A Variable Step-Size for Sparse Nonlinear Adaptive Filters

Alberto Carini
University of Trieste
Trieste, Italy
acarini@units.it

Markus V. S. Lima
Federal University of Rio de Janeiro (UFRJ)
Rio de Janeiro, Brazil
markus.lima@dee.ufrj.br

Hamed Yazdanpanah
University of São Paulo (USP)
São Paulo, Brazil
hamed.yazdanpanah@smt.ufrj.br

Simone Orcioni
Università Politecnica delle Marche
Ancona, Italy
s.orcioni@staff.univpm.it

Stefania Cecchi
Università Politecnica delle Marche
Ancona, Italy
s.cecchi@univpm.it

***Abstract*—The paper deals with the identification of nonlinear systems with adaptive filters. In particular, adaptive filters for functional link polynomial (FLiP) filters, a broad class of linear-in-the-parameters (LIP) nonlinear filters, are considered. FLiP filters include many popular LIP filters, as the Volterra filters, the Wiener nonlinear filters, and many others. Given the large number of coefficients of these filters modeling real systems, especially for high orders, the solution is often very sparse. Thus, an adaptive filter exploiting sparsity is considered, the improved proportionate NLMS algorithm (IPNLMS), and an optimal step-size is obtained for the filter. The optimal step-size alters the characteristics of the IPNLMS algorithm and provides a novel gradient descent adaptive filter. Simulation results involving the identification of a real nonlinear device illustrate the achievable performance in comparison with competing similar approaches.**

***Index Terms*—Adaptive filters, linear-in-the-parameters nonlinear filters, functional link polynomial filters, optimal step-size.**

I. Introduction

The identification of nonlinear systems is a common problem in many areas, such as in image and speech/audio processing [1], [2], in devices using amplifiers and loudspeakers [3]–[5], in biological systems [6], [7], in wireless sensor networks [8], to name just a few examples.

Nonlinear adaptive filters are frequently used to identify unknown nonlinear systems. Very often the adaptive filter belongs to the class of *functional link polynomial* (FLiP) filters [9], [10]. This is a broad class of linear-in-the-parameters (LIP) nonlinear filters that encompasses many popular nonlinear filters like the well known Volterra filters, the Wiener nonlinear filters that derive from the truncation of the Wiener series [11], the Legendre nonlinear filters, the even mirror Fourier nonlinear filters, among many others [10]. The number of coefficients of FLiP filters increases exponentially with the filter order and geometrically with the memory length. Even though only small orders (typically 2 and sometimes 3) are normally used, already with moderate memory lengths the filter has a large number of coefficients, which limits the achievable convergence speed of the adaptive filter. However, in most real systems, many of these coefficients assume zero or almost zero value, since the nonlinear system input-output response is usually sparse [12].

Adaptive algorithms that exploit sparsity can improve the convergence performance, i.e., the convergence speed and/or the steady-state mean square error. Many algorithms exploiting sparsity have been proposed in the literature [13]. A first approach has been that of the proportionate normalized least mean square algorithm (PNLMS), where the coefficients are adapted on the basis of their magnitude [14]. The algorithm is capable of improving convergence in case of sparse filters, but worst performance is observed in case of non-sparse systems. Different approaches have been proposed to improve the performance of the PNLMS algorithm. One of the most effective solutions has been that of the Improved PNLMS (IPNLMS) algorithm [15], which in practice implements a linear combination of the adaptation rules of NLMS and PNLMS algorithms. An alternative to the proportional-update approach is the addition of a sparsity-promoting regularization to the cost function related to a given algorithm. In this second approach, the ℓ_1-norm of the coefficient vector [16]–[18], or even better, an ℓ_0-norm approximation of the coefficient vector [12], [13], [19], [20] are the most widely used regularizations.

A common problem in most adaptive filters, and in particular in nonlinear adaptive filters, is the choice of the step-sizes used in the adaptation. Very often this choice involves a tedious trial and error procedure that produce only suboptimal results. A successful trend in linear adaptive filters has been that of the variable step-sizes, where the step-size is estimated online sample by sample to guarantee optimal convergence, even in time-varying conditions. Among the various solutions, the optimal NLMS step-size has obtained great success in acoustic echo cancellation applications [21]. Attempts have been made to apply these variable step-sizes also to nonlinear filters. An effective variable step-size has been proposed by Kuech and Kellerman for LMS second order Volterra filters in [22]. In [22], the authors first develop a theoretical optimal step-size depending on expectations and non-observable quantities, and then obtain an implementable step-size under the hypothesis that the input signal is a zero-mean independent and identically distributed (iid) process and the coefficient errors are mutually orthogonal. Some strong but effective approximations are applied to the squared values of the coefficient errors. By properly choosing some of the approximating constants, the authors were also able to obtain a PNLMS algorithm. Other variable step-size approaches have been proposed in the literature. In [23], a variable step-size for Volterra filters is proposed, which requires some a priori knowledge of the system to be modeled and is not effective in case of abrupt changes in the unknown system. In [24], an error-dependent step-size control for second order Volterra filters is proposed. The algorithms in [23] and [24] are not able to exploit the sparsity in the nonlinear system and for this reason will not be further considered in the following.

In this paper, motivated by the desire of identifying FLiP filters, exploiting their sparsity and tuning automatically the step-sizes, the IPNLMS algorithm is considered in a nonlinear scenario and an optimal step-size is developed without resorting to any iid assumption on the input signal. It is shown in the paper that the resulting optimal step-size alters the characteristics (the normalization and the PNLMS weight) of the original IPNLMS algorithm, obtaining an original self-tuned adaptive filter. Experimental results considering the identification of a real device with stochastic signals illustrate the achievable performance in comparison with competing approaches.

The rest of the paper is organized as follows: Section II reviews FLiP filters. Section III introduces the optimal step-size for IPNLMS algorithm. Section IV provides some experimental results. Eventually, conclusion is reported in Section V.

The following notation is used throughout the paper. $\mathbb{R}_1$ is the unit interval $[-1,+1]$, $\mathbf{I}$ is an identity matrix of appropriate dimensions, $\mathbf{u}$ is a vector of all ones of appropriate dimensions, diag$[\mathbf{A},\mathbf{B},\mathbf{C}]$ is a block diagonal matrix with matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ on its main diagonal, diag$[\mathbf{v}]$ is a diagonal matrix with the elements of vector $\mathbf{v}$ on the main diagonal, $|\mathbf{v}|$ is the vector formed by the absolute value of the elements of $\mathbf{v}$, $\|\mathbf{v}\|_1$ is the ℓ_1-norm of $\mathbf{v}$ (the sum of the coefficients' absolute value).

II. FLiP filters

FLiP filters are a broad class of LIP nonlinear filters. Their basis functions are formed starting from an ordered set of univariate functions satisfying all requirements of the Stone-Weierstass theorem [25] on $\mathbb{R}_1$,

$$\{g_0[\xi], g_1[\xi], g_2[\xi], ...\}\,, \tag{1}$$

where $g_0[\xi]$ is a function of order 0, usually the constant 1, $g_{2i+1}[\xi]$ for any $i \in \mathbb{N}$ is an odd function of order $2i+1$, and $g_{2i}[\xi]$ for any $i \in \mathbb{N}$ is an even function of order $2i$. By writing the functions in (1) for $\xi \in \{x(n), x(n-1), \dots, x(n-N+1)\}$ and multiplying the terms of different variable in all possible manners, taking care of avoiding repetitions, a set of FLiP basis functions is obtained. Since the basis functions so developed form an algebra that satisfies all requirements of the Stone-Weierstass theorem, the linear combination of these basis functions can arbitrarily well approximate any discrete-time, time invariant, finite memory, causal, continuous nonlinear system

$$y(n) = f[x(n), x(n-1), \dots, x(n-N+1)], \tag{2}$$

where $f : \mathbb{R}_1^N \to \mathbb{R}$ is a continuous function [9].

By definition, the order of a FLiP basis function is the sum of the orders of its factors $g_i[\xi]$. The basis function diagonal number is the maximum time difference between the involved input samples. A FLiP filter of order K, memory N, diagonal number D is given by the linear combination of all FLiP basis functions having order, memory and diagonal number up to K, N, D, respectively.

Any choice of the univariate functions in (1) takes to a different class of FLiP filters. In the experimental results, the popular class of Volterra filters, where $g_i(\xi) = \xi^i$, will be considered, but the theory that follows is applicable to any FLiP filter.

FLiP filters are particularly appealing since they can be implemented in the form of a filter bank:

$$y(n) = \sum_{p=0}^{R-1} \sum_{m=0}^{N_p-1} h_p(m) f_p(n-m), \tag{3}$$

where $f_p(n)$ are the zero lag basis functions, i.e., $f_0(n) = 1$, $f_1(n) = g_1[x(n)]$, $f_2(n) = g_2[x(n)]$, $f_3(n) = g_1[x(n)]g_1[x(n-1)]$, ..., $f_{2+D}(n) = g_1[x(n)]g_1[x(n-D)]$, $f_{3+D}(n) = g_3[x(n)]$, and so on, N_p is the memory length for the basis function $f_p(n)$, which is N minus the diagonal number of $f_p(n)$; R is the total number of zero lag basis functions. As for any LIP filter, they can also be expressed in vector form as the product of a coefficient vector $\mathbf{h}(n)$ and an input data vector $\mathbf{x}(n)$ collecting all basis functions:

$$y(n) = \mathbf{h}^T(n)\mathbf{x}(n). \tag{4}$$

In what follows, the constant basis function $f_0(n)$ will be neglected and the vectors $\mathbf{h}(n)$ and $\mathbf{x}(n)$ will be decomposed as follows,

$$\mathbf{h}(n) = [\mathbf{h}_1^T(n), \mathbf{h}_2^T(n), \dots, \mathbf{h}_K^T(n)]^T, \tag{5}$$

$$\mathbf{x}(n) = [\mathbf{x}_1^T(n), \mathbf{x}_2^T(n), \dots, \mathbf{x}_K^T(n)]^T, \tag{6}$$

where $\mathbf{h}_i(n)$ is the collection of coefficients of order i, and $\mathbf{x}_i(n)$ is the collection of the corresponding basis functions, with $i = 1, \dots, K$.

III. IPNLMS algorithm and its optimal step-size

Let us assume we want to identify a LIP nonlinear filter with input signal $x(n)$ and output signal $d(n)$ using a FLiP adaptive filter of order K, memory N, diagonal number D. The error signal is

$$e(n) = d(n) - \mathbf{h}^T(n)\mathbf{x}(n), \tag{7}$$

and the unknown system is assumed to be representable with that FLiP filter, i.e.,

$$d(n) = \mathbf{h}_o^T(n)\mathbf{x}(n) + \nu(n), \tag{8}$$

where $\mathbf{h}_o(n) = [\mathbf{h}_{1,o}^T(n), \mathbf{h}_{2,o}^T(n), \dots, \mathbf{h}_{K,o}^T(n)]^T$ is the unknown system coefficient vector and $\nu(n)$ is an additive noise, uncorrelated with all other signals.

The FLiP filter is adapted according to the following rule

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \overline{\boldsymbol{\mu}}(n)\frac{\mathbf{K}(n)\mathbf{x}(n)}{\mathbf{x}^T(n)\mathbf{K}(n)\mathbf{x}(n)+\epsilon}e(n), \qquad (9)$$

where $\overline{\boldsymbol{\mu}}(n)$ is a diagonal matrix comprised of the individual variable step-sizes, $\mathbf{K}(n)$ is a diagonal weight matrix and ϵ is a small constant used to avoid divisions by zero.

For $\overline{\boldsymbol{\mu}}(n) = \mathbf{I}$ and $\epsilon = 0$, the adaptation rule in (9) is the exact solution of the following optimization problem [26]:

$$\begin{aligned} &\text{Minimize } [\mathbf{h}(n+1) - \mathbf{h}(n)]^T\mathbf{K}^{-1}(n)[\mathbf{h}(n+1) - \mathbf{h}(n)] \\ &\qquad \text{subject to } d(n) - \mathbf{h}^T(n+1)\mathbf{x}(n) = 0. \end{aligned}$$

In what follows, the different nonlinear kernels are separately weighted considering $\mathbf{K}(n) = \text{diag}[\mathbf{K}_1(n), \dots \mathbf{K}_K(n)]$, with $\mathbf{K}_i(n)$ a diagonal matrix having size compatible with $\mathbf{h}_i(n)$. In the IPNLMS algorithm,

$$\mathbf{K}_i(n) = \text{diag}\left[\frac{1-\alpha}{2L_i}\mathbf{u} + \frac{(1+\alpha)\mathbf{h}_i(n)}{2||\mathbf{h}_i(n)||_1+\delta}\right], \qquad (10)$$

with L_i the number of elements of $\mathbf{h}_i(n)$, $-1 \le \alpha \le +1$, and δ is a small constant to avoid divisions by zero. In (10), the different kernels are separately weighted as proposed in [27]. For $\alpha = -1$, (9) reduces to the classical NLMS algorithm. For $\alpha = 1$, (9) gives the proportionate NLMS (PNLMS), neglecting the additional constants used by PNLMS to avoid computational problems. In [15], to obtain a good convergence speed both in case of sparse and non-sparse systems, the value of $\alpha = -0.5$ or $\alpha = 0$ is suggested.

An optimal step-size matrix (9) is now developed. Let us consider $\mathbf{m}(n) = \mathbf{h}_o(n) - \mathbf{h}(n)$ and the weighted norm

$$||\mathbf{m}(n)||^2_{\mathbf{K}^{-1}(n)} = \mathbf{m}^T(n)\mathbf{K}^{-1}(n)\mathbf{m}(n). \qquad (11)$$

At each iteration we choose the step-size matrix $\boldsymbol{\mu}(n)$ such that

$$||\mathbf{m}(n+1)||^2_{\mathbf{K}^{-1}(n)} - ||\mathbf{m}(n)||^2_{\mathbf{K}^{-1}(n)} < 0. \qquad (12)$$

Considering that

$$\mathbf{m}(n+1) = \mathbf{m}_i(n) - \overline{\boldsymbol{\mu}}(n)\mathbf{K}(n)\mathbf{x}(n)D^{-1}(n)e(n), \qquad (13)$$

with $D(n) = \mathbf{x}^T(n)\mathbf{K}(n)\mathbf{x}(n) + \epsilon$, and replacing (13) in (12) we have

$$\begin{aligned} &-2e(n)D^{-1}(n)\mathbf{m}^T(n)\overline{\boldsymbol{\mu}}(n)\mathbf{x}(n) + \\ &\quad +e^2(n)D^{-2}(n)\mathbf{x}^T(n)\overline{\boldsymbol{\mu}}^2(n)\mathbf{K}(n)\mathbf{x}(n) < 0. \end{aligned} \qquad (14)$$

In what follows, we consider $h_{i,j}$ the j-th element of $\mathbf{h}_i(n)$, $m_{i,j}(n)$, $x_{i,j}(n)$, $k_{i,j}(n)$ and $\overline{\mu}_{i,j}$ the corresponding error coefficient, basis function, weight in $\mathbf{K}(n)$, and step-size, respectively.

Equation (14) can be equivalently written as

$$\begin{aligned} &\text{sign[e(n)]}\Big[-2\mathbf{m}^T(n)\overline{\boldsymbol{\mu}}(n)\mathbf{x}(n) + \\ &\quad +e(n)D^{-1}(n)\mathbf{x}^T(n)\overline{\boldsymbol{\mu}}^2(n)\mathbf{K}(n)\mathbf{x}(n)\Big] < 0. \end{aligned} \qquad (15)$$

Considering expectation and setting to zero the derivative of the left-hand side of (15) with respect to each $\overline{\mu}_{i,j}$, the following optimal step-size is obtained

$$\overline{\mu}_{i,j} = \frac{E\left[m_{i,j}(n)x_{i,j}(n)\text{sign}[e(n)]\right]}{E\left[x^2_{i,j}(n)k_{i,j}(n)e(n)\text{sign}[e(n)]D^{-1}(n)\right]}. \qquad (16)$$

In (16), the denominator can be easily estimated with a time average, while the numerator depends on the unknown quantity $m_{i,j}(n)$. The sign and the magnitude of $m_{i,j}(n)$ are separately approximated in the following. As for the sign of $m_{i,j}(n)$, it is reasonable to assume $\text{sign}[m_{i,j}(n)] = \text{sign}[h_{i,j,o}(n) - h_{i,j}(n)(n)] \simeq \text{sign}[h_{i,j}(n+1) - h_{i,j}(n)] = \text{sign}[x_{i,j}(n)e(n)]$. This is equivalent to approximate $m_{i,j}(n)x_{i,j}(n)\text{sign}[e(n)]$ with $|m_{i,j}(n)||x_{i,j}(n)|$. As for the magnitude of $m_{i,j}(n)$ it is reasonable to assume it depends on the average magnitude of the kernel i coefficients and on the magnitude of the $h_{i,j}(n)$, since the largest coefficients are always affected by the largest errors. Thus, $|m_{i,j}(n)|$ is here approximated with

$$|m_{i,j}(n)| \simeq \gamma\left[||\mathbf{h}_i||_1/L_i + \beta|h_{i,j}(n)|\right], \qquad (17)$$

where γ is an unknown constant, whose knowledge it will be shown to be immaterial, and β is a constant that gives larger or smaller weight to the coefficient $h_{i,j}(n)$ magnitude. It should be noted that in [22] a similar approximation was imposed on the squared magnitude of $m_{i,j}(n)$. The approximation appears better fitted to the magnitude $|m_{i,j}(n)|$. Since $e(n) = \sum_{s,t} m_{s,t}(n)x_{s,t}(n) + \nu(n)$ with $\nu(n)$ uncorrelated with the other signals, $\mu_{i,j}$ in (16) can be written as

$$\overline{\mu}_{i,j} = \frac{E\left[|m_{i,j}(n)||x_{i,j}(n)|\right]}{E\left[\left(\sum_{s,t}|m_{s,t}(n)||x_{s,t}(n)|\right)x^2_{i,j}(n)k_{i,j}(n)D^{-1}(n)\right]}. \qquad (18)$$

Inserting (17) in (18), the unknown constant γ simplifies and the expectations can be computed with time averages. The optimal step-size deriving from (18) has general validity. It can be simplified considering that $D(n)$ and $k_{i,j}(n)$ are slowly varying and can be taken out of expectation. In this case, simplifying $k_{i,j}(n)$, and $D(n)$ in (9) and (18), the following adaptation rule is obtained:

$$h_{i,j}(n+1) = h_{i,j}(n) + \mu_{i,j}(n)x_{i,j}(n)e(n), \qquad (19)$$

with

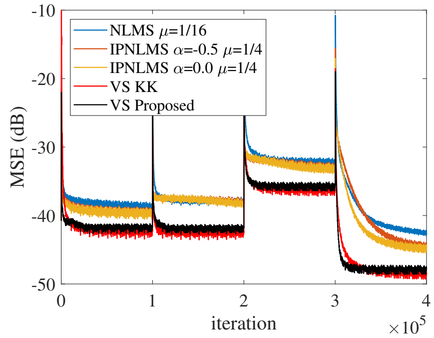
$$\mu_{i,j} = \frac{E\left[|m_{i,j}(n)||x_{i,j}(n)|\right]}{E\left[\left(\sum_{s,t}|m_{s,t}(n)||x_{s,t}(n)|\right)x^2_{i,j}(n)\right]}, \qquad (20)$$

with all expectations computed by time averages and $|m_{i,j}(n)|$ given by (17).

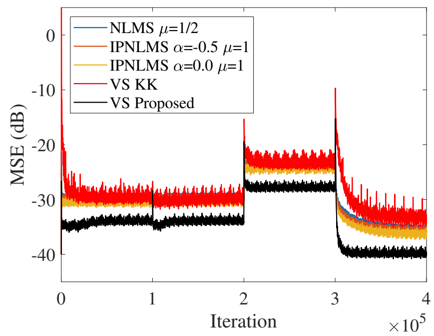
Starting from an IPNLMS algorithm, the optimal step-size has eliminated both the IPNLMS weight matrix $\mathbf{K}(n)$ and the normalization term $D(n)$. The normalization in reality is performed by (20), while the improved proportionate part promoting sparsity in the solution is performed by (17).

IV. Experimental results

We consider some experiments involving the identification of a real device, a Behringer Mic 100 Vacuum Tube Preamplifier, using a Volterra filter. The device has a potentiometer that allows us to control the distortion level on the output signal. Two stochastic signals with the same power and sampling frequency 16 kHz were generated: 1) a signal with eigenvalue spread about 20 obtained by filtering a white noise with an IIR filter having the following poles and zeros,



(a)



(b)

Fig. 1. Learning curves without additional noise: (a) stochastic input 1), (b) stochastic input 2).

$$\text{poles} = [0.5e^{i\cdot\pi/4}, 0.5e^{-i\cdot\pi/4}, 0.55e^{3i\cdot pi/4}, 0.55e^{-3i\cdot pi/4}];$$
$$\text{zeros} = [0.95i, -0.95i];$$

2) a white noise filtered with a first order all-poles filter with single pole at 0.95. The input signals at 16 kHz were upsampled in Matlab at 48 kHz to allow playback and recording on the Focusrite Scarlett 2i2 audio interface and the recorded signals were downsampled back at 16 kHz. Thus, in the experiment a chain composed by the upsampler, the Scarlett digital to analog converter (DAC), the MIC100 preamplifier, the Scarlett analog to digital (ADC), and the downsampler, is identified. At the maximum used amplitude on a sinusoidal signal at 1 kHz having the same power of the stochastic inputs the second, third, and total harmonic distortion are 3.5%, 0.6%, and 4.3%, respectively, and the signal to noise ratio (SNR) is larger than 60 dB. The system was identified with a Volterra filter of order 3, memory 30, and with diagonal number 7 for the second order kernel and 4 for the third order kernel, and 652 coefficients in total. To give an indication of the system sparsity, the filter identified with the proposed approach on stochastic input 1) after 10^6 samples has 160 coefficients with magnitude greater than 1/50 of the maximum coefficient. In order to simulate some abrupt changes in the unknown system to be modelled, at iteration L, the input signal is attenuated by a factor 4 and the output signal is time shifted by two samples; at iteration $2L$, the input signal is attenuated by a factor 2 (amplified by a factor 2 with respect to the previous segment), the output signal is amplified by a factor 2 and the time shift is removed; at iteration $3L$, the input signal resumes the original amplitude, the output signal is attenuated by a factor 2 (4 with respect to the previous segment) and time shifted by two samples. Both $L = 100,000$ and $L = 10,000$ are considered in the following. Different noise conditions are simulated by adding to the output a white Gaussian noise, as detailed in the following. The adaptive filter equipped with the proposed variable step-size is compared with NLMS, IPNLMS Volterra filters with $\alpha = -0.5$, and $\alpha = 0.0$, having constant step-size, and with the variable step-size of [22], implemented as detailed in [28]. It should be noted that the variable step-size of [22] was developed for a second order Volterra filter in an acoustic echo scenario, while here it is used for the identification of a third order Volterra filter using different stochastic signals. In the proposed approach, $\beta = 1$ and all expectations were estimated with exponential time averages with forgetting factor 0.8. The same forgetting factor was considered for the method of [22].

Figure 1 compares the performance of the proposed approach, with the algorithm of [22], the NLMS and IPNLMS algorithm for $L = 100,000$ without adding any additional noise at the output and for the two kinds of stochastic input considered. The learning curves are ensemble averages obtained over 200 runs of the algorithms, filtered with a box-filter of length 100. The curves of NLMS and IPNLMS are those that overall gave the best convergence performance, with the corresponding step-size detailed in the legend. With the first stochastic input, both variable step-sizes provide similar convergence results, which are far superior than the constant step-size algorithms. With the second input, which is much more correlated, the algorithm of [22] provides poor convergence and its learning curve overlaps with that of the constant step-size algorithms, while the proposed approach exhibits superior results.

Figure 2 shows a more detailed comparison of the proposed method with that of [22] for different SNRs considering $L = 10,000$. It can be appreciated that, for high SNR with both inputs the proposed method gives better performance in terms of convergence speed and similar or better steady-state MSE. When the SNR decreases, the convergence speed and the steady-state performance become comparable in both methods, with the method of [22] that provides slightly better steady-state MSE. In reality, the improved performance of [22] is due to a step-size factor 0.5 introduced to take into account the effect of noise as suggested in [28]. By reducing the proposed step-size by a factor 0.5, both algorithms provide the same steady-state MSE for high SNR.

V. Conclusion

The paper has derived an optimal variable step-size for the IPNLMS algorithm applied to FLiP filters. The variable step-size has been obtained without resorting to iid hypothesis on the input signal, or to the orthogonality of the coefficients' errors. It has been shown that the variable step-size has drastically altered the characteristics of the resulting adaptive filter, eliminating both the normalization and the coefficient weighing performed by IPNLMS. The performance of the resulting algorithm equipped with the variable step-size has been compared with NLMS, IPNLMS, and with the variable-

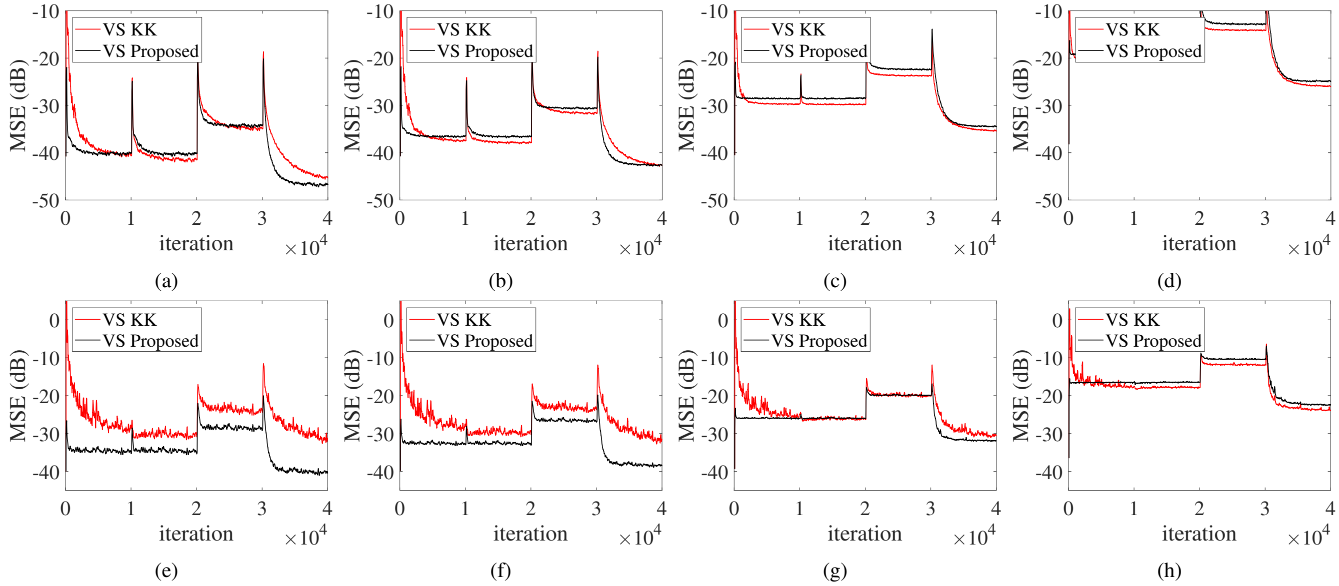


Fig. 2. Learning curves with different noise conditions: (a) and (e) 30 dB SNR, (b) and (f) 20 dB SNR, (c) and (g) 10 dB SNR, (d) and (e) 0 dB SNR, for (a)-(d) stochastic input 1), (e)-(h) stochastic input 2).

step size of [22], considering different inputs and noise conditions. The good performance of the proposed variable step-size makes it a good candidate for nonlinear system identification especially in case of highly correlated signal.

References

[1] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*. New York: Wiley, 2000.

[2] N. V. George and G. Panda, "Advances in active noise control: A survey, with emphasis on recent nonlinear techniques," *Signal Processing*, vol. 93, no. 2, pp. 363–377, Feb. 2013.

[3] M. Zeller and W. Kellermann, "Fast and robust adaptation of DFT-domain Volterra filters in diagonal coordinates using iterated coefficient updates," *IEEE Trans. Signal Processing*, vol. 58, no. 3, pp. 1589–1604, Mar. 2010.

[4] C. Shi and Y. Kajikawa, "Identification of the parametric array loudspeaker with a Volterra filter using the sparse NLMS algorithm," in *Proc. of ICASSP 2015*, Brisbane, Australia, Apr. 2015, pp. 3372–3376.

[5] S. Orcioni, A. Terenzi, S. Cecchi, F. Piazza, and A. Carini, "Identification of Volterra models of tube audio devices using multiple-variance method," *J. Audio Eng. Soc.*, vol. 66, no. 10, pp. 823–838, Oct. 2018.

[6] T. Berger, D. Song, R. Chan, and V. Marmarelis, "The neurobiological basis of cognition: Identification by multi-input, multioutput nonlinear dynamic modeling," *Proceedings of the IEEE*, vol. 98, no. 3, pp. 356–374, Mar. 2010.

[7] V. Marmarelis, *Nonlinear Dynamic Modeling of Physiological Systems*. Hoboken, NJ, USA: John Wiley & Sons, 2004.

[8] R. do Prado, F. Henriques, and D. Haddad, "Sparsity-aware distributed adaptive filtering algorithms for nonlinear system identification," in *Proc. of IJCNN 2018*, Rio de Janeiro, Brazil, July 2018.

[9] G. L. Sicuranza and A. Carini, "Nonlinear system identification using quasi-perfect periodic sequences," *Signal Processing*, vol. 120, pp. 174–184, Mar. 2016.

[10] A. Carini, S. Cecchi, and S. Orcioni, "Orthogonal LIP nonlinear filters," in *Adaptive learning methods for nonlinear system modeling*, D. Comminello and J. C. Príncipe, Eds. Elsevier, 2018, ch. 2.

[11] A. Carini, S. Orcioni, A. Terenzi, and S. Cecchi, "Nonlinear system identification using Wiener basis functions and multiple-variance perfect sequences," *Signal Processing*, vol. 160, pp. 137–149, 2019.

[12] H. Yazdanpanah, A. Carini, and M. V. S. Lima, "L0-norm adaptive Volterra filters," in *2019 27th European Signal Processing Conference (EUSIPCO)*, Sept. 2019, pp. 1–5.

[13] M. V. S. Lima, T. N. Ferreira, W. A. Martins, and P. S. R. Diniz, "Sparsity-aware data-selective adaptive filters," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4557–4572, Sept. 2014.

[14] D. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 508–518, Sept. 2000.

[15] J. Benesty and S. Gay, "An improved PNLMS algorithm," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2002)*, vol. 2, Orlando, FL, USA, May 2002, pp. 1881–1884.

[16] R. Meng, R. de Lamare, and V. Nascimento, "Sparsity-aware affine projection adaptive algorithms for system identification," in *Sensor Signal Processing for Defence*, London, U.K., Sept. 2011, pp. 1–5.

[17] Y. Chen, Y. Gu, and A. Hero, "Sparse LMS for system identification," in *Proc. of ICASSP 2009*, Taipei, Taiwan, Apr. 2009, pp. 3125–3128.

[18] E. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[19] M. V. S. Lima, W. A. Martins, and P. S. R. Diniz, "Affine projection algorithms for sparse system identification," in *Proc. of ICASSP 2013*, Vancouver, Canada, May 2013, pp. 5666–5670.

[20] Y. Gu, J. Jin, and S. Mei, "l_0 norm constraint LMS algorithm for sparse system identification," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 774–777, Sept. 2009.

[21] A. Mader, H. Puder, and G. U. Schmidt, "Step-size control for acoustic echo cancellation filters – an overview," *Signal Processing*, vol. 4, pp. 1697–1719, Sep. 2000.

[22] F. Kuech and W. Kellermann, "Coefficient-dependent step-size for adaptive second-order Volterra filters," in *2004 12th European Signal Processing Conference*, 2004, pp. 1805–1808.

[23] C.-S. Shih and S.-F. Hsieh, "Fast converging nonlinear echo cancellation based on optimum step size," in *IEEE International Symposium on Signal Processing and Information Technology*, 2009, pp. 135–139.

[24] C. Contan, B. S. Kirei, and M. D. Topa, "Error-dependent step-size control of adaptive normalized least-mean-square filters used for nonlinear acoustic echo cancellation," *Signal, Image and Video Processing*, vol. 10, no. 3, pp. 511–518, 2016.

[25] W. Rudin, *Principles of Mathematical Analysis*. New York: McGraw-Hill, 1976.

[26] S. L. Gay, "An efficient, fast converging adaptive filter for network echo cancellation," in *Conference Record of Thirty-Second Asilomar Conference on Signals, Systems and Computers (Cat. No. 98CH36284)*, vol. 1. IEEE, 1998, pp. 394–398.

[27] F. Kuech and W. Kellermann, "Proportionate NLMS algorithm for second-order Volterra filters and its application to nonlinear echo cancellation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC), Kyoto*, vol. 188, 2003.

[28] F. Küch, W. Kellermann, E. Hansler, and G. Schmidt, "Nonlinear acoustic echo cancellation," *Topics in Acoustic Echo and Noise Control*, pp. 205–257, 2006.