Iranian Journal of Astronomy and Astrophysics
Vol. 10, No. 4, Autumn 2023, 335–356
©Available online at http://ijaa.du.ac.ir
DOI: 10.22128/ijaa.2023.694.1150
Online ISSN: 2383–403X

Iranian Journal of Astronomy and Astrophysics

Research Paper

Physics-Informed Deep Learning for Three Dimensional Black Holes

Emad Yaraie[1] · Hossein Ghaffarnejad*[2] · Mohammad Farsam[3]

[1] Instituut-Lorentz for Theoretical Physics, ITP, Leiden University, Niels Bohrweg 2, Leiden 2333 CA, The Netherlands;
email: eyaraie@semnan.ac.ir

[2] Faculty of Physics, Semnan University, P.C. 35131–19111, Semnan, Iran;
*email: hghafarnejad@semnan.ac.ir

[3] Instituut-Lorentz for Theoretical Physics, ITP, Leiden University, Niels Bohrweg 2, Leiden 2333 CA, The Netherlands;
email: mhdfarsam@semnan.ac.ir

Received: 27 June 2023; **Accepted:** 30 December 2023; **Published:** 30 December 2023

Abstract. In this paper, we have designed an artificial neural network architecture to produce metric field of planar BTZ and quintessence black holes applying a data-driven approach andleveraging holography principle (according to AdS/DL (Anti de Sitter/ Deep Learning) correspondence given by [1]). Data has been collected by choosing minimally coupled massive scalar field with quantum fluctuations and we try to process two emergent and ground-truth metrics versus the holographic parameter which plays the role of depth of the neural network. Loss or error function which shows rate of deviation of these two metrics in presence of penalty regularization term reaches to its minimum value when values of the learning rate approach to the observed steepest gradient point. Values of the regularization or penalty term of the quantum scalar field has critical role to matching this two mentioned metric. Also, we design an algorithm which helps us to find optimum value for learning parameter and finally, we understand that loss function convergence heavily depends on the number of epochs and learning rate.

Keywords: Machine learning, Deep learning, Black holes, Three dimensions, BTZ; Optimization, Loss function

1 Introduction

After prediction of black holes evaporation in presence of quantum matter field by Hawking [2] and the black hole entropy by Bekenestein [3] which is related to surface gravity of the black hole, Susskind and t'Hooft stated that the theory of quantum gravity within any region is encoded on the surface of that region [4,5] which is called now as the holographic principle. The best successful theory so far for the holographic principle is the Anti de Sitter-conformal field theory (AdS/CFT) correspondence which is proposed by Maldacena [6] for the first time. This correspondence has two consequences such that the quantum gravity in each slice of AdS spacetime is explained by the data on the boundary slice and

* *Corresponding author*

information which lives on the boundary evolves between the slices of the AdS spacetime by the Hamiltonian of conformal invariant quantum fields. The study of AdS black holes in $d < 4$ dimensions are conducted in a variety of ways [7,8]. For instance, one can see [9,10] for understanding of the dual field theory in the context of AdS/CFT correspondence and [11–16] for studying of effects of the quintessence fields in 1+2 dimensional black holes spacetimes. Compared to 4D case, BTZ black hole has certain good theoretical properties, e.g. this is just an example, not proposal: turning on angular momentum is simpler compared to 4D case, and something new may be checked by utilizing it. In fact, the BTZ black hole was introduced by Maximo Banados, Claudio Teitelboim, and Jorge Zanelli [17,18] (see also [19]). The BTZ black hole geometry is the solution of the vacuum Einstein field equation in $2 + 1$ dimensions in presence of a negative cosmological constant. In fact, its dual field theory is the $1+1$ dimensional CFT, from the AdS/CFT viewpoint, and the presence of the black hole corresponds to finite-temperature effects on the CFT.

Recently, deep learning is used extensively in the solution of computer vision problems. Training of deep neural networks is in fact an optimization problem, for which a loss function should be minimized. By according to some criteria, loss functions must be determined based on what we need the model to learn. Although loss functions play important role in deep learning applications, but an extensive comparison of them is not available in the literature. There are just tree kind in the convolutional Neural Network so called as AlexNet, VGG and GoogleNet. They are reviewed appropriately in the thesis [20]. Deep neural network which is known as deep structured learning is part of a broader family of machine learning methods based on artificial neural networks with representation learning. This is shown that is extended to be applicable for more branches of physical science such as the gravitational and the cosmological context (see [21] for a good review). For instance, one can see some published works about application of deep learning method related with gravity as follows: Yong Yang et al used deep learning method to determine atmospheric parameters of white dwarf stars recently [22]. Christopher J. Shallue and Andrew Vanderburg also used deep learning method to identify exoplanets [23]. Matsuoka et al, apply the deep learning method to estimate parameters of atmospheric gravity wave in reanalysis data sets [25]. In fact, neural networks that work according to the laws of physics are called physics-informed neural networks (see [24] and references therein). This kind of learning algorithm is inspired by information processing and it is distributed by communication nodes in biological systems. Artificial neural networks models have been used since the 1950s [26] and flourished in the 2000s [27]. It is composed of multiple layers to progressively extract higher-level features from the raw input and delivering an output. With respect to the task at hand, the output could be have discrete value or continuous value [27,28]. Recent breakthrough results in computer vision, natural language processing speech recognition, biomedicine and many other domains have produced a massive interest in this direction [29–32]. Hashimoto et al in ref. [1] presented a deep neural network representation for the AdS/CFT correspondence. They demonstrated the emergence of the bulk metric function via the learning process for given data sets of response in boundary quantum field theories. In this approach, the emergent radial (holographic) direction of the bulk is identified with the depth of the layers, and the network itself is interpreted as a bulk geometry. As an application of correspondence of deep learning with Ads/CFT correspondence they used the scalar ϕ^4 theory with unknown mass and coupling, in unknown curved spacetime contained a black hole horizon, and investigated two properties as follows. The first, they demonstrated their deep learning framework to determine the background metric which fit given response data by showing from boundary data generated by the AdS Schwarzschild spacetime. The second step, they demonstrated that the their proposed network with experimental data as an input, can determine the bulk metric, the mass and the quadratic coupling of the holographic model. At last, they use the experimental data of magnetic response of a strongly correlated material $Sm_{0.6}Sr_{0.4}MnO_3$. By looking at their work one can infer that their proposed AdS/DL correspondence not only enables gravity modeling of strongly correlated systems, but also sheds light on a hidden mechanism of the emerging space in both AdS and DL. Hence one can extend their approach by considering other physical effects in production of a black hole geometry.

By according to ideas given in [1], we should provide a deep neural network representation of a scalar field equation moving in curved spacetime where the discretized holographic AdS radial direction is the deep layers. The weights of the neural network are identified with metric of the curved spacetime. The input response data is at the boundary of AdS, and the output binomial data is the black hole horizon condition. Therefore, a successful machine learning results in a concrete metric of a holographic modeling of the system measured by the experiment. This is all which is called as AdS/DL correspondence of a deep neural network by Hashimoto et al. When stress tensor of scalar field has zero barotropic index $w = 0$ (the dust model) then the 3D black hole reads as planar BTZ black hole. In the cosmological regime, the dark sector of the matter/energy proposal play an important role in support of the cosmic inflation and so considering these effects on the BTZ black hole formation should be useful which we like to investigate in this work. For a quintessence phase of the dark fluid which is surrounded the a BTZ black hole, the non vanishing barotropic index is $-1 < w \leq -\frac{1}{3}$ for which the 3D black hole is called as quintessence black hole which we like to produce them by using method given by [1]. The paper is organized as follows:
In section 2, we present brief review of architecturing deep neural network and developing deep neural learning model. In section 3, we provide a brief review of 1+2 dimensional BTZ black hole metric solution. Then, we investigate correspondence between metric components and parameters of deep neural network for the BTZ planar black hole such that the black hole could feed with in input layer by corresponding boundary data which is labeled with respect to the horizon boundary conditions. Then, when data is propagating towards the black hole horizon, the spacetime metric is being reproduced. Section 4 is dedicated to the network architecture, training implementation and data setting. In the last section, we investigate conclusion and outlook of the work.

2 Artificial neural network

A neural network, also sometimes is called an artificial neural network, is a kind of processing structure which their name and structure are inspired by the human brain, mimicking the way where the biological neurons signal to one another. Basic building block of a neural network is made in fact by a neuron. We show schematic diagram of a simple neural network in Figure 1. In this figure, the artificial neuron takes all the inputs $x_{1,2}$, weights W (shown with solid lines) which is a linear transformation between vector components of the neuron as $x_i \to \Sigma_j W_{ij} x_j$, aggregates (not shown) and an activation function $x_i \to \varphi(x_i)$ which is usually a nonlinear transformation on the vector components of the neuron x_i such that it should deliver the output of the neuron at each layer. In fact, the activation function controls value of the output when the neuron is activated. A row of neurons is called layer and a network can have multiple layers. Input layer receives data x_i and delivers output to next layer via two above mentioned transformations as $x_k \to \varphi(W_{kl} x_l)$ and final layer is responsible for delivering values which correspond to result demanded for the problems such that regression, classification and etc. Layers located between first and last ones are called hidden layers. In general, for N layers a deep feed-forward neural network can be

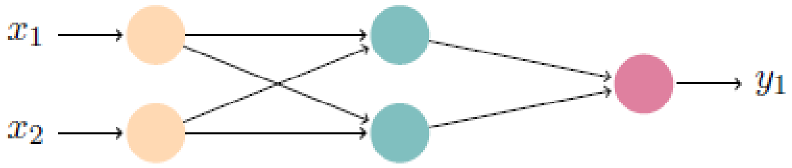


Figure 1: Schematic diagram of a simple neural network in which activation function is shown with the colored circle so that pink (left side), blue (middle) and violet (right side) correspond with input, hidden and output neuron layers respectively. Weights are shown with solid lines. Input data is $x_{1,2}$ and output one is y_1.

constructed as follows

$$y(x^{(1)}) = f_i\varphi(W_{ij}^{(N-1)}\varphi(W_{jk}^{(N-2)}\cdots\varphi(W_{lm}^{(1)}x_m^{(1)}))), \tag{1}$$

where f_i means activation function $x_i \to \varphi(x_i)$ but at last layer which delivers to the target $y(x^{(1)})$. In the learning process, the variables of the Network $(f_i, W_{ij}^{(n)})$ for $n = 1, 2, \ldots, N-1$ are updated by a gradient descent method with a given loss or error function

$$E = \sum_{data} |y(\bar{x}^{(1)}) - \bar{y}| + E_{reg}(W). \tag{2}$$

Here, the sum is over the whole set of pairs $\{(\bar{x}^{(1)}, \bar{y})\}$ of the input data $\bar{x}^{(1)}$ and the output data $\bar{y}$. The regularization penalty term E_{reg} is introduced to require expected properties for the wight [32]. The equation (2) can be evaluated by different optimizing methods such as gradient descent, Adam and etc which in fact is an iterative method for optimization of a function. By moving data from input layers to final layer via feed-forward algorithm with suitable smoothness properties it demonstrates how much predicted values are far from values of ground truth $\bar{y}$? This error is then propagated back through the network by applying back propagation algorithm so that the weights are updated according to the amount that they contributed to the error [33]. Predictions are made by providing the input to the network and by performing a forward pass and then by generating an output. In this view the architecture means how a model can be constructed from two dimensional input data and one dimensional output feature. With respect to the context of our problem this architecture can be extended to more layers and neurons with various kind of activation functions and operations of between layers [29–32].
In the following section, we investigate correspondence between the BTZ black hole metric and neural network components.

3 Neural network for planar BTZ Black holes

In 1992 Banados, Teitelboim and Zanelli investigated and obtained a 3D planer black hole which is called now BTZ black hole solution [17–19] (see also [7]).In absence of the cosmological constant, there is no black hole containing event horizons in 3D curved spacetimes but thanks to the negative cosmological constant there is BTZ black hole metric solution which provides properties similar to ones which are appeared for 4D Schwarzschild black holes. By considering planar topology, general form of metric field in 1+2 dimensional black hole spacetimes is

$$ds^2 = -f(r)dt^2 + f(r)^{-1}dr^2 + r^2dx^2, \tag{3}$$

where x is a planar coordinate, r is the radial coordinate and $f(r)$ stand for the blacking functions. Einstein's field equations can be written as

$$R_{ab} - \frac{1}{2}g_{ab}R - \frac{1}{L^2}g_{ab} = 8\pi T_{ab}, \tag{4}$$

where $a, b = 1, 2, 3$ in 3D spacetimes and L is the AdS radius. The right side stress tensor is assumed to be perfect non viscous fluid such that

$$T_t^t = T_r^r = -\rho, \quad T_x^x = (2w+1)\rho, \tag{5}$$

where ρ and w are energy density and the state parameter of the fluid respectively [7,11]. By substituting the stress tensor (5) and by solving the Einstein's equations (4) with respect to the line element (3) we obtain

$$ds^2 = -\frac{r^2}{L^2}f(r)dt^2 + \frac{L^2}{r^2}f(r)^{-1}dr^2 + \frac{r^2}{L^2}dx^2, \tag{6}$$

where blacking function takes on the following form

$$f(r) = 1 - \left(\frac{r_+}{r}\right)^{\sigma}, \quad \sigma = 2(1+w_q),$$

in which r_+ is radius of the black hole event horizon and for BTZ model $\sigma = 2$ can be written versus the ADM mass of the black hole M and the AdS radius L such that $r_+ = \left(ML^2\right)^{1/2}$ [7,11]. In fact, the BTZ 1+2 dimensional black hole in a Schwarzschild coordinates is stationary and axially symmetric because it has two Killing vectors $J^t\partial_t$ and $J^\varphi\partial_\varphi$ and generically has no other symmetries for which the event horizon is determined by M, L, J^φ. In the above planner line element, we eliminated axially symmetric property of the BTZ black hole by using the planner symmetry and so the constant angular momentum J^φ is negligible. The case $w = 0$ corresponds to the non-quintessence BTZ black hole and $-1 < w < -\frac{1}{3}$ corresponds to quintessence black hole, which in this paper we are interested for particular choices $w = \{0, -\frac{1}{2}, -\frac{3}{4}\}$ and design artificial neural networks in order to represent scalar field in background of them.

In order to facilitate designing neural network architecture, we use the following conformal transformation for r coordinate.

$$dz = f^{-\frac{1}{2}}dr, \tag{7}$$

in which z is holographic direction and by integrating of the above transformation we have

$$r = r_+ \cosh\left(\frac{z}{L}\right). \tag{8}$$

By substituting this into the line element (3), we obtain

$$ds^2 = -f(z)dt^2 + dz^2 + g(z)dx^2, \tag{9}$$

where the BTZ metric components are given versus the holographic z parameter as follows

$$f(z) \equiv \frac{r_+^2}{L^2}\left(\sinh\frac{z}{L}\right)^2, \quad g(z) \equiv \frac{r_+^2}{L^2}\left(\cosh\frac{z}{L}\right)^2. \tag{10}$$

In this conformal frame, the boundary of the AdS is located at infinity $z \to \infty$ for which $f(z) \to g(z) \approx (r_+^2/4L^2)\exp(2z/L) \to \infty$ while the black hole horizon lives at $z_h = 0$ for which $f(z) = 0$ and $g(0) = (r_+^2/L^2)$. As an application of neural network model, we like to study interaction of a scalar field with the BTZ black hole metric as follows. We consider a minimally coupling massive scalar field with self interaction potential $V(\phi)$ which is propagated in the spacetime (9). Dynamics of this field is described by the following Lagrangian density.

$$\mathcal{L} = \sqrt{g}\bigg\{\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - \frac{1}{2}m^2\phi^2 - V(\phi)\bigg\}, \tag{11}$$

in which $g = |\det g_{\mu\nu}|$ is absolute value of determinant of the metric field $g_{\mu\nu}$ and by varying with respect to the field ϕ the corresponding Euler Lagrange equation reads

$$\Box\phi + m^2\phi + \frac{\delta V}{\delta\phi} = 0, \quad \Box \equiv g^{-\frac{1}{2}}\partial_\mu(g^{\frac{1}{2}}g^{\mu\nu}\partial_\nu), \tag{12}$$

which for (9) can be written as the following first order differential equation.

$$\partial_z\pi + R(z)\pi + m^2\phi + \frac{\delta V[\phi]}{\delta\phi} = 0, \tag{13}$$

where $\pi \equiv \frac{\partial\phi(z)}{\partial z}$ is canonical momenta of the field ϕ and

$$R(z) = \frac{1}{2}\frac{d\ln(f(z)g(z))}{dz} = \frac{\sigma - 2 + 2\cosh\left(\frac{2z}{L}\right)}{L\sinh\left(\frac{2z}{L}\right)}, \tag{14}$$

is an effective potential. This potential is singular on the black hole horizon $z_h = 0$ but has finite value $R(\pm\infty) = \pm\frac{2}{L}$ on the AdS boundary. The equation (13) together with $\pi \equiv \frac{\partial\phi(z)}{\partial z}$ can be solved via neural network system by discretization method. To do so, the strategy should be providing a manifestation of scalar field equation in deep neural network scheme [1] where holographic direction z mimics the deep layers and the neurons are shown with 2 components vectors $(\phi(z), \pi(z))$. Correspondence of the field equation with the neural network system is possible by discretizing the equation of motion in holographic direction z such that [1]

$$\begin{aligned} \phi(z+\Delta z) &= \phi(z) + \Delta z\pi(z), \\ \pi(z+\Delta z) &= \pi(z) - \Delta z\bigg(R(z)\pi(z) + m^2\phi(z) + \frac{\delta V(\phi)}{\delta\phi(z)}\bigg), \end{aligned} \tag{15}$$

which can be written with matrix form as follows

$$\begin{pmatrix} \phi(z+\Delta z) \\ \pi(z+\Delta z) \end{pmatrix} = \begin{pmatrix} 1 & \Delta z \\ -m^2\Delta z & 1 - R(z)\Delta z \end{pmatrix}\begin{pmatrix} \phi(z) \\ \pi(z) \end{pmatrix} + \begin{pmatrix} 0 \\ -\frac{\delta V(\phi(z))}{\delta\phi(z)}\Delta z \end{pmatrix}, \tag{16}$$

where Δz is distance of adjacent points in discrete coordinate system with

$$z^{(n)} \equiv (N - n + 1)\Delta z,$$

and N is total number of neural network layers. According to the Figure 1 for the equations (16), we can use $x_1 \equiv \phi(z)$ and $x_2 = \pi(z)$ for components of the vector neurons $(\phi(z), \pi(z))$. Regarding these and linear affine transformation $x_i \to \sum_j W_{ij}x_j$ one can obtain weights matrix W_{ij} for the equations (16) as

$$W^{(n)} = \begin{pmatrix} 1 & \Delta z \\ -m^2\Delta z & 1 - \Delta zR(z^{(n)}) \end{pmatrix}, \tag{17}$$

for n layers and by regarding the nonlinear transformations $x_i \to \phi(x_i)$ for each layer one can obtain activation function for output data on each layer as follows

$$\begin{cases} \phi_1(x_1) = x_1, \\ \phi_2(x_2) \to x_2 - \Delta z \frac{\delta V(x_1)}{\delta x_1}. \end{cases} \tag{18}$$

In fact, the definitions (17) and (18) bring the scalar field system in curved geometry (11) into the form of neural network (1) [1]. Thus, one can infer that architect of a neural network system in this paper corresponds to scalar field equation in BTZ black hole spacetime in which the weights of network play role of the BTZ black hole metric, $\varphi_{1,2}$ take on role of the activation functions and holographic direction should mimic depth of the network. For simplicity, in the rest of the paper, we set $L = 1$, $m = 1$ and $V[\phi] = \frac{\lambda\phi^4}{4}$ (the Higgs potential) with $\lambda = 1$ and number of hidden layers to be 8 which yields to $\Delta z = -0.1$, $z_b = 1$ and $z_h = 0.1$ (the horizon cut off frequency) which is used to regularization of interacting quantum scalar fields. In fact, input data for ϕ originates from quantum fluctuations of the field (see equation 9 in ref. [1]) which whose frequencies approach to infinite value on the black hole horizon and they should be regularized. In the following section, we investigate numerical processing to produce output data or target.

4 The network architecture, training implementation and data setting

The architectures of our neural network setup with total 10 layers is shown schematically in Figure 2 and corresponding data are collected in the Table 1 by designing as 8 hidden layers with two input and output layers.

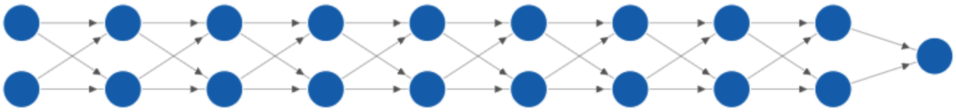


Figure 2: The architectures of our neural network setup composed with 8 hidden layers consists of two neurons in each input and hidden layers and one neuron in output layer respectively.

The architecture is implemented via PyTorch ecosystem [34] in GPU mode. The dataset can be randomly produced by drawing values of independent variables ϕ and π in AdS boundary $z = 1$ for domains $\phi \in [0, 1.5]$ and $\pi \in [-0.2, 0.2]$ respectively and transform them to the black hole horizon $z_h = 0$ by applying the equation of motion (15) for metric potential (14) (see Figure 3). To do so, we choose 1000 positive value data which can be identified by $|F| < 0.1$ as cut off on the horizon and 1000 negative value data corresponding to $|F| > 0.1$ which are labeled with $y_+ = 0$ and $y_- = 1$ respectively. In fact, the boundary condition at the horizon can be used as a classifier to categorize generated dataset into binary classes such that for some positive input data the output at the final layer should satisfy

$$0 = F \equiv \left[\frac{2}{\eta}\pi + m^2\phi + \frac{\delta V(\phi)}{\delta\phi}\right]_{z_{fin}}, \tag{19}$$

Table 1: Architecture used in the networks with batch size 10.

Layer	Transformation	Output	dimension
h_0	affine linear	(ϕ_1, ϕ_2)	2
h_1	affine linear	(ϕ_1, ϕ_2)	2
h_2	affine linear	(ϕ_1, ϕ_2)	2
h_3	affine linear	(ϕ_1, ϕ_2)	2
h_4	affine linear	(ϕ_1, ϕ_2)	2
h_5	affine linear	(ϕ_1, ϕ_2)	2
h_6	affine linear	(ϕ_1, ϕ_2)	2
h_7	affine linear	(ϕ_1, ϕ_2)	2
h_8	affine linear	(ϕ_1, ϕ_2)	2
h_9	linear	f(F)	1

in which $z = z_{fin} << 1$ is the horizon cutoff. Dataset will be injected into the neural network in 200 batches. In other words we choose 100 batches for positive and 100 batches for negative value data respectively which they propagate through the neural network from visible layer (z_h) to the final layer (z_{fin}) via equation of motions. Our final layer is defined by the map F such that the output data is $y_+ = 0$ for a positive answer response data originated from quantum fluctuations of the field [1]. In fact, for limits $z_{fin} = 0$ the condition (19) reads $\pi(z = 0) = 0$. Now we can make the deep neural network to learn the metric component function $h(z)$, the mass parameter of the field m and the interaction potential $V(\phi)$. The training is done by the loss function (2). In fact, experiments provide only positive answer data with $y_+ = 0$, while for the training we need also negative answer data which is to generate false response data and so we assign output $y_- = 1$ for the latter case. According to choice given by [1] we use a function $\tanh|F|$ for the final layer rather than just F, because $\tanh|F|$ provides $y \to 1$ for any negative input. By regarding these choices, the final output of the neural network is made as binary. In this view the activation function of final layer for cases $w = 0,\ w = -\frac{1}{2}$ and $w = -\frac{3}{4}$ can respectively given by [1],

$$f(F) = 1 + 0.5\tanh[100(F - 0.1)] - 0.5\tanh[100(F + 0.1)], \tag{20}$$

$$f(F) = 1 + 0.5\tanh[Q(F - 0.1)] - 0.5\tanh[0.6(F + 0.1)], \tag{21}$$

and

$$f(F) = 1 + 0.5\tanh[1.1(F - 0.1)] - 0.5\tanh[P(F + 0.1)], \tag{22}$$

where $Q = \{0.6, 0.9, 1.1\}$ and $P = \{0.6, 0.8, 0.9\}$. Looking at the Figures 10 and 12, one can infer that the best fit is happened for $Q = 0.6$ and $P = 0.9$. and to choose physically sensible metric among other learned metrics, we use lose function (2) and the penalty or regularization term given by the discrete form of the metric potential (14) as

$$E_{reg} = 3\sum_{n=1}^{N-1}(z^i)^4(R(z^{i+1}) - R(z^i))^2,$$

to plot variation of loss function versus the leaning rate in Figure 4. This diagram shows minimum variant of the error function is happened for learning rate 0.1 approximately. In the error function (2), the quantities $(\bar{x}_i, y_i(\bar{x}))$ are the training dataset and $\bar{y}$ is ground-truth y. The produced errors by loss function can be saved up across all of the training examples and the network can be updated at the end. The hyperparameters or training parameters which we used in this work are as follows: The batch size namely number of training samples

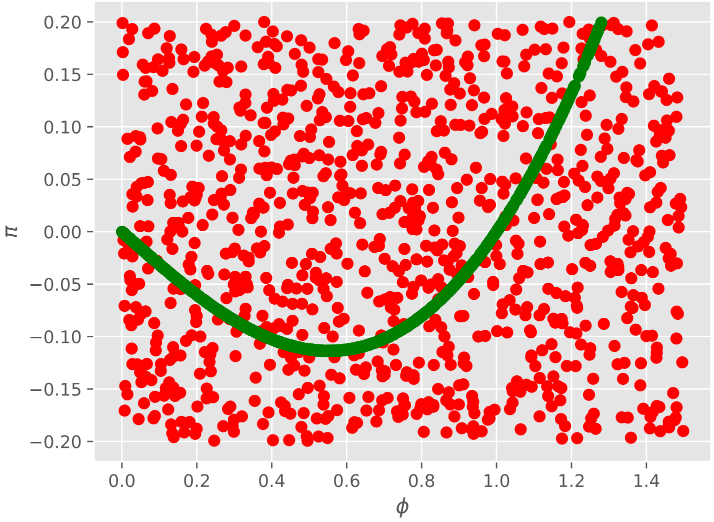


Figure 3: The data generated by the discretized BTZ metric (14) to visualize how numerical values should be used as data processing. The green points correspond to the positive data $y_+ = 0$ and the red points correspond to the negative data $y_- = 1$. This diagram is produced for non quintessence BTZ black hole $w = 0$ and for quintessence cases $w = -\frac{1}{2}, -\frac{3}{4}$ we will have similar diagrams (not shown).

which is used to compute the gradient at each update is 10 for non-quintessence case $w = 0$ and 100 for quintessence cases $w = -\frac{1}{2}, -\frac{3}{4}$. The optimum learning rate hyperparameter is chosen with numeric values 0.0001 and 0.01 for non-quintessence and quintessence cases respectively. (These values for learning rate can be detected by design an algorithm which lead us to an optimal learning rate for making the model. In the following subsection, we will be talking about the procedure of finding optimum learning rate). Looking at the Figure 5 one can infer that presence of an suitable penalty or regularization term is crucial to choose well learned metric among other learned metrics. In Figures 5 and 6, by using tuned values of learning rate and batch size, we have illustrated the impact of epochs on performance of model. It can be seen in Figure 6 with 50,000 epochs in which the emerged metric mimics ground-truth metric pretty well. To check how well the model is learned [36], the optimization learning curves and the performance learning curves are plotted for epoches 10,000, 30,000 and 50,000 respectively in Figures 7, 8, and 9 respectively. In fact, these learning curves are as diagnostic tools for plot of model learning optimization, performance over experience or times.Looking at these diagrams one can infer that it is vivid the model which learned with 50,000 epochs behaves better.

4.1 The dynamics of learning rate

The learning rate hyperparameter controls the speed at which the model learns. A large learning rate allows the model to learn faster and a small learning rate may allow the model to learn better but the price has to be paid is longer learning time. A learning rate that is too large could result in large weight updates which causes the objective function of the model shows an oscillation behavior with respect to the training epochs. The source of oscillating behavior gets back to weights that are diverging. On the other hand, a learning rate that is too small may get stuck on a suboptimal solution. Diagnostic plots can be used to investigate how the learning rate impacts learning dynamics of the model. This is investigated by Leslie N. Smith in [37] in depth. He has demonstrated that if a model be trained initially by a low learning rate and then it get increased exponentially or linearly at each iteration a good learning rate candidate could be achieved but if we monitor the learning at each iteration and then plot the logarithm of learning rate versus loss function, there will be spotted as the learning rate increases and a point is appeared where the loss decreases to stop emerges and then starts to increase again. This minimum point is the point we will be choosing as the learning rate hyperparameter of our model. In order to find minimum value of the error function we utilized Adam optimizer [35] with starting learning rate 0.1 and corresponding exponential decay as $\beta_1 = 0.05$. In fact, the Adam optimizer is an adaptive learning rate optimization algorithm where momentum instead of the gradient of current step is applied to guide the search. In other words, it is combined directly as an estimate of the first order moment of the gradient and accumulates the gradient of the past steps to determine the direction to go. By conducting experiment base on what explained above in order to find optimum learning rate we obtain diagram of Figure 4-a for non quintessence case and Figure 4-b for quintessence case where in both of them a quick drop can be observed in the loss function. In fact, increasing the learning rate further will cause an increase in the loss and even diverge from the minimum because of the parameter updates.

5 Conclusion

In this paper, by leveraging correspondence of AdS/CFT and AdS/SL, we design deep neural network architecture for 3D planar BTZ and quintessence black holes to learn boundary data which lives on conformal field theory side. To do so, we saw that the weights of network play the role of metric and holographic direction mimics the depth of network. Such that data propagates from boundary to horizon of black hole and cause to produce the background metric. We have considered a penalty regularization term for loss function such that to be only sensible with respect to the reality metric to be chosen among other learned metrics. In order to achieve a high-performing model, hyperparameters tuning has been conducted. We have noticed loss function convergence heavily depends on the number of epochs and learning rate. Finding faster convergence for loss function motives us to investigate the impact of learning rate on neural network performance by performing an experiment where we gradually increase exponentially the learning rate to observe for steepest drop in loss function which has guided us to pick up suitable learning rate parameter. The message of our paper is that the emerged spacetime could be a more universal phenomenon and helps to understand emergence of spacetime in holographic three dimensions. In this case one can infer that the ADS/DL correspondence and neural network data processing paradigm could be an applicable model instead of the unknown pure quantum gravity theory. Such that it can say us what is happening at Planck scale of the nature? As we saw, the error function has an integral relationship with the emergent metric function, so the physical parameters of the assumed black hole, such as electric charge, angular momentum, or other physical quantities, (for instance quintessence effect which is considered here), will play an important role to form the loss function and so correspondence of two emergent metric and ground truth metric. Checking of the work for angular momentum effect of the BTZ black hole via deep learning and neural network data processing, is needed more time to produce the numerical processing which we intend to do in the next work.

Authors’ Contributions

All authors have the same contribution.

Data Availability

No data available.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Ethical Considerations

The authors have diligently addressed ethical concerns, such as informed consent, plagiarism, data fabrication, misconduct, falsification, double publication, redundancy, submission, and other related matters.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or nonprofit sectors.

References

[1] Hashimoto, K., Sugishita, S., Tanaka A., & Tomiya, A. 2018, Deep learning and the AdS/CFT correspondence, Phys. Rev. D4, 046019.

[2] Hawking, S. 1976, Black Holes and Thermodynamics, Phys. Rev. D13, 191.

[3] Bekenstein, D. 1973, Black Holes and Entropy, Phys. Rev. D7, 2333.

[4] Hooft, G. 1995, Dimensional Reduction in Quantum Gravity.

[5] Susskind, L. 1995, J. Math. Phys., 36, 6377.

[6] Maldacena, J. M. 1998, Int. J. Theor. Phys., 2, 231.

[7] Banados, M., Teitelboim, C., & Zanelli, J. 1992, Phys. Rev. Lett., 69, 1849.

[8] Sa, P., Kleber, A., & Lemos, J. 1996, Class. Quantum Grav., 13, 125.

[9] Son, D., & Starinets, A. 2002, JHEP, 09, 042.

[10] Abdalla, E., de Oliveira, J., & Lima Santos, A. 2012, Phys. Lett. B709, 276.

[11] Kiselev, V. 2003, Class. Quantum Grav., 20, 1187.

[12] Ghaffarnejad, H., Yaraie, E., & Farsam, M. 2018, Int. J. Theor. Phys., 57, 1671.

[13] Ghaffarnejad, H., Yaraie, E., & Farsam, M. 2020, Advances in High Energy Physics, 9529356.

[14] Ghaffarnejad, H., Yaraie, E., & Farsam, M. 2020, Eur. Phys. J. Plus, 135, 179.

[15] Chen, S., Pan, Q., & Jing, J. 2013, Class. Quantum Grav., 14, 145001.

[16] de Oliveira, J., & Fontana, R. 2018, Phys. Rev. D98, 044005.

[17] Banados, M., Teitelboim, C., & Zanelli, J.1992, Phys. Rev. Lett., 69, 1849.

[18] Banados, M., Henneaux, M., Teitelboim, C., & Zanelli, J. 2013, Erratum Phys. Rev. D, 88, 069902.

[19] Matsueda, H., & Suzuki, T. 2017, J. Phys. Soc. Jpn., 86, 104001.

[20] Bircanoglu, C. 2017, www.researchgate.net, publication: 318588371.

[21] Carleo, G., & et al. 2019, Rev. Mod. Phys., 91, 045002.

[22] Yang, Y., Zhao, J., Zhang, J., Ye, X., & Zhao, G. 2020, The Astron. J., 160, 236.

[23] Christopher, J., & Vanderburg, A. 2018, The Astorn. J., 155, 94.

[24] Karniadakis, G. E., Kevrekidis, I. G., & Yang, L. 2021, Nature Rev. Phys., 3, 422.

[25] Matsuoka, D., Watanabe, S., Sato, K., Kawazoe, S., & et al. 2020, Geophysical Research Letters, 47, 1.

[26] Rosenblatt, F. 1958, Psyc. Rev., 65, 386.

[27] Hinton, G. E., & Salakhutdinov, R. R. 2006, Science, 313, 504.

[28] LeCun, Y., Bengio, Y., & Hinton, G. 2015, Nature, 521, 436.

[29] Rawat, W., & Wang, Z. 2017, Neural Comput., 29, 2352.

[30] Yu, D., & Li, J. 2017, IEEE/CAA J. Autom. Sinica, 4, 396.

[31] Young, T., Hazarika, D., Poria, S., & Cambria, E. 2017, IEEE Comput. Intell. Mag., 13, 55.

[32] Cao, C., & et al. 2018, Genomics Proteomics Bioinform., 16, 17.

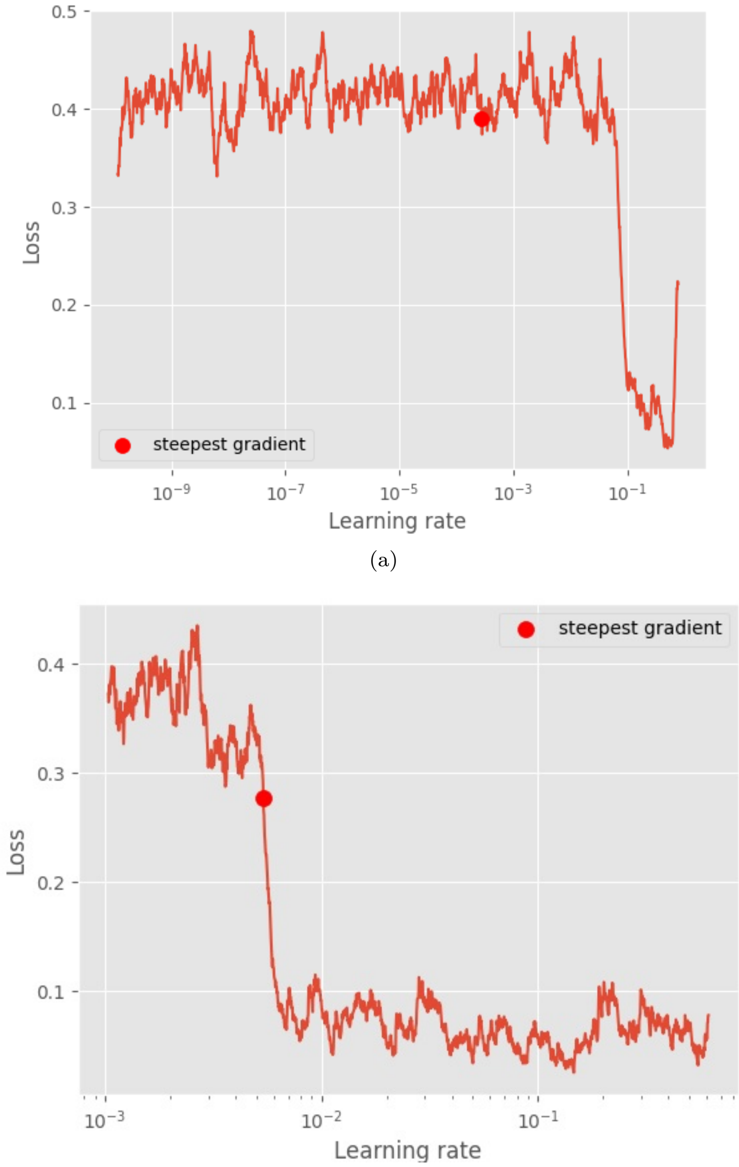
[33] Rumelhart, D. E., Hinton, G. E., Williams, R. J. 1986, Nature. 323, 533.

[34] Paszke, A., Gross, S., Chintala, S., & et al. 2017, Curran Associates.

[35] Kingma, D. P., & Ba, J. 2017, Adam: A Method for Stochastic Optimization, arXiv:1412.6980 [cs.LG].

[36] Anzanello, M., & Fogliatto, F. 2011, Int. J. of Indus. Ergo., 41, 573.

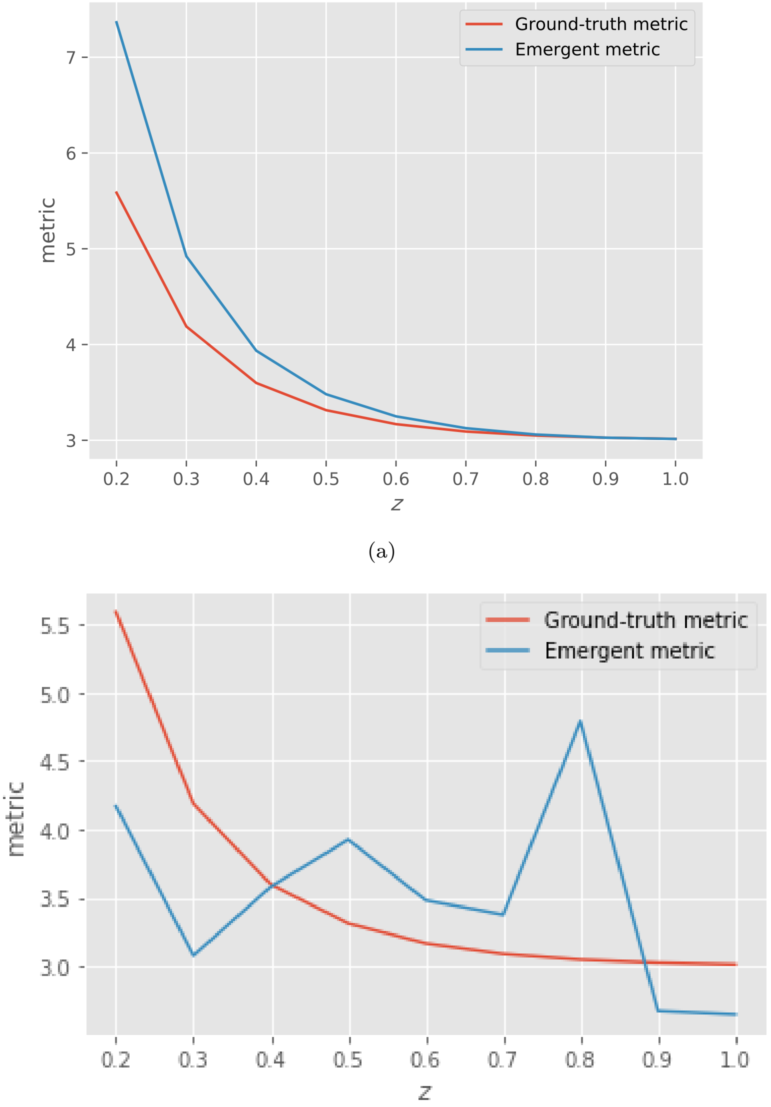
[37] Leslie, S. N. 2017, Cyclical learning rates for training Neural Networks, arXiv:1506.01186 [cs.CV].



(a)

(b)

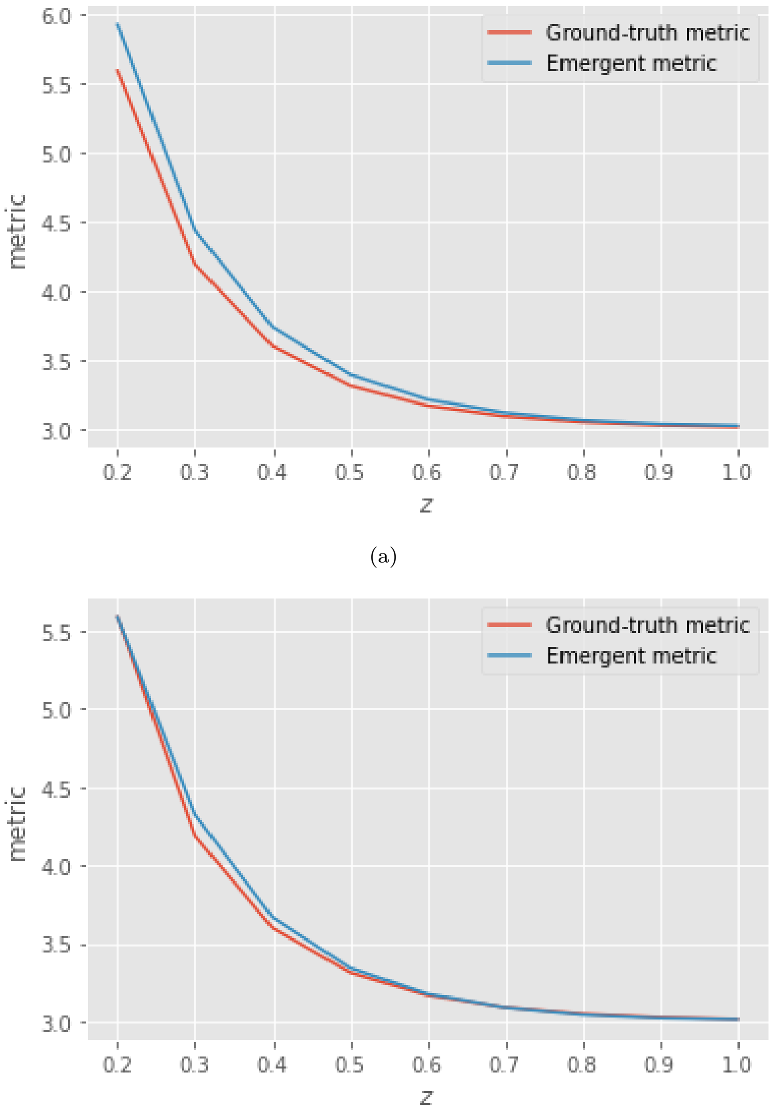
Figure 4: Behavior of the loss function versus the learning rates: (a) is plotted for non-quintessence ($w = 0$) which shows for small learning rates the iterations become large and so minimum of the loss function is happened at long times. (b) is plotted for quintessence ($w = -\frac{1}{2}, -\frac{3}{4}$) and it shows for large learning rates the loss function pass far from the observed steepest gradient point (red dot) and so the ADAM optimizer does not never obtain minimum value for the loss function. But if we choose best value for the learning rates equal to the steepest gradient point the loss function reaches to its minimum value as soon.



(a)

(b)

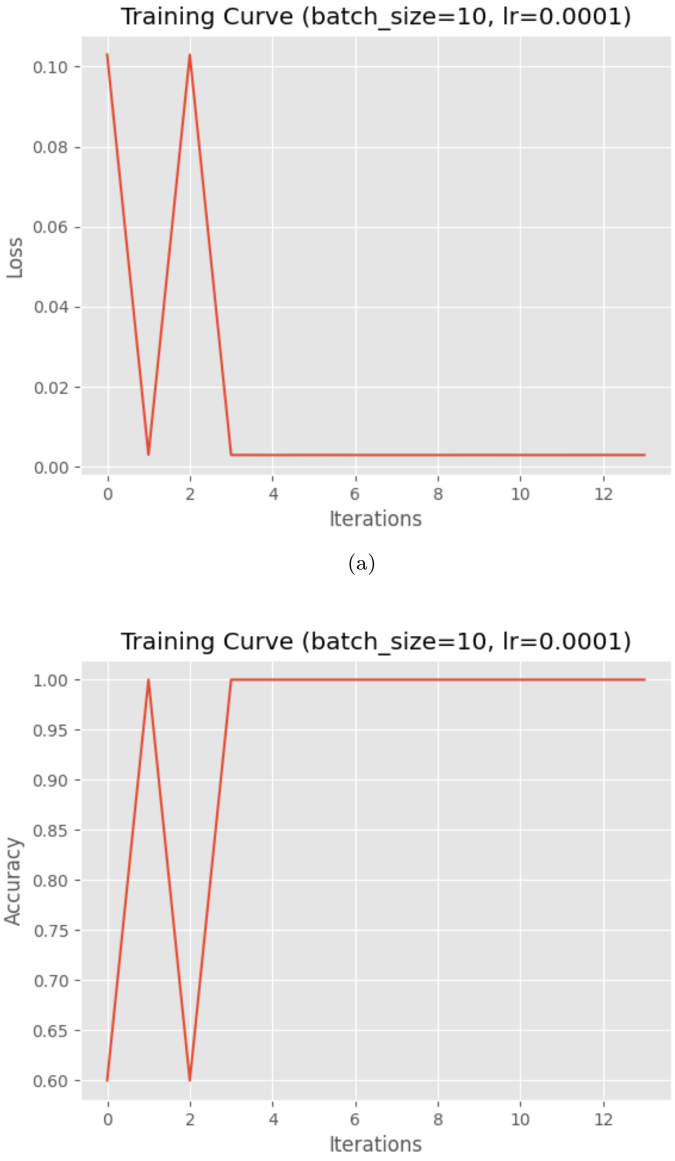
Figure 5: The emerged metric and the ground-truth metric has been portrayed with and without the penalty term respectively at (a) and (b) after 10000 epochs for non quintessence case $w = 0$.



(a)

(b)

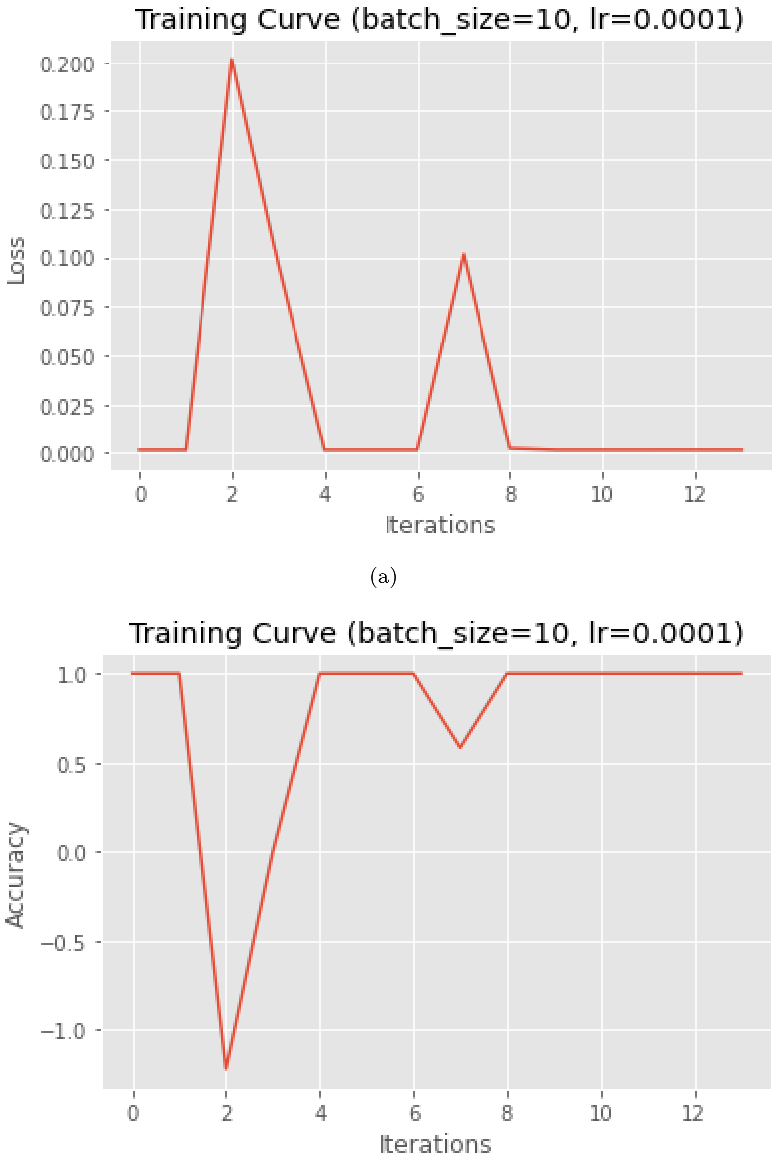
Figure 6: For case of non quintessence $w = 0$, the emerged metric and the ground-truth metric penalty term has been portrayed at (a) and (b) after 30000 and 50000 epochs respectively.



(a)

(b)

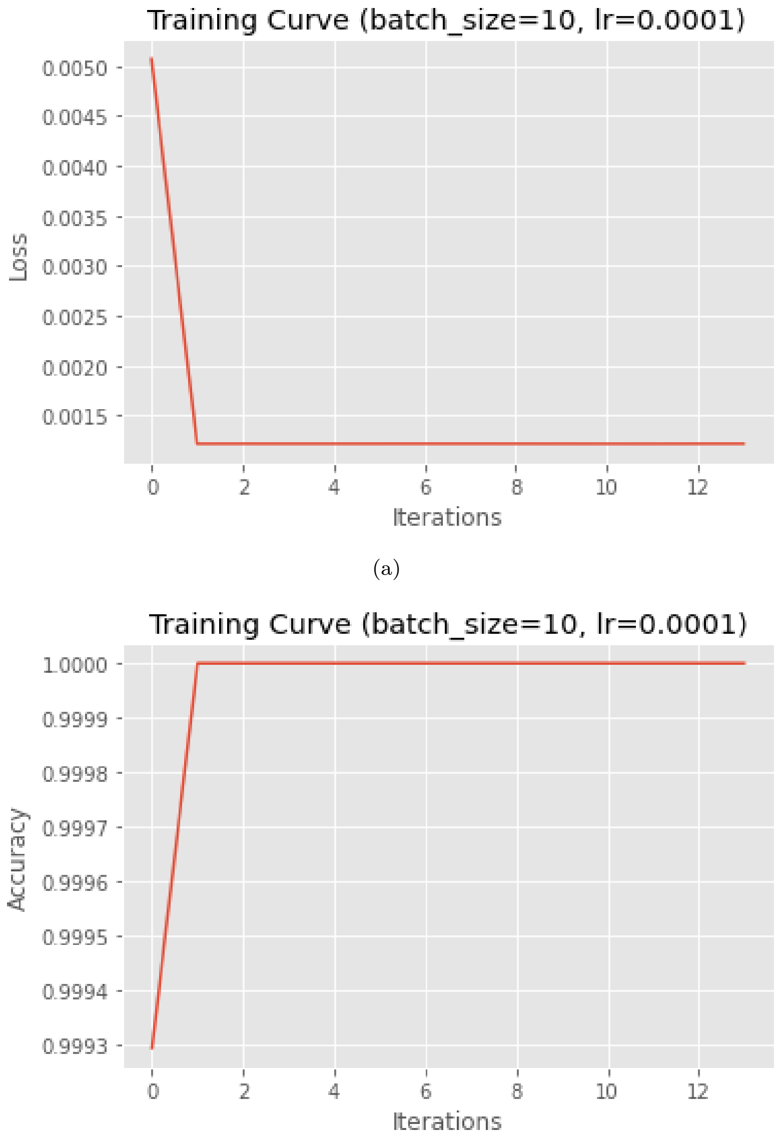
Figure 7: The loss and statistical R-Squared accuracy over 10000 epochs. Each iteration corresponds to the number of epochs to be over 15. This diagrams are plotted for non quintessence case $w = 0$.



(a)

(b)

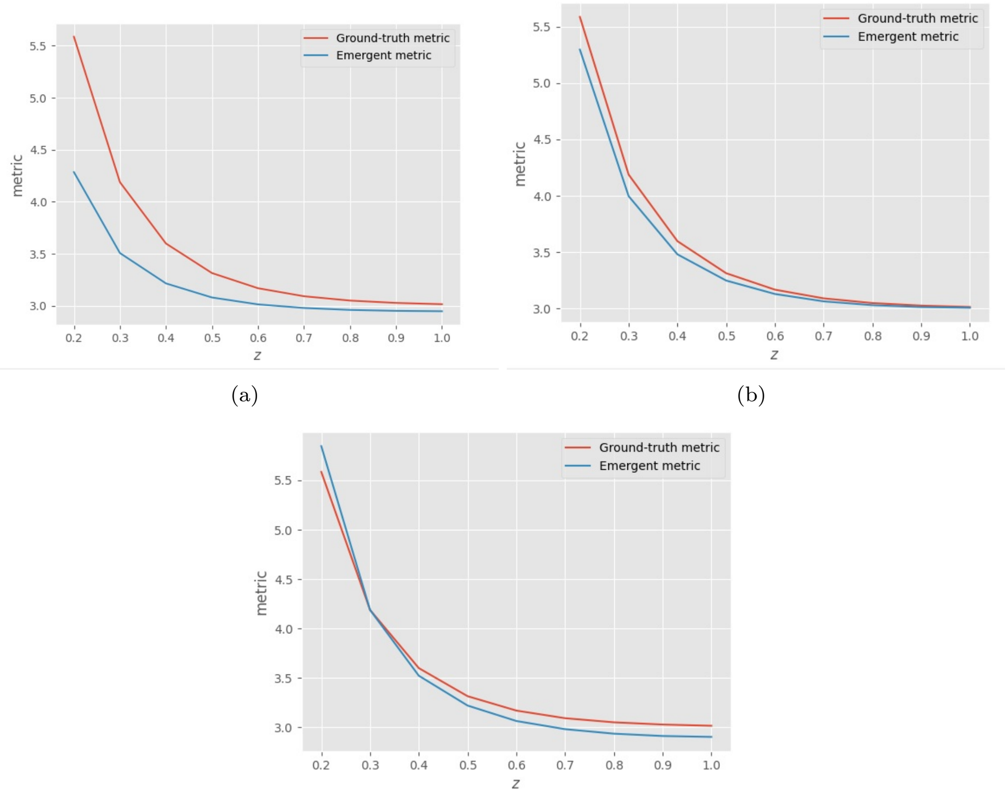
Figure 8: The loss and statistical R-Squared accuracy over 30000 epochs. Each iteration corresponds to the number of epochs to be over 15. This diagrams are plotted for non quintessence case $w = 0$.



(a)

(b)

Figure 9: The loss and statistical R-Squared accuracy over 50000 epochs. Each iteration corresponds to the number of epochs to be over 15. This diagrams are plotted for non quintessence case $w = 0$.

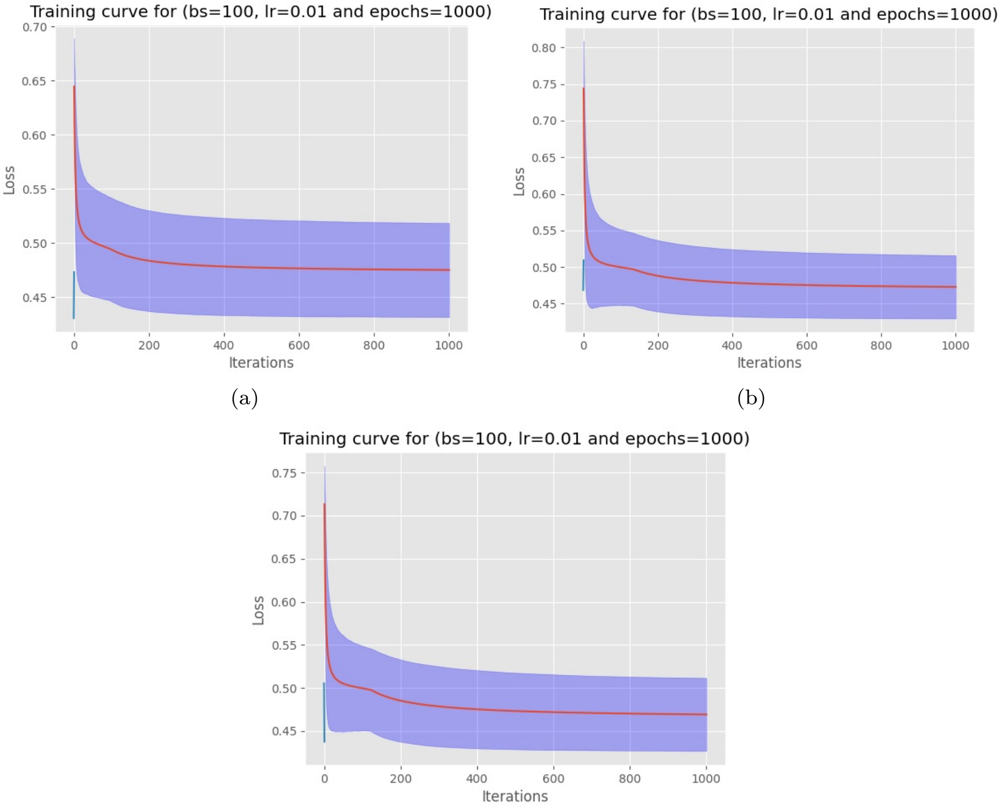


(a)

(b)

(c)

Figure 10: The learned metric with 1000 epochs with learning rate 0.01 with quintessence case $w = -\frac{1}{2}$ for different activation functions of last neurons given by the equation (21) at (a) for $Q = 0.6$, (b) for $Q = 0.9$ and (c) for $Q = 1.1$ respectively.



Training curve for (bs=100, lr=0.01 and epochs=1000)

(a)

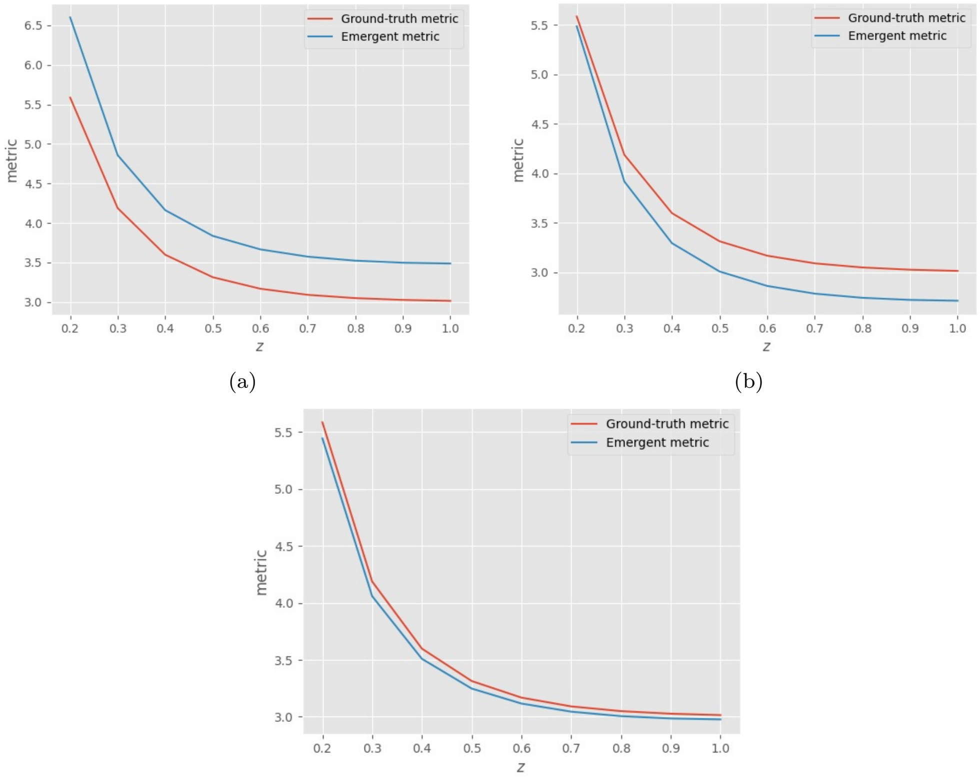
Training curve for (bs=100, lr=0.01 and epochs=1000)

(b)

Training curve for (bs=100, lr=0.01 and epochs=1000)

(c)

Figure 11: The behavior of the loss functions with confidence interval 95 percent for 1000 epochs with learning rate 0.01 and quintessence $w = -\frac{1}{2}$ for different activation functions of last neurons given by the equation (21): (a) for $Q = 0.6$, (b) for $Q = 0.9$ and (c) for $Q = 1.1$ respectively.

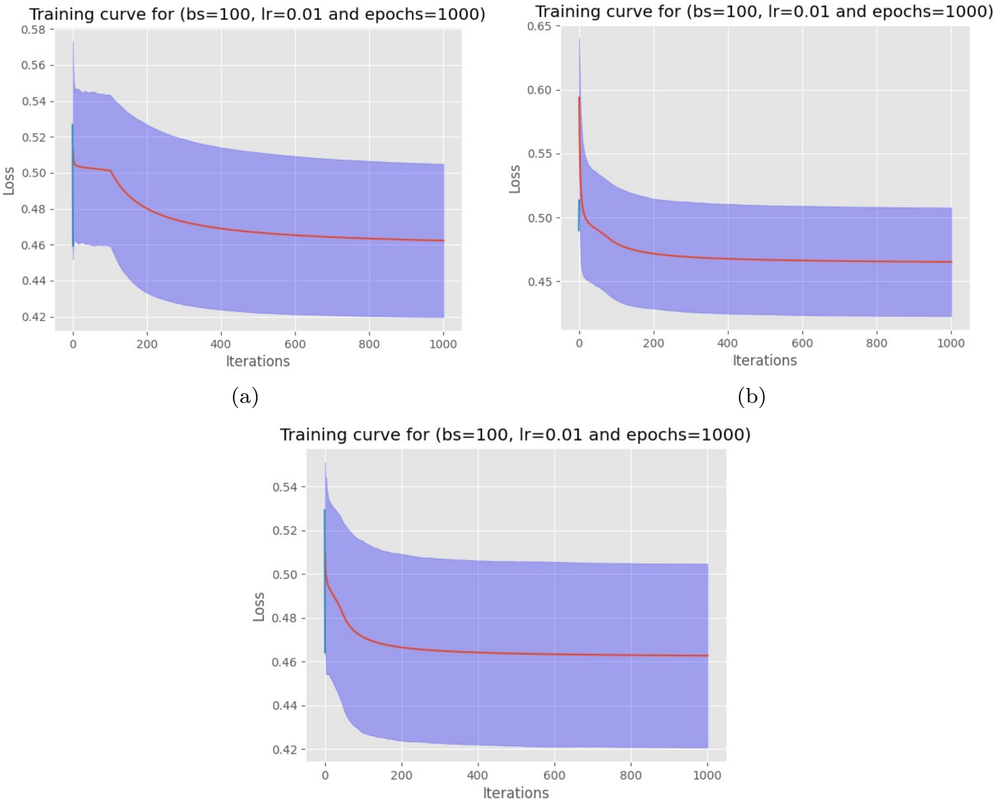


(a)

(b)

(c)

Figure 12: The learned metric with 1000 epochs with learning rate 0.01 for quintessence $w = -\frac{3}{4}$ for different activation functions of last neurons given by the equation (22) at (a) for $P = 0.6$, (b) for $P = 0.8$ and (c) for $P = 0.9$ respectively.



(a)

(b)

(c)

Figure 13: The behavior of the loss functions with confidence interval 95 percent for 1000 epochs with learning rate 0.01 and quintessence $w = -\frac{3}{4}$ for different activation functions of last neurons given by the equation (22): (a) for $P = 0.6$, (b) for $P = 0.8$ and (c) for $P = 0.9$ respectively.