Neuro-Computation Techniques in Sampled-Data Electromagnetic Field Problems

Sied Mehdi Fakhraie, Adalbert Konrad, Kenneth Carless Smith
Department of Electrical and Computer Engineering
University of Toronto
Toronto, Ontario, Canada M5S 1A4

Abstract — In this paper, a technique is introduced by which to extend the applicability of the existing analytic solutions of electromagnetic field problems to cases where random-noisy-sampled data (such as measurement outputs) are available, rather than analytic input functions. We address those problems for which a theoretical solution exists in the form of a superposition of some basis functions. The algorithm introduced employs this same set of basis functions, and finds the expansion coefficients by the use of an iterative error-minimization technique, which resembles those found in the process of training of artificial neural networks. In cases where the expansion functions are orthonormal, guaranteed fast convergence is proved. As well, we show how neuro-computation techniques can be employed to circumvent the effects of various types of measurement errors and noise. Satisfactory performance of the algorithm is shown for a test problem driven by random inputs corrupted with various levels of Gaussian noise.

INTRODUCTION

There are many mathematically well-defined and rigorous solutions to time-harmonic boundary-value problems in electromagnetics [1]. Amongst them is modal analysis , in which we first identify a set of expansion functions satisfying boundary conditions. Then, using available input data such as sources or fields on the boundary, we proceed to find expansion coefficients from which to determine output fields. If the expansion functions satisfy the boundary conditions, the only problem that remains is that of matching the solution to real or equivalent sources.

While this approach is rigorous, difficulties arise in many practical cases (e.g. nondestructive testing, radar and remote sensing, and antenna-pattern estimation using near-field measurements) because instead of crisp-valued input functions, we have a collection of measurements corrupted by various types of error and measurement noise [2]. To resolve this practical difficulty, we propose an optimization-based approach similar to that found in the process of training neural networks, to achieve both input filtering and noise rejection, as well as effective calculation of the output fields in an iterative way.

Manuscript received November 1, 1993.

THE NEURO-COMPUTATION APPROACH

For a typical boundary-value problem, we assume that we have a general form of solution including a series of basis or expansion functions $f_i(x,y,z)$ (in rectangular coordinates) satisfying the boundary conditions, together with some unknown coefficients W_i. The goal is to find the coefficients which match the solution to the sources. However, the practical constraint is that rather than precise real or equivalent sources, we know only some noisy measured values at randomly selected points, e.g. $y^k_{measured}(x_k, y_k, z_k)$; $k=1,...,N$, where N is the number of measured points. In a noise-free case, for an arbitrarily large number M of expansion functions, we want to find values for W_i such that: $y^k_{measured} = y^k_{calculated} = \sum_{i=1}^{M} W_i \cdot f_i(x_k, y_k, z_k)$, for all k. However, in a noisy-data case, we expect the algorithm to fit the data and reject the noise. With the above assumptions, the classical analytical approach (or its numerical counterparts) cannot be used.

Borrowing from the neuro-computation literature, we follow an optimization process, similar to error back-propagation used in the training process of a neural network, to obtain those W_i values which minimize an overall error measure [3]. One possible error-measure is an estimate of Mean-Squared Error (MSE). Our proposed procedure consists of the following steps:

Step 1. Initialize W_i with random small values between 0 and 1, $i = 1,...,M$, where M is the number of expansion functions.

Step 2. If $k \le N$, calculate the output values: $y^k_{calculated} = \sum_{i=1}^{M} W_i \cdot f_i(x_k, y_k, z_k)$; Otherwise go to step 6.

Step 3. Calculate the error using the specified error function. For MSE, the error at each point is considered to be: $E^k = (y^k_{calculated} - y^k_{measured})^2$.

Step 4. Use one of the variations of the gradient-descent technique and change each W_i to a new $(W_i + \Delta W_i)$ such that the error is reduced. Correspondingly, $\Delta W_i(n,k) = -\eta \cdot \frac{\partial E^k}{\partial W_i}$ where k is the index specifying the input point, n is the iteration index, and η is the training rate which controls the adaptation speed. Consequently, $W_i(new) = W_i(old) + \Delta W_i(n,k)$.

Step 5. Increment k and go to step 2.

Step 6. Check the error and n; if $Error_{total} = \sum_{k=1}^{N} E^k$ is less than Target (a pre-specified value), or n is more than Max (the maximum allowable number of iterations), go to step 7. Otherwise increment n, set k=1, and go to step 2.

Step 7. If ($Error_{total}(n) - Error_{total}(n-1)$) is less than ε (a pre-specified small number), a local minimum has been reached, else increase n and try again. In certain situations the result would also be a global minimum (see Appendix 1).

About the error function, the MSE for the set $\mathbf{W}_n$ of weights can be defined as the expected value of the squared error. We assume that the input points are introduced by independent experiments with identical distributions. This assumption yields [5]:

$$E_{\mathbf{W}=\mathbf{W}_n} = \sum_{k=1}^{N} (1/N) (y^k_{calculated} - y^k_{measured})^2 ,$$

which converges to the true statistical average of $E_{\mathbf{W}=\mathbf{W}_n}$ as N approaches infinity.

Here, we are using an instantaneous error and gradient, which under the considered conditions, converge to the true statistical averages [Appendix 1]. The algorithm can accommodate a non-uniform probability distribution function to increase the number of samples where more precise measurements are possible.

The above algorithm is similar to the so-called on-line error back-propagation neural-network training process [3], or to a modified version of Widrow's Least-Mean Square (LMS) algorithm in adaptive filters [9]. For an explanation of the convergence conditions, we refer to Appendix 1.

Noise Filtering and Robustness

In addition to the capability of exploiting randomly measured input data, other major advantages appear when we use the noise-removal and robust techniques discussed in the neuro-computation literature [4]. For example, we can select the error function and adjust the training algorithm to minimize the errors caused by different types of noise. Thus, it has been shown that a mean-squared-error function minimizes the effects of random Gaussian noise [5], and that the Chebyshev norm function does so for uniform noise [6]. Using other variations of error functions can lead to a robust estimation [7]. In terms of statistics, robustness often means allowable deviation from an assumed Gaussian, uniform, or other probability density function. Robust estimators should be able to ignore infrequent observations which are significantly different from the remaining observations. Mean-squared measures tend to overemphasize infrequent samples or erroneous measurements. Alternatively, for example, long-tail noise or the effect of infrequent samples is reduced using an error function such as $2e/(1+e^2)$, where "e" is the difference between the computed and required values.

Numerical Verification

To verify the validity of our approach, we have chosen the classical problem of a step in a waveguide [1]. Fig. 1 illustrates the situation. We assume that only a dominant TE_{10} mode propagates towards the junction from the smaller guide, and we also assume that the larger guide is matched such that only outward-traveling waves exist. For an analytic solution, we suppose that E_y at the junction is that of the incident wave: $f(x,y) = E_y \mid_{z=0} = \sin\pi x/a$ if $y \le c$ and $0 \le x \le a$, and otherwise is zero. Then, we expand the fields in terms of the modes in the rectangular coordinates. Considering boundary conditions, the $E_y \mid_{z=0}$ component of the wave solution in the larger guide should be equal to f(x,y). We have:

$$f(x,y) = \sum_{m=1}^{\infty} \sum_{n=0}^{\infty} E_{mn} \sin\frac{m\pi}{a}x \cos\frac{n\pi}{b}y .$$

This is in the form of a double Fourier series. The E_{mn} are the Fourier coefficients of the expansion of $E_y \mid_{z=0}$, therefore:

$$E_{mn} = \frac{2\varepsilon_n}{ab} \int_0^a \int_0^b f(x,y) \sin\frac{m\pi}{a}x \cos\frac{n\pi}{b}y \, dx \, dy ,$$

where ε_n is Neumann's number and is $\varepsilon_n = 1$ if n=0, and $\varepsilon_n = 2$ if n>0. In the above situation, we have a complete analytic solution for the problem, but it requires complete information about f(x,y) at an infinite number of points. Having a limited number of points, the analytic model doesn't work. Fig. 2 illustrates the analytic stimulus function at the z=0 plane.

In the next step, we remove the strong condition of knowing the analytic stimulation function at the transition, and, instead, assume the input to be a collection of field measurements corrupted by Gaussian noise. We simulate possible measurement conditions by taking assumed analytic values for input (source) fields at the waveguide transition, and adding Gaussian noise with pre-specified variance to it. Using our approach with an MSE error function, we find that the expansion coefficients match the purely analytic ones within a few percent accuracy. As well, we have examined the effect of noise at different power levels. Table I shows an example of the results of our simulations with Gaussian sampling noise. Fig. 3 depicts one of the solutions obtained using our approach. As an example of the convergence speed, for 1000 sampling points, the algorithm converges in 1701 steps (less than two complete passes through data). This takes 5.7 seconds on a Sun SPARCstation 10.

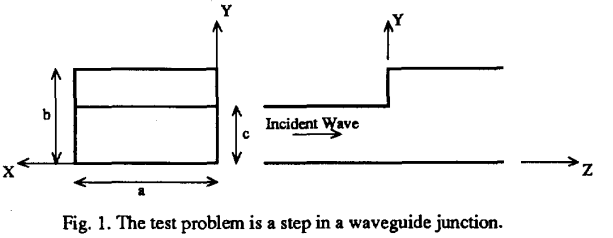

Fig. 1. The test problem is a step in a waveguide junction.

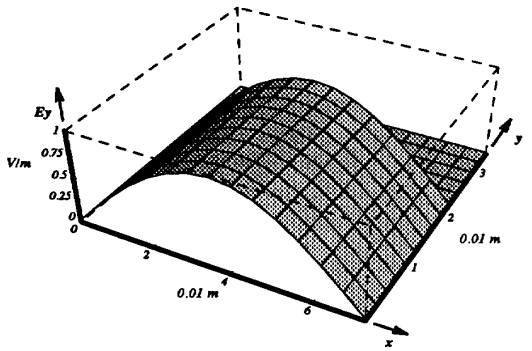

Fig. 2. Analytical stimulus function at the z=0 plane.

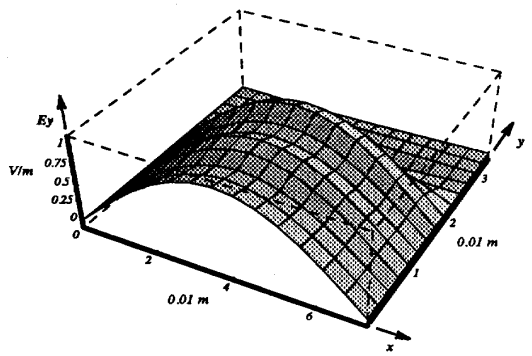

Fig. 3. Stimulation field obtained based on our approach using 700 random data points corrupted with Gaussian noise (variance=0.2).

Conclusions and Future Work

The approach proposed makes it possible for existing analytical solutions to electromagnetic boundary-value problems to be exploited in practical situations having only randomly-sampled input data. Also, it provides a means to minimize the errors caused by various types of noise. Although we have demonstrated its application to modal-analysis examples, it is more generally applicable to other structured-solution problems such as those using the moment method [8].

Acknowledgment

Helpful discussions with Professor S. Pasupathy, Professor D. A. Johns, and Mr. R. Khoini-Poorfard of the Department of Electrical and Computer Engineering at the University of Toronto are greatly appreciated.

References

[1] R. F. Harrington, *Time-Harmonic Electromagnetic Fields*. New York, NY: McGraw-Hill, 1961.

[2] S. Ponnapalli, *The Computation of Far-Field Antenna Patterns from Near-Field Measurements Using an Equivalent Current Approach*. Ph.D. Dissertation, Syracuse University, 1990.

[3] R. Hecht-Nielson, *Neurocomputing*. Reading, Mass. : Addison Wesley, 1990.

[4] B. Kosko, *Neural Networks and Fuzzy Systems*. Englewood Cliffs, NJ: Prentice Hall, 1992.

[5] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York, NY: McGraw-Hill, 1991.

[6] P. Burransco, P. Lucci, "A learning rule in the Chebyshev norm for multi-layer perceptrons," *Proc. Int. Joint Conf. on Neural Networks 1990* , vol. 3, pp. 81-86, San Diego, CA, 1990.

[7] P. J. Hubber, *Robust Statistics*. New York, NY: Wiley, 1981.

[8] R. F. Harrington, *Field Computation by Moment Methods*. Malabar, Florida: Robert E. Krieger Publishing Co. , 1983.

[9] B. Widrow, and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1985.

TABLE I
Simulation Results for Measurements with Gaussian Noise

Number of Measurements	Average Relative Error	Variance of Noise
700	13.72 %	1.0
700	10.5 %	0.5
1000	10.75 %	0.5
200	12.7 %	0.2
500	11.0 %	0.2
700	5.8 %	0.2
1000	5.8 %	0.2

The errors shown are the average of the relative errors in 24 numerically calculated expansion coefficients in comparison with ideal coefficients.

Appendix 1

Proof of Convergence

Fig. 4 illustrates the structure of a network which performs the proposed estimation of the unknown parameters.

The output unit is a linear unit. It adds up all of its inputs and provides a single linear output. $f_1(x,y)$, $\cdots$, $f_N(x,y)$ belong to a set of orthonormal basis functions.

(x,y) are the input variables to the algorithm, which are assumed to be determined by independent random-number-generation trials. We consider an ensemble composed of an infinite number of similar networks with similar weights and each driven by the outcomes of one of the random experiments, by which we determine the random inputs (x,y) to one of the networks. We also assign a corresponding desired output for each pair of inputs. Consequently, (x_k,y_k), with k as the time index, can be considered as a random process which is also assumed to be stationary.

Next, we consider each $f_i(x_k,y_k)$ as a memoryless system whose output is a direct function of its input without any dependence on the past values of the input. Therefore, we conclude that if (x_k,y_k) is in a strict sense stationary, $f_i(x_k,y_k)$ are also stationary [5].

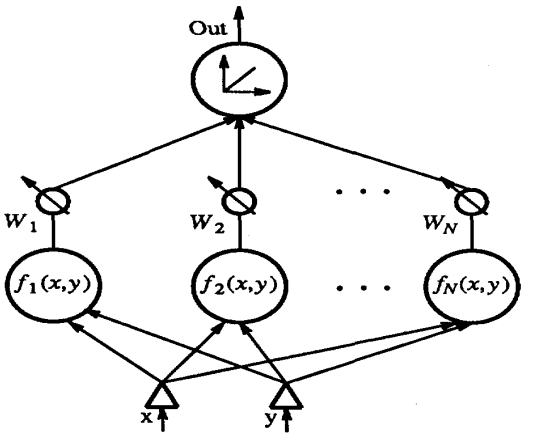

Fig. 4. The network architecture performing the estimation of the unknown parameters.

Then, we prove that at each instant of time (k constant), the functions $f_i(x_k,y_k)$ are uncorrelated. Our assumption is that each input random variable has a uniform probability density function (p.d.f) over a nonzero interval. This means that in our example, for any k, we have: $f_{x_k}(u)=1/a$ if $0\le u\le a$, and $f_{y_k}(v)=1/b$ if $0\le v\le b$; and otherwise both are zero. Based on the independence assumption, for their joint p.d.f we can write: $f_{x_k y_k}(u,v)=f_{x_k}(u)\cdot f_{y_k}(v)=1/ab$ if $0\le u\le a$ and $0\le v\le b$, and otherwise is zero. Assuming a zero-mean set of orthonormal basis functions, for each random variable $f_i(x_k,y_k)$, we find its mean to be zero:

$$E[f_i(x_k,y_k)]=\int_0^a\int_0^b f_i(x_k,y_k)\cdot(1/ab)dx\,dy=0.$$

Next, we calculate the cross-correlation of any two of these random variables. We note that $f_i(x_k,y_k)\cdot f_j(x_k,y_k)$ is again a function of two random variables x_k and y_k (we are talking about the same instant of time):

$$E[f_i(x_k,y_k)\cdot f_j(x_k,y_k)]=\int_0^a\int_0^b f_i(x_k,y_k)\cdot f_j(x_k,y_k)\cdot(1/ab)\,dx\,dy=0\ .$$

This results from the orthogonality of the basis functions and the properties of the assumed uniform distribution. Therefore, $E[f_i(x_k,y_k)\cdot f_j(x_k,y_k)]=E[f_i(x_k,y_k)]\cdot E[f_j(x_k,y_k)]=0$. This implies orthogonality and uncorrelatedness of any two random variables at the nodes of the middle layer at each instant of time.

At the next layer, we have a network similar to the adaptive linear combiner discussed in [9]. The inputs to this linear combiner are stationary random processes $\mathbf{f}=(f_1(x_k,y_k),\ \cdots\ ,f_N(x_k,y_k))$, which at each instant are also uncorrelated. Let y_n be the desired output of the n^{th} network in the ensemble and $\mathbf{W}$ be the common weight vector. Then the error at the n^{th} network is:

$$\varepsilon_n^2=(y_n-\mathbf{f}_n^T\mathbf{W})^2=y_n^2-2y_n\mathbf{f}_n^T\mathbf{W}+\mathbf{W}^T\mathbf{f}_n\mathbf{f}_n^T\mathbf{W}.$$

Averaging over the collection (ensemble) of the adaptive linear combiners yields:

$$E[\varepsilon_n^2]=E[y_n^2]-2E[y_n\mathbf{f}_n^T]\mathbf{W}+\mathbf{W}^TE[\mathbf{f}_n\mathbf{f}_n^T]\mathbf{W}.$$

Following the convention introduced in [9], we define the vector $\mathbf{P}$ as the cross correlation between y_n (the desired response) and the vector $\mathbf{f}_n$, and the matrix $\mathbf{R}$ as the matrix of correlation between f_i functions. Then the mean-squared error is:

$$\xi=E[\varepsilon_n^2]=E[y_n^2]-2\mathbf{P}^T\mathbf{W}+\mathbf{W}^T\mathbf{R}\mathbf{W}\ .$$

The MSE is a quadratic function of the weights. The matrix $\mathbf{R}$ is real, symmetric, and positive definite. Furthermore, In this problem, as we proved, any two f_i and f_j functions are uncorrelated and statistically orthogonal. This means that all the nondiagonal elements of the correlation matrix $\mathbf{R}$ are zero. Considering orthonormality condition, all the diagonal elements are equal and have the following value:

$$E[f_i^2\,f(x_k,y_k)]=\int_0^a\int_0^b f_i^2(x_k,y_k)\cdot 1/ab\,dx\,dy=1/ab.$$

The value (here $1/ab$) depends on the specific p.d.f of the input random variables, however, we can always say that it is a constant value for any uniform distribution. Therefore, $\mathbf{R}$ is invertible and its condition number equals 1, which is the best case considering lower sensitivity in numerical calculations. This also guarantees a unique global minimum,

$$\mathbf{R}=\lambda\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},\ \lambda\ \textit{is a constant value.}$$

The gradient with respect to $\mathbf{W}$ is $\nabla_{\mathbf{W}}\xi=-2\mathbf{P}+2\mathbf{R}\mathbf{W}$. Equating it to zero yields $\mathbf{W}_{optimal}=\mathbf{R}^{-1}\mathbf{P}$.

In practice, we do not have all the statistical information that we need to calculate the matrices directly. The stationarity assumption allows us to use a single network instead of the infinite ensemble of networks and to expose the input data to that single network in a time-serial basis. For each exposure, we can define an instantaneous error. Correspondingly, we use the instantaneous error as an estimate of the statistical average error and calculate an instantaneous gradient to perform a movement in the direction of steepest descent in the error space. The error and the gradient introduced in this paper obey the same assumptions. In [9], it has been shown that the instantaneous gradient is an unbiased estimation of the true gradient. As well, there, it has been shown that the bounds on the training rate η for the convergence of the algorithm are : $0<\eta<1/(trace[\mathbf{R}])$, in which *trace* [**R**] (the summation of the diagonal elements of **R**), here, is equal to the average signal power of our **f** vector.

In cases where the sampling is not uniform or the basis functions are not orthogonal, but if uncorrelatedness of the nondiagonal elements of $\mathbf{R}$ can be proved, the existence of its inverse and a unique global minimum is still guaranteed.