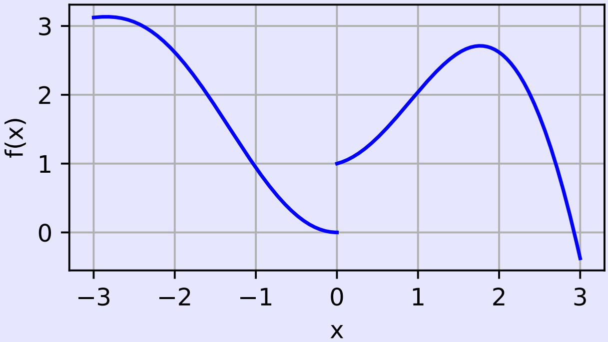
Introduction to functional analysis

Pierre Ablin
ENS, Département de mathématiques et applications

In one dimension

Functions

We consider a function $f : \mathbb{R} \to \mathbb{R}$.
This means that f takes as input a $x \in \mathbb{R}$ and outputs $f(x) \in \mathbb{R}$.
For instance, $f(x) = 3x + 2$: $f(3.2) = 3 \times 3.2 + 2 = 11.6$
A function can be visualized with its graph:



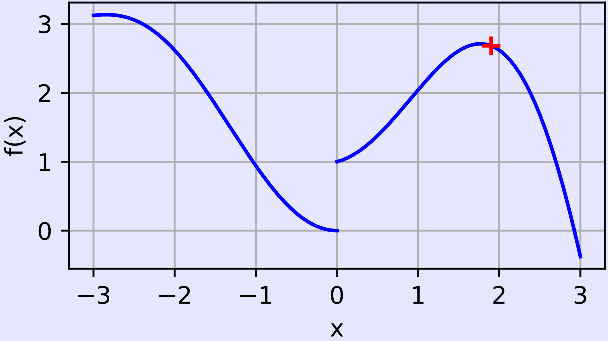
Regularity of functions: continuity

We say that f is continuous at $x \in \mathbb{R}$ if for any converging sequence $x_n \to x$, we have

$$f(x_n) \to f(x)$$

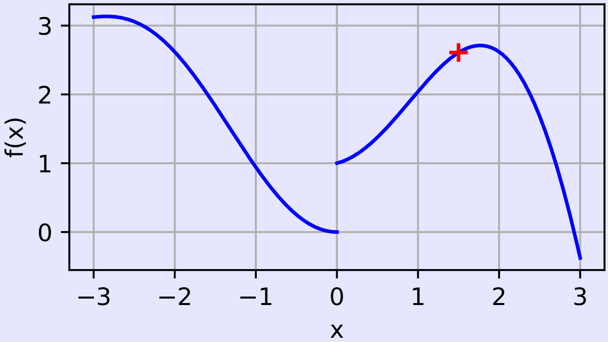
Regularity of functions: continuity

Previous example, $x = 1$



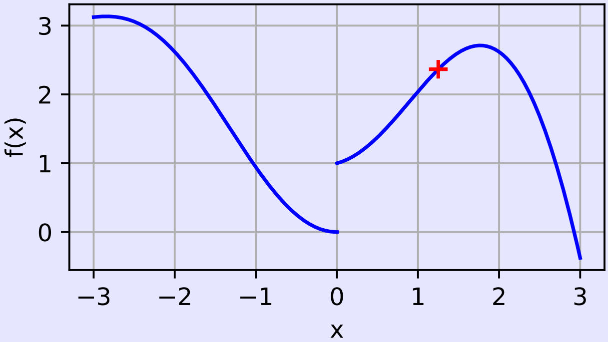
Regularity of functions: continuity

Previous example, $x = 1$



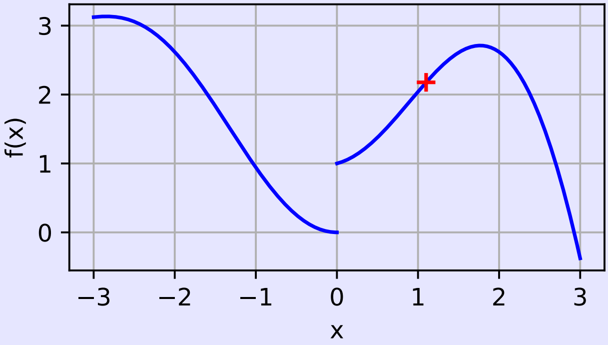
Regularity of functions: continuity

Previous example, $x = 1$



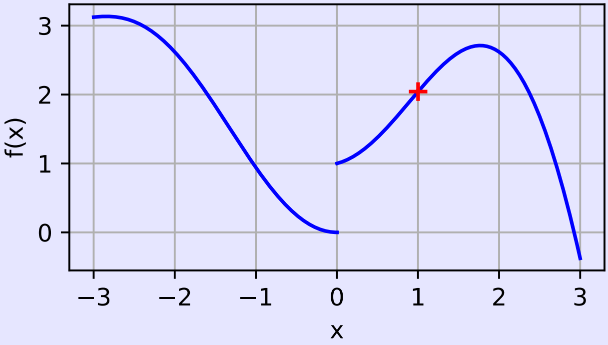
Regularity of functions: continuity

Previous example, $x = 1$



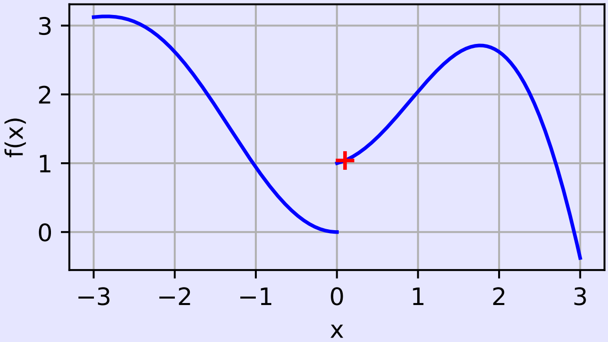
Regularity of functions: continuity

Previous example, $x = 1$



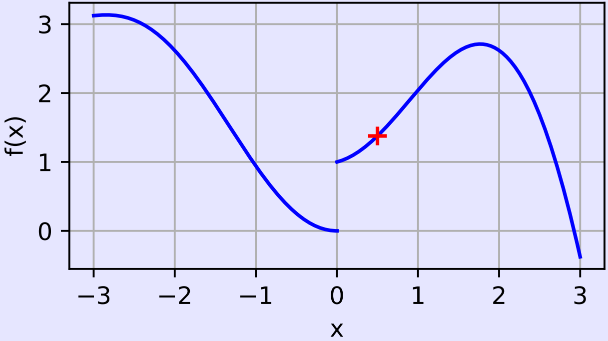
Regularity of functions: continuity

Previous example, $x = 1$



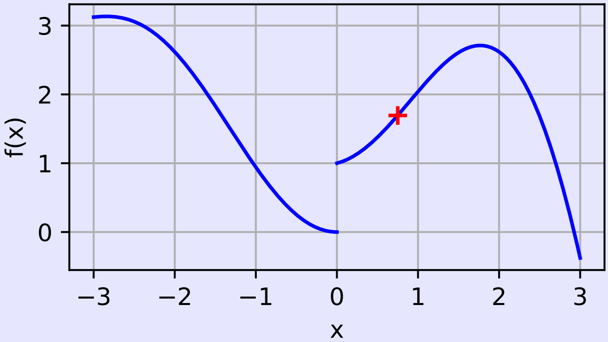
Regularity of functions: continuity

Previous example, $x = 1$



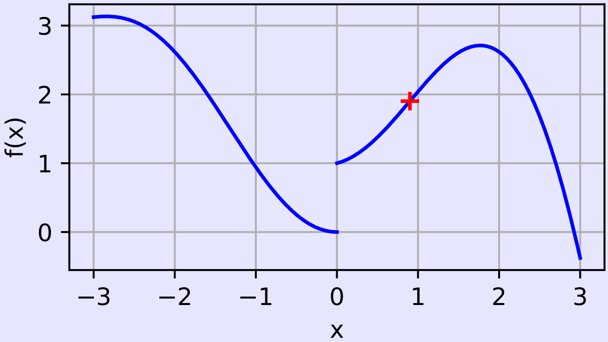
Regularity of functions: continuity

Previous example, $x = 1$



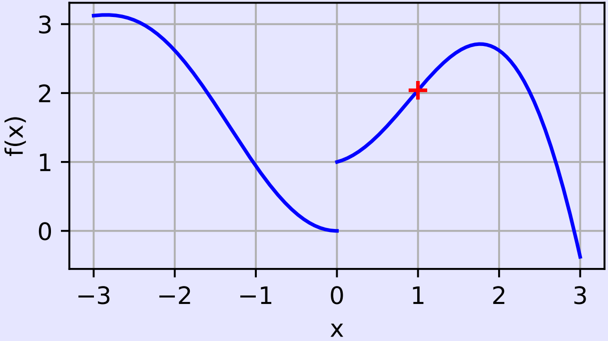
Regularity of functions: continuity

Previous example, $x = 1$



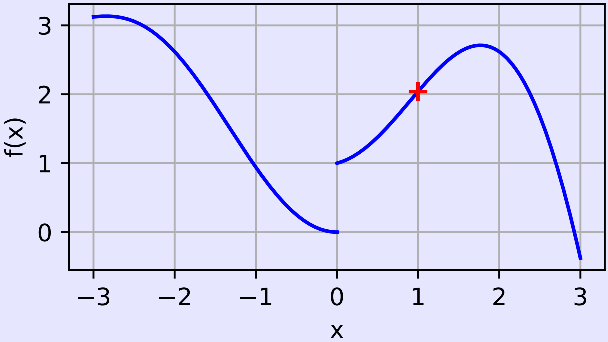
Regularity of functions: continuity

Previous example, $x = 1$



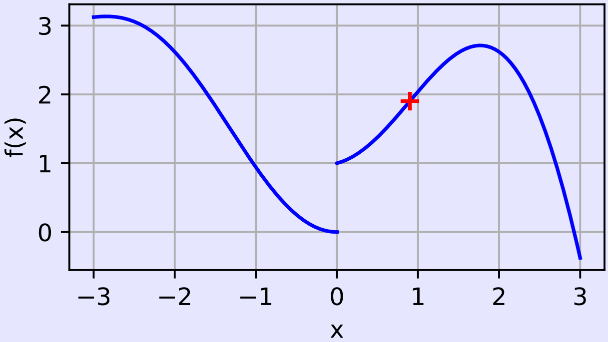
Regularity of functions: continuity

Previous example, $x = 1$. The function is **continuous** at 1 !



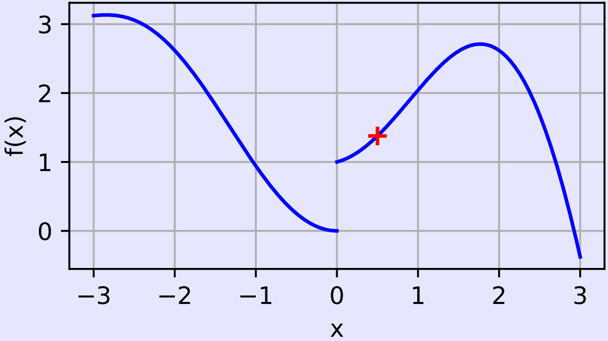
Regularity of functions: continuity

Previous example, $x = 0$



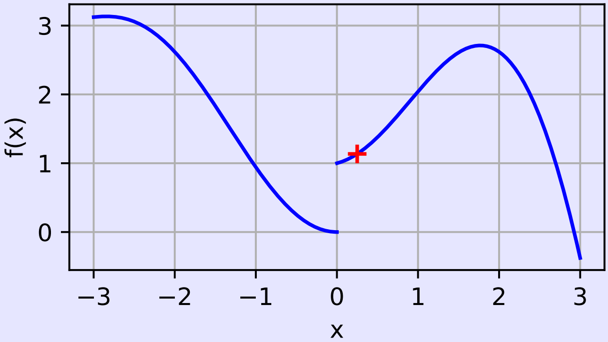
Regularity of functions: continuity

Previous example, $x = 0$



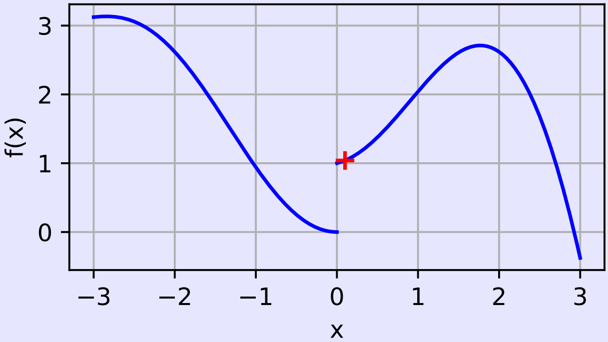
Regularity of functions: continuity

Previous example, $x = 0$



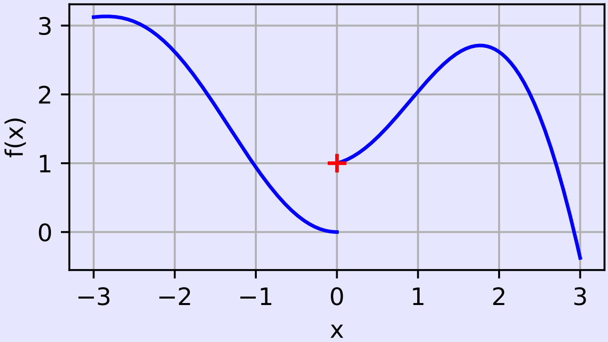
Regularity of functions: continuity

Previous example, $x = 0$



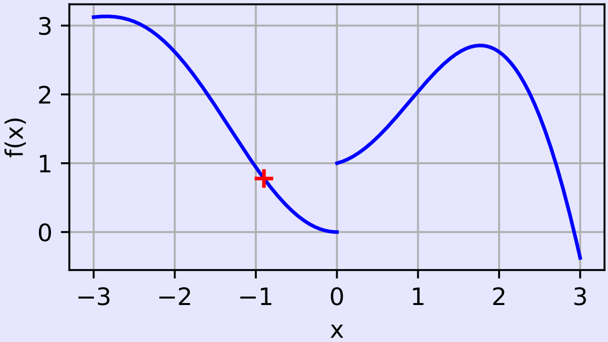
Regularity of functions: continuity

Previous example, $x = 0$



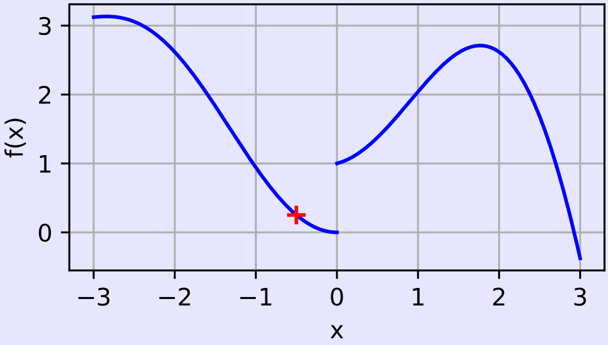
Regularity of functions: continuity

Previous example, $x = 0$



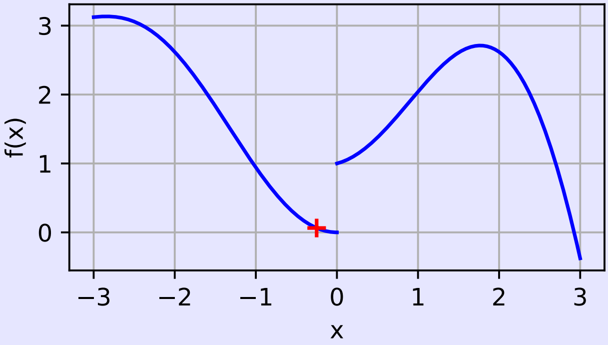
Regularity of functions: continuity

Previous example, $x = 0$



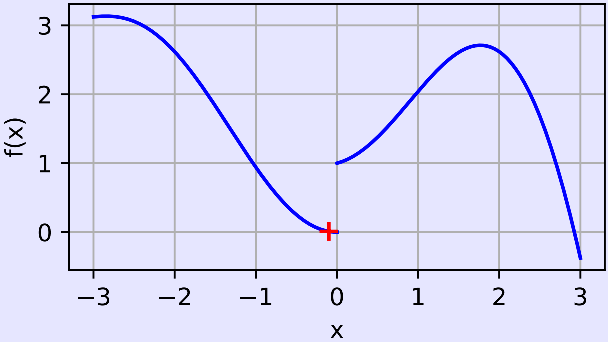
Regularity of functions: continuity

Previous example, $x = 0$



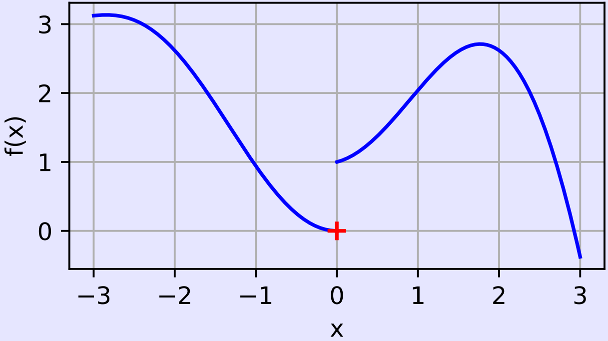
Regularity of functions: continuity

Previous example, $x = 0$



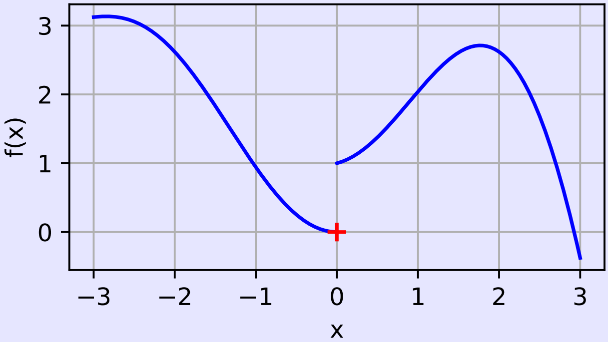
Regularity of functions: continuity

Previous example, $x = 0$



Regularity of functions: continuity

Previous example, $x = 0$. The function is **not continuous** at 0 !



Rules of continuity

- We say that f is continuous on a set $I \subset \mathbb{R}$ if it is continuous at all $x \in I$

Let f, g be two continuous functions on I, then

- $f + g$ is continuous
- $f \times g$ is continuous
- $f \circ g$ is continuous (if $g(x) \in I$ for all $x \in I$)
- $\frac{f}{g}$ is continuous if g does not cancel on I

Most 'basic' functions are continuous

- x^n on $\mathbb{R}$ for all $n \geq 0$
- $\sqrt{x}$ on $\mathbb{R}^+$
- x^n on $\mathbb{R}_*^+$ and $\mathbb{R}_*^-$ for all $n < 0$
- $|x|$, $\exp(x)$, $\log(x)$, $\cos(x)$, $\sin(x)$ where they are defined ...

Exercises

Show that the following functions are continuous on $\mathbb{R}$:

- $\exp(x^3) + x^2$
- $\sqrt{\exp(x) + 1 + \sin(x)}$
- $\frac{\cos(x^2)}{2+\sin(x)}$

Solution 1

- The functions x^3 and $\exp(x)$ are continuous on $\mathbb{R}$.
- Therefore, by composition, $\exp(x^3)$ is continuous on $\mathbb{R}$
- The function x^2 is continuous on $\mathbb{R}$
- Therefore, by summation, $\exp(x^3) + x^2$ is continuous on $\mathbb{R}$.

Solution 2

- The functions $\exp(x)$ and $\sin(x)$ are continuous on $\mathbb{R}$.
- Therefore, by summation, $\exp(x) + 1 + \sin(x)$ is continuous on $\mathbb{R}$
- Furthermore, we have $\sin(x) \geq -1$ and $\exp(x) > 0$, so $\exp(x) + 1 + \sin(x) > 0$.
- The function $\sqrt{x}$ is continuous on $\mathbb{R}^+$
- Therefore, by composition, $\sqrt{\exp(x) + 1 + \sin(x)}$ is continuous on $\mathbb{R}$.

Solution 3

- The functions $\cos(x)$ and x^2 are continuous on $\mathbb{R}$.
- Therefore, by composition, $\cos(x^2)$ is continuous on $\mathbb{R}$.
- The function $\sin(x) + 2$ is continuous and strictly positive > 0
- Therefore, $\frac{\cos(x^2)}{\sin(x)+2}$ is continous on $\mathbb{R}$.

Lipschitz regularity

We say that f is L–Lipschitz on I if for all $x, y \in I$,

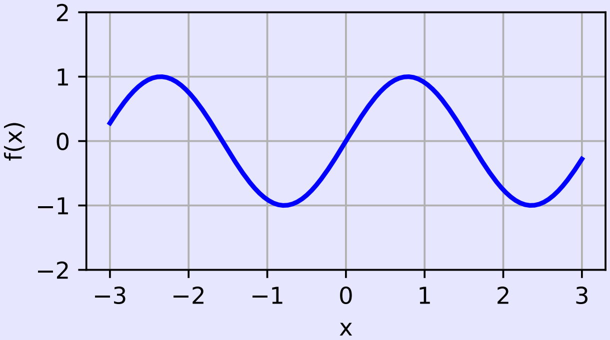
$$|f(x) - f(y)| \leq L|x - y|$$

- It is a form of regularity: it means that if x and y are close, then $f(x)$ and $f(y)$ must also be close, and it allows to control the distance between $f(x)$ and $f(y)$

A more intuitive interpretation

How can we tell that a function is Lipschitz looking at its graph ?

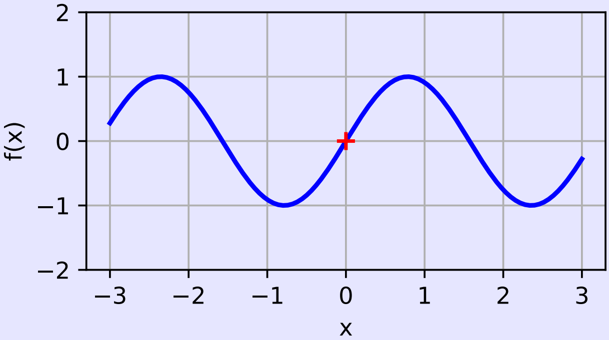
- Take a point



A more intuitive interpretation

How can we tell that a function is Lipschitz looking at its graph ?

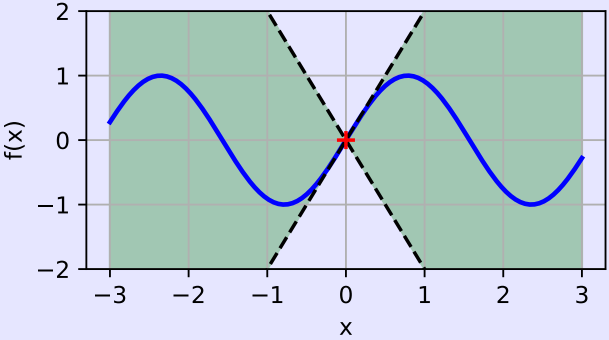
- Take a point



A more intuitive interpretation

How can we tell that a function is Lipschitz looking at its graph ?

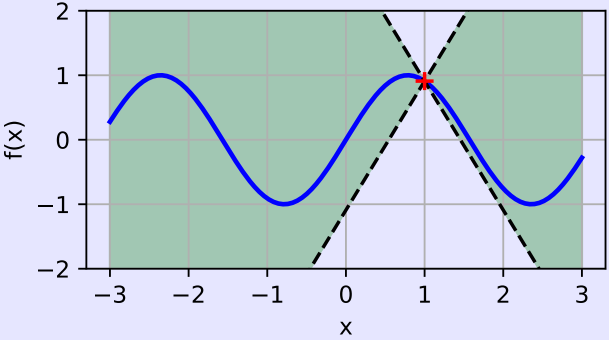
- We must be able to draw two lines that intersect at this point with slope L so that the graph of f is always between these lines
- Here, $L = 2$: the slope of the black lines is ± 2.



A more intuitive interpretation

How can we tell that a function is Lipschitz looking at its graph ?

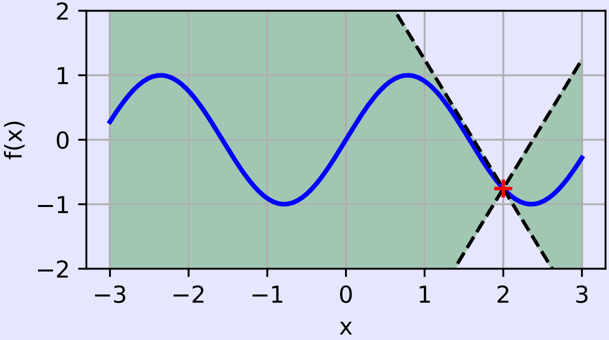
- This slope must work for *any* starting point
- Here, $L = 2$: the slope of the black lines is ± 2.



A more intuitive interpretation

How can we tell that a function is Lipschitz looking at its graph ?

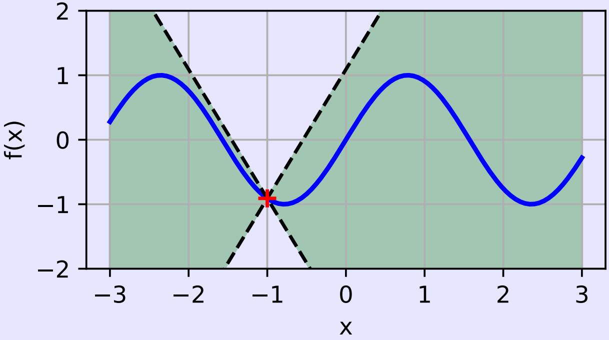
- This slope must work for *any* starting point
- Here, $L = 2$: the slope of the black lines is ± 2.



A more intuitive interpretation

How can we tell that a function is Lipschitz looking at its graph ?

- This slope must work for *any* starting point
- Here, $L = 2$: the slope of the black lines is ± 2.



Exercise

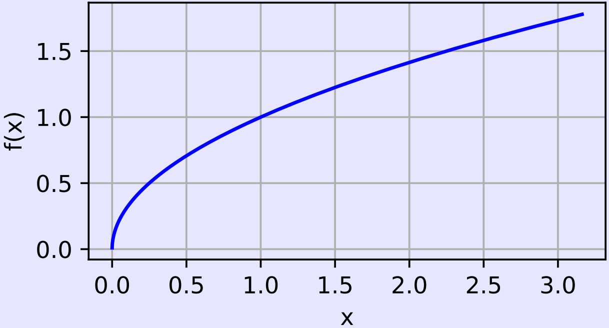
- Show that if f is Lipschitz, then it is continuous. Is the converse true?

Solution

- Assume that f is $L-$Lipschitz, and fix $x \in \mathbb{R}$
- Let x_n that converges to x
- Then, $|f(x_n) - f(x)| \leq L|x_n - x|$
- Since the right hand side goes to 0, the left hand side too
- Therefore, $f(x_n)$ converges to $f(x)$

Solution 2

The function $\sqrt{x}$ is continuous, but not Lipschitz (problem at 0: infinite slope)



Differentiable functions

If f is continuous at x, we know that $f(x+h) - f(x)$ goes to 0 as h goes to 0.
We say that f is **differentiable** at x if

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

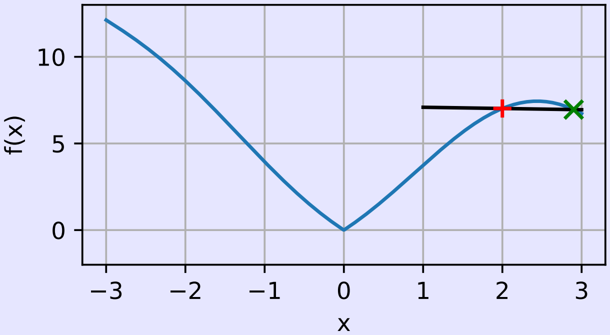
exists.
This means that when h is small:

$$f(x+h) \simeq f(x) + h \times f'(x)$$

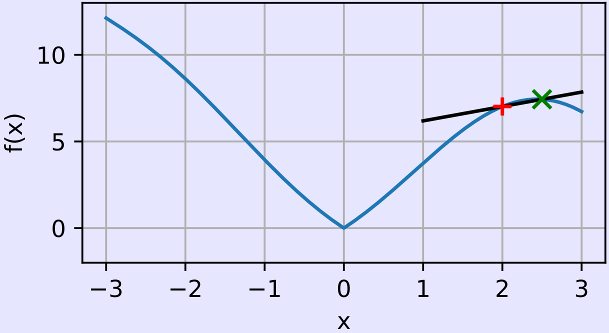
In other words, f is well approximated by an affine function.

- We say that f is differentiable on a set I if it is differentiable for all $x \in I$

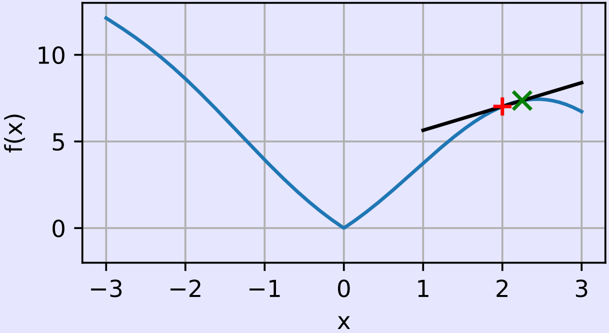
Regularity of functions: differentiability



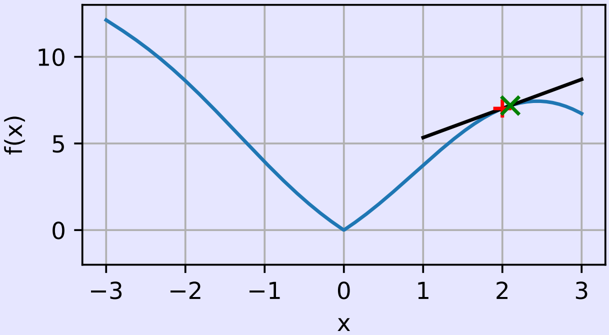
Regularity of functions: differentiability



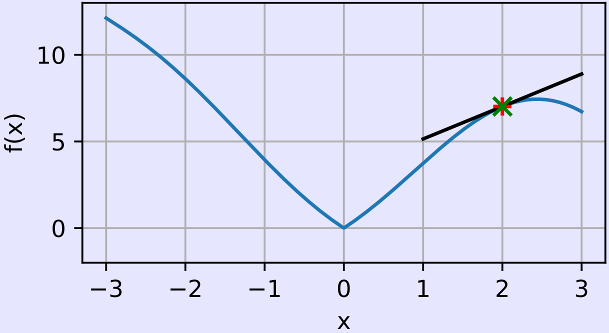
Regularity of functions: differentiability



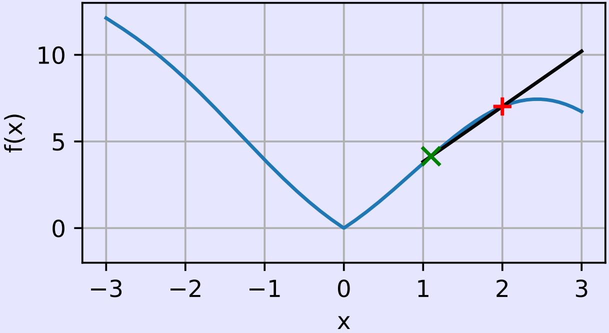
Regularity of functions: differentiability



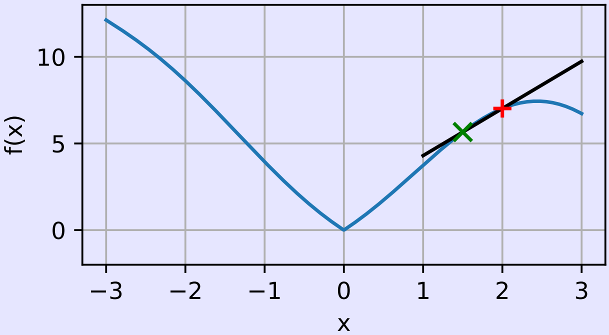
Regularity of functions: differentiability



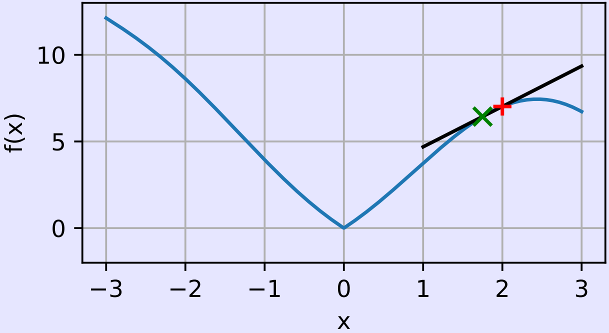
Regularity of functions: differentiability



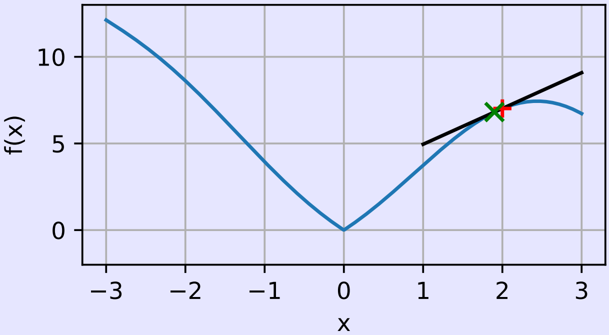
Regularity of functions: differentiability



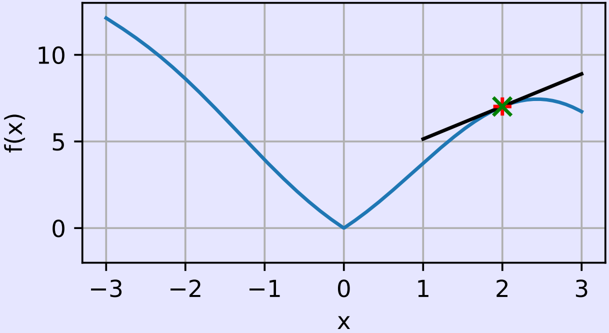
Regularity of functions: differentiability



Regularity of functions: differentiability

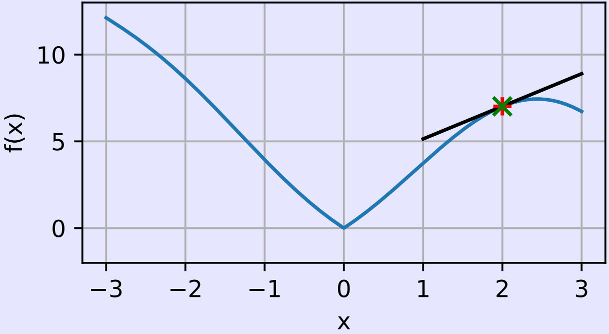


Regularity of functions: differentiability

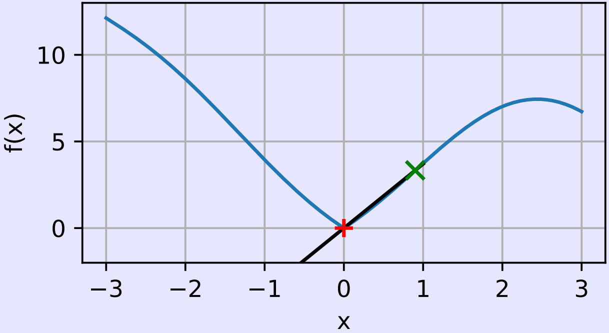


Regularity of functions: differentiability

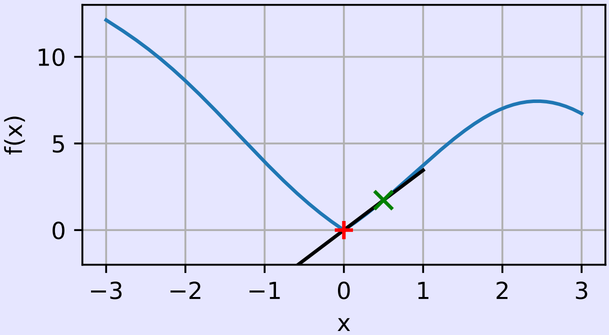
The function is **differentiable** at 2 !



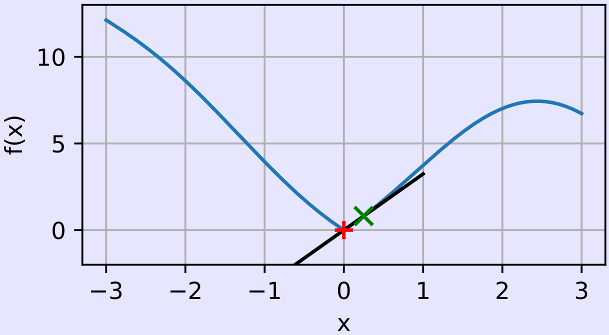
Regularity of functions: differentiability



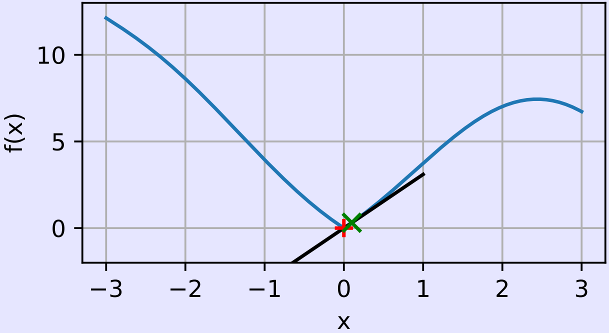
Regularity of functions: differentiability



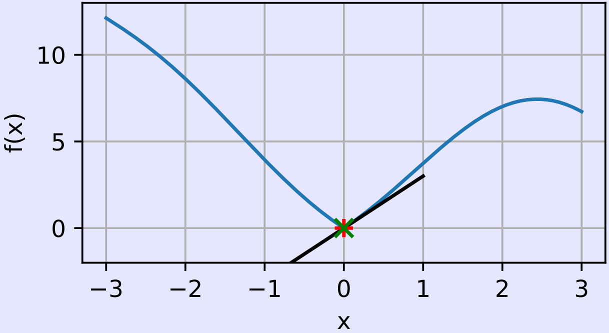
Regularity of functions: differentiability



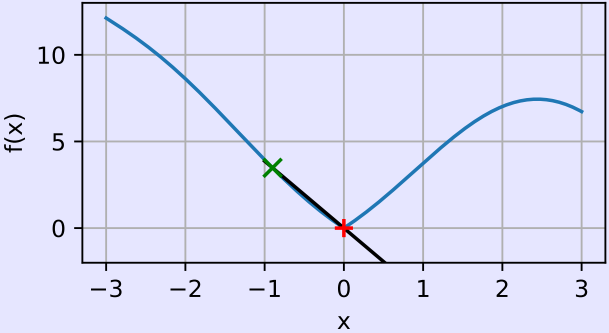
Regularity of functions: differentiability



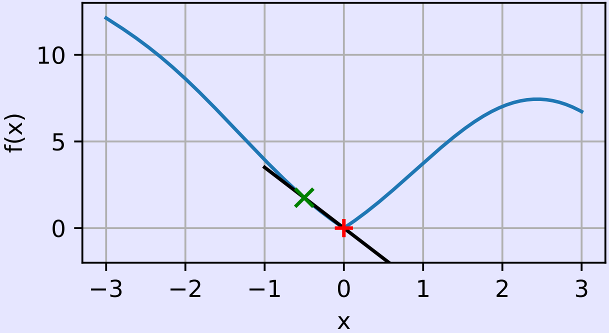
Regularity of functions: differentiability



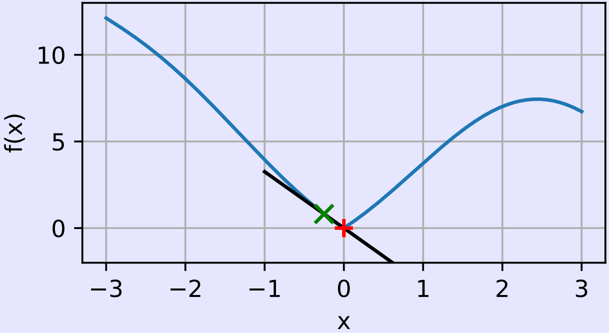
Regularity of functions: differentiability



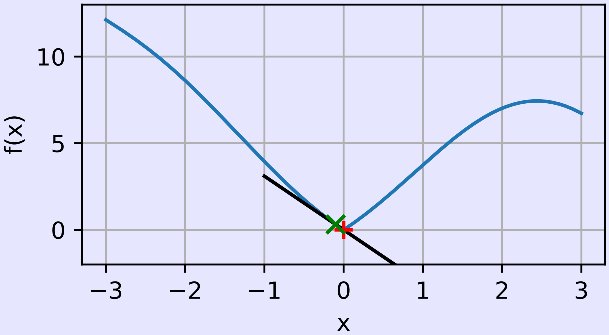
Regularity of functions: differentiability



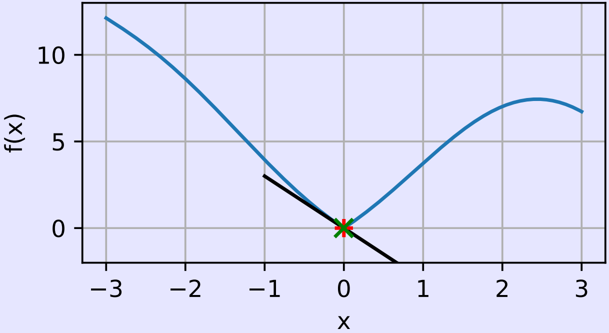
Regularity of functions: differentiability



Regularity of functions: differentiability

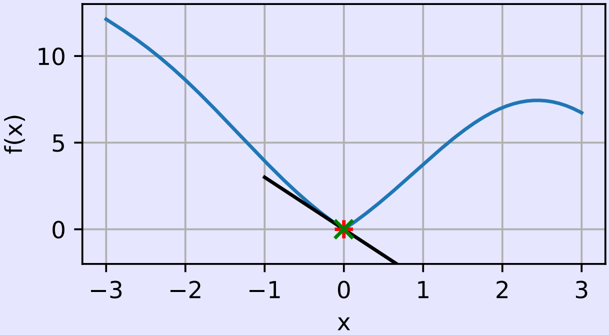


Regularity of functions: differentiability



Regularity of functions: differentiability

The function is **not differentiable** at 0 !



Rules of differentiation

If f and g are two differentiable functions on I then

- $f+g$ is differentiable and $(f+g)' = f' + g'$
- $f \times g$ is differentiable and $(f \times g)' = f' \times g + f \times g'$

The usual functions are differentiable

- The differential of x^n is $n \times x^{n-1}$
- The differential of $\exp(x)$ is $\exp(x)$
- The differential of $\log(x)$ is $\frac{1}{x}$
- The differential of $\sin(x)$ is $\cos(x)$
- The differential of $\cos(x)$ is $-\sin(x)$
- $|x|$ is **not** differentiable at 0

Chain rule

If f and g are differentiable functions, then $f \circ g$ is differentiable and

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

Exercises

Compute the differential of the following functions

- $\exp(x^3) + x^2$
- $x\sqrt{\exp(x) + 1}$

Solution

$$\exp(x^3) + x^2$$

- We need to differentiate $\exp(x^3)$. This is the composition of $\exp$ and x^3
- We apply the chain rule: $\left(\exp(x^3)\right)' = (x^3)' \times \exp'(x^3) = 3x^2 \exp(x^3)$
- The differential of x^2 is $2x$
- By summation, $\left(\exp(x^3) + x^2\right)' = 3x^2 \exp(x^3) + 2x$

Solution

$$x\sqrt{\exp(x)+1}$$

- This is a product, we must differentiate x and $\sqrt{\exp(x)+1}$
- $\sqrt{\exp(x)+1}$ is a composition. We know that $(\sqrt{x})' = \frac{1}{2}x^{-\frac{1}{2}} = \frac{1}{2\sqrt{x}}$
- Furthermore $(\exp(x)+1)' = \exp(x)$
- Using the chain rule: $(\sqrt{\exp(x)+1})' = \exp(x) \times \frac{1}{2\sqrt{\exp(x)+1}} = \frac{\exp(x)}{2\sqrt{\exp(x)+1}}$
- Using the product rule, we find

$$\left(x\sqrt{\exp(x)+1}\right)' = x\frac{\exp(x)}{2\sqrt{\exp(x)+1}} + \sqrt{\exp(x)+1}$$

In higher dimensions

previously we ve seen function that maps scalar to scalar (R to R)
now fct maps mulitple var x1 to xp to a single variable

Multi-dimensional functions

- In many applications, we have more than 1 variable
- $f(x_1, \ldots, x_p) \in \mathbb{R}$
- For instance, in machine learning, $x_1, \ldots, x_p$ can be the parameters of a model.

Multi-dimensional functions

For instance, in dimension 2:

$$f(x_1, x_2) = \frac{x_1}{x_2} + 3 + x_1 + 2x_2$$

$f(4,2) = \frac{4}{2} + 3 + 4 + 2 \times 2 = 13.$

Continuity

We call $\mathbf{x} = [x_1, \ldots, x_p] \in \mathbb{R}^p$ the vector of inputs. The function is $f(\mathbf{x})$.

- We say that f is continuous at $\mathbf{x}$ if when $\mathbf{y} \to \mathbf{x}$, then $f(\mathbf{y}) \to f(\mathbf{x})$.

Usual functions are continuous:

- $f(x_1, \ldots, x_p) = x_1 + \cdots + x_p$
- $f(x_1, \ldots, x_p) = x_1 \times \cdots \times x_p$
- $f(x_1, \ldots, x_p) = x_1$

Linear functions

A particular type of functions are *linear functions*. For a vector $\mathbf{a} = [a_1, \ldots, a_p]$, the associated linear function is

$$f_{\mathbf{a}}(x_1, \ldots, x_p) = a_1 x_1 + \cdots + a_p x_p$$

for instance:

- $f(x_1, x_2) = 3x_1 + 2x_2$ avec vector a = [3, 2]
- $f(x_1, x_2, x_3) = -x_1 + 2x_3$ avec a = [-1, 0, 2]

Scalar products

The **scalar product** between two vectors $\mathbf{x}$ and $\mathbf{y}$ is

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \cdots + x_p y_p$$

So the previous linear functions are

$$f_{\mathbf{a}}(\mathbf{x}) = \langle \mathbf{x}, \mathbf{a} \rangle$$

Differentiability

We can extend the notion of derivative to the higher dimensional setting. We say that f is differentiable at $\mathbf{x}$ if when $\mathbf{y} \to \mathbf{x}$, there exists a vector $\mathbf{g}$ such that we have

$$f(\mathbf{y}) \simeq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle$$

g derivative in 1d
instead of 1 val for detivative we have here p values in vector p -->gradient

- $\mathbf{g}$ is called the **gradient** of f at $\mathbf{x}$ and is denoted $\nabla f(\mathbf{x})$

nabla

Example : gradient of linear functions

If

$$f(x) = \langle \mathbf{x}, \mathbf{b} \rangle$$

Then

$$\nabla f(x) = \mathbf{b}$$

Constant gradient !

Example : gradient of quadratic functions

Let $A \in \mathbb{R}^{p\times p}$ a square, symmetric matrix

$$f(x) = \frac{1}{2}\langle \mathbf{x}, A\mathbf{x}\rangle$$

quadratique car si multiplie x par 2, x et Ax mult par 2 donc au final le total est mult par 4 donc 2²

We take $\varepsilon \in \mathbb{R}$ a small number, $\mathbf{d} \in \mathbb{R}^p$ a vector, and look at

$$f(\mathbf{x}+\varepsilon\mathbf{d}) = \frac{1}{2}\langle \mathbf{x}+\varepsilon\mathbf{d}, A(\mathbf{x}+\varepsilon\mathbf{d})\rangle$$

when e go to zero we will approch to x in the direction d

$$= \frac{1}{2}\langle \mathbf{x}, A\mathbf{x}\rangle + \frac{1}{2}\langle \mathbf{x}, A(\varepsilon\mathbf{d})\rangle + \frac{1}{2}\langle \varepsilon\mathbf{d}, A\mathbf{x}\rangle + \frac{1}{2}\langle \varepsilon\mathbf{d}, A(\varepsilon\mathbf{d})\rangle$$

since sym les 2 termes sont egaux et epsilon peut etre mis devant A car A linear map et aussi dehors du scalar pdct car scalar pdt linear

$$= \underbrace{\frac{1}{2}\langle \mathbf{x}, A\mathbf{x}\rangle}_{f(\mathbf{x})} + \varepsilon\langle A\mathbf{x}, \mathbf{d}\rangle + \frac{\varepsilon^2}{2}\langle \cancel{\varepsilon}\mathbf{d}, A\mathbf{d}\rangle$$

1st order term

2nd order term

To find the gradient, only look at the first order expansion:

$$f(\mathbf{x}+\varepsilon\mathbf{d}) \simeq f(\mathbf{x}) + \varepsilon\langle A\mathbf{x}, \mathbf{d}\rangle$$

2nd order term negligeable car e² tres petit

Example : gradient of quadratic functions

$$f(x) = \frac{1}{2}\langle \mathbf{x}, A\mathbf{x}\rangle$$

$$f(\mathbf{x}+\varepsilon\mathbf{d}) \simeq f(\mathbf{x}) + \varepsilon\langle \underbrace{A\mathbf{x}}_{\nabla f(\mathbf{x})}, \mathbf{d}\rangle$$

We conclude:

$$\nabla f(\mathbf{x}) = A\mathbf{x}$$

- Linear gradient !

Exercise

Let $\mathbf{w}_1, \ldots, \mathbf{w}_n \in \mathbb{R}^p$, and define

fct input vector x

output scalar

$$f(\mathbf{x}) = \sum_{i=1}^{n} \exp(-\langle \mathbf{w}_i, \mathbf{x} \rangle)$$

Compute $\nabla f(\mathbf{x})$.

-->take small perturbation of x and derive first order term of expression

Solution

We take $\varepsilon \in \mathbb{R}$ a small number, $\mathbf{d} \in \mathbb{R}^p$ a vector, and look at $f(\mathbf{x} + \varepsilon \mathbf{d})$

direction vector

$$
\begin{aligned}
f(\mathbf{x} + \varepsilon \mathbf{d}) &= \sum_{i=1}^{n} \exp(-\langle \mathbf{w}_i, \mathbf{x} + \varepsilon \mathbf{d} \rangle) \\
&= \sum_{i=1}^{n} \exp(-\langle \mathbf{w}_i, \mathbf{x} \rangle - \varepsilon \langle \mathbf{w}_i, \mathbf{d} \rangle) \\
&= \sum_{i=1}^{n} \exp(-\langle \mathbf{w}_i, \mathbf{x} \rangle) \exp(-\varepsilon \langle \mathbf{w}_i, \mathbf{d} \rangle) \\
&\simeq \sum_{i=1}^{n} \exp(-\langle \mathbf{w}_i, \mathbf{x} \rangle)(1 - \varepsilon \langle \mathbf{w}_i, \mathbf{d} \rangle) \\
&\simeq \sum_{i=1}^{n} \exp(-\langle \mathbf{w}_i, \mathbf{x} \rangle) - \varepsilon \sum_{i=1}^{n} \exp(-\langle \mathbf{w}_i, \mathbf{x} \rangle) \langle \mathbf{w}_i, \mathbf{d} \rangle
\end{aligned}
$$

dvt limité exp

Solution

We therefore find

$$f(\mathbf{x} + \varepsilon \mathbf{d}) = f(\mathbf{x}) + \varepsilon \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle$$

with

$$\nabla f(\mathbf{x}) = \sum_{i=1}^{n} \exp(-\langle \mathbf{w}_i, \mathbf{x} \rangle) \mathbf{w}_i$$

Another view on the gradient: partial derivatives

We can recover the gradient by computing the derivatives of real functions. We fix $\mathbf{x} \in \mathbb{R}^p$, and define for $x \in \mathbb{R}$

petit x = scalar

$$f_i(x) = f((\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, x, \mathbf{x}_{i+1}, \dots, \mathbf{x}_p))$$

function that moves the coef i of f

- f_i is a function from $\mathbb{R}$ to $\mathbb{R}$, it corresponds to only changing one coordinate in $\mathbf{x}$

The derivative of f_i at $\mathbf{x}_i$ is called the i-th partial derivative of f:

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = f_i'(\mathbf{x}_i) \in \mathbb{R}$$

The gradient at $\mathbf{x}$ is given by

$$\nabla f(\mathbf{x}) = [\frac{\partial f}{\partial x_1}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_p}(\mathbf{x})] \in \mathbb{R}^p$$

Jacobian

now fct that maps multipl input to multi output
avant on avait single variable to single output --> derivative = single number
apres mult vat to sing output --> gradiant = vector
now derivative = jacobian matrix

The gradient is only defined for a *scalar* function, when $f(\mathbf{x})$ is a single real number. However, we will encounter many cases where $f(\mathbf{x}) \in \mathbb{R}^q$.

- The "derivative" of f at $\mathbf{x}$ is called the **Jacobian matrix**, and is noted as $J_f(\mathbf{x}) \in \mathbb{R}^{q \times p}$

It is such that for $\mathbf{y} \to \mathbf{x}$,

$$f(\mathbf{y}) \simeq f(\mathbf{x}) + J_f(\mathbf{x})(\mathbf{y} - \mathbf{x})$$

1st order perturbation

Example: linear functions

Let $A \in \mathbb{R}^{q \times p}$, and consider the function $f : \mathbb{R}^p \to \mathbb{R}^q$ defined by

$$f(\mathbf{x}) = A\mathbf{x}$$

Then, the Jacobian of f at $\mathbf{x}$ is

$$J_f(\mathbf{x}) = A$$