Online Selection of Functional Links for Nonlinear System Identification

Danilo Comminiello, Simone Scardapane, Michele Scarpiniti, Raffaele Parisi, and Aurelio Uncini

Department of Information Engineering, Electronics and Telecommunications (DIET)
"Sapienza" University of Rome
via Eudossiana 18, 00184 Rome, Italy
{danilo.comminiello,simone.scardapane,michele.scarpiniti, raffaele.parisi}@uniroma1.it,aurel@ieee.org

Abstract. This paper introduces a new method for improving nonlinear modeling performance in online learning by using functional link-based models. The proposed algorithm is capable of selecting the useful nonlinear elements resulting from the functional expansion, while setting to zero the ones that does not bring any improvement of the modeling performance. This allows to reduce any gradient noise due to a possible overestimate of the solution, thus preventing any overfitting phenomena. The proposed model is assessed in several nonlinear identification problems, including different levels of nonlinearity, showing significant improvements.

Keywords: Nonlinear Modeling, Functional Links, Nonlinear Transformation, Nonlinear System Identification, Sparse Systems.

1 Introduction

The nonlinearity degree in a signal may depend on several factors related to the signal itself, such as its nonstationary or time-varying nature. Therefore, in nonlinear system identification problems, it becomes very difficult to design *a priori* a nonlinear model to be used without incurring in any overfitting issue. All along the years, in offline learning problems, pruning methods have been widely applied to batch and sequential models, due to the modeling performance improvement that they produce [8,9,12,11,16]. However, in online learning problems, these methods may not be appropriate, sometimes due to expensive computational load, or even due to batch processing. In this case, other methods can be adopted that do not actually prune unnecessary elements, but just perform an online selection of the useful elements.

In this paper, we propose a new method for improving nonlinear modeling performance in online learning, which performs an online selection of the nonlinear elements, thus reducing any gradient noise that may be generated by a possible overestimate of the solution. We focus on a class of linear-in-the-nonlinear adaptive models [18], which are based on a nonlinear transformation of the input signal that projects it in a higher dimensional space. Then, the transformed

S. Bassis et al. (eds.), *Recent Advances of Neural Networks Models and Applications*,
Smart Innovation, Systems and Technologies 37, DOI: 10.1007/978-3-319-18164-6_5

signal can be processed by a linear model. In particular, we take into account the nonlinear *functional link adaptive filters* (FLAFs) [3], in which the nonlinear expansion is carried out by the so-called *functional links* [14,15,19,1], and the subsequent linear model is an adaptive filter.

One of the main advantages of the FLAF model lies in its flexibility, since the setting of several parameters is allowed in order to fit the model to a specific application. In this regard, an important choice in the FLAF design concerns the amount of functional links to be employed for the modeling. This choice is strictly related to the nonlinearity degree introduced by the unknown system. However, when the number of functional links is too high with respect to the nonlinearity to be modeled, it may lead to overfitting phenomena that may cause a decrease of the modeling performance [4]. This is due to the fact that only a portion ,i.e. a sparse representation, of functional links actively contributes to the filtering.

In order to address this problem, we exploit a sparse representation of the functional links [4] to perform an online selection and thus improve the performance. This approach is based on the proportionate adaptive algorithms [7,10,13]. Proportionate algorithms were developed to improve the convergence performance in linear systems when the impulse response to be estimated shows a sparse nature, i.e., many of its coefficients are zero or very close to zero. However, sparsity is not only related to linear systems, but it can also occur in the estimation of nonlinearities, as not all the elements of a nonlinear model may be useful for a correct modeling. Unlike [4], where a *split FLAF* architecture was proposed for nonlinear acoustic echo cancellation, in this paper, we focus on a *μ-law* rule [6] to exploit a sparse functional link representation for nonlinear FLAF. The resulting *μ-law proportionate FLAF* algorithm gives a greater importance to the coefficients that contribute actively to the nonlinear modeling. At the same time, any overfitting phenomenon caused by the unnecessary coefficients is avoided. The effectiveness of the the proposed method is assessed in the nonlinear system identification problems, which requires online processing.

The paper is organized as follows: the nonlinear FLAF model is introduced in Section 2 and the proposed algorithm is detailed in Section 3. Results are discussed in Section 4 and, finally, in Section 5 our conclusions are drawn.

2 A Brief Review on the Nonlinear FLAF

The FLAF model is based on the representation of the input signal in a higher-dimensional space [14], in which an enhanced nonlinear modeling is allowed. Such approach derives from the machine learning theory, more precisely from the Cover's Theorem on the separability of patterns (see for example [8]).

The purely nonlinear FLAF is composed of two main parts: a nonlinear *functional expansion block* (FEB) and a subsequent linear adaptive filter, as depicted in Fig. 1. The FEB consists of a series of functions, which might be a subset of a complete set of orthonormal basis functions satisfying universal approximation constraints. The term "functional links" actually refers to the functions

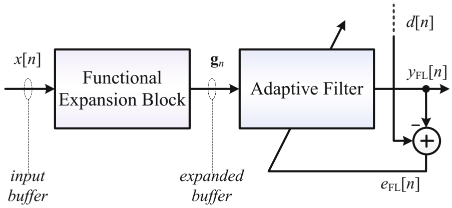


Fig. 1. The nonlinear functional link adaptive filter

contained in the chosen set $\Phi = \left\{\varphi_0(\cdot), \varphi_1(\cdot), \ldots, \varphi_{Q-1}(\cdot)\right\}$, where Q is the number of functional links. At the n-th time instant, the FEB receives the input sample $x[n]$, which is stored in an input buffer $\mathbf{x}_{\mathrm{N},n} \in \mathbb{R}^{M_\mathrm{i}} = \left[x[n]\; x[n-1] \ldots x[n-M_\mathrm{i}+1]\right]^T$, where M_i is defined as the input buffer length. Each element of $\mathbf{x}_{\mathrm{N},n}$ is passed as argument to the chosen set of functions Φ, thus yielding a subvector $\overline{\mathbf{g}}_{i,n} \in \mathbb{R}^Q$:

$$\overline{\mathbf{g}}_{i,n} = \left[\varphi_0(x[n-i])\; \varphi_1(x[n-i]) \ldots \varphi_{Q-1}(x[n-i])\right]. \tag{1}$$

The concatenation of all the subvectors, for $i = 0, \ldots, M_\mathrm{i} - 1$, engenders an *expanded buffer* $\mathbf{g}_n \in \mathbb{R}^{M_\mathrm{e}}$:

$$\mathbf{g}_n = \left[\overline{\mathbf{g}}_{0,n}^T\; \overline{\mathbf{g}}_{1,n}^T \ldots \overline{\mathbf{g}}_{M_\mathrm{i}-1,n}^T\right]^T = \left[g_0[n]\; g_1[n] \ldots g_{M_\mathrm{e}-1}[n]\right]^T \tag{2}$$

where $M_\mathrm{e} \geq M_\mathrm{i}$ represents the length of the expanded buffer. Note that $M_\mathrm{e} = M_\mathrm{i}$ only when $Q = 1$. As functional expansion, we choose a nonlinear trigonometric series expansion:

$$\varphi_j(x[n-i]) = \begin{cases} \sin(p\pi x[n-i]), & j = 2p-2 \\ \cos(p\pi x[n-i]), & j = 2p-1 \end{cases} \tag{3}$$

where $p = 1, \ldots, P$ is the expansion index, being P the *expansion order*, and $j = 0, \ldots, Q-1$ is the functional link index. The trigonometric expansion (3) implies a functional link set Φ composed of $Q = 2P$ functional links. Some convergence properties of the nonlinear FLAF of Fig. 1 can be found in [5]. It is worth noting that (3) actually refers to a *memoryless* expansion, since it does not involve cross-products, but it can be easily extended to a memory expansion (see [3] for a detailed explanation). A way of considering the memory of a nonlinearity is that of taking into account the outer products of the i-th input sample with the functional links of the previous input samples. The FLAF with memory is characterized by a *memory order* K [3].

The expanded buffer $\mathbf{g}_n$ is then fed into a linear adaptive filter $\mathbf{w}_{\mathrm{FL},n} \in \mathbb{R}^{M_\mathrm{e}} = \left[w_{\mathrm{FL},0}[n]\; w_{\mathrm{FL},1}[n] \ldots\; w_{\mathrm{FL},M_\mathrm{e}-1}[n]\right]^T$, thus providing the nonlinear output:

$$y_{\mathrm{FL}}[n] = \mathbf{g}_n^T \mathbf{w}_{\mathrm{FL},n-1}. \tag{4}$$

Thereby, the nonlinear error signal is:

$$e_{\text{FL}}[n] = d[n] - y_{\text{FL}}[n] \tag{5}$$

which is used for the adaptation of $\mathbf{w}_{\text{FL},n}$. In (5), $d[n]$ represents the desired signal for the nonlinear model. Being $\mathbf{w}_{\text{FL},n}$ a conventional linear filter, it can be adapted by any adaptive algorithm based on the minimization of the mean square error [20]. The use of an adaptive filter after the expansion allows to apply the FLAF model to several online learning applications, such as active noise reduction, acoustic echo cancellation [19,2,17,3].

3 The μ-law Proportionate FLAF

Very often, nonlinearities affecting an input signal may vary in time and frequency. This behavior is further stressed when the input signals has a nonstationary nature. This is the reason why not all the nonlinear elements of an expanded buffer may be useful in the same way to model a dynamic nonlinear channel. A possible solution to this drawback is that of using a weighted mask for the nonlinear filter in an attempt to give more prominence to those nonlinear elements of the expanded buffer that have an active role in the modeling of nonlinearities. To this end, we introduce a weighted adaptive algorithm for the nonlinear FLAF, that provides a sparse representation of functional links, thus performing an online selection of them. In order to exploit the sparsity in the expanded buffer, proportionate adaptive algorithms may be adopted [7,10,13]. Among such algorithms, we take into account the *μ-law proportionate normalized least mean square* (MPNLMS) algorithm, which is based on an approximation of the optimal proportionate step size [6]. According to this, the update equation of the MPNLMS-FLAF can be expressed as:

$$\mathbf{w}_{\text{FL},n} = \mathbf{w}_{\text{FL},n-1} + \eta \frac{\mathbf{Q}_n \mathbf{g}_n}{\mathbf{g}_n^T \mathbf{Q}_n \mathbf{g}_n + \delta_{\text{P}}} e_{\text{FL}}[n] \tag{6}$$

where η is the step-size parameter and δ_{rP} is a regularization factor. $\mathbf{Q}_n$ is a diagonal weighting matrix that contains the proportionate coefficients $q_m[n]$, with $m = 0, \ldots, M_{\text{e}} - 1$, whose values are computed according to [6]:

$$q_m[n] = \frac{\gamma_m[n]}{\frac{1}{M_{\text{e}}} \sum_{i=0}^{M_{\text{e}}-1} \gamma_i[n]}, \qquad m = 0, \ldots, M_{\text{e}} - 1. \tag{7}$$

The coefficients $\gamma_m[n]$ can be computed by introducing a function of the estimate of the optimal filter coefficient:

$$\theta_m[n] = \frac{\ln(1 + \mu |w_{\text{FL},m}[n-1]|)}{\ln(1 + \mu)}, \qquad m = 0, \ldots, M_{\text{e}} - 1 \tag{8}$$

where the step size can be represented as $\mu = 1/\varepsilon$. The parameter ε is a very small positive number and its value can be chosen according to the measurement noise. As a default choice, we set $\varepsilon = 0.001$ (i.e., $\mu = 1000$) that means that the noise below -60 dB is negligible. In (8), the constant 1 inside the logarithm has been introduced in order to avoid a singular point when $|w_{\text{FL},m}[n]| = 0$. Moreover, the denominator normalizes the function to be in the range $[0, 1]$. It is worth noting that the function $\theta_m[n]$ is nothing but the μ-law used in nonuniform compression in telecommunication applications [6].

Based on (8), it is possible to define, first, a lower bound for the coefficients $\gamma_m[n]$:

$$\gamma_{\min}[n] = \rho \max\{\xi, \theta_0[n], \ldots, \theta_{M_e-1}[n]\} \tag{9}$$

where ρ is a scaling factor and ξ is a threshold value, usually chosen respectively as $\rho = 0.01$ and $\xi = 0.01$. Then, using (8) and (9), it is possible to define the coefficients $\gamma_m[n]$:

$$\gamma_m[n] = \max\{\gamma_{\min}[n], \theta_m[n]\} \tag{10}$$

that can be finally used to derive the proportionate coefficients in (7) and, thus, to achieve the update equation (6) for the MPNLMS-FLAF.

4 Experimental Results

In this section, we evaluate the nonlinear modeling performance of the proposed MPNLMS-FLAF. We assess the effectiveness of the MPNLMS-FLAF over three different system identification scenarios, which are distinguished according to the nonlinearity degree introduced by an unknown system. In all the scenarios, the plant to be identified is composed of a nonlinear system followed by a linear system, in a Hammerstein configuration as depicted in Fig. 2. The input signal is generated by a first-order autoregressive model, whose transfer function is $\sqrt{1-\theta^2}/\left(1-\theta z^{-1}\right)$, with $\theta = 0.8$, fed with an independent identically distributed (i.i.d.) Gaussian random process. The length of the input signal is $L = 20000$ samples. In each scenario, the linear system is formed by $M = 7$ independent random values between -1 and 1. An additive i.i.d. noise signal $v[n]$ is added at the output of the whole plant, in order to provide 30 dB of signal-to-noise ratio (SNR). Performance is evaluated in terms of the *excess mean square error* (EMSE), in dB:

$$\text{EMSE}[n] = 10\log_{10}\left(\text{E}\left\{(e[n] - v[n])^2\right\}\right) \tag{11}$$

which is averaged over 1000 runs with respect to input and noise. Moreover, in order to facilitate the visualization, curves are smoothed by a moving-average

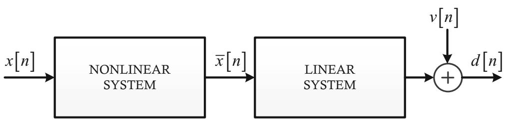
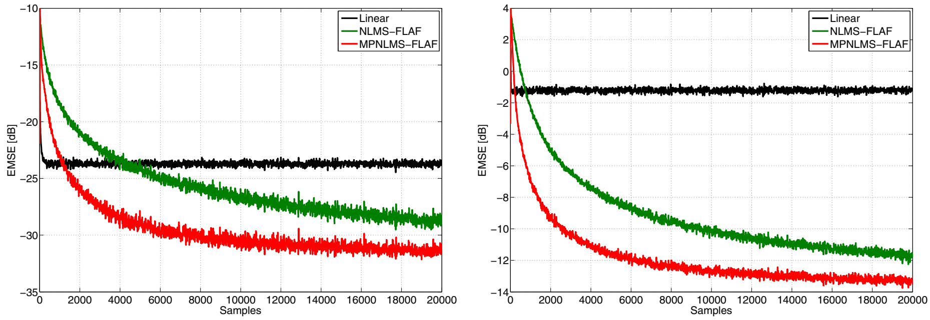


Fig. 2. General scheme for nonlinear system identification scenarios

danilo.comminiello@uniroma1.it



(a) (b)

Fig. 3. Performance behavior in terms of ERLE in the presence of: (a) mild nonlinearity, and (b) strong nonlinearity

filter. In all the scenarios, the proposed MPNLMS-FLAF is compared, in terms of the EMSE, with the standard NLMS-FLAF, in which the FLAF is simply adapted by an NLMS algorithm [3], and with a simple NLMS algorithm, in order to have a linear reference.

In the first scenario, we assume that the nonlinear system applies a symmetrical soft-clipping distortion to the input signal, described by the following equation [4]:

$$\overline{x}[n] = \begin{cases} \frac{2}{3\zeta} x[n] & \text{for } 0 \leq |x[n]| \leq \zeta \\ \operatorname{sign}(x[n]) \frac{3-(2-x[n]/\zeta)^2}{3} & \text{for } \zeta \leq |x[n]| \leq 2\zeta \\ \operatorname{sign}(x[n]) & \text{for } 2\zeta \leq |x[n]| \leq 1 \end{cases} \tag{12}$$

where ζ is a threshold chosen in the range $(0\,,\,0.5]$. For this experiment this threshold is chosen as $\zeta = 0.15$, thus providing a medium/mild degree of nonlinearity. We normalize the input signal to limit its amplitude in the range $[-1, 1]$. The resulting signal $\overline{x}[n]$ is convolved with the linear impulse response, as in Fig. 2. The parameter setting of FLAF-based models for this experiment is: $\mu_{\text{FL}} = 0.2$, $\delta_{\text{FL}} = 10^{-3}$, $M_{\text{i}} = M$, $P = 15$. Memoryless FLAFs are considered for this experiments (i.e., $K = 0$). Results are shown in Fig. 3(a), in which it can be seen that the proposed MPNLMS-FLAF outperforms the NLMS-FLAF.

In the second scenario, as regards the nonlinear system, we consider the same symmetrical soft-clipping distortion of (12), but we choose a threshold $\zeta = 0.05$, which yields a high degree of nonlinearity. Differently from the previous parameter setting, in this case we use a higher expansion order, i.e., $P = 30$. Results are shown in Fig. 3(b), where it is possible to notice that the gap between the linear model and the FLAF-based ones is larger with respect to the previous scenario, since the nonlinearity introduced is stronger than before. Moreover, the proposed MPNLMS-FLAF keeps its performance gain over the NLMS-FLAF.

In the last scenario, we increase the difficulty of the problem by considering a dynamic nonlinearity, whose function involves a temporal dependence.

In particular, the nonlinear system in Fig. 2 is composed of two subsequent blocks. The first block receives the input signal $x[n]$ and applies the following dynamic nonlinearity:

$$\begin{aligned}\overline{u}[n] = &\frac{3}{2}\cos\left(\frac{\pi}{2}x[n]\right) - \frac{3}{10}\cos\left(\frac{\pi}{2}x[n]\right)^2 + \frac{9}{5}\cos\left(\frac{\pi}{2}x[n]\right)\sin\left(\frac{\pi}{4}x[n-1]\right)\\ &+ \frac{1}{2}\sin\left(\frac{\pi}{2}x[n]\right)\cos\left(\frac{\pi}{8}x[n-2]\right) - \frac{2}{5}\sin\left(\frac{\pi}{2}x[n]\right)\cos\left(\frac{\pi}{16}x[n-3]\right)\\ &- \frac{3}{2}\sin\left(\frac{\pi}{4}x[n-1]\right)\cos\left(\frac{\pi}{8}x[n-2]\right)\\ &+ \frac{9}{10}\sin\left(\frac{\pi}{2}x[n]\right)\cos\left(\frac{\pi}{4}x[n-1]\right)\sin\left(\frac{\pi}{16}x[n-3]\right). \end{aligned} \tag{13}$$

The resulting signal $\overline{u}[n]$ is then processed by a 3-rd order Chebyshev filter with transfer function:

$$H(z) = \frac{0.6055 + 0.7785z^{-1} + 0.778z^{-2} + 0.6055z^{-3}}{1 + 0.6416z^{-1} + 0.7692z^{-2} + 0.3574z^{-3}}, \tag{14}$$

thus yielding the nonlinear signal $\overline{u}[n]$ that is then fed into the linear system according to the sceme in Fig. 2. For this experiment, we choose the same expansion order of the previous experiment, $P = 30$, but we take into account also some memory in the functional links in order to better model the time-dependent nonlinearity. In particular, we choose a memory order $K = 5$, which represents a good compromise between performance and computational cost [3]. The other parameters are setting as in the previous experiments. Results are shown in Fig. 4, where it can be seen that, while the linear model is not able to provide any improvement, the FLAF-based models achieve good results, and, in particular, the MPNLMS-FLAF takes advantage of its parameter selection to provide a larger improvement with respect to the NLMS-FLAF. It should be noted that, although the nonlinearity adopted in the third scenario is very strong, it is composed of sine and cosines, thus facilitating the modeling by trigonometric functional links.

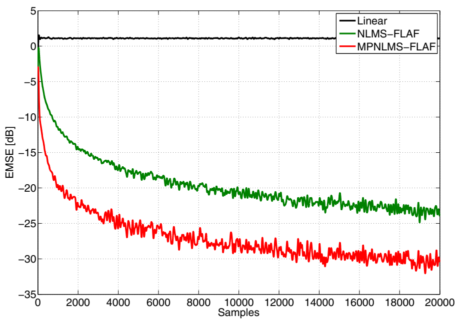


Fig. 4. Performance behavior in terms of the EMSE in the presence of strong dynamic nonlinearity

danilo.comminiello@uniroma1.it

By considering the three experiments above, it is worth noting that the proposed MPNLMS-FLAF achieves an improvement as large as higher the nonlinearity degree. This is due to the fact that a strong nonlinearity needs a larger expansion buffer, whose effectiveness of the nonlinear elements is not uniform but sparse. Moreover, it can be noticed that, unlike the MPNLMS-FLAF, the NLMS-FLAF considers also the useless functional links, which generates overfitting. Therefore, we can conclude that the performance gap of the NLMS-FLAF from the proposed MPNLMS-FLAF is essentially due to the overfitting.

5 Conclusion

In this paper, a new algorithm has been proposed to perform an online selection of functional links in a FLAF. The selection of functional links serves to prevent the occurrence of overfitting phenomena when the system to be identified is unknown. The proposed MPNLMS-FLAF is based on the μ-law proportionate adaptive algorithm and exploits the sparse representation of functional links. This model has been assessed in nonlinear system identification problems. In particular, three scenarios with different nonlinearity degrees have been considered. Results have proved the effectiveness of the proposed method for all the scenarios. In future works, the adopted algorithm may be extended also to more sophisticated FLAF-based architectures for problems like the nonlinear acoustic echo cancellation. Moreover, other kinds of proportionate algorithms can be investigated to online select functional links. The method can be also extended to other classes of linear-in-the-parameters nonlinear models.

References

1. Alhamdoosh, M., Wang, D.: Fast decorrelated neural network ensambles with random weights. Inform. Sciences 264, 104–117 (2014)
2. Comminiello, D., Azpicueta-Ruiz, L.A., Scarpiniti, M., Uncini, A., Arenas-Garcia, J.: Functional link based architectures for nonlinear acoustic echo cancellation. In: Proc. IEEE J. Works. Hands-free Speech Commun. Microph. Arrays (HSCMA 2011), Edinburgh, UK, pp. 180–184 (May 2011)
3. Comminiello, D., Scarpiniti, M., Azpicueta-Ruiz, L.A., Arenas-García, J., Uncini, A.: Functional link adaptive filters for nonlinear acoustic echo cancellation. IEEE Trans. Audio, Speech, Lang. Process. 21(7), 1502–1512 (2013)
4. Comminiello, D., Scarpiniti, M., Azpicueta-Ruiz, L.A., Arenas-García, J., Uncini, A.: Nonlinear acoustic echo cancellation based on sparse functional link representations. To appear in IEEE Trans. Audio, Speech, Lang. Process. (2014), doi:10.1109/TASLP.2014.2324175
5. Comminiello, D., Scarpiniti, M., Parisi, R., Uncini, A.: Convergence properties of nonlinear functional link adaptive filters. IET Electron. Lett. 49(14), 873–875 (2013)
6. Doroslovački, M., Deng, H.: Proportionate adaptive algorithms for network echo cancellation. IEEE Trans. Signal Process. 54(5), 1794–1803 (2006)
7. Duttweiler, D.L.: Proportionate normalized least-mean-squares adaptation in echo cancelers. IEEE Trans. Speech Audio Process. 8(5), 508–518 (2000)

8. Haykin, S.: Neural Networks and Learning Machines, 3rd edn. Prentice Hall, Upper Saddle River (2008)
9. Karnin, E.: A simple procedure for pruning back-propagation trained neural networks. IEEE Trans. Neural Netw. 1(2), 239–242 (1990)
10. Naylor, P., Cui, J., Brookes, M.: Adaptive algorithms for sparse echo cancellation. Signal Process. 86(6), 1182–1192 (2006)
11. Omlin, C., Giles, C.: Pruning recurrent neural networks for improved generalization performance. IEEE Trans. Neural Netw. 5(5), 848–851 (1994)
12. Orlandi, G., Piazza, F., Uncini, A., Ascone, A.: A biological approach to plasticity in artificial neural netowrks. In: Proc. IEEE Int. J. Conf. Neural Netw (IJCNN 1991), Seattle,WA, July 8-12, vol. 2, pp. 583–586 (1991)
13. Paleologu, C., Benesty, J., Ciochină, S.: Sparse Adaptive Filters for Echo Cancellation. Morgan & Claypool Publishers (2010)
14. Pao, Y.H.: Adaptive Pattern Recognition and Neural Networks. Addison-Wesley, Reading (1989)
15. Patra, J.C., Pal, R.N., Chatterji, B.N., Panda, G.: Identification of nonlinear dynamic systems using functional link artificial neural networks. IEEE Trans. Syst., Man, Cybern., B: Cybern. 29(2), 254–262 (1999)
16. Scardapane, S., Nocco, G., Comminiello, D., Scarpiniti, M., Uncini, A.: An effective criterion for pruning reservoir's connections in echo state networks. In: Proc. IEEE WCCI - Int. J. Conf. Neural Netw. (IJCNN 2014), Beijing, China, July 6-11 (2014)
17. Scarpiniti, M., Comminiello, D., Parisi, R., Uncini, A.: A collaborative approach to adaptive noise cancellation. In: Proc. XXII Italian Works. Neural Netw. (WIRN), Vietri Sul Mare, Italy, May 17-19 (2012)
18. Sicuranza, G., Carini, A.: On a class of nonlinear filters. In: Tabus, M., Egiazarian, K. (eds.) Festschrift in Honor of Jaakko Astola on the Occasion of his 60th Birthday. TICPS, vol. 47, pp. 115–144 (2009)
19. Sicuranza, G.L., Carini, A.: A generalized FLANN filter for nonlinear active noise control. IEEE Trans. Audio, Speech, Lang. Process. 19(8), 2412–2417 (2011)
20. Uncini, A.: Fundamentals of Adaptive Signal Processing. Springer (2014) ISBN 978-3-319-02806-4