2013 European Control Conference (ECC)
July 17-19, 2013, Zürich, Switzerland.

An indirect model selection algorithm for nonlinear active noise control

Simone Morici, Emanuele Spiriti, and Luigi Piroddi

Abstract— **Model structure selection is a crucial task in applications where nonlinear black-box models are used, in order to reduce the model size and the associated computational effort. One such application is Active Noise Control (ANC), where nonlinear effects arise due *e.g.* to saturation and distortion of microphones and loudspeakers. Both parameter estimation and model selection are complex in the general nonlinear case if standard algorithms of the Least Mean Squares (LMS) type are used, due to the inherent difficulties in the gradient calculation when the secondary path is nonlinear. A model selection method is here proposed that employs a gradient-free parameter estimation algorithm to tackle the secondary path issue. A virtualization scheme is used to estimate the performance of the model subject to various different structural modifications, in order to select the most appropriate one to apply to the actual control filter. Some simulation examples are discussed to show the effectiveness of the algorithm.**

I. INTRODUCTION

A crucial component of nonlinear black-box identification methods is structure selection, [1]. Indeed, whatever parametric functional expansion is used to represent a nonlinear system, a large number of parameters must be initially assumed to endow the model with the necessary flexibility. For example, a particularly rich class is that of polynomial NARX/NARMAX (Nonlinear AutoRegressive [Moving Average] with eXogenous variable) models, [2], for which several theoretical developments and applications are documented in the literature. These models are in the form of a recursion where the current output is made dependent on previous input and output (and possibly, noise) samples through a polynomial expansion, configuring a linear regression where the regressors are monomials in the mentioned variables. Due to the curse of dimensionality, the size of a polynomial NARX/NARMAX model increases rapidly with the degree of the polynomial expansion and the system order, so that it is mandatory to select the most appropriate regressors to provide accurate and robust models. In the absence of model selection, an overparameterized model is typically obtained, with several unwanted consequences, such as overfitting, parameter fluctuation, poor model generalization capabilities, local minima, and sometimes instability (see, *e.g.*, [3], [4]).

One important application area where the need for adaptive nonlinear model selection methods arises is that of active noise control (ANC). Indeed, ANC is affected by several sources of nonlinearity (*e.g.*, distortion or saturation of microphones, amplifiers, loudspeakers [5], chaotic noise signals [6]), and several nonlinear ANC (NANC) methods have been recently introduced in the literature to tackle these issues, based on different nonlinear control filter model structures (see, *e.g.*, [7], [8], [9], [10], [11], [12]). In this context, model selection can significantly cut down the online computational load (by reducing the size of the estimated models), and improve the noise reduction performance (by estimating the parameters of more precisely matched model structures) [11].

Unfortunately, none of the adaptive model selection approaches available in the literature (see, *e.g.*, [13], [14], [15], [16], [17]) can be used in the NANC context. The reason stands in the particular structure of the identification problem underlying the ANC scheme, which can be characterized as *indirect*, in that the desired filter output signal is not directly accessible due the presence of the secondary path (the electroacoustic system accounting for the actuation and measurement chain of the control loop, as well as the involved acoustic channel) between the control filter and the output. Indeed, even assuming that the control filter is represented by a linear-in-the-parameters model and the secondary path model is time-invariant and known, it is not generally guaranteed that the input/output relation can be still represented as a linear regression in the control filter parameters, [11], a condition required by all the available recursive model selection approaches.

The presence of a nonlinear secondary path also complicates parameter estimation using Least Mean Squares (LMS) type algorithms, since the error gradient computation involves several recursive nonlinear filtering operations, raising both computational and stability issues [11]. This problem can be tackled using a gradient-free optimization method, employing only repeated measurements of the loss function for different perturbations of the parameter vector [18].

Following the same rationale, an innovative model selection method, suitable for indirect identification problems, is here proposed for use in the NANC context to select compact NARX models. The proposed method requires only repeated output error measurements resulting from different model structure modifications. An heuristic policy is used to progressively construct the model by adding and pruning terms. To avoid affecting the control performance during the selection, a model of the secondary path is used to evaluate the appropriateness of each structural modification attempted by estimating its effect on the output performance on a virtual channel. This allows to evaluate different structural modifications in parallel, compatibly with the computational resources available. The model selection task is intrinsically computationally intensive, since it amounts to solving a combinatorial problem (each regressor may be included in or eliminated from the model) where a reliable parameter estimation must be obtained for each tentative model struc-

S. Morici, E. Spiriti and L. Piroddi are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy. piroddi@elet.polimi.it

ture. However, it can be suitably downsampled to fit the computational capacity of the control hardware, while the main control loop provides noise attenuation based on a provisional control filter structure.

The paper is organized as follows. Section II briefly introduces the NARX model class. Section III illustrates the basic ANC scheme and provides its interpretation as an indirect identification problem, discussing the implications in the nonlinear case. The proposed method is presented in Section IV and tested on some simulation examples in Section V, before the final conclusions.

II. THE NARX MODEL CLASS

The deterministic NARX model [2] is an input-output recursive model where the current output is given by a nonlinear functional expansion of lagged inputs and outputs:

$$y(n) = f(y(n-1), ..., y(n-L), x(n), ..., x(n-L)), \quad (1)$$

where $f(\cdot)$ is a generic nonlinear function, which is often parameterized through a (truncated) polynomial expansion, so that (1) can be rewritten as a linear regression:

$$y(n) = \boldsymbol{w}(n)^T \boldsymbol{r}(\boldsymbol{z}(n)), \quad (2)$$

where $\boldsymbol{w}(n) = [w_1(n)\ w_2(n)\ \dots\ w_R(n)]^T$ is the weight vector, $\boldsymbol{r}(\boldsymbol{z}(n)) = [r_1(\boldsymbol{z}(n))\ r_2(\boldsymbol{z}(n))\ \dots\ r_R(\boldsymbol{z}(n))]^T$ the regressor vector, and $r_i(\boldsymbol{z}(n)) = \prod_{j=1}^{2L} z_j(n)^{l_{ij}}$, with $\boldsymbol{z}(n) = [y(n-1)\ \dots\ y(n-L)\ x(n)\ \dots\ x(n-L)]$ and l_{ij} being nonnegative integers such that $0 \leq \sum_{j=1}^{2L} l_{ij} \leq l$, l being the maximum degree of the expansion.

The size of the full NARX model (in terms of the number of regressors – and, consequently, parameters – R) increases rapidly with L and l (curse of dimensionality), but in practice satisfactory models can be obtained with a small fraction of the terms. Therefore, a candidate regressor set is typically assumed that includes all terms of a full NARX model for given L and l values, and the actual model is obtained by means of a selection process that establishes which candidate regressors to include.

Several model selection algorithms have been introduced in the literature both in the batch and the recursive case. Referring to the first class of methods, several algorithms rely on the use of Orthogonal Least Squares (OLS) to decouple the estimation of the parameters, such as the Forward Regression Orthogonal Estimator (FROE) [19] or the Fast Recursive Algorithm (FRA) [20]. An alternative technique based on an output error approach is developed in [4]. It is of obvious interest to develop adaptive model selection algorithms for on-line and possibly time-varying applications. Among these the GFEX (Givens rotation with Forward selection and EXponential windowing, [13]) and GFSL (Givens rotation with Forward selection and SLiding windowing, [14]) algorithms provide nice extensions of the OLS regression method to the adaptive case. Recursive regularization-based identification methods have also been recently proposed in the literature, [15], [16]. Some recent developments in this area are also documented in [17]. It is important to remark that all the mentioned adaptive methods rely on a linear regression structure of the identification problem.

III. ACTIVE NOISE CONTROL

A. The basic ANC scheme

Active Noise Control (ANC) (see, *e.g.*, [21], [22]) concerns the attenuation of acoustic noise using secondary acoustic sources and exploiting the principle of destructive interference. Adaptive feedforward schemes are more often used since sound transmission involves delays and time-varying dynamics, which make the feedback control design more difficult. The basic feedforward ANC scheme (see Fig. 1) requires the availability of a reference input signal $x(n)$, that must be highly correlated with the offending noise $d(n)$. The primary path P describes the model between the reference and noise signals, essentially accounting for the acoustic path involved in the noise transmission. The reference signal is processed by an adaptive digital filter C, whose output $y(n)$ drives a secondary acoustic source, with the aim to generate a secondary signal $y'(n)$ interfering negatively at the error microphone with the primary noise. The secondary path S describes the model between $y(n)$ and $y'(n)$. The control filter parameters (called weights) are adapted so as to minimize the (mean square) measured error $e(n) = d(n) - y'(n)$.

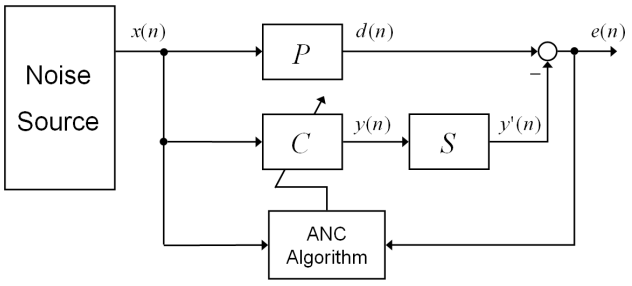

Fig. 1: Block diagram of an ANC system.

The weight update rule is the core of the feedforward ANC scheme, and is typically a simple gradient-based method of the Least Mean Squares (LMS) family, such as the Filtered-x Least Mean Squares (FXLMS) and the Filtered-u LMS (FULMS) algorithms [21].

B. ANC as an indirect identification problem

The basic ANC scheme of Fig. 1 configures a model identification problem, where the objective is to find a model of C such that the series between C and S emulates P, at least to the extent possible given the excitation characteristics of the reference signal $x(n)$. Such identification problem is *indirect*, in that the target output for model C is not accessible, and the only information available is not directly related to the output of C, but to a filtered version of it, through S.

A consequence of this is that, in the full nonlinear case (where both C and S are nonlinear), the overall system with input $x(n)$ and output $y'(n)$ (*i.e.*, the series of C and S) cannot be generally characterized as a linear regression in the parameters $\boldsymbol{w}(n)$, even if both systems are described by

linear-in-the-parameters models, unless very specific structural properties hold for the involved systems. Let C be described by model (1) and, similarly, S be formulated as:

$$\begin{aligned} y'(n) &= g(y'(n-1), ..., y'(n-M), y(n), ..., y(n-M)) \\ &= \boldsymbol{v}^T \boldsymbol{q}(\boldsymbol{z}'(n)), \end{aligned} \tag{3}$$

where $\boldsymbol{v} = [v_1 \ v_2 \ \dots \ v_Q]^T$ is the weight vector, $\boldsymbol{q}(n) = [q_1(n) \ q_2(n) \ \dots \ q_Q(n)]^T$ the regressor vector, and $q_i(n) = \prod_{j=1}^{2M} z'_j(n)^{m_{ij}}$, with $\boldsymbol{z}'(n) = [y'(n-1) \ \dots \ y'(n-L) \ y(n) \ \dots \ y(n-L)]$ and m_{ij} being nonnegative integers such that $0 \leq \sum_{j=1}^{2M} m_{ij} \leq m$.

One special condition where the series of C and S is still a linear regression in the unknown parameters of C occurs, *e.g.*, if both models depend linearly on $y(\cdot)$ [11]. Indeed, in that case, the control filter can be reformulated (in operatorial notation) as

$$A(z)y(n) = \boldsymbol{w}_x(n)^T \boldsymbol{r}_x(\boldsymbol{x}(n)), \tag{4}$$

where $A(z) = 1 - a_1 z^{-1} - \ldots - a_L z^{-L}$ is a polynomial in the unit delay operator z^{-1}, $\boldsymbol{x}(n) = [x(n) \ \dots \ x(n-L)]$ and $\boldsymbol{r}_x$ is the vector of (nonlinear) regressors depending only on $x(\cdot)$. The vector of unknown parameters is $\boldsymbol{w}(n) = [a_1 \ \dots \ a_L \ \boldsymbol{w}_x(n)^T]^T$. Similarly, the secondary path becomes:

$$y'(n) = \boldsymbol{v}_{y'}^T \boldsymbol{q}_{y'}(\boldsymbol{y}'(n)) + B(z)y(n), \tag{5}$$

where $B(z) = 1 - b_1 z^{-1} - \ldots - b_L z^{-L}$ is a polynomial in z^{-1}, $\boldsymbol{y}'(n) = [y'(n-1) \ \dots \ y'(n-L)]$ and $\boldsymbol{q}_{y'}$ is the vector of (nonlinear) regressors depending only on $y'(\cdot)$. In these assumptions, the overall model becomes:

$$A(z)\left(y'(n) - \boldsymbol{v}_{y'}^T \boldsymbol{q}_{y'}(\boldsymbol{y}'(n))\right) + \boldsymbol{w}_x(n)^T \left(B(z)\boldsymbol{r}_x(\boldsymbol{x}(n))\right), \tag{6}$$

which is a recursive equation with input $x(n)$ and output $y'(n)$ ($y(n)$ does not appear anymore) that is linear in the parameters $\boldsymbol{w}(n)$. Then, OLS can be used for parameter estimation, and any of the model selection methods based on linear regressions can be applied.

In the general case, more complex algorithms are required for both parameter estimation and model selection, due to the intrinsically indirect identification setting. The former task has been previously addressed with the Nonlinear Filtered-Gradient LMS (NFGLMS) algorithm [11]. This is a computationally intensive gradient-descent based approach, where the gradient of the cost function is constructed recursively employing the input-output gradient of the secondary path model. As for model structure selection, the available algorithms are not applicable since they are based on a direct identification setting and on a linear regression formulation [17].

IV. THE PROPOSED METHOD

A way to tackle the indirect identification setting, both for parameter estimation and model structure selection purposes, is to employ a gradient-free minimization algorithm, requiring only measurements of the cost function and based on empirical weight perturbation policies. Methods of this type have been regarded as particularly appealing for NANC, in that they can avoid altogether the knowledge of the secondary path model (see, *e.g.*, [23] and [24]). However, since they operate by iteratively applying perturbations on the parameters and assessing their effect on the output, their usage for this purpose provides less than satisfactory ANC performance. Indeed, to reliably establish the effect of a specific weight perturbation, one has to wait for the resulting transient to end, which may protract the adaptation process for an unacceptably long time. In addition, the noise attenuation performance is also affected, since worsening modifications may be tested at times.

On a different line, the secondary path model is here exploited to calculate virtually the effect of a model perturbation, providing an estimate of the control performance as a result of its application. This allows to decouple the parameter estimation and model selection phases from the actual control task, where the controller parameters and structure are updated only in directions for which a performance improvement has been estimated.

A. The virtualization scheme

With reference to the NANC setting where both the control filter and the secondary path are nonlinear systems, assume that the control filter is described by Eq. (1) and that a model for the secondary path is available in the form:

$$\begin{aligned} \hat{y}'(n) &= \hat{g}(\hat{y}'(n-1), ..., \hat{y}'(n-M), y(n), ..., y(n-M)) \\ &= \hat{\boldsymbol{v}}^T \boldsymbol{q}(\boldsymbol{z}'(n)), \end{aligned} \tag{7}$$

with obvious notation (compare with (3)). Then, at each step an estimate of the noise can be obtained as:

$$\hat{d}(n) = e(n) + \hat{y}'(n). \tag{8}$$

The obtained sequence $\hat{d}(\cdot)$ can be used to evaluate at time step n the effect of a given perturbation on the current filter parameters, by simulating the series of (a copy of) the control filter (with perturbed parameters) and (a copy of) the estimated secondary path for T steps starting from $n - T + 1$, T being a time horizon sufficiently long for the assessment of the perturbation effect, and at the same time not too long to significantly affect the overall computational load. In fact, in many simulations $T = 1$ provided more than satisfactory results. The explained virtualization scheme can be exploited to devise an ANC weight update rule based on the parallel testing of different parameter perturbations, allowing to estimate the local gradient direction employing solely the corresponding cost function evaluations.

More precisely, the Empirical Weight Update (EWU) algorithm [18] operates similarly to the Finite Difference Stochastic Approximation (FDSA) algorithm [25], perturbing (in both directions) one parameter coordinate at a time.

EWU iteration:

1) Apply (1) to obtain the control output $y(n)$.
2) Apply (7) to estimate the secondary path output $\hat{y}'(n)$. Then, measure $e(n)$ and estimate the noise sample with (8).
3) For $i = 1$ to R (number of weights) let $\boldsymbol{w}_{i+}(n) = \boldsymbol{w}(n) + \mu_i \boldsymbol{e}_i$ and calculate $J_{i+}(n) = eval(\boldsymbol{w}_{i+}(n), T, n)$.

4) Repeat step (3) with $\boldsymbol{w}_{i-}(n) = \boldsymbol{w}(n) - \mu_i \boldsymbol{e}_i$, calculating $J_{i-}(n) = eval(\boldsymbol{w}_{i-}(n), T, n)$, $i = 1, \ldots, R$.
5) For $i = 1$ to R, let $J_i^{min}(n) = min(J(n), J_{i+}(n), J_{i-}(n))$. Then, if $J(n) = J_i^{min}(n)$ $\delta w_i(n) = 0$ else if $J_{i+}(n) = J_i^{min}(n)$ $\delta w_i(n) = q_i mu_i$ else $\delta w_i(n) = -q_i mu_i$, where $q_i = \frac{J(n) - J_i^{min}(n)}{\sqrt{\sum_{j=1}^{R} (J(n) - J_j^{min}(n))^2}}$ is a scaling factor that encourages perturbations in the most promising directions.
6) Finally, let $\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \delta \boldsymbol{w}(n)$.

The EWU Algorithm employs several time the function $eval$, which evaluates the virtual value of the cost function resulting from a parameter perturbation.

$Function\ \bar{J} = eval(\bar{\boldsymbol{w}}, T, n)$:

1) For $k = n - T + 1$ to n calculate $\bar{y}(k) = \bar{\boldsymbol{w}}^T \boldsymbol{r}(\bar{\boldsymbol{z}}(k))$, where $\bar{\boldsymbol{z}}(k) = [\bar{y}(k-1)\ \ldots\ \bar{y}(k-L)\ x(k)\ \ldots\ x(k-L)]$, and $\bar{y}(k) = y(k)$ for $k < n - T + 1$.
2) For $k = n - T + 1$ to n calculate $\bar{y}'(k) = \hat{\boldsymbol{v}}^T \boldsymbol{q}(\bar{\boldsymbol{z}}'(k))$, where $\bar{\boldsymbol{z}}'(k) = [\bar{y}'(k-1)\ \ldots\ \bar{y}'(k-L)\ \bar{y}(k)\ \ldots\ \bar{y}(k-L)]$, and $\bar{y}'(k) = \hat{y}'(k)$ for $k < n - T + 1$.
3) Calculate $\bar{J} = \frac{1}{T} \sum_{k=n-T+1}^{n} (\hat{d}(k) - \bar{y}'(k))^2$.

To improve the performance, the step-size parameter $\boldsymbol{\mu}$ can also be adapted to enhance promising perturbation directions or to refine the parameter estimate near convergence (a detailed explanation of this auxiliary adaptation is omitted for brevity).

In terms of computational complexity the dominant cost of the EWU scheme is represented by the $2R$ evaluations of the error function, each of which requires an order of $2(R+Q)$ operations, for T samples. In the simplest case, in which T is set to 1, the cost is thus quadratic in the model size.

B. The heuristic iterative model selection policy

Model selection can be carried out exploiting in a slightly different way the virtualization scheme explained previously. Indeed, to test if the current model is improved by the addition of a new regressor, one could simply add the new regressor with initial parameter value equal to 0, retaining all the other parameters at their current values, regarding this as a regular parameter perturbation of the EWU algorithm. However, the optimal parametrization of the augmented model is not necessarily proximal to that of the original model (at least in the common coordinates), so that to establish the effect of the structural variation the parameters of the augmented model need to be adapted as well. A similar reasoning applies to pruning operations. For this reason, each structural modification is tested by applying the EWU algorithm (with $T = 1$, for simplicity) for a given time period τ on a virtual path, starting from null parametrization.

Following the heuristic scheme of [4], the model structure is progressively modified by applying elementary structural modifications (single regressor addition or pruning), as long as these modifications result in an improved model accuracy. At each iteration of the algorithm all possible terms are tested for inclusion in the model and the best one is added if it provides at least a pre-specified improvement. All tests are run in parallel, exploiting the virtualization mechanism. After the specified time period, the corresponding cost functions are compared and the best model is retained. A pruning stage is then executed, with the same rationale, this time testing the elimination of a regressor from the model. A regressor can be pruned if its elimination from the model costs less than the accuracy improvement brought by the added regressor.

In detail, the model selection algorithm operates as follows:

EWU-based model selection:

0) Let $\Phi = \{r_i(\boldsymbol{z}(n)), i = 1, \ldots, R\}$ be the complete set of regressors (*e.g.*, all the monomials of a polynomial expansion of degree l in the arguments $\boldsymbol{z}(n) = [y(n-1)\ \ldots\ y(n-L)\ x(n)\ \ldots\ x(n-L)]$ for a given order L). Let also $\Phi_{in} \subseteq \Phi$ and $\Phi_{out} = \Phi \setminus \Phi_{in}$ be the sets of regressors included in and excluded from the model. For example, one can initially set $\Phi_{in} = \emptyset$. Let also $\boldsymbol{r}_X$ denote the vector of regressors associated to the set X.
1) For each $r_i(\boldsymbol{z}(n)) \in \Phi_{out}$, $i = 1, \ldots, |\Phi_{out}|$, let $X_i = \Phi_{in} \cup \{r_i(\boldsymbol{z}(n))\}$. Apply the EWU algorithm with $T = T_{sel}$ for τ steps on a virtual path to estimate the parameters of the model $y(n) = \boldsymbol{w}(n)^T \boldsymbol{r}_{X_i}$, starting from $\boldsymbol{w}(n) = \boldsymbol{0}$. At the end of the adaptation process calculate $J_i(n) = eval(\boldsymbol{w}(n), T', n)$, where T' is a sufficiently long horizon to evaluate the model performance, as well as the corresponding $J(n)$ of the current model. If $min_i(J_i(n)) \leq J(n) - \delta J$, then $\Phi_{in} = X_i$, $\Phi_{out} = \Phi \setminus \{r_i(\boldsymbol{z}(n))\}$, where δJ is a predefined threshold, else the algorithm ends.
2) If $|\Phi_{in}| \geq 3$, for each $r_i(\boldsymbol{z}(n)) \in \Phi_{in}$, $i = 1, \ldots, |\Phi_{in}|$, let $X_i = \Phi_{in} \setminus \{r_i(\boldsymbol{z}(n))\}$ and proceed as in Step (1). If $min_i(J_i(n)) \leq J(n)$, then $\Phi_{in} = X_i$, $\Phi_{out} = \Phi \cup \{r_i(\boldsymbol{z}(n))\}$. Optional: Repeat Step (2) until no further regressor is pruned.

The added computational complexity (multiple instances of the EWU algorithm must be executed in parallel) can be compensated by suitably downsampling the model selection task, distributing its load on larger time horizons. Notice, however, that the first steps of the model selection procedure are quite fast, because the EWU is applied to (many) small-sized models. In any case, the main benefit of the model selection is to be evaluated in the long run, since a significant reduction in the model size may drastically cut the on-line computational effort (after the selection is completed). Moreover, the unpredictable effects of overparametrization are also avoided.

V. SIMULATION EXAMPLES

A. Analytical example

The first simulation example is designed to test the algorithm in an ideal condition where there exists a parametrization of the control filter such that the series of controller and secondary path coincides exactly with the primary path, in order to see if it can retrieve the exact model structure (and parametrization). The secondary path is given by:

$$y'(n) = 0.1y'(n-1)y(n-2) + 0.3y(n) + 0.2y'(n-1)y(n) + 0.2y'(n-1)y(n)^2$$

and the controller structure is designed as follows:

$$y(n) = ax(n)^2 + by(n-1)x(n-1) + cx(n)$$

The primary path is designed so that there exists an ideal model for the control filter that solves the identification problem. Precisely, the model of the primary path is obtained as the series of the controller, with parameters $a = 0.4$, $b = 0.2$, $c = 0.3$, and the secondary path.

The model selection process performs a step every 50 samples, using a window of 500 samples to evaluate the effect of a model structure modification ($\tau = 500$ and $T_{sel} = 1$). Accordingly, the process starts after 550 samples. New terms are included in the model if they provide a 15% improvement. By inspection of Fig. 2, it is apparent that the correct regressors are picked out early in the selection process and never pruned. Comparatively, other regressors are selected only temporarily and for very small times. Furthermore, when all three essential regressors have been included in the model their coefficients tend to the exact values, while spurious regressors are attributed negligible coefficient values (see Fig. 3).

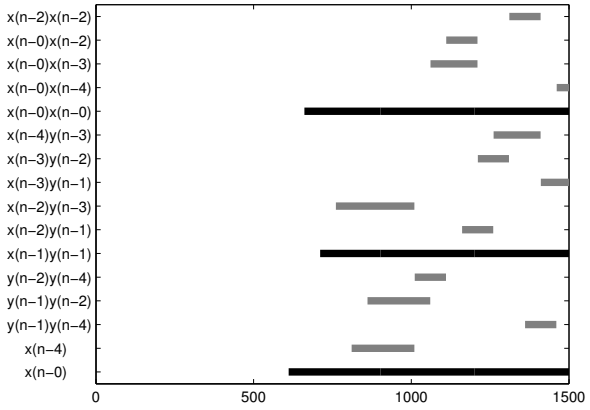

Fig. 2: Example 1: Selected regressors during the model selection process (correct regressors in black).

B. The NANC scheme with saturating microphones

A particularly hard NANC problem concerns a system affected by microphone saturation both at the reference and the error microphone, as represented in Figure 4, [9], [11]. Following [11], the primary and secondary paths are modeled by two FIR filters, taken from [26]:

$$\begin{aligned} d(n) &= 0.0179x(n) + 0.1005x(n-1) + 0.279x(n-2) \\ &\quad + 0.489x(n-3) + 0.586x(n-4) + 0.489x(n-5) \\ &\quad + 0.279x(n-6) + 0.1005x(n-7) + 0.0179x(n-8), \\ y'(n) &= 0.7756y(n) + 0.5171y(n-1) + 0.362y(n-2). \end{aligned}$$

The reference and error microphone are subject to saturation, in the form of a 50% clipping of the signal with respect to its maximum value [9]. The reference signal is the sum of three sine waves at the normalized frequencies of 0.02, 0.04, and 0.08, with sampling frequency set at 8000 Hz. A quadratic NARX model with $L = 16$ (594 terms) is

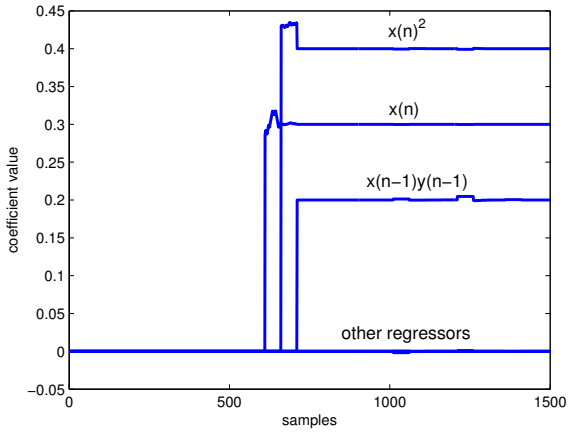

Fig. 3: Example 1: Values of the estimated coefficients during the model selection process.

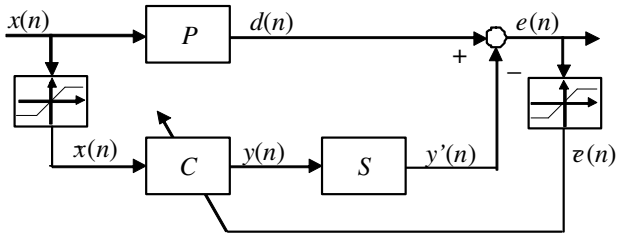

Fig. 4: Block diagram of a NANC system in the presence of saturated reference and error signals.

assumed as a starting point for the model selection phase. The model selection algorithm operates with a test-time equal to 500 and $T_{sel} = 1$. The algorithm performs a step every 50 samples. A core model with 3-4 terms is rapidly obtained that is capable of a significant noise reduction very early. The final selected model includes only 10 terms and is given by

$$\begin{aligned} y(n) &= 0.5564y(n-1) - 0.0447y(n-6) + 0.392x(n-1) \\ &\quad + 0.9098x(n-4) - 0.0314x(n-2)x(n-13) \\ &\quad - 0.0290x(n-1)y(n-7) + 0.0077x(n-2)y(n-9) \\ &\quad - 0.0217x(n-9)y(n-13) + 0.0669x(n-11)y(n-9) \\ &\quad - 0.0025y(n-1)y(n-14). \end{aligned}$$

The following figures 5 and 6 report the attenuation performance achieved by the selected model using the EWU algorithm for online weight update. For comparison purposes, several gradient-based algorithms, namely the NFGLMS, the Volterra FXLMS (VFXLMS, [7]) and the Bilinear FXLMS (BFXLMS, [9]) algorithms have also been tested on the full quadratic NARX (594 terms), the full 2^{nd} order Volterra expansion (170 terms) and the full bilinear models (306 terms), respectively (all with $L = 16$). Apparently, the compact size of the model, besides reducing of at least an order of magnitude the online computational effort, allows the update algorithm to converge faster and the steady-state noise reduction compares favorably with all the other

methods.

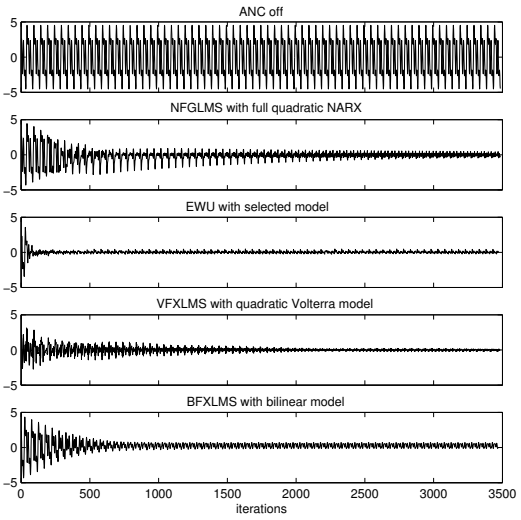

Fig. 5: Example 2: Residual output with different ANC methods.

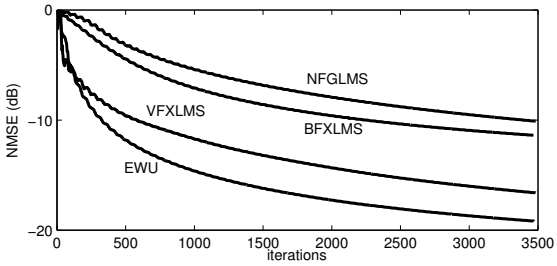

Fig. 6: Example 2: NMSE with different ANC methods.

The result is all the more valuable, considering that in all simulations the EWU algorithm employs the FIR model of the secondary path as its estimate, not taking into account the presence of the saturation.

VI. CONCLUSIONS

A novel model selection algorithm for nonlinear model of the NARX family has been developed, which is especially suited for indirect identification problems, as the NANC scheme. The method is based on the direct evaluation of the output error and does not require the explicit computation of the error gradient with respect to the weights. A model of the secondary path is used to reconstruct the original noise and to estimate virtually the performance associated to a parametric or structural modification of the controller model. The model selection method has been tested on a benchmark case where the system nonlinearity is represented by microphone saturation.

REFERENCES

[1] X. Hong, R. J. Mitchell, S. Chen, C. J. Harris, K. Li, and G. W. Irwin, "Model selection approaches for non-linear system identification: a review," *Int. J. of Systems Science*, vol. 39, no. 10, pp. 925–946, 2008.

[2] I. J. Leontaritis and S. A. Billings, "Input-output parametric models for non-linear systems - Part II: stochastic non-linear systems," *Int. J. of Control*, vol. 41, pp. 329–344, 1985.

[3] L. A. Aguirre and S. A. Billings, "Dynamical effects of over-parametrization in nonlinear models," *Physica D*, vol. 80, pp. 26–40, 1995.

[4] L. Piroddi and W. Spinelli, "An identification algorithm for polynomial NARX models based on simulation error minimization," *Int. J. of Control*, vol. 76, no. 17, pp. 1776–1871, 2003.

[5] S. D. Snyder and N. Tanaka, "Active control of vibration using a neural network," *IEEE Trans. Neural Netw.*, vol. 6, no. 4, pp. 819–828, 1995.

[6] T. Matsuura, T. Hiei, H. Itoh, and K. Torikoshi, "Active noise control by using prediction of time series data with a neural network," in *Proc. of the IEEE Int. Conf. on Syst., Man, Cybern.*, vol. 3, Vancouver, Canada, Oct. 22-25, 1995, pp. 2070–2075.

[7] L. Tan and J. Jiang, "Adaptive volterra filters for active control of nonlinear noise processes," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1667–1676, 2001.

[8] D. P. Das and G. Panda, "Active mitigation of nonlinear noise processes using a novel filtered-s lms algorithm," *IEEE Trans. Speech, Audio Process.*, vol. 12, no. 3, pp. 313–322, 2004.

[9] S. Kuo and H.-T. Wu, "Nonlinear adaptive bilinear filters for active noise control systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 3, pp. 1–8, 2005.

[10] D. Zhou and V. DeBrunner, "Efficient adaptive nonlinear filters for nonlinear active noise control," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 3, pp. 669–681, 2007.

[11] R. Napoli and L. Piroddi, "Nonlinear active noise control with NARX models," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 286–295, 2010.

[12] G. L. Sicuranza and A. Carini, "A generalized FLANN filter for nonlinear active noise control," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 8, pp. 2412–2417, Nov. 2011.

[13] W. Luo, S. A. Billings, and K. M. Tsang, "On-line structure detection and parameter estimation with exponential windowing for nonlinear systems," *European Journal of Control*, vol. 2, pp. 291–304, 1996.

[14] W. Luo and S. A. Billings, "Adaptive model selection and estimation for nonlinear systems using a sliding data window," *Signal Processing*, vol. 46, no. 2, pp. 179–202, 1995.

[15] P. Garrigues and L. El Ghaoui, "An homotopy algorithm for the lasso with online observations," in *Proc. NIPS*, 2008.

[16] D. Angelosante and G. B. Giannakis, "RLS-weighted lasso for adaptive estimation of sparse signals," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, April 2009.

[17] C. Cantelmo and L. Piroddi, "Adaptive model selection for polynomial NARX models," *IET Control Theory and Applications*, vol. 4, no. 12, pp. 2693–2706, 2010.

[18] S. Morici, E. Spiriti, and L. Piroddi, "An empirical weight update approach for nonlinear active noise control," in *Proc. American Control Conference*.

[19] M. Korenberg, S. A. Billings, Y. P. Liu, and P. J. McIlroy, "Orthogonal parameter estimation algorithm for non-linear stochastic systems," *Int. J. of Control*, vol. 48, pp. 193–210, 1988.

[20] K. Li, J.-X. Peng, and G. W. Irwin, "A fast nonlinear model identification method," *IEEE Trans. Autom. Control*, vol. 50, no. 8, pp. 1211–1216, 2005.

[21] S. M. Kuo and D. R. Morgan, *Active Noise Control Systems: Algorithms and DSP Implementations*. New York (NY), USA: Wiley-Interscience, 1996.

[22] S. J. Elliott, *Signal Processing for Active Control*. San Diego (CA), USA: Academic Press, 2001.

[23] Y.-L. Zhou, Q.-Z. Zhang, X.-D. Li, and W.-S. Gan, "On the use of an spsa-based model-free feedback controller in active noise control for periodic disturbances in a duct," *Journal of Sound and Vibration*, vol. 317, pp. 456–472, May 2008.

[24] N. K. Rout, D. P. Das, and G. Panda, "Particle swarm optimization based active noise control algorithm without secondary path identification," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 2, pp. 554–563, Feb. 2012.

[25] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Statist.*, vol. 23, no. 3, pp. 462–466, 1952.

[26] O. Tobias and R. Seara, "Performance comparison of the FXLMS, nonlinear FXLMS and leaky FXLMS algorithms in nonlinear active noise applications," in *Proc. of the 11^{th} IEEE European Signal Process. Conf.*, vol. 1, Toulouse, France, Sep. 3-6, 2002, pp. 155–158.