Vision and Image Processing: Very Brief Recall on Derivatives

François Lauze

Department of Computer Science
University of Copenhagen

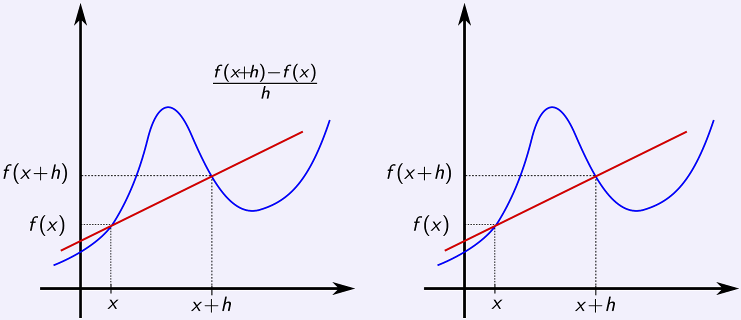
VIP, 24.11.2021

Plan for this lecture

- Derivatives in 1D
- Derivatives in 2D/3D, Gradients

Outline

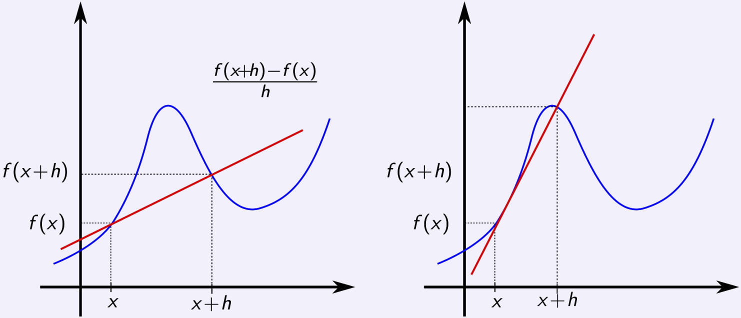
Secants, limits and derivatives



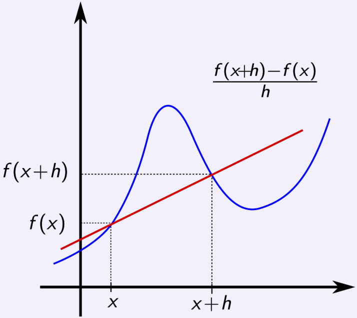
Secant : line joining two point on the graph of f. Slope between $(x, f(x))$ and $(x + h, f(x + h))$:

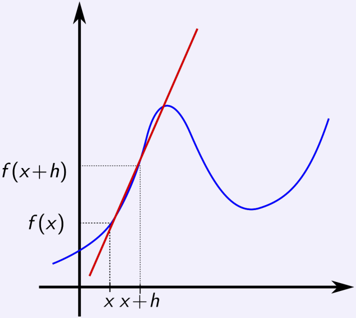
$$\frac{f(x+h)-f(x)}{(x+h)-x} = \frac{f(x+h)-f(x)}{h}$$

Secants, limits and derivatives

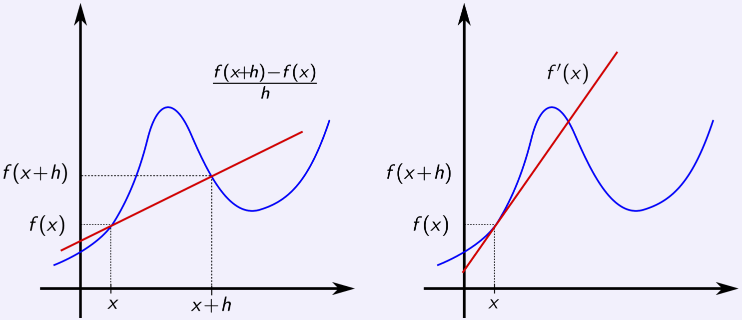


Secants, limits and derivatives





Secants, limits and derivatives



Derivative: limit of the slope when $h \to 0$ (but $h \neq 0$!) Slope of the tangent at x

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

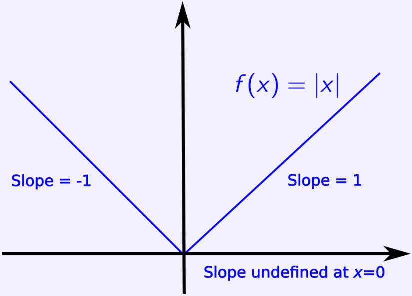
Derivative: a limit on the rate of change of a function.

Differentiable function

- f is *differentiable at* x if the limit above exists.
- f is *differentiable* if the limit exists for all x.

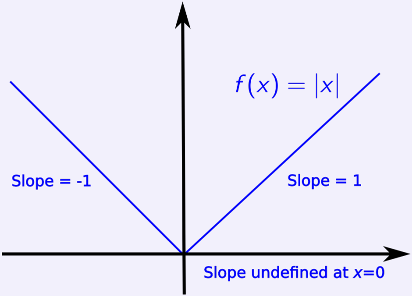
Differentiable function

- f is *differentiable at* x if the limit above exists.
- f is *differentiable* if the limit exists for all x.



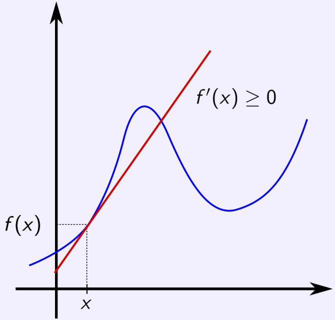
Differentiable function

- f is *differentiable at* x if the limit above exists.
- f is *differentiable* if the limit exists for all x.



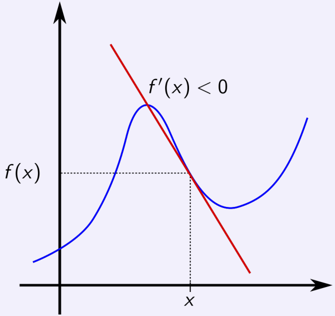
- $f(x) = |x|$ not differentiable at $x = 0$.

Derivatives and variations of a function



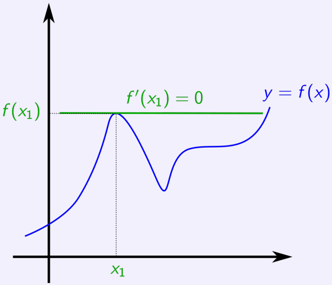
$f'(x) \geq 0$: f is increasing around x

Derivatives and variations of a function



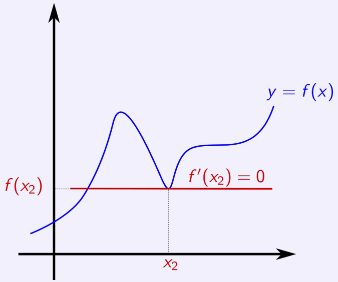
$f'(x) \leq 0$: f is decreasing around x

Null derivative



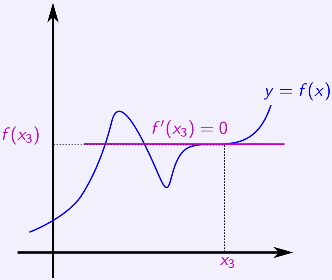
- $f'(x_1) = 0$, f is locally below its tangent: *local maximum*
- $f'(x) > 0$ if $x < x_1$ locally, $f'(x) < 0$ if $x > x_1$ locally.
- x_1 is a *local maximizer* of f.

Null derivative



- $f'(x_2) = 0$, f is locally above its tangent: *local minimum*
- $f'(x) < 0$ if $x < x_2$ locally, $f'(x) < 0$ if $x > x_2$ locally.
- x_2 is a *local minimizer* of f.

Null derivative



- $f'(x_3) = 0$, f is crosses its tangent. neither minimum nor maximum.
- $f'(x) > 0$ if $x < x_3$ locally, $f'(x) > 0$ if $x > x_3$ locally.
- x_3 is an *inflection point*.
- The opposite situation also possible: $f'(x) < 0$ if $x < x_3$ locally, $f'(x) < 0$ if $x > x_3$ locally. Make a picture!

Example: Derivative of x^n

Use the *binomial formula*

$$\begin{aligned}(a+b)^n &= \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \\ &= a^n + na^{n-1}b + \frac{n(n-1)}{2} a^{n-2} b^2 + \dots b^n\end{aligned}$$

Develop

$$\begin{aligned}\frac{(x+h)^n - x^n}{h} &= \frac{x^n + hnx^{n-1} + h^2 \frac{n(n-1)}{2} x^{n-1} + \cdots + h^n - x^n}{h} \\ &= nx^{n-1} + \underbrace{h\left(\frac{n(n-1)}{2} x^{n-1} + \cdots + h^{n-2}\right)}_{\text{goes } \to 0 \text{ when } h \to 0}\end{aligned}$$

The limit when $h \to 0$ is nx^{n-1}.

Some Classical Formulas

function	derivative	domain/remark
x^α	$\alpha x^{\alpha-1}$	if α is not an integer, x should be > 0
e^x	e^x	$x \in \mathbb{R}$
$\ln x$	$\frac{1}{x}$	$x > 0$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$	special case of first rule with $\alpha = \frac{1}{2}$, $x > 0$
$\cos x$	$-\sin x$	$x \in \mathbb{R}$
$\sin x$	$\cos x$	$x \in \mathbb{R}$
$\tan x$	$\frac{1}{\cos^2 x}$	$x \neq k \pm \frac{\pi}{2}, k \in \mathbb{Z}$
$\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$	$-1 < x < 1$
$\arccos x$	$-\frac{1}{\sqrt{1-x^2}}$	$-1 < x < 1$
$\arctan x$	$\frac{1}{1+x^2}$	$x \in \mathbb{R}$

Example: Derivative of a Product – Leibniz Rule

Rule for differentiating $x \mapsto f(x)g(x)$. Write secant ratios:

Example: Derivative of a Product – Leibniz Rule

Rule for differentiating $x \mapsto f(x)g(x)$. Write secant ratios:

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

Example: Derivative of a Product – Leibniz Rule

Rule for differentiating $x \mapsto f(x)g(x)$. Write secant ratios:

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

At the limit when $h \to 0$,

- The first term
$$f(x+h)\frac{g(x+h)-g(x)}{h} \to f(x)g'(x)$$
- The second term
$$\frac{f(x+h)-f(x)}{h}g(x) \to f'(x)g(x)$$
- We get Leibniz Rule: $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$.

Classical Rules for Computing Derivatives

function	derivative	rule name
$\lambda f(x)$	$\lambda f'(x)$	scalar multiplication rule
$f(x) + g(x)$	$f'(x) + g'(x)$	sum rule
$f(x)g(x)$	$f'(x)g(x) + f(x)g'(x)$	Leibniz rule
$\frac{f(x)}{g(x)}$	$\frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$	quotient rule
$f(g(x))$	$f'(g(x))g'(x)$	chain rule
$e^{f(x)}$	$f'(x)e^{f(x)}$	exponentiation rule (chain rule!)
$\ln \lvert f(x) \rvert$	$\frac{f'(x)}{f(x)}$	logarithm rule
$(f(x)g(x))''$	$f''(x)g(x) + 2f'(x)g'(x) + g''(x)$	iterated Leibniz rule

$f''(x)$ is the derivative of $f'(x)$ at x. Second (order) derivative of f.

A Not That Simple Example: $f(x) = x\cos(e^{\sin(x)})$

- Write as $f(x) = g(x)h(x)$ with $g(x) = x$, $h(x) = \cos(e^{\sin(x)})$

$$\begin{aligned} f'(x) &= g'(x)h(x) + g(x)h'(x) && \text{(Leibniz rule)} \\ &= \cos(e^{\sin(x)}) + xh'(x) && (g'(x) = 1) \end{aligned}$$

- Write $h(x)$ as $l(m(x))$ with $l(x) = \cos(x)$, $m(x) = e^{\sin(x)}$.

$$\begin{aligned} h'(x) &= l'(m(x))m'(x) && \text{(chain rule)} \\ &= -\sin(m(x))m'(x) && (\cos'(x) = -\sin(x)) \end{aligned}$$

- Use Exponential rule $e^{a(x)} = a'(x)e^{a(x)}$ for $m(x) = e^{\sin(x)}$

$$\left(e^{\sin(x)}\right)' = \cos(x)e^{\sin(x)}, \quad (\sin'(x) = \cos(x))$$

- Reassemble the parts

$$f'(x) = \cos(e^{\sin(x)}) - x\sin(e^{\sin(x)})\cos(x)e^{\sin(x)}$$

Example: Arithmetic Mean

- Given n real numbers $x_1, \ldots, x_n$.

Example: Arithmetic Mean

- Given n real numbers $x_1, \ldots, x_n$.
- Classical Arithmetic Mean

$$\mathbf{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Example: Arithmetic Mean

- Given n real numbers $x_1, \ldots, x_n$.
- Classical Arithmetic Mean

$$\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Variance function $V(x) = \frac{1}{2n}\sum_{i=1}^{n}(x - x_i)^2$.

Example: Arithmetic Mean

- Given n real numbers $x_1, \ldots, x_n$.
- Classical Arithmetic Mean

$$\mathbf{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Variance function $V(x) = \frac{1}{2n} \sum_{i=1}^{n} (x - x_i)^2$.
- When is the derivative 0?

$$\begin{aligned} V'(x) &= \frac{1}{2n} \sum_{i=1}^{n} \frac{d(x - x_i)^2}{dx} = \frac{1}{2n} \sum_{i=0}^{n} 2(x - x_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} (x - x_i) = \frac{1}{n} \sum_{i=1}^{n} x - \frac{1}{n} \sum_{i=1}^{n} x_i \\ &= x - \mathbf{x} \end{aligned}$$

Example: Arithmetic Mean

- Given n real numbers $x_1, \ldots, x_n$.
- Classical Arithmetic Mean

$$\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Variance function $V(x) = \frac{1}{2n}\sum_{i=1}^{n}(x - x_i)^2$.
- When is the derivative 0?

$$\begin{aligned} V'(x) &= \frac{1}{2n}\sum_{i=1}^{n}\frac{d(x-x_i)^2}{dx} = \frac{1}{2n}\sum_{i=0}^{n} 2(x - x_i) \\ &= \frac{1}{n}\sum_{i=1}^{n}(x - x_i) = \frac{1}{n}\sum_{i=1}^{n} x - \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= x - \mathbf{x} \end{aligned}$$

- $V'(x) = 0$ if and only if $x = \mathbf{x}$. Minimum? Maximum? Inflection point?

Example: Arithmetic Mean

- Given n real numbers $x_1, \ldots, x_n$.
- Classical Arithmetic Mean

$$\mathbf{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Variance function $V(x) = \frac{1}{2n} \sum_{i=1}^{n} (x - x_i)^2$.
- When is the derivative 0?

$$\begin{aligned} V'(x) &= \frac{1}{2n} \sum_{i=1}^{n} \frac{d(x - x_i)^2}{dx} = \frac{1}{2n} \sum_{i=0}^{n} 2(x - x_i) \\ &= \frac{1}{n} \sum_{i=1}^{n} (x - x_i) = \frac{1}{n} \sum_{i=1}^{n} x - \frac{1}{n} \sum_{i=1}^{n} x_i \\ &= x - \mathbf{x} \end{aligned}$$

- $V'(x) = 0$ if and only if $x = \mathbf{x}$. Minimum? Maximum? Inflection point?
- Minimum! Why?

Example: Arithmetic Mean

- Given n real numbers $x_1, \dots, x_n$.
- Classical Arithmetic Mean

$$\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Variance function $V(x) = \frac{1}{2n}\sum_{i=1}^{n}(x - x_i)^2$.
- When is the derivative 0?

$$\begin{aligned} V'(x) &= \frac{1}{2n}\sum_{i=1}^{n}\frac{d(x - x_i)^2}{dx} = \frac{1}{2n}\sum_{i=0}^{n} 2(x - x_i) \\ &= \frac{1}{n}\sum_{i=1}^{n}(x - x_i) = \frac{1}{n}\sum_{i=1}^{n} x - \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= x - \mathbf{x} \end{aligned}$$

- $V'(x) = 0$ if and only if $x = \mathbf{x}$. Minimum? Maximum? Inflection point?
- Minimum! Why?
- The arithmetic mean is the *unique* point where $V(x)$ is minimum!

Outline

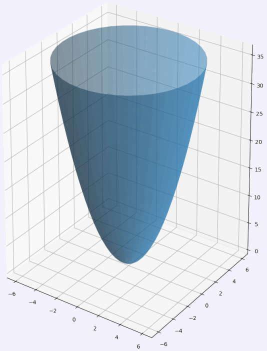
Partial derivatives

- Function $(x, y) \in \mathbb{R}^2 \mapsto f(x, y) \in \mathbb{R}$: two variables x and y.
- partial derivative $\frac{\partial f}{\partial x}(x, y)$ in x-direction: derivative of $x \mapsto f(x, y)$.

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

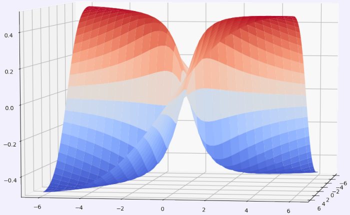
- partial derivative $\frac{\partial f}{\partial x}(x, y)$ in y-direction: derivative of $y \mapsto f(x, y)$

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



- if partial derivatives exist at (x_0, y_0) f is *differentiable* at (x_0, y_0).
- if partial derivatives exist everywhere, f is *differentiable*.
- Example $f(x, y) = x^2 + y^2$ (squared Euclidean distance).
- Partial derivatives $\frac{\partial f}{\partial x} = 2x$, $\frac{\partial f}{\partial y} = 2y$
- f is differentiable.
- Euclidean distance: Is $f(x, y) = \sqrt{x^2 + y^2}$ differentiable?

Example $f(x,y) = \frac{xy}{1+x^2+y^2}$



- Partial derivatives
- in x

$$\frac{\partial f}{\partial x} = \frac{y(1-x^2+y^2)}{(1+x^2+y^2)^2}$$

- in y

$$\frac{\partial f}{\partial x} = \frac{x(1+x^2-y^2)}{(1+x^2+y^2)^2}$$

- Computation in x. y supposed to be fixed: Using quotient rule, power rule,

$$\begin{aligned}\frac{\partial f}{\partial x}(x,y) &= \frac{\frac{\partial xy}{\partial x}(1+x^2+y^2) - xy\frac{\partial(1+x^2+y^2)}{\partial x}}{(1+x^2+y^2)^2} \\ &= \frac{y(1+x^2+y^2) - xy(2x)}{(1+x^2+y^2)^2} = \frac{y(1-x^2+y^2)}{(1+x^2+y^2)^2}\end{aligned}$$

Differential, Gradient

- Differential of $f(x, y)$: the *line vector* made of partial derivatives.

$$Df(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

Differential, Gradient

- Differential of $f(x, y)$: the *line vector* made of partial derivatives.

$$Df(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

- Gradient of $f(x, y)$: the *column vector* made of partial derivatives.

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = Df(x, y)^T$$

Differential, Gradient

- Differential of $f(x, y)$: the *line vector* made of partial derivatives.

$$Df(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

- Gradient of $f(x, y)$: the *column vector* made of partial derivatives.

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = Df(x, y)^T$$

- Difference between Differential and Gradient: not too much *in this course!*

Differential, Gradient

- Differential of $f(x, y)$: the *line vector* made of partial derivatives.

$$Df(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

- Gradient of $f(x, y)$: the *column vector* made of partial derivatives.

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = Df(x, y)^T$$

- Difference between Differential and Gradient: not too much *in this course!*

$Df(x, y)^T Df(x, y)$ is a matrix, $\nabla f(x, y)^T \nabla f(x, y)$ is a real (Linear Algebra)

Differential, Gradient

- Differential of $f(x, y)$: the *line vector* made of partial derivatives.

$$Df(x,y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

- Gradient of $f(x, y)$: the *column vector* made of partial derivatives.

$$\nabla f(x,y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = Df(x,y)^T$$

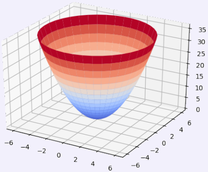
- Difference between Differential and Gradient: not too much *in this course!*

- $Df(x,y)^T Df(x,y)$ is a matrix, $\nabla f(x,y)^T \nabla f(x,y)$ is a real (Linear Algebra)

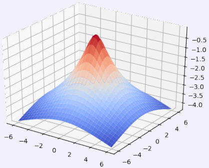
- $\nabla f(x,y)^T \nabla f(x,y) = \|\nabla f(x,y)\|^2$: length of gradient vector.

Gradients and Critical Points

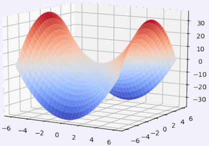
- A point (x, y) is *critical* for f if $\nabla f(x, y) = \vec{0}$. Minima, Maxima, Saddle points...



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



$g(x, y) = -\log(1 + x^2 + y^2)$



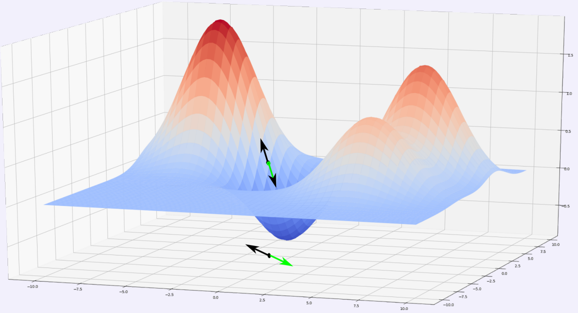
$h(x, y) = x^2 - y^2$

- All the three functions have only one critical point, at $(x, y) = (0, 0)$.
- For f: minimum
- For g: maximum
- For h: saddle point.

Gradients and Optimization

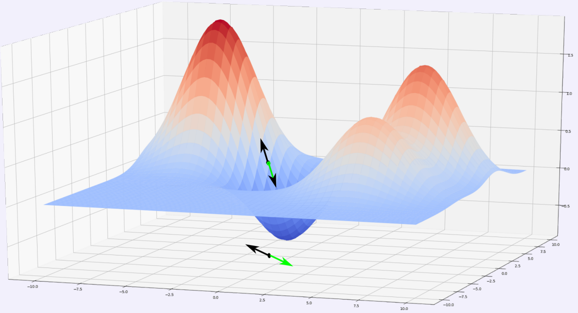
Gradients and Optimization

Fact: The gradient points in the direction of the steepest ascent of the function.



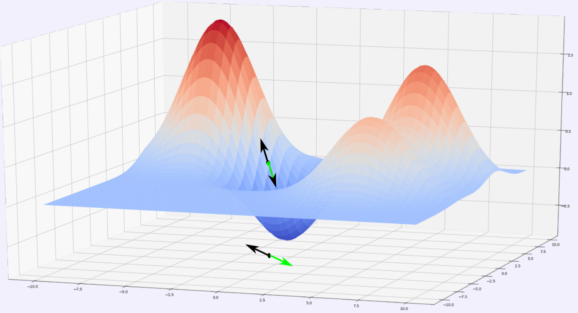
Gradients and Optimization

Fact: The gradient points in the direction of the steepest ascent of the function. Its opposite points in the direction of steepest descent.



Gradients and Optimization

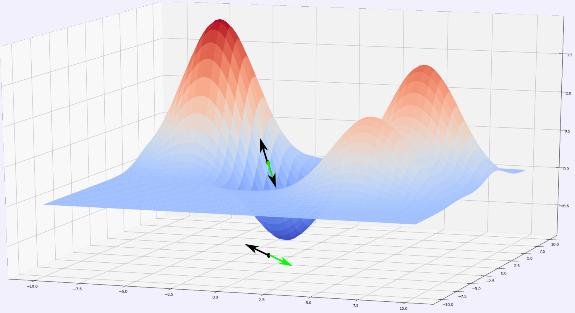
Fact: The gradient points in the direction of the steepest ascent of the function. Its opposite points in the direction of steepest descent.



What can this be useful for?

Gradients and Optimization

Fact: The gradient points in the direction of the steepest ascent of the function. Its opposite points in the direction of steepest descent.



What can this be useful for? **Optimization – gradient descent!**

Second order derivatives

- Second order derivatives in x: Different combinations

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

- *Mixed* Partial Derivatives: (equality from Schwarz' Theorem)

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

- *Hessian matrix* of f: *symmetric* matrix (still function of x and y)

$$\operatorname{Hess} f = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial xy} \\ \frac{\partial^2 f}{\partial xy} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

- *Laplacian* of f:

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \operatorname{Trace} \operatorname{Hess} f$$

Some Examples

- Hessian of $f(x, y) = x^2 + y^2$

$$\text{Hess} f = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = 2I_2, \quad I_2 = 2 \times 2\text{-identity matrix}$$

- $\Delta f = 4$.

- Hessian of $f(x, y) = \frac{xy}{1+x^2+y^2}$: not as nice as previous one!

$$\text{Hess} f = \frac{1}{(x^2 + y^2 + 1)^3} \begin{bmatrix} 2xy\left(x^2 - 3y^2 - 3\right) & 6x^2y^2 - x^4 - y^4 + 1 \\ 6x^2y^2 - x^4 - y^4 + 1 & -2xy\left(3x^2 - y^2 + 3\right) \end{bmatrix}$$

- Laplacian of f

$$\Delta f = -\frac{4xy\left(x^2 + y^2 + 3\right)}{\left(x^2 + y^2 + 1\right)^3}$$

In More Variables

Function $f(x_1, x_2, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$.

- Partial Derivative w.r.t x_i

$$\frac{\partial f}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, x_2, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots x_n) - f(x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots x_n)}{h}$$

- The same, but we need more letters!
- Differentials, Gradients:

$$Df = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right), \quad \nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)^T$$

- Hessian: $n \times n$ symmetric matrices

$$\operatorname{Hess} f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_1 x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 x_n} & \frac{\partial^2 f}{\partial x_2 x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

- Laplacian

$$\nabla f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x^i}$$

That's all Folk!