

# Random notes about Math for Deep Learning Models

Last update: 18 APR 2020

## Contents

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>Introduction</b>   | <b>1</b> |
| <b>2</b> | <b>Differentiation</b>  | <b>2</b> |
| 2.1      | Introduction about Differentiation . . . . .                      | 2        |
| 2.2      | Differentiation Definition in Mathematics . . . . .               | 2        |
| 2.3      | Some Theorems in context of Differentiation . . . . .             | 2        |
| <b>3</b> | <b>What is convolution in mathematics</b>                         | <b>3</b> |
| 3.1      | Convolution definition . . . . .                                  | 3        |
| 3.2      | Convolution commutativity . . . . .                               | 3        |
| 3.3      | Time invariant property of convolution . . . . .                  | 4        |
| 3.4      | Some Convolution Algebraic Properties . . . . .                   | 4        |
| <b>4</b> | <b>What is cross-correlation in math</b>                          | <b>4</b> |
| 4.1      | Cross-Correlation definition . . . . .                            | 4        |
| 4.2      | Cross-Correlation connection with convolution . . . . .           | 5        |
| 4.3      | Some Cross-Correlation Properties not necessary to hold . . . . . | 5        |
| <b>5</b> | <b>To be continue...</b>  | <b>5</b> |

## 1 Introduction

This document is suggested to be evolved in time and contains some general math things in which DL aspects is leveraging on or heuristically or non-heuristically.

Anybody welcome to append into it **pure mathematical** type things more or less relative to Deep Learning.

**It's only in the beginning phase and will be evolved in time.**

## 2 Differentiation

### 2.1 Introduction about Differentiation

This day's people talk about chain rule when prediction schema  $F(x; \theta) : X \rightarrow Y$  constructed as composition of several differentiable functions.

Leveraging in chain-rule for differentiable function allow compute gradient of function in systematic way, such that computation will be correct from math point of view, even numerically it can be various problems.

### 2.2 Differentiation Definition in Mathematics

Function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is differentiable if for arbitrarily  $\Delta x \in \mathbb{R}^n$  which element wise consist of some components (without any inter-dependency in them) we can have:

$$f(x + \Delta x) - f(x) = A\Delta x + \alpha(\Delta x)|\Delta x| \quad (1)$$

**Where:**

1.  $\alpha(\Delta x)$  is infinite small function i.e.  $\lim_{\Delta x \rightarrow 0} \alpha(\Delta x) = 0$
2.  $A$  is a rectangular matrix and its items should not depend on specific direction or values of  $\Delta x$
3.  $\alpha(\Delta x)$  is infinite dimensional function relative to  $\Delta x$

Differentiation is a local property of the function. If function is differentiable in all points in some set  $\chi$  then people say that function is differentiable in the whole set  $\chi$ .

### 2.3 Some Theorems in context of Differentiation

**Theorem 2.1** (Partial Derivatives). *If function  $f$  is differentiable in point then all partial derivatives of the function exist and equal to corresponded elements of the matrix  $A$ .*

To derive this theorem you should vary  $\Delta x = [0, 0, \dots, 0, dx_i, 0, \dots, 0]^T$ . This simple variation in input will allow eliminating every from  $A$ , except  $i$ -th column.

**Theorem 2.2** (Continuity Property). *If function  $f$  is differentiable in point then function is continuous in the point.*

**Theorem 2.3** (Differentiability from existing partial derivatives). *If function  $f$  has all partial derivatives in point  $\mathbf{a}$  and this partial derivative are continuous functions in a point  $\mathbf{a}$  then function is differentiable in point  $\mathbf{a}$ .*

**Theorem 2.4** (Differentiability of function composition). *Also known as the chain rule.*

1. *If function  $f$  is differentiable in point of it's domain  $\mathbf{a}$ .*
2. *If function  $g$  is differentiable in point of it's domain  $\mathbf{b}$ .*
3. *If  $\mathbf{b} = f(\mathbf{a})$*

*Then function  $z = g(f(x))$  is differentiable in points  $a$  and  $z' = g' \cdot f'$*

### 3 What is convolution in mathematics

#### 3.1 Convolution definition

I think you know that convolution is well-defined operation:

$$f * g(x) = \int_{-\infty}^{+\infty} f(y)g(x-y)dy \quad (2)$$

Sometimes limits of integration in (2) can be reduced to some subinterval of  $[-\infty, +\infty]$ .

#### 3.2 Convolution commutativity

$$\begin{aligned} f * g(x) &= \int_{-\infty}^{+\infty} f(y)g(x-y)dy = \\ &= \int_{+\infty}^{-\infty} f(x-z)g(z)(-dz) = \\ &= \int_{-\infty}^{+\infty} g(z)f(x-z)dz = g * f(x) \quad (3) \end{aligned}$$

So we have proved that  $f * g = g * f$

### 3.3 Time invariant property of convolution

$$\begin{aligned}\tau(f * g, b) &= f * g(x - b) = \int_{-\infty}^{+\infty} f(y)g((x - b) - y)dy = \\ &= \int_{-\infty}^{+\infty} f(y)g(x - b - y)dy = \\ &= \int_{-\infty}^{+\infty} f(y)\tau(g, b)(x - y)dy = \\ &= f * \tau(g, b)\end{aligned}\quad (4)$$

So we have proved that  $f * g = g * f$

### 3.4 Some Convolution Algebraic Properties

$$f * g = g * f \quad (5)$$

$$(f * g) * h = f * (g * h) \quad (6)$$

$$f * (g + h) = f * g + f * h \quad (7)$$

$$\frac{d^k(f * g)}{dx^k} = f * \frac{d^k g}{dx^k} \quad (8)$$

$$f * \delta(a) = f(a) \quad (9)$$

## 4 What is cross-correlation in math

### 4.1 Cross-Correlation definition

If consider functions  $\mathbb{R} \rightarrow \mathbb{R}$  then value of cross-correlation for them is **defined point-wise for specific value x** as the following

$$f \star g(x) = \int_{-\infty}^{+\infty} f(y)g(x + y)dy \quad (10)$$

If compared with convolution there is no "flipping and dragging" of convolution kernel in algebraic definition.

**What Neural Network community mean by Convolution is some form of Cross-Correlations**

## 4.2 Cross-Correlation connection with convolution

Let's define function  $f^- := f(-x)$ . This operation takes a function and reverse it w.r.t. to function value axis.

$$\begin{aligned}
 f \star g(x) &= \int_{-\infty}^{+\infty} f(y)g(x+y)dy = \\
 &\quad |y = -z, dy = -dz| = \\
 &\quad \int_{+\infty}^{-\infty} f(-z)g(x-z)(-dz) = \\
 &\quad \int_{-\infty}^{+\infty} f(-z)g(x-z)dz = \\
 &\quad \int_{-\infty}^{+\infty} f^-g(x-z)dz = f^- * g
 \end{aligned} \tag{11}$$

So we derived  $f \star g = f^- * g$

## 4.3 Some Cross-Correlation Properties not necessary to hold

Convolution behaves like multiplication, but as you see  $f \star g = f^- * g$  So due to that different algebraic properties is not necessary to hold. List of some properties which in general does not hold

$$f \star g = f^- * g = g * f^- \neq g^- * f = g \star f \tag{12}$$

$$(f \star g) \star h \neq f \star (g \star h) \tag{13}$$

## 5 To be continue...

Deep Learning models evolve in time and attack composition function representation from different ways and learning too.