Accuracy of a Neural Network Differential Equation Solver

Christopher Monterola & Caesar Saloma
National Institute of Physics
University of the Philippines Diliman, 1101 Quezon City, Philippines
E-mail: chris@nip.upd.edu.ph, csaloma@nip.upd.edu.ph

EXTENDED ABSTRACT

Modeling complex systems is basically a two-fold process. The first step involves formulation of the dynamical relations which usually takes the form of differential equations (DE) while the second step engages in determination of the relationship between dynamical variables. Each of the task is significant and oftentimes, equally difficult. For example, when the relevant DEs of interacting systems are prepared, say the Lotka - Volterra equations for predator-prey models or the hydrodynamical equations for fluid motion, the problem of solving this equation is equally formidable since in many cases DE has no known analytic solution. One way of solving DE is by transforming its coordinates such that it resembles a form with known analytic solution. Such process however is generally tedious, or at worst will never work because such transformation does not exist in the first place. A more common approach is by utilizing numerical techniques (e.g., finite-difference, Runge Kutta, etc.) which generally relies on the accuracy of the sampling interval. However, numerical methods provide an iterative representation of the solution and hence, the error propagates at later iterations. Another way is by using neural networks (NN). Recently we have shown that an unsupervised NN trained using the modified backpropagation method can solve differential equation that: (1) models the propagation of pulse in a nonlinear media, (2) replicates the arrangement of competing biological entities in a given space, and (3) determines the optimum design for a nuclear reactor (Monterola & Saloma, 1998; Monterola & Saloma, 2000). In this study we analyze formally how accurately NN can solve linear DEs, and we propose an approach in increasing the accuracy of the solution obtained by NN for general types of DEs. We utilize a three layer unsupervised NN with inputs $\{ x_1, x_2, \ldots, x_n \}$ in solving the DE which is given by $F(x_1,x_2, \ldots, x_n) = D\Psi(x_1,x_2, \ldots, x_n) = 0$, where the D is a differential operator and x_i's are the dynamical parameters. The s^{th} - output $\psi_s^{(k)}$ corresponding to the k^{th}-set of inputs of the NN, is given by $\psi_s^{(k)} = f_0(\Sigma_{m=1,H} d_{sm}^{(k)} y_m^{(k)})$; $y_m^{(k)} = f_0(\Sigma_{j=1,L} w_{mj}^{(k)} r_j)$ where r_j is the j^{th} input, $y_m^{(k)}$ gives the m^{th} output of the hidden nodes $\{m=1,2 \ldots H\}$, $f_H(z) = \tanh(z)$ and $f_0(z) = z$ represents the hidden and output activation functions, respectively. Parameter $w_{mj}(k)$ describes the interconnection weights between neurons in the m^{th} and the j^{th} layer while $d_{sm}(k)$ gives the synaptic strength of the s^{th} and m^{th} layer.

Learning of NN is accomplished by updating the weights $d_{sm}^{(k)}$ and $w_{mj}^{(k)}$ using the gradient descent method such that it minimizes F and simultaneously satisfies the boundary conditions $B(x_1,x_2, \ldots, x_n) = 0$ (Monterola & Saloma, 1998; Monterola & Saloma, 2000 &Haykin, 1999). Meaning, $w_{mj}^{(k+1)} = w_{mj}^{(k)} - \eta \partial E^{(k)}/\partial w_{mj}^{(k)} - \alpha \partial E^{(k-1)}/\partial w_{mj}^{(k-1)}$ where $E = (F)^2 + (B)^2$. If we can reduce E to exactly zero, then it is guaranteed that the NN solution ψ_{nn} is exactly equivalent to the true analytic solution ψ_{theo} or the normalized mean square error $\xi = \Sigma|\psi_{theo} - \psi_{nn}|^2 / \Sigma|\psi_{theo}|^2$ approaches zero. However in actuality, E can only be lowered up to some finite value d after k iterations owing to both computer and NN architectural limitations. Can we therefore correlate δ with ξ? This is an important and nontrivial problem for a NN DE solver since it will provide a compact stopping rule in a learning NN. Also, for problems with no known analytic solution, it will allow an estimate of the accuracy of ψ_{nn}.

We formally derive a correlation of δ with ξ for any linear DE (i.e., $D = a_n d^n/dx^n + a_{n-1} d^{n-1}/dx^{n-1} + \ldots + a_0$) and show that $\xi(\delta) = \Sigma| (A-1)\psi_{nn} - \delta|^2 / \Sigma |A\psi_{nn} - \delta|^2$ where A is a constant that is fixed using the BCs of the problem. This result is obtained by manipulating the

particular solution of the family of curves defined by F. (Rainville & Bedient, 1981). We test our prediction for a harmonic oscillator from which $D\psi=(d^2/dx^2+1)\psi=0$, with the imposition that $\psi=1$ at $x=0$. Fig. 1 shows that δ (solid square) decreases exponentially as the iteration number k increases, a signature characteristic of a generalizing NN. Comparison of our prediction (filled circle) with the actual ξ (open squares) shows good agreement when δ is reduced to a sufficiently small value ($\delta < 10^{-1}$), the period when the NN learns the solution. When δ is large, weights of the NN undergo large fluctuation and the prediction is expected to fail because of the inherent instability of the NN (Haykin, 1999).

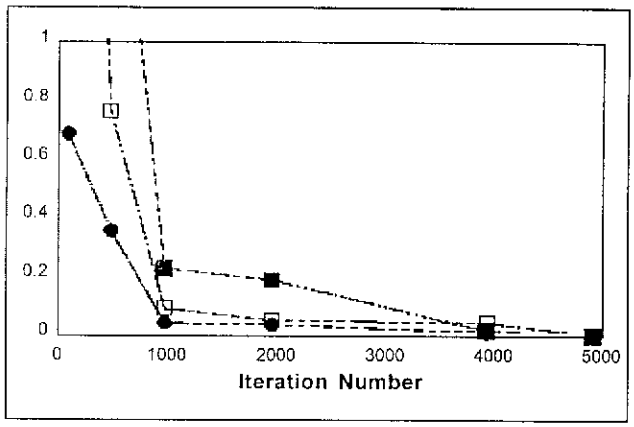

Fig. 1. Linear differential equation. Comparison of the actual normalized mean square error ξ (□) and the predicted error (●) for its corresponding δ (■)

The very diversified forms of nonlinear DEs made the problem of determining an exact relation between δ and ξ seemingly impossible (Rainville & Bedient, 1981). Here, we propose a starting numerical method that can possibly gauge the accuracy of the NN's solution from δ utilizing both extrapolation and curve fitting. In particular we consider the nonlinear Schroedinger equation (NSE) that models the propagation of a pulse in a nonlinear dispersive waveguide. NSE has no known analytic solution for an arbitrary initial pulse and is usually treated by employing numerical methods (Monterola & Saloma, 2000). Shown in Fig. 2 are the plots of δ and ξ against increasing iterations for different values of medium's nonlinearity N when the NN is used to solve NSE. Notice that there are no apparent trends regarding the relation of ξ and δ for N=1,2,3, and 4.

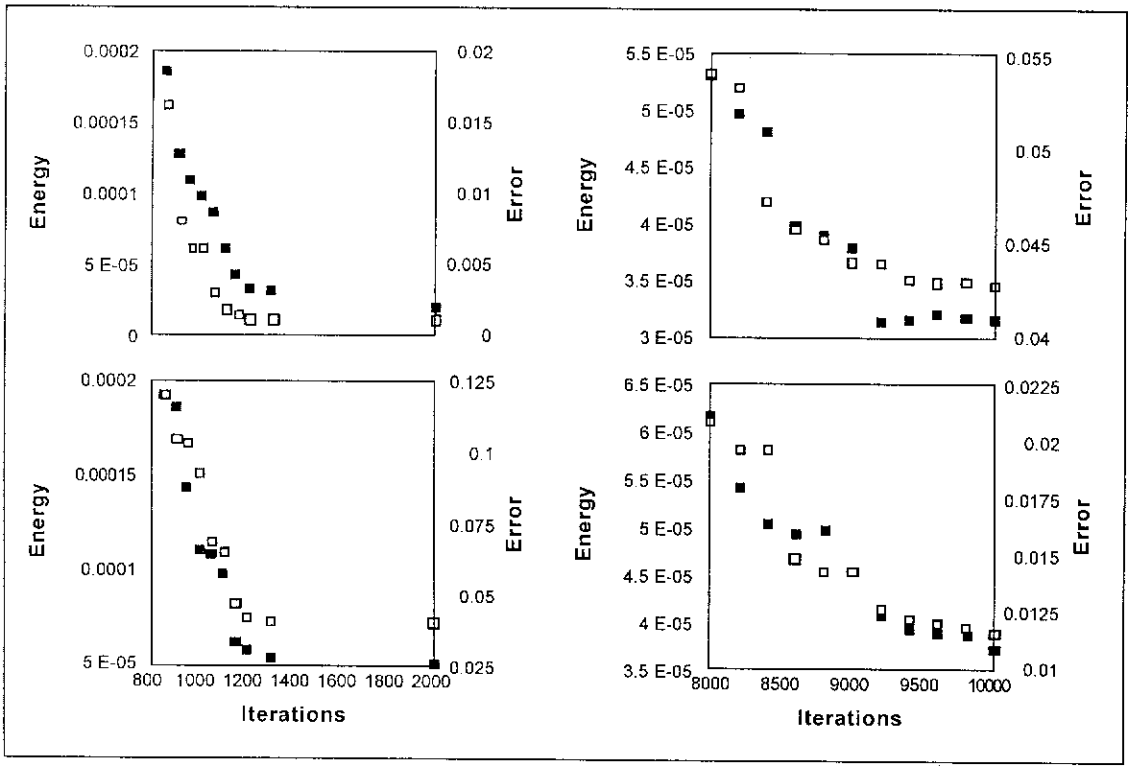

Fig. 2. Nonlinear Schroedinger Equation. The energy function δ(□) and its corresponding normalized mean square error ξ (□) as a function of training iterations for different nonlinearities (N=1, 2, 3 ,4) of the medium in the NSE

Using the information available on the energy δ and the corresponding input-output pairs of the NN, we extrapolate the solution ψ for the case of $\delta = 0$. Since the error variance of each fitting procedure is known, we can determine the curve that best fits the ψ-δ pairs. Thus, the remaining problem is to numerically predict ξ from the best curve-a starting point that reduces the abstraction of the problem and we hope allows us to treat systematically each variety of non linear DE. Even if this might fail however, the procedure already merits from the fact that it increases the accuracy of ψ as shown in Fig. 3, where we fit six functions namely: linear (Ax +B); quadratic (Ax^2+ Bx +C); exponential 1 ($A \exp^{Bx}$); exponential 2($A \exp^{Bx+C}$) and power law (AB^{Cx}), A, B and C are fitting parameters. Note that even without a priori information of ψ_{theo}, we can improve the accuracy by extrapolating the solution when δ is zero. Such result is evidently significant for N=3 and 4. This procedure also reduces the NN training time if our concern is just accuracy of ψ_{theo}.

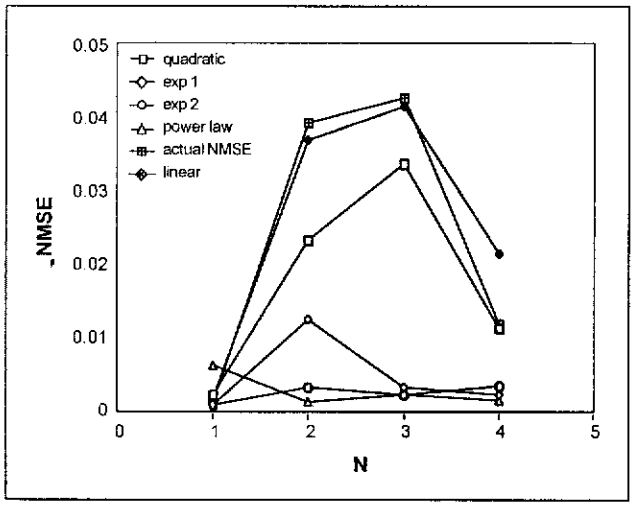

Fig. 3. Nonlinear Schroedinger Equation. Error ξ the NN solution ψ is extrapolated at $\delta = 0$ using various fitting functions.

Solving DE using NN has the following advantages: (1) solution obtained is non- iterative; (2) the mapping is fast because of the inherent parallelism of the architecture; and (3) the method is straightforward since no tedious coordinate transformation is necessary.

ACKNOWLEDGMENT

This project is partially supported by the Commission on Higher Education.

REFERENCES

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. 2nd Ed. New York, Prentice-Hall: 842 p.

Monterola, C. & C. Saloma, 1998. Characterizing the dynamics of complex interacting systems. *Phys. Rev. E*, 57(2): R1247-R1250.

Monterola, C. & C. Saloma, 2000. Solution to nonlinear schroedinger equation using neural networks. (Submitted 2000)

Rainville, E., P. Bedient, & R. Bedient, 1997. Elementary Differential Equations. Sixth Ed. New York, Macmillan Pub. Co.: 530 p.