§16

16.1 Gradient/Directional Derivative

Definition 1 (Directional Derivative). The *directional derivative of a function* $f : \mathcal{D}^n \to \mathbf{R}$ *at* $\boldsymbol{a} \in \mathcal{D}^n$ *in the direction of the nonzero vector* $\boldsymbol{v} \in \mathbf{R}^n$ is:

$$\mathrm{D}_{\boldsymbol{v}} f(\boldsymbol{a}) := \lim_{h \to 0} \frac{f(\boldsymbol{a} + h\boldsymbol{v}/\|\boldsymbol{v}\|) - f(\boldsymbol{a})}{h}$$

Note 1. The vector $\boldsymbol{v}/\|\boldsymbol{v}\|$ is the unit vector in the direction of the (nonzero, by assumption) vector $\boldsymbol{v}$.

Example 1 (Licata 9.2). Compute the directional derivative of $f(x, y) = \frac{1}{xy}$ at $\boldsymbol{a} = (1, 2)$ in the direction of $\boldsymbol{v} = \left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right]$ using the definition.

Solution. The unit vector in the direction of $\boldsymbol{v}$ is

$$\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \frac{\left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right]}{\sqrt{2^2 + 3^2}} = \left[\begin{smallmatrix} 2/\sqrt{13} \\ 3/\sqrt{13} \end{smallmatrix}\right]$$

By definition:

$$\begin{aligned}
\mathrm{D}_{f} f(\boldsymbol{a}) &= \lim_{h \to 0} \frac{f\left(1 + h\frac{2}{\sqrt{13}}, 2 + h\frac{3}{\sqrt{13}}\right) - f(1, 2)}{h} \\
&= \lim_{h \to 0} \frac{\left[\left(1 + h\frac{2}{\sqrt{13}}\right)\left(2 + h\frac{3}{\sqrt{13}}\right)\right]^{-1} - 2^{-1}}{h} \\
&= \lim_{h \to 0} \frac{2 - \left(1 + h\frac{2}{\sqrt{13}}\right)\left(2 + h\frac{3}{\sqrt{13}}\right)}{h2\left(1 + h\frac{2}{\sqrt{13}}\right)\left(2 + h\frac{3}{\sqrt{13}}\right)} \\
&= \lim_{h \to 0} \frac{-h\frac{4}{\sqrt{13}} - h\frac{3}{\sqrt{13}} - h^2\frac{6}{13}}{h2\left(1 + h\frac{2}{\sqrt{13}}\right)\left(2 + h\frac{3}{\sqrt{13}}\right)} \\
&= \lim_{h \to 0} \frac{-\frac{4}{\sqrt{13}} - \frac{3}{\sqrt{13}} - h\frac{6}{13}}{2\left(1 + h\frac{2}{\sqrt{13}}\right)\left(2 + h\frac{3}{\sqrt{13}}\right)} \\
&= \frac{-\frac{7}{\sqrt{13}}}{2(1)(2)} = -\frac{7}{4\sqrt{13}}
\end{aligned}$$

■

When the directional derivative $\mathrm{D}_{\boldsymbol{v}} f(\boldsymbol{a})$ exists, it is a number (scalar) that indicates the rate of change of f at $\boldsymbol{a}$ in the direction of $\boldsymbol{v}$. Directional derivative generalizes partial derivative because $\frac{\partial f}{\partial x_i}(\boldsymbol{a}) = \mathrm{D}_{\boldsymbol{e}_i} f(\boldsymbol{a})$. Conversely, the partial derivatives $\frac{\partial f}{\partial x_1}(\boldsymbol{a}), \ldots, \frac{\partial f}{\partial x_n}(\boldsymbol{a})$ determine the directional derivative $\mathrm{D}_{\boldsymbol{e}_i} f(\boldsymbol{a})$ for any $\boldsymbol{v}$ through what is called the gradient ∇f.

Definition 2 (Gradient). The *gradient of the function* $f : \mathcal{D}^n \to \mathbf{R}$ *at* $\boldsymbol{a}$ is:

$$\nabla f(\boldsymbol{a}) := \begin{bmatrix} \partial_{x_1} f(\boldsymbol{a}) \\ \partial_{x_2} f(\boldsymbol{a}) \\ \vdots \\ \partial_{x_n} f(\boldsymbol{a}) \end{bmatrix}$$

Note 2. Note that for a scalar function $f : \mathcal{D}^n \to \mathbf{R}$, $\nabla f(\boldsymbol{a}) = \mathrm{D} f(\boldsymbol{a})^{\mathrm{T}}$.

The relationship between directional derivatives and the gradient is the following proposition.

Proposition 1. For a scalar function $f : \mathcal{D}^n \to \mathbf{R}$, a vector $\boldsymbol{a} \in \mathcal{D}^n$, and a nonzero vector $\boldsymbol{v}$ in $\mathbf{R}^n$:

$$\mathrm{D}_{\boldsymbol{v}} f(\boldsymbol{a}) = \nabla f(\boldsymbol{a}) \cdot \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$$

The proposition is useful both computationally and theoretically:

- Computational: Find a directional derivative using the gradient (and dot product).
- Theoretical: Maximize/minimize the directional derivative.

Example 2 (Licata 9.6). Compute the directional derivative of $f(x, y) = \frac{1}{xy}$ at $\boldsymbol{a} = (1, 2)$ in the direction of $\boldsymbol{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ using the gradient.

Solution. The gradient is

$$\nabla f(1,2) = \begin{bmatrix} -(x^2 y)^{-1} \\ -(x y^2)^{-1} \end{bmatrix} \Bigg|_{(x,y)=(1,2)} = \begin{bmatrix} -\tfrac{1}{2} \\ -\tfrac{1}{4} \end{bmatrix}$$

and the unit vector in the direction of $\boldsymbol{v}$ is

$$\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \frac{\begin{bmatrix} 2 \\ 3 \end{bmatrix}}{\sqrt{2^2 + 3^2}} = \frac{1}{\sqrt{13}} \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

so the desired direction derivative is:

$$\mathrm{D}_{\boldsymbol{v}} f(\boldsymbol{a}) = \begin{bmatrix} -\tfrac{1}{2} \\ -\tfrac{1}{4} \end{bmatrix} \cdot \left(\frac{1}{\sqrt{13}} \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right) = -\frac{7}{4\sqrt{13}} \qquad \blacksquare$$

Let's explore the theoretical use of the proposition in more detail. If $\nabla f(\boldsymbol{a})$ and $\boldsymbol{v}/\|\boldsymbol{v}\|$ have angle θ between them, then

$$\mathrm{D}_{\boldsymbol{v}} f(\boldsymbol{a}) = \nabla f(\boldsymbol{a}) \cdot \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \|\nabla f(\boldsymbol{a})\| \cos\theta$$

since $\boldsymbol{v}/\|\boldsymbol{v}\|$ has length 1. Therefore:

- $-\|\nabla f(\boldsymbol{a})\| \leq \mathrm{D}_{\boldsymbol{v}} f(\boldsymbol{a}) \leq \|\nabla f(\boldsymbol{a})\|$ for all nonzero vectors $\boldsymbol{v}$ in $\mathbf{R}^n$
- $\mathrm{D}_{\boldsymbol{v}} f(\boldsymbol{a})$ is smallest (equal to $-\|\nabla f(\boldsymbol{a})\|$) when $\boldsymbol{v}$ and $\nabla f(\boldsymbol{a})$ are anti-parallel
- $\mathrm{D}_{\boldsymbol{v}} f(\boldsymbol{a})$ is largest (equal to $\|\nabla f(\boldsymbol{a})\|$) when $\boldsymbol{v}$ and $\nabla f(\boldsymbol{a})$ are parallel

Here are two more properties of the gradient:

- $\mathrm{D}_{\boldsymbol{v}} f(\boldsymbol{a})$ is zero when $\boldsymbol{v}$ and $\nabla f(\boldsymbol{a})$ are orthogonal
- For $\boldsymbol{a}$ in the height c level set $S = f^{-1}(c)$ of f, $\nabla f(\boldsymbol{a})$ is orthogonal to the tangent space[1)] of S at $\boldsymbol{a}$.
  - In particular, for any curve $C \subseteq S = f^{-1}(c)$ passing through $\boldsymbol{a}$, the tangent line to C at $\boldsymbol{a}$ is orthogonal to $\nabla f(\boldsymbol{a})$.
  - Said differently, if $I \subseteq \mathbf{R}$ is an interval and $\boldsymbol{g}\colon I \to \mathbf{R}^n$ is a parametrized curve with $\boldsymbol{g}(I) \subseteq S = f^{-1}(c)$, then $\boldsymbol{g}'(t)$ and $\nabla f(\boldsymbol{g}(t))$ are orthogonal for all $t \in I$.

1) tangent space is the term used for a "tangent plane" that is not necessarily 2-dimensional

16.2 (Local) Extrema of Multivariate Functions

Definition 3. Let $f\colon \mathcal{D}^n \to \mathbf{R}$ be a function, and let $\boldsymbol{a}$ be a point in $\mathcal{D}^n$.

- f has a *local minimum*[2)] at $\boldsymbol{a}$ if there is $\varepsilon > 0$ so that $f(\boldsymbol{x}) \geq f(\boldsymbol{a})$ for all $\boldsymbol{x}$ in $\mathcal{D}^n$ satisfying $\|\boldsymbol{x} - \boldsymbol{a}\| < \varepsilon$.
- f has a *local maximum*[3)] at $\boldsymbol{a}$ if there is $\varepsilon > 0$ so that $f(\boldsymbol{x}) \leq f(\boldsymbol{a})$ for all $\boldsymbol{x}$ in $\mathcal{D}^n$ satisfying $\|\boldsymbol{x} - \boldsymbol{a}\| < \varepsilon$.
- f has a *local extremum*[4)] at $\boldsymbol{a}$ if f has either a local minimum or a local maximum at $\boldsymbol{a}$.

2) plural of minimum is minima

3) plural of maximum is maxima

4) plural of extremum is extrema

Definition 4. Let $f\colon \mathcal{D}^n \to \mathbf{R}$ be a function. A point $\boldsymbol{a}$ in $\mathcal{D}^n$ is a *critical point of* f if one of the following is true:

(i) f is not differentiable

(ii) $\nabla f(\boldsymbol{a}) = \mathbf{0}$

It is not always easy to check if a multivariable function is differentiable. (What is the definition of differentiability for multivariable functions?[5]) The function $f : \mathbf{R}^2 \to \mathbf{R}$ defined by:

$$f(x, y) = \begin{cases} x^2 y/(x^4 + y^2) & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

is *not differentiable at* $(0, 0)$ despite the fact that $\partial_x f$ and $\partial_y f$ both exist everywhere, even at the origin: $\nabla f(0, 0) = \mathbf{0}$).

5) For general maps $\mathcal{D}^n \to \mathbf{R}^m$, the definition of differentiability involves a remainder term and a certain limit. In this case, when $m = 1$, differentiability at $\boldsymbol{a}$ is equivalent to the existence of $\nabla f(\boldsymbol{a})$ and the identity $\lim_{\boldsymbol{x} \to \boldsymbol{a}} \big[f(\boldsymbol{x}) - f(\boldsymbol{a}) - \nabla f(\boldsymbol{a}) \cdot (\boldsymbol{x} - \boldsymbol{a}) \big] = 0$.

For "nice" functions, we may use the following, easier, test to find the critical points.

Proposition 2. Let $f : \mathcal{D}^n \to \mathbf{R}$ be a function such that for each $\boldsymbol{a}$ in $\mathcal{D}^n$, either:

- $\nabla f(\boldsymbol{a})$ is does not exist or
- $\nabla f(\boldsymbol{x})$ is defined near $\boldsymbol{a}$ and continuous at $\boldsymbol{a}$

Then a point $\boldsymbol{a}$ in $\mathcal{D}^n$ is a critical point of f if and only if one of the following is true:

(i) $\nabla f(\boldsymbol{a})$ does not exist or

(ii) $\nabla f(\boldsymbol{a}) = \mathbf{0}$

Proposition 3 (First Derivative Test). If $f : \mathcal{D}^n \to \mathbf{R}$ has a local extremum as $\boldsymbol{a}$ in $\mathcal{D}^n$, then $\boldsymbol{a}$ is either a critical point of f or a "boundary point"[6] of $\mathcal{D}^n$.

6) The term "boundary" will be defined rigorously later.

Note 3. Not every critical point must be a local extremum. We even give a name to a special type of critical point that is not a local extremum.

Definition 5. We call a point $\boldsymbol{a}$ a *saddle point* if:

(i) f is differentiable and

(ii) $\nabla f(\boldsymbol{a}) = \mathbf{0}$

but $\boldsymbol{a}$ is not a local extremum of f.

For "nice" functions, we may use the following, easier, test to find the saddle points.

Proposition 4. Let $f : \mathcal{D}^n \to \mathbf{R}$ be a function such that for each $\boldsymbol{a}$ in $\mathcal{D}^n$, either:

- $\nabla f(\boldsymbol{a})$ is does not exist or
- $\nabla f(\boldsymbol{x})$ is defined near $\boldsymbol{a}$ and continuous at $\boldsymbol{a}$

Then a point $\boldsymbol{a}$ in $\mathcal{D}^n$ is a saddle point for f if and only if $\nabla f(\boldsymbol{a}) = \mathbf{0}$ but but $\boldsymbol{a}$ is not a local extremum of f. ✑

Proposition 5 (Second Derivative Test). Let $f\colon \mathcal{D}^n \to \mathbf{R}$ be a function, and assume $\boldsymbol{a}$ is a point in $\mathcal{D}^n$ such that:

- $\nabla f(\boldsymbol{a}) = \mathbf{0}$
- $\mathsf{H} f(\boldsymbol{x}) = [\partial_{x_i}\partial_{x_j} f(\boldsymbol{x})]$ has entries that are continuous at $\boldsymbol{a}$

Then:

- $\boldsymbol{x}^{\mathsf{T}}[\mathsf{H} f(\boldsymbol{a})]\boldsymbol{x}$ is positive definite $\Rightarrow$ f has a local minimum at $\boldsymbol{a}$
- $\boldsymbol{x}^{\mathsf{T}}[\mathsf{H} f(\boldsymbol{a})]\boldsymbol{x}$ is negative definite $\Rightarrow$ f has a local maximum at $\boldsymbol{a}$
- $\boldsymbol{x}^{\mathsf{T}}[\mathsf{H} f(\boldsymbol{a})]\boldsymbol{x}$ is indefinite $\Rightarrow$ f has a saddle point at $\boldsymbol{a}$

The Second Derivative Test is indeterminate in the semidefinite cases.

Example 3 (Licata 12.5). Let $f(x, y) = (\cos x)(\ln y)$.

(a) Show that $(\frac{\pi}{2}, 1)$ is a critical point of f.

(b) Compute $\mathsf{H} f(\frac{\pi}{2}, 1)$.

(c) Classify $(\frac{\pi}{2}, 1)$ as a local minimum, a local maximum, or a saddle point.

Solution.

(a) The gradient is $\nabla f(\frac{\pi}{2}, 1) = \begin{bmatrix} (-\sin x)(\ln y) \\ (\cos x)(1/y) \end{bmatrix}\Big|_{(x,y)=(\pi/2,1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ so $(\frac{\pi}{2}, 1)$ is a critical point of f.

(b) The Hessian is:

$$\mathsf{H} f(\tfrac{\pi}{2}, 1) = \begin{bmatrix} (-\cos x)(\ln y) & (-\sin x)(1/y) \\ (-\sin x)(1/y) & (\cos x)(-1/y^2) \end{bmatrix}\Big|_{(x,y)=(\pi/2,1)} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

(c) The determinant of $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$ is negative (equal to -1) so the quadratic form associated to the Hessian matrix at $(\frac{\pi}{2}, 1)$ is indefinite. Therefore $(\frac{\pi}{2}, 1)$ is a saddle point for f. ∎