A NONLINEAR ARCHITECTURE INVOLVING A COMBINATION OF PROPORTIONATE FUNCTIONAL LINK ADAPTIVE FILTERS

Danilo Comminiello[*†], *Michele Scarpiniti*[†], *Luis A. Azpicueta-Ruiz*[‡], *Jerónimo Arenas-García*[‡], and *Aurelio Uncini*[†]

[†] Dept. Inform. Eng., Electron. and Telecom.
"Sapienza" University of Rome
Via Eudossiana 18, 00184 Rome, Italy

[‡] Dept. Signal Theory and Communications
Universidad Carlos III de Madrid
28911 Leganés, Spain

ABSTRACT

In this paper, we consider a functional link-based architecture that separates the linear and nonlinear filterings and exploits any sparse representation of functional links. We focus our attention on the nonlinear path in order to improve the modeling performance of the overall architecture. To this end, we propose a new scheme that involves the adaptive combination of filters downstream of the nonlinear expansion. This combination enhances the sparse representation of functional links according to how much distorted the input signal is, thus improving the nonlinear modeling performance in case of time-varying nonlinear systems. Experimental results show the performance improvement produced by the proposed model.

Index Terms— Nonlinear Adaptive Filtering, Functional Links, Linear-in-the-Parameters Nonlinear Filters, Sparse Representations, Adaptive Combination of Filters

1. INTRODUCTION

Online learning for nonlinear system modeling has always drawn a great interest due to a wide range of applications that can be found in this field. One of the most popular models for nonlinear system identification is the class of linear-in-the-parameters (LIP) nonlinear filters [1, 2], which is characterized by a linear filtering of any nonlinear representation of the input signal. Among this class of filters, several models have been proposed, including adaptive Volterra filters [1, 3], even mirror Fourier nonlinear filters [4], Hammerstein spline adaptive filters [5], online extreme learning machines [6] and Legendre nonlinear filters [7], among others.

In this work, we focus on a class of LIP nonlinear adaptive filters based on *functional links*, known as functional link artificial neural network (FLANN) filters [8], or also as functional link adaptive filters (FLAFs) [9]. These filters are characterized by a nonlinear expansion of the input followed by a linear filtering of the expanded signal. In particular, we take into account a split FLAF (SFLAF) architecture [9], which separates the adaptation of linear and nonlinear elements.

Based on the nonlinearity level introduced by any system to be modeled, the functional links involved in a SFLAF may show a sparse representation, i.e., only a selection of them is really useful to model a nonlinear system. This is the reason why the proportionate SFLAF (PSFLAF) has been introduced [10] to exploit such sparse functional link representations. Here, we propose a new architecture that improves the nonlinear modeling performance of the PSFLAF. Such scheme, called combined PSFLAF (cPSFLAF), is characterized by a convex combination of two proportionate adaptive filters [11, 12] downstream of the functional expansion. This leads to a more general model to be used regardless of any nonlinearity level caused by the unknown system. Results show effectiveness and robustness of the proposed model in a time-varying nonlinear system identification scenario.

The rest of the paper is organized as follows: the PSFLAF exploiting sparse functional link representations is described in Section 2. In Section 3, the proposed cPSFLAF scheme is introduced, and, in Section 4, experimental results are shown. Finally, in Section 5 our conclusions are presented.

2. NONLINEAR MODELING BY SPARSE REPRESENTATION OF FUNCTIONAL LINKS

Very often in real-world problems, the response of a system to be identified is produced by any combination of a linear and nonlinear components. In order to model such a system, the *split functional link adaptive filter* (SFLAF) architecture was recently proposed [9]. The SFLAF, depicted in Fig. 1, is a parallel architecture including a linear path and a nonlinear path. The former is simply composed of a linear adaptive filter, which is devoted to model the linear components of an unknown system. This allows the nonlinear path to be focused on the modeling of the nonlinear components of the system.

At n-th time instant the SFLAF receives the input sample $x[n]$, which is stored in the linear input buffer $\mathbf{x}_{\mathrm{L},n} \in \mathbb{R}^M = \left[\begin{array}{ccc} x[n] & \ldots & x[n-M+1] \end{array}\right]^T$, where M is the length of

[*] Corresponding author e-mail: danilo.comminiello@uniroma1.it. The work of Danilo Comminiello was partly funded by bdSound.

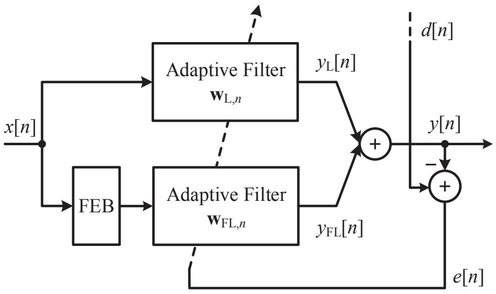

Fig. 1. The split functional link adaptive filter.

the adaptive filter $\mathbf{w}_{\mathrm{L},n} = \left[\begin{array}{ccc} w_{\mathrm{L},0}[n] & \ldots & w_{\mathrm{L},M-1}[n] \end{array}\right]^T$. The adaptive filtering yields the linear output $y_{\mathrm{L}}[n] = \mathbf{x}_{\mathrm{L},n}^T \mathbf{w}_{\mathrm{L},n-1}$. The input $x[n]$ is also stored in $\mathbf{x}_{\mathrm{N},n} = \left[\begin{array}{ccc} x[n] & \ldots & x[n-M_{\mathrm{i}}+1] \end{array}\right]^T$, where M_{i} is defined as the input buffer length. This signal is processed by a functional expansion block (FEB), which consists of a set of *functional links* $\Phi = \left\{\varphi_0(\cdot), \ldots, \varphi_{Q_{\mathrm{f}}-1}(\cdot)\right\}$, where Q_{f} is the number of functional links. Each element of $\mathbf{x}_{\mathrm{N},n}$ is expanded by the chosen set of functions Φ, thus yielding the subvectors $\overline{\mathbf{g}}_{i,n} = \left[\begin{array}{ccc} \varphi_0(x[n-i]) & \ldots & \varphi_{Q_{\mathrm{f}}-1}(x[n-i]) \end{array}\right]$. The concatenation of all the subvectors, for $i = 0, \ldots, M_{\mathrm{i}}-1$, yields an *expanded buffer* $\mathbf{g}_n \in \mathbb{R}^{M_{\mathrm{e}}}$:

$$\begin{aligned} \mathbf{g}_n &= \left[\begin{array}{cccc} \overline{\mathbf{g}}_{0,n}^T & \overline{\mathbf{g}}_{1,n}^T & \cdots & \overline{\mathbf{g}}_{M_{\mathrm{i}}-1,n}^T \end{array}\right]^T \\ &= \left[\begin{array}{cccc} g_0[n] & g_1[n] & \ldots & g_{M_{\mathrm{e}}-1}[n] \end{array}\right]^T, \end{aligned} \tag{1}$$

where $M_{\mathrm{e}} \geq M_{\mathrm{i}}$ represents the length of this expanded buffer. Note that $M_{\mathrm{e}} = M_{\mathrm{i}}$ only when $Q_{\mathrm{f}} = 1$.

Several choices can be made for the functional expansion in the FEB. Here, we use a nonlinear trigonometric series expansion such that:

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

where $p = 1, \ldots, P$ is the expansion index, being P the *expansion order*, and $j = 0, \ldots, Q_{\mathrm{f}}-1$ is the functional link index. In the case of trigonometric expansion, it is easy to verify that the set Φ is composed of $Q_{\mathrm{f}} = 2P$ functional links.

It is worth noting that $\mathbf{g}_n$ is composed of nonlinear elements only, since the linear part of a system to identified can be demanded to the linear path. The achieved expanded buffer $\mathbf{g}_n$ is then fed into an adaptive filter $\mathbf{w}_{\mathrm{FL},n} \in \mathbb{R}^{M_{\mathrm{e}}} = \left[\begin{array}{ccc} w_{\mathrm{FL},0}[n] & \ldots & w_{\mathrm{FL},M_{\mathrm{e}}-1}[n] \end{array}\right]^T$, thus providing the nonlinear output $y_{\mathrm{FL}}[n] = \mathbf{g}_n^T \mathbf{w}_{\mathrm{FL},n-1}$.

The overall error signal of the SFLAF is:

$$\begin{aligned} e[n] &= d[n] - y[n] \\ &= d[n] - \mathbf{x}_{\mathrm{L},n}^T \mathbf{w}_{\mathrm{L},n-1} - \mathbf{g}_n^T \mathbf{w}_{\mathrm{FL},n-1} \end{aligned} \tag{3}$$

where $y[n]$ stands for the overall SFLAF output signal, resulting from the sum of the two path outputs. Both adaptive filters try to minimize the power of $e[n]$, and this can be realized following different adaptation schemes. However, in order to focus on the nonlinear path we choose a classic *normalized least-mean square* (NLMS) algorithm for $\mathbf{w}_{\mathrm{L},n}$. Hence:

$$\mathbf{w}_{\mathrm{L},n} = \mathbf{w}_{\mathrm{L},n-1} + \mu_{\mathrm{L}} \frac{\mathbf{x}_{\mathrm{L},n} e[n]}{\mathbf{x}_{\mathrm{L},n}^T \mathbf{x}_{\mathrm{L},n} + \delta_{\mathrm{L}}} \tag{4}$$

where δ_{L} is a regularization factor and $0 < \mu_{\mathrm{L}} < 2$ is a step-size parameter.

When nonlinearities introduced by the unknown system are varying in time and/or in amplitude, the nonlinear elements generated by the FEB and stored in $\mathbf{g}_n$ may be not all useful in the same way for the modeling, and this may cause a performance decrease. A possible solution is the use of a weighted mask for the filter of the nonlinear path [10], 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. This means taking advantage of a sparse representation of functional links. To this end, the update equation of $\mathbf{w}_{\mathrm{FL},n}$ can be written as [10]:

$$\mathbf{w}_{\mathrm{FL},n} = \mathbf{w}_{\mathrm{FL},n-1} + \mu_{\mathrm{FL}} \frac{\mathbf{Q}_n \mathbf{g}_n}{\mathbf{g}_n^T \mathbf{Q}_n \mathbf{g}_n + \delta_{\mathrm{PFL}}} e[n] \tag{5}$$

where δ_{PFL} is a regularization factor, $0 < \mu_{\mathrm{FL}} < 2 - \mu_{\mathrm{L}}$ is a step-size parameter [10], and

$$\mathbf{Q}_n = \operatorname{diag}\left\{\begin{array}{ccc} q_0[n] & \ldots & q_{M_{\mathrm{e}}-1}[n] \end{array}\right\}, \tag{6}$$

is the proportionate matrix that performs a mask on the filter according to its sparsity, whose goal is to weight the coefficients of $\mathbf{w}_{\mathrm{FL},n}$ proportionally to the contribution they provide to the nonlinear modeling. For this reason, we denote this architecture as *proportionate SFLAF* (PSFLAF).

The diagonal elements of $\mathbf{Q}_n$ are computed by evaluating the nonlinear filter estimate at the previous time instant. The larger a coefficient value of $\mathbf{w}_{\mathrm{FL},n-1}$, the higher the corresponding weighting. Note that when sparsity is not considered, $\mathbf{Q}_n = \mathbf{I}$ and the PSFLAF turns into a SFLAF model [9].

In this paper, the choice of the diagonal elements of (6) derives directly from the *improved proportionate normalized least mean square* (IPNLMS) algorithm [13], but any other proportionate algorithm may be used. Based on the IPNLMS, the diagonal elements of $\mathbf{Q}_n$ can be achieved as:

$$q_l[n] = \frac{1-\rho}{2M_{\mathrm{e}}} + (1+\rho) \frac{|w_{\mathrm{FL},l}[n-1]|}{2\left\|\mathbf{w}_{\mathrm{FL},n-1}\right\|_1 + \xi} \tag{7}$$

with $l = 0, \ldots, M_{\mathrm{e}}-1$ and $-1 \leq \rho \leq 1$; and ξ is a small positive constant that avoids divisions by zero. The *proportionality factor* ρ balances the proportionality, since when its value is close to 1 a high sparsity degree is assumed, while when $\rho = -1$ the adaptation follows the NLMS rule. The proportionality factor also affects the choice of the regularization parameter δ_{PFL} in (5), since $\delta_{\mathrm{PFL}} = \delta_{\mathrm{FL}}(1-\rho)/2M_{\mathrm{e}}$, where δ_{FL} is the regularization factor that might be used for a FLAF.

3. THE COMBINED PROPORTIONATE SFLAF

The SFLAF is the simplest nonlinear architecture based on functional links. However, more complex structures can be thought. Here, we propose a variant of the SFLAF, in which we exploit the combination of adaptive filters to improve the nonlinear modeling performance. The new architecture, denoted as *combined PSFLAF* (cPSFLAF), is depicted in Fig. 2, where it is possible to notice that, similarly to the SFLAF in [9], the overall output signal results from the sum of the outputs of the linear and nonlinear branches. However, the nonlinear output is characterized by the adaptive combination between the two adaptive filters downstream of the FEB.

Such architecture allows to improve the nonlinear modeling performance by exploiting the different properties of the adaptive filters involved in the combination. This can be performed by choosing different adaptation rules for $\mathbf{w}_{\text{FL}1,n}$ and $\mathbf{w}_{\text{FL}2,n}$, or the same algorithm but using different parameter settings, e.g., different step sizes. However, the cPSFLAF is also well-suited to exploit the sparse representations of functional links, and this can be performed by distinguish parameters related to the proportionate algorithms, as we are going to show in the following.

The linear path of the cPSFLAF is processed as in the SFLAF, while the nonlinear path follows the same procedure of the SFLAF until the generation of the expanded buffer $\mathbf{g}_n$. This signal is then fed into both the adaptive filters, thus generating the individual outputs and errors, for $i = 1, 2$:

$$y_{\text{FL}i}[n] = \mathbf{g}_n^T \mathbf{w}_{\text{FL}i,n-1} \quad (8)$$

$$e_{\text{FL}i}[n] = d[n] - (y_{\text{L}}[n] + y_{\text{FL}i}[n]), \quad (9)$$

where error signals (9) have been obtained following a similar procedure than in case of the combination of kernels scheme [3]. These error signals are used to adapt the filters $\mathbf{w}_{\text{FL}i,n-1}$:

$$\mathbf{w}_{\text{FL}i,n} = \mathbf{w}_{\text{FL}i,n-1} + \mu_{\text{FL}i} \frac{\mathbf{Q}_{i,n}\mathbf{g}_n}{\mathbf{g}_n^T \mathbf{Q}_{i,n}\mathbf{g}_n + \delta_{\text{PFL}i}} e_{\text{FL}i}[n] \quad (10)$$

for $i = 1, 2$, where the diagonal elements of $\mathbf{Q}_{i,n}$ are computed similarly to (7), but considering that is possible to choose different parameter values for the two filters. In this way, different compromises regarding the adaptation of the nonlinear filters can be alleviated by means of combined schemes. In particular, related to the sparsity property, we could distinguish ρ_i, for $i = 1, 2$, which consequently yields $q_{i,l}[n]$ and $\delta_{\text{PFL}i}$. By choosing different values for the proportionality factor, it is possible to exploit any sparse representation of the functional link, increasing the robustness of the scheme with respect to the compromise imposed by the selection of parameter ρ. In fact, if we choose a high sparsity degree for the first filter, i.e., ρ_1 close to 1, and a low one for the second filter, i.e., ρ_2 close to -1, it is possible to generalize the model regardless of how sparse is a representation of functional links.

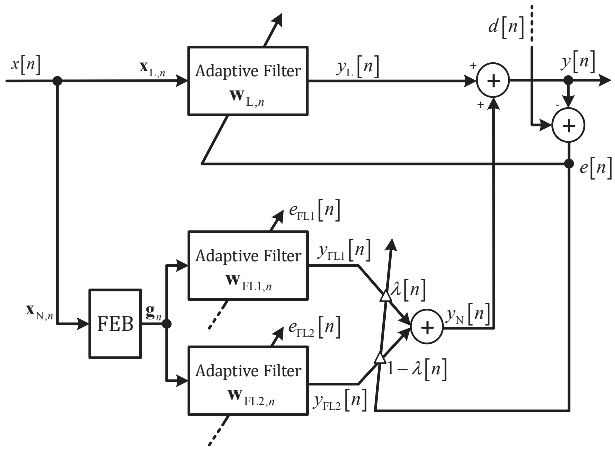

Fig. 2. The combined PSFLAF architecture.

As it is possible to see in Fig. 2, the overall output of the nonlinear branch is achieved by combining convexly the individual filter outputs (8):

$$y_{\text{N}}[n] = \lambda[n]\, y_{\text{FL}1}[n] + (1 - \lambda[n])\, y_{\text{FL}2}[n] \quad (11)$$

where $\lambda[n]$ is an adaptive *mixing parameter* that balances the combination between the two filters $\mathbf{w}_{\text{FL}i}[n]$ $(i = 1, 2)$, giving more importance between the two filters, and even, under certain circumstances, improving the behavior of both of them [11]. Such awareness is obtained according to a mean square error minimization. In particular, the adaptation of $\lambda[n]$ is performed by using an auxiliary adaptive parameter $a[n]$, which is related to $\lambda[n]$ by means of a sigmoidal function that keeps the mixing parameter in the range $[0, 1]$, and defined according to [14, 15] as:

$$\lambda[n] = \beta\left(\frac{1}{1 + e^{-a[n]}} - \alpha\right), \quad (12)$$

where $\alpha = 1/\left(1 + e^4\right)$ and $\beta = 1/(1 - 2\alpha)$. The auxiliary vector is updated by using a gradient descent rule, hence:

$$\begin{aligned} a[n] =& a[n-1] + \frac{\mu_{\text{c}}}{\beta r[n-1]} e[n]\, \Delta y[n] \\ & \cdot (\lambda[n] + \alpha\beta)(\beta - \alpha\beta - \lambda[n]) \end{aligned} \quad (13)$$

where $\Delta y[n] = y_{\text{FL}1}[n] - y_{\text{FL}2}[n]$. Also, in (13), μ_{c} is the step-size parameter of the adaptive combination, $r[n] = \gamma r[n-1] + (1 - \gamma)\Delta y^2[n]$ is the estimated power of $\Delta y[n]$ that permits a normalized adaptation of $a[n]$, and γ is a smoothing factor. The overall error signal $e[n]$, is derived as:

$$e[n] = d[n] - (y_{\text{L}}[n] + y_{\text{N}}[n]) \quad (14)$$

and it is used to update both the linear filter $\mathbf{w}_{\text{L},n}$ and the auxiliary parameter $a[n]$.

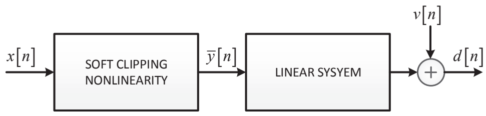

Fig. 3. Nonlinear system to be identified.

4. EXPERIMENTAL RESULTS

We assess the effectiveness of the proposed FLAF-based architecture in nonlinear system identification problems. For all the experiments, the system to be identified has a Hammerstein structure, i.e., it involves a nonlinear block followed by a linear one, as depicted in Fig. 3. The input signal $x[n]$ is distorted by a soft-clipping nonlinearity [10]:

$$\overline{y}[n] = \begin{cases} \frac{2}{3\zeta} x[n] & , \quad 0 \le |x[n]| \le \zeta \\ \text{sign}(x[n]) \frac{3-(2-|x[n]|/\zeta)^2}{3} & , \quad \zeta \le |x[n]| \le 2\zeta \\ \text{sign}(x[n]) & , \quad 2\zeta \le |x[n]| \le 1 \end{cases} \tag{15}$$

where $0 < \zeta \le 0.5$ is a nonlinearity threshold. The signal $\overline{y}[n]$ is then fed into a linear system, formed by $M = 10$ independent random values between -1 and 1. An independent and identically distributed (i.i.d.) noise signal $v[n]$ is added at output of the whole plant, in order to provide 20 dB of signal-to-noise ratio (SNR). The input signal $x[n]$ is generated by means of 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 i.i.d. Gaussian random process. In order to produce a change in the nonlinearity, we choose a clipping threshold $\zeta = 0.25$ for the first half of the experiment, while we set $\zeta = 0.05$ for the second half. In this way, the first part of the experiment is characterized by a mild nonlinearity, which becomes very hard in the second half. Performance is evaluated in terms of the *excess mean-square error* (EMSE), expressed in dB as $\text{EMSE}[n] = 10\log_{10}\left(\text{E}\left\{(e[n] - v[n])^2\right\}\right)$, which is averaged over 100 runs with respect to input and noise. Moreover, in order to facilitate the visualization, curves are smoothed by a moving-average filter. The base parameter setting for the FLAF-based architecture is: $\mu_{\text{L}}[n] = 0.1$, $\delta_{\text{L}} = \delta_{\text{FL}} = 10^{-2}$, $P = 10$, $\mu_{\text{c}} = 0.5$ and $\gamma = 0.9$. Memoryless trigonometric expansions are chosen for the FEB.

There are several strategies to improve the nonlinear modeling for the cPSFLAF. For the first experiment set, we choose one of the most promising strategies deriving from the classic theory of adaptive combination of filters, which involves the use of two filters only differing in the step-size values [11]. In particular, we set $\mu_{\text{FL1}} = 0.05$, $\mu_{\text{FL2}} = 0.5$, while a common intermediate value for the proportionality factor is chosen: $\rho_1 = \rho_2 = 0$. The overall length of this experiment is $L = 80000$ samples. Results in terms of the EMSE are depicted in Fig. 4(a), where it is possible to see that the combined PSFLAF takes advantage of the convergence performance of the PSFLAF with the higher step-size value (i.e., μ_{FL2}), while the higher precision of the PSFLAF with μ_{FL1} is exploited at steady state. This behavior occurs in both the cases of mild and strong nonlinearity, as it is also shown by the mixing parameter evolution in Fig. 4(b) that is close to 0 at transient state, i.e., it selects the PSFLAF with μ_{FL2} according to (11), while approaching 1 at steady state, i.e., when the best performance is provided by the PSFLAF with μ_{FL1}.

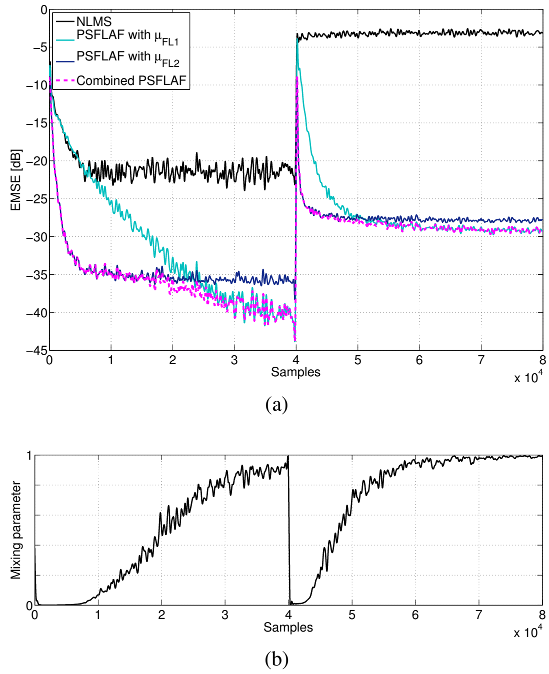

Fig. 4. Evaluation of the nonlinear modeling by using the cPSFLAF with different step-size values (a) and its mixing parameter evolution (b).

Another combination strategy can be chosen for the combination of proportionate filters, and it involves two filters having different proportionality factors, in order to design a system robust with respect to the degree of sparsity of the nonlinear path [12]. In particular, we choose $\rho_1 = 0.9$, $\rho_2 = -0.9$ in order to specialize the architecture regardless of any sparsity degree. The step-size parameter of both the filters on the nonlinear path is chosen as: $\mu_{\text{FL1}} = \mu_{\text{FL2}} = 0.1$. In Fig. 5, it is possible to see the behavior of the cPSFLAF. Unlike the previous case, in which a different choice of the step-size parameter affects both the transient and the steady states, the choice of different proportionality factors has implications mainly on the convergence rate. This is the reason why we choose a shorter length for this experiment, i.e., $L = 40000$. In the first half of the experiment, when the nonlinearity is mild (i.e., $\zeta = 0.25$), there is a high degree of sparsity, since the functional links involved are overdimensioned for the exact modeling of the nonlinear system. In fact, as it is possible to see in Fig. 5, the PSFLAF with ρ_2 (that is the one with the proportionality factor close to -1) achieves the worst convergence performance, even slower than a linear NLMS. This is

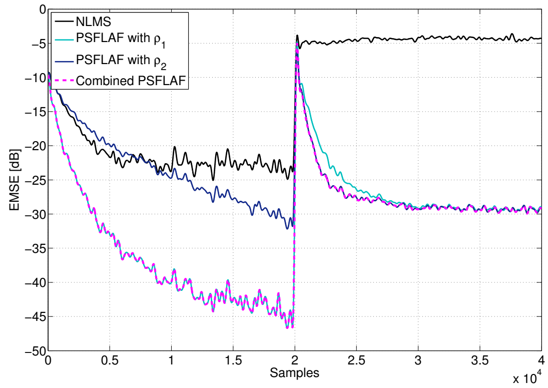

Fig. 5. Evaluation of the nonlinear modeling by using the cPSFLAF with different proportionality factors.

the reason why the cPSFLAF follows the behavior of the best performing filter, which is the PSFLAF with ρ_1. On the other hand, in the second half of the experiment, the nonlinearity level is very strong, therefore a high portion of $\mathbf{g}_n$ is successfully used for the modeling, thus reducing any sparsity of the functional link representation. In this case, the cPSFLAF follows the convergence behavior of the PSFLAF with ρ_2 that has the lower proportionality factor. It should be noted that under computational restrictions, we could select $\rho_2 = -1$, giving rise to an NLMS adaptation instead of an IPNLMS adaptive scheme. With this operation, a similar behavior is expected but with reduced computational cost.

5. CONCLUSIONS

In this paper, a new nonlinear filtering architecture has been proposed that takes advantage of any sparse functional link representation. The proposed model is characterized by an adaptive combination of two proportionate filters on the nonlinear path that yields an improvement of the overall modeling performance. Future research will include the combination of filters also on the linear path and the development of hierarchical schemes that would further generalize this model.

REFERENCES

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

[2] G. L. Sicuranza and A. Carini, "On a class of nonlinear filters," in *Festschrift in Honor of Jaakko Astola on the Occasion of his 60th Birthday*, M. G. I. Tabus and K. Egiazarian, Eds., 2009, vol. TICSP, pp. 115–144.

[3] L. A. Azpicueta-Ruiz, M. Zeller, A. R. Figueiras-Vidal, J. Arenas-García, and W. Kellermann, "Adaptive combination of Volterra kernels and its application to nonlinear acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 1, pp. 97–110, Jan. 2011.

[4] A. Carini and G. L. Sicuranza, "Recursive even mirror Fourier nonlinear filters and simplified structures," *IEEE Trans. Signal Process.*, vol. 62, no. 24, pp. 6534–6544, 2014.

[5] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini, "Hammerstein uniform cubic spline adaptive filters: Learning and convergence properties," *Signal Process.*, vol. 100, pp. 112–123, July 2014.

[6] S. Scardapane, D. Comminiello, M. Scarpiniti, and A. Uncini, "Online sequential extreme learning machine with kernels," *IEEE Trans. Neural Netw., Learn. Syst.*, 2015, In Press. DOI: 10.1109/TNNLS.2014.2382094.

[7] A. Carini, S. Cecchi, S. Romoli, and G. L. Sicuranza, "Legendre nonlinear filters," *Signal Process.*, vol. 109, pp. 84–94, Apr. 2015.

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

[9] D. Comminiello, M. Scarpiniti, L. A. Azpicueta-Ruiz, J. Arenas-García, and A. Uncini, "Functional link adaptive filters for nonlinear acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 7, pp. 1502–1512, July 2013.

[10] D. Comminiello, M. Scarpiniti, L. A. Azpicueta-Ruiz, J. Arenas-García, and A. Uncini, "Nonlinear acoustic echo cancellation based on sparse functional link representations," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 7, no. 22, pp. 1172–1183, July 2014.

[11] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1078–1090, Mar. 2006.

[12] J. Arenas-García and A. R. Figueiras-Vidal, "Adaptive combination of proportionate filters for sparse echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1087–1098, Aug. 2009.

[13] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2002, vol. 2, pp. 1881–1884.

[14] M. Lázaro-Gredilla, L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-García, "Adaptively biasing the weights of adaptive filters," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3890–3895, July 2010.

[15] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini, "Combined adaptive beamforming schemes for nonstationary interfering noise reduction," *Signal Process.*, vol. 93, no. 12, pp. 3306–3318, Dec. 2013.