

Physics Informed Neural Networks.

A complete explanation for beginners

Christian Ortiz

September 28, 2022

Abstract

In this tutorial we first discuss a simple toy model that describes a time-dependent one dimensional system. This toy model allows the reader to not only understand how physics informed neural networks are implemented but also to understand how the *autograd* and the *backpropagation* functions work. In particular, how the former handles second order derivatives and how the latter may be given considering a forward gradient descent. We then extend this model to consider a time-dependent three dimensional system in a deep learning architecture, and explore novel computational approaches.

1 Introduction

Partial derivatives are core ingredients in neural network architectures. However, they have been mostly overlooked... blah blah.

2 Toy Model

To have a solid understanding of physics informed neural networks, we first construct a simple toy model developing our own *autograd* and *backpropagation* functions. We start discussing these functions from a mathematical perspective and then we implement our physics informed architecture. In the following section, which handles a general model, we consider a similar approach, but from a computational perspective. In this sense, here we do not explicitly address the difference between the *forward* and *reverse* modes or the JVP and VJP approaches, instead we just give a general introduction.

2.1 System

The following neural network is considered:

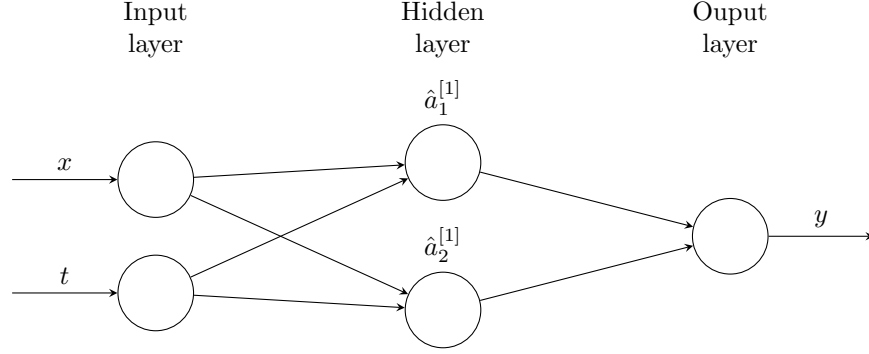


Figure 1: A neural network with 2 input parameters (x, t) , 1 hidden layer with 2 neurons, and 1 output (y) .

For simplicity, we set the following:

$$x = a_1^{[0]}, \quad (1)$$

$$t = a_2^{[0]}, \quad (2)$$

$$y = a_1^{[2]}. \quad (3)$$

Then we have,

$$a_1^{[1]} = g_1(z_1^{[1]}) = g_1(w_{11}^{[1]}a_1^{[0]} + w_{12}^{[1]}a_2^{[0]} + b_1^{[1]}), \quad (4)$$

$$a_2^{[1]} = g_1(z_2^{[1]}) = g_1(w_{21}^{[1]}a_1^{[0]} + w_{22}^{[1]}a_2^{[0]} + b_2^{[1]}), \quad (5)$$

$$a_1^{[2]} = g_2(z_1^{[2]}) = g_2(w_{11}^{[2]}a_1^{[1]} + w_{12}^{[2]}a_2^{[1]} + b_1^{[2]}). \quad (6)$$

$a_i^{[k]} = g_k(z_i^{[k]})$ is the activation function. $w_{ij}^{[k]}$ is the weight, and $b_i^{[k]}$ is the bias. k denotes the layer index and i (j) denotes the row (column) of the matrix. In

vectorial form the above expressions read,

$$\mathbf{a}^{[0]} = \begin{pmatrix} a_1^{[0]} \\ a_2^{[0]} \end{pmatrix}, \quad \mathbf{a}^{[1]} = \begin{pmatrix} a_1^{[1]} \\ a_2^{[1]} \end{pmatrix}, \quad a^{[2]} = a_1^{[2]}, \quad (7)$$

$$\mathbf{z}^{[1]} = \begin{pmatrix} z_1^{[1]} \\ z_2^{[1]} \end{pmatrix}, \quad z^{[2]} = z_1^{[2]}, \quad (8)$$

$$\mathbf{b}^{[1]} = \begin{pmatrix} b_1^{[1]} \\ b_2^{[1]} \end{pmatrix}, \quad b^{[2]} = b_1^{[2]}, \quad (9)$$

$$\mathbf{w}^{[1]} = \begin{pmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} \end{pmatrix}, \quad \mathbf{w}^{[2]} = \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix}, \quad (10)$$

where we have,

$$\mathbf{a}^{[1]} = g_1(\mathbf{z}^{[1]}) = g_1(\mathbf{w}^{[1]} \cdot \mathbf{a}^{[0]} + \mathbf{b}^{[1]}), \quad (11)$$

$$y = a^{[2]} = g_2(z^{[2]}) = g_2(\mathbf{w}^{[2]} \cdot \mathbf{a}^{[1]} + b^{[2]}). \quad (12)$$

Eq. (11) is equivalent to Eqs. (4)-(5) and Eq. (12) is equivalent to Eq. (6).

2.2 Autograd

Considering Eqs. (1)-(6) we have,

$$y = g_2 \left(w_{11}^{[2]} g_1 \left(w_{11}^{[1]} x + w_{12}^{[1]} t + b_1^{[1]} \right) + w_{12}^{[2]} g_1 \left(w_{21}^{[1]} x + w_{22}^{[1]} t + b_2^{[1]} \right) + b_1^{[2]} \right). \quad (13)$$

If we want to compute $\frac{\partial y}{\partial x}$ and $\frac{\partial y}{\partial t}$, in Eq. (13) it is enough to derive with respect to x and t ; however, this procedure is extensive. A better approach in multivariate calculus is to consider the Jacobian matrix. To do so, first, the chain rule differentiation gives us:

$$\frac{\partial y}{\partial \mathbf{a}^{[0]}} = \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial \mathbf{a}^{[0]}}. \quad (14)$$

Each term is given by,

$$\frac{\partial y}{\partial z^{[2]}} = g'_2(z^{[2]}), \quad (15)$$

$$\frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} = \begin{pmatrix} \frac{\partial z^{[2]}}{\partial a_1^{[1]}} & \frac{\partial z^{[2]}}{\partial a_2^{[1]}} \end{pmatrix}, \quad (16)$$

$$\frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} = \begin{pmatrix} \frac{\partial a_1^{[1]}}{\partial z_1^{[1]}} & \frac{\partial a_1^{[1]}}{\partial z_2^{[1]}} \\ \frac{\partial a_2^{[1]}}{\partial z_1^{[1]}} & \frac{\partial a_2^{[1]}}{\partial z_2^{[1]}} \end{pmatrix}, \quad (17)$$

$$\frac{\partial \mathbf{z}^{[1]}}{\partial \mathbf{a}^{[0]}} = \begin{pmatrix} \frac{\partial z_1^{[1]}}{\partial x} & \frac{\partial z_1^{[1]}}{\partial t} \\ \frac{\partial z_2^{[1]}}{\partial x} & \frac{\partial z_2^{[1]}}{\partial t} \end{pmatrix}. \quad (18)$$

In the above expressions, the matrices are the so called Jacobian matrices. The dimension of each Jacobian matrix is $n \times m$, where n (m) refers to the dimension of the numerator (denominator) in the partial derivative. To solve these expressions we consider Eqs. (4)-(6); therefore, we have,

$$\frac{\partial y}{\partial z^{[2]}} = g'_2(z^{[2]}), \quad (19)$$

$$\frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} = \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix}, \quad (20)$$

$$\frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} = \begin{pmatrix} g'_1(z_1^{[1]}) & 0 \\ 0 & g'_1(z_2^{[1]}) \end{pmatrix}, \quad (21)$$

$$\frac{\partial \mathbf{z}^{[1]}}{\partial \mathbf{a}^{[0]}} = \begin{pmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} \end{pmatrix}. \quad (22)$$

Consequently, $\frac{\partial y}{\partial \mathbf{a}^{[0]}}$ reads,

$$\frac{\partial y}{\partial \mathbf{a}^{[0]}} = g'_2(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g'_1(z_1^{[1]}) & 0 \\ 0 & g'_1(z_2^{[1]}) \end{pmatrix} \begin{pmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} \end{pmatrix}. \quad (23)$$

To solve for each component we consider directional unit vectors of the form $\mathbf{e}_x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{e}_t = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, then we have,

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial \mathbf{a}^{[0]}} \cdot \mathbf{e}_x, \quad (24)$$

$$\frac{\partial y}{\partial t} = \frac{\partial y}{\partial \mathbf{a}^{[0]}} \cdot \mathbf{e}_t, \quad (25)$$

which gives us,

$$\frac{\partial y}{\partial x} = g'_2(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g'_1(z_1^{[1]}) & 0 \\ 0 & g'_1(z_2^{[1]}) \end{pmatrix} \begin{pmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (26)$$

$$= g'_2(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g'_1(z_1^{[1]}) & 0 \\ 0 & g'_1(z_2^{[1]}) \end{pmatrix} \begin{pmatrix} w_{11}^{[1]} \\ w_{21}^{[1]} \end{pmatrix} \quad (27)$$

$$= g'_2(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g'_1(z_1^{[1]})w_{11}^{[1]} \\ g'_1(z_2^{[1]})w_{21}^{[1]} \end{pmatrix} \quad (28)$$

$$= g'_2(z^{[2]}) \begin{pmatrix} w_{11}^{[2]}g'_1(z_1^{[1]})w_{11}^{[1]} + w_{12}^{[2]}g'_1(z_2^{[1]})w_{21}^{[1]} \end{pmatrix}. \quad (29)$$

and

$$\frac{\partial y}{\partial t} = g'_2(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g'_1(z_1^{[1]}) & 0 \\ 0 & g'_1(z_2^{[1]}) \end{pmatrix} \begin{pmatrix} w_{11}^{[1]} & w_{12}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (30)$$

$$= g'_2(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g'_1(z_1^{[1]}) & 0 \\ 0 & g'_1(z_2^{[1]}) \end{pmatrix} \begin{pmatrix} w_{12}^{[1]} \\ w_{22}^{[1]} \end{pmatrix} \quad (31)$$

$$= g'_2(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g'_1(z_1^{[1]})w_{12}^{[1]} \\ g'_1(z_2^{[1]})w_{22}^{[1]} \end{pmatrix} \quad (32)$$

$$= g'_2(z^{[2]}) \begin{pmatrix} w_{11}^{[2]}g'_1(z_1^{[1]})w_{12}^{[1]} + w_{12}^{[2]}g'_1(z_2^{[1]})w_{22}^{[1]} \end{pmatrix}. \quad (33)$$

As you can see, in the above expressions we have multiplied the Jacobian matrices from right to left, this is called the *forward* mode, an approach to be discussed in more detail when studying the general model. Now let's consider the second derivative. We previously had,

$$\frac{\partial y}{\partial \mathbf{a}^{[0]}} = \begin{pmatrix} \frac{\partial y}{\partial x} & \frac{\partial y}{\partial t} \end{pmatrix}. \quad (34)$$

The second derivative reads,

$$\left(\frac{\partial}{\partial \mathbf{a}^{[0]}} \right)^T \cdot \left(\frac{\partial y}{\partial \mathbf{a}^{[0]}} \right) = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial t} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial y}{\partial x} & \frac{\partial y}{\partial t} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 y}{\partial x^2} & \frac{\partial^2 y}{\partial x \partial t} \\ \frac{\partial^2 y}{\partial t \partial x} & \frac{\partial^2 y}{\partial t^2} \end{pmatrix}. \quad (35)$$

Replacing Eq. (14) in Eq. (35) we have,

$$\left(\frac{\partial}{\partial \mathbf{a}^{[0]}}\right)^T \cdot \left(\frac{\partial y}{\partial \mathbf{a}^{[0]}}\right) = \left(\frac{\partial}{\partial \mathbf{a}^{[0]}}\right)^T \cdot \left(\frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial \mathbf{a}^{[0]}}\right) \quad (36)$$

$$= \left(\frac{\partial}{\partial t}\right) \cdot \left(\frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x} + \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t}\right) \quad (37)$$

$$= \left(\frac{\partial}{\partial x} \left(\frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x}\right) + \frac{\partial}{\partial x} \left(\frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t}\right) + \frac{\partial}{\partial t} \left(\frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x}\right) + \frac{\partial}{\partial t} \left(\frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t}\right)\right). \quad (38)$$

In the above expression, only the terms that depend on z , intrinsically depend on x and t . As an example let's consider $\frac{\partial^2 y}{\partial x^2}$, then we have:

$$\frac{\partial^2 y}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x}\right) \quad (39)$$

$$= \left(\frac{\partial}{\partial x} \frac{\partial y}{\partial z^{[2]}}\right) \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x} + \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \left(\frac{\partial}{\partial x} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}}\right) \frac{\partial \mathbf{z}^{[1]}}{\partial x}. \quad (40)$$

The terms in parenthesis require a chain rule differentiation. Each term reads,

$$\frac{\partial}{\partial x} \frac{\partial y}{\partial z^{[2]}} = \left(\frac{\partial}{\partial z^{[2]}} \frac{\partial y}{\partial z^{[2]}}\right) \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x}, \quad (41)$$

$$\frac{\partial}{\partial x} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} = \left(\frac{\partial}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}}\right) \frac{\partial \mathbf{z}^{[1]}}{\partial x}. \quad (42)$$

Replacing Eqs. (41) and (42) in Eq. (40) gives us,

$$\frac{\partial^2 y}{\partial x^2} = \left(\frac{\partial^2 y}{\partial z^{2[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x}\right) \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x} + \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \left(\frac{\partial^2 \mathbf{a}^{[1]}}{\partial \mathbf{z}^{2[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x}\right) \frac{\partial \mathbf{z}^{[1]}}{\partial x}. \quad (43)$$

The first order terms in Eq. (43) are already solved in Eqs. (19)-(22). The second order derivatives are solved as follow:

$$\frac{\partial^2 y}{\partial z^{2[2]}} = \frac{\partial}{\partial z^{[2]}} \frac{\partial y}{\partial z^{[2]}} = \frac{\partial}{\partial z^{[2]}} g'_2(z^{[2]}) = g''_2(z^{[2]}) \quad (44)$$

Notice that in Eq. (44), the solution is just the second derivative of the activation function, i.e., $g''_2(z^{[2]})$. The solution of $\frac{\partial^2 \mathbf{a}^{[1]}}{\partial \mathbf{z}^{2[1]}}$ is a bit more interesting. We have,

$$\frac{\partial}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} = \frac{\partial}{\partial \mathbf{z}^{[1]}} \begin{pmatrix} g_1'(z_1^{[1]}) & 0 \\ 0 & g_1'(z_2^{[1]}) \end{pmatrix} \quad (45)$$

$$\equiv \frac{\partial}{\partial \mathbf{z}^{[1]}} \begin{pmatrix} g_1'(z_1^{[1]}) \\ 0 \\ 0 \\ g_1'(z_2^{[1]}) \end{pmatrix} \quad (46)$$

$$\equiv \begin{pmatrix} \frac{\partial g_1'(z_1^{[1]})}{\partial z_1^{[1]}} & \frac{\partial g_1'(z_1^{[1]})}{\partial z_2^{[1]}} \\ 0 & 0 \\ 0 & 0 \\ \frac{\partial g_1'(z_2^{[1]})}{\partial z_1^{[1]}} & \frac{\partial g_1'(z_2^{[1]})}{\partial z_2^{[1]}} \end{pmatrix} \quad (47)$$

$$\equiv \begin{pmatrix} g_1''(z_1^{[1]}) & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & g_1''(z_2^{[1]}) \end{pmatrix}. \quad (48)$$

In Eq. (48), to have a better visualization, we have reshaped our Jacobian matrix, i.e., $\frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}}$ is a 2×2 matrix but we express it as a 4×1 matrix. To fully understand that reshaping does not affect our final result, let's solve the second parenthesis in Eq. (43), i.e.,

$$\frac{\partial}{\partial x} \left(\frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \right) = \frac{\partial^2 \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]2}} \frac{\partial \mathbf{z}^{[1]}}{\partial x} = \begin{pmatrix} g_1''(z_1^{[1]}) & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & g_1''(z_2^{[1]}) \end{pmatrix} \begin{pmatrix} w_{11}^{[1]} \\ w_{21}^{[1]} \end{pmatrix} \quad (49)$$

$$= \begin{pmatrix} g_1''(z_1^{[1]})w_{11}^{[1]} \\ 0 \\ 0 \\ g_1''(z_2^{[1]})w_{21}^{[1]} \end{pmatrix} \quad (50)$$

$$\equiv \begin{pmatrix} g_1''(z_1^{[1]})w_{11}^{[1]} & 0 \\ 0 & g_1''(z_2^{[1]})w_{21}^{[1]} \end{pmatrix}. \quad (51)$$

As you can see in Eq. (51), in order to preserve the correct dimensionality, we have reshaped back our Jacobian matrix. Replacing Eqs. (19)-(22), (44), and (51) in Eq. (43) gives us,

$$\frac{\partial^2 y}{\partial x^2} = \left(\frac{\partial^2 y}{\partial z^{2[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x} \right) \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x} + \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \left(\frac{\partial^2 \mathbf{a}^{[1]}}{\partial \mathbf{z}^{2[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x} \right) \frac{\partial \mathbf{z}^{[1]}}{\partial x}. \quad (52)$$

$$= \left(g_2''(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1'(z_1^{[1]})w_{11}^{[1]} \\ g_1'(z_2^{[1]})w_{21}^{[1]} \end{pmatrix} \right) \cdot \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1'(z_1^{[1]})w_{11}^{[1]} \\ g_1'(z_2^{[1]})w_{21}^{[1]} \end{pmatrix} \quad (53)$$

$$+ g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1''(z_1^{[1]})w_{11}^{2[1]} \\ g_1''(z_2^{[1]})w_{21}^{2[1]} \end{pmatrix}. \quad (54)$$

Considering similar steps we can compute $\frac{\partial^2 y}{\partial t^2}$, $\frac{\partial^2 y}{\partial x \partial t}$, and $\frac{\partial^2 y}{\partial t \partial x}$. We therefore have,

$$\frac{\partial^2 y}{\partial t^2} = \left(\frac{\partial^2 y}{\partial z^{2[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t} \right) \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t} + \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \left(\frac{\partial^2 \mathbf{a}^{[1]}}{\partial \mathbf{z}^{2[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t} \right) \frac{\partial \mathbf{z}^{[1]}}{\partial t}. \quad (55)$$

$$= \left(g_2''(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1'(z_1^{[1]})w_{12}^{[1]} \\ g_1'(z_2^{[1]})w_{22}^{[1]} \end{pmatrix} \right) \cdot \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1'(z_1^{[1]})w_{12}^{[1]} \\ g_1'(z_2^{[1]})w_{22}^{[1]} \end{pmatrix} \quad (56)$$

$$+ g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1''(z_1^{[1]})w_{12}^{2[1]} \\ g_1''(z_2^{[1]})w_{22}^{2[1]} \end{pmatrix}, \quad (57)$$

$$\frac{\partial^2 y}{\partial t \partial x} = \left(\frac{\partial^2 y}{\partial z^{2[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t} \right) \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x} + \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \left(\frac{\partial^2 \mathbf{a}^{[1]}}{\partial \mathbf{z}^{2[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t} \right) \frac{\partial \mathbf{z}^{[1]}}{\partial x}. \quad (58)$$

$$= \left(g_2''(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1'(z_1^{[1]})w_{12}^{[1]} \\ g_1'(z_2^{[1]})w_{22}^{[1]} \end{pmatrix} \right) \cdot \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1'(z_1^{[1]})w_{11}^{[1]} \\ g_1'(z_2^{[1]})w_{21}^{[1]} \end{pmatrix} \quad (59)$$

$$+ g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1''(z_1^{[1]})w_{12}^{[1]}w_{11}^{[1]} \\ g_1''(z_2^{[1]})w_{22}^{[1]}w_{21}^{[1]} \end{pmatrix}, \quad (60)$$

$$\frac{\partial^2 y}{\partial t \partial x} = \left(\frac{\partial^2 y}{\partial z^{[2]^2}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x} \right) \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t} + \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \left(\frac{\partial^2 \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[2]^2}} \frac{\partial \mathbf{z}^{[1]}}{\partial x} \right) \frac{\partial \mathbf{z}^{[1]}}{\partial t}. \quad (61)$$

$$= \left(g_2''(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1'(z_1^{[1]}) w_{11}^{[1]} \\ g_1'(z_2^{[1]}) w_{21}^{[1]} \end{pmatrix} \right) \cdot \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1'(z_1^{[1]}) w_{12}^{[1]} \\ g_1'(z_2^{[1]}) w_{22}^{[1]} \end{pmatrix} \quad (62)$$

$$+ g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1''(z_1^{[1]}) w_{11}^{[1]} w_{12}^{[1]} \\ g_1''(z_2^{[1]}) w_{21}^{[1]} w_{22}^{[1]} \end{pmatrix}. \quad (63)$$

Notice that $\frac{\partial^2 y}{\partial x \partial t} \equiv \frac{\partial^2 y}{\partial t \partial x}$.

2.3 Backpropagation

Backpropagation of PDE is

$$\frac{\partial y}{\partial t} = \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t} \quad (64)$$

$$\frac{\mathcal{J}_{pde}}{\partial \mathbf{w}^{[2]}} = \frac{1}{m_p} (2p_n) \left(\frac{\partial}{\partial \mathbf{w}^{[2]}} \left(\frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t} \right) - \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{w}^{[2]}} + 2y \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{w}^{[2]}} \right) \quad (65)$$

$$\left(\frac{\partial^2 y}{\partial z^{[2]^2}} \frac{\partial z^{[2]}}{\partial \mathbf{w}^{[2]}} \right) \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t} + \frac{\partial y}{\partial z^{[2]}} \left(\frac{\partial}{\partial \mathbf{w}^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \right) \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t} \quad (66)$$

Backpropagation, properly speaking, is an algorithm for training feedforward neural networks, or better say, is a reverse-mode automatic differentiation algorithm. Here we just pay attention to the mathematical steps and disregard the computational solution, which is better understood when discussing the general model.

In the previous section, we derived the first and second derivatives of y with respect to the input parameters (x, t) . In this section we derive the first derivative of \mathcal{J} with respect to the weights and the biases. To do so, we have,

$$\mathcal{J} = \frac{1}{m} \sum_i^m \mathcal{L}(y^{(i)}, u^{(i)}), \quad (67)$$

where \mathcal{J} is the cost function, \mathcal{L} is the loss function. y is the predicted output from our neural network, u is the real value, m represents the number of samples, and upperindex i refers to the i th sample. In linear regression problems the most commonly used loss function is the least squared error. Here we keep it general

in order to study other types of loss functions. Replacing Eq. (13) in Eq. (67) gives us,

$$\mathcal{J} = \frac{1}{m} \sum_i^m \mathcal{L} \left(g_2 \left(w_{11}^{[2]} g_1 \left(w_{11}^{[1]} x^{(i)} + w_{12}^{[1]} t^{(i)} + b_1^{[1]} \right) + w_{12}^{[2]} g_1 \left(w_{21}^{[1]} x^{(i)} + w_{22}^{[1]} t^{(i)} + b_2^{[1]} \right) + b_1^{[2]} \right), u^{(i)} \right). \quad (68)$$

To compute the derivatives with respect to the weights and biases, in Eq. (68) it is enough to derive J with respect to $w_{11}^{[2]}$, $w_{12}^{[2]}$, $w_{11}^{[1]}$, $w_{12}^{[1]}$, $w_{21}^{[1]}$, $w_{22}^{[1]}$, $b_1^{[1]}$, and $b_1^{[2]}$; however, this procedure is extensive. A better approach in multivariate calculus is to consider the Jacobian matrix. To start with, we will consider the case where each iteration is given with $m = 1$ (the case $m > 1$ is implemented when discussing the general model), then $\mathcal{J} = \mathcal{L}(y, u)$ and the chain rule differentiation gives us

$$\frac{\partial J}{\partial \mathbf{w}^{[2]}} = \frac{\partial \mathcal{J}}{\partial y} \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{w}^{[2]}}, \quad (69)$$

$$\frac{\partial J}{\partial b^{[2]}} = \frac{\partial \mathcal{J}}{\partial y} \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial b^{[2]}}, \quad (70)$$

$$\frac{\partial J}{\partial \mathbf{w}^{[1]}} = \frac{\partial \mathcal{J}}{\partial y} \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial \mathbf{w}^{[1]}}. \quad (71)$$

$$\frac{\partial J}{\partial \mathbf{b}^{[1]}} = \frac{\partial \mathcal{J}}{\partial y} \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial \mathbf{b}^{[1]}}. \quad (72)$$

In Eqs. (69)-(72), The components read,

$$\frac{\partial \mathcal{J}}{\partial y} = \frac{\partial \mathcal{L}(y, u)}{\partial y} = \mathcal{L}'(y, u), \quad (73)$$

$$\frac{\partial y}{\partial z^{[2]}} = g'_2(z^{[2]}), \quad (74)$$

$$\frac{\partial z^{[2]}}{\partial \mathbf{w}^{[2]}} = \begin{pmatrix} \frac{\partial z^{[2]}}{\partial w_{11}^{[2]}} & \frac{\partial z^{[2]}}{\partial w_{12}^{[2]}} \end{pmatrix} = \begin{pmatrix} a_1^{[1]} & a_2^{[1]} \end{pmatrix}, \quad (75)$$

$$\frac{\partial z^{[2]}}{\partial b^{[2]}} = 1 \quad (76)$$

$$\frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} = \begin{pmatrix} \frac{\partial z^{[2]}}{\partial a_1^{[1]}} & \frac{\partial z^{[2]}}{\partial a_2^{[1]}} \end{pmatrix} = \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \quad (77)$$

$$\frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} = \begin{pmatrix} \frac{\partial a_1^{[1]}}{\partial z_1^{[1]}} & \frac{\partial a_1^{[1]}}{\partial z_2^{[1]}} \\ \frac{\partial a_2^{[1]}}{\partial z_1^{[1]}} & \frac{\partial a_2^{[1]}}{\partial z_2^{[1]}} \end{pmatrix} = \begin{pmatrix} g'_1(z_1^{[1]}) & 0 \\ 0 & g'_1(z_2^{[1]}) \end{pmatrix} \quad (78)$$

$$\frac{\partial \mathbf{z}^{[1]}}{\partial \mathbf{w}^{[1]}} = \begin{pmatrix} \frac{\partial z_1^{[1]}}{\partial w_{11}^{[1]}} & \frac{\partial z_1^{[1]}}{\partial w_{12}^{[1]}} & \frac{\partial z_1^{[1]}}{\partial w_{21}^{[1]}} & \frac{\partial z_1^{[1]}}{\partial w_{22}^{[1]}} \\ \frac{\partial z_2^{[1]}}{\partial w_{11}^{[1]}} & \frac{\partial z_2^{[1]}}{\partial w_{12}^{[1]}} & \frac{\partial z_2^{[1]}}{\partial w_{21}^{[1]}} & \frac{\partial z_2^{[1]}}{\partial w_{22}^{[1]}} \end{pmatrix} = \begin{pmatrix} x & t & 0 & 0 \\ 0 & 0 & x & t \end{pmatrix} \quad (79)$$

$$\frac{\partial \mathbf{z}^{[1]}}{\partial \mathbf{b}^{[1]}} = \begin{pmatrix} \frac{\partial z_1^{[1]}}{\partial b_1^{[1]}} & \frac{\partial z_1^{[1]}}{\partial b_2^{[1]}} \\ \frac{\partial z_2^{[1]}}{\partial b_1^{[1]}} & \frac{\partial z_2^{[1]}}{\partial b_2^{[1]}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (80)$$

Then, replacing these expression back in Eqs. (69)-(72) we have,

$$\frac{\partial J}{\partial \mathbf{w}^{[2]}} = \mathcal{L}'(y, u) g_2'(z^{[2]}) \begin{pmatrix} a_1^{[1]} & a_2^{[1]} \end{pmatrix}, \quad (81)$$

$$\frac{\partial J}{\partial b^{[2]}} = \mathcal{L}'(y, u) g_2'(z^{[2]}), \quad (82)$$

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{w}^{[1]}} &= \mathcal{L}'(y, u) g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1'(z_1^{[1]}) & 0 \\ 0 & g_1'(z_2^{[1]}) \end{pmatrix} \begin{pmatrix} x & t & 0 & 0 \\ 0 & 0 & x & t \end{pmatrix} \\ &= \mathcal{L}'(y, u) g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1'(z_1^{[1]})x & g_1'(z_1^{[1]})t & 0 & 0 \\ 0 & 0 & g_1'(z_2^{[1]})x & g_1'(z_2^{[1]})t \end{pmatrix} \\ &= \mathcal{L}'(y, u) g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]})x & w_{11}^{[2]} g_1'(z_1^{[1]})t & w_{12}^{[2]} g_1'(z_2^{[1]})x & w_{12}^{[2]} g_1'(z_2^{[1]})t \end{pmatrix} \\ &\equiv \mathcal{L}'(y, u) g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]})x & w_{11}^{[2]} g_1'(z_1^{[1]})t \\ w_{12}^{[2]} g_1'(z_2^{[1]})x & w_{12}^{[2]} g_1'(z_2^{[1]})t \end{pmatrix} \end{aligned} \quad (83)$$

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{b}^{[1]}} &= \mathcal{L}'(y, u) g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1'(z_1^{[1]}) & 0 \\ 0 & g_1'(z_2^{[1]}) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \mathcal{L}'(y, u) g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} & w_{12}^{[2]} \end{pmatrix} \begin{pmatrix} g_1'(z_1^{[1]}) & 0 \\ 0 & g_1'(z_2^{[1]}) \end{pmatrix} \\ &= \mathcal{L}'(y, u) g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]}) & w_{12}^{[2]} g_1'(z_2^{[1]}) \end{pmatrix} \\ &\equiv \mathcal{L}'(y, u) g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]}) \\ w_{12}^{[2]} g_1'(z_2^{[1]}) \end{pmatrix} \end{aligned} \quad (84)$$

Similar to the previous section, in Eqs. (83)-(84) we have reshaped (and reshaped back) our Jacobian matrices to preserve the dimensionality of the derivative terms. What follows in gradient descent is to update the parameters considering,

$$\mathbf{w}^{[2]} = \mathbf{w}^{[2]} - \alpha \frac{\partial J}{\partial \mathbf{w}^{[2]}}, \quad (85)$$

$$b^{[2]} = b^{[2]} - \alpha \frac{\partial J}{\partial b^{[2]}}, \quad (86)$$

$$\mathbf{w}^{[1]} = \mathbf{w}^{[1]} - \alpha \frac{\partial J}{\partial \mathbf{w}^{[1]}}, \quad (87)$$

$$\mathbf{b}^{[1]} = \mathbf{b}^{[1]} - \alpha \frac{\partial J}{\partial \mathbf{b}^{[1]}}, \quad (88)$$

where α is the learning rate. Since we have taken the case $m = 1$ for each iteration, this is referred to as the stochastic gradient descent.

To end the discussion of this section, let's pay attention to the following derivatives,

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial x}, \quad \frac{\partial y}{\partial t} = \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial t}. \quad (89)$$

$$\frac{\partial J}{\partial \mathbf{w}^{[1]}} = \frac{\partial \mathcal{J}}{\partial y} \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial \mathbf{w}^{[1]}}. \quad \frac{\partial J}{\partial \mathbf{b}^{[1]}} = \frac{\partial \mathcal{J}}{\partial y} \frac{\partial y}{\partial z^{[2]}} \frac{\partial z^{[2]}}{\partial \mathbf{a}^{[1]}} \frac{\partial \mathbf{a}^{[1]}}{\partial \mathbf{z}^{[1]}} \frac{\partial \mathbf{z}^{[1]}}{\partial \mathbf{b}^{[1]}}. \quad (90)$$

Taking into account that in this example $\frac{\partial \mathcal{J}}{\partial y}$ is a real number, then all the expressions given above are identical except for the last terms. From a computational point of view, it means that we just need to compute the set of partial derivatives once and duplicate it in all the other expressions.

2.4 Forward gradient descent

In this section let's focus on a novel approach to perform gradient descent. In contrast to backpropagation, here the gradients are obtained in the forward mode, but forward gradient descent is not just a forward automatic differentiation, it is an stochastic approach that uses directional vectors. Here we focus on the mathematical steps, the advantages from a computational perspective are discussed when studying the general model. To start with, let's consider the results from the previous section. We have,

$$\frac{\partial J}{\partial \mathbf{w}^{[2]}} = \mathcal{L}' g_2'(z^{[2]}) \begin{pmatrix} a_1^{[1]} & a_2^{[1]} \end{pmatrix}, \quad (91)$$

$$\frac{\partial J}{\partial b^{[2]}} = \mathcal{L}' g_2'(z^{[2]}), \quad (92)$$

$$\frac{\partial J}{\partial \mathbf{w}^{[1]}} = \mathcal{L}' g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]}) x & w_{11}^{[2]} g_1'(z_1^{[1]}) t & w_{12}^{[2]} g_1'(z_2^{[1]}) x & w_{12}^{[2]} g_1'(z_2^{[1]}) t \end{pmatrix}, \quad (93)$$

$$\frac{\partial J}{\partial \mathbf{b}^{[1]}} = \mathcal{L}' g_2'(z^{[2]}) \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]}) & w_{12}^{[2]} g_1'(z_2^{[1]}) \end{pmatrix}. \quad (94)$$

Notice that in the above expressions we have not reshaped back the Jacobian matrices. To compute each term we consider the basis vectors; therefore have,

$$\frac{\partial J}{\partial w_{11}^{[2]}} = \mathcal{L}' g_2' \begin{pmatrix} a_1^{[1]} & a_2^{[1]} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mathcal{L}' g_2' a_1^{[1]}, \quad (95)$$

$$\frac{\partial J}{\partial w_{12}^{[2]}} = \mathcal{L}' g_2' \begin{pmatrix} a_1^{[1]} & a_2^{[1]} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \mathcal{L}' g_2' a_2^{[1]}, \quad (96)$$

$$\frac{\partial J}{\partial b^{[2]}} = \mathcal{L}' g_2', \quad (97)$$

$$\begin{aligned} \frac{\partial J}{\partial w_{11}^{[1]}} &= \mathcal{L}' g_2' \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]})x & w_{11}^{[2]} g_1'(z_1^{[1]})t & w_{12}^{[2]} g_1'(z_2^{[1]})x & w_{12}^{[2]} g_1'(z_2^{[1]})t \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\ &= \mathcal{L}' g_2' w_{11}^{[2]} g_1'(z_1^{[1]})x, \end{aligned} \quad (98)$$

$$\begin{aligned} \frac{\partial J}{\partial w_{12}^{[1]}} &= \mathcal{L}' g_2' \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]})x & w_{11}^{[2]} g_1'(z_1^{[1]})t & w_{12}^{[2]} g_1'(z_2^{[1]})x & w_{12}^{[2]} g_1'(z_2^{[1]})t \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \\ &= \mathcal{L}' g_2' w_{11}^{[2]} g_1'(z_1^{[1]})t, \end{aligned} \quad (99)$$

$$\begin{aligned} \frac{\partial J}{\partial w_{21}^{[1]}} &= \mathcal{L}' g_2' \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]})x & w_{11}^{[2]} g_1'(z_1^{[1]})t & w_{12}^{[2]} g_1'(z_2^{[1]})x & w_{12}^{[2]} g_1'(z_2^{[1]})t \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \\ &= \mathcal{L}' g_2' w_{12}^{[2]} g_1'(z_2^{[1]})x, \end{aligned} \quad (100)$$

$$\begin{aligned} \frac{\partial J}{\partial w_{22}^{[1]}} &= \mathcal{L}' g_2' \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]})x & w_{11}^{[2]} g_1'(z_1^{[1]})t & w_{12}^{[2]} g_1'(z_2^{[1]})x & w_{12}^{[2]} g_1'(z_2^{[1]})t \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \\ &= \mathcal{L}' g_2' w_{12}^{[2]} g_1'(z_2^{[1]})t \end{aligned} \quad (101)$$

$$\frac{\partial J}{\partial b_1^{[1]}} = \mathcal{L}' g_2' \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]}) & w_{12}^{[2]} g_1'(z_2^{[1]}) \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \mathcal{L}' g_2' w_{11}^{[2]} g_1'(z_1^{[1]}), \quad (102)$$

$$\frac{\partial J}{\partial b_2^{[1]}} = \mathcal{L}' g_2' \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]}) & w_{12}^{[2]} g_1'(z_2^{[1]}) \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \mathcal{L}' g_2' w_{12}^{[2]} g_1'(z_2^{[1]}), \quad (103)$$

The above expressions show a conventional way to compute the gradients in forward mode; however, instead of choosing the basis vectors we can consider directional vectors. To do this, we first define the following functions,

$$\mathbf{g}_w^{[2]} = \left(\frac{\partial J}{\partial \mathbf{w}^{[2]}} \cdot \mathbf{v}_w^{[2]} \right) \mathbf{v}_w^{[2]}, \quad (104)$$

$$g_b^{[2]} = \left(\frac{\partial J}{\partial b^{[2]}} v_b^{[2]} \right) v_b^{[2]}, \quad (105)$$

$$\mathbf{g}_w^{[1]} = \left(\frac{\partial J}{\partial \mathbf{w}^{[1]}} \cdot \mathbf{v}_w^{[1]} \right) \mathbf{v}_w^{[1]}, \quad (106)$$

$$\mathbf{g}_b^{[1]} = \left(\frac{\partial J}{\partial b^{[1]}} \cdot \mathbf{v}_b^{[1]} \right) \mathbf{v}_b^{[1]}. \quad (107)$$

$\mathbf{v}_w^{[2]} = \begin{pmatrix} v_{w1}^{[2]} & v_{w2}^{[2]} \end{pmatrix}^T$, $\mathbf{v}_w^{[1]} = \begin{pmatrix} v_{w1}^{[1]} & v_{w2}^{[1]} & v_{w3}^{[1]} & v_{w4}^{[1]} \end{pmatrix}^T$, and $\mathbf{v}_b^{[1]} = \begin{pmatrix} v_{b1}^{[1]} & v_{b2}^{[1]} \end{pmatrix}^T$ are called directional vectors. For the particular case of $g_b^{[2]}$ we consider instead the scalar $v_b^{[2]}$. Replacing Eqs. (91)-(94) in Eqs. (104)-(107) we have,

$$\mathbf{g}_w^{[2]} = \left[\mathcal{L}' g_2' \begin{pmatrix} a_1^{[1]} & a_2^{[1]} \end{pmatrix} \begin{pmatrix} v_{w1}^{[2]} \\ v_{w2}^{[2]} \end{pmatrix} \right] \begin{pmatrix} v_{w1}^{[2]} \\ v_{w2}^{[2]} \end{pmatrix} = \mathcal{L}' g_2' \begin{pmatrix} a_1^{[1]} v_{w1}^{[2]} + a_2^{[1]} v_{w1}^{[2]} v_{w2}^{[2]} \\ a_2^{[1]} v_{w2}^{[2]} + a_1^{[1]} v_{w1}^{[2]} v_{w2}^{[2]} \end{pmatrix}, \quad (108)$$

$$g_b^{[2]} = \mathcal{L}' g_2' v_b^{[2]}, \quad (109)$$

$$\begin{aligned} \mathbf{g}_w^{[1]} &= \left[\mathcal{L}' g_2' \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]}) x & w_{11}^{[2]} g_1'(z_1^{[1]}) t & w_{12}^{[2]} g_1'(z_2^{[1]}) x & w_{12}^{[2]} g_1'(z_2^{[1]}) t \end{pmatrix} \begin{pmatrix} v_{w1}^{[1]} \\ v_{w2}^{[1]} \\ v_{w3}^{[1]} \\ v_{w4}^{[1]} \end{pmatrix} \right] \begin{pmatrix} v_{w1}^{[1]} \\ v_{w2}^{[1]} \\ v_{w3}^{[1]} \\ v_{w4}^{[1]} \end{pmatrix} \\ &= \mathcal{L}' g_2' \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]}) x v_{w1}^{[1]} + w_{11}^{[2]} g_1'(z_1^{[1]}) t v_{w1}^{[1]} v_{w2}^{[1]} + w_{12}^{[2]} g_1'(z_2^{[1]}) x v_{w1}^{[1]} v_{w3}^{[1]} + w_{12}^{[2]} g_1'(z_2^{[1]}) t v_{w1}^{[1]} v_{w4}^{[1]} \\ w_{11}^{[2]} g_1'(z_1^{[1]}) t v_{w2}^{[1]} + w_{11}^{[2]} g_1'(z_1^{[1]}) x v_{w1}^{[1]} v_{w2}^{[1]} + w_{12}^{[2]} g_1'(z_2^{[1]}) x v_{w2}^{[1]} v_{w3}^{[1]} + w_{12}^{[2]} g_1'(z_2^{[1]}) t v_{w2}^{[1]} v_{w4}^{[1]} \\ w_{12}^{[2]} g_1'(z_2^{[1]}) x v_{w3}^{[1]} + w_{11}^{[2]} g_1'(z_1^{[1]}) x v_{w1}^{[1]} v_{w3}^{[1]} + w_{11}^{[2]} g_1'(z_1^{[1]}) t v_{w2}^{[1]} v_{w3}^{[1]} + w_{12}^{[2]} g_1'(z_2^{[1]}) t v_{w3}^{[1]} v_{w4}^{[1]} \\ w_{12}^{[2]} g_1'(z_2^{[1]}) t v_{w4}^{[1]} + w_{11}^{[2]} g_1'(z_1^{[1]}) x v_{w1}^{[1]} v_{w4}^{[1]} + w_{11}^{[2]} g_1'(z_1^{[1]}) t v_{w2}^{[1]} v_{w4}^{[1]} + w_{12}^{[2]} g_1'(z_2^{[1]}) x v_{w3}^{[1]} v_{w4}^{[1]} \end{pmatrix} \end{aligned} \quad (110)$$

$$\begin{aligned} \mathbf{g}_b^{[1]} &= \left[\mathcal{L}' g_2' \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]}) & w_{12}^{[2]} g_1'(z_2^{[1]}) \end{pmatrix} \begin{pmatrix} v_{b1}^{[1]} \\ v_{b2}^{[1]} \end{pmatrix} \right] \begin{pmatrix} v_{b1}^{[1]} \\ v_{b2}^{[1]} \end{pmatrix} \\ &= \mathcal{L}' g_2' \begin{pmatrix} w_{11}^{[2]} g_1'(z_1^{[1]}) v_{b1}^{[1]} + w_{12}^{[2]} g_1'(z_2^{[1]}) v_{b1}^{[1]} v_{b2}^{[1]} \\ w_{12}^{[2]} g_1'(z_2^{[1]}) v_{b2}^{[1]} + w_{11}^{[2]} g_1'(z_1^{[1]}) v_{b1}^{[1]} v_{b2}^{[1]} \end{pmatrix}. \end{aligned} \quad (111)$$

These new functions are also cost function gradients; therefore, to update the weights and biases we consider,

$$\mathbf{w}^{[2](i+1)} := \mathbf{w}^{[2](i)} - \alpha \mathbf{g}_w^{[2](i)}, \quad (112)$$

$$b^{[2](i+1)} := b^{[2](i)} - \alpha g_b^{[2](i)}, \quad (113)$$

$$\mathbf{w}^{[1](i+1)} := \mathbf{w}^{[1](i)} - \alpha \mathbf{g}_w^{[1](i)}, \quad (114)$$

$$\mathbf{b}^{[1](i+1)} := \mathbf{b}^{[1](i)} - \alpha \mathbf{g}_b^{[1](i)}. \quad (115)$$

In the above expressions upperindex i refers to the i th iteration. In stochastic (forward) gradient descent, which is the case that we discuss here, this is also the i th sample. As you may notice, except for the last term, the solution is quite similar to the conventional update given in Eqs. (85)-(88). To better understand how this solution also reduces the cost function, let's consider first $b^{[2]}$, we then have,

$$b^{[2](i+1)} := b^{[2](i)} - \alpha \frac{\partial J}{\partial b^{[2](i)}} \quad (116)$$

$$b^{[2](i+1)} := b^{[2](i)} - \alpha g_b^{[2](i)} \quad (117)$$

In Eqs. (116)-(117) we are comparing the conventional update of $b^{[2]}$ with the forward gradient descent method. If we consider m iterations we would have,

$$b^{[2](m+1)} := b^{[2](1)} - \alpha \sum_{k=1}^m \frac{\partial J}{\partial b^{[2](k)}} \quad (118)$$

$$b^{[2](m+1)} := b^{[2](1)} - \alpha \sum_{k=1}^m g_b^{[2](k)} \quad (119)$$

Replacing Eq. (97) and (109) in Eqs. (118) and (119), respectively, we have,

$$b^{[2](m+1)} := b^{[2](1)} - \alpha \sum_{k=1}^m \mathcal{L}'^{(k)} g_2'^{(k)}, \quad (120)$$

$$b^{[2](m+1)} := b^{[2](1)} - \alpha \sum_{k=1}^m \mathcal{L}'^{(k)} g_2'^{(k)} v_b^{2[2](k)}. \quad (121)$$

Notice that if $v_b^{2[2](k)} = 1$ for all k 's, then Eqs. (120)-(121) would be identical. This gives us a hint of what we are looking for. To do so, we consider now that our directional vector components are random variables from a normal distribution with zero mean and unit variance. What does it mean? It means that every time we choose a random number, $v_b^{[2]}$, the probability of choosing a value near zero is high. In fact, 95% of those numbers should lie in the interval $[-2, 2]$. In this sense we could say that $v_b^{[2]} \sim 0$ is likely to happen. This is the expected value and we call it $\mathbb{E}[v_b^{[2]}] = 0$. But what can we say about the

squared of the number, $v_b^{2[2]}$. To find the answer, let's proceed with a proper definition where the probability density of the normal distribution with zero mean and unit variance reads,

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (122)$$

Then the expected value in this case is,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad (123)$$

which is an odd function; therefore, we straightforward get $\mathbb{E}[X] = \mathbb{E}[v_b^{2[2]}] = 0$. Now let's calculate the expected value of X^2 . We have,

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (124)$$

To solve this let's perform a trick. First we know that

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} = \sqrt{2\pi}. \quad (125)$$

Then, considering $u = e^{-\frac{x^2}{2}}$ and $v = x$, integration by parts gives us,

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = x e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -x^2 e^{-\frac{x^2}{2}} dx \quad (126)$$

$$\sqrt{2\pi} = x e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx. \quad (127)$$

In Eq. (127), the first term on the right hand side is zero, then reordering we have,

$$\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = 1, \quad (128)$$

$$\mathbb{E}[X^2] = \mathbb{E}[v_b^{2[2]}] = 1. \quad (129)$$

Consequently if we expand Eq. (121) we will have,

$$b^{[2](m+1)} := b^{[2](1)} - \alpha \left(\mathcal{L}'^{(1)} g_2'^{(1)} v_b^{2[2](1)} + \mathcal{L}'^{(2)} g_2'^{(2)} v_b^{22} + \dots + \mathcal{L}'^{(m)} g_2'^{(m)} v_b^{2[2](m)} \right). \quad (130)$$

In this expansion, each $v_b^{2[2](i)} \sim 1$ and therefore the expression resembles to some extent to the conventional gradient descent. Note however that this is

an stochastic procedure and likely not as efficient as the conventional gradient descent. What we are sure from Eq. (129) is that for a large m we have

$$\frac{1}{m} \sum_{k=1}^m v_b^{2[2](k)} = 1. \quad (131)$$

However, comparing Eqs. (130) and (131) we can see that the extra terms, i.e., \mathcal{L}' and g'_2 , make it very challenging to determine in advance the efficiency of this method.

Nonetheless to conclude this section let's proceed to determine the other parameters. In contrast to $v_b^{[2]}$, the other terms, given in Eqs. (108)-(111), exhibit the following form,

$$Av_i^2 + \sum_{i \neq j} Bv_i v_j. \quad (132)$$

It is clear from Eqs. (108)-(111) that if $v_i^2 = 1$ and $v_i v_j = 0$ for all possible values, then forward gradient descent reduces to the conventional gradient descent. But this is not the case. Similar to what we did for $v_b^{[2]}$, every other directional vector component in forward gradient descent is an independent random variable from a normal distribution with zero mean and unit variance. Consequently, in Eq. (132) we already showed that $\mathbb{E}[v_i^2] = 1$. What we are missing to show is that $\mathbb{E}[v_i v_j] = 0$. We therefore have,

$$\mathbb{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dx dy, \quad (133)$$

Since each function is odd in the above expression, then the global integral is zero. As a result, it is true that $\mathbb{E}[XY] = \mathbb{E}[v_i v_j] = 0$.

This method needs to be revisited to see if it may perform well in a physics informed neural network.

2.5 Physics informed neural networks

A second-order linear partial differential equation with two independent variables (x, t) in its general form reads:

$$a_{(x,t)} \frac{\partial^2 u}{\partial x^2} + b_{(x,t)} \frac{\partial^2 u}{\partial x \partial t} + c_{(x,t)} \frac{\partial^2 u}{\partial t^2} + \alpha_{(x,t)} \frac{\partial u}{\partial x} + \beta_{(x,t)} \frac{\partial u}{\partial t} + \gamma_{(x,t)} u + \delta_{(x,t)} = 0 \quad (134)$$

Eq. (134) is a time-dependent one dimensional system that exhibits a physical meaning if we include initial and boundary conditions. To ease the discussion

let's consider that the system is constraint along the x -axis in the interval $[0, 2\pi]$, then our conditions read:

$$u(x, 0) = u_0(x), \quad (135)$$

$$u(0, t) = u(2\pi, t). \quad (136)$$

Eq. (135) is our initial condition and Eq. (136) is our boundary condition. Notice that both conditions are imposed without taking into account an analytical solution to the PDE.

To start with, let's consider that the real solution to Eq. (134) is $u = f(x, t)$, with $f: \mathbb{R}^2 \rightarrow \mathbb{R}$. Similarly, let's construct a neural network, such as the one given in Fig. 1, where we had 2 input parameters and one output. The solution is given in Eq. (13) where we had $y = g_2(x, t)$, with $g_2: \mathbb{R}^2 \rightarrow \mathbb{R}$. As you can see, g_2 and f are quite similar, they both have two input parameters and one output. If we manage to construct a neural network such that its output is always similar to $f(x, t)$, then we would have obtained a proper neural network approximation to the PDE. To do so, there are three important steps to follow. Let's start with the simplest one.

2.5.1 Initial condition

The first step is to properly set the initial condition. As you can see in Eq. (135) in this particular case $t = 0$ for all values of x . We therefore have, $u(x_i, 0) = u_0(x_i)$, where x_i is a particular point in the interval $[0, 2\pi]$. For simplicity we consider m_x points, i.e., $x_i = \{x_1, x_2, x_3, \dots, x_{m_x}\}$. Following a similar approach for a neural network, we would have m_x input parameters of the form $(x_i, 0)$ and the output would be $y(x_i, 0)$. If we would want our neural network to output the same result as the PDE, then we should have $y(x_i, 0) - u(x_i, 0) \approx 0$; however, because we are considering m_x samples, it is better to speak of the least squared error, which is the cost function that we already defined in Eq. (67). In detail we have,

$$\mathcal{J}_{init} = \frac{1}{m_x} \sum_i^{m_x} (y(x_i, 0) - u(x_i, 0))^2. \quad (137)$$

If our goal would only be to construct a neural network that outputs $\sim u(x_i, 0)$. Then we would have already solved the problem because all what is left is to perform a gradient descent to reduce the cost function. But this is not the case, let's consider now the second step.

2.5.2 Boundary condition

The second step is a bit more complex. The condition that we want to fulfill is given in Eq.(136), which can be re-expressed as,

$$h(t_i) = u(0, t_i) - u(2\pi, t_i) = 0. \quad (138)$$

Similar to the previous step, we can consider m_t points such that $t_i = \{t_1, t_2, \dots, t_{m_t}\}$. Then, in our neural network we would have $y(0, t_i)$ and $y(2\pi, t_i)$. Notice that, in contrast to the previous case, here we perform the forward propagation twice, one with $x = 0$ and the second one with $x = 2\pi$. The difference reads,

$$h_n(t_i) = y(0, t_i) - y(2\pi, t_i). \quad (139)$$

Notice that we have defined the function h_n similar to Eq. (138). If we would want our neural network to output the same result as the PDE, then we should have $h_n(t_i) - h(t_i) \approx 0$; however, because we are considering m_t samples, it is better to consider the cost function, which reads,

$$\mathcal{J}_{bc} = \frac{1}{m_t} \sum_i^{m_t} (h_n(t_i) - h(t_i))^2. \quad (140)$$

As you may notice, $h(t_i) = 0$ always, then the cost function reduces to,

$$\mathcal{J}_{bc} = \frac{1}{m_t} \sum_i^{m_t} (h_n(t_i))^2 = \frac{1}{m_t} \sum_i^{m_t} (y(0, t_i) - y(2\pi, t_i))^2. \quad (141)$$

At this point, if we would like our neural network to fulfill both, the initial and the boundary conditions, then we need both errors to approach to zero. We therefore sum up the cost functions,

$$\mathcal{J} = \mathcal{J}_{init} + \mathcal{J}_{bc}, \quad (142)$$

and perform a gradient descent. However, as you may have noticed, we have not yet taken into account the PDE. This is the third and final step.

2.5.3 Partial differential equation

Similar to the previous section we set a new function of the form,

$$p(x, t) = a_{(x,t)} \frac{\partial^2 u}{\partial x^2} + b_{(x,t)} \frac{\partial^2 u}{\partial x \partial t} + c_{(x,t)} \frac{\partial^2 u}{\partial t^2} + \alpha_{(x,t)} \frac{\partial u}{\partial x} + \beta_{(x,t)} \frac{\partial u}{\partial t} + \gamma_{(x,t)} u + \delta_{(x,t)} = 0, \quad (143)$$

As you may notice, this function is always zero; however, in contrast to Eq. (138), p depends in both, x and t . To obtain a similar expression for our neural network we set,

$$p_n(x, t) = a_{(x,t)} \frac{\partial^2 y}{\partial x^2} + b_{(x,t)} \frac{\partial^2 y}{\partial x \partial t} + c_{(x,t)} \frac{\partial^2 y}{\partial t^2} + \alpha_{(x,t)} \frac{\partial y}{\partial x} + \beta_{(x,t)} \frac{\partial y}{\partial t} + \gamma_{(x,t)} y + \delta_{(x,t)} = 0, \quad (144)$$

As you can see in Eq.(144), $a, b, c, \alpha, \beta, \gamma$, and δ are easily known because they depend solely on the input parameters. For example, we can randomly choose m_p samples of the form (x_i, t_i) . In this case, the difficult part relies in the derivative terms; however, these terms were already solved in the *autograd* section. Consequently, Eq.(144) is numerically solved. If we would want our neural network to output the same result as the PDE, then we should have $p_n(x_i, t_i) - p(x_i, t_i) \approx 0$; however, because we are considering m_p samples, it is better to consider the cost function, which reads,

$$\mathcal{J}_{pde} = \frac{1}{m_p} \sum_i^{m_p} (p_n(x_i, t_i) - p(x_i, t_i))^2. \quad (145)$$

Since $p(x_i, t_i) = 0$ always, then the cost function reduces to,

$$\mathcal{J}_{pde} = \frac{1}{m_p} \sum_i^{m_p} (p_n(x_i, t_i))^2 \quad (146)$$

At this point, if we would like our neural network to fulfill all conditions, then we need the total error to approach to zero. We therefore sum up the cost functions,

$$\mathcal{J} = \mathcal{J}_{init} + \mathcal{J}_{bc} + \mathcal{J}_{pde}, \quad (147)$$

and perform a gradient descent. As you may notice, these three steps are the building blocks to construct a physics informed neural network and can be extended to add additional constraints. It is important to keep in mind that in this section we have considered a simple neural network, then even if the formalism is correct, the error will be huge. What we need is a neural network with several hidden layers. This is discussed in the following section where we study a time-dependent three dimensional PDE. The open questions are:

- 1) How big a neural network needs to be to solve a fourth dimensional PDE? Is there a correlation between the number of hidden layers and the PDE complexity?
- 2) Second order derivatives in the *autograd* function have higher order terms that are quite small compared to the first order components. Disregarding these terms may speed up the computation time while not affecting that much the gradient descent operation. This must be revisited.
- 3) Is it possible to use a general directional vector instead of the basis vectors in the forward mode? Forward gradient descent is one alternative. Can we construct a novel approach?

3 General Model

blah blah