

Solving differential equations with neural networks: Applications to the calculation of cosmological phase transitions

Maria Laura Piscopo,^{*} Michael Spannowsky,[†] and Philip Waite[‡]

*Institute for Particle Physics Phenomenology, Department of Physics,
Durham University, Durham DH1 3LE, United Kingdom*



(Received 2 May 2019; published 9 July 2019)

Starting from the observation that artificial neural networks are uniquely suited to solving optimization problems, and most physics problems can be cast as an optimization task, we introduce a novel way of finding a numerical solution to wide classes of differential equations. We find our approach to be very flexible and stable without relying on trial solutions, and applicable to ordinary, partial and coupled differential equations. We apply our method to the calculation of tunneling profiles for cosmological phase transitions, which is a problem of relevance for baryogenesis and stochastic gravitational wave spectra. Comparing our solutions with publicly available codes which use numerical methods optimized for the calculation of tunneling profiles, we find our approach to provide at least as accurate results as these dedicated differential equation solvers, and for some parameter choices, even more accurate and reliable solutions. In particular, we compare the neural network approach with two publicly available profile solvers, CosmoTransitions and BubbleProfiler, and give explicit examples where the neural network approach finds the correct solution while dedicated solvers do not. We point out that this approach of using artificial neural networks to solve equations is viable for any problem that can be cast into the form $\mathcal{F}(\vec{x}) = 0$, and is thus applicable to various other problems in perturbative and nonperturbative quantum field theory.

DOI: [10.1103/PhysRevD.100.016002](https://doi.org/10.1103/PhysRevD.100.016002)

I. INTRODUCTION

A neural network is an algorithm designed to perform an optimization procedure, where the loss function provides a measure of the performance of the optimization. Thus, if a physics problem can be cast into the form $\mathcal{F}(\vec{x}) = 0$, then its solution can be calculated by minimizing the loss function of a neural network. While this approach is applicable to any function \mathcal{F} , we attempt to apply this observation to the solution of differential equations and to the nonperturbative calculation of tunneling rates of electroweak phase transitions.

Solving differential equations is a profound problem, relevant for all areas of theoretical physics. For large classes of differential equations, analytic solutions cannot be found. Thus, numerical or approximative methods are needed to solve them. Standard methods to solve differential equations numerically include the Runge-Kutta

method, linear multistep methods, finite-element or finite-volume methods, and spectral methods [1]. We instead propose a novel approach to solving differential equations using artificial neural networks.

In recent years, machine-learning algorithms have become increasingly popular in extracting correlations in high-dimensional parameter spaces. Even for a small number of dimensions, e.g., $n_{\text{dim}} \geq 3$, it becomes very difficult to visualize data such that a human can extract correlations to a high degree of accuracy. Machine-learning algorithms, and in particular neural networks, prove to be faster and more precise and allow a parametric improvement of the precision in how well the region of interest is interpolated. As a result, various neural network architectures have been designed, e.g., convolutional neural networks, recurrent neural networks, deep neural networks, etc., to perform various increasingly complex tasks.

In particle physics such tasks include the classification of signal-to-background events based on event selection cuts [2–5], or the classification of complex objects such as jets, according to the image their radiation imprints in the detector [6–14]. In other applications neural networks are used to regress between data points [15–19] or are trained on a well-known sample to identify outliers or anomalies [20–22]. However, in all aforementioned applications that can be characterized as classification and regression, the neural network is applied to an output sample, trying to extract

^{*}maria.l.piscopo@durham.ac.uk

[†]michael.spannowsky@durham.ac.uk

[‡]p.a.waite@durham.ac.uk

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

information on the parameters that determine the input. In particle physics that would mean to analyze the radiation profile as recorded by a particle detector to learn the parameters of the underlying model, e.g., the Standard Model. Input and output are connected through quantum field theory, i.e., a nontrivial set of differential and integral equations.

We propose to use these powerful artificial neural network algorithms in a different way, namely to directly find solutions to differential equations. We then apply these methods to calculate the solution of the nonperturbative quantum-field-theoretical description of tunneling processes for electroweak phase transitions. The fast and reliable calculation of tunneling rates of the electroweak phase transitions within and in extensions of the Standard Model is of importance to assessing if the model allows for a strong first-order phase transition during the evolution of the early Universe. This could explain baryogenesis [23,24] in such a model as the source of matter-antimatter asymmetry in the Universe, and further lead to a stochastic gravitational wave signal which could potentially be measured at future gravitational wave experiments [25,26], e.g., eLISA [27].

The universal approximation theorem [28,29] allows us to expect a neural network to perform well in solving complicated mathematical expressions. It states that an artificial neural network containing a single hidden layer can approximate any arbitrarily complex function with enough neurons. We make use of this property by proposing a neural network model where the output of the network alone solves the differential equation, subject to its boundary conditions. In contrast to previous approaches [30–35] where the neural network is part of a full trial solution which is fixed to satisfy the boundary conditions, our approach includes the boundary conditions as additional terms in the loss function. The derivatives of the network output with respect to its inputs are calculated and passed to the loss function, and the network is optimized via back-propagation to regress to the solution of the differential equation. The network then gives a fully differentiable function which can be evaluated at any point within the training domain, and in some cases, extrapolated to further points (although we do not explore the extrapolation performance here).

We will begin by describing the method in detail and showcasing how it can be used to solve differential equations of varying complexity, before applying it to the calculation of cosmological phase transitions.

II. THE METHOD

A. Design of the network and optimization

We consider an artificial feedforward neural network (NN) with n inputs, m outputs and a single hidden layer with k units. The outputs of the network, $N_m(\vec{x}, \{w, \vec{b}\})$, can be written as

$$N_m(\vec{x}, \{w, \vec{b}\}) = \sum_{k,n} w_{mk}^f g(w_{kn}^h x_n + b_k^h) + b_m^f, \quad (1)$$

where the activation function $g: \mathbf{R}^k \mapsto \mathbf{R}^k$ is applied element-wise to each unit, and h and f denote the hidden and final layers, respectively. We use a single neural network with m outputs to predict the solutions to m coupled differential equations, and for the case of one differential equation, we use $m = 1$.

A set of m coupled j th order differential equations can be expressed in the general form,

$$\mathcal{F}_m(\vec{x}, \phi_m(\vec{x}), \nabla \phi_m(\vec{x}), \dots, \nabla^j \phi_m(\vec{x})) = 0, \quad (2)$$

with boundary or initial conditions imposed on the solutions $\phi_m(\vec{x})$. Writing the differential equations in such a way allows us to easily convert the problem of finding a solution into an optimization one. An approximate solution $\hat{\phi}_m(\vec{x})$ is one which approximately minimizes the square of the left-hand side of Eq. (2), and thus the analogy can be made to the loss function of a neural network. In previous approaches [30–33], $\hat{\phi}_m(\vec{x})$ is a trial solution composed of two parts: one which satisfies the boundary conditions, and one which is a function of the output of a neural network and vanishes at the boundaries. However, this requires one to choose a special form of the trial solution which is dependent on the boundary conditions. Furthermore, for some configurations of boundary conditions, finding such a trial solution is a very complex task, e.g., in the case of phase transitions. Instead, we identify the trial solution with the network output, $\hat{\phi}_m(\vec{x}) \equiv N_m(\vec{x}, \{w, \vec{b}\})$, and include the boundary conditions as extra terms in the loss function. If the domain is discretized into a finite number of training points \vec{x}^i , then approximations to the solutions, $\hat{\phi}_m(\vec{x})$, can be obtained by finding the set of weights and biases, $\{w, \vec{b}\}$, such that the neural network loss function is minimized on the training points. For i_{\max} training examples, the full loss function that we use is

$$\begin{aligned} \mathcal{L}(\{w, \vec{b}\}) = & \frac{1}{i_{\max}} \sum_{i,m} \hat{\mathcal{F}}_m(\vec{x}^i, \hat{\phi}_m(\vec{x}^i), \dots, \nabla^j \hat{\phi}_m(\vec{x}^i))^2 \\ & + \sum_{\text{B.C.}} (\nabla^p \hat{\phi}_m(\vec{x}_b) - K(\vec{x}_b))^2, \end{aligned} \quad (3)$$

where the second term represents the sum of the squares of the boundary conditions, defined at the boundaries \vec{x}_b .¹ These can be Dirichlet or Neumann, or they can be initial conditions if defined at the initial part of the domain.

¹Here, p represents the order of derivative for which the boundary condition is defined, and K is a function on the boundary. For example, for the condition $\frac{d}{dx}\phi(0) = 1$ the second term would be $(\frac{d}{dx}\hat{\phi}(0) - 1)^2$.

The problem is then to minimize $\mathcal{L}(\{w, \vec{b}\})$ by optimizing the weights and biases in the network, for a given choice of network setup. To calculate the loss, it is necessary to compute the derivatives of the network output with respect to its input. Since each part of the network, including the activation functions, are differentiable, then the derivatives can be obtained analytically. Reference [33] outlines how to calculate these derivatives. The optimization can then proceed via backpropagation by further calculating the derivatives of the loss itself with respect to the network parameters. We use the Keras framework [36] with a TensorFlow [37] backend to implement the network and perform the optimization of the loss function.

As with any neural network, the choice of hyperparameters will have an effect on the performance. For our setup, the important parameters are the number of hidden layers, the number of units in each hidden layer, the number of training points $\vec{x}^{(i)}$ (corresponding to the number of anchors in the discretization of the domain of the differential equation), the activation function in each hidden layer, the optimization algorithm, the learning rate, and the number of epochs the network is trained for. Furthermore, a choice must be made for the size of the domain that contains the points that the network is trained on, but this will usually be determined by the problem being solved.

In all the examples, we use the Adam optimizer [38] with learning rate reduction on plateau—i.e., when the loss plateaus, the learning rate is reduced—and an initial learning rate of 0.01. We find that the network is not susceptible to overfitting—the training points are chosen exactly from the domain that one is trying to find the solution to, and are not subject to statistical fluctuations; thus, finding a solution for which the loss at every training point is zero would not limit the generalization of the solution to other points within the domain. Therefore, we use a large number of epochs such that the training loss becomes very small. For all examples we use a conservative number of 5×10^4 epochs. Furthermore, we use the entire set of training points in each batch so that the boundary conditions in the loss are included for each update of the network parameters. We also find that, in general, a single hidden layer with a small number of units [$\mathcal{O}(10)$] is sufficient to obtain very accurate solutions.

In order to assess and improve the stability and performance in certain cases, there are some additional technical methods which we employ beyond the basic setup. First, the differentiability of the network solution allows us to calculate the differential contribution, $\hat{\mathcal{F}}$, to the loss across the entire training domain. This shows the degree of accuracy to which each part of the network solution satisfies the differential equation, and can be used for assessing the performance in cases where the analytic solution is not known. Second, for coupled differential equations with initial conditions, we find that the stability of the solution can be improved by iteratively training on

increasing domain sizes. Finally, for the calculation of phase transitions, we employ a two-step training where initially the boundaries are chosen to be the true and false vacua, before the correct boundary conditions are used in the second training. This prevents the network from finding the trivial solution where the field is always in the false vacuum.

B. Ordinary differential equation examples

To show how well the method can solve ordinary differential equations (ODEs), we apply it to both a first and a second order ODE, which have known analytic solutions. The equations we study are

$$\frac{d\phi}{dx} + \left(x + \frac{1 + 3x^2}{1 + x + x^3} \right) \phi - x^3 - 2x - x^2 \frac{1 + 3x^2}{1 + x + x^3} = 0, \quad (4)$$

with the boundary condition $\phi(0) = 1$ in the domain $x \in [0, 2]$, and

$$\frac{d^2\phi}{dx^2} + \frac{1}{5} \frac{d\phi}{dx} + \phi + \frac{1}{5} e^{-\frac{x}{5}} \cos x = 0, \quad (5)$$

with boundary conditions $\phi(0) = 0$ and $\frac{d}{dx}\phi(0) = 1$ in the domain $x \in [0, 2]$.

As a simple neural network structure, we choose a single hidden layer of 10 units with sigmoid activation functions, and we discretize the domain into 100 training examples. It is then just a case of passing the differential equations and boundary conditions to the loss function, as described in Eq. (3), and proceeding with the optimization. Figure 1 shows the results of the neural network output, compared to the analytic solutions of Eqs. (4) and (5). The middle panel of Fig. 1 shows the absolute numerical difference between the numerical and analytic solutions. This difference can be reduced further by increasing the number of epochs, the number of anchors in the discretization of the domain, or the number of units in the hidden layer. Thus, the neural network provides handles to consistently improve the numerical accuracy one aims to achieve.

The lower panel of Fig. 1 shows the differential contribution to the loss function, i.e., how much each training example contributes to the loss. As we will describe in the next section, if the solution is not analytically known, this provides a measure to assess whether the found solution is the correct one or if a numerical instability led the network to settle in a local minimum for the loss.

C. Coupled differential equation example

When discussing the calculation of cosmological phase transitions, we will study the solution of coupled nonlinear differential equations, for which no closed analytic form is known. Here, we will first show that such solutions can be

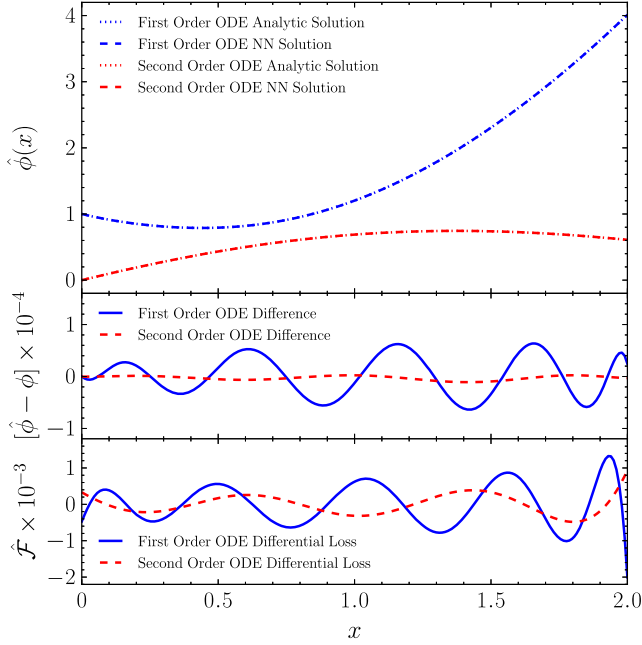


FIG. 1. The upper panel shows the solutions to the first and second order ODEs of Eqs. (4) and (5), with boundary conditions as outlined in the text. The middle panel shows the numerical difference between the analytic solution and the NN predicted solution for both cases. The lower panel shows the differential contribution $\hat{\mathcal{F}}$ to the loss across the entire domain.

obtained with our approach, for a case where an analytic solution is known. We consider

$$\begin{aligned} \frac{d\phi_1}{dx} - \cos x - \phi_1^2 - \phi_2 + 1 + x^2 + \sin^2 x &= 0, \\ \frac{d\phi_2}{dx} - 2x + (1 + x^2) \sin x - \phi_1 \phi_2 &= 0, \end{aligned} \quad (6)$$

with boundary conditions

$$\phi_1(0) = 0, \quad \phi_2(0) = 1. \quad (7)$$

If the boundary conditions are set on one end of the domain, e.g., here at $x = 0$, it requires an increasingly elaborate network to maintain numerical stability for the solution over a large domain, e.g., where $x \gg 1$. This is due to small numerical instabilities during backpropagation because of the complexity of the loss hypersurface. If such numerical instability leads the network to choose a path that is in close proximity to the true solution, the NN can settle on a local minimum with a small value for the loss function. To solve this problem, we propose to incrementally extend the domain on which a solution should be found, by partitioning the training examples and increasing the number of partitions the NN is trained on in each step. If the weights the NN has learned in the previous step are then retained before training for the next step—i.e., the network only has to learn the function on the part of the

domain that was incrementally increased—we find that one can achieve numerical stability for an arbitrarily large domain.

We show this mechanism in Fig. 2, where we have partitioned the full domain containing 100 training examples into three regions each of size 1. The network structure again consists of a single hidden layer of 10 units with sigmoid activation functions, and with two units in the final layer, since there are two coupled equations. The upper panel shows the solutions for ϕ_1 and ϕ_2 for each iterative step. While the first iteration only allows a solution to be found on a smaller domain, i.e., here from 0 to 1, subsequent steps, and in particular the third step, allow an accurate solution to be found over the entire domain. Again, the differential $\hat{\mathcal{F}}$ proves to be a good indicator of whether the calculated solution is satisfying the differential equation over the entire domain (see the lower panel of Fig. 2).

D. Partial differential equation example

While we do not study partial differential equations (PDEs) in the later physics examples of calculating phase transitions, we showcase here the flexibility of our NN method. With the same network architecture as used for the

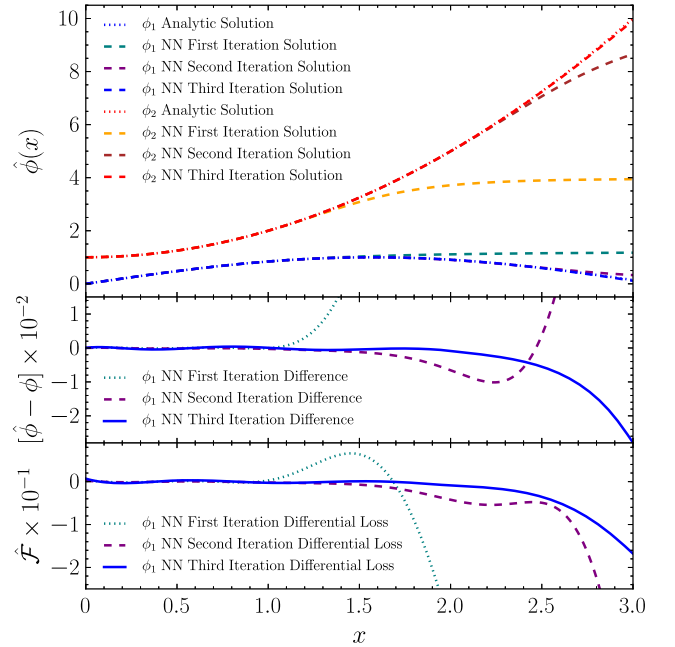


FIG. 2. The upper panel shows the solutions for the functions ϕ_1 and ϕ_2 to the coupled differential equation of Eq. (6). The middle panel displays the numerical difference between the analytic solution and the NN predicted solution for ϕ_1 . The lower panel shows the differential contribution $\hat{\mathcal{F}}$ to the loss across the entire domain, from the equation for ϕ_1 . The three NN curves in each panel correspond to the first, second and third iteration steps in the training of the network, with iterative increase of the training domain, as described in the text.

solution of the ordinary differential equations (except for an extra input unit for each additional variable), we can apply our approach to the solution of partial differential equations. The precise solution of such equations is a widespread problem in physics, e.g., in mechanics, thermodynamics and quantum field theory. As an example, we choose the second order partial differential equation,

$$\nabla^2 \phi - e^{-x}(x - 2 + y^3 + 6y) = 0, \quad (8)$$

with boundary conditions

$$\begin{aligned} \phi(0, y) &= y^3, & \phi(1, y) &= (1 + y^3)e^{-1}, \\ \phi(x, 0) &= xe^{-x}, & \phi(x, 1) &= e^{-x}(x + 1), \end{aligned} \quad (9)$$

for which an exact analytic solution is known. In Fig. 3 we show the difference between the numerical solution as predicted by the NN and the analytic solution over the domain $(x, y) \in [0, 1] \times [0, 1]$. The 100 training examples were chosen from an evenly spaced 10×10 grid. As the value of $\phi(x, y)$ is of $\mathcal{O}(1)$ for most of the domain, the relative and absolute accuracies are similar, so we only show the absolute accuracy here. Across the entire domain, we find a numerical solution with very good absolute and relative accuracy for this second order partial differential equation. However, with a deeper NN, e.g., a second layer with 10 tanh units, we find that the accuracy improves by an order of magnitude further. Deeper and wider networks result in even better accuracies.

III. CALCULATION OF PHASE TRANSITIONS DURING THE EARLY UNIVERSE

Electroweak baryogenesis is a candidate for solving the baryon asymmetry puzzle, the observed abundance of

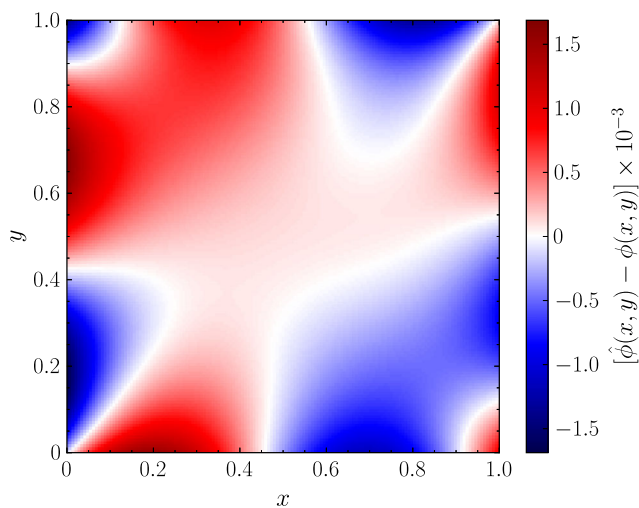


FIG. 3. Numerical difference between the analytic solution and the NN predicted solution of Eq. (8), with boundary conditions as given in Eq. (9), over the domain $(x, y) \in [0, 1] \times [0, 1]$.

matter over antimatter in the Universe [24,39]. The need for a dynamical generation of baryon asymmetry is dictated by inflation—the entropy production occurring in the reheating phase of inflation would have washed out any asymmetry already present at the beginning of the Universe’s evolution [40]. A model of baryogenesis was proposed by Sakharov in 1967 [23], and must now be accommodated by any fundamental theory capable of addressing the baryon asymmetry problem. This is commonly translated into three necessary conditions: (i) baryon number violation, (ii) C - and CP -violation, and (iii) loss of thermal equilibrium. While the first condition can be satisfied in the SM, the second and third conditions require it to be extended [41–43]. Departure from thermal equilibrium can be obtained during a strong first-order phase transition, which is usually accompanied by a sudden change of symmetry [44]. Within the SM, this could have occurred during electroweak symmetry breaking when the Universe had the temperature $T \sim 100$ GeV [45,46]. In order to assess whether this might have been the case, it is crucial to discuss the conditions for scalar-field phase transitions at finite temperature.

Quantum fluctuations allow the transition between two vacua of the potential $V(\vec{\phi})$.² When these are not degenerate, the configuration which corresponds to a local minimum, the false vacuum $\vec{\phi}_F$, becomes unstable under barrier penetration, and can decay into the true vacuum $\vec{\phi}_T$ of the potential. The tunneling process converts a homogeneous region of false vacuum into one of approximate true vacuum—a bubble. Far from this region the false vacuum persists undisturbed [47]. The Euclidean action for this process reads

$$\mathcal{S}_4(\vec{\phi}) = \int d\tau d^3x \left[\frac{1}{2} \left(\frac{d\vec{\phi}}{d\tau} \right)^2 + \frac{1}{2} (\nabla \vec{\phi})^2 + V(\vec{\phi}) \right]. \quad (10)$$

The description of the tunneling action at finite temperatures follows from the equivalence between the quantum statistics of bosons (fermions) at $T \neq 0$ and Euclidean quantum field theory, periodic (antiperiodic) in the Euclidean time τ with period T^{-1} . In the calculation of $\mathcal{S}_4(\vec{\phi})$, the integration over τ is replaced by multiplication by T^{-1} [48], leaving the three-dimensional Euclidean action

$$\mathcal{S}_3(\vec{\phi}) = \int d^3x \left[\frac{1}{2} (\nabla \vec{\phi})^2 + V(\vec{\phi}, T) \right], \quad (11)$$

with $\mathcal{S}_4(\vec{\phi}) = T^{-1} \mathcal{S}_3(\vec{\phi})$. Suggested by the symmetry of the physical problem, we assume $\vec{\phi}(\vec{x})$ to be invariant under

²Without loss of generality we consider an n -dimensional real scalar field $\vec{\phi}(\vec{x})$.

three-dimensional Euclidean rotations (see Ref. [49] for a rigorous demonstration in the case of one single scalar field), and define $\rho = \sqrt{x^2}$. The bubble configuration $\vec{\phi}_b(\rho)$ is the solution to the Euler-Lagrange equation of motion,

$$\frac{d^2 \vec{\phi}}{d\rho^2} + \frac{2}{\rho} \frac{d\vec{\phi}}{d\rho} = \nabla V, \quad (12)$$

where the gradient of the potential is with respect to the field $\vec{\phi}$. The boundary conditions are

$$\frac{d}{d\rho} \vec{\phi}(0) = 0, \quad \lim_{\rho \rightarrow \infty} \vec{\phi}(\rho) = \vec{\phi}_F. \quad (13)$$

The solution thus minimizes the action. The probability per unit time and unit volume for the metastable vacuum to decay is given by

$$\frac{\Gamma}{V} = A e^{-B/T}. \quad (14)$$

This is maximized by the bounce,

$$B = \mathcal{S}_3(\vec{\phi}_b) - \mathcal{S}_3(\vec{\phi}_F), \quad (15)$$

where $\mathcal{S}_3(\vec{\phi}_F)$ is the action evaluated at the stationary configuration $\vec{\phi}_F$. A complete expression for the factor A in Eq. (14) would require complex computations of differential operator determinants, for which we refer the reader to Ref. [40]. An estimate can be obtained from dimensional analysis, which gives $A \sim T^4$ [50].

Dedicated methods for calculating the nucleation rate, by finding a solution for the bubble profile $\vec{\phi}$ to the nonlinear coupled differential equations of Eq. (12), exist and have been implemented in publicly available codes, e.g., CosmoTransitions [51] and BubbleProfiler [52]. For the single-field case, both CosmoTransitions and BubbleProfiler use variants of the overshooting and undershooting method. In the multiple-field case, BubbleProfiler applies the Newton-Kantorovich method [53], as described in [54]. CosmoTransitions instead uses a method that splits the equation of motion into a parallel and perpendicular component along a test path through field space. Then the path is varied until a configuration is found that simultaneously solves both directions of the equations of motion. A further code to calculate the tunneling rates is given in Ref. [55]. An approach using neural networks to directly learn bounce actions from potentials was described in Ref. [56]. Recently, a novel approximative approach for single [57] and multiple fields [58] was proposed, and a new method based on exact analytic solutions of piecewise linear potentials is outlined in Ref. [59]. Older numerical approaches to calculating bubble profiles and tunneling rates include Refs. [60–63].

A. Phase transition with a single scalar field

As a first application of our method to the computation of cosmological phase transitions, we consider the case of a single scalar field. Equation (12) then has a straightforward classical analogy—it describes the motion of a particle with coordinate $\phi(\rho)$ subject to the inverted potential $-V(\phi)$ and to a peculiar looking damping force which decreases with time. The problem reduces to finding the initial position ϕ_0 , in the vicinity of ϕ_T , such that the particle stops at ϕ_F as $\rho \rightarrow \infty$.

Existence of a solution was proven in Ref. [47]. Starting too close or too far from ϕ_T would lead to missing the final configuration ϕ_F , due to overshooting and undershooting, respectively. Continuity of $\phi(\rho)$ thus implies that there must exist an intermediate initial position ϕ_0 which solves the boundary conditions in Eq. (13). The solution presents two limiting profiles, determined by the ratio of $\Delta \equiv V(\phi_F) - V(\phi_T)$ to the height of the potential barrier $V(\phi_{\text{bar}})$. If this ratio is $\gtrsim 1$, which corresponds to the thick-wall case, the particle will overshoot unless its initial energy is similar to $V(\phi_F)$. Conversely, if this ratio is small, corresponding to the thin-wall case, in order to avoid undershooting, the particle must wait close to ϕ_T until the time $\rho \approx R$, when the damping force has become negligible. The value of R can be determined exactly in the thin-wall limit [47] using

$$R = \frac{2\sigma}{\Delta}, \quad (16)$$

where the surface tension σ is given by

$$\sigma = \lim_{\Delta \rightarrow 0} \int_{\phi_F}^{\phi_T} d\phi \sqrt{2[V(\phi) - V(\phi_F)]}. \quad (17)$$

We test our method on the potential [52],

$$V(\phi) = \frac{\lambda}{8}(\phi^2 - a^2)^2 + \frac{\epsilon}{2a}(\phi - a), \quad (18)$$

and set $\lambda = a = 1$. Two distinct and nondegenerate minima exist for $0 < \epsilon \lesssim 0.3$, with the upper bound representing the thick-wall limit and smaller values of ϵ representing progressively thinner cases. A plot of the potential is shown in Fig. 4, for the values $\epsilon = 0.01$ and $\epsilon = 0.3$ which we consider as our thin-wall and thick-wall cases, respectively.

For the boundary conditions in Eq. (13), it is clearly not possible to implement an infinite domain for the training of a neural network, and the divergence in the second term of Eq. (12) prevents the equation from being evaluated at $\rho = 0$. Therefore, a training domain $\rho \in [\rho_{\min}, \rho_{\max}]$ must be chosen. Since the solution approaches the boundaries exponentially, it can be safely assumed that the numerical uncertainties induced by these choices can be neglected, provided that ρ_{\min} is sufficiently small and ρ_{\max} is

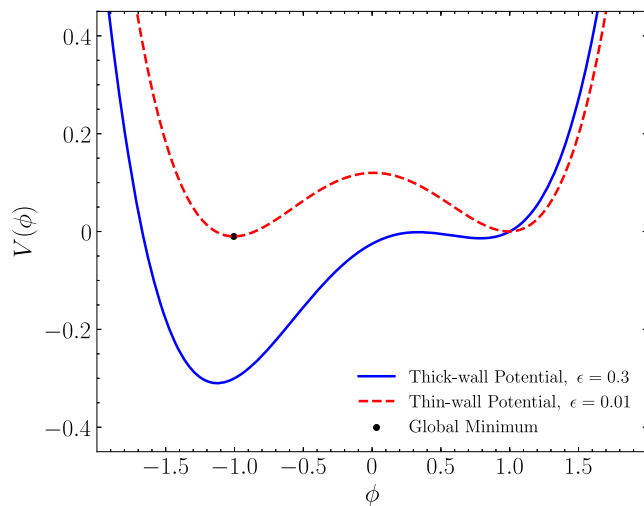


FIG. 4. Plot of the potential in Eq. (18), with $\lambda = \alpha = 1$, for the thick-wall (blue solid) and the thin-wall (red dashed) cases. For the latter, the position of the global minimum is also marked by the black dot for clarity.

sufficiently large. To help in choosing this domain, the identification of ϵ in Eq. (18) with Δ in Eq. (16) can be made, and the bubble radius R calculated. We then use $\rho_{\max} = 5R$ for the thick-wall case, and $\rho_{\max} = 2R$ for the thin-wall case [since Eq. (16) underestimates the true radius for thick-wall cases]. Furthermore, we use $\rho_{\min} = 0.01$ for both cases. Although these choices may seem arbitrary, we find that the solution converges provided that the transition point is contained well inside the domain, and the result remains stable even if larger domains are used. The boundary conditions then read

$$\frac{d}{d\rho}\phi(\rho_{\min}) = 0, \quad \phi(\rho_{\max}) = \phi_F. \quad (19)$$

Our NN method can then be applied to find the bubble profile by solving the Euler-Lagrange equation (12). In this context, the NN method corresponds to an approach where the neural network attempts to apply the minimum action principle to the Euclidean action of Eq. (11). The test-field configuration, defined by the output layer of the neural network, is then adjusted using backpropagation until the classical trajectory is found. We discretize the domain into 500 training points and choose a network with a single hidden layer. For the thick-wall case, we use 10 hidden units with a sigmoid activation function, as was used in earlier examples; however, for the thin-wall case, we find that a single tanh unit is sufficient to achieve very good performance since the solution itself closely resembles a tanh function. To prevent the network from finding the trivial solution where the field remains in the false vacuum forever, we first train the network with the boundary condition at ρ_{\min} modified to $\phi(\rho_{\min}) = \phi_T$ so that the network finds a solution in the vicinity of the correct

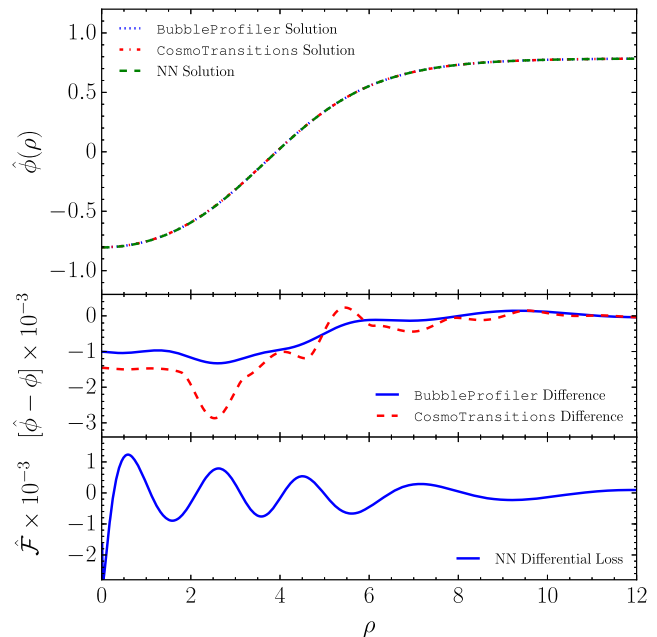


FIG. 5. The upper panel shows the bubble profile for the thick-wall potential ($\epsilon = 0.3$) in Eq. (18) for one scalar field, as obtained by our NN method, BubbleProfiler and CosmoTransitions. The middle panel displays the numerical difference between the NN predicted solution and the solutions from the other two codes. The lower panel shows the differential contribution $\hat{\mathcal{F}}$ to the loss.

solution, since the starting point is close to the true vacuum, before training again with the correct boundary conditions. We use this two-step training for all phase transition calculations.

Our results for the thick-wall and thin-wall cases are shown in Figs. 5 and 6, respectively, together with the CosmoTransitions and BubbleProfiler solutions. While all three methods agree very well for the thick-wall case, there is a disagreement in CosmoTransitions compared to BubbleProfiler and the NN approach in the thin-wall case. The dotted vertical line indicates where the bubble radius should be according to Eq. (16). Both BubbleProfiler and NN find a solution that matches the analytic calculation for the bubble radius. CosmoTransitions instead finds a solution with a smaller bubble radius, and therefore a smaller action and a larger tunneling rate.

For thin-wall cases, numerical stability is difficult to achieve. It is possible for an approximate solution to be found, which transitions at a much earlier ρ than it should, since a translated solution also approximately solves the differential equation [55]. For our method, $\hat{\mathcal{F}}$ can be monitored during the course of the training. During the early stages of the training where the solution does not yet have the correct transition point, $\hat{\mathcal{F}}$ will be sharply distributed in the region of the incorrect transition. As the training proceeds and the solution converges, the function will flatten out until an accurate solution is found across the entire domain.

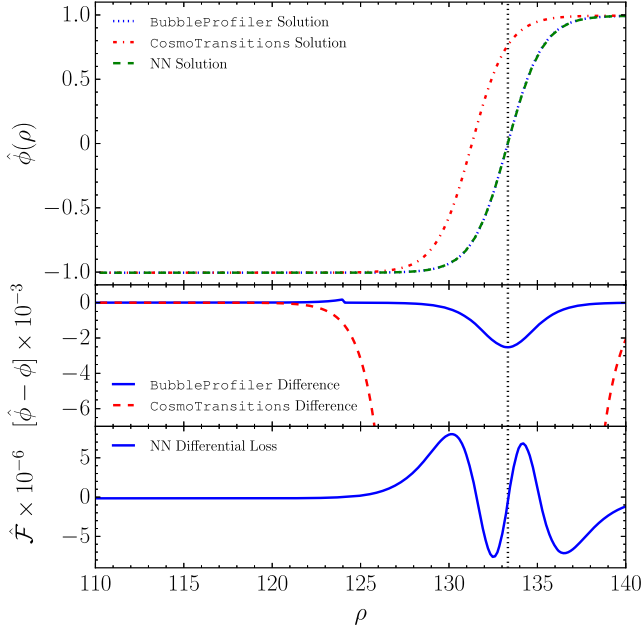


FIG. 6. The upper panel shows the bubble profile for the thin-wall potential ($\epsilon = 0.01$) in Eq. (18) for one scalar field, as obtained by our NN method, BubbleProfiler and CosmoTransitions. The middle panel displays the numerical difference between the NN predicted solution and the solutions from the other two codes. The lower panel shows the differential contribution $\hat{\mathcal{F}}$ to the loss. The dotted vertical line shows the analytic location of the bubble radius, which agrees with the radius found by both the NN and BubbleProfiler.

We have shown that the NN achieved very good stability for the thin-wall case using a single tanh function. We also explored the idea of using an adaptive distribution of training examples, such that more examples are distributed close to the region where the transition of the NN solution happens, and this distribution is then modified over the course of the training. A larger contribution to the loss in this region will be amplified by having more training examples, which can speed up learning. We found that the results can be improved by using this procedure, and this is an idea which could be investigated further in future work.

B. Phase transition with two scalar fields

To investigate how well the NN approach can solve the differential equation of Eq. (12) for multiple fields, we consider a potential for two scalar fields [51],

$$V(\phi_1, \phi_2) = (\phi_1^2 + \phi_2^2) \left[\frac{9}{5}(\phi_1 - 1)^2 + \frac{1}{5}(\phi_2 - 1)^2 - \delta \right], \quad (20)$$

which has a local minimum at $\phi_1 = \phi_2 = 0$ and a global minimum near $\phi_1 \approx \phi_2 \approx 1$. We focus again on the thick- and thin-wall cases, setting $\delta = 0.4$ for the former and

$\delta = 0.02$ for the latter. For the thick-wall potential, we solve the coupled equations in (12) with the boundary conditions,

$$\begin{aligned} \frac{d}{d\rho} \phi_1(\rho_{\min}) &= 0, & \phi_1(\rho_{\max}) &= 0, \\ \frac{d}{d\rho} \phi_2(\rho_{\min}) &= 0, & \phi_2(\rho_{\max}) &= 0, \end{aligned} \quad (21)$$

in the training domain $\rho \in [0.01, 6]$ with 500 training points. Again, the NN is built with 10 units in a single hidden layer with a sigmoid activation function. Since there are two fields, the NN has two units in the final layer. The two components ϕ_1 and ϕ_2 of the bubble solution, and the associated path through field space, are shown in Figs. 7 and 8, respectively. Once more, BubbleProfiler and the NN predictions agree very well, both for the one-dimensional profiles for ϕ_1 and ϕ_2 , and for the path in the (ϕ_1, ϕ_2) plane. CosmoTransitions shows a slightly different shape for the solutions of $\phi_1(\rho)$ and even more so for $\phi_2(\rho)$, resulting in a slightly modified escape path in Fig. 8. The behavior and small numerical value of the differential contribution $\hat{\mathcal{F}}$ to the loss suggests that the NN has converged to a correct solution for the profiles. Since it also agrees very closely with the result from BubbleProfiler, we conclude that in this

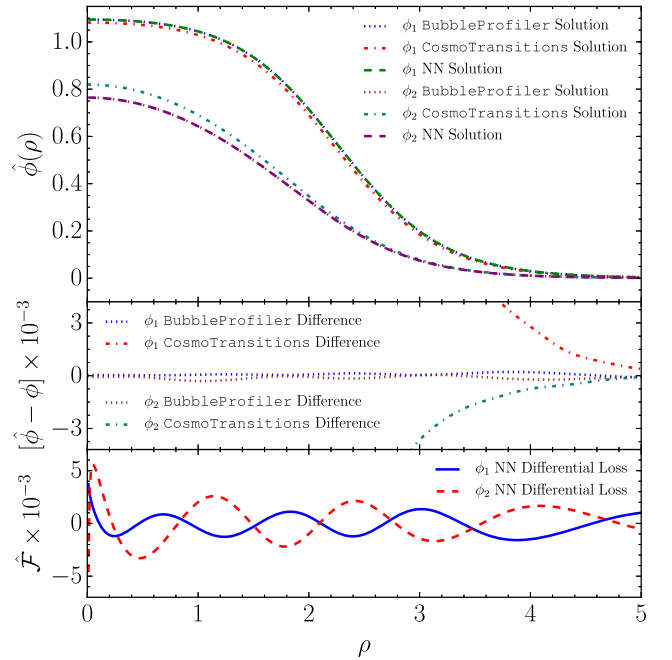


FIG. 7. The upper panel shows the bubble profiles for the thick-wall potential in Eq. (20) for two scalar fields, as obtained by our NN method, BubbleProfiler and CosmoTransitions. The middle panel displays the numerical difference between the NN predicted solutions and the solutions from the other two codes. The lower panel shows the differential contribution $\hat{\mathcal{F}}$ to the loss from ϕ_1 and ϕ_2 .

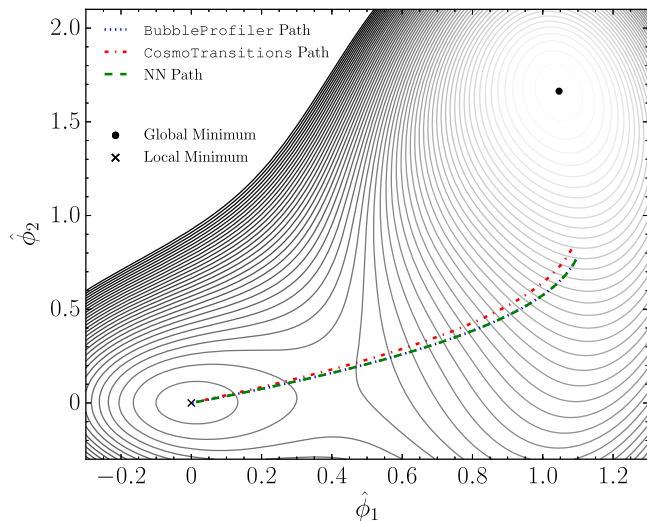


FIG. 8. Calculated solutions for the tunneling path for NN, BubbleProfiler and CosmoTransitions. The paths range from the local minimum to the exit point of the tunneling barrier. Also shown are the contours of the potential, where the global minimum is denoted by the black dot and the local minimum by the black cross.

case the BubbleProfiler result is correct. We note that our NN solution has found initial positions for the fields which agree with those from BubbleProfiler. In thick-wall cases, these can differ significantly from the true vacuum ϕ_T —these initial positions have been independently found by the network during optimization and have not been used as an input during training.

For the thin-wall potential we find that the performance can be significantly improved if a deeper network is used. BubbleProfiler instead does not find a solution at all, while the NN agrees very well with the path found by CosmoTransitions. Since there is not a solution from all three codes, we do not show the plot here.

Thus, in this section we have shown examples where CosmoTransitions or BubbleProfiler fail to provide a correct result, while the NN approach can cope well with both the thick-wall and the thin-wall solutions.

C. Singlet-scalar extended Standard Model with finite-temperature contributions

As a final example, we study a scenario of phenomenological interest, namely the extension of the SM Higgs sector by a single scalar field.³ Despite its simplicity, the singlet-scalar extended Standard Model (SSM) could potentially provide solutions to puzzles such as the existence of dark matter [66–68] and electroweak baryogenesis [69–72], where a crucial requirement is a strong electroweak phase transition, as discussed previously. The tree-level potential reads

³For projected and existing limits on this model, see Refs. [64,65] and references therein.

$$V^{(0)}(h, s) = -\frac{1}{2}\mu_h^2 h^2 + \frac{1}{4}\lambda_h h^4 + \frac{1}{2}\mu_s^2 s^2 + \frac{1}{4}\lambda_s s^4 + \frac{1}{4}\lambda_m s^2 h^2, \quad (22)$$

where h denotes the Higgs field and s the additional \mathbb{Z}_2 -symmetric scalar field.⁴ It is possible to consider a scenario in which the potential barrier separating the electroweak symmetric and the broken phase is generated already at tree level [73]. In this scenario, to study the evolution of parameters with T , it is enough to include only the high-temperature expansion of the one-loop thermal potential, which results in thermal corrections to the mass parameters [61],

$$V^{(1)}(h, s, T) = \left(\frac{1}{2}c_h h^2 + \frac{1}{2}c_s s^2 \right) T^2, \quad (23)$$

where

$$c_h = \frac{1}{48} [9g^2 + 3g'^2 + 2(6h_t^2 + 12\lambda_h + \lambda_m)], \quad (24)$$

$$c_s = \frac{1}{12} (2\lambda_m + 3\lambda_s), \quad (25)$$

with g and g' being the $SU(2)_L$ and $U(1)_Y$ gauge couplings, respectively, and h_t is the top Yukawa coupling. We then consider Eq. (12) with the potential

$$V(h, s, T) = V^{(0)}(h, s) + V^{(1)}(h, s, T). \quad (26)$$

At high temperatures the thermal contribution in Eq. (23) dominates, and the global minimum is the \mathbb{Z}_2 and electroweak symmetric configuration ($h = 0, s = 0$). The behavior as T decreases is determined by the choice of parameters. These are constrained to the parameter region in which the potential develops a strong tree-level barrier at the critical temperature T_C [73]. In particular, at $T > T_C$ after \mathbb{Z}_2 -symmetry breaking, s acquires a nonzero vacuum expectation value, $\langle s \rangle = w$, along the $\langle h \rangle = 0$ direction. This configuration constitutes a global minimum for the potential. At $T = T_C$ a second degenerate minimum appears at the electroweak symmetry breaking phase $\langle h \rangle = v$ and at the restored \mathbb{Z}_2 -symmetric vacuum, $\langle s \rangle = 0$. Finally, at $T < T_C$ the electroweak minimum $(v, 0)$ represents the only energetically favorable configuration. The nucleation temperature at which the phase transition from $\vec{\phi}_F = (0, w)$ to $\vec{\phi}_T = (v, 0)$ occurs is found from the requirement $\mathcal{S}_3(T_N)/T_N \simeq 140$ [41,74].

As an example parameter configuration, we consider $T_C = 110$ GeV, $\lambda_m = 1.5$ and $\lambda_s = 0.65$, as used in

⁴This condition could also be relaxed, since in models with no \mathbb{Z}_2 -symmetry the most general renormalizable potential would have three more parameters [64,73].

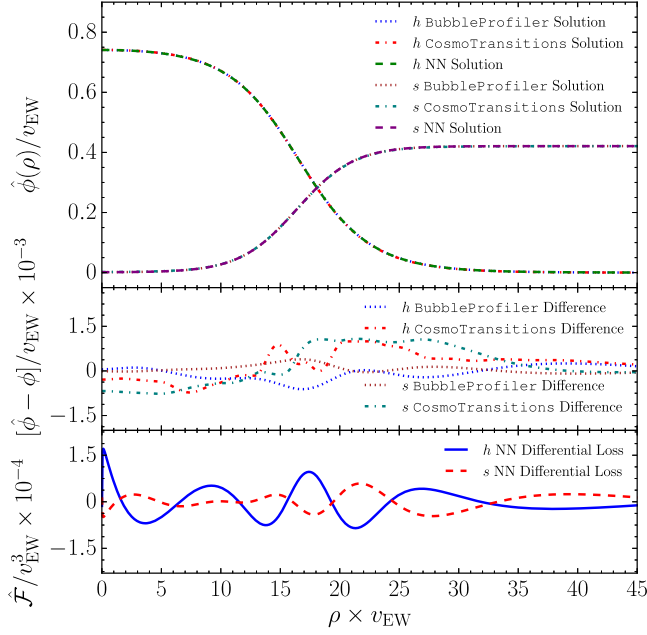


FIG. 9. The upper panel shows the bubble profiles for the singlet-scalar extended Standard Model potential in Eq. (26), as obtained by our NN method, BubbleProfiler and CosmoTransitions. The middle panel displays the numerical difference between the NN predicted solutions and the solutions from the other two codes. The lower panel shows the differential contribution $\hat{\mathcal{F}}$ to the loss from h and s .

Ref. [52], and a temperature of $T = 85$ GeV, which is the nucleation temperature that BubbleProfiler finds. We thus solve Eq. (12) with the boundary conditions

$$\begin{aligned} \frac{d}{d\rho} h(\rho_{\min}) &= 0, & h(\rho_{\max}) &= 0, \\ \frac{d}{d\rho} s(\rho_{\min}) &= 0, & s(\rho_{\max}) &= w. \end{aligned} \quad (27)$$

We use a neural network with 10 units in a single hidden layer with a sigmoid activation function, on a training domain of $\rho \in [0.01, 50]$ with 500 training points. To avoid large numerical values in the loss function, we scale all mass parameters in the potential by the electroweak symmetry breaking vacuum expectation value at zero temperature, v_{EW} . Our result, along with the comparison to CosmoTransitions and BubbleProfiler, is shown in Fig. 9. We find very good agreement between all three methods to calculate the bubble profiles $h(\rho)$ and $s(\rho)$, and the small values of $\hat{\mathcal{F}}$ across the domain show that good convergence has been achieved.

IV. CONCLUSIONS

By building on the capabilities of an artificial neural network in solving optimization problems, we have proposed a novel way to find solutions to differential equations.

Our method extends existing approaches on several accounts: (i) We avoid trial functions by including the boundary conditions directly into the loss function; (ii) the differential shape of $\hat{\mathcal{F}}$ is an excellent indicator of whether a good solution to \mathcal{F} has been found over the entire domain; (iii) in regions of numerical stability we propose increasing the domain iteratively to find stable solutions over arbitrarily large domains; (iv) for solutions that vary quickly over a small part of the domain, we find that it can be numerically beneficial to self-adaptively distribute more anchors in such regions.

We applied this approach to finding fully differentiable solutions to ordinary, coupled and partial differential equations, for which analytic solutions are known. Various network architectures have been studied, and even relatively small networks showed a very good performance.

To show how this method can be applied to a task of direct phenomenological interest, we used it to calculate the tunneling profiles of electroweak phase transitions and compared them to those obtained by CosmoTransitions and BubbleProfiler. We have presented explicit examples where the neural network method finds correct solutions, while either CosmoTransitions or BubbleProfiler fails. We find an optimized neural network to be very flexible and reliable, and is able to converge to solutions for all the examples tested with an accuracy that is competitive with the dedicated programs for calculating the bubble profiles. However, further work, e.g., in developing an approach to choosing the domain sizes for phase transitions in a more robust way, would be required to develop a fully automated tool using this approach.

As this method could be straightforwardly extended beyond the calculation of differential equations, we envision it to be applicable to a wide range of problems in perturbative and nonperturbative quantum field theory.

ACKNOWLEDGMENTS

We would like to thank Mikael Chala for help with the CosmoTransitions code. M. S. acknowledges the generous hospitality of Barbara Jaeger and her group at the University of Tuebingen and support of the Humboldt Society during the completion of this work.

- [1] Springer Verlag GmbH, European Mathematical Society, Encyclopedia of Mathematics, <https://www.encyclopediaofmath.org/>.
- [2] H. M. McGlone, Neural network analysis in Higgs search using $i\bar{i}H, H \rightarrow b\bar{b}$ and TAG database development for ATLAS, Report No. CERN-THESIS-2009-193, 2009, <https://cds.cern.ch/record/1356221>.
- [3] S. Chatrchyan *et al.* (CMS Collaboration), *Phys. Rev. D* **87**, 072001 (2013).
- [4] M. Aaboud *et al.* (ATLAS Collaboration), *Phys. Rev. D* **97**, 072016 (2018).
- [5] A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Rev. Lett.* **122**, 021801 (2019).
- [6] L. G. Almeida, M. Backovi, M. Cliche, S. J. Lee, and M. Perelstein, *J. High Energy Phys.* **07** (2015) 086.
- [7] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, *J. High Energy Phys.* **07** (2016) 069.
- [8] J. Barnard, E. N. Dawe, M. J. Dolan, and N. Rajcic, *Phys. Rev. D* **95**, 014018 (2017).
- [9] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, *J. High Energy Phys.* **01** (2017) 110.
- [10] A. Lenz, M. Spannowsky, and G. Tetlalmatzi-Xolocotzi, *Phys. Rev. D* **97**, 016001 (2018).
- [11] E. M. Metodiev, B. Nachman, and J. Thaler, *J. High Energy Phys.* **10** (2017) 174.
- [12] G. Kasieczka, T. Plehn, M. Russell, and T. Schell, *J. High Energy Phys.* **05** (2017) 006.
- [13] A. Butter, G. Kasieczka, T. Plehn, and M. Russell, *SciPost Phys.* **5**, 028 (2018).
- [14] S. Chang, T. Cohen, and B. Ostdiek, *Phys. Rev. D* **97**, 056009 (2018).
- [15] R. D. Ball, L. DelDebbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali (NNPDF Collaboration), *Nucl. Phys.* **B809**, 1 (2009); **B816**, 293(E) (2009).
- [16] G. Louppe, M. Kagan, and K. Cranmer, [arXiv:1611.01046](https://arxiv.org/abs/1611.01046).
- [17] R. T. D'Agnolo and A. Wulzer, *Phys. Rev. D* **99**, 015014 (2019).
- [18] C. Englert, P. Galler, P. Harris, and M. Spannowsky, *Eur. Phys. J. C* **79**, 4 (2019).
- [19] J. Brehmer, K. Cranmer, G. Louppe, and J. Pavez, *Phys. Rev. Lett.* **121**, 111801 (2018).
- [20] M. Farina, Y. Nakai, and D. Shih, [arXiv:1808.08992](https://arxiv.org/abs/1808.08992).
- [21] J. Hajer, Y.-Y. Li, T. Liu, and H. Wang, [arXiv:1807.10261](https://arxiv.org/abs/1807.10261).
- [22] J. H. Collins, K. Howe, and B. Nachman, *Phys. Rev. Lett.* **121**, 241803 (2018).
- [23] A. D. Sakharov, *Pis'ma Zh. Eksp. Teor. Fiz.* **5**, 32 (1967).
- [24] V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov, *Phys. Lett.* **155B**, 36 (1985).
- [25] A. Kosowsky, M. S. Turner, and R. Watkins, *Phys. Rev. Lett.* **69**, 2026 (1992).
- [26] C. Grojean and G. Servant, *Phys. Rev. D* **75**, 043507 (2007).
- [27] C. Caprini *et al.*, *J. Cosmol. Astropart. Phys.* **04** (2016) 001.
- [28] K. Hornik, M. Stinchcombe, and H. White, *Neural Netw.* **2**, 359 (1989).
- [29] K. Hornik, *Neural Netw.* **4**, 251 (1991).
- [30] H. Lee and I. S. Kang, *J. Comput. Phys.* **91**, 110 (1990).
- [31] A. Meade and A. Fernandez, *Math. Comput. Modell.* **19**, 1 (1994).
- [32] A. Meade and A. A. Fern, *Math. Comput. Modell.* **20**, 19 (1994).
- [33] I. E. Lagaris, A. Likas, and D. I. Fotiadis, [arXiv:physics/9705023](https://arxiv.org/abs/physics/9705023).
- [34] S. Mall and S. Chakraverty, *Adv. Artificial Neural Syst.* **2013**, 1 (2013).
- [35] A. Iftikhar and M. Bilal, *Am. J. Comput. Math.* **04**, 223 (2014).
- [36] F. Chollet *et al.*, Keras, <https://github.com/fchollet/keras> (2015).
- [37] M. Abadi *et al.*, [arXiv:1603.04467](https://arxiv.org/abs/1603.04467).
- [38] D. P. Kingma and J. Ba, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [39] M. E. Shaposhnikov, *JETP Lett.* **44**, 465 (1986).
- [40] A. D. Linde, *Contemp. Concepts Phys.* **5**, 1 (1990).
- [41] G. W. Anderson and L. J. Hall, *Phys. Rev. D* **45**, 2685 (1992).
- [42] M. Dine, R. G. Leigh, P. Huet, A. D. Linde, and D. A. Linde, *Phys. Lett. B* **283**, 319 (1992).
- [43] P. Huet and E. Sather, *Phys. Rev. D* **51**, 379 (1995).
- [44] R. K. Pathria, *Statistical Mechanics*, 2nd ed. (Elsevier, New York, 1996).
- [45] V. A. Rubakov and M. E. Shaposhnikov, *Usp. Fiz. Nauk* **166**, 493 (1996) [*Phys. Usp.* **39**, 461 (1996)].
- [46] D. E. Morrissey and M. J. Ramsey-Musolf, *New J. Phys.* **14**, 125003 (2012).
- [47] S. R. Coleman, *Phys. Rev. D* **15**, 2929 (1977); **16**, 1248(E) (1977).
- [48] A. D. Linde, *Nucl. Phys.* **B216**, 421 (1983); **B223**, 544(E) (1983).
- [49] S. R. Coleman, V. Glaser, and A. Martin, *Commun. Math. Phys.* **58**, 211 (1978).
- [50] R. Apreda, M. Maggiore, A. Nicolis, and A. Riotto, *Nucl. Phys.* **B631**, 342 (2002).
- [51] C. L. Wainwright, *Comput. Phys. Commun.* **183**, 2006 (2012).
- [52] P. Athron, C. Balzs, M. Bardsley, A. Fowlie, D. Harries, and G. White, [arXiv:1901.03714](https://arxiv.org/abs/1901.03714).
- [53] L. V. Kantorovich, *Usp. Math. Nauk* **3**, 89 (1948).
- [54] S. Akula, C. Balzs, and G. A. White, *Eur. Phys. J. C* **76**, 681 (2016).
- [55] A. Masoumi, K. D. Olum, and B. Shlaer, *J. Cosmol. Astropart. Phys.* **01** (2017) 051.
- [56] R. Jinno, [arXiv:1805.12153](https://arxiv.org/abs/1805.12153).
- [57] J. R. Espinosa, *J. Cosmol. Astropart. Phys.* **07** (2018) 036.
- [58] J. R. Espinosa and T. Konstandin, *J. Cosmol. Astropart. Phys.* **01** (2019) 051.
- [59] V. Guada, A. Maiezza, and M. Nemevek, *Phys. Rev. D* **99**, 056020 (2019).
- [60] P. John, *Phys. Lett. B* **452**, 221 (1999).
- [61] J. M. Cline, G. D. Moore, and G. Servant, *Phys. Rev. D* **60**, 105035 (1999).
- [62] T. Konstandin and S. J. Huber, *J. Cosmol. Astropart. Phys.* **06** (2006) 021.
- [63] J.-h. Park, *J. Cosmol. Astropart. Phys.* **02** (2011) 023.
- [64] S. Profumo, M. J. Ramsey-Musolf, and G. Shaughnessy, *J. High Energy Phys.* **08** (2007) 010.
- [65] J. M. No and M. Spannowsky, *Eur. Phys. J. C* **79**, 467 (2019).
- [66] A. Kusenko, *Phys. Rev. Lett.* **97**, 241301 (2006).
- [67] C. P. Burgess, M. Pospelov, and T. ter Veldhuis, *Nucl. Phys.* **B619**, 709 (2001).
- [68] J. McDonald, *Phys. Rev. Lett.* **88**, 091304 (2002).

- [69] J. M. Cline and K. Kainulainen, *J. Cosmol. Astropart. Phys.* **01** (2013) 012.
- [70] M. Chala, G. Nardini, and I. Sobolev, *Phys. Rev. D* **94**, 055006 (2016).
- [71] S. J. Huber, T. Konstandin, T. Prokopec, and M. G. Schmidt, *Nucl. Phys.* **B757**, 172 (2006).
- [72] S. J. Huber and M. G. Schmidt, *Nucl. Phys.* **B606**, 183 (2001).
- [73] J. R. Espinosa, T. Konstandin, and F. Riva, *Nucl. Phys.* **B854**, 592 (2012).
- [74] J. M. Moreno, M. Quiros, and M. Seco, *Nucl. Phys.* **B526**, 489 (1998).