

MLP-BASED EQUALIZATION AND PRE-DISTORTION USING AN ARTIFICIAL IMMUNE NETWORK

Romis R. de F. Attux¹, L. T. Duarte¹, R. Ferrari¹, C. M. Panazio², L. N. de Castro³, F. J. Von Zuben¹, J. M. T. Romano¹

¹FEEC – State University of Campinas (Unicamp), Campinas, Brazil, ²CNAM – Conservatoire National des Arts et Métiers, Paris, France, ³UniSantos – Catholic University of Santos – Santos, Brazil.

E-mail: {romisri, ltduarte, rferrari, romano}@decom.fee.unicamp.br, cpanazio@cnam.fr, lnunes@unisantos.edu.br, vonzuben@dca.fee.unicamp.br

ABSTRACT

Due to its universal approximation capability, the multilayer perceptron (MLP) neural network has been applied to several function approximation and classification tasks. Despite its success in solving these problems, its training, when performed by a gradient-based method, is sometimes hindered by the existence of unsatisfactory solutions (local minima). In order to overcome this difficulty, this paper proposes a novel approach to the training of a MLP based on a simple artificial immune network model. The application domain for assessing the performance of the proposed technique is that of digital communications, in particular, the problems of channel equalization and pre-distortion. The obtained simulation results demonstrate that the proposal is capable of efficiently solving the problems tackled.

1. INTRODUCTION

A most relevant issue in digital communication systems is to find countermeasures to the impairments introduced by the channel and the devices that are essential to their proper operation. The idea of using a filter to reconstruct the original message from its corrupted version results in two possibilities: *equalization* and *pre-distortion*. The former approach is based on the design of a reception filter that mitigates the harmful effects of the medium, whereas the latter introduces a beneficial and controlled distortion in the signal before its transmission.

When the channel is nonlinear, or even if it engenders a complex linear problem, it becomes imperative that nonlinear filters be part of the solutions to both of the aforementioned problems. One of the most popular and traditional members of this class of structures is the MLP, a feedforward neural network with demonstrated universal approximation capability.

Notwithstanding these desirable features, the nonlinear character of the MLP renders the adaptation of

its parameters via classical gradient-based techniques rather complicated due to the multimodal character of the resulting MSE cost function. Another relevant issue is that the adaptation of any pre-distorter according to the classical framework requires *channel estimation (identification)*, which is a rather undesirable demand in some practical cases.

Several non gradient-based methods to train MLP networks have already been proposed in the literature (e.g. [1-3]). Among these approaches, evolutionary techniques play a major role because they perform a broad exploration of the search-space, thus being able to determine quasi-optimal weight sets (and other parameters, such as network architecture). In the present paper, we propose a new evolutionary technique to train MLP equalizers and pre-distorters, the basis of which is the artificial immune network model introduced in [4]. This proposal has several advantages over the standard back-propagation algorithm: it does not require the calculus of the derivative of the objective function; it provides a good balance between exploration and exploitation of the search-space; and it eliminates the need for channel identification in pre-distortion. The proposal will be applied to a nonlinear equalization problem and, as an original paradigm, to a pre-distortion problem.

2. MLP-BASED EQUALIZATION AND PRE-DISTORTION

The main objective of a communication system is to assure a proper message interchange between a transmitter and a receiver, both of which are interconnected by a channel, as shown in Figure 1.



Figure 1: Simplified Model of a Communication System.

The existence of a medium whereby information is sent implies that the message $s(n)$ will be received as a distorted version $r(n)$. Any attempt to reconstruct the

original information directly from this corrupted signal would probably lead to unacceptable bit error rates (BERs).

A direct approach to mitigate these noxious effects is to build a device that is, in general terms, the channel inverse. From this primary notion, two immediate strategies arise: 1) to place a filter at the receiver, i.e., to process $r(n)$ in order to obtain a sequence as close as possible to the transmitted message; or 2) to produce a controlled distortion that, when combined with the effects of the channel, give rise to a received signal that be equivalent to the original information. The first strategy is called *equalization*, while the second one is referred to as *pre-distortion*. Let us discuss both in more detail.

2.1 – Equalization

Due to its straightforwardness, the idea of *equalization* is as old as the modern communication theory. Its rationale is indeed very simple: to filter the distorted signal using a device that produces a mapping inverse to that originated by the channel. Figure 2 illustrates this process.

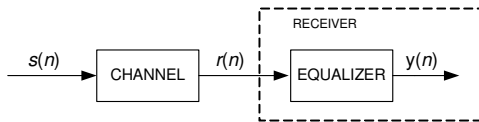


Figure 2: Equalization Scheme.

In the context of linear systems, the consequences thereof are immediate: channel and equalizer will have “opposite” spectral features. However, devices of this kind may be insufficient if the channel is nonlinear, or even if the linear problem is particularly complex (e.g. due to a bad choice of the equalization delay). In such cases, nonlinear structures like neural networks and fuzzy filters become attractive alternative approaches.

After selecting the adequate structure of the neural (or fuzzy) equalizer, it will be necessary to determine its parameters in a suitable way. If the channel is known *a priori*, this problem becomes relatively simple. However, when the channel is unknown, it is necessary to devise a cost function that guides the adaptation process. If there is a reference signal, a straightforward solution is to employ a *mean-square error* (MSE) cost function:

$$J_{EQ} = E[(s(n-d) - y(n))^2] \quad (1)$$

where d is the equalization delay and $E(\cdot)$ the expectation operator.

The MSE is, as attested by countless instances reported in the literature, a very solid basis for the adaptation process. Nevertheless, the use of filters derived from the field of computational intelligence opened a new perspective: to analyze the problem of equalizing a

channel as if it were a classification task, thereby taking advantage of the fact that, in digital systems, the transmitted samples always belong to a finite alphabet. However, this new perspective requires a formulation distinct from that we have just presented: instead of inverting the channel, we must conceive the equalizer as a device capable of minimizing the decision error. Using the MAP (Maximum a Posteriori) criterion and assuming the existence of AWGN (Additive White Gaussian Noise), it is possible to obtain the input-output relation of the optimal solution, viz, the *Bayesian equalizer* [5].

The Bayesian filter is the sample-by-sample equalizer *par excellence*, the benchmark against which every candidate equalizer should be measured. Two structures are privileged by their capacity of emulating the optimal device: the RBF (Radial Basis Function) neural network and a fuzzy filter with Gaussian membership functions [5].

The main drawback associated with this approach is that the number of noiseless channel states, which corresponds, for instance, to the number of neurons in the hidden layer of the RBF network, grows exponentially with the orders of both the channel and equalizer. This drawback attracted our interest towards another kind of neural network: the multilayer perceptron (MLP) [6], which will be discussed in more detail later.

2.2 – Pre-Distortion

The problem of pre-distortion can be understood as that of building a device that, by modifying the transmitted message in a controlled way, counterbalances the pernicious effects of the channel, as shown in Figure 3.

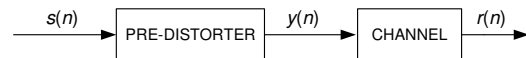


Figure 3: Simplified pre-distortion scheme.

The main advantage of this approach is that, in a broadcast system, it would be necessary to introduce a single device in the transmitter – the pre-distorter – instead of one equalizer per receiver. On the other hand, there is a relevant limitation: in such case it would be impossible for the pre-distorter to mitigate effects produced externally to the transmitter.

Pre-distorters (PD) are traditionally used in satellite communications to compensate for the nonlinear effects of amplifiers that, in order to increase their power efficiency, work near saturation [7]. The standard design of a PD follows two stages: 1) identification of the amplifier model; and 2) creation of a model inverse to that obtained in the first stage.

This twofold training procedure has, in our opinion, two major weaknesses: it requires, at least partially, a

priori knowledge of the features of the amplifier, and it renders essential the apparatus inherent to an identification problem. Therefore, it would be desirable to obtain the parameters of the optimal PD directly from, for instance, an error signal. A straightforward possibility is to use the traditional MSE cost function:

$$J_{PD} = E[(s(n) - r(n))^2] \quad (2)$$

Notwithstanding its inherent solidity, this supervised approach requires, in the context of pre-distortion, that a major difficulty be overcome: the derivatives of the cost function will, due to the relative positions in the scheme shown in Figure 3, depend on the channel model. Apparently, the project becomes a sort of Sisyphean task: the demand for system identification has arisen once more.

If our repertoire of optimization techniques were restricted to methods based on cost function differentiation, this would be a major difficulty in solving the pre-distortion problem. Nevertheless, there are search procedures that do not require this sort of mathematical manipulation. The so-called evolutionary algorithms, whose *modus operandi* depends solely on the definition of a fitness function, are examples of this kind of technique.

In this work, we will focus our attention on an evolutionary technique that belongs to the field of artificial immune systems: the opt-aiNet [4]. The opt-aiNet is a valuable tool for solving a wide range of optimization problems for two main reasons: 1) it presents a good balance between exploration and exploitation of the search-space; and 2) differently from other evolutionary proposals, it contains a mechanism devised to regulate population size and to maintain the diversity. Furthermore, this tool proved itself to be capable of successfully adapting a number of filtering structures in similar contexts [9], [10]. The opt-aiNet will be discussed in detail in section 3.

The synergy between the MSE cost function and the opt-aiNet is the key to the establishment of a new pre-distortion paradigm. Although the proposal be independent of the choice of any particular pre-distorter, it will be assessed in this work using an MLP neural network, in accordance with previous efforts reported in the literature [7] and the discussion presented in Section 2.1.

3. THE PROPOSED HYBRID SYSTEM: MLP TRAINED VIA OPT-AINET

3.1 – The Multilayer Perceptron

The MLP is one of the most popular neural network architectures for feedforward processing [6]. Its fundamental building block is the well-established artificial neuron model composed of a linear combiner and

a sigmoidal memoryless nonlinearity that works as activation function. The layered interconnection of a set of these units gives rise to a structure with multiple layers, the number of which is a design variable. As described in [6], an MLP with a single hidden layer and a linear output neuron has the ability of universal approximation, thus being complex enough for our purposes. The input-output relation of this single layer neural network can be written as:

$$y(n) = \mathbf{c}^T \varphi(\mathbf{A}\mathbf{x}(n) + \mathbf{b}) + k, \quad (3)$$

where \mathbf{A} is the weight matrix of the hidden layer, \mathbf{b} is the bias vector of the hidden layer, \mathbf{c} is the weight vector of the output layer, $\varphi(\cdot)$ is the sigmoidal activation function of the hidden layer, and k is the bias term of the output neuron.

Given the strong nonlinear character of this kind of filter, the resulting MSE cost function may contain multiple minima. This means that there will be a significant chance of convergence to a poor local optimum if a gradient-based technique, such as the back-propagation algorithm, is used to train the MLP network.

3.2 – The opt-aiNet Algorithm for Training MLP Networks

The acronym opt-aiNet stands for “Optimization version of an Artificial Immune Network” [4]. It is a particular type of artificial immune system (AIS) developed to solve optimization problems. AIS constitute a brand new computational intelligence approach that takes ideas from the immune system in order to develop new tools for solving multi-modal problems [4],[8]. Several theories, principles and models of the vertebrate immune system give rise to highly abstract models (algorithms) that can be applied to many domains, from biology to computing.

The opt-aiNet algorithm borrows ideas from two main theories about how the immune system operates, namely, clonal selection and affinity maturation and the immune network theory. In brief, clonal selection and affinity maturation state that the immune system is composed of sets of cells and molecules that carry receptors for antigens (disease-causing agents). When some of these receptors recognize an antigen, they are stimulated to proliferate and differentiate into either memory cells or effector cells. During proliferation, most cells suffer a controlled mutation process that followed by a natural selection mechanism allow the most adapted offspring cells to survive in detriment of the least adapted ones. This whole process has some peculiar features: cellular reproduction is asexual; the number of offspring a given cell may have is a direct function of its affinity (degree of recognition) with the antigen; the mutation suffered during

reproduction is inversely proportional to affinity (well-adapted individuals tend to be less disturbed).

In contrast to the clonal selection theory, to which the idea of cells acting in isolation is subjacent, the immune network theory describes the immune system as being composed of cells and molecules that interact with each other in a network-like form. These self-interaction patterns suggest a dynamic immune system with eigen-behaviors even in the absence of foreign stimulation (antigens). An antigen would thus be responsible for disturbing a self-organizing and self-sustainable system.

The opt-aiNet algorithm, which was developed from these two theories, clonal selection and immune network, can be summarized as follows (further details are provided in the sequence).

1. *Initialization*: randomly create an initial population of network cells;
2. *Local search*: while stopping criterion (see below) is not met, do:
 - 2.1 *Clonal expansion*: for each network cell, determine its fitness (an objective function to be optimized) and normalize the fitness vector. Generate a set of N_c antibodies, named *clone*, which are the exact copies of their parent cell;
 - 2.2 *Affinity maturation*: mutate each clone with a rate that is inversely proportional to the fitness of its parent antibody, which itself is kept unmutated. The mutation follows Eq. (4) below. For each mutated clone, select the one with highest fitness and calculate the average fitness of the selected cells;
 - 2.3 *Local convergence*: if the average fitness of the population does not vary significantly from one iteration to the other, go to the next step; else, return to Step 2;
3. *Network interactions*: determine the affinity (similarity) between each pair of network antibodies (see below);
4. *Network suppression*: eliminate all network cells whose affinity with each other is greater than a pre-specified threshold (σ_s), and determine the number of remaining cells in the network; these are named memory cells;
5. *Diversity*: introduce a number of new randomly generated cells into the network and return to Step 2.

When the opt-aiNet algorithm described above is used to train an MLP, each cell corresponds to an

individual of the population and is represented as a real-valued vector in a Euclidean space. This vector contains all the weights of the MLP network to be adapted; for example, if the MLP network has one input, two hidden units and one output, then there are seven connections (weights) to be updated, all of which are concatenated into a single vector.

The fitness of a given cell corresponds to the value of the objective function (Eqs. (7) and (9)) at the point characterized by its parameter vector. The affinity between two cells corresponds to their Euclidean distance. Clones are offspring cells that are the progenies of a single cell; they are originally identical to the parent cell, but, after mutation, they become slight variations of their parent cell. The fitness-proportional mutation of Step 2.2 is performed according to the following expressions:

$$\begin{aligned} c' &= c + \alpha N(0,1), \\ \alpha &= (1/\beta) \exp(-f^*) \end{aligned} \quad (4)$$

where c' is a mutated cell c , $N(0,1)$ is a Gaussian random variable of zero mean and standard deviation $\sigma=1$, β is a parameter that controls the decay of the inverse exponential function, and f^* is the fitness of an individual normalized in the interval $[0,1]$. A mutation is accepted only if the mutated antibody c' is within the domain range. Finally, the standard stopping criterion is based on the cardinality of the memory population.

4. RESULTS

In order to evaluate the performance of the opt-aiNet algorithm in the MLP-based equalization context, consider the following nonlinear channel model:

$$r(n) = x(n) + 0.3x(n)^2 + v(n), \quad (5)$$

where $r(n)$ is the received signal, $v(n)$ is a zero mean additive white Gaussian noise and $x(n)$ represents the channel ISI (inter-symbol interference), which is given by:

$$x(n) = 0.5547s(n) + 0.8320s(n-1) \quad (6)$$

The transmitted signal $s(n)$ is considered to be an independent and identically distributed (i.i.d.) sequence of symbols belonging to a binary $\{\pm 1\}$ alphabet. Since the opt-aiNet algorithm is an optimization tool that searches for multiple maxima of a given function, it is necessary to define a new cost function, J_{FIT_EQ} , whose maxima are the minima of J_{EQ} :

$$J_{FIT_EQ} = \frac{1}{1 + J_{EQ}} \quad (7)$$

The simulations will investigate the effectiveness of the proposed paradigm in terms of the estimated BER for several signal-to-noise ratio (SNR) values. We shall

compare our proposal to an MLP trained by the classical back-propagation algorithm [6] and to the optimal Bayesian solution. The parameters in the simulations, including those related to MLP structure and training, were defined after several trials. In the simulations related with the equalization problem, we used an MLP equalizer with 6 neurons in the single hidden layer and 2 inputs, $r(n)$ and $r(n-1)$. The equalization delay was fixed at $d = 0$, since this is the worst-case scenario. The opt-aiNet parameters were set to the values shown in the column “Equalization” of Table 1 and, for each SNR value, the MLP equalizer was trained by the opt-aiNet algorithm

Table 1: Opt-aiNet parameters.

Opt-aiNet Parameter	Equalization	Pre-Distortion
Initial Population Size	5	20
σ_s	5	3
Number of offspring per cell	5	25
β	30	80

using 3,000 transmitted symbols. The cost function (7) was estimated through time-averaging, using a window of 150 past samples of $s(n)$ and $r(n)$. The back-propagation algorithm used in the simulations is a standard implementation with learning rate 0.04 and *momentum constant* 0.1. In this case, the MLP equalizer was trained using 20,000 transmitted symbols for each SNR value. After the training procedures, the MLP parameters were fixed and the BER was measured by means of continuous symbol transmission until either 3,000 errors were observed or a maximum of 10^7 symbols were sent.

The simulation results, including the Bayesian equalizer curve, are shown in Figure 4. The curve of the opt-aiNet training represents an average of 10 experiments, while the back-propagation results are presented in two curves: one for the best achieved performance and another for the average of 50 experiments. The number of experiments was such that a reliable estimate of the mean curve could be computed. The average performance of the MLP equalizer trained via opt-aiNet algorithm is nearly the same as that of the Bayesian equalizer, which demonstrates the robustness and the solidity of the proposed paradigm. A similar behavior is observed in the curve of the best back-propagation result, but, nevertheless, due to the existence of local optima, there are experiments in which the back-propagation algorithm has a poor performance, which explains the high BER values present in the average curve. On the other hand, the computational complexity of the opt-aiNet algorithm is higher than that of the back-

propagation. In order to improve the average results in training based on iterative nonlinear optimization, it is possible to use more elaborate initialization strategies [11] and also second-order optimization methods [12]. However, the problem related to the convergence to possibly poor local optima will persist.

Now, consider the problem of designing a pre-distorter to compensate for the nonlinear distortions of a power amplifier, whose AM-AM and AM-PM responses are given respectively by [7]:

$$A(|s(n)|) = \frac{\alpha_A |s(n)|}{1 + \beta_A |s(n)|^2}, \quad \Phi(|s(n)|) = \frac{\alpha_\Phi |s(n)|}{1 + \beta_\Phi |s(n)|^2}, \quad (8)$$

where α_A , β_A , α_Φ and β_Φ are model constants. For this simulation, the values of the amplifier parameters are set to $\alpha_A = 11.534$, $\beta_A = 1.6242$, $\alpha_\Phi = 11.431$ and $\beta_\Phi = 39.071$. The transmitted signal $s(n)$ is considered to be an i.i.d. sequence belonging to a 16 QAM alphabet. Both the transmitted symbols constellation and the power amplifier output without pre-distortion are shown in Figure 5. In the experiments, we used a complex-valued MLP with a single input and 5 neurons in the single hidden layer as pre-distorter. The perceptron was trained by the opt-aiNet algorithm to maximize the following cost function:

$$J_{FIT_PD} = \frac{1}{1 + J_{PD}} \quad (9)$$

The cost function (9), which represents the fitness of each individual in the population of the opt-aiNet algorithm, was evaluated using all the 16 possible $s(n)$ values and the respective amplifier output $r(n)$. A J_{FIT_PD} value of 0.99 was set as the stopping criterion to the algorithm. The opt-aiNet parameters were set to the values shown in the column “Pre-Distortion” of Table 1. Figure 6 shows a typical amplifier output with pre-distortion together with the 16 QAM constellation. As we can see, the pre-distorter performs its task very well, removing almost all the distortions and recovering the original 16 QAM constellation.

5. CONCLUSIONS

This work introduced a new paradigm for neural network training based on an immune-evolutionary technique called opt-aiNet. In order to verify the performance of the proposed paradigm, we chose two representative problems in digital communications: channel equalization and pre-distortion.

The average performance of the proposal in the context of equalization of a nonlinear channel was very close to the optimal solution. This attests the robustness of the immune-evolutionary approach, as well as its remarkable capability of avoiding local convergence,

which is a major drawback of gradient-based training techniques.

The proposed training scheme also gave rise to a new pre-distortion strategy that does not require knowledge of the amplifier model. The results confirm both the viability and the effectiveness of the proposal.

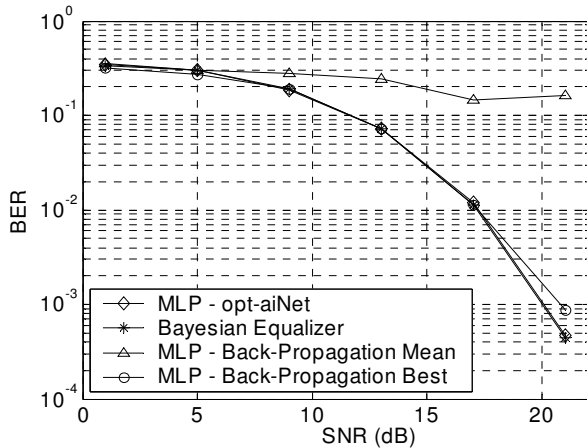


Figure 4: BER curves for the MLP equalizers and for the Bayesian equalizer.

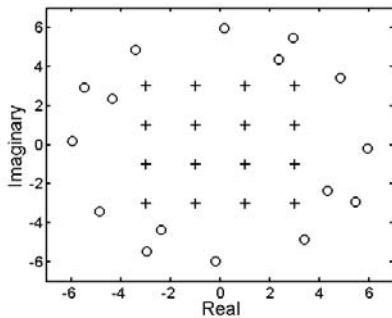


Figure 5: 16 QAM constellation (+) and amplifier output without PD (o)

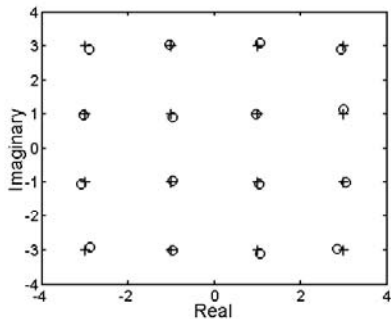


Figure 6: 16 QAM constellation (+) and amplifier output with PD (o).

ACKNOWLEDGEMENTS

The authors thank FAPESP and CNPq for the financial support.

REFERENCES

- [1] A. Doering, M. Galicki and H. Witte, "Structure Optimization of Neural Networks with the A*-Algorithm", *IEEE Transactions on Neural Networks*, 8(6):1434-1445, 1997.
- [2] Xin Yao, Evolutionary artificial neural networks. In A. Kent and J. G. Williams, editors, *Encyclopedia of Computer Science and Technology*, 33:137-170. Marcel Dekker Inc., New York, 1995.
- [3] V. G. Gudise and G. K. Venayagamoorthy, "Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks" in *Proceedings of the IEEE Swarm Intelligence Symposium 2003 (SIS 2003)*, Indianapolis, Indiana, USA, pp. 110-117, 2003.
- [4] L. N. de Castro and J. I. Timmis, "Artificial Immune Systems: A New Computational Intelligence Approach", Springer-Verlag, London, 2002.
- [5] S. Chen, B. Mulgrew and P. M. Grant, "A Clustering Technique for Digital Communications Channel Equalization Using Radial Basis Function Networks", *IEEE Trans. on Neural Networks*, 4(4):570-579, 1993.
- [6] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, 1998.
- [7] M. Ibnkahla, "Neural network predistortion technique for digital satellite communications" in *Proceedings of ICASSP*, 6:5-9, June 2000.
- [8] L. N. de Castro and J. I. Timmis, "Artificial Immune Systems as a Novel Soft Computing Paradigm", *Soft Computing Journal*, 7 (8):526-544, 2003.
- [9] R. R. de F. Attux, M. B. Loiola, R. Suyama, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano, "Blind Search for Optimal Wiener Equalizers Using an Artificial Immune Network Model", *EURASIP Journal of Applied Signal Processing*, 2003(8):740:747 2003.
- [10] R. R. de F. Attux, L. N. de Castro, F. J. Von Zuben, J. M. T. Romano "A Paradigm for Blind IIR Equalization Using the Constant Modulus Criterion and an Artificial Immune Network," in *Proceedings of the IEEE NNSP*, Toulouse, France, 2003.
- [11] L. N. de Castro and F. J. Von Zuben, "A hybrid paradigm for weight initialization in supervised feedforward neural network learning", *ICS-Workshop on Artificial Intelligence*, pp. 30-37, Tainan, Taiwan, December 1998.
- [12] R. Battiti, "First- and Second-Order Methods for Learning: Between Steepest Descent and Newton's Method", *Neural Computation*, 4(2):141-166, 1992.