Function of several variables
Session 2

Edouard Marchais

EPITA

September 2023

Themes :

- Directional derivatives and gradients
- Chain rules
- Maxima, minima and saddle points

Directional derivatives and gradients

- In two dimensions, we can construct a derivative in **another direction** than Ox ou Oy.

- For a function $f(x, y)$, we are talking of a **directional derivative** given by

$$D_{\boldsymbol{u}} f = \frac{\partial f}{\partial x}\, u_1 + \frac{\partial f}{\partial y}\, u_2$$

  where $\boldsymbol{u} = u_1\, \boldsymbol{i} + u_2\, \boldsymbol{j}$ is a **unit vector** ($\|\boldsymbol{u}\| = 1$).

- If $D_{\boldsymbol{u}} f = 0$ then f is a constant in the direction $\boldsymbol{u}$.

- $D_{\boldsymbol{u}} f$ is **bigger** (absolute value) as the slope is getting **steeper** in the direction $\boldsymbol{u}$.

- It is usual to express the directional derivative in terms of the **gradient** of the function $f(x,y)$ given by

$$\boldsymbol{\nabla} f = \begin{pmatrix} f_x \\ f_y \end{pmatrix}$$

- We then have the simple expression (we always have $\|\boldsymbol{u}\| = 1$)

$$D_{\boldsymbol{u}} f = \boldsymbol{\nabla} f \cdot \boldsymbol{u}$$

- If $\boldsymbol{u}$ and $\boldsymbol{\nabla} f$ are colinear then

$$\boldsymbol{u} = \frac{\boldsymbol{\nabla} f}{\|\boldsymbol{\nabla} f\|}$$

- The gradient formulation possess the advantage of not referring to any specific **coordinates system**.

- We consider a point (x, y) moving, over time, along a path in the xy plane.

- At this point we note $\boldsymbol{T}$ the tangent vector to the trajectory and therefore its direction will change over time.

- If f is a function of x, y then is also a function of the time t. It then makes sense to compute its **rate of change** $\frac{df}{dt} = f'(t)$. We have the expression

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = \boldsymbol{\nabla} f \cdot \boldsymbol{v}$$

  where $\boldsymbol{v}$ is the **speed vector** associated to the point (x, y).

- If we introduce a **position parameter** s along the path, then the slope with respect to this path is given by

$$\frac{df}{ds} = \frac{\partial f}{\partial x}\frac{dx}{ds} + \frac{\partial f}{\partial y}\frac{dy}{ds} = \boldsymbol{\nabla} f \cdot \boldsymbol{T}$$

- The $\boldsymbol{T}$ vector is **normalised** and **tangent** to the trajectory. It is defined by

$$\boldsymbol{T} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$$

Chain rules

We consider **three particularly important cases** :

- **Composition** of one and two variables functions :

$$f\left(g(x,y)\right) \qquad \Rightarrow \qquad f_x = ? \quad , \quad f_y = ?$$

- **Dependance** of variables with respect to a *third parameter*

$$f\left(x(t), y(t)\right) \qquad \Rightarrow \qquad f'(t) = ?$$

- **Dependance** of variables with respect to *two other parameters* (change of variables)

$$f\left(x(t,u), y(t,u)\right) \qquad \Rightarrow \qquad f_t = ? \quad , \quad f_u = ?$$

If $f = f(g(x,y))$

- The partial derivatives de $f(g(x,y))$ are obtained **simplyt** en by following a *chain rule* using **Leibniz's notation**

$$f_x = \frac{\partial f}{\partial x} = \underbrace{\frac{dg}{dg}}_{=1} \frac{\partial f}{\partial x} \quad \text{then} \quad f_x = \frac{df}{dg}\frac{\partial g}{\partial x} = f'(g)\, g_x$$

- We remark that we replaced ∂f by df because f depends « only » on g. The symbol ∂g is replaced by ∂g because g depends on only x but also on y.

- By repeating this reasoning for f_y yields

$$f_y = \frac{\partial f}{\partial y} = \frac{df}{dg}\frac{\partial g}{\partial y} = f'(g)\, g_y$$

If $f = f(x(t), y(t))$

- To compute $f'(t)$ we just need to come back to the definition of a **differential** of $f(x, y)$

$$df = \frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy \quad \Rightarrow \quad \frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

- This allows for a trivial **generalisation**

$$f'(t) = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt}$$

- If t appear **explicitly** in f then the formula becomes

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial t}$$

If $f = f(x(t,u), y(t,u))$

- The partial derivatives are obtained in a similar way to the previous derivative $f'(t)$, by replacing d by ∂, to wit

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\,\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\,\frac{\partial y}{\partial t} \quad \text{et} \quad \frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\,\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\,\frac{\partial y}{\partial u}$$

- Warning : when several variables are part of a derivative computation, special attention must be brought to the **interdependance** of these variables (see course document).

- For a function $f(x, y)$, the formula

$$\frac{\partial f}{\partial x}\frac{\partial x}{\partial f} = 1$$

is **not true**!

Maxima, minima and points-cols

- The **extrema** of a function $f(x, y)$ are characterised by

$$\frac{\partial f}{\partial x} = 0 \qquad \text{et} \qquad \frac{\partial f}{\partial y} = 0$$

  and the points satisfying these conditions are usually called **stationary points**.

- If les partial derivatives don't exist at a given point, we talk about **rough points**.

- If a point is located on the frontier of a research domain for an extremum, we talk about **end points**.

- The nature of a stationary point (minimum, maximum, etc..) is then determined by its **(partial) second derivatives** at this point.

- Geometrically, this amounts to study the position of the surface associated to f with respect to its **tangent plan** at the stationary point (similar to 1D).

- In arbitrary dimension, the **condition of stationarity** is simply given by the relation

$$\boldsymbol{\nabla} f = \mathbf{0}$$

where $f \equiv f(x_1, \ldots, x_n)$ and $\mathbf{0}$ is the nul vector of $\mathbb{R}^n$.

- **Stationary point on the frontier**

  - Example : we want to minimise/maximise the following function

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

    on the domain $y \leq 0$.

  - The conditions $f_x = f_y = 0$ give the point $(1/3, 1/3)$ which is **clearly not** in the research domain.

  - In that case we set before hand $y = 0$, i.e we are at the **frontier** of the research domain.

  - The condition $f_x(x, 0) = 0$ leads to an extremum located (on the frontier) at the point $(1/2, 0)$.

- Warning : The geometry of a domain frontier may be **very complicated** ...

- In such case it is useful to do a **substitution** and use a **position parameter** along the frontier.

- Generally, the search of an extremum on a frontier can quickly become **complex problem**...

- In such situation, we may use the method of **Lagrange multipliers**.

- **Nature of a stationary point**

  - We consider first the simple case of a **quadratic function** which stationary point is $(0,0)$

$$f(x,y) = \mathsf{r}\,x^2 + 2\mathsf{s}\,xy + \mathsf{t}\,y^2$$

  - Depending on **Monge parameters** r, s, t, one can determine the nature of the point $(0,0)$.

  - To show this, it suffices to write the function $f(x,y)$ under **canonical form** or **reduced form** in function of its parameters.

- **Test for a function quadratic**

  - If $\mathsf{s}^2 - \mathsf{rt} < 0$ then the origin is a **extrémum** local.
    - If $\mathsf{r} > 0$ then the origin is a local **minimum** .
    - If $\mathsf{r} < 0$ then the origin is a local **maximum** .

  - If $\mathsf{s}^2 - \mathsf{rt} > 0$ then the origin is not a local extremum, it is a **saddle point**.

  - If $\mathsf{s}^2 - \mathsf{rt} = 0$ then **one cannot conclude** directly.
    In this case we can use other means
    (for instance use a graphic, etc...)

• Case of an arbitrary function

- The previous test remains **unchanged** for a arbitrary function $f(x, y)$ and a stationary point located at (x_0, y_0).

- Only the signification of the parameters r, s, t changes

$$\mathsf{r} = \left(\frac{\partial^2 f}{\partial x^2}\right)_0 \quad , \quad \mathsf{s} = \left(\frac{\partial^2 f}{\partial x \partial y}\right)_0 \quad , \quad \mathsf{t} = \left(\frac{\partial^2 f}{\partial y^2}\right)_0$$

  where the index 0 means that teh derivatives are evaluated at the **point of coordinates $(\boldsymbol{x_0}, \boldsymbol{y_0})$**.

- These second derivatives are also important in the Taylor's development of $f(x, y)$...

- **Taylor series of a function of two variables**
  - The series of $f(x,y)$ about $(0,0)$ takes the general form of a **double sum**

$$f(x,y) \;\; = \;\; \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\left(\frac{\partial^{n+m} f}{\partial x^n \partial y^m}\right)_0 \frac{x^n}{n!}\frac{y^m}{m!}$$

  - The double sum can be written in a more **concise** manner

$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} = \sum_{n+m\geq 0}^{\infty}$$

  - For an **expansion about (a,b)** we just make substitutions

$$x^n \mapsto (x-a)^n \quad , \quad y^m \mapsto (y-b)^m \quad , \quad (\ldots)_0 \mapsto (\ldots)_{(a,b)}$$

To compute the successive terms of the double sum we have to consider the successive values of $n + m$.

- $n + m = 0$: 1 solution
  - $n = m = 0 \;\Rightarrow\; (\partial_x^0 \partial_y^0 f)_{(a,b)} = f(a, b)$
- $n + m = 1$: 2 solutions (linear terms)
  - $n = 1$ and $m = 0 \Rightarrow\; (\partial_x^1 \partial_y^0 f)_{(a,b)} = \partial_x f(a, b)$
  - $n = 0$ and $m = 1 \Rightarrow\; (\partial_x^0 \partial_y^1 f)_{(a,b)} = \partial_y f(a, b)$
- $n + m = 2$: 3 solutions (quadratic terms)
  - $n = 2$ and $m = 0 \Rightarrow\; (\partial_x^0 \partial_y^2 f)_{(a,b)} = \partial_x^2 f(a, b)$
  - $n = 1$ and $m = 1 \Rightarrow$ $$\begin{aligned}(\partial_x^1 \partial_y^1 f)_{(a,b)} &= \partial_{xy}^2 f(a, b) \\ &= \tfrac{1}{2}\partial_{xy}^2 f(a, b) + \tfrac{1}{2}\partial_{yx}^2 f(a, b)\end{aligned}$$
  - $n = 0$ and $m = 2 \Rightarrow\; (\partial_x^0 \partial_y^2 f)_{(a,b)} = \partial_y^2 f(a, b)$
- etc...

- **Conclusives remarks**

  - Note : We recognise the calculational scheme of **binomial coefficients** that show up in Newton's formula

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

  - Alternatively we can use **numerical) Newton method** to solve simultaneously $f_x = f_y = 0$ where second derivatives have to be evaluated.

  - When these second derivatives become to complicated to compute we can use the method of **gradient descent** (see optimization).