right)\right)\Bigg|_{t=0} \\
&= (2)\left(\frac{1}{\sqrt{3}}\right)(1)(0) \\
&+ (1)^2\left(\frac{1}{\sqrt{3}}\right)(0) \\
&+ (1)^2(1)\left(-\frac{1}{\sqrt{3}}\right) \\
&= -\frac{1}{\sqrt{3}}
\end{aligned}
$$

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 $\mathbf{x}$ is given by*

$$\nabla f(\mathbf{x}) \cdot \mathbf{u}$$

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

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

so that

$$
\begin{aligned}
\gamma_1(t) &= x_1 + tu_1 \\
\gamma_2(t) &= x_2 + tu_2 \\
&\vdots \\
\gamma_n(t) &= x_n + tu_n
\end{aligned}
$$

and

$$f(\mathbf{x} + t\mathbf{u}) = f(\gamma(t))$$

By the chain rule we have

$$\begin{aligned}
\mathbf{D}_{\mathbf{u}} f(\mathbf{x}) &= \left.\frac{d}{dt} f(\mathbf{x}+t\mathbf{u})\right|_{t=0} \\
&= \left.\frac{d}{dt}(f\circ\gamma)\right|_{t=0} \\
&= \mathbf{D}(f\circ\gamma)|_{t=0} \\
&= \mathbf{D}f(\gamma(0))\,\mathbf{D}\gamma(0) \\
&= \begin{pmatrix} \frac{\partial f}{\partial x_1}(\gamma(0)) & \frac{\partial f}{\partial x_2}(\gamma(0)) & \cdots & \frac{\partial f}{\partial x_n}(\gamma(0)) \end{pmatrix} \begin{pmatrix} \frac{d\gamma_1}{dt}(0) \\ \frac{d\gamma_2}{dt}(0) \\ \vdots \\ \frac{d\gamma_n}{dt}(0) \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) & \frac{\partial f}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \\
&= \frac{\partial f}{\partial x_1}(\mathbf{x})u_1 + \frac{\partial f}{\partial x_2}(\mathbf{x})u_2 + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x})u_n \\
&= \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{x})\right)\cdot(u_1, u_2, \ldots, u_n) \\
&= \nabla f(\mathbf{x})\cdot\mathbf{u}
\end{aligned}$$

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 $\mathbf{u} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right)$ at the point $(1,1,0)$.

$$\begin{aligned}
\nabla f(1,1,0) &= \left.\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)\right|_{(1,1,0)} \\
&\left.\left(2xyz, x^2z, x^2y\right)\right|_{(1,1,0)} \\
&= (0,0,1)
\end{aligned}$$

So

$$\begin{aligned}
\left.\frac{d}{dt} f((1,1,0)+t\mathbf{u})\right|_{t=0} &= \nabla f(1,1,0)\cdot\mathbf{u} \\
&= (0,0,1)\cdot\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right) \\
&= -\frac{1}{\sqrt{3}}
\end{aligned}$$

The gradient ∇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 $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function and assume that $\nabla f(\mathbf{x}) \neq \mathbf{0}$. Then the direction of $\nabla f(\mathbf{x})$ coincides with the direction in which $f(\mathbf{x})$ is increasing most rapidly.*

Proof. We want to determine the direction $\mathbf{u}$ in which a directional derivative

$$\mathbf{D}_{\mathbf{u}} f(\mathbf{x}) = \left.\frac{d}{dt} f(\mathbf{x}+t\mathbf{u})\right|_{t=0}$$

is maximized. Using the preceding theorem we have

$$\begin{aligned}\mathbf{D}_{\mathbf{u}}f(\mathbf{x}) &= \nabla f(\mathbf{x}) \cdot \mathbf{u} \\ &= \|\nabla f(\mathbf{x})\| \, \|\mathbf{u}\| \cos(\theta)\end{aligned}$$

where θ is the angle between $\nabla f(\mathbf{x})$ and $\mathbf{u}$. Since $\mathbf{u}$ is, by definition, a unit vector $\|\mathbf{u}\| = 1$ and so

$$\left.\frac{d}{dt} f(\mathbf{x} + t\mathbf{u})\right|_{t=0} = \nabla f(\mathbf{x}) \cos(\theta)$$

The right hand side is obviously maximized when $\theta = 0$; i.e. when $\mathbf{u}$ points in the same direction as $\nabla f(\mathbf{x})$.

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 $\nabla f(\mathbf{x})$ corresponds to the direction uphill at the point $\mathbf{x}$.

Here is another application of the gradient.

THEOREM 8.8. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function and let $\mathbf{x}_0$ be a point on the level surface*

$$S = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = k\}$$

Then $\nabla f(\mathbf{x}_0)$ is normal to the surface S at the point $\mathbf{x}_0$ in the following sense: if $\mathbf{v}$ is the tangent vector at $t = 0$ to any curve $\gamma(t)$ that lies within S and satisfies $\gamma(t) = 0$, then $\mathbf{v} \cdot \nabla f(\mathbf{x}_0) = 0$.

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,

$$\begin{aligned}0 &= \left.\frac{d}{dt}(f \circ \gamma)\right|_{t=0} \\ &= \mathbf{D}f(\gamma(0)\mathbf{D}\gamma(0) \\ &= \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \end{pmatrix} \begin{pmatrix} \frac{d\gamma_1}{dt}(0) \\ \frac{d\gamma_2}{dt}(0) \\ \vdots \\ \frac{d\gamma_n}{dt}(0) \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \\ &= \nabla f(\mathbf{x}_0) \cdot \mathbf{v}\end{aligned}$$

Because the gradient of f at the point $\mathbf{x}_0$ is perpendicular to the tangent vector at $\mathbf{x}_0$ to any curve $\gamma(t)$ that lives in a level surface $S = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = k\}$ it is reasonable to define the plane tangent to the surface S at the point $\mathbf{x}_0$ in terms of the gradient.

DEFINITION 8.9. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a diffetentiable function and let S be a surface in $\mathbb{R}^n$ of the form $S = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = k\}$, the the* ***tangent plane to** S **at the point*** *$\mathbf{x}_0$ is defined by the equation*

$$\nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) = 0$$