Inverse Problems, Design and Optimization Symposium
Rio de Janeiro, Brazil, 2004

USING NEURAL NETWORKS TO OBTAIN INITIAL ESTIMATES FOR THE SOLUTION OF INVERSE HEAT TRANSFER PROBLEMS

Francisco J. C. P. Soeiro
Department of Mechanical Engineering
Universidade do Estado do Rio de Janeiro, UERJ
Rio de Janeiro, RJ, Brazil
soeiro@uerj.br

Patrícia Oliva Soares
Instituto Politécnico, IPRJ,
Universidade do Estado do Rio de Janeiro, UERJ,
C.P. 97282, 28601-970, Nova Friburgo, RJ, Brazil
pattyoliva@uol.com.br

Haroldo F. Campos Velho
Laboratório Associado de Matemática Aplicada e Computacional, LAC,
Instituto Nacional de Pesquisas Espaciais, INPE,
CEP 12245-970, S. José dos Campos, SP, Brazil
haroldo@lac.inpe.br

Antônio J. Silva Neto
Department of Mechanical Engineering and Energy, Instituto Politécnico, IPRJ,
Universidade do Estado do Rio de Janeiro, UERJ,
C.P. 97282, 28601-970, Nova Friburgo, RJ, Brazil
ajsneto@iprj.uerj.br

ABSTRACT

In the present work a multi-layer perceptron neural network is used to obtain initial estimates to solve inverse heat transfer problems using a gradient based method (Levenberg-Marquardt).

INTRODUCTION

In recent years it has been observed an increasing interest in the analysis and solution of inverse heat transfer problems. Implicit formulations of inverse problems of parameter estimation, in which a cost function is minimized, have largely been employed for the solution of such problems [1-5]. In most cases gradient based methods have been used for the minimization of the cost function, but when convergence occurs it may in fact lead to a local minimum. On the other hand, the main advantage of such methods is a good rate of convergence.

Another approach for the minimization of the cost function involves global optimization methods, with an increasing interest towards stochastic methods. With the proper implementation of these methods one gets very close to the global minimum, but the computational cost is usually very high.

In previous works [6-8] combinations of gradient based methods and stochastic optimization methods were used. In that approach, stochastic methods were employed to obtain good estimates to gradient methods, trying to get the best of each approach, i.e., the computational efficiency of gradient methods and the assurance of getting a good approximation to the global minimum by the careful implementation of the stochastic optimization methods.

In the present work a faster alternative was tried to obtain a better first guess to the Levenberg-Marquardt (LM) method. A simple multi-layer perceptron (MLP) artificial neural network (ANN) with back-propagation algorithm was used for this purpose. A reduced number of neurons in the network and a reduced number of patterns for training were used to get with low computational cost the initial estimates for the LM method. Differently from previous works dealing with ANNs for the inverse solution in heat conduction problems for function estimation [9-11], the goal here is to focus on the parameter estimation problem. A literature review has shown that there is a large number of publications in damage/defects identification using ANNs. Nonetheless, there is still a lot to be done on inverse heat transfer problems.

Bokar [12] used an ANN fot the estimation of parameters in a inverse radiative transfer problem, and Boillereaux et al. [13] estimated thermal properties of food samples using an ANN.

HEAT CONDUCTION PROBLEM

Consider a cylindrical sample of radius *R*, with a line source at $r = 0$, and long enough such that heat conduction can be considered only in the radial direction. The sample is initially at the same temperature of the ambient, $T = T_{amb}$, and the heat source, whose intensity may vary with time, $g(t)$, starts to release its energy at $t = 0$.

The mathematical formulation of the physical situation described is given by

$$\frac{1}{r}\frac{\partial}{\partial r}\left(kr\frac{\partial T}{\partial r}\right)+g(t)=\rho c_p\frac{\partial T(r,t)}{\partial t}$$

$$\text{in } 0<r<R\text{, for } t>0 \qquad \text{(1a)}$$

$$-k\frac{\partial T}{\partial r}=h(T-T_{\text{amb}}) \quad \text{in } r=R\text{, for } t>0 \qquad \text{(1b)}$$

$$T(r,0)=T_{\text{amb}} \quad \text{in } 0<r<R\text{, for } t=0 \qquad \text{(1c)}$$

where h is the heat transfer coefficient, ρ the density, k the thermal conductivity, and c_p the heat capacity.

Measured data on the temperature at a given location of the sample at different times $\{W_i\}$ is considered available.

RADIATIVE TRANSFER PROBLEM

Consider the problem of radiative transfer in an absorbing, isotropically scattering, plane-parallel, and gray medium of optical thickness τ_0, between two diffusely reflecting boundary surfaces. The mathematical formulation of the direct problem with azymuthal symmetry is given by

$$\mu\frac{\partial I(\tau,\mu)}{\partial\tau}+I(\tau,\mu)=\frac{\omega}{2}\int_{-1}^{1}I(\tau,\mu')d\mu',$$

$$0\le\tau\le\tau_0,\ -1\le\mu\le 1 \qquad \text{(2a)}$$

$$I(0,\mu)=A_1+2\rho_1\int_0^1 I(0,-\mu')\mu'd\mu'\ ,\mu>0 \qquad \text{(2b)}$$

$$I(\tau_0,-\mu)=A_2+2\rho_2\int_0^1 I(\tau_0,\mu')\mu'd\mu'\ ,\mu>0 \qquad \text{(2c)}$$

where $I(\tau,\mu)$ is the dimensionless radiation intensity, τ the optical variable, μ the direction cosine, ω the single scattering albedo, and ρ_1 and ρ_2 the diffuse reflectivities. The illumination from the outside is supplied by isotropically incident radiation given by the terms A_1 and A_2. Measured exit intensities $\{Y_i\}$, at different polar angles, are considered available.

SOLUTION OF THE INVERSE PROBLEM WITH THE LEVENBERG-MARQUARDT METHOD

In the heat conduction problem we are interested in the estimation of k and c_p, represented by the vector of unknowns

$$\vec{Z}=\{k,c_p\}^T \qquad (3)$$

using experimental data on the temperature acquired inside the medium, W_i, $i = 1, 2,\ldots,N$.

As the number of measured data, N, is usually much larger than the number of parameters to be estimated, $M = 2$, the problem is solved as a finite dimensional optimization problem in which we seek to minimize

$$Q=\sum_{i=1}^{N}\left[T_i(k,c_p)-W_i\right]^2=\vec{F}^T\vec{F} \qquad (4)$$

with the elements of vector $\vec{F}$ given by

$$F_i=T_i(k,c_p)-W_i, \qquad i=1,2,\ldots,N \qquad (5)$$

where T_i are the calculated temperatures. The index i represents the discretization of the time interval in which the temperature measurements are taken.

The inverse radiative transfer problem we are interested in can be stated as: by utilizing the measured data $\{Y_i\}$, $i = 1, 2,\ldots, N$, determine the elements of the unknown vector $\vec{Z}$, with $M=2$, defined as

$$\vec{Z}=\{\omega,\tau_0\}^T \qquad (6)$$

A squared residues norm, similar to Eq. (4) is also minimized, with the elements of the residue vector given by

$$F_i = I_i(\tau_0, \omega, \rho_1, \rho_2) - Y_i, \quad i = 1, 2, \ldots, N \qquad (7)$$

where I_i are the computed exit intensities.

The minimization of the cost function Q with the Levenberg-Marquardt method consists on constructing an iterative procedure that starts with an initial guess $\vec{Z}^0$, and new estimates are obtained with

$$\vec{Z}^{n+1} = \vec{Z}^n + \Delta\vec{Z}^n \quad , \quad n = 0, 1 2, \ldots \qquad (8)$$

being the variation $\Delta\vec{Z}^n$ calculated from

$$\Delta\vec{Z}^n = -\left(\left(J^T\right)^n J^n + \lambda^n \Gamma\right)^{-1} \left(J^T\right)^n \vec{F}^n \qquad (9)$$

where λ is the damping parameter, Γ the identity matrix, and the elements of the Jacobian matrix J for the heat conduction problem are

$$J_{ij} = \frac{\partial T_i}{\partial Z_j}, \; i = 1, 2, ..., N \text{ and } j = 1, 2, ..., M \quad (10a)$$

and for the radiative transfer problem are

$$J_{ij} = \frac{\partial I_i}{\partial Z_j}, \; i = 1, 2, .., N \text{ and } j = 1, 2, .., M \quad (10b)$$

The iterative procedure of sequentially calculating $\Delta\vec{Z}^n$ and $\vec{Z}^{n+1}$ from Eqs. (8) and (9), is continued until the convergence criterion

$$\left|\Delta Z_j^n\right| < \varepsilon, \qquad \text{for } j = 1, 2, ..., M \qquad (11)$$

is satisfied, where ε is a small number, say 10^{-5}.

The damping factor λ^n is varied during the iterative procedure, such that when convergence is achieved its value is close to zero. Here we follow the approach proposed by Marquardt [14].

THE MULTI-LAYER PERCEPTRON NEURAL NETWORK

The multi-layer perceptron (MLP) [12] is a collection of connected processing elements called nodes or neurons, arranged in layers (Fig. 1). Signals pass into the input layer nodes, progress forward through the network hidden layers and finally emerge from the output layer. Each node i is connected to each node j in its preceding layer through a connection of weight w_{ij}, and similarly to nodes in the following layer. A weighted sum is performed at i of all the signals x_j from the preceding layer, yielding the excitation of the node; this is then passed through a nonlinear activation function, f , to emerge as the output of the node x_i to the next layer, as shown in the equation

$$y_i = f(\sum_j w_{ij} x_j) \qquad (12)$$

Various choices for the function f are possible. In this work the hyperbolic tangent function $f(x)$ = tanh (x) is used.

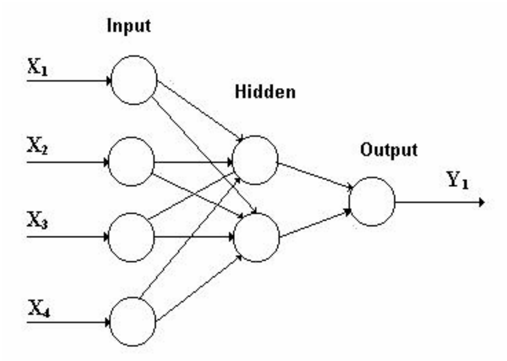

Figure 1 – Multi-layer perceptron network.

The first stage of using an ANN to model an input-output system is to establish the appropriate values for the connection weights w_{ij}. This is the "training" or learning phase. Training is accomplished using a set of network inputs for which the desired outputs are known. These are the so called patterns, which are used in the training stage of the ANN. At each training step, a set of inputs are passed forward through the network yielding trial outputs which are then compared to the desired outputs. If the comparison error is considered small enough, the weights are not adjusted. Otherwise the error is passed backwards through the net and a training algorithm uses the error to adjust the connection weights. This is the back-propagation algorithm used in the present work.

Once the comparison error is reduced to an acceptable level over the whole training set, the training phase ends and the network is established. The parameters of a model (output) can be, then, determined using the real experimental data, which are inputs of the established neural network. This is the generalization stage in the use of the ANN.

RESULTS AND DISCUSSION

Inverse heat conduction problem

For the thermal characterization of a new polymeric material with 25% in mass of lignin, which is obtained from sugar cane bagasse, we have used real transient temperature measured data acquired with a hot wire experimental apparatus [16]. Using the Levenberg-Marquardt (LM) method the following results were obtained: $k = 0.07319 \pm 0.00006\, \mathrm{W/m\ K}$ and c_p=1563 ± 3 $\mathrm{J/kg\ K}$.

In Case (1) of Table 1 are presented the estimates obtained at each iteration of the LM method starting with the initial guess $\vec{Z}^0 = (k^0, c_p{}^0) = (0.05, 1000)$. The value of the cost function, or objective function, Q, at each iteration is also presented. In Case (2), Table 2, LM starts from a different initial guess $\vec{Z}^0 = (k^0, c_p{}^0) = (0.05, 3500)$ and doesn't converge. The solution is probably a local minimum.

Table 1. Case 1 - $\vec{Z}^0$ = (0.05, 1000) – LM

Iteration	k (W/m K)	c_p (J/kg K)	Q (K^2)
0	0.05	1000.0	4285.75
1	0.06584	1361.6	262.53
2	0.07245	1537.7	2.6314
3	0.07319	1563.0	0.1179
4	0.07319	1563.0	0.1179

Table 3 shows the solution using a stochastic method alone (Simulated Annealing - SA). It converged after 35 cycles, which requires more than 7000 function evaluations. Table 4 represents a combination of LM with SA. After running SA for only five cycles we obtain the initial guess for the LM method. The combination worked well, with less computational cost than the one required to obtain the solution shown in Table 3.

Table 2. Case 2 - $\vec{Z}^0$ = (0.005, 3500) – LM

Iteration	k(W/m K)	c_p(J/kg K)	Q (K^2)
0	0.0050	3500.00	42124.58
1	0.0106	5602.83	5461.03
2	0.0224	6838.62	264.75
3	0.0418	4629.64	17.53
4	0.0418	4629.64	17.53

Table 3. Case 2 - $\vec{Z}^0$ = (0.005, 3500) – SA

Cycle	k(W/mK)	c_p (J/kg K)	Q (K^2)
0	0.0050	3500.0	42124.58
1	0.0509	3492.1	6.415
2	0.0539	3160.3	4.441
3	0.0576	2752.3	2.625
5	0.0708	1705.6	0.164
35	0.0732	1563.3	0.118

Table 4. Case 2 - $\vec{Z}^0$ = (0.005, 3500) – LM after five cycles of SA (Combination SA-LM)

Iteration	k(W/m K)	c_p (J/kg K)	Q (K^2)
0	0.0708	1705.6	0.164
1	0.0731	1561.7	0.132
2	0.0732	1563.0	0.118

In Table 5, the results using ANNs to obtain initial estimates for the LM method are presented. A decreasing number of patterns was used to train the network. Even with a small number of patterns the results obtained from the ANN are good enough to be used as initial estimates for the LM method. The patterns were generated using a finite difference approximation for problem (1).

The computational cost of the combination ANN-LM is lower than the one for the combination SA-LM.

The experimental data used for the solution of the inverse problem consisted of a set of 26 real temperature measurements taken at different times. Therefore, the input layer of the ANN consisted of $N = 26$ entries, and for the hidden layer of the ANN we started with $2N$ neurons, i.e. 52. After that we considered N neurons, i.e. 26. In this work only one hidden layer was used. As shown in Table 5, in all cases attempted the ANN generated, at a low cost, good initial guesses to be used in the LM method. Therefore, the combination ANN-LM is very promising.

Table 5. Case 2 – Using Neural Networks to obtain initial estimates for LM (Combination ANN-LM)

Number of patterns	N_{hidden}	k(W/m K)	c_p(J/kg K)
50	52	0.0703	1740.44
25	52	0.0715	1937.90
10	26	0.0690	2177.29
5	26	0.0698	2016.63

Remark: With the above initial estimates (ANN), LM converged to the expected values, $k = 0.0732$ W/m K and $c_p = 1563$ J/kg K with less than five iterations in all cases.

Inverse radiative transfer problem

As real experimental data was not available we have considered synthetic experimental data for the solution of the inverse radiative transfer problem. In this case we make

$$Y_i = I_{calc}\left(\vec{Z}_{exact}\right) + 2.576\, r\sigma \qquad (13)$$

where I_{calc} is obtained from the solution of problem (2) using the exact values for the unknowns, $\vec{Z}_{exact}$, r is a random number in the range [-1, 1], and σ represents the standard deviation of the measurement errors.

In Table 6 we present the results obtained using ANNs for the estimation of the single scattering albedo and optical thickness for the particular case with $\omega = 0.64$ and $\tau_o = 2.25$.

The results shown in Table 6 were obtained using simulated noiseless data, i.e. $\sigma = 0$ in Eq. (13).

The patterns used in the training stage of the ANNs were generated using a discrete ordinates method for the solution of problem (2).

A few test cases were also run considering noisy data. The results are shown in Table 7.

As expected some degradation on the estimation is observed when noisy data is used, but these estimates are still good enough to be used as initial guesses for the Levenberg-Marquardt method.

Table 6. Inverse radiative transfer problem – Using Neural Networks to obtain initial estimates for LM (Combination ANN-LM)

Number of patterns	N_{hidden}	ω	τ_0
50	40	0.5999	3.056
25	40	0.6237	2.427
10	20	0.6374	2.670
5	20	0.5752	2.457

Remark: With the above initial estimates (ANN), LM converged to the expected values, $\omega = 0.64$ and $\tau_0 = 2.25$, with less than ten iterations in all cases.

Table 7. Results for the inverse radiative transfer problem using ANNs and noisy data.

Number of patterns	Error			
	2 %		5 %	
	ω	τ_0	ω	τ_0
5	0.6437	2.819	0.6413	2.858
10	0.6543	2.643	0.6371	2.858
25	0.6308	3.121	0.6566	3.020
50	0.6279	2.8894	0.6374	3.176

Silva Neto and Soeiro [6,7] solved inverse radiative transfer problems using stochastic methods (SA - Simulated Annealing and GA - Genetic Algorithms), and combinations of these stochastic methods with the Levenberg-Marquardt method, i.e. SA-LM and GA-LM, but in all cases the computational cost was higher than the one required for the combination ANN-LM.

CONCLUSIONS

In the present work neural networks were used to obtain initial estimates for the LM method. This approach worked well for the proposed problems, showing promising results when the design space is complex with several local minima. The computational cost was also decreased when compared with the combination of a stochastic global optimization method, such as SA and a gradient based method, such as LM.

Acknowledgements

The authors acknowledge the financial support provided by CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico -, FAPERJ – Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro - and FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo.

REFERENCES

1. Alifanov, O. M., Solution of an Inverse Problem of Heat Conduction by Iteration Methods, *J. of Engineering Physics*, Vol. 26, pp. 471-476, 1974.

2. Beck, J. V., Blackwell, B. and St. Clair Jr., C. R., Inverse Heat Conduction, *Wiley*, New York, 1985.

3. Vogel, J., Sára, L. and Krejcí, L., A simple inverse heat conduction method with optimization, *Int. J. Heat Mass Transfer*, Vol. 36, No. 17, pp. 4215-4220, 1993.

4. Frankel, J. I. and Keyhani, M., A global time treatment for inverse heat conduction problems, *Journal of Heat Transfer*, Vol. 199, pp. 673-683, 1997.

5. Silva Neto, A. J., Explicit and implicit formulations for inverse radiative transfer problems, *Mini-Symposium on Computational Treatment of Inverse Problems in Mechanics - 5th World Congress on Computational Mechanics*, Vienna, Austria, 2002.

6. Silva Neto, A. J. and Soeiro, F. J. C. P., Estimation of phase function of anisotropic scattering with a combination of gradient based and stochastic global optimization methods, *5th World Congress on Computational Mechanics*, Vienna, Austria, 2002.

7. Silva Neto, A. J. and Soeiro, F.J.C.P., Solution of implicitly formulated inverse heat transfer problems with hybrid methods, *Mini-Symposium Inverse Problems from Thermal Fluids and Solid Mechanics Applications – 2nd MIT Conference on Computational Fluid and Solid Mechanics*, Cambridge, EUA, 2003.

8. Chiwiacowsky, L.D., Campos Velho, H.F. and Gasbarri, P., The damage identification problem: A hybrid approach, *2nd Topic Congress on Dynamics and Control*, pp. 1393-1402, São José dos Campos, Brazil.

9. Krejsa, J., Woodbury, K. A., Ratliff, J. D. and Raudensky, M., Assessment of strategies and potential for neural networks in the inverse heat conduction problem, *Inverse Problems in Engineering*, Vol. 7, pp. 197-213, 1999.

10. Shiguemori, E.H., Harter, F.P., Campos Velho, H.F. and Silva, J.D.S., Estimation of boundary conditions in heat transfer by neural networks, *TEMA: Tendências em Matemática Aplicada e Computacional*, Vol. 3, pp. 189-195, 2002.

11. Shiguemori, E.H., Campos Velho, H.F. and Silva, J.D.S., Estimation of initial condition in heat conduction by neural network, *4th International Conference on Inverse Problems in Engineering: Theory and Practice* (ICIPE-2002), Vol. II, pp. 251-258, Angra dos Reis, Brazil, 2002.

12. Bokar, J.C., The estimation of spatially varying albedo and optical thickness in a radiating slab using artificial neural networks, *Int. Comm. Heat Mass Transfer*, Vol.26, No.3, pp.359-367, 1999.

13. Boillereaux, L., Cadet, C. and Le Bail, A., Thermal properties estimation during thawing via real-time neural network learning, *Journal of Food Engineering*, Vol.57, pp.17-23, 2003.

14. Marquardt, D.W., An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Indust. Appl. Math*, Vol.11, pp.431-441, 1963.

15. Haykin, S., “Neural Networks: A Comprehensive Foundation”, Prentice Hall, Inc., 1999.

16. Carvalho, G. and Silva Neto, A. J., An inverse analysis for polymers thermal properties estimation, *Proc. 3rd International Conference on Inverse Problems in Engineering: Theory and Practice*, Port Ludlow, USA, 1999.