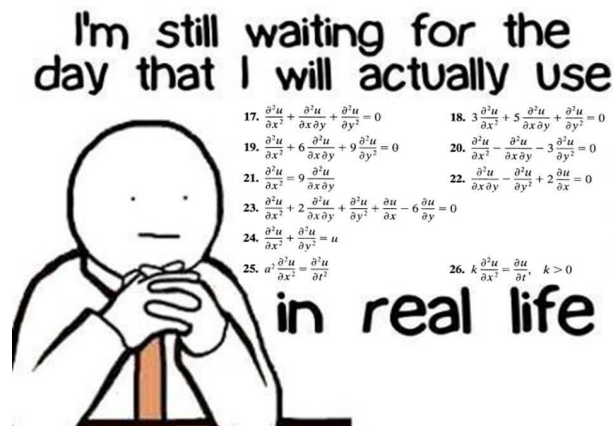


## Part I

# Quick outline of the course

## 0.1 Why mathematics?

Why mathematics?



Why should we learn mathematics?



If someone can take up this position (painlessly), what do you say to yourself? [0.2cm] Good! I'd like to be agile as she is ... [0.2cm] OR [0.2cm] Hm, I didn't need such a daredevil position in my life, I am going to train sitting on a chair instead, that's what I do ...

Understanding

**MATHEMATICS**  
is not about  
numbers, equations,  
computations, or  
algorithms:  
it is about  
**UNDERSTANDING.**

*William Paul Thurston*

- 
1. Petroleum Engineering (\$155,000 – after some time)
  2. Physics (\$101,800)
  3. Applied Mathematics (\$98,600 “Jobs in this field can be found in nearly every sector.”)
  4. Computer Science (\$97,900)
  5. Biomedical Engineering (\$97,800)
  6. Statistics (\$93,800)
  7. Civil Engineering (\$90,200)
  8. Mathematics (\$89,900)
  9. Environmental Engineering (\$88,600)
  10. Software Engineering (\$87,800)
  11. Finance (\$87,300)

15 Majors that  
Will Make You  
Rich (measured  
by money)

12. Construction Management (\$85,200)
13. Biochemistry (\$84,700)
14. Geology (\$83,300)
15. Management Information Systems (\$82,200)

source: <http://likes.com/misc/15-majors-that-will-make-you-rich>

---

Famous names  
...

- George Stibitz (Ph.D. in mathematical physics)

He was a Bell Labs researcher known for his work in the 1930s and 1940s on the realization of Boolean logic digital circuits using electromechanical relays as the switching element.

- breaking of German codes during WWII: Tommy Flowers (electroengineer), Alan Turing (mathematician), Max Newman (mathematician)
- Claude Shannon (founder of information theory, mathematician)

Shannon is famous for having founded information theory with one landmark paper published in 1948. But he is also credited with founding both digital computer and digital circuit design theory in 1937, when, as a 21-year-old master's student at MIT, he wrote a thesis demonstrating that electrical application of Boolean algebra could construct and resolve any logical, numerical relationship.

- John von Neumann (Von Neumann architecture)

---

Dennis Ritchie,  
Linus Torvalds

Ritchie graduated from Harvard University with degrees in physics and applied mathematics.

source: [wikipedia.org](http://wikipedia.org)

Linus Torvalds was born to Nils and Anna Torvalds, who were both journalists. However, he was highly influenced by his maternal grandfather to pursue his career in computers. Since childhood, Linus **was brilliant in mathematics**. Life of Linus Torvalds in computers began at the University of Helsinki in 1988 where he studied **computer science**. Linus is from a minority group in Finland and thus his first language is not Finnish but Swedish. For this reason, his pronunciation of Linux in Swedish were not understood or often taken as an error.

source: <http://www.mapsofworld.com/cities/finland/helsinki/linus-torvalds.html>

---

Bill Gates

In his sophomore year, Gates **devised an algorithm for pancake sorting** as a solution to one of a series of unsolved problems presented in a combinatorics class by Harry Lewis, one of his professors. Gates' solution held the record as the fastest version for over thirty years; its successor is faster by only one percent. His solution was later formalized in a published paper in collaboration with Harvard computer scientist Christos Papadimitriou.

source: [wikipedia.org](http://wikipedia.org)

---

Larry Page,  
Sergey Brin

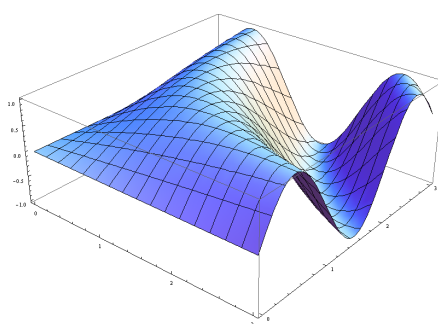
The company was founded by Larry Page and Sergey Brin while they were both attending Stanford University. . . . After enrolling for a Ph.D. program in computer science at Stanford University, Larry Page was in search of a dissertation theme and considered exploring the **mathematical properties of the World Wide Web**, understanding its link structure as a huge graph. . . . Sergey Brin earned his undergraduate degree at the University of Maryland, following in his father's and grandfather's footsteps by studying mathematics, as well as computer science. After graduation, he moved to Stanford University to acquire a Ph.D in computer science.

source: [wikipedia.org](http://wikipedia.org)

## 0.2 Short overview of included topics

Multivariate  
functions and  
optimization

- Many problems can be formulated as optimization problems: we maximize/minimize some functions that determines gain/cost/time/distance  
...
- If the function is given analytically, we know how to find the optimum.



[2mm]  $\sin(x \cdot y)$

General algebra

Notions from general algebra are one of the basic mathematical tools.

·	1	2	3	4	5	6	7	8	9	10	11	12
1	1	2	3	4	5	6	7	8	9	10	11	12
2	2	4	6	8	10	12	1	3	5	7	9	11
3	3	6	9	12	2	5	8	11	1	4	7	10
4	4	8	12	3	7	11	2	6	10	1	5	9
5	5	10	2	7	12	4	9	1	6	11	3	8
6	6	12	5	11	4	10	3	9	2	8	1	7
7	7	1	8	2	9	3	10	4	11	5	12	6
8	8	3	11	6	1	9	4	12	7	2	10	5
9	9	5	1	10	6	2	11	7	3	12	8	4
10	10	7	4	1	11	8	5	2	12	9	6	3
11	11	9	7	5	3	1	12	10	8	6	4	2
12	12	11	10	9	8	7	6	5	4	3	2	1

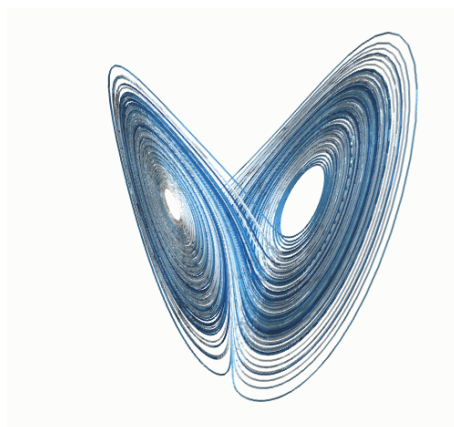
[1mm] Cayley

table of the group  $\mathbb{Z}_{13}^{\times}$

Besides a general introduction, we will focus on finite groups and fields, which form the basis for cryptography, hash functions, etc.

Numerical  
mathematics

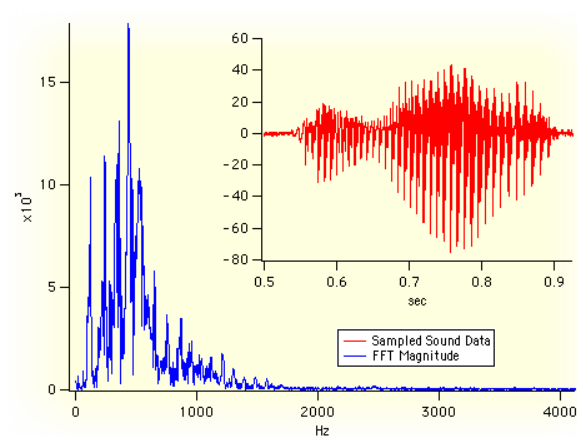
- continuous mathematics using the computer, stability of numerical algorithms ...



---

Discrete Fourier transform

- basic tool for frequency analysis



source: [www.wavemetrics.com](http://www.wavemetrics.com)

## Part II

# Multivariate optimisation



## 0.3 Examples

---

**Problem 1.** *Imagine the following situation: You have created a program that processes a text input by a user. You know from theoretical analysis of the source code and algorithms used within the program that it is impossible to determine the exact time needed to process a text of length  $k$ . However, you know that it is **approximately** proportional to the length of the text.*

Duration of a  
text processing  
program (1 of 6)

**Mathematically:** Denote  $t(k)$  the “average” number of seconds needed to process a text of length  $k$ . We know that

$$t(k) \approx \alpha k \quad \text{for some } \alpha \in \mathbb{R}.$$

**Problem:** The **proportionality constant**  $\alpha$  is unknown. How would you reasonably estimate its value?

---

### Sketch of a solution:

Duration of a  
text processing  
program (2 of 6)

1. Run the program for several, say  $n$ , texts of various lengths and measure the actual running times. This gives us  $n$  couples of measurements  $(k_1, t_1), (k_2, t_2), \dots, (k_n, t_n)$ .
2. For a given  $\alpha$ , we can measure the approximation error  $t(k) \approx \alpha k$  by this function:

$$e(\alpha) = (t_1 - \alpha k_1)^2 + (t_2 - \alpha k_2)^2 + \dots + (t_n - \alpha k_n)^2 = \sum_{i=1}^n (t_i - \alpha k_i)^2.$$

3. In order to find the best approximating proportionality constant  $\alpha$ , we find the value of  $\alpha$  for which the error  $e(\alpha)$  is minimal:

an optimal value of  $\alpha$  is a minimum point of the function  $e(\alpha)$ .

---

### How to find a minimum point of $e(\alpha)$ :

Duration of a  
text processing  
program (3 of 6)

1. Find the **first derivative**  $e'(\alpha)$ :

$$e'(\alpha) = \left( \sum_{i=1}^n (t_i - \alpha k_i)^2 \right)' = \sum_{i=1}^n -2k_i(t_i - \alpha k_i).$$

2. Find the **critical points**, i.e., the points  $\alpha_0$  where  $e'(\alpha_0)$  is zero or does not exist:

$$e'(\alpha_0) = 0 \Leftrightarrow \sum_{i=1}^n -2k_i(t_i - \alpha_0 k_i) = 0 \Leftrightarrow \sum_{i=1}^n k_i t_i = \alpha_0 \sum_{i=1}^n k_i^2 \Leftrightarrow \alpha_0 = \frac{\sum_{i=1}^n k_i t_i}{\sum_{i=1}^n k_i^2}$$

3. The critical points are our candidates for the points of (local) minimal or maximal values of the function  $e$ . To be sure that the value of  $\alpha$  we found is a minimum we need the **second derivative**:

$$e''(\alpha) = \left( \sum_{i=1}^n -2k_i(t_i - \alpha k_i) \right)' = \sum_{i=1}^n 2k_i^2.$$

... continues ...

---

Duration of a text processing program (4 of 6)

We know that if  $e''(\alpha_0) > 0$  (resp.  $e''(\alpha_0) < 0$ ), then the critical point  $\alpha_0$  is a local strict minimum (resp. strict maximum) point. If  $e''(\alpha_0) = 0$ , then  $\alpha$  is neither of these two cases (it may just be an inflexion point).

**Solution:** based on our measurements  $(k_1, t_1), (k_2, t_2), \dots, (k_n, t_n)$ , we get the best approximation  $t(k) \approx \alpha k$  for

$$\alpha = \alpha_0 = \frac{\sum_{i=1}^n k_i t_i}{\sum_{i=1}^n k_i^2}.$$

Indeed, this  $\alpha_0$  is the unique (why unique?) global (why global?) minimum point of the approximation error function  $e(\alpha)$  since the second derivative

$$e''(\alpha_0) = \sum_{i=1}^n 2k_i^2 \text{ is positive.}$$

---

Duration of a text processing program (5 of 6)

**Problem 2** (slight modification). *Imagine the following situation: You have created a program that processes a text input by a user. You know from theoretical analysis of the source code and algorithms used within the program that it is impossible to determine precisely the time needed to process a text of length  $k$ . However, you know that it is **approximately** proportional to the length of the text **and to the frequency of the processor**.*

**Mathematically:** Denote by  $t(k, f)$  the “average” number of seconds needed to process a text of length  $k$ , and the frequency of the processor by  $f$ . We know that

$$t(k, f) \approx \alpha k + \beta f \quad \text{for some } \alpha, \beta \in \mathbb{R}.$$

**Problem:** The constants  $\alpha$  and  $\beta$  are unknown. How would you reasonably estimate their values?

---

Duration of a text processing program (6 of 6)

**Sketch of solution:**

1. Run the program for several, say  $n$ , texts of various lengths on computers with different frequencies and measure the actual running times. This gives us  $n$  triplets of measurements  $(k_1, t_1, f_1), (k_2, t_2, f_2), \dots, (k_n, t_n, f_n)$ .
2. For a given  $\alpha$  and  $\beta$ , we can measure the approximation error  $t(k, f) \approx \alpha k + \beta f$  by this **two-variable** function:

$$\begin{aligned} e(\alpha, \beta) &= (t_1 - \alpha k_1 - \beta f_1)^2 + (t_2 - \alpha k_2 - \beta f_2)^2 + \dots + (t_n - \alpha k_n - \beta f_n)^2 = \\ &= \sum_{i=1}^n (t_i - \alpha k_i - \beta f_i)^2. \end{aligned}$$

3. In order to find the best approximating constants  $\alpha$  and  $\beta$ , we find values of  $\alpha$  and  $\beta$  for which the error  $e(\alpha, \beta)$  is minimal: an optimal value of  $\alpha$  and  $\beta$  is the “two-dimensional” minimum point of  $e(\alpha, \beta)$ .

---

Comments

**Why “optimization”?**

A typical situation in physics, engineering, economy, chemistry... is that you have a function that measures your profit, your loss, the energy of something... The value of such function is given by one or more inputs and the

relation between inputs and the resulting value is usually stated as a mathematical formula since all these sciences uses mathematical models to understand and quantify their subject of interest. An example of such function is our function  $e(\alpha, \beta)$  that measures the approximation error.

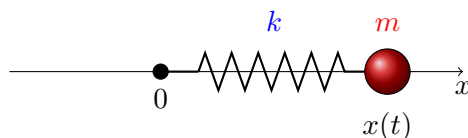
Typically, we want to maximize or minimize such functions (maximize the profit, the energy, minimize the loss, the error) which leads to the problem of finding **optimal** values of the inputs. Therefore the name “optimization”.

---

Comments

There is **another very important usage** of the derivative.

Derivatives measure the rate of change of a function. This helps us to describe the behaviour of a **dynamical systems** like a ball on a spring:



Position of the ball at time  $t$  is a function  $x(t)$  satisfying the differential equation

$$x''(t) + \omega^2 x(t) = 0.$$

Solution of this equation is

$$x(t) = x_0 \cos(\omega t) + \frac{v_0}{\omega} \sin(\omega t), \quad t \in \mathbb{R},$$

where  $x_0 = x(0)$  and  $v_0$  are the position and the speed of the ball at time  $t = 0$ . This model is known as **harmonic oscillator**.

## 0.4 Univariate optimization

### Derivative

---

How do we differentiate?

**Example 3.** Find the first derivative of  $f(x)$ , where

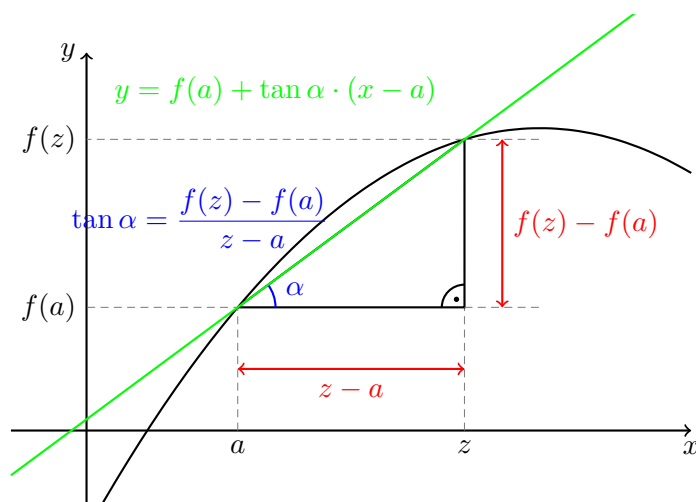
(a)  $f(x) = x^3 + 4x^2 + 6,$

(b)  $f(x) = \sin(x^3),$

(c)  $f(x) = e^x \sin x.$

Solutions:

- (a)  $f'(x) = 3x^2 + 8x$ ,  
 (b)  $f'(x) = 3x^2 \cos(x^3)$ ,  
 (c)  $f'(x) = e^x \sin x + e^x \cos x$ .



Geom. meaning  
of the deriva-  
tive: tangent  
line (1 of 2)

- The tangent line to the graph of a function  $f(x)$  at a point  $x_0$  is a straight line that “just touches” the curve at that point.
- Any straight line has equation  $y = ax + b$ , where  $a$  is the **slope** of the line.
- The slope of the tangent line to  $f(x)$  at the point  $x_0$  equals the first derivative evaluated at  $x_0$ :  $f'(x_0)$ .
- The tangent line at the point  $x_0$  satisfies the equation

$$y = f'(x_0)(x - x_0) + f(x_0).$$

Geom. meaning  
of the deriva-  
tive: tangent  
line (2 of 2)

With this geometrical explanation it is easy to see that the following statements are true:

- If  $f'(x_0)$  is positive, then  $f(x)$  is increasing at  $x_0$ .
- If  $f'(x_0)$  is negative, then  $f(x)$  is decreasing at  $x_0$ .
- If  $x_0$  is a local minimum/maximum point of  $f(x)$ , then  $f'(x_0) = 0$  or  $f'(x_0)$  does not exist. Such points are called **critical points**.

**Example 4.** Find all critical points of

$$f(x) = \frac{x^3}{3} + 2x^2 + 3x + 6.$$

What does it mean that the second derivative  $f''(x)$  is positive?

- The second derivative is a derivative of the first derivative; therefore the fact that  $f''(x)$  is positive implies that  $f'(x)$  is increasing (at the point  $x$ ).
- If  $f'(x)$  is increasing, then the function  $f(x)$  is more and more increasing (if  $f'(x) > 0$ ) or less and less decreasing (if  $f'(x) < 0$ ).
- An illustrative example of function with positive second derivative is  $f(x) = x^2$ .

Again, if we understand the geometrical meaning of the second derivative, we can easily see that the following statements are true:

**Theorem 5.** Let  $x_0$  be a critical point of a function  $f(x)$  such that  $f'(x_0) = 0$  and  $f''(x_0)$  exists.

- If  $f''(x_0) > 0$ , then the function is **convex** at  $x_0$  and  $x_0$  is a point of a (strict) minimum.
- If  $f''(x_0) < 0$ , then the function is **concave** at  $x_0$  and  $x_0$  is a point of a (strict) maximum.

Question: what can happen if  $f''(x_0) = 0$ ?

Universal cook-  
book of univari-  
ate optimization

Given a function  $f(x)$ , we want to find its extremal values.

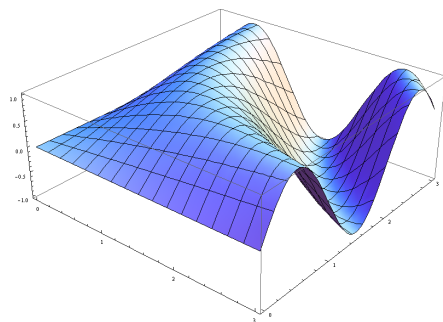
1. Find the first derivative  $f'(x)$ .
2. Find the critical points: solve the equation  $f'(x) = 0$  and find the points where the derivative **does not exist**.
3. Find the second derivative  $f''(x)$ .
4. If possible, for all critical points  $x_0$  evaluate  $f''(x_0)$  and decide whether this point is a point of minimum or maximum or whether it is an inflexion point. (Other critical points have to be treated by hand.)

The goal of this and the next lecture is to understand what happens when we have more than 1 variable. We shall build a similar cookbook for such functions.

## 0.5 Multivariate optimization

Graph of  
multivariate  
functions (1 of  
2)

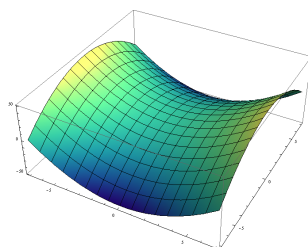
For a univariate function  $f(x)$ , its graph is the set of points  $(x, f(x))$  which can be depicted in Cartesian coordinate system (typically with  $x$ - and  $y$ -axis). [2mm] What if the function depends on more variables? For instance:  $f(x, y)$ .



Graph of a two-variable function  
 $\sin(x \cdot y)$ : the set of points  $(x, y, \sin(x \cdot y))$ .

- To depict a graph of a two-variable function we need a third axis (typically  $z$ -axis) and a 3-dimensional figure. Such graph is in general some surface.
- It is impossible to (directly) depict graphs of functions of more than 2 variables since we cannot make 4 or more dimensional figures.

**Example 6.** How does the graph of  $f(x, y) = x^2 - y^2$  look?



Given the function  $f(x, y) = x^2 + xy + y^2$ .

- If we fix the value of the variable  $y$  to 3, we obtain a univariate function  $f(x) = x^2 + x3 + 9$  having its derivative equal to  $2x + 3$ .
- We can fix the value of  $y$  not only to a specific number: **we just treat  $y$  as a constant**. Then we get univariate function  $f(x) = x^2 + xy + y^2$  and its derivative is  $2x + y$ .
- This derivative is called **partial derivative of  $(x, y)$  with respect to  $x$**  and denoted by

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

- In the same way we define the partial derivative of  $f(x, y)$  with respect to  $y$ :

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

- In general  $\frac{\partial f}{\partial x}(x, y)$  and  $\frac{\partial f}{\partial y}(x, y)$  are two-variate functions.



The derivative of a (single variate) function  $f(x)$  is the following limit (if it exists):

$$f'(x) = \lim_{\delta \rightarrow 0} \frac{f(x + \delta) - f(x)}{\delta}.$$

Partial derivatives are defined similarly:

**Definition 7.** *The partial derivative of  $f(x_1, x_2, \dots, x_n)$  with respect to  $x_i$  at the point  $(x_1, x_2, \dots, x_n)$  is defined by (if the limit below exists)*

$$\begin{aligned} \frac{\partial f}{\partial x_i}(x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) &= \\ &= \lim_{\delta \rightarrow 0} \frac{f(x_1, x_2, \dots, x_{i-1}, x_i + \delta, x_{i+1}, \dots, x_n) - f(x_1, x_2, \dots, x_n)}{\delta}. \end{aligned}$$

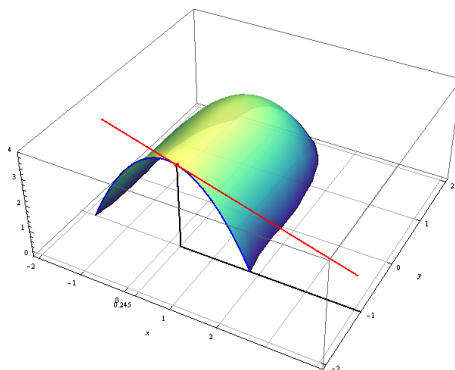
Since the definition is similar, even the geometrical meaning is analogous.

The partial derivatives of  $f(x, y)$  can be in short denoted by

$$f_x(x, y) = \frac{\partial f}{\partial x}(x, y) \quad \text{and} \quad f_y(x, y) = \frac{\partial f}{\partial y}(x, y).$$

The number  $f_x(x, y)$  for given values of  $x$  and  $y$  is again the slope of a tangent line, but a surface has infinitely many tangent lines in all possible directions at any point, so which one is this one?

It is the only tangent line which is parallel to the  $x$ -axis.



**Definition 8.** For a function  $f(x_1, x_2, \dots, x_n)$  we define second partial derivatives

$$f_{x_j x_i}(x_1, x_2, \dots, x_n) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x_1, x_2, \dots, x_n) = \frac{\partial}{\partial x_j} \left( \frac{\partial f}{\partial x_i}(x_1, x_2, \dots, x_n) \right),$$

in particular, for  $i = j$  we have

$$f_{x_i x_i}(x_1, x_2, \dots, x_n) = \frac{\partial^2 f}{\partial x_i^2}(x_1, x_2, \dots, x_n) = \frac{\partial}{\partial x_i} \left( \frac{\partial f}{\partial x_i}(x_1, x_2, \dots, x_n) \right).$$

---

Partial derivatives – exercises

**Example 9.** Find partial derivatives with respect to all variables

(a)  $f(x, y) = xy + e^x \cos y,$

(b)  $f(x, y) = x^2 y^3 + x^3 y^4 - e^{xy^2},$

(c)  $f(x, y, z) = \sin(xy/z).$

**Example 10.** Find all second partial derivatives of the functions

(a)  $f(x, y) = x^2 + xy^2 + 3x^3 y,$

(b)  $f(x, y, z) = e^{xz} + y \cos x,$

(c)  $f(x, y, z) = z \cos(xy) + x \sin(yz).$

---

Equality of mixed partial derivatives

The fact that the mixed partial derivatives are equal is not a coincidence:

**Theorem 11.** If a function  $f(x, y)$  has continuous second partial derivatives, then the mixed second derivatives are equal, i.e.,

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}.$$

This theorem is not true in general, a counterexample is the function

$$f(x, y) = \begin{cases} 0 & \text{at point } (0, 0) \\ \frac{xy(x^2 - y^2)}{x^2 + y^2} & \text{otherwise.} \end{cases}$$

