ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

Statistical properties of the 'Hopfield estimator' of dynamical systems

Miguel Atencia[1] and Gonzalo Joya[2] *

1- University of Málaga - Dept of Applied Mathematics
Campus de Teatinos, 29071-Málaga - Spain
matencia@uma.es

2- University of Málaga - Dept of Electronics Technology
gjoya@uma.es

Abstract. This paper analyses the statistical properties of a method for estimating the parameters of systems defined by ordinary differential equations. Previously, this estimator was defined as an adapted version of Hopfield neural networks, and its convergence and robustness with respect to signal disturbances were proved, even when parameters are time-varying. This contribution aims at analysing the estimation error by performing a set of simulations where a random noise with known probability distribution is added to signals. It is shown that, asymptotically, the estimator is unbiased and its variance vanishes. Further theoretical work is being undertaken in order to rigourously support these empirical findings.

1 Introduction

System identification [1] can be defined as the characterization of a dynamical system, by observing its measurable behaviour. Identification has been studied from a variety of viewpoints, such as statistical regression, signal processing, and adaptive control. In this work we deal with systems defined by Ordinary Differential Equations (ODEs), although the method proposed is amenable to be extended to more general models, such as Stochastic Differential Equations and Delay Differential Equations. Regardless of the way a model is obtained, either from the intuition of experts or from the application of physical laws, it usually contains some parameters whose numerical value is uncertain and, possibly, time-varying. Classical methods for parameter estimation include least squares techniques and gradient based methods [2], but these algorithms present some limitations when dealing with time-varying parameters.

There exist two main approaches to the analysis of algorithms for parameter estimation: deterministic and probabilistic. Since estimation algorithms are usually described by a recurrent definition, they constitute dynamical systems whose analytical properties contribute to assess the estimation performance. In particular, a desirable property is the convergence of the estimations towards the actual values of the parameters or, at least, towards a bounded region around

*This work has been partially supported by Projects No. TIN2008-04985 by the Spanish Ministerio de Ciencia e Innovación, No. P08-TIC-04026 by the Junta de Andalucía, and No. D/023835/09 by the Agencia Española de Cooperación Internacional para el Desarrollo. The useful suggestions of the anonymous reviewers are gratefully acknowledged.

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

these values, i.e. the asymptotic boundedness of the estimation error. A different, although related, view concerns the probabilistic properties of the estimation when the system states are considered to be affected by random noise modelled by a probability distribution. Then, the analysis aims at obtaining the distribution of the estimation or, at least, its mean and variance. A satisfactory practical performance is expected when an estimator is unbiased (its mean is the actual value of the parameter) and efficient (its variance is minimal) or, at least, if these favourable properties are attained asymptotically.

An estimation method has been proposed [3], based upon an adapted version of continuous Hopfield networks [4] showing a remarkable ability to deal with time-varying paraments. A rigourous theoretical analysis proved the asymptotic boundedness of the estimation error. It was also proved [5] that the estimation error remains bounded when the system states are affected by disturbances. These results are briefly reviewed in Section 2. Despite convergence results provide a valuable insight into the behaviour of the estimation algorithm, the methods of proof are essentially nonconstructive, thus no hint is provided about the size of the bounded region that the error converges to. Indeed, this bound could have no practical significance, because the estimation error could be unacceptably high. Therefore, in this paper, we start a probabilistic analysis of the estimator, aimed at establishing the practical performance of the estimator, when signals are disturbed by noise with a known probability distribution.

The estimator is applied to a system of ODEs, which was proposed as a model of HIV epidemics in Cuba [6]. Simulated data are obtained by numerically integrating the system and perturbing the system states with gaussian noise of known variance. This experimental setting is repeatedly performed in order to obtain a statistically significant sample of the distribution of the estimations. The analysis of this sample shows that the mean of the estimations converges to the actual values of the parameters, even though the noise does not vanish. This result, discussed in Section 3, suggests that the estimator is asymptotically unbiased and efficient. This is only a first step in the probabilistic analysis of the estimator, so we have pointed out future directions of research in Section 4.

2 A 'neural' method for parametric identification

In this section we define a method for parameter estimation and state some of its dynamical properties. To this end, consider a model defined by a system of ODEs, which we assume to be in the *Linear In the Parameters (LIP)* form:

$$\frac{d\,\boldsymbol{x(t)}}{d\,t} = \boldsymbol{A}\left(\boldsymbol{x}(t)\right)\,\boldsymbol{\theta}\left(t\right) + \boldsymbol{b}\left(\boldsymbol{x}(t)\right) \tag{1}$$

where $\boldsymbol{\theta}$ is a vector of parameters. The notation can be simplified by defining the vector $\boldsymbol{y} = \dfrac{d\,\boldsymbol{x}}{d\,t} - \boldsymbol{b}$, so the system model is described by the equation $\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{\theta}$. Most estimation techniques proceed by determining estimations $\hat{\boldsymbol{\theta}}$ that minimize the *prediction error* $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{A}\,\hat{\boldsymbol{\theta}} = \boldsymbol{A}\,\tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ is the estimation error.

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

Thus, consider the optimization problem $\min_{\hat{\boldsymbol{\theta}}} V$ where the target function V is the squared norm of the prediction error:

$$V = \frac{1}{2}\,\|\boldsymbol{e}\|^2 = \frac{1}{2}\,\boldsymbol{e}^\top\boldsymbol{e} = \frac{1}{2}\,\left(\boldsymbol{A}\,\tilde{\boldsymbol{\theta}}\right)^\top\left(\boldsymbol{A}\,\tilde{\boldsymbol{\theta}}\right) = \frac{1}{2}\,\tilde{\boldsymbol{\theta}}^\top\,\boldsymbol{A}^\top\,\boldsymbol{A}\,\tilde{\boldsymbol{\theta}} \tag{2}$$

An estimation method results from defining a particular optimization algorithm to minimize the prediction error. The proposed method stems from the optimization capability [7] of the Hopfield neural network [8, 4], which in the Abe formulation [9] is defined by the following system of ODEs:

$$\frac{d\,u_i}{d\,t} = \sum_j w_{i\,j} s_j - I_i\;;\quad s_i = \tanh\frac{u_i}{\beta} \tag{3}$$

where s_i is the state of neuron i, $w_{i\,j}$ and I_i are the network weights and biases, respectively, and β is a design variable than can be, in principle, fixed arbitrarily. The procedure for the application of Hopfield networks to optimization consists in matching the target function to the network Lyapunov function:

$$V(\boldsymbol{s}) = -\frac{1}{2}\sum_i\sum_j w_{ij}\,s_i\,s_j + \sum_i I_i\,s_i$$

so that the weights and biases are obtained. Then, the network is implemented until it reaches a minimum of the target. The comparison of this Lyapunov function and the target function from Equation (2), leads to the definition: $\boldsymbol{W} = -\boldsymbol{A}^\top\boldsymbol{A}$ and $\boldsymbol{I} = -\boldsymbol{A}^\top\boldsymbol{y}$. In this case, unlike the application of Hopfield networks to combinatorial optimization problems [10], weights and biases are time-varying, since they are computed from the states of the dynamical system.

The dynamical properties of the defined estimator have been theoretically analysed. Under mild assumptions, it was proved that the estimations converges towards a bounded region around the actual values of parameters, even when parameters are time-varying [11] and the measurements $\boldsymbol{y}$, $\boldsymbol{A}$ of the dynamical model are perturbed by a bounded disturbance [5]. However, the analysis so far is deterministic, so the size of such bounded region has not been related to the magnitude of the signal disturbance. Thus, in the rest of this contribution we aim at characterizing the statistical properties of the estimation with regard to the probability distribution of the noise.

3 Experimental results

By way of benchmark, we apply the estimator to the identification of a system of ODEs, which has been proposed as a model of HIV epidemics in Cuba [6]:

$$\begin{aligned}
\frac{d\,x}{d\,t} &= (\lambda - k_1)\,x + \lambda'\,(y+z) - k_2\frac{x(y+z)}{x+y+z} \\
\frac{d\,y}{d\,t} &= k_1\,x - \mu\,y \\
\frac{d\,z}{d\,t} &= -\,\mu\,z + k_2\frac{x(y+z)}{x+y+z}
\end{aligned} \tag{4}$$

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

where x, y and z represent different infected populations; λ, λ' and μ are parameters assumed to be known; and, finally, k_1 and k_2 are parameters that must be estimated, since they provide an assessment of health policies. In order to test the estimation in several situations, k_1 is assumed to be time-varying, whereas k_2 is constant. The model is simulated from the initial value $(x_0, y_0, z_0) = (200, 0, 0)$, and data points are recorded at discrete intervals $\Delta t = \frac{7}{365.25}$ years. Then, the computed populations are disturbed by noise with a normal distribution $\mathcal{N}(0, 2)$. This setting is intended to mimic to some extent the real available data.

The population data obtained by the procedure above described, is supplied to the proposed estimation method, which is in turn numerically integrated along the same time period. Since only discrete data have been recorded, they are interpolated in order to match the finer time step of the integrator. First of all, the estimator was applied to a noiseless data set, resulting a perfect fit, with a negligible Sum of Squared Errors (SSE) of $SSE = \sum_t \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^2 < 10^{-15}$. In this experiment, the design variable β, which appeared in Equation (3), was set to 10^{-5}. Once the performance of the estimator has been established in a deterministic context, the estimator was simulated 100 times, by providing disturbed signals. The obtained estimations were averaged along the 100 instances and the results are shown in Figure 1. It is obvious that the average estimations present a significant coincidence with the actual parameter values. Further, the mean estimation error converges towards zero, which is an empirical support for considering the proposed estimator as asymptotically unbiased. Also, the value of the SSE, where the sum is carried over all the time instants, remains low.

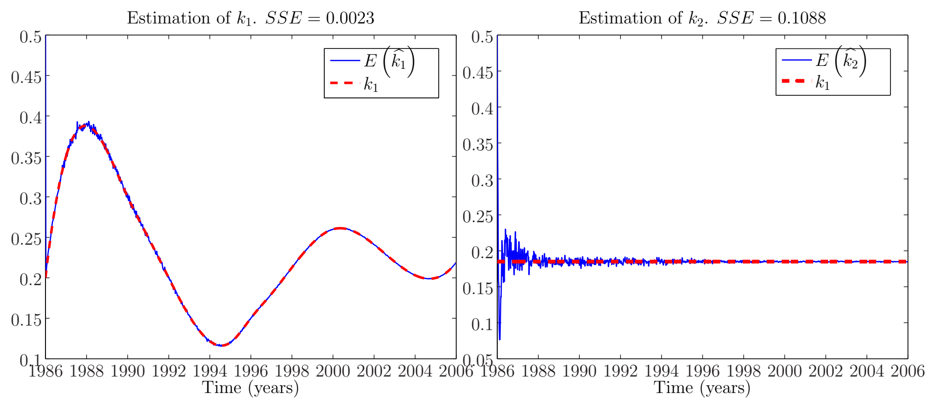

Fig. 1: Average estimation of parameters k_1 and k_2.

The satisfactory performance of the estimator is confirmed by the analysis of the variance of the estimations, shown in Figure 2. It is now evident that the probability distribution of the estimator converges towards a degenerate distribution coincident with the parameter values, despite the fact that the noise does not vanish. It is interesting to compare these findings with the known theoretical results [12]. A fundamental result states that there is a lower limit,

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

called the Cramér-Rao bound, to the values of the variance of any unbiased estimator. Further, under some conditions on the correlation of the data samples, it is proved that this minimal variance is proportional to $\frac{1}{\sqrt{n}}$, where n is the number of data instances. Although theoretical research is needed in order to prove these properties, these simulations provide preliminary evidence that the estimator is asymptotically efficient, whereas data that stems for the model are *informative enough* to help the estimation convergence. This theoretical context also provides an interpretation for the fact that the variance of the estimation of k_2 is significantly larger than that of k_1 or, equivalently, the convergence of the former is much slower than the latter: the variance of the estimator depends on the *sensitivity* of the system solution with respect to the parameter. It is reasonable to expect that the dependance of the populations on the parameter k_2 is critical, due to the stronger influence of the nonlinear terms that multiply k_2 in the model given by Equation (4).

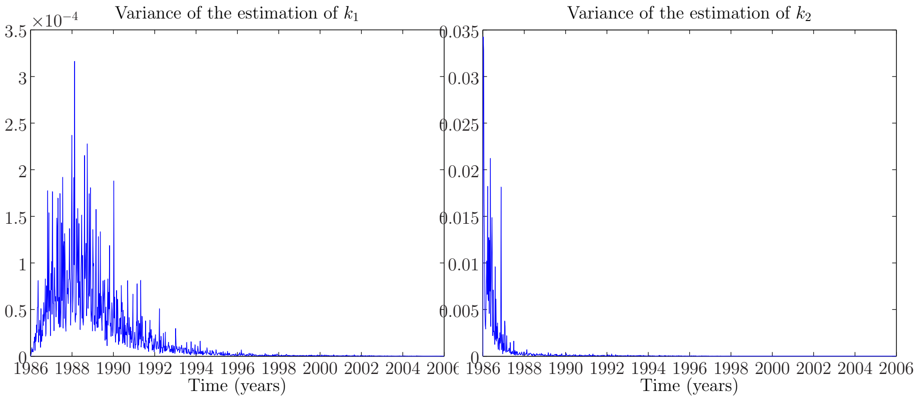

Fig. 2: Variance of the estimation of parameters k_1 and k_2.

A final remark concerns the choice of the design variable β that appears in the definition of the estimator, Equation (3). The initial set of experiments was attempted with the same value that performed satisfactorily with noiseless data, namely $\beta = 10^{-5}$. However, numerical instabilities appeared, so that either the estimations converged towards erroneous values or they did not converge at all. Then, the value of β was progressively increased until $\beta = 0.1$, when satisfactory results were attained.

4 Conclusions and future directions

We have reviewed the statistical properties of an algorithm based upon the optimization capability of Hopfield networks, previously proposed for parameter estimation of dynamical systems. Empirical results suggest that the mean of the estimations converges to the actual values of parameters, whereas the variance converges to zero, i.e. the estimator is asymptotically unbiased and efficient.

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence
and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

The present work is currently being expanded in four directions. Firstly, comparisons are being carried out to assess the proposed estimator with regard to classical techniques, such as least squares. Preliminary results show that the 'Hopfield estimator' provides lower error, faster convergence and less computational cost. Secondly, the analytical expression of the variance is being developed, in order to compare its value to that of the theoretical minimum given by the Cramér-Rao bound. Besides, the experiments are being reproduced with a variety of systems and noise levels with the aim to confirm that the findings of this paper can be extended with wider generality. Finally, we are exploring the optimal choice of the design variable β, which seems to play the role of a regularization parameter. Also, different values of β for the estimators of the different parameters k_1, k_2, could be defined in accordance to the diverse sensitivities of the model states. This rationale could lead to a global enhancement of the estimation results.

References

[1] Lennart Ljung. *System identification : theory for the user.* Prentice Hall, 1999.

[2] Jean-Jacques Slotine and Weiping Li. *Applied nonlinear control.* Prentice Hall, 1991.

[3] Miguel Atencia, Gonzalo Joya, and Francisco Sandoval. Hopfield Neural Networks for Parametric Identification of Dynamical Systems. *Neural Processing Letters*, 21(2):143–152, 2005.

[4] J J Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 81(10):3088 –3092, May 1984.

[5] Hugo Alonso, Teresa Mendonça, and Paula Rocha. Hopfield neural networks for on-line parameter estimation. *Neural Networks*, 22(4):450–462, 2009.

[6] M. Atencia, G. Joya, and F. Sandoval. Modelling the HIV-AIDS Cuban Epidemics with Hopfield Neural Networks. In José Mira and José Álvarez, editors, *Artificial Neural Nets Problem Solving Methods*, volume 2687 of *Lecture Notes in Computer Science*, pages 1053–1053. Springer, 2003.

[7] J. J Hopfield and D. W Tank. "Neural" computation of decisions in optimization problems. *Biological Cybernetics*, 52:141–152, 1985.

[8] J J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8):2554 –2558, April 1982.

[9] Shigeo Abe. Theories on the Hopfield Neural Networks. In *Proc. IEE International Joint Conference on Neural Networks*, volume I, pages 557–564, 1989.

[10] Miguel Atencia, Gonzalo Joya, and Francisco Sandoval. Dynamical Analysis of Continuous Higher Order Hopfield Networks for Combinatorial Optimization. *Neural Computation*, 17(8):1802–1819, 2005.

[11] Miguel Atencia, Gonzalo Joya, and Francisco Sandoval. Parametric Identification of Robotic Systems with Stable Time-varying Hopfield Networks. *Neural Computing & Applications*, 13(4):270–280, 2004.

[12] Adriaan van den Bos. *Parameter estimation for scientists and engineers.* Wiley-Interscience, 2007.