# KL divergence is not Lipschitz smooth

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft : June 24, 2019
Last update : October 16, 2019

# KL divergence

Given two scalar $x, y$, the KL divergence is defined as

$$f_{\mathsf{KL}}(x,y) = x\log\frac{x}{y} - x + y.$$

- dom$f = \ ]0,\infty[$.
  i.e. the domain of $f$ is the whole real line except $0$ and negative numbers
- $f$ is not symmetric : $f(x,y) \neq f(y,x)$
- $f$ is not a distance function : it does not fulfill the definition of distance
- $f$ is not Lipschitz smooth $\leftarrow$ the focus of this document

# Lipschitz smooth

**Definition**. A function $f$ is called Lipschitz smooth if the gradient $\nabla f$ is Lipschitz.

**Definition**. A function $g$ is called Lipschitz if and only if there exists a finite valued $L$ such that

$$\frac{\|g(\mathbf{a}) - g(\mathbf{b})\|}{\|\mathbf{a} - \mathbf{b}\|} \leq L$$

for all $\mathbf{a}, \mathbf{b} \in \text{dom} g$.

The key points of the definition are :

1. The gradient $\nabla g$ exists.
2. $L \in [0, \infty[$ : it has to be finite (cannot be $\infty$). That is, the fraction $\frac{\|g(\mathbf{a}) - g(\mathbf{b})\|}{\|\mathbf{a} - \mathbf{b}\|}$ is bounded between $0$ and $\infty$.
3. Conditions (1) and (2) have to hold for all $\mathbf{a}$,$\mathbf{b}$ inside dom$g$ : if there exists one pair of $(\mathbf{a}, \mathbf{b})$ that either (1) and/or (2) is false, then $g$ is not Lipschitz.

# LK divergence is not Lipschitz smooth

KL divergence is not Lipschitz smooth because of point (3) : there is a pair of point such that $L$ is unbounded.

Proof : the idea is to show there is a pair $(\mathbf{a}, \mathbf{b})$ such that

$$\frac{\|\nabla f(\mathbf{a}) - \nabla f(\mathbf{b})\|}{\|\mathbf{a} - \mathbf{b}\|} \leq L,$$

where $L \to \infty$ (unbounded).

Note. $f$ for KL divergence is a scalar function of two variables so we have $\dfrac{\partial f}{\partial x}$, $\dfrac{\partial f}{\partial y}$.

# The partial derivatives of KL-divergence

Given two scalars $x$, $y \in \ ]0,\infty[$, we have

$$f(x,y) = x\log\frac{x}{y} - x + y$$

Assume $\log$ is natural logarithm so that $\dfrac{d\log x}{dx} = \dfrac{1}{x}$, we have

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

We also have

$$\frac{\partial f(x,y)}{\partial x} = \log x - \log y$$

# The proof of KL-divergence is not Lipschitz smooth

Fix $x$, at any two point $y_1, y_2 \in \mathrm{dom} f$ we have

$$\frac{\partial f(x,y_1)}{\partial y} - \frac{\partial f(x,y_2)}{\partial y} = \left(1 - \frac{x}{y_1}\right) - \left(1 - \frac{x}{y_2}\right) = x\left(\frac{1}{y_2} - \frac{1}{y_1}\right)$$

$$\frac{\frac{\partial f(x,y_1)}{\partial y} - \frac{\partial f(x,y_2)}{\partial y}}{y_1 - y_2} = \frac{x}{y_1 y_2} \to \infty \text{ if } y_1 \to 0$$

Hence there exists a pair of $y_1, y_2$ that

$$\frac{\left\|\frac{\partial f(x,y_1)}{\partial y} - \frac{\partial f(x,y_2)}{\partial y}\right\|}{\|y_1 - y_2\|} \leq L$$

is unbounded. Therefore, $f$ is not Lipschitz smooth.

For the direction of fixing $y$ with the pair $(x_1, x_2)$, the same result can be drawn (which requires to use L'Hospital rule).

-END-