
Deep ToC: A New Method for Estimating the Solutions of PDEs

Carl Leake¹

Abstract

This article presents a new methodology called deep ToC that estimates the solutions of partial differential equations (PDEs) by combining neural networks with the Theory of Connections (ToC). ToC is used to transform PDEs with boundary conditions into unconstrained optimization problems by embedding the boundary conditions into a “constrained expression” that contains a neural network. The loss function for the unconstrained optimization problem is taken to be the square of the residual of the PDE. Then, the neural network is trained in an unsupervised manner to solve the unconstrained optimization problem. This methodology has two major advantages over other popular methods used to estimate the solutions of PDEs. First, this methodology does not need to discretize the domain into a grid, which becomes prohibitive as the dimensionality of the PDE increases. Instead, this methodology randomly samples points from the domain during the training phase. Second, after training, this methodology represents a closed form, analytical, differentiable approximation of the solution throughout the entire training domain. In contrast, other popular methods require interpolation if the estimated solution is desired at points that do not lie on the discretized grid. The deep ToC method for estimating the solution of PDEs is demonstrated on four problems with a variety of boundary conditions.

1. Introduction

Partial differential equations are a powerful mathematical tool that is used to model physical phenomena, and their solutions are used in the design and verification processes of a variety of systems. For example, finite element analysis, which is a popular method used to approximate the solution of PDEs, may be used to verify that the loads experienced

by a beam throughout its lifetime will not cause the beam to fail. Many methods exist to approximate the solutions of PDEs. Probably the most famous of these methods is the finite element method (FEM) (Argyris & Kelsey, 1954; Turner et al., 1956; Clough, 1960). FEM has been incredibly successful in approximating the solution to PDEs in a variety of fields including structures, fluids, and acoustics. However, FEM does have some drawbacks.

FEM discretizes the domain into elements. This works well for low-dimensional cases, but the number of elements grows exponentially with the number of dimensions. Therefore, the discretization becomes prohibitive as the number of dimensions increases. Another issue is that FEM solves the PDE at the elements, but if the solution is needed between elements, an interpolation scheme must be used.

(Sirignano & Spiliopoulos, 2018) explored the use of neural networks to solve PDEs, and showed that the use of neural networks avoids these problems. Rather than discretizing the entire domain into a number of elements that grows exponentially with the dimension, neural networks can sample points randomly from the domain. Moreover, once the neural network is trained, it is a closed form, analytical, differentiable approximation of the PDE. Therefore, no interpolation schemes are needed when estimating the solution to points that did not appear during training. In addition, (Sirignano & Spiliopoulos, 2018) compared the neural network method with FEM on points that did not appear during training, and showed that the solution obtained by the neural network generalized well to points outside of the training data. In fact, the maximum error on the test set of data was never more than the maximum error on the training set of data. In contrast, the FEM had more error on the test set than on the training set. In one case, the test set had approximately 3 orders of magnitude more error than the training set.

(Sirignano & Spiliopoulos, 2018) is not the first article to explore the use of neural networks to solve PDEs. One of the early papers on the topic was (Lagaris et al., 1998). Although (Sirignano & Spiliopoulos, 2018) improved what was presented in (Lagaris et al., 1998) in almost every way, one item that was missing was exact satisfaction of the boundary constraints. The solution for PDEs presented by (Lagaris et al., 1998) used a method similar to Coon’s Patch

¹Ph.D. Student, Aerospace Engineering, Texas A&M University, College Station, TX. Correspondence to: Carl Leake <leakec@tamu.edu>.

(Coons, 1967) to satisfy the boundary constraints exactly.

Exact boundary constraint satisfaction is of interest for a variety of problems, especially when the boundary value information is known to a high degree of precision. Take for example, using the heat equation to describe the temperature within a rod. If the temperature on the edges of the rod is known to a high degree of precision, and the heat equation is being used to estimate the temperature inside the rod, then the only information that is known for sure a priori is the boundary conditions. Hence, it is desired to have the boundary conditions of the rod met exactly. Moreover, embedding the boundary conditions means that the neural network only needs to sample points from the interior of the domain, not the domain and the boundary. Thus, a method for embedding boundary conditions for PDEs of arbitrary dimension is desired.

The Theory of Connections (ToC) is a framework that is able to meet many types of boundary conditions while maintaining a function that can be freely chosen. This free function can be chosen, for example, to satisfy a differential equation. The theory of connections has already been able to solve differential equations with initial value constraints, boundary value constraints, relative constraints, integral constraints, and infinite constraints (Mortari, 2017b;a; Mortari et al., 2018; Johnston & Mortari, 2018). Recently, the framework has been extended to n dimensions and for constraints on the initial value, final value, derivatives of the initial and final value, and all combinations thereof for all independent variables. This means the framework can now create constrained expressions to solve multidimensional PDEs.

2. Theory of Connections

The Theory of Connections (ToC) is a framework that takes constrained problems and turns them into unconstrained optimization problems. This technique is especially useful when solving PDEs. ToC has two major steps: 1) embed the boundary conditions of the problem into the constrained expression 2) use the free function in the constrained expression to solve the new problem which is an unconstrained optimization problem. The paragraphs that follow will explain these steps in more detail.

The ToC framework is easiest to understand when explained via an example, like a simple harmonic oscillator. Equation (1) gives an example of a simple harmonic oscillator problem.

$$m\ddot{y} + ky = 0 \quad \text{subject to:} \quad \begin{cases} y(0) = y_0 \\ \dot{y}(0) = \dot{y}_0 \end{cases} \quad (1)$$

Step one of the ToC framework is to build the constrained expression. The constrained expression consists of two

parts; the first part is a function that satisfies the boundary constraints, and the second part projects the free-function, $g(x)$, onto the hyper-surface of functions that are equal to zero at the boundaries. For problems with boundary and derivative constraints, the constraint function can be written compactly using Eq. (2).

$$\begin{aligned} f(\mathbf{x}) &= \mathcal{M}_{i,j,\dots,n}(b(\mathbf{x}))v_i(x_1)v_j(x_2)\dots v_n(x_n) + \\ g(\mathbf{x}) &- \mathcal{M}_{i,j,\dots,n}(g(\mathbf{x}))v_i(x_1)v_j(x_2)\dots v_n(x_n) \end{aligned} \quad (2)$$

where \mathbf{x} is a vector of the independent variables, \mathcal{M} is an n th order tensor containing the boundary conditions $b(\mathbf{x})$, $v_i \dots v_n$ are first order tensors whose elements are functions of the independent variables $x_1 \dots x_n$, g is the free-function that can be chosen to be anything (as long as it and its derivatives exist or are tabulated at the coordinates being solved for), and $f(\mathbf{x})$ is the constrained expression. The first term in $f(\mathbf{x})$ satisfies the boundary conditions, and the final two terms project the free-function, $g(\mathbf{x})$, onto the hyper-surface of functions that are equal to zero on the boundaries.

For 1-dimension with independent variable t on the domain $[0, 1]$ and for boundary and first-derivative constraints only, \mathcal{M} has the following form,

$$\mathcal{M}(b(t)) = [b(0) \quad b_t(0) \quad b(1) \quad b_t(1)].$$

The values of \mathcal{M} that are unused are eliminated from the matrix, and the vector, v , is created afterwards with the appropriate size. For the 1-dimensional simple harmonic oscillator, \mathcal{M} and v are given by Eq. (3).

$$\begin{aligned} \mathcal{M}(b(t)) &= [b(0) \quad b_t(0)] \\ v &= \begin{bmatrix} p^1(t) \\ p^2(t) \end{bmatrix} \end{aligned} \quad (3)$$

where $p^1(t)$ and $p^2(t)$ are the functions that need to be solved for. Substituting everything into the constrained expression, $f(t)$, and simplifying yields Eq. (4).

$$f(t) = p^1(t)(b(0) - g(0)) + p^2(t)(b_t(0) - g_t(0)) + g(t) \quad (4)$$

The functions $p_1(t)$ and $p_2(t)$ are solved for by setting the constrained function, $f(t)$, equal to the boundary constraints at the boundaries. For the simple harmonic oscillator this yields the set of simultaneous equations given by Eqs. (5) and (6).

$$b(0) = p^1(0)(b(0) - g(0)) + p^2(0)(b_t(0) - g_t(0)) + g(0) \quad (5)$$

$$b_t(0) = p_t^1(0)(b(0) - g(0)) + p_t^2(0)(b_t(0) - g_t(0)) + g_t(0) \quad (6)$$

Equation (5) shows that $p^1(0) = 1$ and $p^2(0) = 0$, and Eq. (6) shows that $p_t^1(0) = 0$ and $p_t^2(0) = 1$. One solution to this set of simultaneous equations is $p^1(t) = 1$ and $p^2(t) = t$. Substituting these results back into the constrained expression and substituting $b(0) = y_0$ and $b_t(0) = \dot{y}_0$ yields Eq. (7).

$$f(t) = g(t) + y_0 - g(0) + t(\dot{y}_0 - g_t(0)) \quad (7)$$

Notice that for any function $g(t)$, the boundary conditions will always be satisfied exactly. Therefore, solving the differential equation has now become an unconstrained optimization problem. This unconstrained optimization problem could be cast in the following way. Let the function to be minimized, \mathcal{L} , be equal to the square of the residual of the differential equation,

$$\mathcal{L}(t) = \left(m\ddot{f}(t) + kf(t) \right)^2.$$

This function is to be minimized by varying the function $g(t)$. One way to do this is to make $g(t)$ the summation of a series of basis functions, and minimize the coefficients that multiply those basis functions via least-squares or some other optimization technique. For an example of this implementation via Chebyshev Orthogonal Polynomials see (Mortari, 2017a). Also, (Mortari et al., 2018) shows how this can be done using non-linear least squares for non-linear ODEs.

2.1. Two-Dimensional Case

This subsection will give an in depth example for a two-dimensional ToC case. The example is originally from (Lagaris et al., 1998) problem 5, and is one of the PDE problems analyzed in this article. The problem is shown in Eq. (8).

$$\begin{aligned} \nabla^2 z(x, y) &= e^{-x}(x - 2 + y^3 + 6y) \quad \text{subject to:} \\ \begin{cases} z(x, 0) = xe^{-x} \\ z(0, y) = y^3 \\ z(x, 1) = e^{-x}(x + 1) \\ z(1, y) = (1 + y^3)e^{-1} \end{cases} \end{aligned} \quad (8)$$

where $(x, y) \in [0, 1] \times [0, 1]$

In two dimensions, the \mathcal{M} tensor is a second order tensor, and its elements for the case of boundary values and their first derivatives are given in matrix form in Eq. (9) for independent variables x and y . Similar to the one dimensional case, only the rows and columns that correspond to the boundary conditions of interest need to be kept. For the example problem in Eq. (8), \mathcal{M} becomes

$$\mathcal{M} = \begin{bmatrix} 0 & b(x, 0) & b(x, 1) \\ b(0, y) & -b(0, 0) & -b(0, 1) \\ b(1, y) & -b(1, 0) & -b(1, 1) \end{bmatrix}.$$

For the two dimensional case, there are two, first order tensors, v_x and v_y . Their elements are shown in vector form in Eq. (10).

$$v_x = [1 \quad p^1(x) \quad p^2(x)], \quad v_y = \begin{bmatrix} 1 \\ q^1(y) \\ q^2(y) \end{bmatrix} \quad (10)$$

Now, the function $f(x, y)$ can be constructed using Eq. (2); the result is shown in Eq. (11). Four equations can be constructed using the boundary conditions given in Eq. (8). These equations can be used to solve for the unknown functions $p^1(x)$, $p^2(x)$, $q^1(y)$, and $q^2(y)$. One solution to these equations is shown in Eq. (12).

$$v_x = [1 \quad 1 - x^2 \quad x^2], \quad v_y = \begin{bmatrix} 1 \\ 1 - y^2 \\ y^2 \end{bmatrix}. \quad (12)$$

Plugging in the result from Eq. (12) and the boundary conditions from Eq. (8) into Eq. (11) yields the constrained expression for the two-dimensional example, which is shown in Eq. (13).

Notice, that Eq. (13) will always satisfy the boundary conditions of the problem regardless of the value of $g(x, y)$. Thus, the problem has been transformed into an unconstrained optimization problem where the cost function, \mathcal{L} , is the square of the residual of the PDE,

$$\mathcal{L}(x, y) = \left(\nabla^2 f(x, y) - e^{-x}(x - 2 + y^3 + 6y) \right)^2.$$

For the one-dimensional ODEs, the minimization of the cost function was accomplished by making g the summation of orthogonal polynomials, and performing least-squares or some other optimization technique to find the coefficients that multiply those orthogonal polynomials. For two dimensions, one could make $g(x, y)$ the product of two sums of these orthogonal polynomials, calculate all of the cross-terms, and then solve for the coefficients that multiply all terms and cross-terms using least-squares or non-linear least-squares. However, this will become prohibitive as the dimension increases. An alternative solution, and the one explored in this article, is to make the free function, $g(x, y)$, a neural network.

2.2. Remarks for n Dimensions

In n dimensions, there are two main challenges encountered when constructing the constrained expression: 1) constructing the n -th order tensor \mathcal{M} 2) finding the functions that make up the elements of the n first order tensors, v .

Although in the two-dimensional case it was convenient to construct an \mathcal{M} tensor that contained all of the boundary conditions and then keep only the rows and columns that pertained to the boundary conditions of a given problem,

$$\mathcal{M} = \begin{bmatrix} 0 & b(x, 0) & b_y(x, 0) & b(x, 1) & b_y(x, 1) \\ b(0, y) & -b(0, 0) & -b_y(0, 0) & -b(0, 1) & -b_y(0, 1) \\ b_x(0, y) & -b_x(0, 0) & -b_{xy}(0, 0) & -b_x(0, 1) & -b_{xy}(0, 1) \\ b(1, y) & -b(1, 0) & -b_y(1, 0) & -b(1, 1) & -b_y(1, 1) \\ b_x(1, y) & -b_x(1, 0) & -b_{xy}(1, 0) & -b_x(1, 1) & -b_{xy}(1, 1) \end{bmatrix} \quad (9)$$

$$\begin{aligned} f(x, y) = & g(x, y) + q^1(y) \left(b(x, 0) - g(x, 0) \right) + q^2(y) \left(b(x, 1) - g(x, 1) \right) + \\ & p^1(x) \left(b(0, y) - g(0, y) + q^1(y) \left(g(0, 0) - b(0, 0) \right) + q^2(y) \left(g(0, 1) - b(0, 1) \right) \right) + \\ & p^2(x) \left(b(1, y) - g(1, y) + q^1(y) \left(g(1, 0) - b(1, 0) \right) + q^2(y) \left(g(1, 1) - b(1, 1) \right) \right) \end{aligned} \quad (11)$$

in n dimensions the \mathcal{M} tensor can become prohibitively large. Therefore, it is recommended that in n dimensions the \mathcal{M} tensor be constructed for a given problem's boundary conditions. Luckily, there is a step-by-step method for constructing the \mathcal{M} tensor that will be illustrated via a 3-dimensional example that has Dirichlet boundary conditions in x and initial conditions in y and z on the domain $x, y, z \in [0, 1] \times [0, 1] \times [0, 1]$. The \mathcal{M} tensor is constructed using the following three rules.

1. The element of \mathcal{M} for all indices equal to 1 is 0 (i.e. $\mathcal{M}_{111} = 0$).
2. The first order tensor obtained by keeping one dimension's index and setting all other dimension's indices to 1, consists of the value 0 and the boundary conditions for that dimension. Using the example boundary conditions,

$$\begin{aligned} \mathcal{M}_{i11} &= [0, b(0, y, z), b(1, y, z)]^T \\ \mathcal{M}_{1j1} &= [0, b(x, 0, z), b_x(x, 0, z)]^T \\ \mathcal{M}_{11k} &= [0, b(x, y, 0), b_y(x, y, 0)]^T. \end{aligned} \quad (14)$$

3. The remaining elements of the \mathcal{M} tensor are the geometric intersection¹ of the boundary condition elements of the first order tensors given in Eq. (14). For example, the element \mathcal{M}_{223} would be the intersection

¹Technically, it is the geometric intersection of the boundary condition elements plus a sign (+ or -) that is determined by the number of elements being intersected.

of the elements \mathcal{M}_{211} , \mathcal{M}_{121} , and \mathcal{M}_{113} . The geometric intersection of the elements can be easily found by using the list of four rules that follows:

- (a) The intersection of the boundary conditions will have derivatives on all of the variables that the boundary conditions being intersected have.
- (b) The intersection of the boundary conditions will have a numeric value for a variable if any of the boundary conditions being intersected has a numeric value for that variable.
- (c) Any remaining variables that do not have numeric values will be left as variables.
- (d) The sign in front of the element is equal to $(-1)^{n-1}$ where n is the number of boundary conditions being intersected. For example, intersecting the elements 0, $b_x(x, 0, z)$, and $b_y(x, y, 0)$ will be negative, whereas intersecting the elements $b(1, y, z)$, $b_x(x, 0, z)$, and $b(x, y, 0)$ will be positive.

The following are some examples to help illustrate:

$$\begin{aligned} M_{133} &= -b_{xy}(x, 0, 0) \\ M_{221} &= -b(0, 0, z) \\ M_{332} &= b_x(1, 0, 0) \end{aligned}$$

Finding the elements that make up the first order tensors v is done by constructing as many equations as there are boundary conditions, and consequently, as many equations as there are functions in the first order tensors v . Then, constraints on the functions in the first order tensors v can

$$\begin{aligned} f(x, y) = & g(x, y) + \frac{x^2 y^2 (y - 1)}{e} + e^{-x} (x + y^2) + (1 - x^2) \left(g(0, 0) + y^2 \left(g(0, 1) + y - g(0, 0) - 1 \right) \right) + \\ & (x^2 - 1) g(0, y) + x^2 \left(y^2 g(1, 1) + (1 - y^2) g(1, 0) \right) - x^2 g(1, y) + (y^2 - 1) g(x, 0) - y^2 g(x, 1) \end{aligned} \quad (13)$$

$$\begin{aligned}
 f(x, y) = & \mathcal{N}(x, y; \theta) + \frac{x^2 y^2 (y - 1)}{e} + e^{-x} (x + y^2) + \\
 & (1 - x^2) \left(\mathcal{N}(0, 0; \theta) + y^2 \left(\mathcal{N}(0, 1; \theta) + y - \mathcal{N}(0, 0; \theta) - 1 \right) \right) + (x^2 - 1) \mathcal{N}(0, y; \theta) + \\
 & x^2 \left(y^2 \mathcal{N}(1, 1; \theta) + (1 - y^2) \mathcal{N}(1, 0; \theta) \right) - x^2 \mathcal{N}(1, y; \theta) + (y^2 - 1) \mathcal{N}(x, 0; \theta) - y^2 \mathcal{N}(x, 1; \theta)
 \end{aligned} \tag{15}$$

be calculated and the functions can be solved for, as was done in Eqs. (5) and (6). The only recommendation that this author can give at this time, based on all of the problems solved using ToC to date, is the simplest solution is to assume a polynomial form for the functions in v .

3. PDE Solution Methodology

As with the previous section, the easiest way to describe the methodology is with an example. The example used throughout this section will be the PDE given in Eq. (8).

As mentioned in the previous section, deep ToC approximates solutions to PDEs by finding the constrained expression for the PDE and choosing a neural network as the free function. For all of the problems analyzed in this article, a simple, fully connected neural network was used. These networks consist of non-linear activation functions composed with affine transformations of the form $\mathcal{A} = W \cdot x + b$, where W are neuron weights, b are neuron biases, and x is a vector of inputs from the previous layer (or the inputs to the neural network if it is the first layer). The weights and biases of the entire neural network make up the tunable parameters, θ . In this article, neural networks will be represented with the symbol \mathcal{N} . For example, a neural network with inputs x and y would be given as $\mathcal{N}(x, y; \theta)$. Thus, the constrained expression, given originally in Eq. (13), now has the form given in Eq. (15).

In order to estimate the solution to the PDE, the parameters of the neural network have to be optimized to minimize the loss function, which is taken to be the square of the residual of the PDE. For this example,

$$\begin{aligned}
 \mathcal{L}_i(x_i, y_i) = & \left(\nabla^2 f(x_i, y_i) - e^{-x_i} (x_i - 2 + y_i^3 + 6y_i) \right)^2, \\
 \mathcal{L} = & \sum_i^N \mathcal{L}_i.
 \end{aligned}$$

The attentive reader will notice that training the neural network will require, for this example, taking two second order partial derivatives of $f(x, y)$ to calculate \mathcal{L}_i , and then taking gradients of \mathcal{L} with respect to the neural network parameters, θ , in order to train the neural network.

To take these higher order derivatives, TensorFlow'sTM gradients function was used (TensorFlowTM). This function uses automatic differentiation (Baydin et al., 2015) to com-

pute these derivatives. However, one must be conscientious when using the gradients function to ensure they get the correct gradients.

When taking the gradient of a vector, y_j , with respect to another vector, x_i , TensorFlowTM computes,

$$z_i = \frac{\partial \left(\sum_{j=1}^N y_j \right)}{\partial x_i},$$

where z_i is a vector of the same size as x_i . The only place where this may be an issue in the example used in this section is when computing $\nabla^2 f_i$. The desired output of this calculation is the following vector,

$$z_i = \left\{ \frac{\partial^2 f_1}{\partial x_1^2} + \frac{\partial^2 f_1}{\partial y_1^2}, \dots, \frac{\partial^2 f_N}{\partial x_N^2} + \frac{\partial^2 f_N}{\partial y_N^2} \right\},$$

where z_i has the same size as f_i and (x_i, y_i) is the pair used to generate f_i . TensorFlow'sTM gradients function will compute the following vector,

$$z_i = \left\{ \frac{\partial^2 \left(\sum_{j=1}^N f_j \right)}{\partial x_1^2} + \frac{\partial^2 \left(\sum_{j=1}^N f_j \right)}{\partial y_1^2}, \dots, \frac{\partial^2 \left(\sum_{j=1}^N f_j \right)}{\partial x_N^2} + \frac{\partial^2 \left(\sum_{j=1}^N f_j \right)}{\partial y_N^2} \right\}.$$

However, because f_i only depends on (x_i, y_i) and the derivative operator commutes with the sum operator, TensorFlow'sTM gradients function will compute the desired vector. Moreover, the size of the output vector will be correct because the input vectors, x_i and y_i , have the same size as f_i .

3.1. Training the Neural Network

Three methods were tried when optimizing the parameters of the neural networks:

1. Adam optimizer (Kingma & Ba, 2014)
2. Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization algorithm (Fletcher, 1987)
3. Hybrid method

The first method, Adam, is a variant of stochastic gradient descent (SGD) that combines the advantages of two other

popular SGD variants: AdaGrad (Duchi et al., 2011) and RMSProp (Tieleman & Hinton).

The second method, BFGS, is a quasi-Newton method designed for solving unconstrained, non-linear optimization problems. This method was chosen based on its performance when optimizing neural network parameters to estimate PDE solutions in (Lagaris et al., 1998). The hybrid method uses both methods in series. For all four problems shown in this article, the BFGS optimizer gave the best results.

4. Results

This section compares the estimated solution found using deep ToC with with the analytical solution. Four PDE problems are analyzed. The first is the example PDE given in Eq. (8), and the second is the wave equation. The third and fourth PDEs are simple solutions to the incompressible Navier-Stokes equations.

4.1. Problem 1

The first problem analyzed was the PDE given by Eq. (8), copied below for the readers' convenience.

$$\nabla^2 z(x, y) = e^{-x}(x - 2 + y^3 + 6y) \quad \text{subject to:}$$

$$\begin{cases} z(x, 0) = xe^{-x} \\ z(0, y) = y^3 \\ z(x, 1) = e^{-x}(x + 1) \\ z(1, y) = (1 + y^3)e^{-1} \end{cases}$$

$$\text{where } (x, y) \in [0, 1] \times [0, 1]$$

The neural network used to estimate the solution to this PDE was a fully connected neural network with five hidden layers and 30 neurons per layer. The non-linear activation function used was the hyperbolic tangent. Other neural network sizes and non-linear activation functions were tried, but this size and activation function combination did the best. The biases of the neural network were all initialized as zeros, and the weights were initialized using TensorFlow's™ implementation of the Xavier initialization with uniform random initialization (Glorot & Bengio, 2010). Training pairs, (x, y) , were created for this problem by sampling x and y as independent and identically distributed (IID) random variables on the uniform distribution on $[0, 1]$. The network was trained using the BFGS method on a batch size of 10,000 training pairs.

Figure 1 shows the difference between the analytical solution and the estimated solution using deep ToC on a grid of 100 evenly distributed points (10 per independent variable). This grid represents the test set.

The maximum error on the test set was 4.806×10^{-6} meters

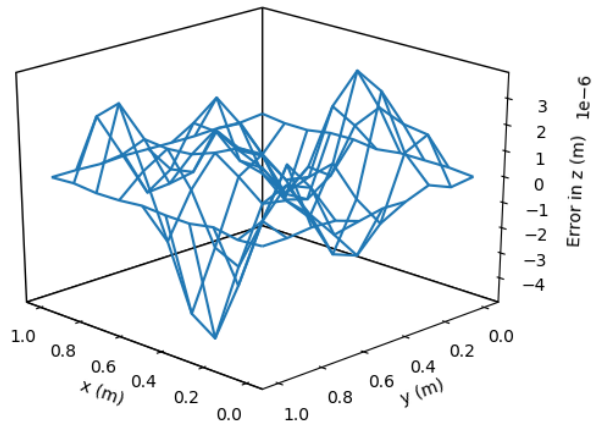


Figure 1. Problem 1 Solution Error

and the average error was 8.872×10^{-7} meters. The maximum error is relatively low, five orders of magnitude lower than the solution values, which are on the order of 10^{-1} meters. However, the maximum solution error using deep ToC is not as low as the maximum solution error obtained in (Lagaris et al., 1998), which was on the order of 10^{-7} meters. Although there are many possible explanations for this discrepancy, for example, different implementations of the optimizer or different weight initialization functions, the author was concerned that it may be the assumed solution form, $f(x, y)$, that was causing the increase in estimated solution error. The solution form created using deep ToC is more complex both in the number of terms and the number of times the neural network appears within the assumed solution form.

To investigate, a comparison was made between the solution form posed in (Lagaris et al., 1998) and the solution form posed in this article, while keeping all other variables constant. For this comparison, the neural network architecture consisted of one hidden layer with 10 neurons, and the non-linear activation function used was the hyperbolic tangent function². The neural network was trained using the BFGS optimizer. The training pairs were created by randomly sampling 10,000 point pairs (x, y) where x and y are IID random variables on the uniform distribution $[0, 1]$. The test set was a grid of 100 evenly distributed points (10 per independent variable).

Figure 2 was created using the solution form posed in (Lagaris et al., 1998). The maximum error on the test set was 5.481×10^{-6} meters and the average error on the test set was 1.131×10^{-6} meters.

²The sigmoid function was also tried, which is what is used in (Lagaris et al., 1998); however, the average error and maximum error for both solution forms, deep ToC and (Lagaris et al., 1998), was almost doubled compared to using the hyperbolic tangent function.

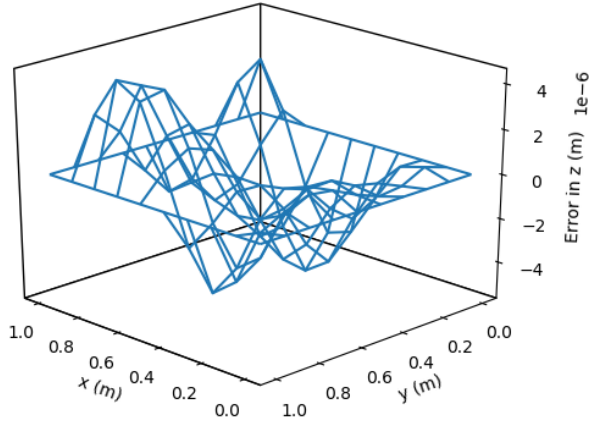


Figure 2. Problem 1 Solution Error Using (Lagaris et al., 1998) Solution Form

Figure 3 was created using the deep ToC solution form. The maximum error on the test set was 8.124×10^{-6} meters and the average error on the test set was 1.696×10^{-6} meters.

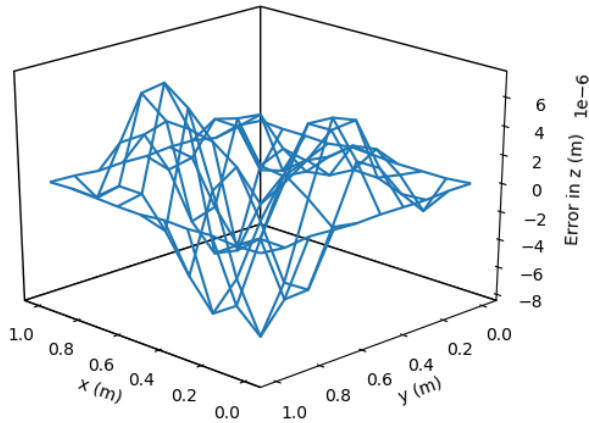


Figure 3. Problem 1 Solution Error Using Deep ToC Solution Form

Comparing Figs. 2 and 3 shows that the solution form from (Lagaris et al., 1998) does slightly better in terms of average error and maximum error for this problem, but the difference is not significant. However, neither of the tests conducted here is able to reduce the error to the level reported in (Lagaris et al., 1998). Thus, there must be some other factor, initialization, optimizer method, etc., that is causing this difference. One important take-away from this comparison is that despite the more complex solution form created using deep ToC, which can easily be applied to higher dimensions, the final solution accuracy is very similar to the simpler solution form that was designed for lower-dimensional problems.

4.2. Problem 2

The second problem analyzed was the wave equation, shown in Eq. (16).

$$\frac{\partial^2 u}{\partial t^2}(x, t) = c^2 \frac{\partial^2 u}{\partial x^2}(x, t) \quad \text{subject to:}$$

$$\begin{cases} u(0, t) = 0 \\ u(1, t) = 0 \\ u(x, 0) = x(1 - x) \\ u_t(x, 0) = 0 \end{cases} \quad (16)$$

where $(x, t) \in [0, 1] \times [0, 1]$

where the constant, c , was chosen to be 1. The constrained expression for this problem is shown in Eq. (17). The neural network used to estimate the solution to this PDE was a fully connected neural network with two hidden layers and 30 neurons per layer. The non-linear activation function used was the hyperbolic tangent. The biases and weights were initialized using the same method as problem 1. The training points, (x, t) , were created by sampling x and t IID from the uniform distribution on $[0, 1]$. The network was trained using the BFGS method on a batch size of 10,000 training pairs.

Figure 4 shows the difference between the analytical solution and the estimated solution using deep ToC on a grid of 100 evenly distributed points (10 per independent variable). This grid represents the test set.

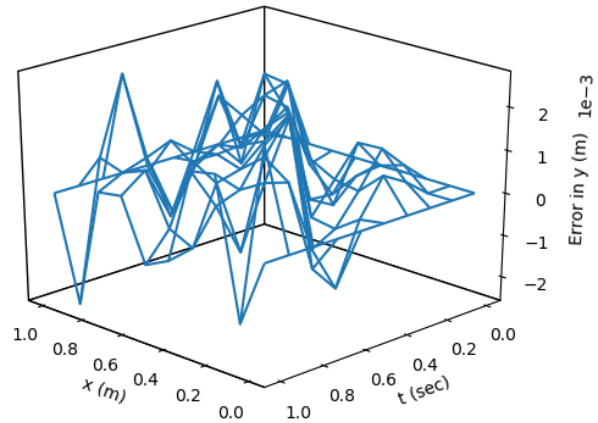


Figure 4. Problem 2 Solution Error

The maximum error on the test set was 2.701×10^{-3} meters and the average error on the test set was 6.978×10^{-4} meters. The error of this solution is larger than in the problem one, while the solution values are on the same order of magnitude, 10^{-1} meters, as in problem one. The larger relative error in problem two is most likely due to the more oscillatory nature of the solution.

$$f(x, t) = (1 - x) \left[\mathcal{N}(0, 0) - \mathcal{N}(0, t) \right] + x \left[\mathcal{N}(1, 0) - \mathcal{N}(1, t) \right] - \mathcal{N}(x, 0) + x(1 - x) + \mathcal{N}(x, t) + t \left[(1 - x) \mathcal{N}_t(0, 0) + x \mathcal{N}_t(1, 0) - \mathcal{N}_t(x, 0) \right] \quad (17)$$

$$\begin{aligned} f^u(x, y, t) &= \mathcal{N}(x, y, t; \theta) - \mathcal{N}(x, y, 0; \theta) + \mathcal{N}(0, y, 0; \theta) - \mathcal{N}(0, y, t; \theta) + \frac{1}{2H} \left(\frac{HP \left(y^2 - \frac{H^2}{4} \right)}{\mu} \right. \\ &\quad - (H - 2y) \left(\mathcal{N} \left(0, -\frac{H}{2}, 0; \theta \right) - \mathcal{N} \left(x, -\frac{H}{2}, 0; \theta \right) \right) + (H - 2y) \left(\mathcal{N} \left(0, -\frac{H}{2}, t; \theta \right) - \mathcal{N} \left(x, -\frac{H}{2}, t; \theta \right) \right) \\ &\quad \left. - (H - 2y) \left(\mathcal{N} \left(0, \frac{H}{2}, 0; \theta \right) - \mathcal{N} \left(x, \frac{H}{2}, 0; \theta \right) \right) + (H - 2y) \left(\mathcal{N} \left(0, \frac{H}{2}, t; \theta \right) - \mathcal{N} \left(x, \frac{H}{2}, t; \theta \right) \right) \right) \\ f^v(x, y, t) &= \mathcal{N}(x, y, t; \theta) - \mathcal{N}(x, y, 0; \theta) + \mathcal{N}(0, y, 0; \theta) - \mathcal{N}(0, y, t; \theta) + \frac{1}{2H} \left(\right. \\ &\quad - (H - 2y) \left(\mathcal{N} \left(0, -\frac{H}{2}, 0; \theta \right) - \mathcal{N} \left(x, -\frac{H}{2}, 0; \theta \right) \right) + (H - 2y) \left(\mathcal{N} \left(0, -\frac{H}{2}, t; \theta \right) - \mathcal{N} \left(x, -\frac{H}{2}, t; \theta \right) \right) \\ &\quad \left. - (H - 2y) \left(\mathcal{N} \left(0, \frac{H}{2}, 0; \theta \right) - \mathcal{N} \left(x, \frac{H}{2}, 0; \theta \right) \right) + (H - 2y) \left(\mathcal{N} \left(0, \frac{H}{2}, t; \theta \right) - \mathcal{N} \left(x, \frac{H}{2}, t; \theta \right) \right) \right) \end{aligned} \quad (19)$$

4.3. Problem 3

The third problem analyzed was a known solution to the incompressible Navier-Stokes equations, called Poiseuille flow. The problem solves the flow velocity in a two-dimensional pipe in steady-state with a constant pressure gradient applied in the longitudinal axis. Equation (18) shows the associated equations and boundary conditions.

$$\begin{aligned} \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= 0 \\ \rho \left(\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} \right) &= -\frac{\partial P}{\partial x} + \mu \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) \\ \rho \left(\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} \right) &= \mu \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right) \end{aligned}$$

subject to:

$$\left\{ \begin{array}{l} u(0, y, t) = u(x, y, 0) = \frac{1}{2\mu} \frac{\partial P}{\partial x} \left(y^2 - \left(\frac{H}{2} \right)^2 \right) \\ u(x, \frac{H}{2}, t) = u(x, -\frac{H}{2}, t) = 0 \\ \frac{\partial u}{\partial t}(0, y, t) = 0 \\ v(0, y, t) = \frac{\partial v}{\partial t}(0, y, t) = v(x, y, 0) = 0 \\ v(0, \frac{H}{2}, t) = v(0, -\frac{H}{2}, t) = 0 \end{array} \right. \quad (18)$$

where u and v are velocities in the x and y directions respectively, H is the height of the channel, P is the pressure, ρ is the density, and μ is the viscosity. For this problem, the values $H = 1$ m, $\rho = 1$ kg/m³, $\mu = 1$ Pa·s, and $\frac{\partial P}{\partial x} = -5$ N/m³ were chosen. The constrained expressions for the u -velocity, called $f^u(x, y, t)$, and v -velocity, called $f^v(x, y, t)$, are shown in Eq. (19). The neural network used

to estimate the solution to this PDE was a fully connected neural network with four hidden layers and 30 neurons per layer. The non-linear activation function used was the sigmoid. The biases and weights were initialized using the same method as problem 1. The training points, (x, y, t) , were created by sampling x , y , and t IID from the a uniform distribution that spanned the range of the associated independent variable. For x , the range was $[0, 1]$. For y , the range was $[-\frac{H}{2}, \frac{H}{2}]$, and for t , the range was $[0, 1]$. The network was trained using the BFGS method on a batch size of 1,000 training pairs. The loss function used was the sum of the squares of the residuals of the three PDEs in Eq. (18).

For one particular run³, the maximum error in the u -velocity was 3.308×10^{-7} , the average error in the u -velocity was 9.998×10^{-8} , the maximum error in the v -velocity was 5.575×10^{-7} , and the average error in the v -velocity was 1.542×10^{-7} . The maximum error and average error for this problem are much lower than in problems 1 and 2. However, the constrained expression for this problem essentially encodes the solution, because the initial flow condition at time zero is the same as the flow condition throughout the spatial domain at any time. Thus, if the neural network outputs a value of zero for all inputs, the problem will be solved exactly. Although the neural network does output a very small value for all inputs, it is interesting to note that none of the layers have weights or biases that are at or near zero.

³All errors are given in meters per second.

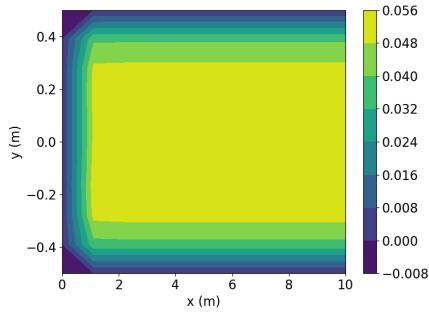


Figure 5. U-Velocity in Meters Per Second at 0.01 Seconds

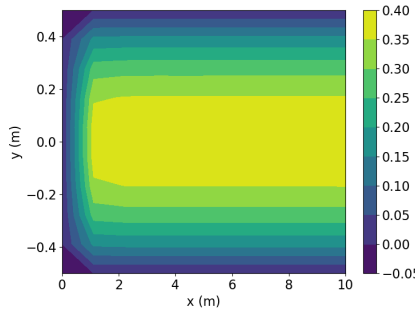


Figure 6. U-Velocity in Meters Per Second at 0.1 Seconds

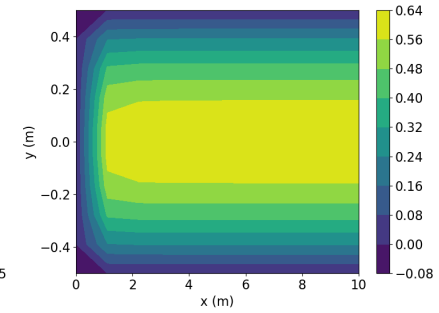


Figure 7. U-Velocity in Meters Per Second at 3.0 Seconds

4.4. Problem 4

The fourth problem is another solution to the Navier-Stokes equations, and is very similar to the third. The only difference is that in this case, the fluid is not in steady state, it starts from rest. Mathematically, this problem looks the same as the one given in Eq. (18), except that the boundary condition at time zero is given by $u(x, y, 0) = 0$ and the boundary condition at $x = 0$ is given by $u(0, y, t) = 0$. This problem was created to avoid encoding the solution to the problem into the constrained expression, as was the case in the previous problem. The constrained expression for the problem looks the same as the constrained expression given in Eq. (19), except $f^u(x, y, t)$ does not contain the term $\frac{HP(y^2 - \frac{H^2}{4})}{\mu}$.

The neural network used in this problem is exactly the same as the neural network used in problem three. Problem 4 used 2,000 training points that were selected the same way as in problem three, except the new ranges for the independent variables were $[0, 10]$ for x , $[0, 3]$ for t , and $[-\frac{H}{2}, \frac{H}{2}]$ for y .

Figures 5 through 7 show the u -velocity of the fluid throughout the domain at three different times. Qualitatively, the solution should look as follows. The solution should be symmetric about the line $y = 0$, and the solution should develop spatially and temporally such that after a period of time and sufficiently far from the inlet, $x = 0$, the u -velocity will be equal or very close to the steady state u -velocity of problem 3. Qualitatively, the u -velocity field looks correct at all points. Quantitatively, the u -velocity at $x = 10$ from Fig. 7 was compared with the known steady state u -velocity, and had a maximum error of 5.530×10^{-4} meters per second and an average error of 2.742×10^{-4} meters per second.

5. Conclusions

This article demonstrated how to combine neural networks with the theory of connections into a new methodology, called deep ToC, that was used to estimate the solutions

of PDEs. Results on four problems were presented that display how accurately relatively simple neural networks can approximate the solutions to some well known PDEs. The difficulty of the PDEs in these problems ranged from linear, two-dimensional PDEs to coupled, non-linear, three-dimensional PDEs.

Future work should investigate the performance of different neural network architectures on the estimated solution error. For example, (Sirignano & Spiliopoulos, 2018) suggests a neural network architecture where the hidden layers contain element-wise multiplications and sums of sub-layers. The sub-layers are more standard neural network layers like the fully connected layers used in the neural networks of this article.

Another topic for investigation is reducing the estimated solution error by sampling the training points based on the loss function values for the training points of the previous iteration. For example, one could create batches where half of the new batch consists of half of the points in the previous batch that had the largest loss function value and the other half are randomly sampled from the domain. This should consistently give training points that are in portions of the domain where the estimated solution is farthest from the real solution.

References

- Argyris, J. and Kelsey, S. Energy theorems and structural analysis: A generalized discourse with applications on energy principles of structural analysis including the effects of temperature and nonlinear stress-strain relations. *Aircraft Engineering and Aerospace Technology*, 26(10):347–356, 1954. doi: 10.1108/eb032482. URL <https://doi.org/10.1108/eb032482>.
- Baydin, A. G., Pearlmutter, B. A., and Radul, A. A. Automatic differentiation in machine learning: a survey. *CoRR*, abs/1502.05767, 2015. URL <http://arxiv.org/abs/1502.05767>.

- Clough, R. W. The finite element method in plane stress analysis. pp. 345–378, 1960. URL <http://www.e-periodica.ch/cntmng?pid=bse-me-001:1968:28::26>.
- Coons, S. A. Surfaces for computer-aided design of space forms. Technical report, Massachusetts Institute of Technology, Cambridge, MA, USA, 1967.
- Duchi, J., Hazan, E., and Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12:2121–2159, July 2011. ISSN 1532-4435. URL <http://dl.acm.org/citation.cfm?id=1953048.2021068>.
- Fletcher, R. *Practical Methods of Optimization*. Wiley, New York, 2nd edition, 1987.
- Glorot, X. and Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Teh, Y. W. and Titterton, M. (eds.), *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pp. 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR. URL <http://proceedings.mlr.press/v9/glorot10a.html>.
- Johnston, H. and Mortari, D. Linear differential equations subject to multivalued, relative and/or integral constraints with comparisons to chebfun. *SIAM Journal of Numerical Analysis*, 2018. Submitted.
- Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. URL <http://arxiv.org/abs/1412.6980>.
- Lagaris, I. E., Likas, A., and Fotiadis, D. I. Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks*, 9(5):987–1000, Sept 1998. ISSN 1045-9227. doi: 10.1109/72.712178.
- Mortari, D. Least-squares Solutions of Linear Differential Equations. *MDPI Mathematics*, 5(48), 2017a.
- Mortari, D. The Theory of Connections: Connecting Points. *MDPI Mathematics*, 5(57), 2017b.
- Mortari, D., Johnston, H., and Smith, L. Least-squares solutions of nonlinear differential equations. In *2018 AAS/AIAA Space Flight Mechanics Meeting Conference, Kissimmee, FL, January 8–12, 2018*. AAS/AIAA, 2018.
- Sirignano, J. and Spiliopoulos, K. DGM: A deep learning algorithm for solving partial differential equations, September 2018.
- TensorFlow™. Automatic differentiation and gradient tape. https://www.tensorflow.org/tutorials/eager/automatic_differentiation, September 2018. Accessed: 11-04-2018.
- Tieleman, T. and Hinton, G. Lecture 6.5 - RMSProp, COURSE: Neural Networks for Machine Learning. Technical report, 2012.
- Turner, M. J., Clough, R. W., Martin, H. C., and Topp, L. J. Stiffness and deflection analysis of complex structures. *Journal of the Aeronautical Sciences*, 23(9):805–823, sep 1956. doi: 10.2514/8.3664.