Derivatives for multivariate functions

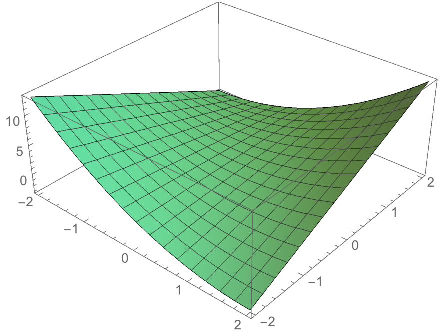
Philipp Warode

October 1, 2019

- A function depending on more than one variable is called **multivariate** function.

Example:

$$f(x, y) := x^2 + 2xy + 1$$



- Let $f(x) = f(x_1, x_2, \ldots, x_n)$ be a multivariate function
- The **partial derivative** in the x_i direction in $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})$ is

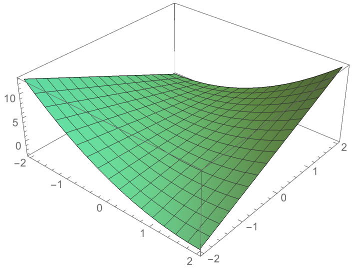
$$\frac{\partial}{\partial x_i} f(x^{(0)}) := \lim_{x \to x_i^{(0)}} \frac{f(x_1^{(0)}, \ldots, x_{i-1}^{(0)}, x, x_{i+1}^{(0)}, \ldots, x_n^{(0)}) - f(x^{(0)})}{x - x_i^{(0)}}$$

- Let $f(x) = f(x_1, x_2, \ldots, x_n)$ be a multivariate function
- The **partial derivative** in the x_i direction in $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})$ is

$$\frac{\partial}{\partial x_i} f(x^{(0)}) := \lim_{x \to x_i^{(0)}} \frac{f(x_1^{(0)}, \ldots, x_{i-1}^{(0)}, x, x_{i+1}^{(0)}, \ldots, x_n^{(0)}) - f(x^{(0)})}{x - x_i^{(0)}}$$

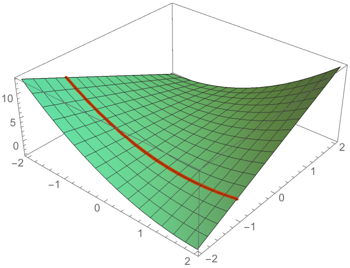
- The partial derivative is just the derivative wrt. x_i where the other values are treated as constants.

Function f



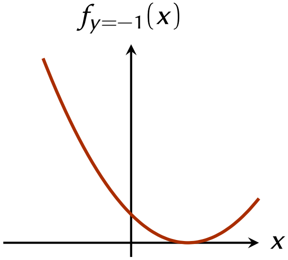
$$f(x, y) = x^2 + 2xy + 1$$

Function f



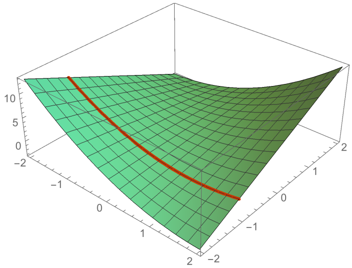
$$f(x, y) = x^2 + 2xy + 1$$

Function f with fixed $y_0 = -1$



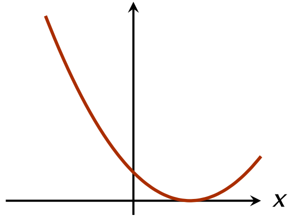
$$f_{y=-1}(x) = x^2 - 2x + 1$$

Function f



$f(x, y) = x^2 + 2xy + 1$

Function f with fixed $y_0 = -1$

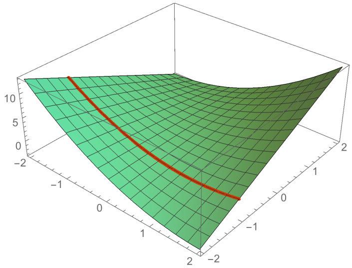


$f_{y=-1}(x) = x^2 - 2x + 1$

$f'_{y=0}(x) = 2x - 2$

Partial derivatives

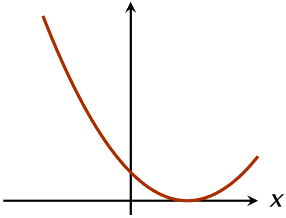
Function f



$$f(x, y) = x^2 + 2xy + 1$$

$$\frac{\partial}{\partial x} f(x, y) = 2x + 2y$$

Function f with fixed $y_0 = -1$



$$f_{y=-1}(x) = x^2 - 2x + 1$$

$$f'_{y=0}(x) = 2x - 2$$

- The partial derivative $\frac{\partial}{\partial x_i} f(x)$ again depends on the whole vector $x = (x_1, \ldots, x_n)$.
- The vector of all partial derivatives

$$\nabla f(x) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{pmatrix}$$

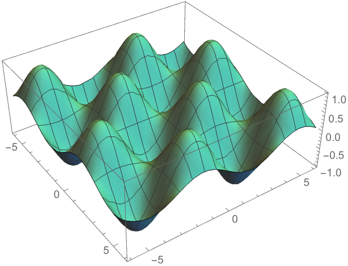
  is called the **gradient** of f.
- The gradient is a vector that points in the direction of steepest increase

A point $x^{(0)} \in \mathbb{R}^n$ is a **local maximum** of $f : \mathbb{R}^n \to \mathbb{R}$ if there is a $\epsilon > 0$ with

$$f(x_0) \geq f(x) \text{ for all } x \text{ with } \|x - x_0\| < \epsilon$$

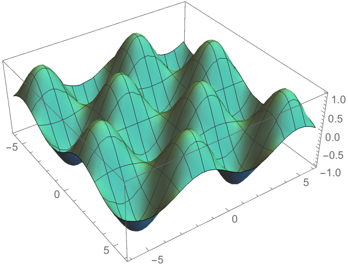
A point $x^{(0)} \in \mathbb{R}^n$ is a **local maximum** of $f : \mathbb{R}^n \to \mathbb{R}$ if there is a $\epsilon > 0$ with

$$f(x_0) \geq f(x) \text{ for all } x \text{ with } \|x - x_0\| < \epsilon$$



A point $x^{(0)} \in \mathbb{R}^n$ is a **local minimum** of $f : \mathbb{R}^n \to \mathbb{R}$ if there is a $\epsilon > 0$ with

$$f(x_0) \leq f(x) \text{ for all } x \text{ with } \|x - x_0\| < \epsilon$$



Theorem

If $x^{(0)}$ is a extreme point, then

$$\nabla f(x^{(0)}) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x^{(0)}) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x^{(0)}) \end{pmatrix} = \mathbf{0}.$$

Theorem

If $x^{(0)}$ is a extreme point, then

$$\nabla f(x^{(0)}) := \begin{pmatrix} \frac{\partial}{\partial x_1} f(x^{(0)}) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x^{(0)}) \end{pmatrix} = \mathbf{0}.$$

$\nabla f(x) = 0$ is only a **necessary condition** for an extreme point.

Definition

The symmetric matrix

$$H^f(x^{(0)}) := \begin{pmatrix} \frac{\partial^2}{\partial^2 x_1} f(x^{(0)}) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x^{(0)}) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x^{(0)}) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x^{(0)}) & \frac{\partial^2}{\partial^2 x_2} f(x^{(0)}) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f(x^{(0)}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x^{(0)}) & \frac{\partial^2}{\partial x_n \partial x_2} f(x^{(0)}) & \cdots & \frac{\partial^2}{\partial^2 x_n} f(x^{(0)}) \end{pmatrix}$$

is called **Hessian matrix** of f at x_0.

Theorem

A function f has a local maximum/minimum at x_0 if

- $\nabla f(x^{(0)}) = 0$ *and*
- $H^f(x^{(0)})$ *is positive/negative definite*

Definition

The symmetric matrix

$$H^f\left(x^{(0)}\right) := \begin{pmatrix} \frac{\partial^2}{\partial^2 x_1} f\left(x^{(0)}\right) & \frac{\partial^2}{\partial x_1 \partial x_2} f\left(x^{(0)}\right) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f\left(x^{(0)}\right) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f\left(x^{(0)}\right) & \frac{\partial^2}{\partial^2 x_2} f\left(x^{(0)}\right) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} f\left(x^{(0)}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f\left(x^{(0)}\right) & \frac{\partial^2}{\partial x_n \partial x_2} f\left(x^{(0)}\right) & \cdots & \frac{\partial^2}{\partial^2 x_n} f\left(x^{(0)}\right) \end{pmatrix}$$

is called **Hessian matrix** of f at x_0.

Theorem

A function f has a local maximum/minimum at x_0 if

- $\nabla f\left(x^{(0)}\right) = 0$ *and*
- $H^f\left(x^{(0)}\right)$ *is positive/negative definite*
  $\Leftrightarrow$ *All eigenvalues of $H^f\left(x^{(0)}\right)$ are positive/negative*

Criteria for local extrema for $n = 2$

$f : \mathbb{R}^2 \to \mathbb{R}^2$ has a local maximum/minimum at $(x^{(0)}, y^{(0)})$ if

- $\frac{\partial}{\partial x} f(x^{(0)}, y^{(0)}) = \frac{\partial}{\partial y} f(x^{(0)}, y^{(0)}) = 0$ (**Necessary condition**)
- $\det\left(H^f(x^{(0)}, y^{(0)})\right) > 0$ and
  - $\frac{\partial^2}{\partial^2 x} f(x^{(0)}, y^{(0)}) < 0$ or (**Sufficient condition for maximum**)
  - $\frac{\partial^2}{\partial^2 x} f(x^{(0)}, y^{(0)}) > 0$ (**Sufficient condition for minimum**)
- The determinant of H^f can be computed as

$$\det\left(H^f(x^{(0)}, y^{(0)})\right) = \frac{\partial^2}{\partial^2 x} f(x^{(0)}, y^{(0)}) \cdot \frac{\partial^2}{\partial^2 y} f(x^{(0)}, y^{(0)}) - \left(\frac{\partial^2}{\partial x \partial y} f(x^{(0)}, y^{(0)})\right)^2$$