Mathematics for Economists: Lecture 3

Juuso Välimäki

Aalto University School of Business

Spring 2020

This lecture covers

1. Multivariate functions
   1.1 As a function of a single coordinate variable: partial derivative
   1.2 Linear approximation of a function around a point: the derivative
   1.3 Change in the value of the function in a given direction: directional derivative
   1.4 Fastest change: the gradient
   1.5 No change: approximating the level curves
2. Vector-valued multivariate functions
   2.1 The derivative of a vector-valued multivariate function
   2.2 The chain rule: Euler's theorem for homogenous functions

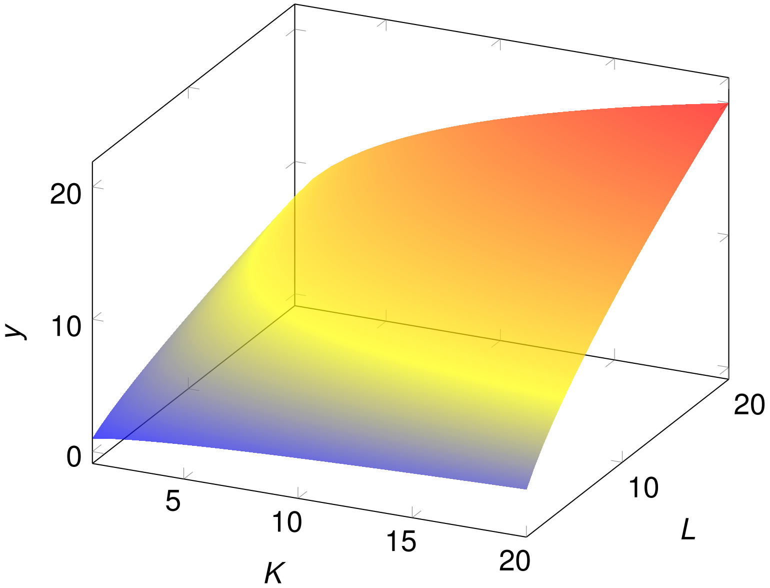
Partial derivative

- A firm uses capital, K, and labor L to produce output y.
- $\hat{K}$ units of capital and $\hat{L}$ units of labor produce $\hat{K}^{\frac{1}{3}}\hat{L}^{\frac{2}{3}}$ units of output
- We say that the firm has the *production function*

$$y = f(K, L) = K^{\frac{1}{3}}L^{\frac{2}{3}}.$$

- Let's draw the graph of the production function

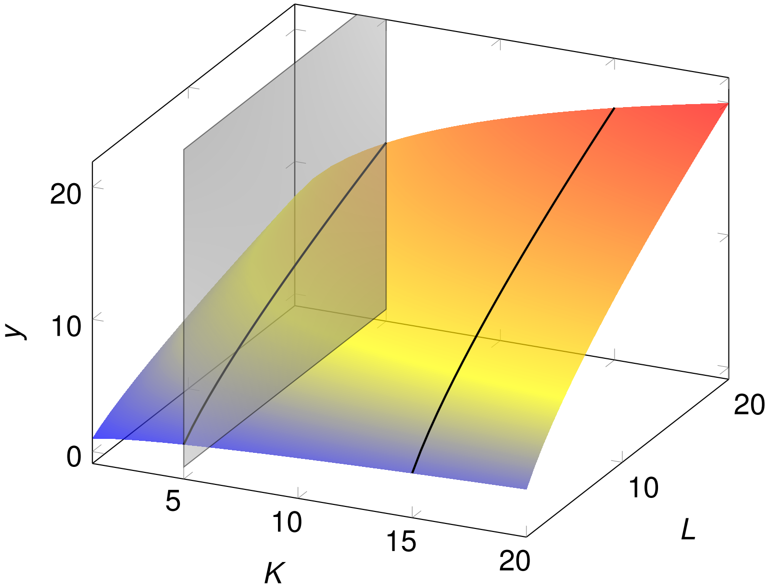
Graph for $y = f(K, L) = K^{\frac{1}{3}} L^{\frac{2}{3}}$



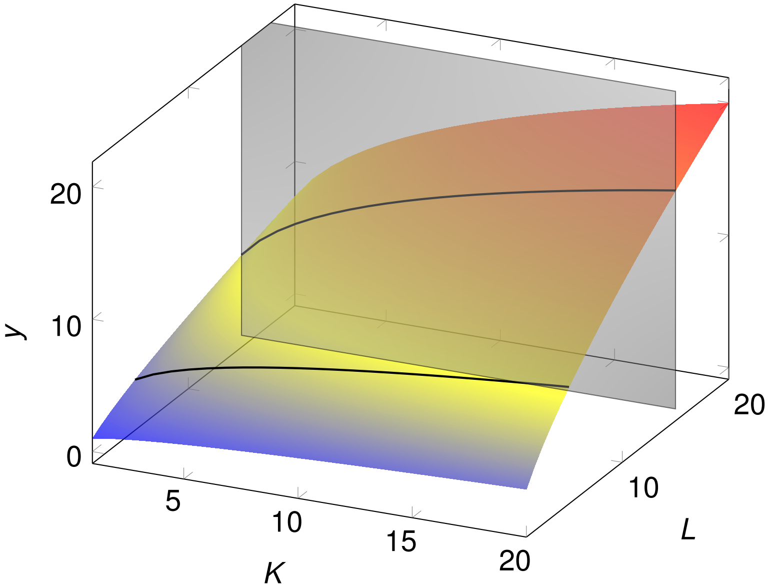
Slicing the production function

- Imagine that capital K is fixed at $\hat{K}$ for the firm
  - Then the production function is a function of L only
- Similarly for fixed $\hat{L}$, production depends only on K
- We can view the slices of the production function as follows:

Graph for $y = f(\hat{K}, L) = \hat{K}^{\frac{1}{3}} L^{\frac{2}{3}}$ for $\hat{K} = 5, 15$



Graph for $y = f(K, \hat{L}) = K^{\frac{1}{3}} L^{\frac{2}{3}}$ for $\hat{L} = 5, 15$



Partial derivative

- More systematically, start with $f : \mathbb{R}^n \to \mathbb{R}$, where we write

$$y = f(x) = f(x_1, x_2, ..., x_n) .$$

- For $j \neq i$, fix the values of the other variables at $x_j = \hat{x}_j$.
- Then we have the function

$$f(\hat{x}_1, ..., \hat{x}_{i-1}, x_i, \hat{x}_{i+1}, ..., \hat{x}_n).$$

- Note that $f(\hat{x}_1, ..., \hat{x}_{i-1}, x_i, \hat{x}_{i+1}, ..., \hat{x}_n)$ is a function of a single variable x_i
- We can then define the derivative of f with respect to x_i exactly as before around the point $\widehat{x} = (\widehat{x}_1, ..., \widehat{x}_n)$.

Partial derivative

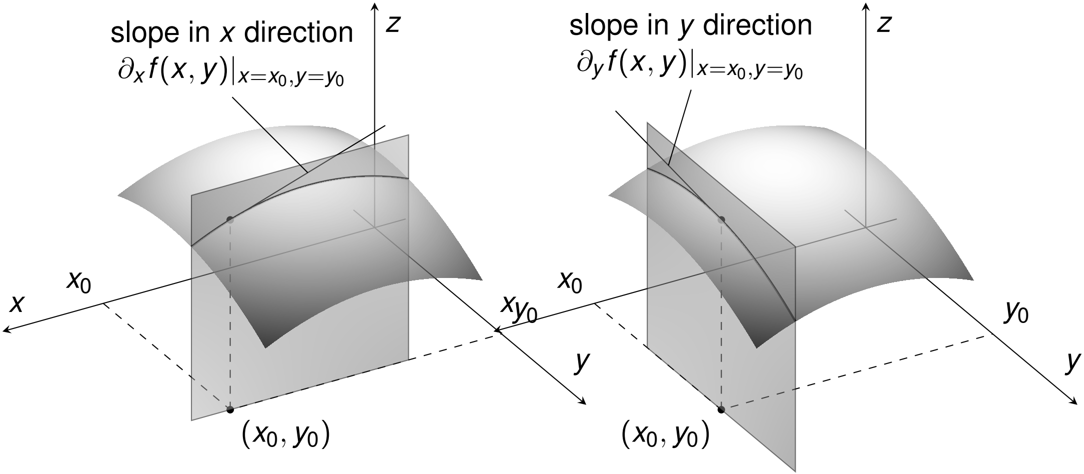
- Define the *partial derivative* of f with respect to x_i at $\hat{x}$ as the derivative of $f(\hat{x}_1, ..., \hat{x}_{i-1}, x_i, \hat{x}_{i+1}, ..., \hat{x}_n)$ at $\hat{x}$:

$$\frac{\partial f\left(\widehat{x}\right)}{\partial x_i} = \lim_{x \to \hat{x}_i} \frac{f\left(\widehat{x}_1, \widehat{x}_2, ..., \widehat{x}_{i-1}, x_i, \widehat{x}_{i+1}, ..., \widehat{x}_n\right) - f\left(\widehat{x}\right)}{x_i - \hat{x}_i}.$$

- Since we have defined the partial derivative as the derivative of $f(\hat{x}_1, ..., \hat{x}_{i-1}, x_i \hat{x}_{i+1}, ..., \hat{x}_n)$, all the rules for computing derivatives are valid for partial derivatives too
- For example, we have that if $\frac{\partial f(\widehat{x})}{\partial x_i} > 0$, then f is strictly increasing in variable x_i near point $\hat{x}$.
- Let's compute the partial derivative with respect to L at $(\hat{K}, \hat{L}) = (5, 10)$.

$$\frac{\partial f(\hat{K}, \hat{L})}{\partial L} = \frac{2}{3} 5^{\frac{1}{3}} 10^{\frac{-1}{3}} = \frac{2}{3} (\frac{1}{2})^{\frac{1}{3}} = 0.53.$$

Partial derivatives for $z = f(x, y)$



The derivative of a multivariate real-valued function

- We can also write:

$$\frac{\partial f\left(\widehat{x}\right)}{\partial x_i} = \lim_{h \to 0} \frac{f\left(\widehat{x} + h\mathbf{e}_i\right) - f\left(\widehat{x}\right)}{h},$$

- Here $\mathbf{e}_i$ is the i^{th} unit vector and $h = x_i - \hat{x}_i$.
- How to define the derivative of a multivariate function f at $\widehat{x}$? We could try with something along the lines:

$$Df\left(\widehat{x}\right) = \lim_{x \to \hat{x}} \frac{f\left(x\right) - f\left(\widehat{x}\right)}{x - \hat{x}}.$$

- This is nonsense since on the right hand side, we have a real number divided by a vector and this is not defined.

The derivative of a multivariate real-valued function

- How about finding a linear function at $\hat{x}$ to approximate the non-linear function f?
- From this point of view, we could say that f has a derivative at $\hat{x}$ if there exists a linear function $Df(\hat{x})$ such that for small $\Delta x = (\Delta x_1, ..., \Delta x_n)$:

$$f\left(\widehat{x} + \Delta x\right) - f(\hat{x}) = Df\left(\widehat{x}\right)\Delta x$$

The derivative of a multivariate real-valued function

- Approximation for multivariate functions
  - How to make the above statement mathematically precise?
  - We say that a vector $y \in \mathbb{R}^n$ is small if $\|y\| := \sqrt{\sum_i^n y_i^2}$ is small.
  - In particular, if $\|y\| < \varepsilon$, then $y_i < \varepsilon$ for all $i \in \{1, ..., n\}$.
  - Notice that by the Pythagorean theorem, $\|y\|$ measures the distance of y form the origin. This is also called the norm of vector y.
- The exact mathematical formulation for the statement above is thus the following: $\forall \varepsilon > 0, \exists \delta > 0$ such that

$$\frac{\left|f\left(\widehat{x} + \Delta x\right) - f(\hat{x}) - Df\left(\widehat{x}\right)\Delta x\right|}{\|\Delta x\|} < \varepsilon \text{ whenever } \|\Delta x\| < \delta.$$

- From now on, this is the formal mathematical meaning for our statements that assert equality of two objects whenever some quantity is small.

The derivative of a multivariate real-valued function

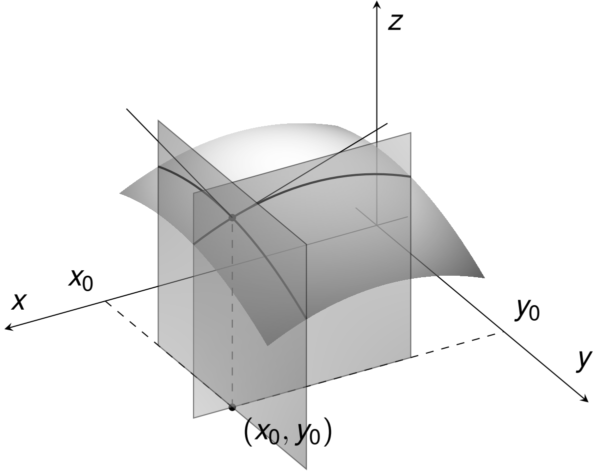
- Linear functions from $\mathbb{R}^n$ to $\mathbb{R}$ are given by an inner product with a vector $a \in \mathbb{R}^n$.
- In other words, $Df\left(\widehat{x}\right)$ is a row vector.
- Since $Df(\hat{x})$ is defined as a linear function, it is completely determined by its values for unit vectors.
- This means that we can write the derivative as:

$$Df\left(\widehat{x}\right) = \left(\frac{\partial f\left(\widehat{x}\right)}{\partial x_1}, \ldots, \frac{\partial f\left(\widehat{x}\right)}{\partial x_n}\right).$$

The derivative of a multivariate real-valued function

- One can also show that if all partial derivatives of f exist and if they are continuous in $\hat{x}$ at $\hat{x}$, then f has a derivative at $\hat{x}$.
- The following alternative way of denoting the partial derivative of f at $\widehat{x}$ is often used:

$$\frac{\partial f\left(\widehat{x}\right)}{\partial x_i} = f_{x_i}\left(\widehat{x}\right)$$


z
x
x0
y0
y
(x0, y0)

The gradient

- Sometimes one needs the column vector of partial derivatives at $\hat{x}$.
- It is called the gradient of f at $\hat{x}$ and denoted by:

$$\nabla f\left(\widehat{x}\right) = \begin{pmatrix} \frac{\partial f\left(\widehat{x}\right)}{\partial x_1} \\ \vdots \\ \frac{\partial f\left(\widehat{x}\right)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} f_{x_1}\left(\widehat{x}\right) \\ \vdots \\ f_{x_n}\left(\widehat{x}\right) \end{pmatrix} = Df(\hat{x})^\top .$$

The gradient

- The gradient can be given a geometric interpretation as the direction of fastest growth of f at $\widehat{x}$.
- We can ask for the direction Δx, that achieves the biggest change in the value of f.

$$f\left(\widehat{x} + \Delta x\right) - f\left(\widehat{x}\right) = Df\left(\widehat{x}\right) \cdot \Delta x,$$

- The quantity $Df\left(\widehat{x}\right) \cdot \Delta x$ is sometimes called the directional derivative of f at $\hat{x}$ in direction Δx.
- By the geometric interpretation of dot product, you will recall from high school that for fixed length vectors, the dot product is maximized when the vectors are parallel:

$$\Delta x = \nabla f\left(\widehat{x}\right).$$

- Gradient at $\hat{x}$ gives the direction of steepest increase for the value of f at $\hat{x}$.

Level curves

- In what direction is the value of a function $f : \mathbb{R}^n \to \mathbb{R}$ constant near a point $\hat{x}$?
- The directional derivative should be zero:

$$f(\hat{x} + \Delta x) - f(\hat{x}) = Df(\hat{x})\Delta x.$$

- Therefore for small Δx,

$$f(\hat{x} + \Delta x) - f(\hat{x}) = 0 \text{ if } \nabla f\left(\widehat{x}\right) \cdot \Delta x = 0.$$

- In other words, the value of the function does not change in the directions that are orthogonal to the gradient.
- Can you relate this to the slope of indifference curves? We will cover this in the next lecture.

The derivative of a vector-valued function

- A vector valued function $f : \mathbb{R}^n \to \mathbb{R}^m$ can be written as a vector of real valued multivariate functions:

$$y = f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix},$$

  where for each i,

$$f_i(x) : \mathbb{R}^n \to \mathbb{R}.$$

- Recall that for real-valued, f has a derivative at $\widehat{x}$ if there is a linear function $Df(\widehat{x})$ such that for small Δx,

$$f(\widehat{x} + \Delta x) - f(\widehat{x}) = Df(\widehat{x}) \cdot \Delta x.$$

- Now we want to apply the same definition to all component functions of f.

The derivative of a vector-valued function

- If f has a derivative at $\widehat{x}$, the results in the previous section imply that:

$$f_i\left(\widehat{x}+\Delta x\right)-f_i\left(\widehat{x}\right)=Df_i(\widehat{x})\cdot\Delta x.$$

- Then we can write the derivative of f at $\hat{x}$ as:

$$Df(\widehat{x})=\begin{pmatrix} Df_1\left(\widehat{x}\right) \\ \vdots \\ Df_m\left(\widehat{x}\right) \end{pmatrix}.$$

In other words,

$$Df(\widehat{x})=\begin{pmatrix} \frac{\partial f_1\left(\widehat{x}\right)}{\partial x_1} & \cdots & \frac{\partial f_1\left(\widehat{x}\right)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m\left(\widehat{x}\right)}{\partial x_1} & \cdots & \frac{\partial f_m\left(\widehat{x}\right)}{\partial x_n} \end{pmatrix}.$$

The derivative of a vector-valued function

- Again, it can be shown that when all the partial derivatives $\frac{\partial f_m(\hat{x})}{\partial x_n}$ exist and are continuous in $\hat{x}$, then the derivative $Df(\hat{x})$ exists.
- Since all rules for computing derivatives are valid for computing partial derivatives, many rules have multidimensional generalizations.
- In particular, the chain rule remains valid.
- If $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^k$, are both differentiable at $\hat{x}$ and $f(\hat{x})$ respectively, then the function $h(x) := g(f(x))$ is differentiable at $\hat{x}$ and

$$Dh(\hat{x}) = Dg(f(\hat{x}))Df(\hat{x}).$$

- Write out the sum in terms of partial derivatives to see the total effect of a change in x_i on a $z_j = g_j(f_1(x_1, ..., x_n), ..., f_m(x_1, ..., x_n))$.

Example: Euler's theorem on linearly homogenous functions

Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is called *linearly homogenous* or *homogenous of degree 1* if for all $\lambda \in \mathbb{R}$ and for all $x \in \mathbb{R}^n$,

$$f(\lambda x) = \lambda f(x).$$

Linear homogeneity is often assumed for production functions.

Proposition

If $f : \mathbb{R}^n \to \mathbb{R}$ is linearly homogenous, then

$$f(x) = \sum_{i=1}^{n} x_i \frac{\partial f(x)}{\partial x_i}.$$

Example: Euler's theorem on linearly homogenous functions

Proof.

Consider the two sides in the definition of linear homogeneity as functions of λ and take derivatives with respect to λ. The derivative of the right-hand side is $f(x)$. Denote the left-hand side by $h(\lambda)$. Then for fixed x,

$$h(\lambda) = f(g(\lambda)), g(\lambda) = \lambda x.$$

Notice that $g : \mathbb{R} \to \mathbb{R}^n$ and $Dg(\lambda) = x$. Since $f : \mathbb{R}^n \to \mathbb{R}$, $Df(x) = (\frac{\partial f(x)}{\partial x_1}, ..., \frac{\partial f(x)}{\partial x_n})$, and the claim follows by chain rule and evaluating the derivative at $\lambda = 1$ since

$$Dh(\lambda) = Df(\lambda x)Dg(\lambda) = (\frac{\partial f(\lambda x)}{\partial x_1}, ..., \frac{\partial f(\lambda x)}{\partial x_1}) \cdot x = \sum_{i=1}^{n} x_i \frac{\partial f(\lambda x)}{x_i}.$$

This is a very useful result for the analysis of firms and consumers. □

Applications: unconstrained optimization in $\mathbb{R}^n$

- A real-valued function $f(x)$ has a maximum at $\widehat{x}$ if

$$f(y) \leq f(\widehat{x}) \text{ for all } y \in \mathbb{R}^n.$$

- Since

$$y = \widehat{x} + (y - \widehat{x}),$$

  we can write:

$$f(\widehat{x} + (y - \widehat{x})) - f(\widehat{x}) \leq 0.$$

- For $\|y - \hat{x}\|$ small, we have

$$f(\widehat{x} + (y - \widehat{x})) - f(\widehat{x}) = Df(\widehat{x}) \cdot (y - \widehat{x}).$$

- If $\hat{x}$ is a maximum, then

$$Df\left(\widehat{x}\right) \cdot \left(y - \widehat{x}\right) \leq 0.$$

- Since y is arbitrary, we get a necessary condition for the optimum:

$$Df\left(\widehat{x}\right) = 0,$$

  or

$$\frac{\partial f\left(\widehat{x}\right)}{\partial x_i} = 0 \text{ for all } i.$$

- We will soon see examples of unconstrained optimization.

Application: comparative statics

- Assume that the endogenous variables $y \in \mathbb{R}^n$ and exogenous variables $x \in \mathbb{R}^k$ satisfy the equations

$$\begin{aligned} f_1(y, x) &= 0, \\ &\vdots \\ f_n(y, x) &= 0, \end{aligned} \tag{1}$$

at point $(\widehat{y}, \widehat{x})$.

- How do small changes in the exogenous variables affect the endogenous variables?
- Tool: Implicit function theorem. This is the main topic in the next few lectures. We are interested in two questions: i) Do solutions $(y(x), x)$ to 1 exist for x near $\hat{x}$ (i.e. $\|x - \hat{x}\|$ small)? ii) How can we characterize

$$y_1(x_1, ..., x_k),$$
$$\vdots$$
$$y_n(x_1, ..., x_k).$$

- To answer both of these questions, we must consider the derivative of f.

Computing the derivative

1. Compute at $(x_1, x_2, x_3) = (1, 2, 1)$ the derivative of the following function:

$$f(x_1, x_2, x_3) = x_1 \ln x_2 + \sqrt{x_2 x_3}.$$

- Real-valued function f: its derivative is the row vector of its partial derivatives at an arbitrary point $x = (x_1, x_2, x_3)$:

$$Df(x) = \left(\frac{\partial f(x_1, x_2, x_3)}{\partial x_1}, \frac{\partial f(x_1, x_2, x_3)}{\partial x_2}, \frac{\partial f(x_1, x_2, x_3)}{\partial x_3} \right)$$

$$= \left(\ln x_2, \frac{x_1}{x_2} + \frac{1}{2} x_2^{-\frac{1}{2}} x_3^{\frac{1}{2}}, \frac{1}{2} x_2^{\frac{1}{2}} x_3^{-\frac{1}{2}} \right).$$

- Evaluating at $(1, 2, 1)$

$$Df(1, 2, 1) = \left(\ln 2, \frac{1}{2} + \frac{1}{2\sqrt{2}}, \frac{\sqrt{2}}{2} \right).$$

2. Consider the function

$$f : \mathbb{R}^3 \to \mathbb{R}^2,$$

where

$$f(x,y,z) = \begin{pmatrix} f_1(x,y,z) \\ f_2(x,y,z) \end{pmatrix} = \begin{pmatrix} x^2 y + yz \\ \frac{y-z}{x} \end{pmatrix}.$$

- Our task is to compute the derivative of f at $(x,y,z) = (1,1,1)$.
- Form the derivative of f as a matrix of partial derivatives at an arbitrary $x \in \mathbb{R}^3$:

$$\begin{pmatrix} \frac{\partial f_i(x,y,z)}{\partial x} & \frac{\partial f_i(x,y,z)}{\partial y} & \frac{\partial f_i(x,y,z)}{\partial z} \\ \frac{\partial f_2(x,y,z)}{\partial x} & \frac{\partial f_2(x,y,z)}{\partial y} & \frac{\partial f_2(x,y,z)}{\partial z} \end{pmatrix} = \begin{pmatrix} 2xy & x^2 + z & y \\ -\frac{y-z}{x^2} & \frac{1}{x} & -\frac{1}{x} \end{pmatrix}.$$

- Evaluate the derivative at $(1,1,1)$:

$$Df(1,1,1) = \begin{pmatrix} 2 & 2 & 1 \\ 0 & 1 & -1 \end{pmatrix}.$$

3. The CES utility function

$$u(x_1, x_2, x_3) = \left(x_1^\rho + x_2^\rho + x_3^\rho\right)^{\frac{1}{\rho}}.$$

- Compute the partial derivatives at $x = (x_1, x_2, x_3)$.

$$\frac{\partial u}{\partial x_i}(x_1, x_2, x_3) = \frac{1}{\rho}\left(x_1^\rho + x_2^\rho + x_3^\rho\right)^{\frac{1}{\rho}-1} \rho x_i^{\rho-1} = x_i^{\rho-1}\left(x_1^\rho + x_2^\rho + x_3^\rho\right)^{\frac{1}{\rho}-1}.$$

- The gradient at x is then:

$$\nabla u(x_1, x_2, x_3) = \begin{pmatrix} x_1^{\rho-1}\left(x_1^\rho + x_2^\rho + x_3^\rho\right)^{\frac{1}{\rho}-1} \\ x_2^{\rho-1}\left(x_1^\rho + x_2^\rho + x_3^\rho\right)^{\frac{1}{\rho}-1} \\ x_3^{\rho-1}\left(x_1^\rho + x_2^\rho + x_3^\rho\right)^{\frac{1}{\rho}-1} \end{pmatrix}.$$

- Can you see the chain rule at work here?