Mathematics for Economists: Lecture 3

Juuso Välimäki

Aalto University School of Business

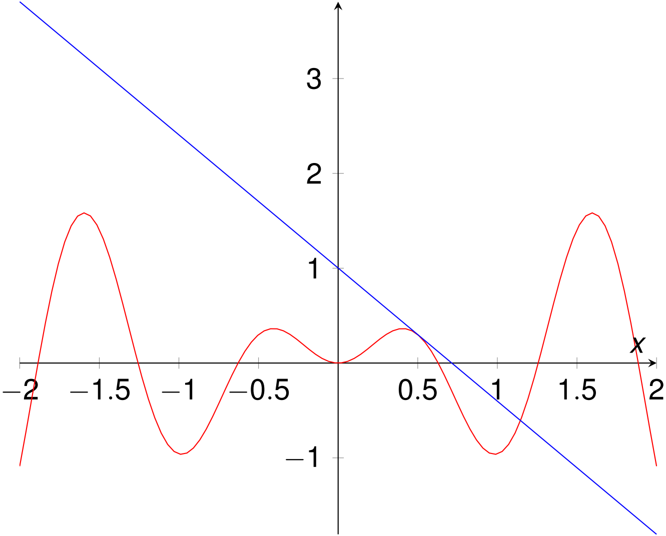
Spring 2021

Content of Lecture 3

- In Lecture 2, Gaussian elimination and linear models in economics
- This Lecture:
  1. Linear approximation of functions of a real variable: the derivative
  2. Visualizing multivariate functions
  3. Linear approximations to multivariate functions
  4. Linear approximations and partial derivatives
  5. Directions of increase and level curves
  6. Non-linear models in economics: first examples

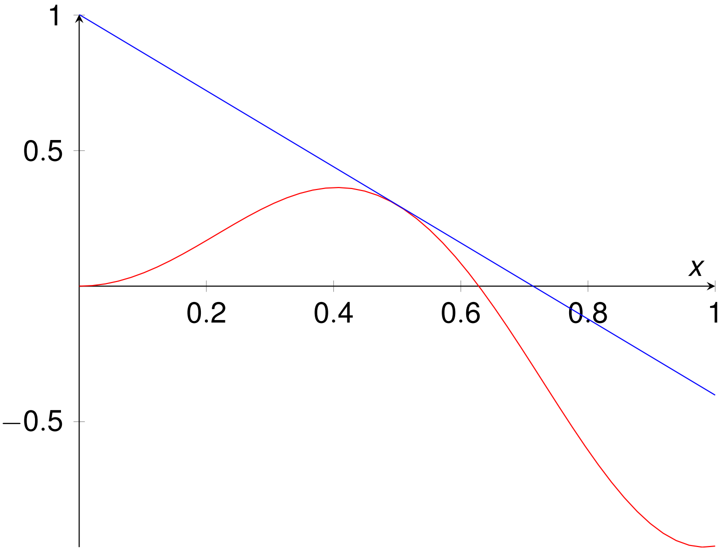
Linear approximation of univariate functions

Approximate $f(x) = xsin(5x)$ by its tangent at $x = 0.5$. Not a great success:



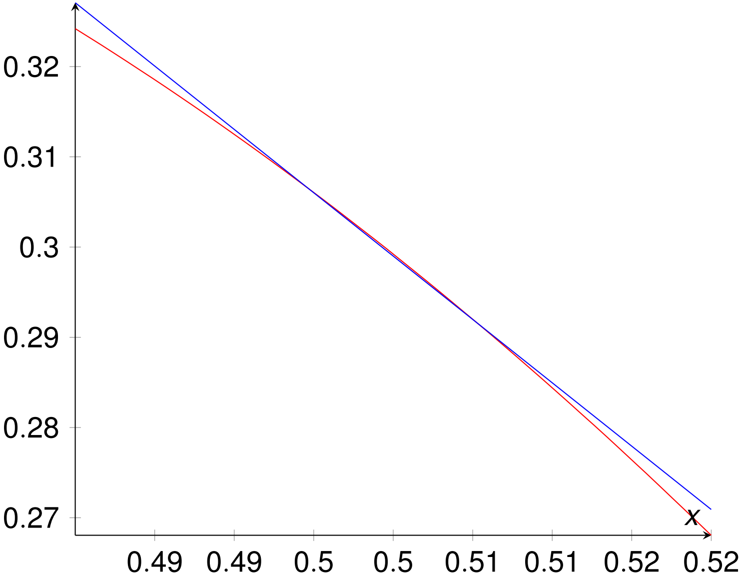
Linear approximation of univariate functions

The function is a bit less variable over the interval $[0, 1]$:



Linear approximation of univariate functions

On the interval $[0.48, 0.52]$, it looks almost linear:



Linear approximation of univariate functions

- Recall the definition of the derivative of a function $f : \mathbb{R} \to \mathbb{R}$ at x_0:

$$Df(x_0) = \frac{df(x_0)}{dx} = f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

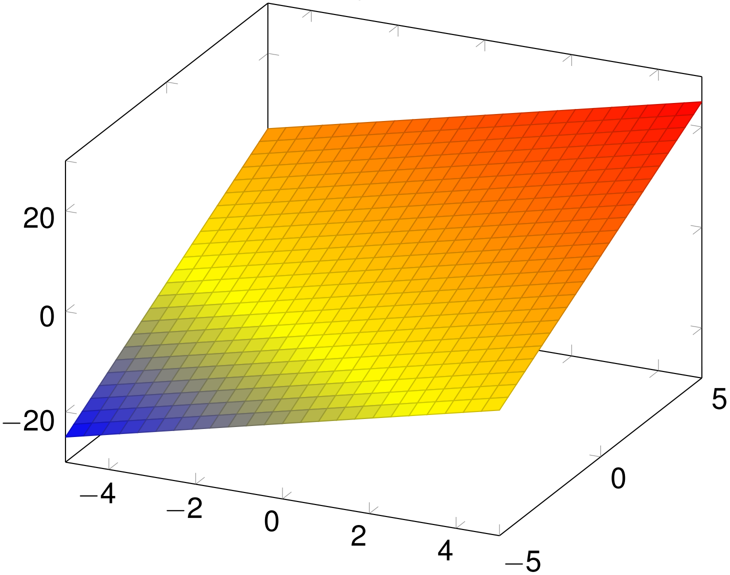
- If the limit exists, we also have (from the definition of limits) that for all $\varepsilon > 0$, there is a $\delta > 0$ such that:

$$\frac{|f(x_0 + h) - f(x_0) - Df(x_0)h|}{|h|} < \varepsilon,$$

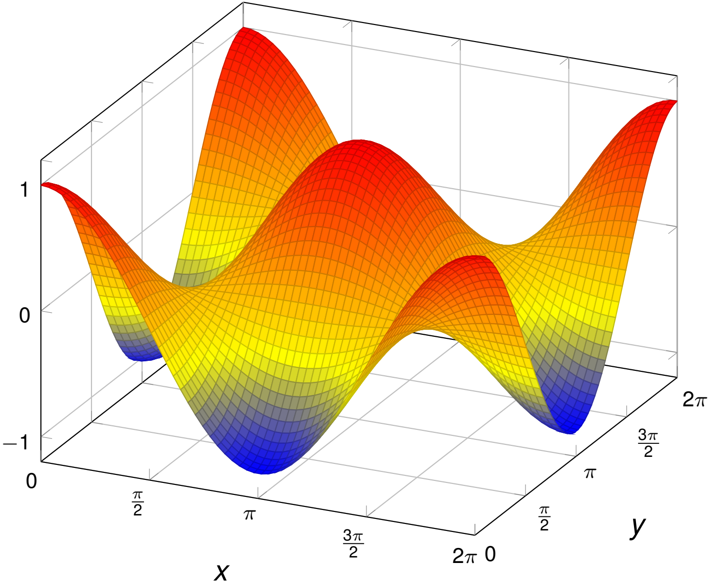
  whenever $|h| < \delta$.
- We say then that $Df(x_0)h$ approximates $f(x) - f(x_0)$ well near x_0.
- $Df(x_0)h$ is a linear (in h) approximation of the changes in the value of the function near x_0.
- We want to generalize this idea to multivariate functions.

Graphing functions of two variables

A linear function: $z = 2x - 3y$.



The graph of $f(x, y) = cos(x)cos(y)$



2-d slices of the 3-d graph

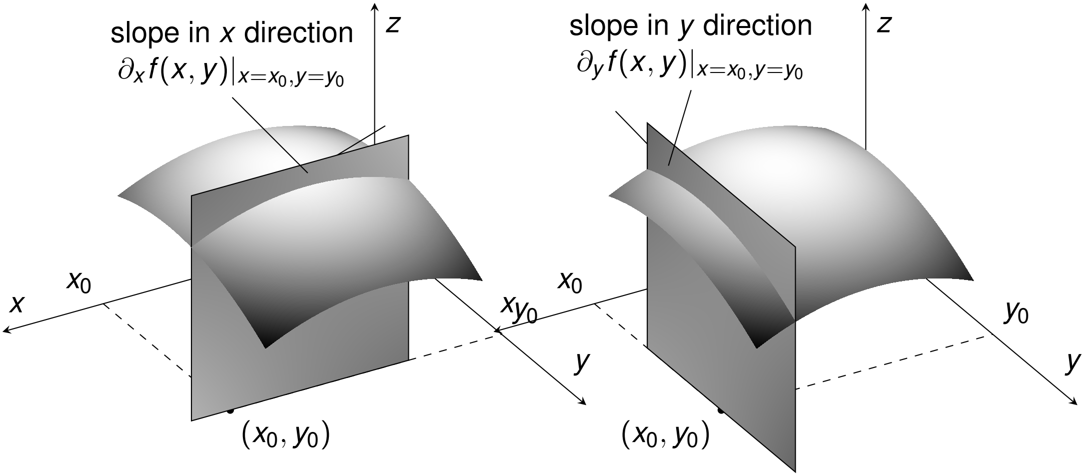


Figure: Cross sections of $f(x, y)$ in the x and y direction at (x_0, y_0).

Level curves of a bivariate function

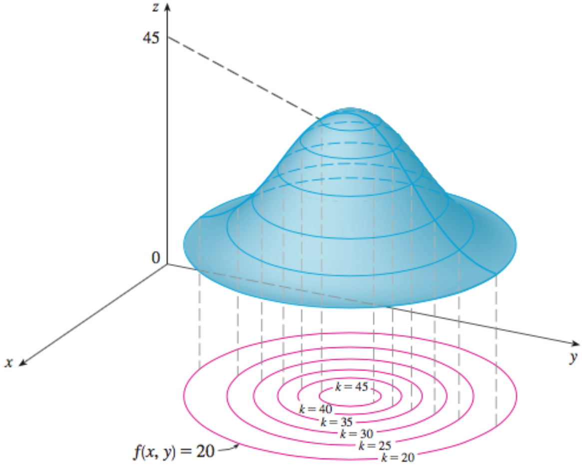


Figure: Some level curves of f.

All in one picture

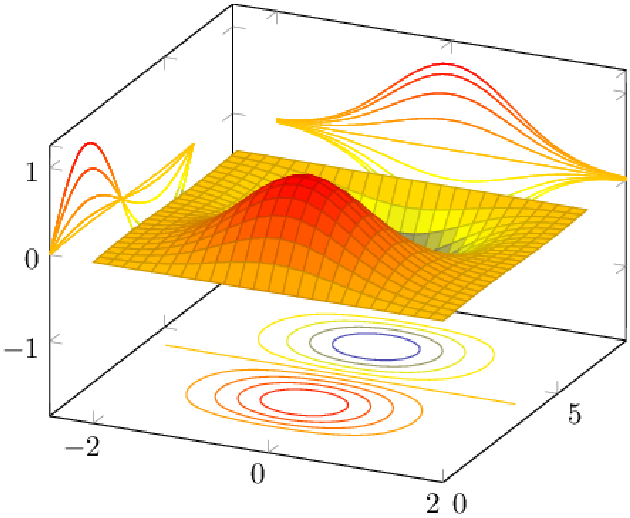


Figure: The graph of f together with some of its cross sections and level curves.

Linear functions

- Recall that a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be linear if for all $\lambda \in /R$ and for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, i) $f(\lambda \boldsymbol{x}) = \lambda f(\boldsymbol{x})$, ii) $f(\boldsymbol{x} + \boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y})$.
- For $f : \mathbb{R}^n \to \mathbb{R}$, let $f(\boldsymbol{e}^i) = a_i$, where $\boldsymbol{e}^i = (e_1^1, ..., e_n^1)$ is the i^{th} unit vector, i.e. $e_j^i = 0$ if $i \neq j$ and $e_i^i = 1$.
- Then we have

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} f(x_i \boldsymbol{e}^i) = \sum_{i=1}^{n} x_i f(\boldsymbol{e}^i) = \sum_{i=1}^{n} a_i x_i = \boldsymbol{a} \cdot \boldsymbol{x},$$

  where $\boldsymbol{a} = (a_1, ..., a_n)$.
- For $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$, let $f(\boldsymbol{e}^i) = \boldsymbol{a}^i \in \mathbb{R}^m$. By the same reasoning as above, any linear $\boldsymbol{f}$ takes the form:

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x},$$

  where the ij^{th} element of $\boldsymbol{A}$ is the i^{th} row element of $\boldsymbol{a}^j$.

Linear approximation of multivariate functions

- If we want to find a linear approximation to $f : \mathbb{R}^n \to \mathbb{R}$ near $\boldsymbol{x}_0$, we look for a vector $\boldsymbol{a}$ such that $\frac{|f(\boldsymbol{x}_0+h)-f(\boldsymbol{x}_0)-\boldsymbol{a}\cdot\boldsymbol{x}|}{\|\boldsymbol{x}-\boldsymbol{x}_0\|}$ is small whenever $\|\boldsymbol{x}-\boldsymbol{x}_0\|$ is small.
- If such an $\boldsymbol{a}$ exists, we call it the derivative $D_{\boldsymbol{x}}f(\boldsymbol{x}_0)$ of f at $\boldsymbol{x}_0$ and we say that f is differentiable at $\boldsymbol{x}_0$.
- When the derivative exists, we have for small $\|\boldsymbol{x}-\boldsymbol{x}_0\|$:

$$f(\boldsymbol{x}) - f(\boldsymbol{x}_0) \approx D_{\boldsymbol{x}}f(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0).$$

- Our next task is to identify the components of $D_{\boldsymbol{x}}f(\boldsymbol{x}_0)$.

Partial derivatives

- Consider changes in the direction of a coordinate axis: $\boldsymbol{x} = \boldsymbol{x}_0 + h\boldsymbol{e}^i$.
- Since all the other coordinates of $\boldsymbol{x}$ remain fixed, we can compute:

$$\lim_{h \to 0} \frac{f(\boldsymbol{x}_0 + h\boldsymbol{e}^i) - f(\boldsymbol{x}_0)}{h},$$

  exactly as in the case of univariate functions,
- We call this limit the i^{th} partial derivative of f at $\boldsymbol{x}_0$ and denote it by:

$$D_{x_i} f(\boldsymbol{x}_0) =: \frac{\partial f(\boldsymbol{x}_0)}{\partial x_i} := \lim_{h \to 0} \frac{f(\boldsymbol{x}_0 + h\boldsymbol{e}^i) - f(\boldsymbol{x}_0)}{h}.$$

Partial derivatives

Recall the picture from before:

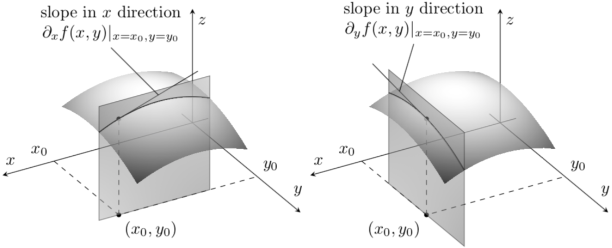


Figure: Partial derivatives of f at (x_0, y_0).

From partial derivatives to linear approximation

- But this is all we need if a linear approximation exists!
- A linear approximation in the direction $\Delta \boldsymbol{x} = \boldsymbol{e}^i$ must coincide with $\frac{\partial f(\boldsymbol{x}_0)}{\partial x_i}$.
- But each direction $\Delta \boldsymbol{x}$ can be written as:

$$\Delta \boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}^i,$$

so by linearity we get for all $\Delta \boldsymbol{x}$ that

$$f(\boldsymbol{x}_0 + \Delta \boldsymbol{x}) - f(\boldsymbol{x}_0) = D_{\boldsymbol{x}} f(\boldsymbol{x}_0) \cdot \Delta \boldsymbol{x} + \text{ h.o.t. },$$

where $D_{\boldsymbol{x}} f(\boldsymbol{x}_0)$ is the row vector of partial derivatives

$$D_{\boldsymbol{x}} f(\boldsymbol{x}) = \left(\frac{\partial f(\boldsymbol{x}_0)}{\partial x_1}, \ldots, \frac{\partial f(\boldsymbol{x}_0)}{\partial x_n}\right).$$

- A linear approximation exists and f is differentiable at $\boldsymbol{x}_0$ if all of its partial derivatives exist at $\boldsymbol{x}_0$ and are continuous in $\boldsymbol{x}$.

Planar approximation to a non-linear surface

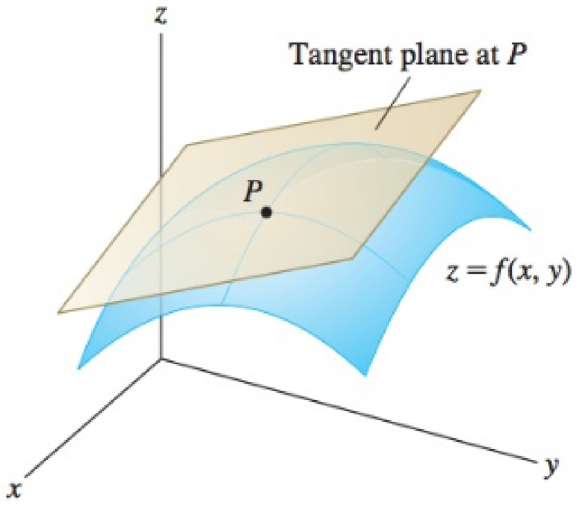


Figure: Linear approximation to f at point P.

Computing the derivative: an example

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

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

- Since we have a real-valued function f, its derivative is the row vector of its partial derivatives evaluated $x = (x_1, x_2, x_3)$:

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

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

Evaluating at $(1, 2, 1)$

$$D_{\boldsymbol{x}} f(1, 2, 1) = \left(\ln 2, \frac{1}{2} + \frac{1}{2\sqrt{2}}, \frac{\sqrt{2}}{2} \right).$$

Utility functions: marginal utilities

- Utility function $u : \mathbb{R}^n_+ \to \mathbb{R}$ assigns a numerical value $u(\boldsymbol{x})$ for each possible (positive) consumption vector $\boldsymbol{x} \in \mathbb{R}^n_+ := \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{x} \geq 0\}$.
- $u(\boldsymbol{x}) \geq u(\boldsymbol{y})$ if and only if the consumer considers $\boldsymbol{x}$ at least as good as $\boldsymbol{y}$
- Consider first the case with two goods, i.e. $n = 2$.
- Partial derivatives of the utility function are called the *marginal utilities* denoted by MU_{x_i}.

$$MU_{x_i}(\hat{x}_1, \hat{x}_2) := \frac{\partial u(\hat{x}_1, \hat{x}_2)}{\partial x_i}.$$

- If $\frac{\partial u(\hat{x}_1, \hat{x}_2)}{\partial x_1} > 0$, then $u(\hat{x}_1 + h, \hat{x}_2) > u(\hat{x}_1, \hat{x}_2)$ for small $h > 0$, and we say that utility is strictly increasing in good 1 at $(\hat{x}_1, \hat{x}_2)$.
- If this holds at all (x_1, x_2), we say simply that utility is strictly increasing. Typically it is assumed that utility is strictly increasing in all goods.

Utility functions: marginal utilities

- For small consumption changes $(\Delta x_1, \Delta x_2)$, we can approximate the change in utility by using the derivative $D_{\mathbf{x}} u(\hat{x}_1, \hat{x}_2)$:

$$u(\hat{x}_1 + \Delta x_1, \hat{x}_2 + \Delta x_2) - u(\hat{x}_1, \hat{x}_2) = D_{\mathbf{x}} u(\hat{x}_1, \hat{x}_2)(\Delta x_1, \Delta x_2)$$
$$= \frac{\partial u(\hat{x}_1, \hat{x}_2)}{\partial x_1}\Delta x_1 + \frac{\partial u(\hat{x}_1, \hat{x}_2)}{\partial x_2}\Delta x_2.$$

- Recall from Principles 1 that $(\hat{x}_1 + \Delta x_1, \hat{x}_2 + \Delta x_2)$ and $(\hat{x}_1, \hat{x}_2)$ are on the same indifference curve if they are equally good to the consumer: $u(\hat{x}_1 + \Delta x_1, \hat{x}_2 + \Delta x_2) = u(\hat{x}_1, \hat{x}_2)$.
- But then we have:

$$\frac{\partial u(\hat{x}_1, \hat{x}_2)}{\partial x_1}\Delta x_1 + \frac{\partial u(\hat{x}_1, \hat{x}_2)}{\partial x_2}\Delta x_2 = 0,$$

or

$$\Delta x_2 = -\frac{\frac{\partial u(\hat{x}_1, \hat{x}_2)}{\partial x_1}}{\frac{\partial u(\hat{x}_1, \hat{x}_2)}{\partial x_2}}\Delta x_1.$$

Utility functions: MRS

- The consumer is willing to give up $\frac{\frac{\partial u(\hat{x}_1,\hat{x}_2)}{\partial x_1}}{\frac{\partial u(\hat{x}_1,\hat{x}_2)}{\partial x_2}}$ units of good 2 to get an additional unit of good 1 at $(\hat{x}_1, \hat{x}_2)$.
- Hence marginal rate of substitution at $(\hat{x}_1, \hat{x}_2)$ is captured in the ratio of marginal utilities:

$$MRS_{x_1,x_2}(\hat{x}_1, \hat{x}_2) = \frac{\frac{\partial u(\hat{x}_1,\hat{x}_2)}{\partial x_1}}{\frac{\partial u(\hat{x}_1,\hat{x}_2)}{\partial x_2}} = \frac{MU_{x_1}(\hat{x}_1, \hat{x}_2)}{MU_{x_2}(\hat{x}_1, \hat{x}_2)}.$$

Utility functions: MRS

- If $n > 2$ we can ask how many (small) units of good j the consumer would be willing to give up in order to get an additional (small) unit of good i.
- If all the other goods remain fixed at $\hat{\boldsymbol{x}}$ and

$$u(\hat{\boldsymbol{x}} + \Delta x_i \boldsymbol{e}^i + \Delta x_j \boldsymbol{e}^j) = u(\hat{\boldsymbol{x}}),$$

then

$$\frac{\partial u(\hat{x}_1, \hat{x}_2)}{\partial x_i} \Delta x_i + \frac{\partial u(\hat{x}_1, \hat{x}_2)}{\partial x_j} \Delta x_j = 0,$$

and we have:

$$MRS_{x_i, x_j}(\hat{\boldsymbol{x}}) = \frac{\frac{\partial u(\hat{\boldsymbol{x}})}{\partial x_i}}{\frac{\partial u(\hat{\boldsymbol{x}})}{\partial x_j}} = \frac{MU_{x_i}(\hat{\boldsymbol{x}})}{MU_{x_j}(\hat{\boldsymbol{x}})}.$$

The gradient

- The gradient of the utility function denoted by $\nabla u(\boldsymbol{x})$ is the transpose of its derivative:

$$\nabla u(\boldsymbol{x}) = (\frac{\partial u(\boldsymbol{x})}{\partial x_1}, ..., \frac{\partial u(\boldsymbol{x})}{\partial x_n}).$$

- Does the gradient have any particular interpretation?
- A first observation is that when $n = 2$, the gradient is orthogonal to the indifference curve:

$$(1, -\frac{\frac{\partial u(\hat{x}_1,\hat{x}_2)}{\partial x_1}}{\frac{\partial u(\hat{x}_1,\hat{x}_2)}{\partial x_2}}) \cdot (\frac{\partial u(\hat{x}_1,\hat{x}_2)}{\partial x_1}, \frac{\partial u(\hat{x}_1,\hat{x}_2)}{\partial x_2}) = \frac{\partial u(\hat{x}_1,\hat{x}_2)}{\partial x_1} - \frac{\frac{\partial u(\hat{x}_1,\hat{x}_2)}{\partial x_1}}{\frac{\partial u(\hat{x}_1,\hat{x}_2)}{\partial x_2}} \frac{\partial u(\hat{x}_1,\hat{x}_2)}{\partial x_2} = 0.$$

The gradient

- The gradient at $\boldsymbol{x}$ gives the direction in which the utility function increases the fastest near $\boldsymbol{x}$.
- To see this, consider the change in the utilityusing the linear approximation:

$$u(\hat{\boldsymbol{x}} + \Delta \boldsymbol{x}) - u(\hat{\boldsymbol{x}}) = D_{\boldsymbol{x}}(\hat{\boldsymbol{x}})\Delta \boldsymbol{x} + h.o.t.$$

- For a unit length (or norm) of $\Delta \boldsymbol{x}$, the change in utility is maximized at

$$\Delta \boldsymbol{x} = \frac{1}{\|\nabla u(\hat{\boldsymbol{x}})\|} \nabla u(\hat{\boldsymbol{x}})$$

by Cauchy's inequality.

Computing marginal utilities and the MRS

- Linear utility:

$$u(\boldsymbol{x}) = \boldsymbol{a} \cdot \boldsymbol{x} = \sum_{i=1}^{n} a_i x_i.$$

Then $MU_{x_i}(\boldsymbol{x}) = a_i$ for all i and all $\boldsymbol{x}$, and $MRS_{x_i,x_j} = \frac{a_i}{a_j}$ for all $i \neq j$ and all $\boldsymbol{x}$.

- Quasilinear utility:

$$u(x_1, x_2) = v(x_1) + x_2,$$

for some increasing function v.

$$MU_{x_2} = 1, MU_{x_1} = MRS_{x_1,x_2} = v'(x_1).$$

For example if $v(x_1) = \ln(x_1)$, then

$$MU_{x_1} = MRS_{x_1,x_2} = \frac{1}{x_1}.$$

Computing marginal utilities and the MRS

- Cobb-Douglas utility:

$$u(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha} \text{ for } \alpha \in (0, 1).$$

In this case,

$$MU_{x_1}(x_1, x_2) = \alpha(\frac{x_2}{x_1})^{1-\alpha}, \quad MU_{x_2}(x_1, x_2) = (1-\alpha)(\frac{x_1}{x_2})^{\alpha},$$

and therefore:

$$MRS_{x_1, x_2} = \frac{\alpha(\frac{x_2}{x_1})^{1-\alpha}}{(1-\alpha)(\frac{x_1}{x_2})^{\alpha}} = \frac{\alpha x_2}{(1-\alpha)x_1}.$$

Computing marginal utilities and the MRS

- For $n > 2$, we have:

$$u(\boldsymbol{x}) = \prod_{i=1}^{n} x_i^{\alpha_i} \text{ for } \alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i = 1.$$

Denote $y = \prod_{i=1}^{n} x_i^{\alpha_i}$. Then

$$MU_{x_i} = \alpha_i \frac{y}{x_i},$$

and therefore

$$MRS_{x_i,x_j}(\boldsymbol{x}) = \frac{\alpha_i x_j}{\alpha_j x_i}.$$

Exercise: Compute the marginal utilities for $u(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i \ln x_i$.

Computing marginal utilities and the MRS

- Constant elasticity of substitution utility (CES utility function): for $\rho < 1$ and $\rho \neq 0$,
$$u(x_1, x_2) = (a_1 x_1^\rho + a_2 x_2^\rho)^{\frac{1}{\rho}}.$$
Denote $y(x_1, x_2) = (a_1 x_1^\rho + a_2 x_2^\rho)$ and we have $u(x_1, x_2) = y(x_1, x_2)^{\frac{1}{\rho}}$. We get by chain rule that
$$MU_{x_i}(x_1, x_2) = \frac{1}{\rho} y(x_1, x_2)^{\frac{1-\rho}{\rho}} \rho a_i x_i^{\rho-1}.$$
Therefore the marginal rates of substitution are quite simple:
$$MRS_{x_1,x_2}(x_1, x_2) = \frac{a_1}{a_2} (\frac{x_1}{x_2})^{\rho-1}.$$

Next Lecture

- Vector valued multivariate functions
- Implicit function theorem
- Comparative statics in economic models