Mathematics for Economists: Lecture 2

Juuso Välimäki

Aalto University School of Business

Spring 2020

This lecture covers

1. Solving linear models
   1.1 One equation
   1.2 Many equations
2. Non-linear functions of a single variable
   2.1 Examples
   2.2 Derivative
   2.3 Rules for derivatives
3. Multivariate functions
   3.1 How to picture them
   3.2 Graph vs. level curves

Linear models

- A single variable: $b \in \mathbb{R}$ and $x \in \mathbb{R}$ with

$$b = ax.$$

- Solving: $x = a^{-1}b$. When is this valid, i.e. when does a solution exist? If $a \neq 0$. If $a = 0$, then there are no solutions if $b \neq 0$ and infinitely many solutions if $b = 0$.
- The only possibilities are: a single solution, 0 solutions and infinitely many solutions.
- We note that if $a \neq 0$, $f = ax$ is injective (since $x \neq x' \Rightarrow x \neq x'$) and surjective (since $a \times a^{-1}b = b$) and it is neither injective nor surjective nor surjective when $a = 0$.
- We want to generalize this idea to m linear equations in n real variables.

Multidimensional linear models

- Warm-up: let $b \in \mathbb{R}$ and consider the function $f(x_1, x_2) = a_1 x_1 + a_2 x_2$ and the equation:
$$a_1 x_1 + a_2 x_2 = b.$$
- This equation has no solutions if $a_1 = a_2 = 0$ and $b \neq 0$. In all other cases it has infinitely many solutions. f in never injective and surjective only if $a_1 \neq$ or $a_2 \neq 0$.
- At the other end, we have the model where $x \in \mathbb{R}$, but $f(x) \in \mathbb{R}^2$, say $y = f(x) = (y_1, y_2) = (a_1 x, a_2 x)$. Recall that we take vectors to be column vectors so you should really visualize this as:
$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_1 x \\ a_2 x \end{pmatrix}$$

Multidimensional linear models

- The corresponding set of equations for $b = (b_1, b_2) \in \mathbb{R}^2$ is:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 x \\ a_2 x \end{pmatrix}$$

- This can have solutions only if $\frac{b_1}{b_2} = \frac{a-1}{a_2}$ or $b = (b_1 = b_2) = 0$. Hence f is not surjective. It is injective if $a_1 \neq 0$ or $a_2 \neq 0$.
- If we want to have any hope for a bijective linear function $f : \mathbb{R}^n \to \mathbb{R}^m$, then we must have $n = m$
- When is $f : \mathbb{R}^2 \to \mathbb{R}^2$ bijective?

Multidimensional linear models

- Write f using the matrix

$$f(x) = Ax = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

- Let's see if the following set of equations has a solution for all $b \in \mathbb{R}^2$.

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

- Writing the equation in full:

$$\begin{array}{ccccc} a_{11}x_1 & + & a_{12}x_2 & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & = & b_2 \end{array}$$

Multidimensional linear models

- Solve for x_1 from the first equation and substitute into the second to get

$$-\frac{a_{21}a_{12}}{a_{11}}x_2 + a_{22}x_2 = -\frac{a_{21}}{a_{11}}b_1 + b_2.$$

- This is the same as multiplying the first equation by $-\frac{a_{21}}{a_{11}}$ and adding to the second equation.
- This is valid only if $a_{11} \neq 0$. If not, then solve similarly x_1 from the second equation. If also $a_{21} = 0$, then we are back to the case on the previous slide.
- we can solve for x_2 from the equation above if $a_{22} \neq \frac{a_{21}a_{12}}{a_{11}}$, i.e. if the determinant $a_{11}a_{22} - a_{21}a_{12} \neq 0$.
- Once we have x_2, we can get x_1 from the first equation.
- Notice that the determinant is zero only if $\frac{a_{11}}{a_{21}} = \frac{a_{12}}{a_{22}}$, i.e. row 2 is equal to row 1 multiplied by some non-zero real number.
- If the determinant is zero, then a solution exists only if $\frac{a_{11}}{a_{21}} = \frac{b_1}{b_2}$.

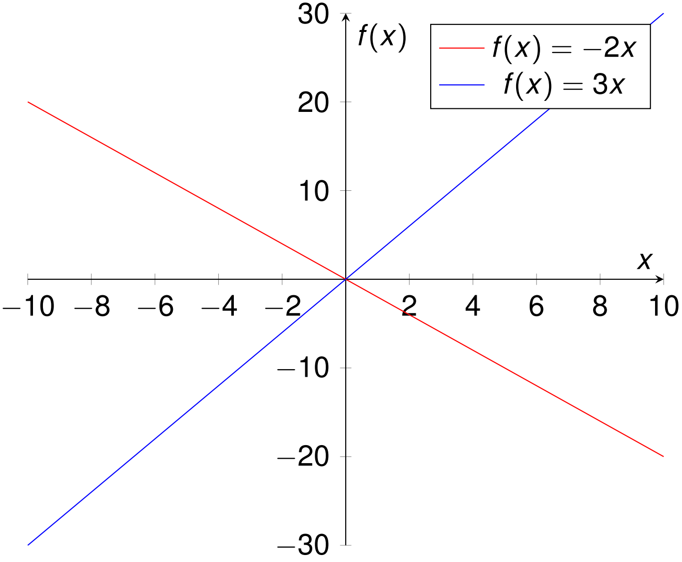
Multidimensional linear models

- We can extend this process of eliminating variables to systems of n linear systems in n variables and it is called Gaussian elimination.
- The principle is simple: If an equation holds so that its left-hand side equals its right-hand side, then the equation with both sides multiplied by an arbitrary real number also holds.
- If two equations hold simultaneously, then the equation obtained by summing together the two left-hand sides and the two right-hand sides also holds.
- The solution to a system of equations is unchanged if two rows in the system are swapped.
- Gaussian elimination is covered in the notes on Basic Linear Algebra and you can ask for additional demonstrations of the method in the review session tomorrow.

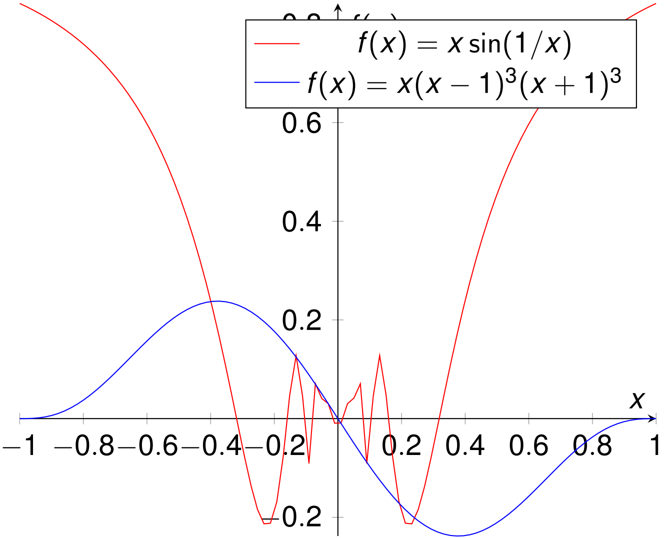
Multidimensional linear models

- The rows in a matrix are linearly dependent if one of the rows can be expressed as a sum of the other rows multiplied by real numbers.
- A system of equations has a unique solution if and only if the rows in the matrix representing the system are linearly independent (i.e. not dependent).
- Gaussian elimination determines the dependence of the rows.
- The row rank of a matrix is the maximal number of linearly independent rows in the matrix,
- The function $f(x) = Ax$ is bijective if and only A has full rank.
- An $n \times n$ -matrix has full rank if and only its determinant is non-zero.
- With these reminders from matrix algebra, we leave the linear models for now.

Non-linear functions of a single variable



Non-linear functions of a single variable



Non-linear functions of a single variable

- As you can see, the linear functions are very regular: they look the same at all x
- Non-linear ones can be quite irregular: their behavior is very different at different x.
- One of the most basic questions you can ask about a function of a real variable is if it increases in x or decreases in x
- f is said to be *increasing* if $x > x' \Rightarrow f(x) \geq f(x')$. It is said to be *strictly increasing* if $x > x' \Rightarrow f(x) > f(x')$
- f is said to be *decreasing* if $x > x' \Rightarrow f(x) \leq f(x')$. It is said to be *strictly decreasing* if $x > x' \Rightarrow f(x) < f(x')$
- By eyeballing, you can see that non-linear functions are sometimes increasing, sometimes whereas linear function are always either increasing or always decreasing.

The derivative

- Consider a function $f : \mathbb{R} \to \mathbb{R}$ near point $x_0 \in \mathbb{R}$.
- Is the function increasing or decreasing there?
- Form the difference quotient

$$\frac{f(x) - f(x_0)}{x - x_0}$$

and consider the limit

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

If the limit exists, we denote it by

$$Df(x_0)$$

or sometimes by

$$f'(x_0)$$

and call it the derivative of f at x_0.

The derivative

- If f has a derivative at x_0, we say that it is *differentiable at* x_0. If it differentiable at all points $x_0 \in \mathbb{R}$, we just say that it is differentiable.
- When a derivative exists, we have from the definition of limits:

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ such that whenever } |x - x_0| < \delta,$$

$$\text{we have: } \left|\frac{f(x) - f(x_0)}{x - x_0} - Df(x_0)\right| < \varepsilon.$$

- In this case, near x_0, the changes in $f(x)$ are well approximated by the linear function $Df(x_0)(x - x_0)$ of changes in x.
- This view of the derivative generalizes to functions of many variables and also to vector-valued functions.
- From now on, we simply say for differentiable functions that for x near x_0 or for small $|x - x_0|$,

$$f(x) = f(x_0) + Df(x_0)(x - x_0).$$

The derivative

- Only continuous functions can be differentiable
- Not all continuous functions are differentiable
- If a function is increasing near x_0, then $Df(x_0) \geq 0$
- If f is decreasing near x_0, then $Df(x_0) \leq 0$
- The derivative of a linear function $f = ax$ is a at all x so the function looks th same everywhere
- For nonlinear functions such as $f(x) = x^2$, the derivative $2x_0$ is different at all points x_0

Rules for computing derivatives

1. If $f(x) = a$ for all x, then $Df(x_0) = 0$ for all x_0.
2. If $f(x) = x$ then $Df(x_0) = 1$ for all x_0.
   (Linear homogeneity) If $g(x) = af(x)$, then $Dg(x_0) = aDf(x_0)$.
3. (Additivity) Let $h(x) = f(x) + g(x)$. Then $Dh(x_0) = Df(x_0) + Dg(x_0)$.

Rules for computing derivatives

4. (Product rule) Let
$$\phi(x) = f(x)\, g(x)\,.$$
Then
$$D\phi(x_0) = Df(x)\, g(x)\, f'(x_0)\, g(x_0) + f(x_0)\, g'(x_0)\,.$$
5. (Chain rule) Let
$$\zeta(x) = g(f(x))\,.$$
Then
$$D\zeta(x_0) = Dg(f(x)) = g'(f(x_0))\, f'(x_0)$$

Rules for computing derivatives

6. With these formulas, we can compute most derivatives that we need. For example, the derivative $D\phi(x_0)$ at x_0 for the function

$$\phi(x) = x^2$$

is obtained from the product rule

$$f(x) = g(x) = x.$$

We get

$$D\phi(x_0) = x_0 + x_0 = 2x_0.$$

By 'mathematical induction', we can see that for

$$f(x) = x^a,$$

$$Df(x_0) = ax_0^{a-1}.$$

By additivity and linear homogeneity, we can extend this to get the derivatives of all polynomial functions.

Rules for computing derivatives

7. The rule for derivatives of quotients follows from the product rule. For $g(x) \neq 0$,
$$h(x) = \frac{f(x)}{g(x)}$$
can be written as:
$$h(x)g(x) = f(x).$$
Therefore
$$h'(x_0)\, g(x_0) = f'(x_0) - h(x_0)\, g'(x_0)$$
and therefore
$$h'(x_0) = \frac{f'(x_0)}{g(x_0)} - \frac{f(x_0)\, g'(x_0)}{(g(x_0))^2}.$$

Rules for computing derivatives

8. The inverse function rule is a consequence of the chain rule: For all x, we have:

$$f^{-1}(f(x)) = x$$

Taking derivatives on both sides and denoting $y_0 = f(x_0)$:

$$Df^{-1}(y_0)\, Df(x_0) = 1$$

and therefore:

$$Df^{-1}(y_0) = \frac{1}{Df(x_0)}.$$

Rules for computing derivatives

9. A case that is not covered by the previous ones is the exponential function:

$$f(x) = e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} \dots$$

We have not proved this, but for convergent power series as the one above, we may differentiate element by element to get:

$$Df(x_0) = 0 + 1 + x_0 + \frac{x_0^2}{2} = \sum_{n=1}^{\infty} \frac{x_0^{n-1}}{(n-1)!} = e^{x_0}.$$

Rules for computing derivatives

10. The logarithmic function denoted by $\ln(y)$ is the inverse function of the exponential function (defined for strictly positive y):

$$g(y) = \ln y,$$

$$f(x) = e^x,$$

$$g(f(x)) = x.$$

By chain rule:

$$Dg(y_0)\, Df'(x_0) = 1, \text{ for all } x_0 \text{ and } y_0 = e^{x_0}.$$

Therefore

$$Dg(y_0) = \frac{1}{Df(x_0)} = \frac{1}{f(x_0)} = \frac{1}{y_0}.$$

So we have:

$$D \ln y_0 = \frac{1}{y_0}.$$

Rules for computing derivatives

11. Trigonometric functions etc. can be differentiated using their representations as complex power series or via direct limit arguments using basic identities from trigonometry.

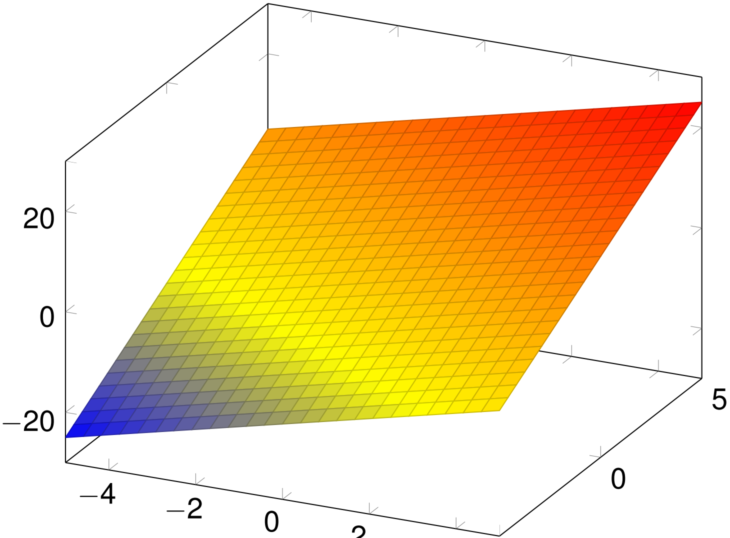
Pictures of multivariate functions

- How to picture multivariate functions? For functions on $\mathbb{R}^2$ need 3-d pictures
- Linear functions again easy since they look everywhere the same
- For non-linear functions, not so easy
- We end with some pictures and we take up the analytical tools on Monday

Graphs for functions of two variables

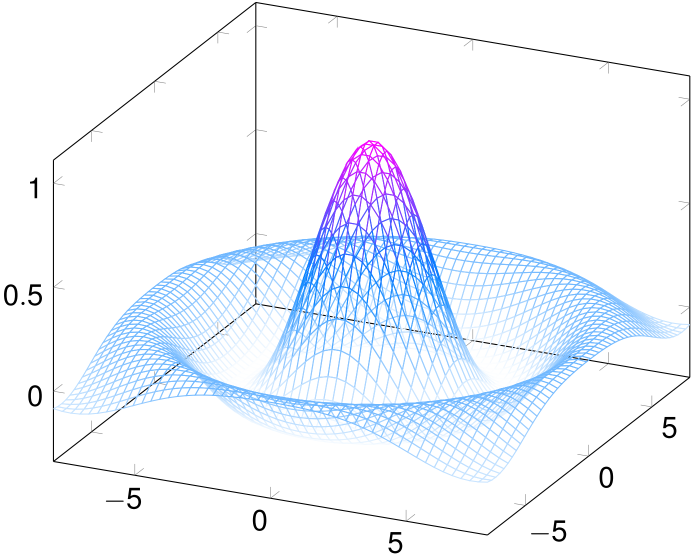
A linear function:

$$y = 2x_1 - 3x_2.$$

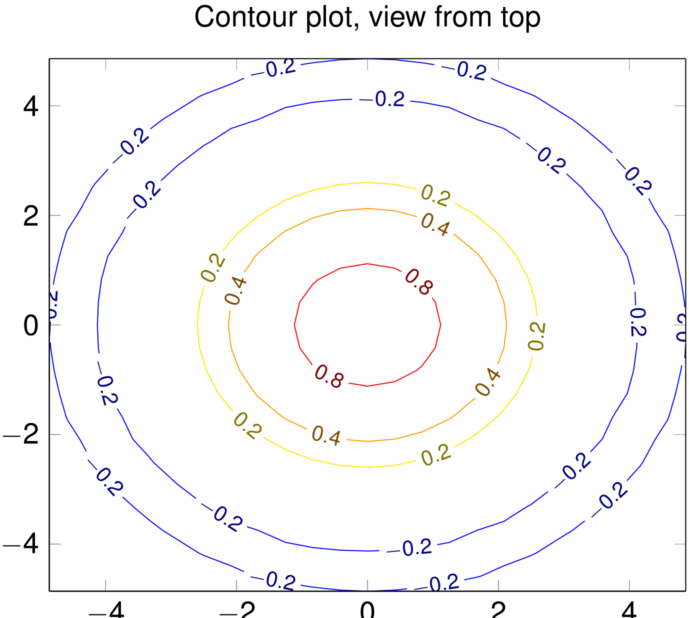


Graphs for functions of two variables

And a non-linear one: $y = \frac{sin(\sqrt{x_1^2+x_2^2})}{\sqrt{x_1^2+x_2^2}}$



Alternatively level sets for the function



Key points of the lecture

1. Representing linear systems of equations and linear functions by matrices
   1.1 Number of solutions to linear equations
   1.2 Existence of inverse matrices and inverse functions (rank condition, linear independence, determinant)
2. Non-linear functions of a single variable
   2.1 Derivative as a linear approximation
   2.2 Rules for derivatives
3. Multivariate functions
   3.1 Pictures