GRAY BOX IDENTIFICATION WITH HOPFIELD NEURAL NETWORKS

Miguel Atencia[1], Departamento de Matemática Aplicada, E.T.S.I. Informática

Gonzalo Joya y Francisco Sandoval, Departamento de Tecnología Electrónica, E.T.S.I. Telecomunicación

Universidad de Málaga, Málaga, Spain

ABSTRACT

In this work, a novel method, based upon Hopfield neural networks, is proposed for parameter estimation in the context of system identification. This subject is a very active field of research, because even when a model of a physical system is available, some parameters may be uncertain or time varying. In our methodology, identification is formulated as an optimization problem, profiting from the applicability of Hopfield networks to this kind of problems. In order to compare the novel technique and the classical gradient method, simulations have been carried out for a linearly parameterized system, and results show that the Hopfield network is more efficient than the gradient estimator, obtaining lower error and less oscillations. Also, the neural technique is applied with encouraging results to non-linearly parameterized systems, for which few methods have been proposed.

Key words: System identification, Hopfield neural networks, parameter estimation, adaptive control, optimization.

RESUMEN

En este trabajo se propone un método novedoso, basado en redes neuronales de Hopfield, para estimación de parámetros en el contexto de identificación de sistemas dinámicos. Es este un tema sujeto a gran actividad investigadora, ya que incluso cuando se dispone de un modelo de un sistema físico, puede haber parámetros inciertos o variables. Con nuestra metodología, la identificación se formula como un problema de optimización, sacando partido de la aplicabilidad de las redes de Hopfield a este tipo de problemas. Para comparar la nueva técnica con el clásico método de gradiente, se han realizado simulaciones de un sistema linealmente parametrizado, y los resultados muestran que la red de Hopfield es más eficiente que el estimador de gradiente, obteniendo un error más bajo y menores oscilaciones. Asimismo, la técnica neuronal se ha aplicado, con resultados alentadores, a sistemas parametrizados no linealmente, para los que actualmente hay pocos métodos propuestos.

1. INTRODUCTION

System identification can be defined as the characterization of a dynamical system, by observing its measurable behaviour. Identification has been studied since decades (see Abe (1989)) for a recent review) from a variety of viewpoints and research communities, such as statistical regression -estimation-, signal processing -filtering-, and control engineering -adaptive control-. When a priori information about the rules that govern a system either do not exist or are too complex, e.g. in electrocardiographic signals, identification techniques are used to build a model by only observing input-output data. Then, the system is called a "black box" since the internal behaviour is unknown. On the contrary, in many physical systems, our knowledge of mechanical, chemical or electrical laws enables us to formulate a model, which is the fundamental tool for studying the system, either analytically or through simulations. However, no matter how deep is our physical insight, the parameters of any model present inaccuracies. For instance, measurement devices involve some error, transmission systems produce noise, a robot can take a stone with unknown mass or the friction

[1]**E-mail**: matencia@ctima.uma.es

coefficient of some material can only be approximately known. Due to these perturbations, although the model is still valid, the numerical value of its parameters must be confronted with observed data, and the system is called a "gray box" since identification must be applied to validate or adapt the a priori formulated model, which is regarded as a *parametric model* (Unbehauen (1996)).

This paper is focused in on-line identification for gray-box models, i.e. continuously obtaining an estimation of the system parameters. This subject has received increased attention after the successful development of adaptive control methods (Marino (1997)) that consist of an identification module and a controller that would achieve optimal control, should the actual system match the identified one.

The mathematical formulation of a dynamical system is a -possibly vectorial- ordinary differential equation[1] (ODE) $\dot{x} = f(x, u, \theta)$ where x(t) is the state vector, u(t) is the input vector, which can be regarded as a control signal, and the parameter vector (t) may be uncertain and/or time-varying. From this formulation, gray box identification is achieved by parameter estimation, i.e. by calculating at each time instant a value $\hat{\theta}$ that approaches the actual value as much as possible, thus minimizing the estimation error $\tilde{\theta} = \hat{\theta} - \theta$. Therefore, the continuous estimation evolves together with the observation of the dynamical system, so the estimation is a dynamical system itself given by an ODE $\dot{\hat{\theta}} = g(\hat{\theta}, x, u)$.

In the above paragraph, we have formulated identification as minimization of the estimation error $\tilde{\theta}$. In doing so, we emphasize that the appropriate framework for identification is optimization. Indeed, most proposed techniques for identification are somehow based on optimization methods. On one hand, the simplest method for on-line identification is the gradient method, which is supported by the same rationale as gradient descent optimization: the estimation should evolve in the direction that best minimizes the error, which is the -negative- gradient of the error function. However, the actual parameter value θ is unknown and so it is the estimation error $\tilde{\theta}$. Thus, the norm of the output prediction error $\|e\| = \|f(x, u, \theta) - \dot{x}\|$, or equivalently its square $\|e\|^2$, is instead used as the target function, and the gradient method leads to the estimator dynamics $\dot{\hat{\theta}} = -k \nabla \|e(\hat{\theta})\|^2$. The tuning of the estimation gain k is a key aspect of the estimator design and must be repeated for each particular system. If k is small, the convergence to the actual parameter is slow. If k is too large, oscillations may appear that cause not only larger error but also high frequency noise, which may excite unmodelled dynamics if the estimation is a part of a closed loop controller. On the other hand, several methods have been proposed for batch estimation, i.e. for identification that is performed only once, after the observation of the complete evolution of the system. Most classical batch estimators are based upon minimization of least square error, while heuristic techniques have recently been proposed based upon genetic algorithms (Pedroso (2002)), which are known to be global optimizers. It is remarkable that most identification techniques have been applied to the simpler "linear in the parameters" (LIP) systems, with the form f(x, u, θ) = A(x, u) θ, while identification of nonlinearly parameterized systems is still almost unexplored.

In this contribution, we apply the optimization methodology of Tank and Hopfield (1985) to the identification problem, by designing a Hopfield neural network whose Lyapunov function is made coincident with the prediction error, so that the network evolution approaches a minimum of the error. When compared to the gradient method, the proposed technique is shown to alleviate the problematic tuning of gain, achieving fast convergence of estimations to the actual values of parameters, as well as less oscillations. With respect to batch estimators, the ability of the Hopfield estimator to track time varying parameters is shown. Besides, a method to deal with non-LIP models, based upon the higher order Hopfield neural network (Samad-Harper (1990)), is presented with promising results. The only requisite of the proposed methods is the knowledge of a bounded region where the actual values of parameters remain.

---

[1]Along the paper, we adhere to the notation $\dot{x}$ for the time derivative.

The organization of this paper is as follows: the well known gradient method is briefly reviewed in Section 2, together with its usual application to linear in the parameters systems. In Section 3, after describing the application of Hopfield neural networks to optimization, parameter estimation with Hopfield networks is presented. Section 4 is dedicated to the application of higher order Hopfield networks to nonlinearly parameterized systems. The efficiency of these methods is assessed by simulations in Section 5. Finally, conclusions are summarized in Section 6.

2. GRADIENT ESTIMATION

In gray box identification, a system model is known but it includes some unknown, uncertain and/or time varying parameters. In the simplest and most usual case, the model is in the LIP form $\dot{x} = A(x,u)\theta$. It will be useful for our purposes to consider the "expected" or "nominal" –known- values of the parameter vector θ_n, so that the system model results in:

$$y = \dot{x} = A(t) \ (\theta_n + \theta(t))$$

where y is the output, θ is the unknown -possibly time dependant- deviation from the nominal values and A is a matrix that depends on the input u and the state x. Both y and A are assumed to be physically measurable. Identification is accomplished by producing an estimation $\hat{\theta}(t)$ that is intended to continuously minimize the estimation error $\tilde{\theta} = \hat{\theta} - \theta$. An additional important objective is to reduce the output prediction error that, in the linear in the parameters form, reduces to $e = A\left(\theta_n + \hat{\theta}\right) - y = A\,\hat{\theta}$. Equivalently, the square of the norm of the prediction error can be chosen as the target function:

$$\|e\|^2 = e^T e = \left(A\,\tilde{\theta}\right)^T\left(A\,\tilde{\theta}\right) = \tilde{\theta}^T\,A^T\,A\,\tilde{\theta}$$

A common technique for on-line estimation is the gradient method that, due to its simplicity, presents some advantages over least mean squares algorithms in estimation of time varying parameters. In the gradient method, the estimation is continuously modified in the direction that best minimizes the prediction error, i.e. the direction of the gradient of $\|e\|^2$ as a function of the estimation vector $\hat{\theta}$, thus leading to the dynamical equation of the estimator:

$$\dot{\hat{\theta}} = -k\,\nabla(e^T e) = -2k\,A^T e = -2k\,A^T\left[A(\theta_n + \hat{\theta}) - y\right]$$

$$= (-2k\,A^T A)\hat{\theta} - 2k\,A^T(A\theta_n - y)$$

where the fact $\partial e/\partial\hat{\theta} = \partial e/\partial\tilde{\theta} = A$ has been used, and k is a design parameter, which must be critically chosen as a trade-off between small values -slow convergence- and large values -oscillations-. The last equality emphasizes the fact that the gradient method leads to a linear differential equation, when applied to a LIP system. This simplification is not preserved in more complex models, and that is a reason why the gradient estimation is not usually used for non-LIP models. The asymptotical convergence to zero of the prediction error can be proved if A and y are constant, but the result is also valid as long as A and y change slowly (Slotine-Li (1991)), which we assume in the sequel. However, for the estimation to converge to the actual values, some condition of "persistent excitation" must hold. This is due to the possible existence of several estimation values that annihilate Eq. , apart from the obvious actual value $\hat{\theta} = \theta$, whereas if the input signal is rich enough then A and y are so complex that the only value of $\hat{\theta}$ that leads to $\dot{\hat{\theta}} = 0$ is the actual

parameter. This richness of the input signal, called persistent excitation in the control literature, will also be assumed.

3. ESTIMATION WITH HOPFIELD AND TANK NETWORKS

The neural network paradigm comprises a variety of computational models, among which Hopfield networks (1982 and 1984) are feedback systems for which a Lyapunov function has been found. This feature guarantees the stability of the network, which can then be useful for the solution of two important problems: associative memory and optimization. In the Abe (1989) formulation, the evolution of a continuous Hopfield network is governed by the following system of differential equations:

$$\dot{\mu} = net(s); \qquad s = \tanh(\mu/\beta)$$

where β is a parameter that controls the slope of the hyperbolic tangent function and the *net* term depends linearly on every state s_i, so that in the general high-order form Samad-Harper (1990), it is a multilinear function of the state vector:

$$net_i = \sum_{j=1}^{q-1} \frac{1}{j!} \sum_{i_1,i_2,...i_j} w_{i,i_1,i_2,...i_j} \, s_{i_1} s_{i_2} ,... s_{i_j} - I_i \tag{5}$$

The computational ability of Hopfield networks stems from the fact that they are stable dynamical systems, which is proved by the existence of a Lyapunov function:

$$V = -\sum_{j=2}^{q} \frac{1}{j!} \sum_{i_1,i_2,...i_j} w_{i_1,i_2...i_j} s_{i_1} s_{i_2} ... s_{i_j} + \sum_i I_i s_i \tag{6}$$

Note that the definitions are chosen so that the identity $\partial V/\partial s_i = - net_i$ holds, which leads to $\dot{V} = \sum_{i=1}^{n} \frac{\partial V}{\partial s_i} \dot{s}_i = \sum_{i=1}^{n} \frac{\partial V}{\partial s_i} \frac{ds_i}{d\mu_i} \dot{\mu}_i = \sum_{i=1}^{n} (net_i)^2 \tanh'(\mu_i) \le 0,$ and the other conditions for V being a Lyapunov function are trivially satisfied (see Joya **et al**. (2002)) for a discussion on Hopfield dynamics). The parameters w and I are called weights and biases, respectively. A remarkable property of Hopfield dynamics, is that states remain in a bounded region, namely the hypercube $s_i \in$ [-1, 1], due to the presence of the bounded function tanh. Therefore, the intended solution should belong to this hypercube, otherwise it will never be attained.

The application of the Hopfield model to optimization is a consequence of its stability: since the network seeks a minimum of its Lyapunov function, it can be regarded as a minimization method, as long as the target function can be identified with the Lyapunov function. The parameters w, I are then obtained from the numeric values in the target function (Tank-Hopfield (1985)). In order to perform system identification, the squared norm of the prediction error, as defined in Eq. , is chosen as the target function. Consider the Lyapunov function candidate V = ½e^T that, after some algebra, results in:

$$V = \frac{1}{2} \hat{\theta}^T A^T A \hat{\theta} + \hat{\theta}^T (A^T A \theta_n - A^T y) + V_1 \tag{7}$$

where V_1 does not depend on $\hat{\theta}$ so it can be neglected without changing the position of the minima of V. Therefore, this equation is identical to the standard Lyapunov function of a first order Hopfield network -Eq. (6) with q = 2- whose states represent the estimation, as long as the weight matrix and the bias vector are appropriately defined:

$$W = -A^T A; \qquad I = A^T A\theta_n - A^T y \tag{8}$$

It is remarkable that the only difference between the obtained Hopfield dynamics and the already known gradient method given by Eq. is the presence of the nonlinear sigmoid-like function tanh. A requirement for the application of this methodology is the knowledge of a bounded region around θ_n, because the network states can not abandon the hypercube $s_i \in [-1, 1]$. However, this region is not required to coincide with the hypercube, since it could be mapped into the hypercube by an appropriate scaling. The proposed method is simulated and its performance is compared to the gradient method in Section 5.

4. NON LINEARLY PARAMETERIZED SYSTEMS

The physical insight about some systems suggest that parameters enter nonlinearly into the model, i.e. the model is not in the LIP form. Some methods have been proposed to deal with non linearly parameterized models, mainly based on their reduction to the LIP form, or in the context of adaptive control (Marino-Tomei (1993)). However, no general technique has been shown to be effective for a wide range of systems. Our aim is to develop an extension of the Hopfield estimator, presented in the previous section, which is able to perform parameter estimation in non-LIP models. As observed in Eq. (6), the Lyapunov function of a high-order Hopfield network is not limited to the quadratic case, but any multilinear function matches this structure. Thus minimization may be performed on multilinear functions by the already known procedure of identifying the target function and the Lyapunov function. Although this remark extends the applicability of Hopfield optimization networks, it does not completely solve the problem of non linearly parameterized models. However, it is well known that any function can be approximated to any desired degree of exactness by a multilinear function, which is simply the Taylor series of the original function.

Let a general system model be defined by the ODE $\dot{x} = f(x,u,\theta)$. The proposed method for parameter estimation consists in, first, approximating f by its Taylor polynomial at $\theta = 0$, regarding f as a function of θ:

$$f(x,u,\theta) \approx f(x,u,0) + \sum_i \left.\frac{\partial f(x,u,\theta)}{\partial \theta_i}\right|_{\theta=0} \theta_i + \frac{1}{2}\sum_i \sum_j \left.\frac{\partial^2 f(x,u,\theta)}{\partial \theta_i \partial \theta_j}\right|_{\theta=0} \theta_i \theta_j + \ldots \tag{9}$$

The choice $\theta = 0$ for the point of expansion is a reasonable assumption, because the "expected" value of θ is zero, as long as the "expected" value of the parameter is θ_n. Once the expansion of f has been performed, the approximation is a multilinear function, and so it is the prediction error e = $f(x,u,\theta) - y$. Then, the target function $V = e^T e$ is identified with the Lyapunov function in Eq. (6), so obtaining the values of the weights and biases. In order to apply this method, bounds of parameters must be known in advance, as in the previous section.

5. SIMULATION RESULTS

5.1. A linear in the parameters model

We will show the efficiency of the proposed method by simulating an idealized single link manipulator (Figure 1), modelled by the following equation:
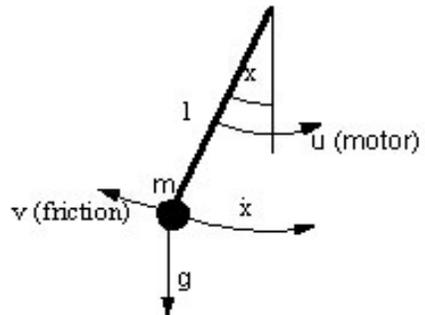
Figure 1. Single link manipulator.
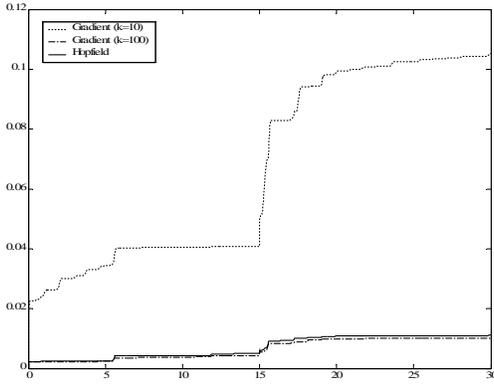
Figure 3. Integral of prediction error.

$$y = \ddot{x} = -\frac{g}{l}\sin x - \frac{v}{ml^2}\dot{x} + \frac{1}{ml^2}u$$

(10)

where x is the angle between the manipulator and the vertical,

$\dot{x}$ is its angular velocity, g is the gravity, l is the manipulator arm length, v is the friction coefficient, m is the mass that is being transported at the arm extreme and u is the torque produced by a motor, which can be regarded as the input or control signal. If the gravity is assumed constant, this model can be converted into the "linear in the parameters" form by the transformation $(\theta_n + \theta)^T = (-g/l, -v/ml^2, 1/ml^2)$, $A = (\sin x, \dot{x}, u)$ so that equation (1) holds. The control signal u is defined as a sum of three sinusoids displaced according to the Schroeder phase Ljung (1995), to achieve persistent excitation without excessive amplitude.

The setting of the simulated experiment is as follows: At the beginning, the estimator vector $\hat{\theta}$ is null, but the actual parameters differ from the nominal values so that $\theta(t = 0) = (0.89 \quad 0.25 \quad -0.29)^T$. Besides, the system is subject to a variable load, the transported mass abruptly changing at t = 15 sec. The gains k = 10 and k = 100 have been chosen for the gradient estimator, while no special attention has been paid to the selection of an optimal value for the parameter β of the Hopfield network. The performance of the Hopfield network is graphically compared to the gradient method. In particular, the error in the first component θ_1 is plotted in Figure 2, where we observe a short transient during which all three estimators oscillate. Then, the Hopfield network suddenly converges to the actual value, the gradient estimator with k = 100 oscillates wildly and the gradient with k = 10 also converges, but its convergence is so slow that it has not reached the actual value when the parameter change at t = 15 sec occurs. After the load change, both gradient estimations are distorted, while the network is practically unaffected, which is logical as θ_1 does not depend on the mass. It is clear the problematic trade-off of the gradient method between oscillatory answer (k = 100) and slow convergence (k = 10), while the neural network converges quickly and its oscillations are appropriately damped. We also compare the integral of the prediction error of the Hopfield estimator to that of the gradient method in Figure 3. The Hopfield network and the gradient with k = 100 both produce an accurate prediction, but the gradient with k = 10 accumulates significant error, mainly after the load change at t = 15 sec .

5.2. A non linearly parameterized model

As an illustrative example of the application of high-order Hopfield network to parameter estimation, consider the following non-LIP system:

$$y = \dot{x} = f(x, u, \theta) = u + x\theta_1 e^{\theta_2 x} \qquad (11)$$

whose Taylor approximation is
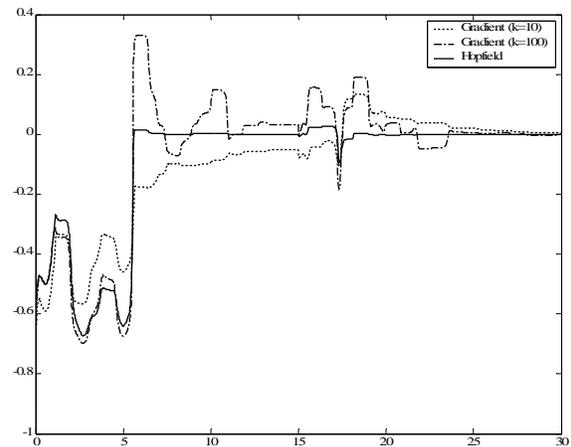
Figure 2. Estimation error for θ_1.

$f(x,u,\theta) \approx u + x\theta_1 + \frac{1}{2} x^2\theta_1\theta_2 + \frac{1}{2} x^2\theta_1\theta_2$. In order to apply the Hopfield and Tank optimization methodology, the prediction error is identified to the Lyapunov function:

$$V \equiv \|e\|^2 = \|x - f(x,u,\theta)\|^2 \approx (2ux - 2x\dot{x})\theta_1 + x^2\theta_1^2 + (2ux^2 - 2x^2\dot{x})\theta_1\theta_2 + 2x^3\theta_1^2\theta_2 + x^4\theta_1^2\theta_2^2 \quad (12)$$

which results in the following weights and biases (null parameters have been omitted):

$$I_1 = -(2ux - 2x\dot{x}) \quad w_{11} = 2x^2 \quad w_{112} = w_{121} = w_{211} = -2x^3$$

$$(13)$$

$$w_{12} = w_{21} = -(2ux^2 - 2x^2\dot{x}) \quad w_{1122} = \cdots = w_{2211} = -x^4$$

Figure 4. Parameter estimation for θ_1.

Figure 5. Parameter estimation for θ_2.

The identification performance of the high-order Hopfield estimator, based upon the Taylor approximation, has been tested on this system. The simulation comprises an initial parameter error, as well as a sudden parameter change at t = 20 sec . In Figures 4 and 5, the ability of the estimator to track parameter variations is shown. The estimation of θ_1, which appears linearly in the model, is quite accurate. However, the estimation of θ_2 suffers from a degradation when this parameter moves far from the nominal value, although finally the estimation converges to the actual value. This behaviour is due to θ_2 entering nonlinearly into the model, which produces a smaller influence of θ_2 in the Taylor expansion, as is observed in Eq. (12) and the fact

$I_2 = w_{22} = 0$. In order to deal with this problem, further research is being developed, trying to determine the optimal selection of the parameter β as well as the appropriate order of the Taylor series. On the other hand, the ability of the neural estimator to predict the output is excellent: In Figure 6 predicted and actual output are not distinguishable, as can be realized by the almost constantly null prediction error. This is a consequence of using output prediction error as target function for optimization.

6. CONCLUSIONS

In this work, the optimization methodology of Hopfield and Tank is applied to parameter estimation, in the context of system identification. When restricting to linear in the parameters systems, simulation results show that this technique presents faster convergence as well as oscillation reduction when compared to the known gradient
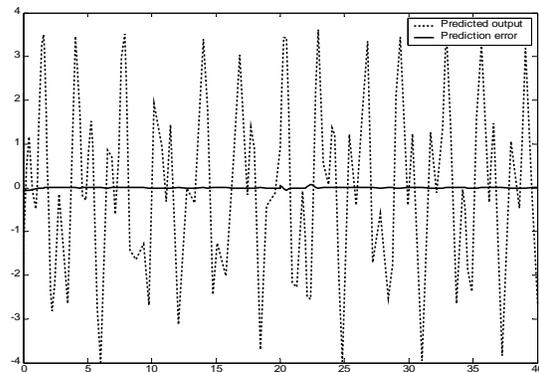
Figure 6. Predicted and actual output.

method. For nonlinearly parameterized systems, the gradient method would lead to a nonlinear dynamical equation, and few other methods have been proposed. However, the high-order Hopfield network can be applied to general systems, with essentially the same methodology as in the linearly parameterized model. The needed adaptation is the approximation of the model function by its Taylor series so that it can be cast into the multilinear form of the Lyapunov function. Despite this approximation, the high-order Hopfield estimator presents reduced error and fast convergence in the estimation of time varying parameters. Simulations have been performed for simplified examples with promising results. Current research is directed towards the study of an optimal choice of the parameter β, the determination of the appropriate order of the Taylor expansion, the inclusion of the estimator into an adaptive control system and the application of neural estimators to real-size problems.

REFERENCES

ABE, S.(1989): "Theories on the Hopfield neural networks", In **Proc. IEE International Joint Conference on Neural Networks**, I, 557–564.

HOPFIELD, J. (1982): "Neural networks and physical systems with emergent collective computational abilities", **Proc. Natl. Acad. Sci.** USA, 79:2554–2558.

_____ (1984): "Neurons with graded response have collective computational properties like those of two-state neurons", **Proc. Natl. Acad. Sci.** USA, 81:3088-3092.

JOYA, G.; M. A. ATENCIA and F. SANDOVAL (2002): "Hopfield neural networks for optimization: Study of the different dynamics", **Neurocomputing** 43(1-4):219–237.

LJUNG, L. (1999): **System Identification. Theory for the User**, Prentice Hall.

MARINO, R. (1997): "Adaptive control of nonlinear systems: Basic results and applications", **Annual Reviews in Control**, 21:55–66.

MARINO, R. and P. TOMEI (1993): "Global adaptive output-feedback control of nonlinear systems, part II: Nonlinear parameterization", **IEEE Transactions on Automatic Control**.

PEDROSO, L.M. (2002): Aplicación de algoritmos evolutivos a la solución del problema de estimación de parámetros en modelos definidos por ecuaciones diferenciales ordinarias, Master's thesis, Universidad de La Habana.

SAMAD, T. and P. HARPER (1990): "High-order Hopfield and Tank optimization networks", **Parallel Computing**, 16:287–292.

SLOTINE, J.J. and W. LI. (1991): **Applied Nonlinear Control**, Prentice Hall.

TANK, D. and J. HOPFIELD (1985): "_Neural_ computation of decisions in optimization problems", **Biological Cybernetics**, 52:141–152.

UNBEHAUEN, H. (1996): "Some new trends in identification and modeling of nonlinear dynamical systems", **Applied Mathematics and Computation**, 78:279–297.