Data Mining - Mathematics

Dr. Jean-Michel RICHER

université angers

FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

2018

jean-michel.richer@univ-angers.fr

Outline

1. Introduction
2. Matrices and vectors
3. Derivative
4. Gradient
5. Lagrangian
6. Exercises

1. Introduction

What we will cover

Mathematical background needed for the understanding of Machine Learning techniques:

- matrix and vector operations
- derivative
- gradient
- lagrangian

Difficulty of mathematics

- use of **symbols** that represent expressions
- some symbols have different meaning depending on the **context**
- sometimes strict syntax and sometimes not

Example of a prime number

n is a prime number if it has only two divisors 1 and itself (but with restriction that $x \neq 1$)

- implies that we deal with integers
- implies the notion of **divisibility**:

$$\forall\, n, p, q, r \in \mathbb{N}$$

$$n = p \times q + r$$

- n is divisible by q if (and only if) $r = 0$ and $p \geq 1$

2. Matrices and vectors

université angers
FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

Vector

- a series of values that can be identified by their index in an array of length p
- $x(p)$ or simply $x \in \mathbb{R}^p$
- for computer scientists a 1D array of length p
- $x(p) = (x_1, x_2, \ldots, x_p) = [x_1, x_2, \ldots, x_p]$, $x_i \in \mathbb{R}$
- can also be represented vertically (for mathematicians):

$$x(p) = x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \quad x^T = [x_1, x_2, \ldots, x_p]$$

in this case x^T (T for Transposition) will be the horizontal representation

FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

Operations on vectors

- vectors must have the same length
- $+, -, \times, /$
- the dot (or scalar) product of two vectors

$$xy = x \cdot y = x \odot y = x^T y = \sum_{i=1}^{p} x_i \times y_i$$

- norm (or length) of a vector $||x|| = \sqrt{x \cdot x}$

Example of operations on vectors

Examples of operations on vectors

$$
\begin{array}{lcll}
x & = & [1, -2, 3] & \\
y & = & [-1, 4, -7] & \\
x + y & = & [1 + (-1), -2 + 4, 3 + (-7)] & = [0, 2, -4] \\
x \times y & = & [1 \times -1, -2 \times 4, 3 \times -7] & = [-1, -8, -21] \\
xy & = & (1 \times -1) + (-2 \times 4) + (3 \times -7) & = -30
\end{array}
$$

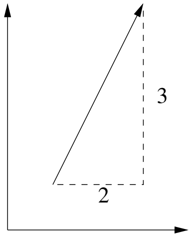
$$
x^T y = [1, -2, 3] \cdot \begin{bmatrix} -1 \\ 4 \\ -7 \end{bmatrix} = -30
$$

université angers — FACULTÉ DES SCIENCES Unité de formation et de recherche DÉPARTEMENT INFORMATIQUE

Examples of operations on vectors

Norm of a vector in 2D

$$\begin{aligned} x \cdot x &= (2 \times 2) + (3 \times 3) \\ &= 4 + 9 = 13 \\ ||x|| &= \sqrt{13} = 3.6055 \end{aligned}$$



Matrix

- $X(p) = (x_1, x_2, \ldots, x_p)$ where $x_i \in \mathbb{R}^n$, notation that can lead to confusion
- for computer scientists a 2D array defined by :
  - n rows
  - and p columns
- can be seen as an array of vectors or vector of vectors

$$X(n,p) = \begin{bmatrix} x_1^1 & x_1^2 & \ldots & x_1^p \\ x_2^1 & x_2^2 & \ldots & x_2^p \\ \vdots & \ddots & \ddots & \vdots \\ x_n^1 & x_n^2 & \ldots & x_n^p \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} x_1^1 \\ x_2^1 \\ \vdots \\ x_n^1 \end{bmatrix} \begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{bmatrix} \ldots \begin{bmatrix} x_1^p \\ x_2^p \\ \vdots \\ x_n^p \end{bmatrix} \end{bmatrix}$$

x_i^j is the element in row i and column j

Square matrices

- a square matrix is such that $n = p$
  - the diagonal is a separation line
  - called **lower triangular** if all the entries above the main diagonal are zero
  - called **upper triangular** if all the entries under the main diagonal are zero
- I or I_n is the identity matrix

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

Matrix sum

- matrices must have the same dimensions

$$X(n,p) = \begin{bmatrix} x^1_1 & x^2_1 & \dots & x^p_1 \\ x^2_1 & x^2_2 & \dots & x^p_2 \\ \vdots & \vdots & \ddots & \vdots \\ x^1_n & x^2_n & \dots & x^p_n \end{bmatrix} + Y(n,p) = \begin{bmatrix} y^1_1 & y^2_1 & \dots & y^p_1 \\ y^2_1 & y^2_2 & \dots & y^p_2 \\ \vdots & \vdots & \ddots & \vdots \\ y^1_n & y^2_n & \dots & y^p_n \end{bmatrix}$$

$$= Z(n,p) = \begin{bmatrix} x^1_1 + y^1_1 & x^2_1 + y^2_1 & \dots & x^p_1 + y^p_1 \\ x^2_1 + y^1_2 & x^2_2 + y^2_2 & \dots & x^p_2 + y^p_2 \\ \vdots & \vdots & \ddots & \vdots \\ x^1_n + y^1_n & x^2_n + y^2_n & \dots & x^p_n + y^p_n \end{bmatrix}$$

Operations on matrices

Matrix product

- $X(n,p) \times Y(p,q) = Z(n,q)$
- number of columns of X = number of rows of Y

$$z_i^j = \sum_{k=1}^{p} a_i^k \times b_k^j$$

```
for (int i = 0; i < n; ++i)
  for (int j = 0; j < q; ++j) {
    double sum = 0;
    for (int k = 0; k < p; ++k) {
      sum += a[i][k] * b[k][j];
    }
    c[i][j] = sum;
  }
```

université angers — FACULTÉ DES SCIENCES Unité de formation et de recherche DÉPARTEMENT INFORMATIQUE

Matrix product

Note that generally:

$$AB \neq BA$$

Example of matrix product

$$X(3,2) = \begin{bmatrix} 11 & 12 \\ 21 & 22 \\ 31 & 32 \end{bmatrix} \times Y(2,3) = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

Operations on matrices

Example of matrix product

$$X(3,2) = \begin{bmatrix} 11 & 12 \\ 21 & 22 \\ 31 & 32 \end{bmatrix} \times Y(2,3) = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$Z(3,3) = \begin{bmatrix} 11 \times 1 + 12 \times 4 & 11 \times 2 + 12 \times 5 & 11 \times 3 + 12 \times 6 \\ 21 \times 1 + 22 \times 4 & 11 \times 1 + 22 \times 4 & 21 \times 1 + 22 \times 4 \\ 31 \times 3 + 32 \times 6 & 31 \times 3 + 32 \times 6 & 31 \times 3 + 32 \times 6 \end{bmatrix}$$

Example of matrix and vector product

$$X(3,2) = \begin{bmatrix} 11 & 12 \\ 21 & 22 \\ 31 & 32 \end{bmatrix} \times y(2) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Example of matrix and vector product

$$X(3,2) = \begin{bmatrix} 11 & 12 \\ 21 & 22 \\ 31 & 32 \end{bmatrix} \times y(2) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$Z(3,3) = \begin{bmatrix} 11 \times 1 + 12 \times 2 \\ 21 \times 1 + 22 \times 2 \\ 31 \times 1 + 32 \times 2 \end{bmatrix} = \begin{bmatrix} 35 \\ 65 \\ 95 \end{bmatrix}$$

Matrix transpose

- to make it simple: exchange of values from both sides of the diagonal of the matrix

$$X(n,p) = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^p \\ x_1^2 & x_2^2 & \dots & x_2^p \\ \vdots & \ddots & \ddots & \vdots \\ x_n^1 & x_n^2 & \dots & x_n^p \end{bmatrix} \quad X^T(p,n) = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_n^1 \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \vdots & \ddots & \ddots & \vdots \\ x_1^p & x_2^p & \dots & x_n^p \end{bmatrix}$$

Operations on matrices

Matrix transpose

$$X(3,3) = \begin{bmatrix} 11 & 12 & 13 \\ 21 & 22 & 23 \\ 31 & 32 & 33 \end{bmatrix} \quad X^T(3,3) = \begin{bmatrix} 11 & 21 & 31 \\ 12 & 22 & 32 \\ 13 & 23 & 33 \end{bmatrix}$$

$$X(3,5) = \begin{bmatrix} 11 & 12 & 13 & 14 & 15 \\ 21 & 22 & 23 & 24 & 25 \\ 31 & 32 & 33 & 34 & 35 \end{bmatrix} \quad X^T(5,3) = \begin{bmatrix} 11 & 21 & 31 \\ 12 & 22 & 32 \\ 13 & 23 & 33 \\ 14 & 24 & 34 \\ 15 & 25 & 35 \end{bmatrix}$$

Inverse of a matrix

The inverse matrix A^{-1} of a matrix A is such that

$$A^{-1} \times A = I$$

and is used for example to solve linear equation systems:

$$A \times x = b \quad \text{then} \quad x = A^{-1} \times b$$

where x and b are vectors

université angers
FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

Inverse of a matrix

$$A(3,3) = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 2 & -3 \\ -7 & 6 & -5 \end{bmatrix}$$

$$A^{-1}(3,3) = \begin{bmatrix} 0.125 & 0.4375 & -0.1875 \\ 0.25 & 0.25 & 0 \\ 0.125 & -0.3125 & 0.0625 \end{bmatrix}$$

$$A \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = b = \begin{bmatrix} 14 \\ -6 \\ -10 \end{bmatrix} \quad A^{-1}b = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

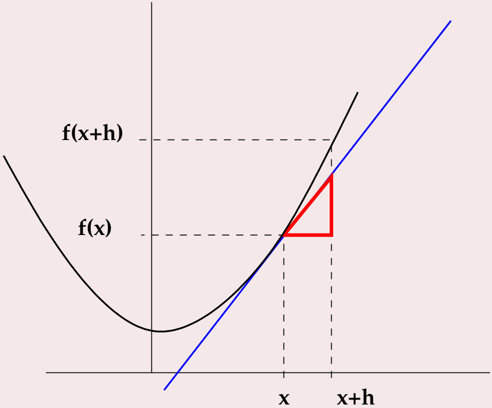
3. Derivative

Derivative of $f(x)$

- the slope of the tangent of a curve in a given point
- if positive: the curve will increase
- if negative: the curve will decrease
- if zero: won't increase or decrease

Derivative

Derivative of $f(x)$



Derivative

More formally the derivative can be defined as

$$lim_{h \to 0} \frac{f(x+h) - f(x)}{(x+h) - x} = lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

Notations:

$$f'(x) \qquad \text{or} \qquad \frac{df(x)}{dx}$$

$\frac{df(x)}{dx}$ means the variation of $f(x)$ if we increase x by a small value

Université Angers — Faculté des Sciences, Unité de formation et de recherche, Département Informatique

Properties of the derivative

Property of addition and product of the derivative

$$(f + g)' = f' + g'$$

$$(f \times g)' = f' \times g + f \times g'$$

As an exercise you could try to prove the result of $(f \times g)'$

Property of the composition

$$(g \circ f)' = (g' \circ f) \times f'$$

in other words:

$$g(f(x))' = g'(f(x)) \times f'(x) \qquad \text{(DerivComp)}$$

Property of the composition

Find the derivative of

$$h(x) = sin(3x^2 + 2)$$

- let $g(x) = sin(x)$, then $g'(x) = cos(x)$
- let $f(x) = 3x^2 + 2$, then $f'(x) = 6x$

So

$$h'(x) = g(f(x))' = g'(f(x)) \times f'(x)$$
$$h'(x) = cos(3x^2 + 2) \times 6x$$

Property of the inverse function

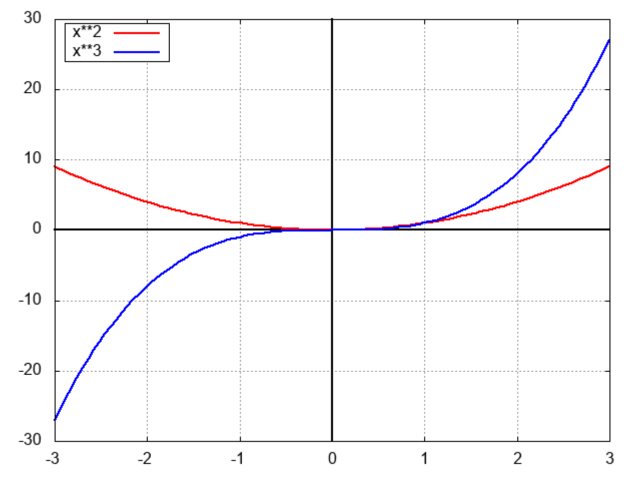
Let $f^{-1}(x)$ be the inverse function of $f(x)$, i.e. $f^{-1}(f(x)) = x$

$$
\begin{aligned}
[f^{-1}(f(x))]' &= (x)' = 1 \\
f^{-1'}(f(x)) \times f'(x) &= 1 \qquad \text{from (DerivComp)}
\end{aligned}
$$

$$f^{-1'}(f(x)) = \frac{1}{f'(x)} \qquad \text{(DerivInv)}$$

Derivative of x^n

$f(x) = x^n$



$f(x) = x^n$

The derivative should have the following behaviour

X	$-\infty$	0	∞
x^{2k}	$-$	0	$+$
x^{2k+1}	$+$	0	$+$

with n even $(2k)$ or odd $(2k+1)$

$(x+a)^n$

Remember that $x^0 = 1$

$$
\begin{aligned}
(x+a)^2 &= x^2 + 2ax + a^2 \\
(x+a)^3 &= x^3 + 3ax^2 + 3a^2x + a^3 \\
(x+a)^4 &= x^4 + 4ax^3 + 6a^2x^2 + 3a^3x + a^4 \\
\dots \\
(x+a)^n &= \alpha_{0,n}a^0x^n + \cdots + \alpha_{i,j}a^ix^j + \cdots + \alpha_{n,0}a^nx^0 \\
(x+a)^n &= \sum_{i=0,j=n-i}^{i=n} \alpha_{i,j}a^ix^j
\end{aligned}
$$

the coefficients $\alpha_{i,j}$ of a^ix^j are given by Pascal's triangle

A bit of history

Blaise Pascal (fr) (1623-1662)

- was a French mathematician, physicist, inventor, writer and catholic theologian
- wrote a significant treatise on the subject of projective geometry at the age of 16
- work on the principles of hydraulic fluids (hydraulic press and the syringe)
- theological work, referred to posthumously as the Pensées (*Thoughts*)

Pascal's triangle

n	x^n	ax^{n-1}	$\dots$		
0	1	0			
1	1	**1**	0		
2	1	**2**	1	0	
3	1	**3**	3	1	0
4	1	**4**	6	4	1

Obviously, the coefficient $\alpha_{1,n-1}$ of ax^{n-1} is n

Derivative of x^n

$$f(x+h) - f(x) = (x+h)^n - x^n$$

Derivative of x^n

$$\begin{aligned} f(x+h) - f(x) &= (x+h)^n - x^n \\ &= (x^n + nhx^{n-1} + \alpha_{2,n-2}h^2x^{n-2} + \ldots) - x^n \end{aligned}$$

Derivative of x^n

$$
\begin{aligned}
f(x+h) - f(x) &= (x+h)^n - x^n \\
&= (x^n + nhx^{n-1} + \alpha_{2,n-2}h^2x^{n-2} + \ldots) - x^n \\
&= nhx^{n-1} + \alpha_{2,n-2}h^2x^{n-2} + \ldots
\end{aligned}
$$

université angers
FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

Derivative of x^n

$$
\begin{aligned}
f(x+h) - f(x) &= (x+h)^n - x^n \\
&= (x^n + nhx^{n-1} + \alpha_{2,n-2}h^2x^{n-2} + \ldots) - x^n \\
&= nhx^{n-1} + \alpha_{2,n-2}h^2x^{n-2} + \ldots \\
\frac{f(x+h)-f(x)}{h} &= \frac{nhx^{n-1}+\alpha_{2,n-2}h^2x^{n-2}+\ldots}{h}
\end{aligned}
$$

Derivative of x^n

$$
\begin{aligned}
f(x+h) - f(x) &= (x+h)^n - x^n \\
&= (x^n + nhx^{n-1} + \alpha_{2,n-2}h^2x^{n-2} + \ldots) - x^n \\
&= nhx^{n-1} + \alpha_{2,n-2}h^2x^{n-2} + \ldots \\
\frac{f(x+h)-f(x)}{h} &= \frac{nhx^{n-1}+\alpha_{2,n-2}h^2x^{n-2}+\ldots}{h} \\
&= nx^{n-1} + \underbrace{\alpha_{2,n-2}hx^{n-2} + \ldots}_{0}
\end{aligned}
$$

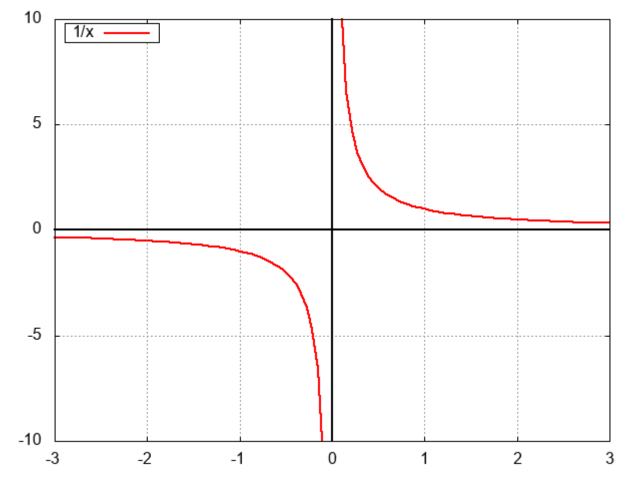
FACULTÉ DES SCIENCES Unité de formation et de recherche DÉPARTEMENT INFORMATIQUE

université angers

Derivative of x^n

$$
\begin{aligned}
f(x+h) - f(x) &= (x+h)^n - x^n \\
&= (x^n + nhx^{n-1} + \alpha_{2,n-2}h^2x^{n-2} + \ldots) - x^n \\
&= nhx^{n-1} + \alpha_{2,n-2}h^2x^{n-2} + \ldots \\
\frac{f(x+h)-f(x)}{h} &= \frac{nhx^{n-1}+\alpha_{2,n-2}h^2x^{n-2}+\ldots}{h} \\
&= nx^{n-1} + \underbrace{\alpha_{2,n-2}hx^{n-2} + \ldots}_{0} \\
lim_{h\to 0}\frac{f(x+h)-f(x)}{h} &= nx^{n-1}
\end{aligned}
$$

Function $1/x$



$f(x) = 1/x$

The derivative should have the following behaviour

X	$-\infty$	0	∞
x	$-$	NA	$-$

Derivative of $1/x$

$$
\begin{aligned}
f(x+h) - f(x) &= \frac{1}{x+h} - \frac{1}{x} \\
&= \frac{x-(x+h)}{x(x+h)} \\
&= \frac{-h}{x^2+hx} \\
\frac{f(x+h)-f(x)}{h} &= \frac{\frac{-h}{x^2+hx}}{h} \\
&= \frac{-h}{h(x^2+hx)} \\
lim_{h\to 0} \frac{f(x+h)-f(x)}{h} &= -\frac{1}{x^2}
\end{aligned}
$$

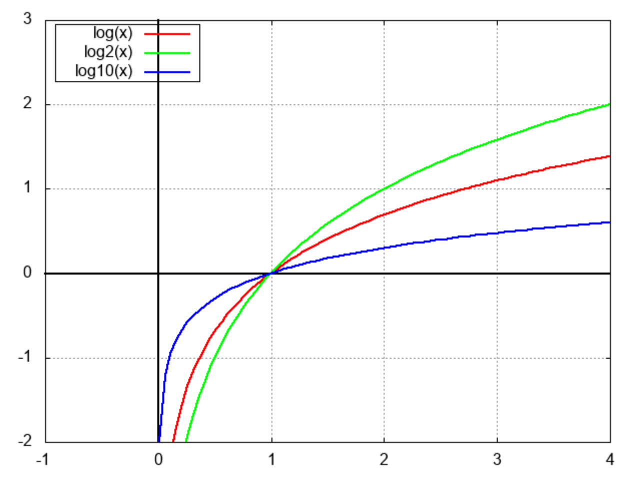
Function $log(x)$

Function $log(x)$

- $ln(x)$ or $log(x)$
- the natural logarithm of x is the power to which $e = 2.718281\ldots$ would have to be raised to equal x
- for example $ln(7.5) = 2.0149\ldots$, because $e^{2.0149\ldots} = 7.5$
- used to replace products by sums
- other functions:

$$log_n(x) = \frac{ln(x)}{ln(n)}$$

Function $log(x)$



Derivative of $log(x)$

$f(x) = log(x)$

The derivative should have the following behaviour

X	$-\infty$	0	∞
x	NA	$-\infty$	$+$

Properties of the function $log(x)$

$$
\begin{aligned}
log(1) &= 0 \\
log(e) &= 1 \\
log(x \times y) &= log(x) + log(y) \\
log(x/y) &= log(x) - log(y) \\
log(x^n) &= n \times log(x)
\end{aligned}
$$

Derivative of $log(x)$

Derivative of $log(x)$

By definition

$$log(a) = \int_1^a \frac{1}{x} dx$$

so the derivative of $log(x)$ is $\frac{1}{x}$

Derivative of $log(x)$

$$f(x+h) - f(x) = log(x+h) - log(x)$$

Derivative of $log(x)$

$$\begin{aligned} f(x+h) - f(x) &= log(x+h) - log(x) \\ &= log(\frac{x+h}{x}) \end{aligned}$$

Derivative of $log(x)$

$$
\begin{aligned}
f(x+h) - f(x) &= log(x+h) - log(x) \\
&= log(\frac{x+h}{x}) \\
&= log(1 + \frac{h}{x})
\end{aligned}
$$

Derivative of $log(x)$

$$
\begin{aligned}
f(x+h) - f(x) &= log(x+h) - log(x) \\
&= log(\tfrac{x+h}{x}) \\
&= log(1 + \tfrac{h}{x}) \\
\frac{f(x+h)-f(x)}{h} &= \frac{1}{h} \times log(1 + \frac{h}{x})
\end{aligned}
$$

Derivative of $log(x)$

$$
\begin{aligned}
f(x+h) - f(x) &= log(x+h) - log(x) \\
&= log(\frac{x+h}{x}) \\
&= log(1 + \frac{h}{x}) \\
\frac{f(x+h)-f(x)}{h} &= \frac{1}{h} \times log(1 + \frac{h}{x}) \\
&= log((1 + \frac{h}{x})^{\frac{1}{h}})
\end{aligned}
$$

Derivative of $log(x)$

$$
\begin{aligned}
f(x+h) - f(x) &= log(x+h) - log(x) \\
&= log(\frac{x+h}{x}) \\
&= log(1 + \frac{h}{x}) \\
\frac{f(x+h)-f(x)}{h} &= \frac{1}{h} \times log(1 + \frac{h}{x}) \\
&= log((1 + \frac{h}{x})^{\frac{1}{h}}) \\
lim_{h \to 0} \frac{f(x+h)-f(x)}{h} &= log((1 + \frac{h}{x})^{\frac{1}{h}})
\end{aligned}
$$

Derivative of $log(x)$

$$
\begin{aligned}
f(x+h) - f(x) &= log(x+h) - log(x) \\
&= log(\frac{x+h}{x}) \\
&= log(1 + \frac{h}{x}) \\
\frac{f(x+h)-f(x)}{h} &= \frac{1}{h} \times log(1 + \frac{h}{x}) \\
&= log((1 + \frac{h}{x})^{\frac{1}{h}}) \\
lim_{h \to 0} \frac{f(x+h)-f(x)}{h} &= log((1 + \frac{h}{x})^{\frac{1}{h}}) \\
&= log(lim_{h \to 0}(1 + \frac{h}{x})^{\frac{1}{h}}) = log(e^{\frac{1}{x}})
\end{aligned}
$$

FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

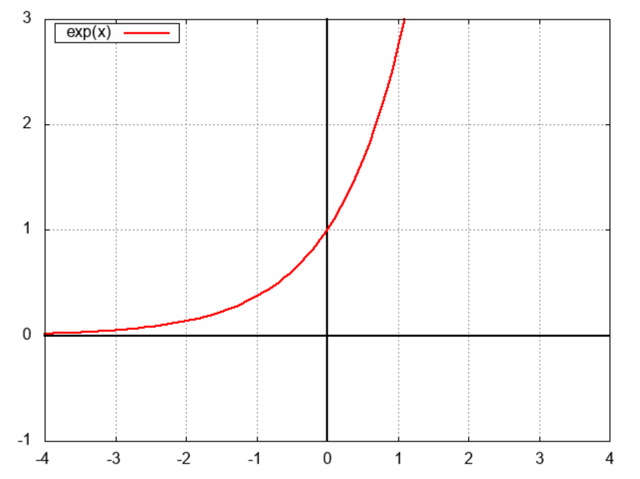
Function $exp(x)$ (Jakob Bernoulli (ch), 1654-1705)

- the exponential function aka the antilogarithm
- $exp(x) = e^x = lim_{n \to \infty}(1 + \frac{x}{n})^n$
- $e^1 = 2.718281\ldots$

Jakob Bernoulli (ch), 1654-1705)

- family, of Belgium origin, were refugees fleeing from persecution by the Spanish rulers of the Netherlands
- swiss mathematician and astronomer
- theory of permutations and combinations (Bernoulli numbers), by which he derived the exponential series
- Law of large numbers, in statistics, 1713

Function $exp(x)$



Derivative of $exp(x)$

$f(x) = exp(x)$

The derivative should have the following behaviour

X	$-\infty$	0	∞
x	$+$	1	$+$

université angers
FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

Properties of the function e^x

Properties of the function e^x

$$
\begin{aligned}
e^0 &= 1 \\
e^x &> 0 \quad \forall x \\
e^{-x} &= \frac{1}{e^x} \\
e^{x+y} &= e^x \times e^y \\
e^{x-y} &= \frac{e^x}{e^y} \\
e^{x \times y} &= (e^x)^y
\end{aligned}
$$

Derivative of e^x

By definition $log(e^x) = x$. So we compute the derivative of this last expression:

$$log(e^x) = x$$

Derivative of e^x

By definition $log(e^x) = x$. So we compute the derivative of this last expression:

$$\begin{aligned} log(e^x) &= x \\ (log(e^x))' &= (x)' \end{aligned}$$

Derivative of e^x

Derivative of e^x

By definition $log(e^x) = x$. So we compute the derivative of this last expression:

$$\begin{aligned} log(e^x) &= x \\ (log(e^x))' &= (x)' \\ \frac{1}{e^x} \times e'(x) &= 1 \qquad by(\text{DerivInv}) \end{aligned}$$

Derivative of e^x

By definition $log(e^x) = x$. So we compute the derivative of this last expression:

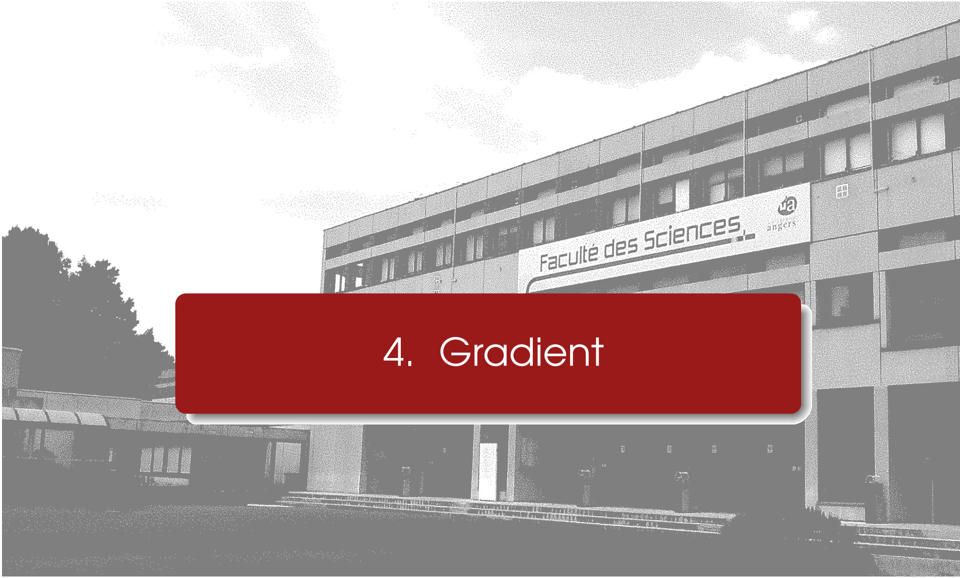
$$\begin{aligned} log(e^x) &= x \\ (log(e^x))' &= (x)' \\ \frac{1}{e^x} \times e'(x) &= 1 \quad by(\text{DerivInv}) \end{aligned}$$

$$e'(x) = e(x)$$

Derivatives

$$
\begin{aligned}
(x^n)' &= n\,x^{n-1} \\
\left(\frac{1}{x}\right)' &= -\frac{1}{x^2} \\
(log(x))' &= \frac{1}{x} \\
(e(x))' &= e^x
\end{aligned}
$$

4. Gradient



université angers
FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

Definition of the gradient

Given a function of several variables $f(x, y, z)$, the gradient is the **vector of the partial derivatives** of f

$$\nabla f(x, y, z) = \nabla f = [\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}]$$

The **partial derivative** $\frac{\partial f}{\partial x}$ is the derivative of $f(x, y, z)$ when y and z are considered as constants:

$$\frac{\partial f}{\partial x} = \frac{\partial f(x, y, z)}{\partial x} = \frac{df(x, y, z)_{|y,z}}{dx}$$

université angers
FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

Gradient

Property of the gradient

- the gradient ∇f gives the direction toward which you can increase the value of the function
- conversly $-\nabla f$ gives the direction toward which you can decrease the value of the function

Finding the minimum of a function

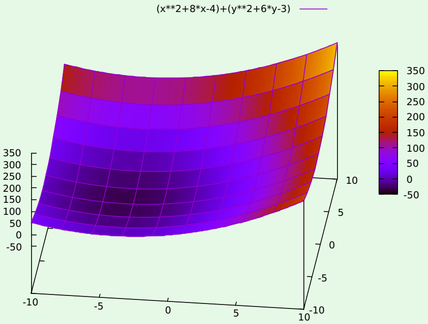
- the gradient can be used to find the minimum of a function by progressively decreasing the coordinates by substracting
- a fraction of the value of the gradient

université angers
FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

A convex quadratic function

Consider the following function:

$$f(x, y) = (x^2 + 8 \times x - 4) + (y^2 + 6 \times y - 3)$$



A convex quadratic function

The gradient of the function is:

$$\frac{\partial f}{\partial x} = 2 \times x + 8$$

$$\frac{\partial f}{\partial y} = 2 \times y + 6$$

The minimum is found for $\frac{\partial f}{\partial x} = 0$ and $\frac{\partial f}{\partial y} = 0$

$$\frac{\partial f}{\partial x} = 0 \Rightarrow x = -4$$

$$\frac{\partial f}{\partial y} = 0 \Rightarrow y = -3$$

Gradient descent algorithm

Data: $f(x)$
Result: $x^{\star}$: the minimum of the function
initialise vector x;
while **not** *terminate_condition* **do**
    compute gradient ∇f;
    $x = x - \alpha \times \nabla f$;
end

Algorithm 1: A very simple descent algorithm

- note that the *Terminate Condition* can be defined in different ways (improvement, number of iterations)
- $\alpha = 0.1$ for example, if too big the algorithm won't find the solution

Descent for a convex quadratic function

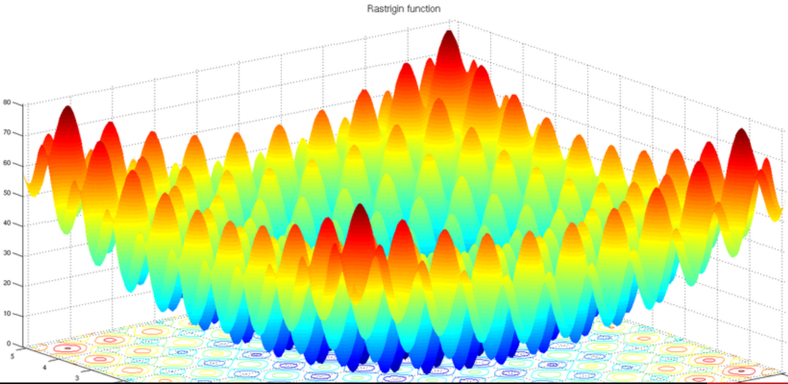
For the previous convex function we obtain this:

```
x0= 3, y0= 5, alpha=0.1
gradient=(14, 16)
x1= 1.5999, y1= 3.4
gradient=(11.2, 12.8)
x2= 0.48, y2 = 2.1199
gradient=(8.96, 10.2399)
...
x48 = -3.99984389478361, y48 = -2.999821594038412
gradient=(0.000312210432779735 9, 0.0003568119231758 26)
x49 = -3.999875115826888, y49 = -2.9998572752307298
```

Difficulty of finding the minimum

it becomes more difficult to find the minimum if

- the function is not convex
- the function has many minima (Rastrigin or Himmelblau functions)



Gradient

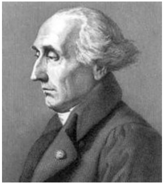
In Python

```
from scipy import optimize

def f(x):
        return x[0]**2+8*x[0]-4+x[1]**2+6*x[1]-3

def fprime(x):
        return np.array([(2*x[0]+8), (2*x[1]+6)])

z = optimize.fmin_bfgs(f, [3, 5], fprime=fprime)
print(z)
```

```
Optimization terminated successfully.
         Current function value: -32.000000
         Iterations: 2
         Function evaluations: 4
         Gradient evaluations: 4
[-4. -3.]
```

université angers
FACULTÉ DES SCIENCES Unité de formation et de recherche DÉPARTEMENT INFORMATIQUE

5. Lagrangian



Joseph-Louis Lagrange

- born **Guiseppe Lodovico Lagrangia** (it,fr) (1736 - 1813) was a franco-italian mathematician and astronomer
- made significant contributions to the fields of analysis, number theory, and both classical and celestial mechanics
- in **1787**, at age 51, moved from Berlin to Paris and became a member of the **French Academy of Sciences**
- remained in France until the end of his life

Lagrangian multipliers

Principle

- you want to minimize or maximize $f(x)$ subject to $g(x) = 0$
- under certain conditions
  - $f(x)$ is a quadratic function
  - $g(x)$ are linear constraints
- define the function

$$\mathcal{L}(x, \alpha) = f(x) + \alpha g(x)$$

where $\alpha \geq 0 \ \in \mathbb{R}$ is called the **lagrangian multiplier**

Resolution

- a solution of $\mathcal{L}(x, \alpha)$ is a point of gradient 0
- so compute and solve

$$\frac{\partial \mathcal{L}(x,\alpha)}{\partial x} = 0$$

$$\frac{\partial \mathcal{L}(x,\alpha)}{\partial \alpha} = g(x) = 0$$

- or reuse in $\mathcal{L}(x, \alpha)$

Statement of the example

Suppose you want to put a fence around some field which as a form of a rectangle (x, y) and you want to maximize the area knowing that you have P meters of fence:

$$\begin{cases} Max & x \times y \\ \text{such that} & P = 2x + 2y \end{cases}$$

then $f(x, y) = xy$ and $g(x) = P - 2x - 2y = 0$

$$\begin{cases} Max & xy \\ \text{such that} & P - 2x - 2y = 0 \end{cases}$$

Method of Lagrange - example

Lagrange formulation

$$\mathcal{L}(x, y, \alpha) = xy + \alpha(P - 2x - 2y)$$

the derivatives give us

$$\frac{\partial \mathcal{L}(x, y, \alpha)}{\partial x} = y - 2\alpha = 0$$

$$\frac{\partial \mathcal{L}(x, y, \alpha)}{\partial y} = x - 2\alpha = 0$$

$$\frac{\partial \mathcal{L}(x, y, \alpha)}{\partial \alpha} = P - 2x - 2y = 0$$

Resolution

- the first two constraints give us $y = 2\alpha = x$, so $x = y$
- in other words, the area is a square
- and the last one that $P = 4x = 4y$
- consequently $\alpha = P/8$ because $\alpha = x/2 = y/2$

6. Exercises

université angers
FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

Matrix - Matrix multiplication

Consider the following matrices:

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 4 & 5 & -3 \\ 6 & -3 & -2 \end{bmatrix} \quad B = \begin{bmatrix} 0.5 & 7 & 3 \\ 8 & -6 & 2 \\ 1 & -2 & 3 \end{bmatrix}$$

Perform the products by hand and compare:

$$A \times B \quad \stackrel{?}{=} \quad B \times A$$

Matrix - Vector multiplication

Consider the following matrix and vector:

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 4 & 5 & -3 \\ 6 & -3 & -2 \end{bmatrix} \quad x = \begin{bmatrix} 0.5 \\ 8 \\ 1 \end{bmatrix}$$

Perform the products by hand and compare:

$$A \times x \quad \stackrel{?}{=} \quad x^T \times A$$

Vector - Vector multiplication

Consider the following vectors:

$$x = \begin{bmatrix} 1 \\ 4 \\ 6 \end{bmatrix} \quad y = \begin{bmatrix} -3 \\ 8 \\ 1 \end{bmatrix}$$

Perform the following operations:

$$x \times y \quad \text{and} \quad x \odot y$$

Check results in python

Check results in python

Write a program in python to check the results of the different matrix and vector products

Derivative rules

Remember to use the following rules:

$$(f+g)' = f' + g' \qquad (dSum)$$

$$(f \times g)' = f' \times g + f \times g' \qquad (dProd)$$

$$f(g(x))' = f'(g(x)) \times g'(x) \qquad (dComp)$$

Derivatives

write a python program to draw the following functions (use matplotlib) and compute their derivatives by hand:

$$
\begin{aligned}
f_1(x) &= \frac{1+\frac{1}{x}}{x-3} \\
f_2(x) &= \frac{1}{x^2+e^x} \\
f_3(x) &= \frac{x \times e^x}{1+e^x}
\end{aligned}
$$

Derivatives

The results are the following

$$
\begin{aligned}
f_1'(x) &= -\frac{x^2 + 2x - 3}{(x-3)^2 x^2} \\
f_2'(x) &= -\frac{e^x + 2x}{(e^x + x^2)^2} \\
f_3'(x) &= \frac{e^x (e^x + x + 1)}{(e^x + 1)^2}
\end{aligned}
$$

Derivative of $f_1(x)$

$$f_1(x) = (1 + \frac{1}{x}) \times \frac{1}{x-3}$$

$$f_1(x) = F(x) \times G(x)$$

So we need to apply the formula (*dProd*)

Derivative of $f_1(x)$

$$F(x) = (1 + \frac{1}{x})$$

$$F'(x) = -\frac{1}{x^2}$$

$$G(x) = \frac{1}{x-3} = H(K(x))$$

with

$$H(z) = \frac{1}{z}$$

$$K(z) = z - 3$$

Derivative of $f_1(x)$

$$G'(x) = H'(K(x)) \times K'(x)$$

$$G'(x) = -\frac{1}{(x-3)^2} \times 1$$

Derivative of $f_1(x)$

Finally

$$f_1'(x) = -\frac{1}{x^2} \times \frac{1}{x-3} + (1 + \frac{1}{x}) \times -\frac{1}{(x-3)^2}$$

$$f_1'(x) = -\frac{1}{x^2} \times \frac{x-3}{x-3} + 1 + \frac{1}{x} \times -\frac{x}{(x-3)^2}$$

$$f_1'(x) = -\frac{(x-3)+(x+1)x}{x^2(x-3)^2}$$

$$f_1'(x) = -\frac{x^2+2x-3}{x^2(x-3)^2}$$

Derivative of $f_2(x)$

$$f_2(x) = \frac{1}{x^2 + e^x}$$

$$f_2(x) = F(G(x))$$

So we need to appy the formula (*dComp*) with

$$F(z) = \frac{1}{z}$$

$$G(z) = z^2 + e^z$$

Derivative of $f_2(x)$

$$
\begin{aligned}
F'(z) &= -\frac{1}{z^2} \\
G'(z) &= 2z + e^z \quad (dSum)
\end{aligned}
$$

Derivative of $f_2(x)$

Finally

$$f_2'(x) = -\frac{1}{x^2 + e^x} \times (2x + e^x)$$

$$f_2'(x) = -\frac{2x + e^x}{x^2 + e^x}$$

Derivatives

- check the results with the derivative calculator
- write a program in python using sympy to compute the derivatives of the functions

Gradient

- determine where $f_1(x)$ is minimum
- using the gradient method try to determine where $f_2(x)$ and $f_3(x)$ are maximum or minimum

Gradient/derivative of f_1

The derivative of f_1 is:

$$f_1'(x) = -\frac{x^2 + 2x - 3}{x^2(x-3)^2}$$

There are extremum (maximum or minimum) where $f_1'(x) = 0$
The function is not defined if the denominator is equal to $x^2(x-3)^2 = 0$:

$$\begin{cases} x^2 = 0 & \Rightarrow \quad x = 0 \\ (x-3)^2 = 0 & \Rightarrow \quad x = 3 \end{cases}$$

Gradient/derivative of f_1

The derivative is equal to 0 if:

$$x^2 + 2x - 3 = 0$$

$$\begin{cases} \Delta = b^2 - 4ac & = & 2^2 - 4 \times 1 \times -3 = 16 \\ x_1 = \dfrac{-b - \sqrt{\Delta}}{2a} & = & \dfrac{-2 - 4}{2 \times 1} = -3 \\ x_2 = \dfrac{-b + \sqrt{\Delta}}{2a} & = & \dfrac{-2 + 4}{2 \times 1} = +1 \end{cases}$$

then

$$x^2 + 2x - 3 = (x - 1)(x + 3)$$

Gradient/derivative of f_2

The derivative of f_2 is:

$$f_2'(x) = -\frac{2x + e^x}{x^2 + e^x}$$

The denominator $x^2 + e$ is always positive so we need to

solve:

$$2x + e^x = 0$$

which is not possible by analytical methods we need to find the root by using an approximation method

Gradient/derivative of f_2 (1/2)

```
import numpy as np

def gradient_ascent(x, df):
        delta = 0.0000001
        alpha = 0.1
        while True:
                x_n = x + alpha * df(x)
                if math.fabs(x-x_n) < delta:
                        break
                x = x_n
        return x
```

Gradient/derivative of f_2 (2/2)

```
def f2(x):
        return 1/(x*x + np.exp(x))

def df2(x):
        return  (-2*x - math.exp(x))/(x**2 + math.exp(
            x))**2

x2_star = gradient_ascent2( -0.5, df2)
print("f2:␣", x2_star, "␣=>␣", f2(np.asarray([x2_star
    ])))
```

Lagrangian

Consier the following problem

$$\begin{cases} Max\ x^2 + y^2 + z^2 \\ \text{such that} \\ x + 2y + z = 1 \\ 2x - y - 3z = 4 \end{cases}$$

- use the method of Lagrange to solve it to determine x, y and z

Lagrangian resolution

We define

$$\begin{aligned}\mathcal{L}(x, y, z, \alpha_1, \alpha_2) &= f(x, y, z) &+ \\ &\quad \alpha_1(x + 2y + z - 1) &+ \\ &\quad \alpha_2(2x - y - 3z - 4)\end{aligned}$$

with $f(x, y, z) = x^2 + y^2 + z^2$

Lagrangian resolution

We need to compute the partial derivatives of $\mathcal{L}$ for each variable :

$$
\begin{aligned}
\frac{\partial \mathcal{L}(x,y,z,\alpha_1,\alpha_2)}{\partial x} &= 2x + \alpha_1 + 2\alpha_2 = 0 \quad (1)\\
\frac{\partial \mathcal{L}(x,y,z,\alpha_1,\alpha_2)}{\partial y} &= 2y + 2\alpha_1 - \alpha_2 = 0 \quad (2)\\
\frac{\partial \mathcal{L}(x,y,z,\alpha_1,\alpha_2)}{\partial z} &= 2z + \alpha_1 - 3\alpha_2 = 0 \quad (3)\\
\frac{\partial \mathcal{L}(x,y,z,\alpha_1,\alpha_2)}{\partial \alpha_1} &= x + 2y + z - 1 = 0 \quad (4)\\
\frac{\partial \mathcal{L}(x,y,z,\alpha_1,\alpha_2)}{\partial \alpha_2} &= 2x - y - 3z - 4 = 0 \quad (5)
\end{aligned}
$$

Lagrangian resolution

Express α_1 from x and α_2 in (1)

$$
\begin{aligned}
-2x - 2\alpha_2 = \alpha_1 \quad & (1) \\
2y + 2\alpha_1 - \alpha_2 = 0 \quad & (2) \\
2z + \alpha_1 - 3\alpha_2 = 0 \quad & (3) \\
x + 2y + z - 1 = 0 \quad & (4) \\
2x - y - 3z - 4 = 0 \quad & (5)
\end{aligned}
$$

Lagrangian resolution

Then replace in (2) and (3)

$$-2x - 2\alpha_2 = \alpha_1 \quad (1)$$

$$2y + 2(-2x - 2\alpha_2) - \alpha_2 = 0 \quad \Rightarrow \quad 2y - 4x - 5\alpha_2 = 0 \quad (2)$$

$$2z + (-2x - 2\alpha_2) - 3\alpha_2 = 0 \quad \Rightarrow \quad 2z - 2x - 5\alpha_2 = 0 \quad (3)$$

$$x + 2y + z - 1 = 0 \quad (4)$$

$$2x - y - 3z - 4 = 0 \quad (5)$$

FACULTÉ DES SCIENCES Unité de formation et de recherche DÉPARTEMENT INFORMATIQUE

université angers

Lagrangian resolution

By substracting (2) and (3) we get $-2x + 2y - 2z = x - y + z = 0$. And finally we have a system of 3 equations with 3 variables:

$$
\begin{aligned}
x - y + z &= 0 && (6) = (2) - (3) \\
x + 2y + z - 1 &= 0 && (4) \\
2x - y - 3z - 4 &= 0 && (5)
\end{aligned}
$$

Lagrangian resolution

Then we compute (6) − (4) and obtain $y = \frac{1}{3}$.
The rest of the resolution is obvious and we should get

$$x = \frac{16}{15}$$

$$y = \frac{1}{3}$$

$$z = \frac{-11}{15}$$

$$\alpha_1 = -\frac{52}{75}$$

$$\alpha_2 = -\frac{54}{75}$$

Lagrangian resolution

In Python:

```
import numpy as np
A = np.asarray([[2,0,0,-1,-2], [0,2,0,-2,1],
    [0,0,2,-1,3], [1,2,1,0,0], [2,-1,-3,0,0]])
b = np.asarray([0,0,0,1,4])
x = np.linalg.solve(A, b)
print(x)
```

6. End

université angers
FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

FACULTÉ DES SCIENCES
Unité de formation et de recherche
DÉPARTEMENT INFORMATIQUE

UA - Angers
2 Boulevard Lavoisier 49045 Angers Cedex 01
Tel: (+33) (0)2-41-73-50-72