CONSIDERING THE MEASUREMENT NOISE FOR A NONLINEAR SYSTEM IDENTIFICATION WITH EVOLUTIONARY ALGORITHMS

Zoé Sigrist[1,3], *Pierrick Legrand*[2], *Eric Grivel*[1], *Benoît Alcoverro*[3]

1 Université Bordeaux 1 - IPB ENSEIRB-MATMECA - IMS - UMR CNRS 5218,
2 Université Bordeaux Segalen - ALEA Team - INRIA Bordeaux Sud-Ouest - IMB - UMR CNRS 5251,
351 cours de la libration 33405 Cedex, TALENCE, France.
3 Commissariat à l'Energie Atomique CEA CESTA BP 2 - 33114 LE BARP, France.

ABSTRACT

This paper deals with the identification of a nonlinear system modelled by a nonlinear output error (NOE) model when the system output is disturbed by an additive zero-mean white Gaussian noise. In that case, standard on-line or off-line least squares methods may lead to poor results. Here, our approach is based on evolutionary algorithms. Although their computational cost can be higher than the above methods, these algorithms present some advantages, which often lead to an "effortless" optimisation. Indeed, they do not need an elaborate formalisation of the problem. When their parameters are correctly tuned, they avoid to get stuck at a local optimum. To take into account the influence of the additive noise, we investigate different approaches and we suggest a whole protocol including the selection of a fitness function and a stop rule. Without loss of generality, simulations are provided for two nonlinear systems and various signal-to-noise ratios.

Index Terms— nonlinear output-error (NOE), biased estimates, genetic algorithms, differential evolution.

1. INTRODUCTION

System identification plays a key role to understand, to analyse and/or to predict the behavior of "real world" systems. For instance, when studying mechanical structures under vibrating environment, the purpose is to retrieve the resonant frequency and the quality factor from the model parameters.

Two main tasks must be addressed for system identification: searching mathematical models and estimating model parameters.

1/ Modelling: the observed data can take the form of time series. In the nonlinear case, the family of nonlinear autoregressive moving average with exogenous input (NARMAX) models allows the system input and system output to be related by using a nonlinear mapping function. Among the most popular models, the Volterra series which are linear regarding the model parameters but nonlinear regarding the input have the advantage of being stable in the bounded input-bounded output (BIBO) sense. However, the memory and the order must be often set to large values. Therefore, a NARX model -with a polynomial form for convenience- can be prefered.

2/ Parameter estimation: for a given model structure, the set of model parameters has to be estimated. This can be done in the least squares (LS) sense, in the maximum likelihood (ML) sense, by using instrumental variables (IV), etc.

When dealing with models linear regarding the parameters and when noise-free data are available, standard on-line or off-line LS methods can be used. When both the system input and output are disturbed by additive white Gaussian noises (AWGNs), the resulting model parameter estimates can be biased. For instance, when dealing with the Volterra model and if the system input is disturbed by noise, the standard least mean squares (LMS) algorithm or the off-line "Yule-Walker equation" lead to poor results, especially when the signal-to-noise ratio (SNR) is low. Noise compensated approaches like the ones we have recently proposed in [1] can be considered. However, the bias of the parameter vector estimate must be preliminary expressed. It usually depends on a matrix, whose structure varies according to the memory and the order of the model and whose entries depend on the additive noise statistics. Alternative solutions can be based on optimal filtering. In that case, both the noise-free input and output and the parameters must be estimated from the noisy observations. For this purpose, methods such as the extended Kalman fillter (EKF) or the extended H_∞ filter, which are based on a 1^{st}-order Taylor expansion of the state space equation around the last state vector estimate, or the 2^{nd}-order EKF (SO-EKF) based on a 2^{nd}-order Taylor expansion can be considered. Nevertheless, this requires the definition of the state space representation, the computation of Jacobian and Hessian matrices, etc. Expectation-maximization algorithm can be also used, but it strongly depends on the initial conditions. Whatever the above methods, if the practitioner changes the model, he has to study the situation again.

To avoid this drawback, alternative solutions can be considered and consist in generating the data from the model and in comparing them with the observations. This can be done by selecting the model parameters by means of a grid search ap-

proach to explore the solution space and to find the best candidate. However, this method has a high computational cost. To reduce this latter, one can apply probabilistics techniques such as evolutionary algorithms (EAs). Starting with a random population of parameter candidates taken in the search space Ω, new populations are iteratively created. These algorithms aim at exploring Ω and evolve towards the candidate that minimizes a beforehand defined fitness function. Among EAs, simulated annealing is due to metallurgy whereas ant colony algorithms and particle swarm optimization take their origins in ethology. The genetic algorithm (GA) and the differential evolution (DE) use mechanisms inspired by biological evolution, such as reproduction, crossover and mutation, with probability p_r, p_c and p_m respectively. It should be noted that those algorithms are often required for nonlinear estimation issues as in neural sciences [2] or image processing [3]. Although their computational costs can be higher than the above standard methods, these algorithms present several advantages. Firstly, they do not need an elaborate formalisation of the problem unlike optimal-filtering based methods. If the model changes, the user has "just" to modify the way the data are generated in the algorithm. Secondly, if the quantities such as the population cardinal, p_m, p_c, etc. are corretly tuned, this kind of approach avoids to get stuck at a local optimum.

Remark: for the last years, a great deal of interest has been paid to model structure selection. Thus, machine learning, such as genetic programming[1] (GP) or artificial neural networks (ANNs), aims at selecting the functions that can be used to map the input data into the output data. Both methods are of interest especially for simulation methods. Concerning GP, the model parameters can be jointly estimated with the model structure in the GP process. Otherwise, for each tree, the corresponding parameters can be estimated by using an orthogonal least squares algorithm (OLS) [4], an EA, etc. Nethertheless, this step is not an easy task with noisy observations. Takagi-Sugano neuro fuzzy networks (TS) could also be considered. As TS can be seen as a linear parameter varying system, LS techniques can be used to estimate the corresponding model parameters. With noisy observations, IV techniques have been proposed [5]. However, although this approach seems to be efficient, it requires a partitionning of the input and output data space by using on-line maximum distance based clustering such as the evolving clustering methods, etc. This results in a very sophisticated method that anybody cannot easily use. For this reason, we will focus on *a priori* model based methods or GP in this paper. In both cases, EAs are of interest for parameter estimation.

[1]GP consists in initially generating a population of computer programs that can be representated by trees. The trees define the way to combine the leafs, namely the data that can be constants and/or observation signals etc., by using nodes which correspond to functions (exponential, logarithm, etc.). For each tree, an evaluation is computed. Then, crossovers which consist in swapping the sub-trees between selected individuals make the population evolve. This process is iterated until the stop criterion is reached.

This paper deals with the identification of a nonlinear mechanical system with a single degree of freedom (SDOF) modelled by a nonlinear output error (NOE) model. This kind of system is chosen because it represents the simplest vibratory system. Then, the model parameters are estimated by using a stochastic optimization algorithm. Here, the output observations are disturbed by a zero-mean AWGN. For this purpose, we investigate several fitness functions when there is only a finite set of noisy observations available to see the noise impact on the model parameter estimation. This work is complementary to the study presented in [6] where the performances of GA and LS methods are compared when noise-free data are available.

The remainder of the paper is organized as follows: in section 2, we describe the system and present the stochastic optimization. In section 3, we study various fitness functions using noisy output and deduce rules for the estimation algorithm. In section 4, simulation results are given for two models and for various SNRs without loss of generality.

2. MODEL PARAMETER ESTIMATION FROM NOISY OBSERVATIONS OF A NARX SYSTEM

The system to be identified is a single-input single-output (SISO) NARX discrete-time system:

$$y(n) = f\left(y(n-1), \ldots, y(n-n_y), u(n-d), \ldots, u(n-n_u)\right) \quad (1)$$

where u is the exogenous input considered here as the system input signal, y the output signal of the system, d the system input delay, n_u and n_y the memory of the system output and the system input respectively and f a nonlinear function.

As the system output is disturbed by a zero-mean AWGN b with variance σ_b^2 (see Fig. 1), we suggest estimating $\hat{y}$ by using the previous estimates $\hat{y}$ instead of the noisy observations z. This representation is known as a NOE model:

$$\hat{y}(n) = f\left(\hat{y}(n-1), \ldots, \hat{y}(n-n_y), u(n-d), \ldots, u(n-n_u)\right) \quad (2)$$

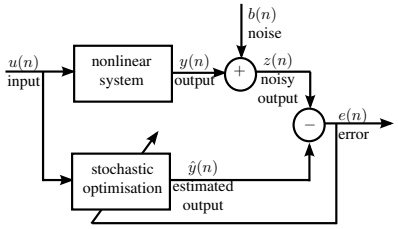

Fig. 1. Stochastic optimization with noisy measurements.

Let us consider a SDOF base excitation mechanical system, whose motion equation is the following in the continuous-time domain, see Fig. 2:

$$m\ddot{y}(t) + c\dot{x}(t) + kx(t) + \varphi\left(x(t)\right) = 0 \quad (3)$$

where m denotes the mass, k the stiffness, c the damping coefficient, y the mass position in the Galilean reference frame, $x = y - u$ the mass position with respect to the base position u and where φ represents a nonlinear polynomial function.

This function can be seen as the so-called stiffness nonlinearity often appearing in some mechanical systems. In this paper, $\varphi(x(t)) = px(t)^3$. Note that the total harmonic distorsion (THD) [7] is smaller than 5% for mechanical structure.

In the discrete-time domain, (3) becomes:

$$\begin{aligned} y(n) = & a_1 y(n-1) + a_2 y(n-2) + a_3 u(n) + a_4 u(n-1) \\ & + a_5 u(n-2) + a_6 (y(n-1) - u(n-1))^3 \end{aligned} \quad (4)$$

where $\{a_i\}_{i=1,\dots,6}$ are the parameters that can be expressed from the quadruplet (m, k, c, p) and the sampling period T_s:

$$\begin{cases} a_1 = \frac{\frac{2m}{T_s^2} - k}{\gamma};\ a_2 = \frac{-\frac{m}{T_s^2} + \frac{c}{2T_s}}{\gamma};\ a_3 = \frac{\frac{c}{2T_s}}{\gamma};\ a_4 = \frac{k}{\gamma} \\ a_5 = -a_3;\ a_6 = \frac{-p}{\gamma},\ \text{with } \gamma = \frac{m}{T_s^2} + \frac{c}{2T_s} \end{cases} \quad (5)$$

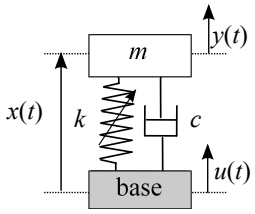

Fig. 2. Nonlinear SDOF with base excitation.

The physical parameters (m, k, c, p) characterize the nonlinearity with the ratio $\frac{p}{m}$ [8] and are related to the resonant frequency f_0 equal to $\frac{1}{2\pi}\sqrt{\frac{k}{m}}$ and the damping ratio ζ_0 equal to $\frac{1}{2}\frac{c}{m}\sqrt{\frac{m}{k}}$. However, there is no unicity of the quadruplet leading to the quantities of interest: f_0, ζ_0 and $\frac{p}{m}$. Therefore the vector $\underline{\theta}_t = \left[\frac{k}{m}, \frac{c}{m}, \frac{p}{m}\right]^T$ is rather considered.

Remark: when dealing with the Duffing oscillator [8], the system ouput only depends on the input signal in (4).

To estimate the model parameters from noisy measurements, a stochastic optimization based on any variant of a GA, a DE, etc. can be used. The N_g model parameters $\{\theta_i\}_{1 \leq i \leq N_g}$ to be estimated , also called genes, are stored in a $(N_g \times 1)$ parameter vector $\underline{\theta}$, which defines a candidate. Then, N_c candidates are randomly chosen in Ω to generate the initial population.

At that stage, the fitness fucntion choice is crucial, as seen in [9]. A discussion on the estimation protocol is given in the next section when dealing with noisy observations.

3. IMPROVING THE ESTIMATION PROTOCOL

In the following, $r_{vw}(\tau)$ denotes the cross-correlation function between v and w for lag τ and $\hat{r}_{vw}(\tau)$ is its unbiased estimate based on N samples:

$$\hat{r}_{vw}(\tau) = \frac{1}{N - |\tau|} \sum_{n=\tau+1}^{N} v(n) w(n-\tau) \quad (6)$$

Then, P_v represents the power of v:

$$P_v = \frac{1}{N} \sum_{n=1}^{N} v^2(n) = \hat{r}_{vv}(0) \quad (7)$$

3.1. Approach #1

Let us look at the following criterion in the noise-free case:

$$J_1 = E\left[(y(n) - \hat{y}(n))^2\right] = E\left[\tilde{y}(n)^2\right] \quad (8)$$

where $E[.]$ denotes the expectation. However, when b disturbs the system output, the fitness becomes:

$$J_{1,noisy} = E\left[(z(n) - \hat{y}(n))^2\right] = J_1 + \sigma_b^2 \quad (9)$$

Therefore, minimizing $J_{1,noisy}$ or J_1 should lead to the same solution. However, in practical case and with one realization of the noise-free data, the following fitness function is rather commonly used:

$$J_{1,N} = \sum_{n=1}^{N} (y(n) - \hat{y}(n))^2 = \sum_{n=1}^{N} \tilde{y}(n)^2 \quad (10)$$

After convergence of the stochastic optimization algorithm, $J_{1,N}$ is usually very small ($i.e. < 10^{-10}$). It has been confirmed by various simulation tests we carried out.

In the noisy case, the fitness function (10) becomes:

$$\begin{aligned} J_{1,N,noisy} &= \sum_{n=1}^{N} (z(n) - \hat{y}(n))^2 \\ &= J_{1,N} + \sum_{n=1}^{N} b(n)^2 + 2\sum_{n=1}^{N} \tilde{y}(n) b(n) \end{aligned} \quad (11)$$

Therefore, at any generation of the stochastic optimization and for a solution candidate leading to $\hat{y}(n)$, the difference $J_{1,N,noisy} - J_{1,N}$ is the sum of the two following terms:

- $\sum_{n=1}^{N} b(n)^2 = \sigma_b^2 \sum_{n=1}^{N} \left(\frac{b(n)}{\sigma_b}\right)^2 = N P_b$ is a random variable following a χ^2 law with N degrees of freedom. When N is high (e.g. higher than 100), it can be approximated by a Gaussian variable, the mean of which is $N\sigma_b^2$ and the variance of which is $2N\sigma_b^4$.
- $2\sum_{n=1}^{N} \tilde{y}(n) b(n) = 2N \hat{r}_{\tilde{y}b}(0)$. According to the Cauchy-Schwartz inequality, the absolute value of this quantity is smaller than $2N\sqrt{P_{\tilde{y}}}\sqrt{P_b}$.

Then, among the criteria stopping an EA, one could define a fixed number of generations g_{max} to be reached. However, it has the disadvantage of either generating useless populations or stopping the algorithm too early. As an alternative, one could check if the minimum fitness function value changes from one generation to another [11]. However, the algorithm can be trapped in a local minimum. Therefore, we first suggest introducing a threshold under which the minimum value of the fitness function must be. It should be noted that, when $4P_{\tilde{y}} < P_b$, the quantity $2\sum_{n=1}^{N} \tilde{y}(n) b(n)$ is negligible compared to $\sum_{n=1}^{N} b(n)^2$. Therefore, provided that the convergence of the stochastic algorithm is guaranteed, in most of the cases (e.g. 99%), the minimum value of $J_{1,N,noisy}$

should be in the interval $\hat{\sigma}_b^2 \left[N-3\sqrt{2N}; N+3\sqrt{2N}\right]$, where $\hat{\sigma}_b^2$ is the estimated noise variance[2]. Thus, under the threshold $T = \hat{\sigma}_b^2 \left(N + 3\sqrt{2N}\right)$, the algorithm could enter in a 'confidence' area where the minimum value of $J_{1,N,noisy}$ should be. Whatever the input and the system studied, T is the same.

Once this first convergence criterion is satisfied, we propose to check if the gene population variance is low, i.e. if the maximum variance of the N_g genes between the N_c individuals in the population is lower than 10^{-3} (according to our tests). If both criteria are satisfied, one can consider that the stochastic optimization algorithm has converged. Otherwise, the estimation must be restarted by changing N_c, increasing p_m, etc. The best chromosome that was obtained can be kept as an individual in the new initial population. See Fig. 3.

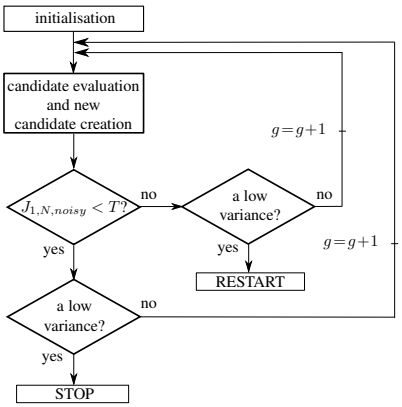

Fig. 3. Protocol for stochastic optimization.

Nevertheless, as the difference between $J_{1,N,noisy}$ and $J_{1,N}$ depends on $\hat{y}(n)$, we cannot always guarantee that $J_{1,N,noisy}$ and $J_{1,N}$ lead to the same solution. Therefore, in the next subsection, we propose fitness functions based on cross-correlation.

3.2. Approach #2

In this section, we suggest considering fitnesses in which the influence of the AWGN is weakened. Let us first consider:

$$J_2 = \sum_{\substack{\tau=-K \\ \tau\neq 0}}^{K} \left(r_{yy}(\tau) - r_{\hat{y}\hat{y}}(\tau)\right)^2 \quad (12)$$

with K as high as possible to take into account the spectral features of the signals. Generally, when dealing with the estimated unbiased cross-correlation function, K does not exceed $\frac{1}{3}N$ to have confidence on the estimate.

[2]Various methods can be used to estimate σ_b^2. 1/ the practitioner has the opportunity to record the system output when no input signal is imposed. 2/ one can study the eigenvalues of the autocorrelation matrix of the noisy output. The predominant values correspond to the signal subspace whereas the lower define the noise subspace and hence allow the noise variance to be estimated. 3/ the M.A.D estimator based on wavelet coefficients $c_{j,k}$ at the level j and the time k can be considered. In that case, with l the last level, one has [10]: $\hat{\sigma}_b = \frac{median_k(c_{l,k} - median_i(c_{l,i}))}{0.6745}$.

However, in the noisy case, it leads to:

$$\begin{aligned} J_{2,noisy} &= \sum_{\substack{\tau=-K \\ \tau\neq 0}}^{K} \left(r_{yy}(\tau) - r_{\hat{y}\hat{y}}(\tau)\right)^2 \\ &\quad + \left(r_{yy}(\tau) + \sigma_b^2 - r_{\hat{y}\hat{y}}(\tau)\right)^2 \\ &= J_2 + \sigma_b^4 + 2\sigma_b^2 \left(r_{yy}(\tau) - r_{\hat{y}\hat{y}}(\tau)\right) \end{aligned} \quad (13)$$

Due to $\sigma_b^4 + 2\sigma_b^2 \left(r_{yy}(\tau) - r_{\hat{y}\hat{y}}(\tau)\right)$ in (13), we cannot guarantee that J_2 and $J_{2,noisy}$ lead to the same minimum.

As a consequence, we suggest the following fitnesses:

$$J_{2,bis,noisy} = \sum_{\substack{\tau=-K \\ \tau\neq 0}}^{K} \left(r_{zz}(\tau) - r_{\hat{y}\hat{y}}(\tau)\right)^2 \quad (14)$$

$$J_{2,ter,noisy} = \sum_{\tau=-K}^{K} \left(r_{zz}(\tau) - \hat{\sigma}_b^2\delta(\tau) - r_{\hat{y}\hat{y}}(\tau)\right)^2 \quad (15)$$

Remark: If the lag $\tau = 0$ is not included, it is true that the signal power is not considered in the fitness.

However, when only N samples are available, the cross-correlation function is replaced by its estimate. Thus N should be as high as possible.

Note that alternative fitnesses could also be:

$$J_{3,bis,N,noisy} = \sum_{\substack{\tau=-K \\ \tau\neq 0}}^{K} \left(\hat{r}_{zz}(\tau) - \hat{r}_{z\hat{y}}(\tau)\right)^2 \quad (16)$$

$$J_{3,ter,N,noisy} = \sum_{\tau=-K}^{K} \left(\hat{r}_{zz}(\tau) - \hat{\sigma}_b^2\delta(\tau) - \hat{r}_{z\hat{y}}(\tau)\right)^2 \quad (17)$$

4. SIMULATION RESULTS

In this section, we compare the performances of standard methods, such as the OLS or the recursive LS (RLS), with EAs in order to estimate the parameters of a SDOF system. DEs and a GA variant[3] initially proposed by Chang [12] are tested.

The SDOF theoretical parameters $\underline{\theta}_t$ values are given in Table 1. The system input is a zero-mean white Gaussian noise of variance 0.56. Note that for nonlinear systems, the input signal should be persistent in order to excite all the system modes. Then, the sample number N is set to 5000 and the sampling frequency F_s is equal to 2 kHz. The genetic variables are the following: $N_c = 60$, $p_m = 0.1$, $p_c = 0.8$, $p_r = 0.1$, $s_{i=1,2,3} = 0.001$ and $\underline{\phi}_{i=1,2,3} \in [-0.01; 0.01]^3$.

[3]In that case, the $p_r \times N_c$ worst chromosomes are replaced by the $p_r \times N_c$ best ones to define the parent population. Then, three randomly selected chromosomes $(\theta_1, \theta_2, \theta_3)$, where θ_1 provides the best fitness function value, are used to define new candidates as follows: $\underline{\theta}_i \leftarrow \underline{\theta}_i + r_i\left(2\underline{\theta}_1 - \underline{\theta}_2 - \underline{\theta}_3\right) + s_i\underline{\phi}_i$, with $i \in (1,2,3)$, r_i a random value taken in $[0;1]$, s_i a positive constant and $\underline{\phi}_i \in R^{N_c}$ a $(N_g \times 1)$ random noise vector. This could be seen as a DE due to the mutations based on $r_i\left(2\underline{\theta}_1 - \underline{\theta}_2 - \underline{\theta}_3\right)$, but there is no one-to-one competition. Thus, it combines a GA and a DE.

The results have been obtained on 100 runs. Then, when using $J_{1,N,noisy}$, the stochastic optimisation follows the protocol described in section 3.1 and $4P_{\tilde{y}} < P_b$. For the approach #2, K is set to 1600 and 200 generations are made.

	$\frac{k}{m}$	$\frac{c}{m}$	$\frac{p}{m}$	f_0 (Hz)	ζ_0	THD %
model 1	3.948×10^5	6.283×10^1	4.343×10^3	100	0.05	2.01
model 2	8.883×10^5	3.770×10^1	-6.218×10^3	150	0.02	1.78
minimum	0	0	-1×10^5			
maximum	1×10^6	8×10^2	1×10^5			

Table 1. Theoretical parameters of the SDOF system and the minimum and maximum values of the search space Ω.

Let us define the following error criterion which is the relative 1-norm error: $e_1 = \frac{|\hat{\vartheta} - \vartheta_t|}{|\vartheta_t|}$, with $\hat{\vartheta}$ the estimated parameter to be compared with the theoretical parameter ϑ_t.

According to Table 2, the parameter estimation error increases with the SNR. Note that the errors obtained with the DEs are not given for the sake of space, but they are slightly the same as the ones obtained with the GA variant [12].

Concerning the approach #1: in Fig. 4, $J_{1,N,noisy}$ of the best candidate reaches a value under T at the convergence. Note that for the sake of space, only one figure is given. For the protocole described in section 3.1, to satisfy both convergence criteria, 150 generations are necessary with the GA variant whereas more than 200 generations are required with DEs. Increasing the variance threshold in the stochastic optimisation algorithm can decrease this generation number.

Concerning the approach #2: the errors on f_0 and ζ_0 are higher with the fitness $J_{2,N,noisy}$ than the corresponding fitnesses (14) and (15). We can draw the same conclusions with the fitness 3. Note that the difficulty is to select K.

Given the simulations we carried out with both approaches, the errors with $J_{1,N,noisy}$ are the lowest in most of cases. Then, with LS methods, the errors of f_0 and ζ_0 are around 100 times higher than when using any DEs or the GA variant. Increasing N does not change those tendancies.

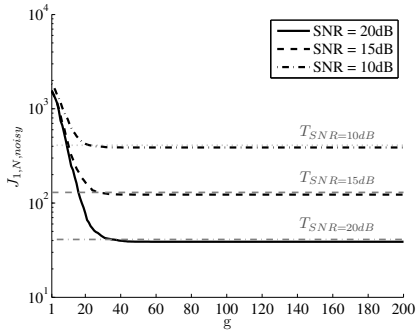

Fig. 4. Fitness function $J_{1,N,noisy}$ of the best candidate and the threshold T for different SNR on 100 runs.

5. CONCLUSIONS AND PERSPECTIVES

To use EAs with noisy data, we here suggest two approaches. The first one based on a threshold has the lower computational cost. Those approaches always provide more reliable results than standard LS methods. In future works, we will combine this kind of methods with GP.

Method	SNR (dB)	$\frac{k}{m}$	$\frac{c}{m}$	$\frac{p}{m}$	f_0	ζ_0
model 1						
$J_{1,N,noisy}$	20	$\mathbf{2.33 \times 10^{-4}}$	$\mathbf{2.66 \times 10^{-3}}$	1.00×10^{-2}	$\mathbf{1.16 \times 10^{-4}}$	$\mathbf{2.54 \times 10^{-3}}$
	15	$\mathbf{4.27 \times 10^{-4}}$	$\mathbf{4.74 \times 10^{-3}}$	1.82×10^{-2}	$\mathbf{2.13 \times 10^{-4}}$	$\mathbf{4.53 \times 10^{-3}}$
	10	$\mathbf{7.97 \times 10^{-4}}$	$\mathbf{8.45 \times 10^{-3}}$	3.34×10^{-2}	$\mathbf{3.98 \times 10^{-4}}$	$\mathbf{8.05 \times 10^{-3}}$
$J_{2,N,noisy}$	20	1.45×10^{-3}	$\mathbf{2.52 \times 10^{-3}}$	$\mathbf{6.65 \times 10^{-3}}$	7.26×10^{-4}	$\mathbf{3.25 \times 10^{-3}}$
	15	2.57×10^{-3}	$\mathbf{4.45 \times 10^{-3}}$	$\mathbf{1.14 \times 10^{-2}}$	1.28×10^{-3}	$\mathbf{5.75 \times 10^{-3}}$
	10	4.47×10^{-3}	$\mathbf{6.83 \times 10^{-3}}$	$\mathbf{1.68 \times 10^{-2}}$	2.24×10^{-3}	$\mathbf{9.09 \times 10^{-3}}$
$J_{2,bis,N,noisy}$	20	1.44×10^{-3}	$\mathbf{2.63 \times 10^{-3}}$	$\mathbf{6.53 \times 10^{-3}}$	7.21×10^{-4}	$\mathbf{3.35 \times 10^{-3}}$
	15	2.52×10^{-3}	$\mathbf{4.57 \times 10^{-3}}$	$\mathbf{1.04 \times 10^{-2}}$	1.26×10^{-3}	$\mathbf{5.84 \times 10^{-3}}$
	10	4.34×10^{-3}	$\mathbf{7.77 \times 10^{-3}}$	$\mathbf{1.49 \times 10^{-2}}$	2.17×10^{-3}	$\mathbf{9.97 \times 10^{-3}}$
$J_{2,ter,N,noisy}$	20	1.44×10^{-3}	$\mathbf{2.64 \times 10^{-3}}$	$\mathbf{6.52 \times 10^{-3}}$	7.20×10^{-4}	$\mathbf{3.36 \times 10^{-3}}$
	15	2.51×10^{-3}	$\mathbf{4.58 \times 10^{-3}}$	$\mathbf{1.04 \times 10^{-2}}$	1.26×10^{-3}	$\mathbf{5.85 \times 10^{-3}}$
	10	4.34×10^{-3}	$\mathbf{7.80 \times 10^{-3}}$	$\mathbf{1.49 \times 10^{-2}}$	2.17×10^{-3}	$\mathbf{9.99 \times 10^{-3}}$
OLS/RLS	20	1.82×10^{-1}	4.53×10^{-1}	5.33×10^{-1}	8.68×10^{-2}	2.94×10^{-1}
	15	5.39×10^{-1}	1.40	1.07	2.35×10^{-1}	4.68×10^{-1}
	10	1.34	1.29×10^{1}	2.17	5.13×10^{-1}	7.78×10^{-1}
model 2						
$J_{1,N,noisy}$	20	$\mathbf{2.48 \times 10^{-6}}$	$\mathbf{2.84 \times 10^{-3}}$	$\mathbf{2.99 \times 10^{-4}}$	$\mathbf{1.24 \times 10^{-6}}$	$\mathbf{2.84 \times 10^{-3}}$
	15	$\mathbf{5.91 \times 10^{-6}}$	$\mathbf{5.08 \times 10^{-3}}$	$\mathbf{5.68 \times 10^{-4}}$	$\mathbf{2.95 \times 10^{-6}}$	$\mathbf{5.08 \times 10^{-3}}$
	10	$\mathbf{1.49 \times 10^{-5}}$	$\mathbf{9.11 \times 10^{-3}}$	$\mathbf{1.12 \times 10^{-3}}$	$\mathbf{7.48 \times 10^{-6}}$	$\mathbf{9.11 \times 10^{-3}}$
$J_{2,N,noisy}$	20	5.34×10^{-4}	$\mathbf{1.70 \times 10^{-3}}$	1.02×10^{-2}	2.67×10^{-4}	$\mathbf{1.97 \times 10^{-3}}$
	15	9.67×10^{-4}	$\mathbf{3.42 \times 10^{-3}}$	1.80×10^{-2}	4.83×10^{-4}	$\mathbf{3.90 \times 10^{-3}}$
	10	1.78×10^{-3}	$\mathbf{6.84 \times 10^{-3}}$	3.22×10^{-2}	8.90×10^{-4}	$\mathbf{7.73 \times 10^{-3}}$
$J_{2,bis,N,noisy}$	20	5.20×10^{-4}	$\mathbf{1.59 \times 10^{-3}}$	1.00×10^{-2}	2.60×10^{-4}	$\mathbf{1.85 \times 10^{-3}}$
	15	9.28×10^{-4}	$\mathbf{3.08 \times 10^{-3}}$	1.77×10^{-2}	4.63×10^{-4}	$\mathbf{3.54 \times 10^{-3}}$
	10	1.66×10^{-3}	$\mathbf{5.82 \times 10^{-3}}$	3.14×10^{-2}	8.32×10^{-4}	$\mathbf{6.65 \times 10^{-3}}$
$J_{2,ter,N,noisy}$	20	5.15×10^{-4}	$\mathbf{1.87 \times 10^{-3}}$	9.76×10^{-3}	2.57×10^{-4}	$\mathbf{2.13 \times 10^{-3}}$
	15	9.24×10^{-4}	$\mathbf{3.38 \times 10^{-3}}$	1.74×10^{-2}	4.62×10^{-4}	$\mathbf{3.84 \times 10^{-3}}$
	10	1.66×10^{-3}	$\mathbf{6.15 \times 10^{-3}}$	3.11×10^{-2}	8.32×10^{-4}	$\mathbf{6.98 \times 10^{-3}}$
OLS/RLS	20	6.20×10^{-2}	8.46×10^{-1}	1.60×10^{-1}	3.04×10^{-2}	9.04×10^{-1}
	15	2.13×10^{-1}	1.77	3.47×10^{-1}	9.82×10^{-2}	1.44
	10	6.42×10^{-1}	5.30	7.59×10^{-1}	2.56×10^{-1}	2.54

Table 2. e_1 error for standard LS methods and GA [12].

6. REFERENCES

[1] Z. Sigrist, E. Grivel and B. Alcoverro, "Estimating second-order Volterra system parameters from noisy measurements based on an LMS variant or an errors-in-variables method," *Signal Processing*, vol. 92, no. 4, pp. 1010–1020, 2012.

[2] L. Buhry, A. Giremus, E. Grivel, S. Saighi and S. Renaud, "New Variant of Differential Evolution Algorithm: Application for Neuroscientists," *EUSIPCO 2009*, 2009.

[3] L. Trujillo, P. Legrand, G. Olague and C. Pérez, "Optimization of the Hölder Image Descriptor using a Genetic Algorithm" *GECCO 2010*, pp. 1147–1154, 2010.

[4] L. Santos Coelho and M. Wicthoff Pessoa, "Nonlinear Model Identification of an Experimental Ball-and-Tube System Using a Genetic Programming Approach," *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1434–1446, 2009.

[5] G. Serra and C. Bottura, "An IV-QR Algorthim for Neuro-Fuzzy Multivariable Online Identification," *IEEE Trans. on Fuzzy Systems*, vol. 15, no. 2, pp. 200–210, 2007.

[6] V. Duong and A. R. Stubberud, "System identification by Genetic Algorithm," *IEEE Aerospace Conference Proceedings*, vol. 5, pp. 2331–2337, 2002.

[7] L. Voishvillo, A. Terekhov, E. Czerwinski and S. Alewandrov, "Graphing, Interpretation, and Comparison of Results of Loudspeaker Nonlinear Distortion Measurements," *Journal of the Audio Engineering Society*, vol. 52, no. 4, pp. 332–357, 2004.

[8] K. Worden and G. Tomlinson, "Nonlinearity in Structural Dynamics: Detection, Identification, and Modelling," *IoP*, 2001.

[9] Q. Chen, K. Worden, P. Peng and A. Y. T. Leung, "Genetic Algorithm with an Improved Fitness Function for (N)ARX modeling," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 994–1007, 2007.

[10] D. Donoho and J. Johnstone, "Ideal Spatial Adaptation by Wavelet Shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[11] X.-P. Zeng, Y.-M. Li and J. Qin, "A Dynamic Chain-like Agent Genetic Algorithm for Global Numerical Optimization and Feature Selection," *Neurocomputing*, vol. 72, no. 4-6, pp. 1214–1228, 2009.

[12] W.-D. Chang, "An Improved Real-Coded Genetic Algorithm for Parameters Estimation of Nonlinear Systems," *Mechanical Systems and Signal Processing*, vol. 20, no. 1, pp. 236–246, 2006.