Sparse Functional Link Adaptive Filter Using an ℓ_1-Norm Regularization

Danilo Comminiello, Michele Scarpiniti, Simone Scardapane, and Aurelio Uncini
Department of Information Engineering, Electronics and Telecommunications (DIET)
"Sapienza" University of Rome, Via Eudossiana 18, 00184 Rome, Italy
Email: danilo.comminiello@uniroma1.it

***Abstract*—Linear-in-the-parameters nonlinear adaptive filters often show some sparse behavior due to the fact that not all the coefficients are equally useful for the modeling of any nonlinearity. Recently, proportionate algorithms have been proposed to leverage sparsity behaviors in nonlinear filtering. In this paper, we deal with this problem by introducing a proportionate adaptive algorithm based on an ℓ_1-norm penalty of the cost function, which regularizes the solution, to be used for a class of nonlinear filters based on functional links. The proposed algorithm stresses the difference between useful and useless functional links for the purpose of nonlinear modeling. Experimental results clearly show faster convergence performance with respect to the standard (i.e., non-regularized) version of the algorithm.**

I. Introduction

In system identification problems, the impulse response of an unknown system may show some sparsity behavior. This often occurs when a small part of the system response is characterized by a large magnitude, such that the coefficients of the rest of the response show negligible values. In terms of energy, this means that a system is sparse when most of its energy is contained in a small part of it [1]. In several applications sparsity is often exploited by introducing a regularizing penalty that leads to an improvement of the overall performance of a model. Some examples can be recently found in compressive sensing [2], [3], ultrasonic beamforming [4], dynamic power network models [4], cognitive radio [5], Alzheimer's disease diagnosis [6], brain dopamine recording [7], underwater communication [8], among others.

Sparsity may also characterize nonlinear systems. In particular, in system identification problems, nonlinear expansions may introduce useless coefficients for the purpose of nonlinear modeling, thus leading to overfitting phenomena and performance degradation [9]. In order to address this problem, it would be desirable to select only the most significant nonlinear elements, i.e., the ones which provide the best modeling performance. However, when a nonlinearity is dynamical, nonstationary or depending on a time-varying signal, it is impossible to select *a priori* the most significant nonlinear coefficients. To this end, sparse adaptive algorithms can be used to detect and select the most useful coefficients in an *online* fashion [10]. In the literature, sparse nonlinear algorithms have been proposed, including kernel-based methods [11] and polynomial methods [12], among others.

In the recent years, sparse algorithms have been developed also for the class of nonlinear filters known as functional link adaptive filters (FLAFs) [9], [10], [13]–[15]. Functional link-models are widespread in the literature due to their flexibility and the large range of application in which they can be used [16]–[21]. They are basically characterized by a nonlinear transformation of the input by any nonlinear series expansion. Very often, the number of functional links may be rather high, thus causing any sparseness degree, which can be exploited by imposing any sparsity constraint. In that sense, ℓ_2-norm regularization has been mainly considered to develop sparse FLAFs achieving outstanding results [9], [10], [13]. In this paper, we investigate on the use of an ℓ_1-norm constraint to develop a new sparse functional link filter.

The ℓ_1-norm penalty has been widely used in the literature, since the introduction of the LASSO algorithm [22], due to the its desirable properties of inducing sparsity while keeping the convexity of the cost function. One of the most successfully used class of algorithms showing an ℓ_1 relaxation is based on zero-attracting proportionate adaptation [23]–[26]. In this work, we extend these results by deriving a class of sparse functional link adaptive filters based on a joint optimization problem [15] involving an ℓ_1-norm penalty. Experimental results show the effectiveness of the resulting zero-attracting (ZA) based FLAFs in nonlinear system identification problems.

The paper is organized as follows. Section II briefly reviews the FLAF model for nonlinear system modeling, while the proposed family of ZA-based FLAFs is introduced in Section III, in which we formulate the ℓ_1-norm constrained optimization problem for the FLAF model and we introduce some significant variations of the ZA-based algorithms. Experimental results are shown in Section IV to prove the capabilities of the proposed algorithms in leveraging sparse behaviors of functional links. Concluding remarks are drawn in Section V.

II. Nonlinear Modeling with Sparse FLAF

A. A Review of the FLAF Model

The functional link adaptive filter (FLAF) model [17] is a linear-in-the-parameters nonlinear filter that nonlinearly expands a linear input signal $\mathbf{x}_{\mathrm{L},n} \in \mathbb{R}^M = \left[\ x[n] \quad x[n-1] \quad \ldots \quad x[n-M+1]\ \right]^{\mathrm{T}}$, being M the length of the regression vector, in order to filter it in a higher dimensional space. The main component of the FLAF is a functional expansion block (FEB), which has the role of

978-1-5386-4881-0/18/$31.00 ©2018 IEEE

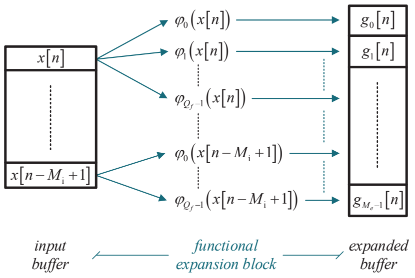

Fig. 1. Memoryless functional link expansion.

applying a chosen expansion series to the input samples in order to produce the nonlinear signal. The nonlinear expansion functions of the FEB are known as functional links, whose set can be denoted as $\Phi = \left\{\varphi_0(\cdot), \ldots, \varphi_{Q_\mathrm{f}-1}(\cdot)\right\}$, where Q_f is the number of the chosen functional links. One of the most popular choice for the set of functional links is to use trigonometric series expansion, which can be described as:

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

for $i = 0, \ldots, M-1$. In (1), $p = 1, \ldots, P$ is the expansion index, where P is the expansion order, and $j = 0, \ldots, Q_\mathrm{f} - 1$ is the functional link index. For trigonometric memoryless expansion, the overall number of functional links contained in the set Φ is equal to $Q_\mathrm{f} = 2P$. The functional link expansion of the input signal is represented in Fig. 1.

The overall expanded vector resulting from the application of the functional link set to the input signal is denoted as $\mathbf{g}_n \in \mathbb{R}^{M_\mathrm{e}} = \begin{bmatrix} g_0[n] & g_1[n] & \ldots & g_{M_\mathrm{e}-1}[n] \end{bmatrix}^\mathrm{T}$, where $M_\mathrm{e} \geq M$ is the length of the expanded vector. Then, the signal $\mathbf{g}_n$, which represent the input in a higher dimensional space, can be processed by any adaptive filter to achieve the system output.

The functional link set Φ may contain both linear and nonlinear functions. However, we assume that all the functional links are nonlinear, such that the resulting expanded vector $\mathbf{g}_n$ is completely composed of nonlinear elements. This allows to increase the flexibility in those problems which require the modeling of linear and nonlinear components, since we can devote one adaptive filter for the non-expanded linear input $\mathbf{x}_{\mathrm{L},n}$ and one adaptive filter for $\mathbf{g}_n$, and we can choose two different learning algorithms for the two filters. This method is described in [17] and called split FLAF (or SFLAF). Therefore, the output of the FLAF model can be obtained by the sum of the linear and nonlinear contributions, i.e.:

$$y[n] = y_\mathrm{L}[n] + y_\mathrm{FL}[n]. \quad (2)$$

In particular, in (2), $y_\mathrm{L}[n] = \mathbf{x}_{\mathrm{L},n}^\mathrm{T}\mathbf{w}_{\mathrm{L},n-1}$, being $\mathbf{w}_{\mathrm{L},n} \in \mathbb{R}^M = \begin{bmatrix} w_{\mathrm{L},0}[n] & \ldots & w_{\mathrm{L},M-1}[n] \end{bmatrix}^\mathrm{T}$ is the coefficient vector for the linear input samples, and $y_\mathrm{FL}[n] = \mathbf{g}_n^\mathrm{T}\mathbf{w}_{\mathrm{FL},n-1}$, being $\mathbf{w}_{\mathrm{FL},n} \in \mathbb{R}^{M_\mathrm{e}} = \begin{bmatrix} w_{\mathrm{FL},0}[n] & \ldots & w_{\mathrm{FL},M_\mathrm{e}-1}[n] \end{bmatrix}^\mathrm{T}$ is the coefficient vector for the nonlinear expanded samples. In this scheme, both the adaptive filters $\mathbf{w}_{\mathrm{L},n}$ and $\mathbf{w}_{\mathrm{FL},n}$ are updated by using the same error signal $e[n] = d[n] - y[n]$, where $d[n]$ is the nonlinear reference signal.

B. Sparsity in FLAF

It is very clear from the literature how sparsity may affect the coefficients of a linear filter, therefore we omit any analysis of the sparsity for the coefficient vector $\mathbf{x}_{\mathrm{L},n}$. However, a major attention is required for a linear-in-the-parameter nonlinear filter, as in the case of $\mathbf{w}_{\mathrm{FL},n}$.

In order to understand how sparsity behaviors occur in FLAFs, it is sufficient to analyze the energy of the coefficient vector $\mathbf{w}_{\mathrm{FL},n}$ at steady state [9], i.e., for $n \to \infty$. As a result, the early functional links of the set Φ, which have small values of the expansion order (i.e., p close to 1), generate the most significant nonlinear elements for the purpose of the nonlinear modeling. On the other hand, the remaining functional links of Φ produce only minor, or even negligible, variations in the modeling results. This is physically motivated by the fact that high-order functional links (i.e., with $p \to P$) aim at modeling those nonlinear components that not always appear in a distortion (e.g., high-order harmonics), thus resulting in a slight improvement that might not always occur. In terms of energy, it is possible to describe the sparsity behavior in functional links as an exponential decay from early to late elements, in which the larger the expansion order the longer the tail. Such behavior occurs for each input sample that is expanded by the functional link set Φ, therefore, overall, we may have a periodic sparse behavior for $\mathbf{w}_{\mathrm{FL},\infty}$.

III. Zero-Attracting Algorithms for FLAF

In this section, we derive the sparse FLAFs using an ℓ_1-norm regularization.

A. Derivation of the ℓ_1-Regularized FLAF

In order to derive the optimization algorithms for the sparse functional links, we start considering a full split FLAF scheme, as introduced in [15]. To this end, we define the overall input and coefficient vectors, both in $\mathbb{R}^{M+M_\mathrm{e}}$:

$$\mathbf{x}_n = \begin{bmatrix} \mathbf{x}_{\mathrm{L},n} \\ \mathbf{g}_n \end{bmatrix}, \qquad \mathbf{w}_n = \begin{bmatrix} \mathbf{w}_{\mathrm{L},n} \\ \mathbf{w}_{\mathrm{FL},n} \end{bmatrix}. \quad (3)$$

from which we can derive the output signal as $y[n] = \mathbf{x}_n^\mathrm{T}\mathbf{w}_{n-1}$, which is equivalent to (2) at this point.

We can define the difference weight vector as $\widetilde{\mathbf{w}}_n = \mathbf{w}_n - \mathbf{w}_{n-1}$. Considering the least-perturbation property and the natural gradient adaptation, as suggested in [13], we can express the ℓ_1-constrained optimization problem as:

$$\begin{aligned} &\arg\min_{\mathbf{w}_n} \|\widetilde{\mathbf{w}}_n\|^2_{\mathbf{Q}_n^{-1}} + \gamma \left\|\mathbf{Q}_n^{-1}\mathbf{w}_n\right\|_1 \\ &\text{s.t.} \quad \varepsilon[n] = 0 \end{aligned} \quad (4)$$

where $\mathbf{Q}_n^{-1}$ is a distance correction matrix with respect to the Euclidean metric, γ is a very small constant and the constraint $\varepsilon[n] = 0$ can be derived from the *a posteriori* output estimation error signal $\varepsilon[n] = d[n] - \mathbf{x}_n^\mathrm{T}\mathbf{w}_n$. It is worth noting that the problem in (4) involves an ℓ_1-norm penalty term that aims at scaling the coefficient vector $\mathbf{w}_n$ by the distance correction matrix $\mathbf{Q}_n^{-1}$ and thus regularizing the solution by mainly exploiting inactive coefficients.

The problem (4) can be solved by deriving the function:

$$\mathcal{L}[n] = \widetilde{\mathbf{w}}_n^{\mathrm{T}} \mathbf{Q}_n^{-1} \widetilde{\mathbf{w}}_n + \gamma \left\| \mathbf{Q}_n^{-1} \mathbf{w}_n \right\|_1 + \lambda \left(d[n] - \mathbf{x}_n^{\mathrm{T}} \mathbf{w}_n \right) \quad (5)$$

where λ is a Lagrange multiplier. We compute the gradient of (5) with respect to $\mathbf{w}_n$ and λ, and solve it yielding:

$$\mathbf{w}_n - \mathbf{w}_{n-1} = \frac{1}{2} \left(\lambda \mathbf{Q}_n \mathbf{x}_n - \gamma \operatorname{sgn}(\mathbf{w}_n) \right). \quad (6)$$

where $\operatorname{sgn}(\cdot)$ is a component-wise sign function defined for the i-th entry of $\mathbf{w}_n$ as:

$$\operatorname{sgn}(w_i[n]) = \begin{cases} w_i[n]/|w_i[n]|, & w_i[n] \neq 0 \\ 0, & w_i[n] = 0 \end{cases} \quad (7)$$

We left-multiply both sides of (6) by $\mathbf{x}_n^{\mathrm{T}}$ and take into account that, due to the constraint in (4), $\mathbf{x}_n^{\mathrm{T}} \mathbf{w}_n = d[n]$, from which $d[n] - \mathbf{x}_n^{\mathrm{T}} \mathbf{w}_{n-1} = e[n]$. Therefore, we can solve for the Lagrange multiplier λ as:

$$\lambda = \frac{2e[n] + \gamma \mathbf{x}_n^{\mathrm{T}} \operatorname{sgn}(\mathbf{w}_n)}{\mathbf{x}_n^{\mathrm{T}} \mathbf{Q}_n \mathbf{x}_n}. \quad (8)$$

We replace the Lagrange multiplier (8) in (6), thus obtaining:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \frac{\mathbf{Q}_n \mathbf{x}_n e[n]}{\mathbf{x}_n^{\mathrm{T}} \mathbf{Q}_n \mathbf{x}_n} - \gamma \left(\mathbf{I} - \frac{\mathbf{x}_n \mathbf{x}_n^{\mathrm{T}} \mathbf{Q}_n}{\mathbf{x}_n^{\mathrm{T}} \mathbf{Q}_n \mathbf{x}_n} \right) \operatorname{sgn}(\mathbf{w}_n), \quad (9)$$

where $\mathbf{I}$ is the identity matrix.

The above equation can be refined by taking some considerations. First, we can assume that most of the weights do not change their sign at each iteration, especially those weights that are inactive, i.e, that are close to zero at steady state. Therefore, in (9), the term $\operatorname{sgn}(\mathbf{w}_n)$, which contains the unknown new update, can be approximated by $\operatorname{sgn}(\mathbf{w}_{n-1})$ without affecting the convergence behavior [23], [26]. Moreover, analyzing the third term of (9), it is possible to note that the matrix subtracted to the identity matrix $\mathbf{I}$ has very small values with respect to $\mathbf{I}$, thus being negligible. Finally, in (9), we can introduce two parameters to ensure a correct adaptation: a regularization factor δ and a diagonal matrix of the step-size parameters $\mathbf{\Lambda}$. Hence, the update equation of the ℓ_1-constrained zero-attracting ZA FLAF can be written as:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mathbf{\Lambda} \frac{\mathbf{Q}_n \mathbf{x}_n e[n]}{\mathbf{x}_n^{\mathrm{T}} \mathbf{Q}_n \mathbf{x}_n + \delta} - \mathbf{\Gamma} \operatorname{sgn}(\mathbf{w}_{n-1}), \quad (10)$$

where $\mathbf{\Gamma} = \gamma \mathbf{\Lambda}$. The elements of $\mathbf{\Lambda}$ can be chosen as $\mu_k = \mu_{\mathrm{L}}$, for $k = 0, \ldots, M-1$, and $\mu_k = \mu_{\mathrm{FL}}$, for $k = M, \ldots, M + M_{\mathrm{e}} - 1$, in order to preserve the flexibility of the split scheme in choosing different parameter settings for the coefficients related to linear and nonlinear input samples [15], [17].

B. A Variable Step-Size Reweighted ZA FLAF

With respect to ℓ_2-norm based proportionate FLAFs [9], [15], the FLAF based on the ZA algorithm involves a further term, i.e., the third one of (10), which is called zero attractor since it shrinks the inactive weights of $\mathbf{w}_n$ to zero. However, such term may lose its effectiveness as the sparsity of a system decreases, i.e., the number of active coefficients increases. To overcome this limitations, a robust version of the ZA algorithm has been proposed according to the process of reweighting in compressive sampling [23], [27], in which the ℓ_1 regularization term $\left\| \mathbf{Q}_n^{-1} \mathbf{w}_n \right\|_1$ in (4) is replaced by a log-sum penalty term $\left\| \mathbf{Q}_n^{-1} \log(1 + \epsilon \mathbf{w}_n) \right\|_1$, where ϵ is a small constant. The new penalty term is more similar to an ℓ_0-norm and it leads to the update equation of the reweighted ZA (RZA) FLAF:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mathbf{\Lambda} \frac{\mathbf{Q}_n \mathbf{x}_n e[n]}{\mathbf{x}_n^{\mathrm{T}} \mathbf{Q}_n \mathbf{x}_n + \delta} - \mathbf{\Gamma}_{\mathrm{R}} \frac{\operatorname{sgn}(\mathbf{w}_{n-1})}{1 + \epsilon |\mathbf{w}_{n-1}|}, \quad (11)$$

where $\mathbf{\Gamma}_{\mathrm{R}} = \epsilon \gamma \mathbf{\Lambda}$. The reweighted zero attractor, i.e., the third term of (11), aims at shrinking to zero those coefficients whose magnitude is comparable to $1/\epsilon$, thus preserving the most active coefficients. This usually results in an improvement in terms of convergence rate and steady state performance.

A further improvement of the ZA algorithm can be made by considering a variable step size (VSS) [26]. The idea is to relate the step sizes directly to the squared error behavior in order to reduce the values of the step sizes as the error decreases, thus leading to an improvement of the convergence performance. We adopt a nonparametric VSS [28] that was proved to be robust in several cases [26], [28]–[30] and it can be expressed as:

$$\mu[n] = \left| 1 - \frac{\sqrt{\left| \widehat{\sigma}_d^2[n] - \widehat{\sigma}_y^2[n] \right|}}{\widehat{\sigma}_e^2[n] + \xi} \right|. \quad (12)$$

In (12), the general parameter $\widehat{\sigma}_\theta^2[n]$ represents the power estimate of the sequence $\theta[n]$, being $\theta = \{d, y, e\}$, and it can be computed as $\widehat{\sigma}_\theta^2[n] = \beta \widehat{\sigma}_\theta^2[n-1] + (1-\beta)\theta^2[n]$ where $\beta \to 1$ is a forgetting factor. ξ ia s small positive constant that avoids divisions by zero. The VSS value (12) is applied to (11) to obtain the VZA FLAF:

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mathbf{\Lambda} \frac{\mathbf{Q}_n \mathbf{x}_n e[n]}{\mathbf{x}_n^{\mathrm{T}} \mathbf{Q}_n \mathbf{x}_n + \delta} - \gamma_{\mathrm{R}}[n] \frac{\operatorname{sgn}(\mathbf{w}_{n-1})}{1 + \epsilon |\mathbf{w}_{n-1}|}, \quad (13)$$

where $\gamma_{\mathrm{R}}[n] = \epsilon \gamma \mu[n]$. It is worth noting that in this case we lose the flexibility of the split scheme in having two different values for the fixed step sizes, but this is compensated by the intrinsic flexibility of the VSS.

C. Choice of the weighting matrix

The matrix $\mathbf{Q}_n$ in (10), also known as proportionate matrix, aims at weighting the coefficients of $\mathbf{w}_n$ proportionally to the contribution they provide to the modeling. $\mathbf{Q}_n$ can be chosen as a diagonal matrix $\mathbf{Q}_n \in \mathbb{R}^{M+M_{\mathrm{e}}} = \operatorname{diag}\left\{ q_0[n] \;\; \ldots \;\; q_{M+M_{\mathrm{e}}-1}[n] \right\}$, whose diagonal elements are derived according to the coefficients at the time instant $n-1$.

One of the most popular choice for the derivation of the diagonal elements of $\mathbf{Q}_n$ follows the proportionate rule [26], [31], which can be defined for ZA(a) FLAF schemes as:

$$q_k[n] = \frac{l_k[n]}{\sum_{i=a}^{b} l_k[n]} \quad (14)$$

for $k = 0, \ldots, M + M_{\mathrm{e}} - 1$, with

$$l_k[n] = \max\left[\rho \max\left[\chi, |w_a[n-1]|, \ldots, |w_b[n-1]| \right], |w_k[n-1]| \right] \quad (15)$$

where $a = 0$, $b = M - 1$ for $k < M$, and $a = M$, $b = M + M_e - 1$ for $k \geq M$. In (15), ρ and χ are constants [31].

Another popular choice is provided by the improved proportionate rule [32], whose application to ZA(b) FLAF schemes can be expressed as [15]:

$$q_k[n] = \begin{cases} \frac{1-\alpha_{\text{L}}}{2M} + (1+\alpha_{\text{L}}) \frac{|w_k[n-1]|}{\xi + 2\sum_{i=a}^{b} w_k[n-1]}, & k < M \\ \frac{1-\alpha_{\text{FL}}}{2M_e} + (1+\alpha_{\text{FL}}) \frac{|w_k[n-1]|}{\xi + 2\sum_{i=a}^{b} w_k[n-1]}, & k \geq M \end{cases} \tag{16}$$

where the scalar ξ is a small positive value. In (16), the proportionality factors $-1 \leq \alpha_{\text{L}}, \alpha_{\text{FL}} \leq 1$ assume a value close to 1 when a high degree of sparseness is expected, while, on the contrary, a low degree is expected when the proportionality factors are close to -1.

IV. Simulation Results

We evaluate the performance of the proposed algorithms and we show their capability in exploiting sparsity. We consider an unknown system to be identified composed by a nonlinear block and a linear one in cascade. The former block applies a soft-clipping nonlinearity on the input signal [13]:

$$\overline{y}[n] = \begin{cases} 2x[n]/3\zeta & , \quad 0 \leq |x[n]| \leq \zeta \\ \frac{3-(2-|x[n]|/\zeta)^2}{3} \operatorname{sgn}(x[n]) & , \quad \zeta \leq |x[n]| \leq 2\zeta \\ \operatorname{sgn}(x[n]) & , \quad 2\zeta \leq |x[n]| \leq 1 \end{cases} \tag{17}$$

where $0 < \zeta \leq 0.5$ is a nonlinearity threshold. The linear system is formed by $M = 10$ independent random values between -1 and 1. A white Gaussian noise signal $v[n]$ is added at output of the system, providing 30 dB of signal-to-noise ratio (SNR). The input signal $x[n]$ derives from an independent and identically distributed (i.i.d.) Gaussian random process, with length $L = 45000$, filtered by a first-order autoregressive model having transfer function $\sqrt{1-\theta^2}/\left(1-\theta z^{-1}\right)$, with $\theta = 0.8$. The *excess mean-square error* (EMSE) is used to evaluate performance and it is 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 1000 runs with respect to input and noise. We set the parameters as follows: $\mu_{\text{L}} = 0.1$, $\mu_{\text{FL}} = 0.1$, $\delta = 10^{-3}$, $P = 30$, $\gamma = 10^{-5}$, $\epsilon = 10^{-2}$, $\xi = 10^{-4}$, $\beta = 0.99$, $\rho = 10^{-2}$, $\chi = 10^{-3}$, $\alpha_{\text{L}} = \alpha_{\text{FL}} = 0$.

In order to assess the proposed algorithms in different sparsity conditions and to further assess their tracking abilities, we evaluate the EMSE in a changing environment, in which for two times we modify the sparseness degree of both the linear and nonlinear subsystems. In particular, in the three parts of the experiment we choose three different values for the number of null coefficients in the linear impulse response and for the nonlinearity threshold in order to have different sparsity behaviors of both the linear coefficients and the functional links [9]. We set respectively $8, 6, 3$ zero coefficients, randomly chosen over M, for the linear response, and we set the nonlinearity threshold ζ respectively to the values $0.15, 0.22, 0.35$.

Results in terms of the EMSE are depicted in Fig. 2, where it is possible to see that all the zero-attracting based FLAFs achieve a large improvement over the standard non-constrained FLAF. In particular, it is possible to note the progressive improvements of the ZA algorithms in terms of the convergence rate due to the reweighting (i.e., RZA(a) and RZA(b) FLAFs), and in terms of the steady-state precision due to the introduction of the variable step size (i.e., for the VZA(a) and VZA(b) FLAFs). For all the algorithms, it is worth noting that the use of the improved proportionate rule of (16) for the computation of the diagonal elements of $\mathbf{Q}_n$ leads to better performance with respect to the same ZA-based FLAFs using the proportionate rule of (14). In particular, it is worth noting the difference between the two rules increases as the sparseness degree of the overall system decreases, i.e., in the second and even more in the third parts of the experiment.

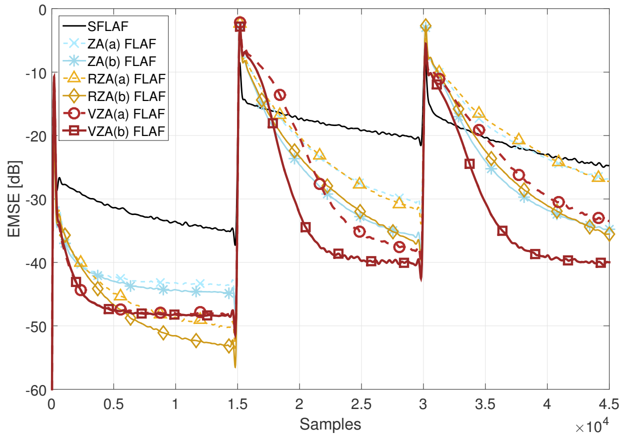

Fig. 2. Performance results in terms of the ERLE with varying sparseness degree (from high to low).

V. Conclusion

In this paper, new functional link adaptive filtering algorithms have been developed based on an ℓ_1-norm regularization that exploits any sparsity in both the linear filter and of the linear-in-the-parameter nonlinear filter involved in the FLAF. The resulting zero-attracting based FLAF algorithm, and its robust versions have been evaluated in nonstationary environment with different sparseness degrees, in which they have shown improved nonlinear modeling performance. Future research will focus on leveraging the different positive properties of the proposed algorithms (e.g, convergence rate and steady-state precision) by using adaptive combined architectures involving regularized FLAFs with different ZA algorithms (e.g., RZA and VSS based FLAFs) or also with different penalty constraints (e.g., ℓ_1 and ℓ_2 norms).

Acknowledgment

This work is partially supported by the Italian National Project "GAUChO - A Green Adaptive Fog Computing and Networking Architecture", under grant number 2015YPXH4W.

978-1-5386-4881-0/18/$31.00 ©2018 IEEE

References

[1] Y. Huang, J. Benesty, and J. Chen, Eds., *Acoustic MIMO Signal Processing*, Signals and Communication Technology. Springer-Verlag, Berlin, Heildelberg, 2006.

[2] D. E. Bellasi and L. Benini, "Energy-efficient analysis of analog and digital compressive sensing in wireless sesnors," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 62, no. 11, pp. 2718–2729, Nov. 2015.

[3] J. K. Pant and S. Krishnan, "Two-pass ℓp-regularized least-squares algorithm for compressive sensing," in *IEEE Int. Symp. Circuits and Syst. (ISCAS)*, Baltimore, MD, May 2017, pp. 1–4.

[4] E. Ozkan, V. Vishnevsky, and O. Goksel, "Inverse problem of ultrasound beamforming with sparsity constraints and regularization," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. PP, no. 99, pp. 1–1, 2017.

[5] A. López Parrado and J. Velasco-Medina, "Cooperative wideband spectrum sensing based on sub-Nyquist sparse fast Fourier transform," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 63, no. 1, pp. 39–43, Jan. 2016.

[6] B. Lei, P. Yang, T. Wang, S. Chen, and D. Ni, "Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1102–1112, Apr. 2017.

[7] H. Zamani, H. Bahrami, P. A. Garris, and P. Mohseni, "On the use of compressive sensing (CS) for brain dopamine recording with fast-scan cyclic voltammetry (FSCV)," in *IEEE Int. Symp. Circuits and Syst. (ISCAS)*, Baltimore, MD, May 2017, pp. 1–4.

[8] E. Panayirci, H. Senol, M. Uysal, and H. V. Poor, "Sparse channel estimation and equalization for OFDM-based underwater cooperative systems with amplify-and-forward relaying," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 214–228, Jan. 2016.

[9] D. Comminiello, M. Scarpiniti, L. A. Azpicueta-Ruiz, J. Arenas-García, and A. Uncini, "Combined nonlinear filtering architectures involving sparse functional link adaptive filters," *Signal Process.*, vol. 135, pp. 168–178, June 2017.

[10] D. Comminiello, S. Scardapane, M. Scarpiniti, R. Parisi, and A. Uncini, "Online selection of functional links for nonlinear system identification," in *Advances in Neural Networks: Computational and Theoretical Issues*, S. Bassis, A. Esposito, and F. C. Morabito, Eds., vol. 37 of *Smart Innovation, Systems and Technologies*, pp. 39–47. Springer-Verlag, 2015.

[11] W. Liu, I. Park, and J. C. Príncipe, "An information theoretic approach of designing sparse kernel adaptive filters," *IEEE Trans. Neural Netw.*, vol. 20, no. 12, pp. 1950–1961, Dec. 2009.

[12] V. Kekatos and G. B. Giannakis, "Sparse Volterra and polynomial regression models: Recoverability and estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5907–5920, Dec. 2011.

[13] 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, Language Process.*, vol. 7, no. 22, pp. 1172–1183, July 2014.

[14] D. Comminiello, M. Scarpiniti, L. A. Azpicueta-Ruiz, J. Arenas-García, and A. Uncini, "A block-based combined scheme exploiting sparsity in nonlinear acoustic echo cancellation," in *IEEE Int. Workshop Mach. Learning and Signal Process. (MLSP)*, Vietri Sul Mare, Italy, 2016, pp. 1–6.

[15] D. Comminiello, M. Scarpiniti, L. A. Azpicueta-Ruiz, J. Arenas-García, and A. Uncini, "Full proportionate functional link adaptive filters for nonlinear acoustic echo cancellation," in *25th Eur. Signal Process. Conf. (EUSIPCO)*, Kos Island, Greece, Aug. 2017, pp. 1185–1189.

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

[17] 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, Language Process.*, vol. 21, no. 7, pp. 1502–1512, July 2013.

[18] V. Patel, V. Gandhi, S. Heda, and N. V. George, "Design of adaptive exponential functional link network-based nonlinear filters," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 9, pp. 1434–1442, Sept. 2016.

[19] A. Carini and D. Comminiello, "Introducing complex functional link polynomial filters," in *42nd IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, New Orleans, LA, Mar. 5-9 2017, pp. 4656–4660.

[20] Vasundhara, N. B. Puhan, and G. Panda, "De-correlated improved adaptive exponential FLAF-based nonlinear adaptive feedback cancellation for hearing aids," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. PP, pp. 1–13, 2017.

[21] D. M. Sahoo and S. Chakraverty, "Functional link neural network learning for response prediction of tall shear buildings with respect to earthquake data," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. PP, no. 99, pp. 1–10, 2017.

[22] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[23] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Taipei,Taiwan, Apr. 2009, pp. 3125–3128.

[24] R. L. Das and M. Chakraborty, "A zero attracting proportionate normalized least mean square algorithm," in *Proc. Asia Pacific Signal and Inform. Process. Assoc. Annu. Summit and Conf. (APSIPA ASC)*, Hollywood, CA, Dec. 2012, pp. 1–4.

[25] J. Chen, C. Richard, Y. Song, and D. Brie, "Transient performance analysis of zero-attracting LMS," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1786–1790, 2016.

[26] R. L. Das and M. Chakraborty, "Improving the performance of the PNLMS algorithm using l_1 norm regularization," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 7, pp. 1280–1290, July 2016.

[27] E. J. Candés, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 877–905, Dec. 2008.

[28] C. Paleologu, J. Benesty, and S. Ciochină, "A variable step-size affine projection algorithm dedesign for acoustic echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 8, pp. 1466–1478, Nov. 2008.

[29] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini, "A novel affine projection algorithm for superdirective microphone array beamforming," in *IEEE Int. Symp. Circuits and Syst. (ISCAS)*, Paris, France, May 2010, pp. 2127–2130.

[30] D. Comminiello, M. Scarpiniti, S. Scardapane, R. Parisi, and A. Uncini, "A nonlinear acoustic echo canceller with improved tracking capabilities," in *Recent Advances in Nonlinear Speech Processing*, A. Esposito, M. Faundez-Zanuy, A. M. Esposito, G. Cordasco, T. Drugman, J. Solé-Casal, and F. C. Morabito, Eds., vol. 48 of *Smart Innovation, Systems and Technologies*, pp. 235–243. Springer, 2016.

[31] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Audio, Speech, Language Process.*, vol. 8, no. 5, pp. 508–518, Sept. 2000.

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