A Low-Complexity Linear-in-the-Parameters Nonlinear Filter for Distorted Speech Signals

Danilo Comminiello, Michele Scarpiniti, Simone Scardapane, 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@uniroma1.it

Abstract. In this paper, the problem of the online modeling of nonlinear speech signals is addressed. In particular, the goal of this work is to provide a nonlinear model yielding the best tradeoff between performance results and required computational resources. Functional link adaptive filters were proved to be an effective model for this problem, providing the best performance when trigonometric expansion is used as a nonlinear transformation. Here, a different functional expansion is adopted based on the Chebyshev polynomials in order to reduce the overall computational complexity of the model, while achieving good results in terms of perceived quality of processed speech. The proposed model is assessed in the presence of nonlinearities for both simulated and real speech signals.

Keywords: Nonlinear Modeling, Functional Links, Chebyshev Polynomials, Loudspeaker Distortions, Nonlinear System Identification

1 Introduction

In the recent years, a widespread availability of commercial hands-free speech communication systems has occurred, also due to the development of immersive speech communication techniques [4, 7]. However, such devices often mount low-cost components, which may affect the quality of the perceived speech. In particular, poor-quality loudspeakers, vibrations of plastic shells, D/A converters and power amplifiers may introduce a significant amount of nonlinearity in speech signals, especially during large signal peaks.

In online learning applications related to hands-free speech communications, such as nonlinear acoustic echo cancellation (NAEC) and active noise control (ANC), linear-in-the-parameters (LIP) nonlinear filters represent an effective and flexible solution [6, 8–10, 12, 18]. However, the modeling and compensations by LIP nonlinear filters may require a large number of nonlinear elements, which involve a high computational load that may represent a problem of real-time applications like NAEC.

In order to address this problem, in this paper, we propose a LIP nonlinear filter that provides the best tradeoff between performance results and required computational resources. In particular, we take into account the nonlinear *functional link adaptive filters* (FLAFs) [6], which is based on a nonlinear expansion of the input signal by the so-called *functional links* [11, 13, 16], and an adaptive filtering of the transformed signal in cascade.

One of the most important advantages of the FLAF is its flexibility, since it is possible to set the different parameters of the FLAF individually in order to fit the model at best for a specific application. In the design of an FLAF, an important choice is the number of functional links to be adopted in the model. This choice is strictly related to the nonlinearity degree introduced by the unknown system and with the chosen type of functional expansion. Therefore, in order to reduce the computational complexity we directly aim at designing a suitable and efficient *functional expansion block* to be used for the modeling of nonlinear speech signals.

In particular, *Chebyshev functional links* are assessed within NAEC problems and compared with other classic functional expansions. Performance is evaluated in terms of both error-based criteria and speech quality measures, while considering the minimum possible computational load. Results are achieved over both simulated and real data and show the effectiveness of Chebyshev functional links to be used for a low-complexity FLAF model.

The paper is organized as follows: the FLAF-based model for the modeling of speech signals is introduced in Section 2, while Chebyshev functional links and their properties are described in Section 3. Results are discussed in Section 4 and, finally, in Section 5 our conclusions are drawn.

2 A Functional Link-Based Nonlinear Model for NAEC

The FLAF model is purely nonlinear, since the adaptive filter receives as input a transformed nonlinear signal. However, very often in acoustic speech signal processing, there is also a linear component to be modeled, as in the case of the presence of an acoustic impulse response in NAEC. To this end, we adopt a filtering scheme based on the FLAF that includes both linear and nonlinear filtering, called *split functional link adaptive filter* (SFLAF) [6].

The SFLAF architecture, depicted in Fig. 1, involves a linear branch and a nonlinear branch in parallel. The former is nothing but a linear adaptive filter totally aiming at modeling the linear components of the system to be identified. On the other hand, the nonlinear branch is a nonlinear FLAF. The output signal of the SFLAF is obtained from the sum of the outputs of the two parallel branches:

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

The error signal is then obtained as:

$$e[n] = d[n] - y[n] \tag{2}$$

where $d[n]$ is the desired signal that may includes any background noise. The error signal (2) is used to adapt both the adaptive filters. In (1), $\mathbf{x}_n \in \mathbb{R}^M =$

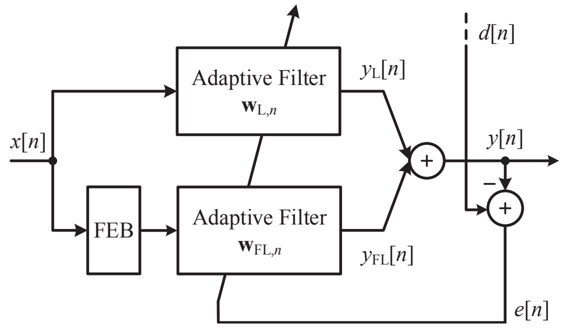

Fig. 1. The split functional link adaptive filter.

$\left[x\left[n\right]\, x\left[n-1\right] \ldots x\left[n-M+1\right]\right]^T$ is the input to the filter on the linear branch, with M being the length of the input vector. Also, in (1), the vector $\mathbf{g}_n \in \mathbb{R}^{M_\mathrm{e}} = \left[g_0\left[n\right]\, g_1\left[n\right] \ldots g_{M_\mathrm{e}-1}\left[n\right]\right]^T$ is the *expanded buffer*, i.e., the output of the *functional expansion block* (FEB) whose length is $M_\mathrm{e} \geq M_\mathrm{i}$.

Both the adaptive filters $\mathbf{w}_{\mathrm{L},n}$ and $\mathbf{w}_{\mathrm{FL},n}$ in (1) can be updated by using any linear adaptive algorithm. Here, we use a *normalized least-mean square* (NLMS) algorithm [17], so that:

$$\mathbf{w}_{\mathrm{L},n} = \mathbf{w}_{\mathrm{L},n-1} + \mu_\mathrm{L} \frac{\mathbf{x}_n e\left[n\right]}{\mathbf{x}_n^T \mathbf{x}_n + \delta} \tag{3}$$

$$\mathbf{w}_{\mathrm{FL},n} = \mathbf{w}_{\mathrm{FL},n-1} + \mu_\mathrm{FL} \frac{\mathbf{g}_n e\left[n\right]}{\mathbf{g}_n^T \mathbf{g}_n + \delta} \tag{4}$$

where μ_L and μ_FL are the step-size parameters and δ is a regularization factor.

3 Chebyshev Functional Link Expansion

3.1 Functional expansion block

One of the most important part in the FLAF-based model is the FEB, which contains a series of functions satisfying universal approximation properties. Such functions, called "functional links", are collected in a chosen set $\Phi = \{\varphi_0\left(\cdot\right), \varphi_1\left(\cdot\right), \ldots, \varphi_{Q-1}\left(\cdot\right)\}$, with Q being the number of functional links. The FEB receives the first $M_\mathrm{i} \leq M$ samples of $\mathbf{x}_n$, which are transformed and expanded in a higher-dimensional space by the chosen set of functional links, thus yielding the nonlinear expanded buffer $\mathbf{g}_n$:

$$
\begin{aligned}
g[n] &= \varphi_0(x[n]) \\
&\vdots \\
g[n-Q+1] &= \varphi_{Q-1}(x[n]) \\
&\vdots \\
g[n-M_e+1] &= \varphi_{Q-1}(x[n-M_i+1])
\end{aligned}
$$

where P is the *order* of the functional link.

3.2 Chebyshev polynomial expansion

The chosen set of functional links must satisfy the universal approximation properties, and it can be a subset of orthogonal polynomials, such as Chebyshev, Legendre, Laguerre and trigonometric polynomials [2, 5, 13, 19] or just approximating functions, such as sigmoid functions [11, 15]. Among such functional expansions, *trigonometric polynomials* represent one of the most popular expansions, especially for applications involving audio and speech input signals [6, 16], since at best of their capabilities they provide the best performance results [5]. However, in this paper we focus on Chebyshev polynomial expansion to reduce the computational load.

Chebyshev polynomials are widely used in different fields of application due to their powerful nonlinear approximation capabilities. These properties were proved in [13, 19] within an artificial neural network (ANN), which also shows faster convergence than a multi-layer perceptron (MLP) network. Chebyshev polynomials involve functions of previously computed functions, thus increasing their effectiveness in dynamic problems. Moreover, being derived from a power series expansion, Chebyshev functional links may approximate a nonlinear function with a very small error near the point of expansion. On the other hand, the drawback is that, far from the point of expansion, the error often increases rapidly. Compared with other power series, Chebyshev polynomials show lower computational complexity and higher efficiency, when the polynomial order is rather low.

Considering the i-th input $x[n-i]$ of the nonlinear buffer, with $i = 0, \ldots, M_i$, the Chebyshev polynomial expansion can be expressed as:

$$
\varphi_j(x[n-i]) = 2x[n-i]\,\varphi_{j-1}(x[n-i]) - \varphi_{j-2}(x[n-i]) \tag{5}
$$

for $j = 0, \ldots, P-1$. It can be noted from (5) that the number of Chebyshev functional links is equal to the expansion order, i.e., $Q = P$. Initial values in (5) (i.e., for $j = 0$) are:

$$
\begin{aligned}
\varphi_{-1}(x[n-i]) &= x[n-i] \\
\varphi_{-2}(x[n-i]) &= 1.
\end{aligned} \tag{6}
$$

3.3 Properties of Chebyshev polynomials

Chebyshev polynomials are endowed with some interesting properties [3]. They are orthogonal in $\mathbb{R}_1$ with respect to the a weighting function $1/\pi\sqrt{1-x^2\left[n-i\right]}$:

$$\int_{-1}^{1} \varphi_j\left(x\left[n-i\right]\right)\varphi_k\left(x\left[n-i\right]\right)\frac{1}{\pi\sqrt{1-x^2\left[n-i\right]}}\mathrm{d}x = \begin{cases} 0, & j \neq k \\ 1, & j = k = 0 \\ 1/2, & j = k \neq 0 \end{cases}. \tag{7}$$

For any $x\left[n-i\right] \in \mathbb{R}_1$, also $\varphi_j\left(x\left[n-i\right]\right) \in \mathbb{R}_1$, with values comprises in the range $[-1,1]$. Therefore, $\varphi_j\left(x\left[n-i\right]\right)$ are *equiripple functions* in $\mathbb{R}_1$.

Moreover, any polynomial with order P, $p\left(x\left[n-i\right]\right)$, can be derived as a *linear combination* of Chebyshev polynomials [3]:

$$p\left(x\left[n-i\right]\right) = \sum_{j=0}^{P-1} c_j\varphi_j\left(x\left[n-i\right]\right). \tag{8}$$

The last property is important since a linear combination of Chebyshev polynomials can arbitrarily well approximate any real continuous function $f\left(x\left[n-i\right]\right)$. This can be proved via the *Stone-Weierstrass theorem* [14], as shown in [3].

Moreover, the approximation of a continuous function $f\left(x\left[n-i\right]\right)$ with a linear combination of Chebyshev polynomials, $p\left(x\left[n-i\right]\right)$, up to a degree P is very close to a min-max approximation [3]. Indeed, the approximation error is:

$$\epsilon\left[n-i\right] = f\left(x\left[n-i\right]\right) - p\left(x\left[n-i\right]\right) = \sum_{j=P}^{+\inf} c_j\varphi_j\left(x\left[n-i\right]\right). \tag{9}$$

For a continuous and differentiable function, the coefficients c_j converge to 0 rapidly, and therefore, $\epsilon\left[n-i\right] \approx c_P\varphi_P\left(x\left[n-i\right]\right)$, which corresponds to an equiripple function.

Some of the above properties are proved in [3] for Chebyshev polynomials.

3.4 Analysis of the computational complexity

We briefly report the computational complexity of the Chebyshev functional link expansions with respect to other standard expansions, like trigonometric and Legendre series expansions. In order to provide a fair view of the computational resources required by the expansions, we do not consider additional cost of the SFLAF structure but we focus only on the operations made by the FEB.

The Chebyshev functional link expansion in (5) involves for each iteration $2PM_\mathrm{i}$ multiplications and PM_i additions. Similarly, the complexity of Legendre and trigonometric functional link expansions is derived in [5]. In terms of the expanded buffer length, we can consider that $M_\mathrm{e} = PM_\mathrm{i}$ for Chebyshev and Legendre functional link expansions and $M_\mathrm{e} = 2PM_\mathrm{i}$ for trigonometric expansion. A comparison of the computational complexity, in terms of multiplications only, is summarized in Table 1. If we fix the expanded buffer length M_e and we consider

Table 1. Computational cost comparison of different functional link expansions in terms of multiplications.

Expansion Type	**No. Multiplications**
Chebyshev Polynomial Expansion	$2M_{\mathrm{e}}$
Legendre Polynomial Expansion	$4M_{\mathrm{e}}$
Trigonometric Series Expansion	$M_{\mathrm{e}}/2+P$

that $P << M_{\mathrm{e}}$, then it is easy to note that the trigonometric expansion involves the smallest number of multiplications. Therefore, in order to achieve the best tradeoff between performance and complexity using Chebyshev functional link, we necessary need to obtain superior performance than trigonometric functional links. As an alternative, we should try to obtain the same performance of trigonometric functional links but with a smaller number of nonlinear elements.

4 Experimental Results

We assess the proposed Chebyshev SFLAF within NAEC scenarios, comparing results with those obtained by trigonometric and Legendre series expansions. For each experiment we show the best possible SFLAF configuration, in terms of the chosen parameters, yielding the optimal tradeoff between performance and computational complexity.

We evaluate the performance results in terms of the *echo return loss enhancement* (ERLE), which describes the amount of echo canceled by the microphone signal, and it is defined as:

$$\mathrm{ERLE}\left[n\right]=10\log_{10}\left(\frac{\mathrm{E}\left\{d^2\left[n\right]\right\}}{\mathrm{E}\left\{e^2\left[n\right]\right\}}\right) \tag{10}$$

The ERLE denotes how much echo signal is canceled, but this does not always correspond to a real signal enhancement in terms of perceived quality. Therefore, besides using the ERLE, we also consider another quality measure suitably designed for speech signals that may denote how "well" an SFLAF model produces a reliable estimate of the echo signal. In particular, we consider one of the most used objective measures for the speech quality evaluation that is the *perceptual evaluation of speech quality* (PESQ) [1], which estimates the overall loudness difference between the original signal and its estimation. Such signals are equalized to a reference listening level and then processed by a filter having a similar response to a standard telephone handset. An auditory transform is then applied to obtain the loudness spectra. The loudness difference between the two signals is averaged over time and frequency in order to achieve a prediction of subjective quality rating [5]. The PESQ score may be comprise in the range $[1.0, 4.5]$, where 4.5 indicates the best possible quality.

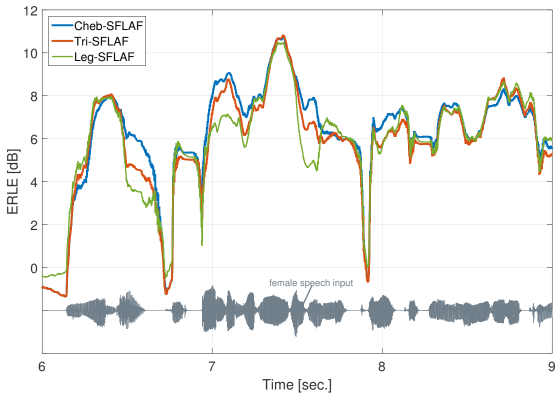

Fig. 2. Performance comparison in terms of ERLE between SFLAFs with different expansions for speech input affected by a soft-clipping nonlinearity.

Table 2. Performance comparison in terms of PESQ and processing time between SFLAFs with different expansions for speech input affected by a soft-clipping nonlinearity.

SFLAF Type	M_{e}	**PESQ**	**Sec.**
Chebyshev SFLAF	$PM_{\mathrm{i}} = 600$	3.765	5.339
Legendre SFLAF	$PM_{\mathrm{i}} = 1200$	2.868	7.876
Trigonometric SFLAF	$PM_{\mathrm{i}} = 1200$	3.582	5.672

4.1 Simulated NAEC scenario

The first experiment is conducted in a simulated teleconferencing environment with reverberation time of $T_{60} \approx 150$ ms, in which the acoustic impulse response between the loudspeaker and the microphone is measured at 8 kHz sampling rate. A desktop computer equipped with an i3 CPU at 3.07 GHz is used for simulations. Female speech signal is used as input. Additive Gaussian noise is considered at the microphone signal, with 20 dB of *signal-to-noise ratio* (SNR). The simulated distortion applied to the female speech is a symmetrical soft-clipping nonlinearity, aiming at simulating a classic loudspeaker saturation effect, described by [5]:

$$\overline{x}[n] = \begin{cases} \frac{2}{3\zeta} x[n] & \text{for } 0 \leq |x[n]| \leq \zeta \\ \operatorname{sign}(x[n]) \frac{3-(2-|x[n]|/\zeta)^2}{3} & \text{for } \zeta \leq |x[n]| \leq 2\zeta \\ \operatorname{sign}(x[n]) & \text{for } 2\zeta \leq |x[n]| \leq 1 \end{cases} \tag{11}$$

where the clipping threshold $0 < \zeta \leq 0.5$ determines the nonlinearity level. Here, we consider a strong distortion, provided by using $\zeta = 0.1$ in (11).

We set the step sizes at $\mu_{\mathrm{L}} = \mu_{\mathrm{FL}} = 0.5$, and $M_{ri} = M$ for all the SFLAF but we use the minimum possible expansion order for Chebyshev, trigonometric and

Legendre series such that results can be comparable in terms of the ERLE. In particular, we have chosen $P = 2$ for both Chebyshev and trigonometric SFLAF and $P = 4$ for Legendre SFLAF. Such results are shown in Fig. 2 where it is possible to see that Chebyshev SFLAF provides the best performance keeping the complexity contained. For better readability of the figures, we show a window of 3 out of 10 seconds of the ERLE behavior. This result is more evident by evaluating the quality measures in terms of the PESQ, which are reported in Table 2, where it can be also seen that Chebyshev SFLAF achieves the best PESQ score, while adapting the lowest number of nonlinear elements and, thus, involving the lowest computational time.

4.2 Real NAEC scenario

In a second experiment, we evaluate the performance of the proposed method on real data from a classic scenario of acoustic echo cancellation, i.e. a hands-free desktop teleconference. For this experiment we consider a typical office room with a relatively low level of background noise, which guarantees sufficiently high signal-to-noise ratio (SNR). In this way it is possible to evaluate the proposed canceller fairly, thus avoiding external interferences that could require further processing modules (e.g., double-talk detectors). For the same reason, we used a high-quality microphone (AKG C562 CM), so that the most significant nonlinearities in the system are those produced by the loudspeaker. To this end, 40 cm far from the microphone, we placed a low-cost commercial loudspeaker, capable of introducing significant distortions. The input signal is male speech recorded at 16 kHz sampling frequency. The length of the experiments is 20 seconds. We consider a typical volume level of a quiet speech conversation, when usually loudspeaker distortions are mild and they cannot be perceived by the user. However, they do affect the echo cancellation, thus degrading the performance in the absence of a nonlinear modeling.

For this experiment, we use the same setting of the previous one, but with different filter lengths. In particular, we use $M = 200$ for the filter on the linear path, and $M_{\mathrm{i}} = 50$ for the functional expansion. Therefore, the number of parameters for the filter on the nonlinear path is $M_{\mathrm{e}} = PM_{\mathrm{i}} = 100$ for the Chebyshev SFAF, and $M_{\mathrm{e}} = 200$ for both the trigonometric and Legendre SFLAFs. Results in terms of the ERLE are shown in Fig 3, showing a window of 3 out of 20 seconds, in which Chebyshev SFLAF clearly shows the best performance with respect to the other methods. This result is confirmed by the PESQ scores in Table 3, where Chebyshev SFLAF outperforms the other methods, while showing the lowest computational complexity.

5 Conclusion

In this paper, a Chebyshev functional link adaptive filter has been introduced as a low complexity model for the modeling of distorted speech signals. The proposed model exploits the properties of Chebyshev polynomial expansions

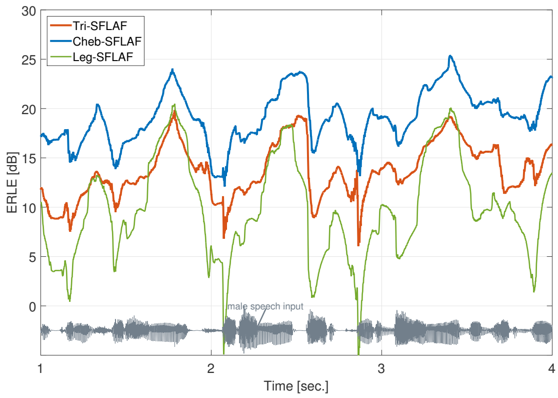

Fig. 3. Performance comparison in terms of ERLE between SFLAFs with different expansions for speech input in a real NAEC scenario.

Table 3. Performance comparison in terms of PESQ and processing time between SFLAFs with different expansions for speech input in a real NAEC scenario.

SFLAF Type	M_e	**PESQ**	**Sec.**
Chebyshev SFLAF	$PM_i = 100$	3.857	2.132
Legendre SFLAF	$PM_i = 200$	2.742	3.626
Trigonometric SFLAF	$PM_i = 200$	3.124	2.951

and takes advantage of the fact that it achieves acceptable performance even with low expansion order, thus resulting the best possible functional link-based model when a low computational complexity is required by a specific problem, like NAEC. Performance are evaluated in terms of an error-based measure, i.e., the ERLE, but also in terms of a speech quality measure, i.e., the PESQ. Overall results proved that Chebyshev SFLAF is the best performing method when the minimum possible computational resources are available for the nonlinear modeling.

References

1. Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs (2000)
2. Carini, A., Cecchi, S., Romoli, L., Sicuranza, G.L.: Legendre nonlinear filters. Signal Processing 109, 84–94 (Apr 2015)
3. Carini, A., Sicuranza, G.L.: A study about Chebyshev nonlinear filters. Signal Processing 122, 24–32 (May 2016)
4. Comminiello, D., Cecchi, S., Scarpiniti, M., Gasparini, M., Romoli, L., Piazza, F., Uncini, A.: Intelligent acoustic interfaces with multisensor acquisition for im-

mersive reproduction. IEEE Transactions on Multimedia 17(8), 1262–1272 (Aug 2015)
5. Comminiello, D., Scardapane, S., Scarpiniti, M., Parisi, R., Uncini, A.: Functional link expansions for nonlinear modeling of audio and speech signals. In: International Joint Conference on Neural Networks (IJCNN). Killarney, Ireland (Jul 2015)
6. Comminiello, D., Scarpiniti, M., Azpicueta-Ruiz, L.A., Arenas-García, J., Uncini, A.: Functional link adaptive filters for nonlinear acoustic echo cancellation. IEEE Transactions on Audio, Speech, and Language Processing 21(7), 1502–1512 (Jul 2013)
7. Comminiello, D., Scarpiniti, M., Parisi, R., Uncini, A.: Intelligent acoustic interfaces for immersive audio. In: 134th Audio Engineering Society Convention. Rome, Italy (May 4-7 2013)
8. Gil-Cacho, J.M., Signoretto, M., Van Waterschoot, T., Moonen, M., Jensen, S.H.: Nonlinear acoustic echo cancellation based on a sliding-window leaky kernel affine projection algorithm. IEEE Transactions on Audio, Speech, and Language Processing 21(9), 1867–1878 (Sep 2013)
9. Guerin, A., Faucon, G., Le Bouquin-Jeannes, R.: Nonlinear acoustic echo cancellation based on Volterra filters. IEEE Transactions on Speech and Audio Processing 11(6), 672–683 (Nov 2003)
10. Hofmann, C., Guenther, M., Huemmer, C., Kellermann, W.: Efficient nonlinear acoustic echo cancellation by partitioned-block significance-aware Hammerstein group models. In: 24th European Signal Processing Conference (EUSIPCO). pp. 1783–1787. Budapest, Hungary (Aug 2016)
11. Pao, Y.H.: Adaptive Pattern Recognition and Neural Networks. Addison-Wesley, Reading, MA (1989)
12. Patel, V., George, N.V.: Design of dynamic linear-in-the-parameters nonlinear filters for active noise control. In: 24th European Signal Processing Conference (EUSIPCO). pp. 16–20. Budapest, Hungary (Aug 2016)
13. Patra, J.C., Kot, A.C.: Nonlinear dynamic system identification using Chebyshev functional link artificial neural networks. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 32(4), 505–511 (Aug 2002)
14. Rudin, W.: Principles of Mathematical Analysis. McGraw-Hill, New York, NY (1976)
15. Scardapane, S., Comminiello, D., Scarpiniti, M., Uncini, A.: Benchmarking functional link expansions for audio classification tasks. In: Bassis, S., Esposito, A., Morabito, F.C., Pasero, E. (eds.) Advances in Neural Networks: Computational Intelligence for ICT, Smart Innovation, Systems and Technologies, vol. 54, pp. 133–141. Springer (Jun 2016)
16. Sicuranza, G.L., Carini, A.: A generalized FLANN filter for nonlinear active noise control. IEEE Transactions on Audio, Speech, and Language Processing 19(8), 2412–2417 (Nov 2011)
17. Uncini, A.: Fundamentals of Adaptive Signal Processing. Signal and Communication Technology, Springer International Publishing AG, Cham, Switzerland (2015), ISBN 978-3-319-02806-4
18. Van Vaerenbergh, S., Azpicueta-Ruiz, L.A., Comminiello, D.: A split kernel adaptive filtering architecture for nonlinear acoustic echo cancellation. In: 24th European Signal Processing Conference (EUSIPCO). pp. 1768–1772. Budapest, Hungary (Aug 2016)
19. Zhao, H., Zhang, J.: Functional link neural network cascaded with Chebyshev orthogonal polynomial for nonlinear channel equalization. Signal Processing 88(8), 1946–1957 (Aug 2008)