The Cost-Accuracy Trade-Off in Operator Learning with Neural Networks

Maarten V. de Hoop\textsuperscript{1}, Daniel Zhengyu Huang\textsuperscript{2}, Elizabeth Qian\textsuperscript{2}, and Andrew M. Stuart\textsuperscript{2}

\textsuperscript{1}Rice University, Houston, TX, USA.
\textsuperscript{2}California Institute of Technology, Pasadena, CA, USA.

Abstract. The term ‘surrogate modeling’ in computational science and engineering refers to the development of computationally efficient approximations for expensive simulations, such as those arising from numerical solution of partial differential equations (PDEs). Surrogate modeling is an enabling methodology for many-query computations in science and engineering, which include iterative methods in optimization and sampling methods in uncertainty quantification. Over the last few years, several approaches to surrogate modeling for PDEs using neural networks have emerged, motivated by successes in using neural networks to approximate nonlinear maps in other areas. In principle, the relative merits of these different approaches can be evaluated by understanding, for each one, the cost required to achieve a given level of accuracy. However, the absence of a complete theory of approximation error for these approaches makes it difficult to assess this cost-accuracy trade-off. The purpose of the paper is to provide a careful numerical study of this issue, comparing a variety of different neural network architectures for operator approximation across a range of problems arising from PDE models in continuum mechanics.

Keywords: Computational partial differential equations, Surrogate modeling, Operator approximation, Neural networks, Computational complexity.

1 Introduction

In many problems in computational partial differential equations (PDEs) the fundamental driver in deciding which approximation methodology to employ is the shape of the cost-accuracy curve: this determines what computational resources are required to achieve a desired level of accuracy, a measure of computational complexity. On this basis some methods may be shown to clearly outperform others, guiding computational practice. In the numerical analysis of PDEs there is a deep literature addressing this issue. This literature comprises two main components: (i) an analysis of the error as a function of the resolution of the finite dimensional approximation \cite{1-5}, and (ii) analysis of the cost of running the model, at a given level of finite-dimensional resolution, often dominated by matrix inversion and/or matrix-vector multiplies \cite{6,7} and/or by time-stepping and

\begin{itemize}
  \item [\textsuperscript{1}] mdehoop@rice.edu.
  \item [\textsuperscript{2}] dzhuang@caltech.edu.
  \item [\textsuperscript{3}] eqian@caltech.edu.
  \item [\textsuperscript{4}] astuart@caltech.edu.
\end{itemize}

https://www.global-sci.com/jml

Global Science Press
iteration count for nonlinear solvers. Theoretical results in (i) and (ii) may be combined
to determine the cost-accuracy curve for different methods and thereby inform the choice
of method for a given problem. For certain classes of equations, multi-resolution methods
have emerged which are near optimal in terms of minimizing cost for a given error [8,9].

Data-driven approximation of mappings/operators between function spaces provides
a way to learn cheap-to-evaluate surrogates which can bypass the need for employing
PDE solvers, after an initial training phase in which data are generated. These surro-
gates then enable efficient many-query analyses of PDE-based problems in computational
science and engineering. However, the theory for data-driven approximations is in its in-
fancy and cost-accuracy curves are not analytically understood. The goal of this work is
to provide a numerical study of the cost accuracy trade-off, for a range of operator neural
network architectures, including PCA-based neural networks (PCA-Net) [10,11], Deep-
ONet [12,13], pointwise evaluation (PARA-Net, defined in this paper), and the Fourier
neural operator (FNO) [14,15]. The numerical studies are conducted on four test prob-
lems: (1) the two-dimensional incompressible Navier-Stokes equation, (2) the Helmholtz
equation, (3) a structural mechanics test problem, and (4) the linear advection equation.

There are four sources of error in these operator learning problems: a) discretization of
the input and output spaces; b) parameterization of the operator approximators; c) finite
data volume; d) the optimizer used in training. In this paper we concentrate on b) and c)
and study the cost-accuracy trade-off in relation to data volume and number of parame-
ters in the neural network. The reason for not studying a) in this work is that, if properly
designed, operator approximators have the property of discretization invariance, mean-
ing that they are defined to act between function spaces and training of parameters for one
discretization can therefore be used for other discretizations [11,14,15]. In this setting a sin-
gle set of parameters will provide good approximations for all resolutions for which the
discretization error is small enough. As for the role of the optimizer d), while there exists
numerical evidence that stochastic gradient descent methods can be effective in driving
the loss function (close) to its global minimum [16–18], this work is far from being theo-
retically well-understood and, furthermore, is not in the context of operator learning and
partial differential equations. There are also other optimization approaches, for example
using second-order information [19,20], or ensemble methods [21,22], that may produce
different results. However, in order to limit the scope of our numerical studies, we em-
ploy stochastic gradient descent using fixed standard choices of the optimization hyper-
parameters for all test cases. When there is evidence that the optimization itself limits the
accuracy achieved, we will highlight this in our discussion. With this caveat, we primarily
focus our study on b) and c) and extract clear signals from numerical experiments, laying
foundation for future studies which delve into the interactions with a) and d). Our numer-
ical experiments will disentangle the roles of errors caused by b) and by c). The seminal
work of Giles [23,24] on multilevel Monte Carlo methods demonstrates that a theoretical
understanding of errors incurred through the interaction of finite sampling and finite
dimensional approximation leads to highly efficient methods which use different sam-
ple sizes for different finite dimensional approximations of expectations. Future analysis
studying the interaction between the sources of error arising from b) and c) would be very
valuable in the field of operator approximation and could lead to improved complexity