2017 21st International Conference on System Theory, Control and Computing (ICSTCC)

Adaptive Kernel Bandwidth Method for Kernel Normalized LMS Adaptive Algorithm

Kiyoshi Nishikawa
Department of System Design
Tokyo Metropolitan University, Japan
Email: knishikawa@m.ieice.org

Felix Albu, Dinu Coltuc, and Henri Coanda
Valahia University of Targoviste
Targoviste, Romania
Email: felix_albu@ieee.com

***Abstract*—In this paper, an adaptive adjustment method for the kernel parameter used in the kernel adaptive filters (KAFs) is proposed. The KAF is one of the linear-in-the-parameters (LIP) nonlinear filters, and is based on the kernel method used in machine learning. Typically, the Gaussian kernel function is used, but there is no effective method for automatically adjusting its parameter that influences the convergence characteristics of the KAFs. An adaptive adjustment method for this parameter is proposed in the paper. The proposed method uses the difference of ℓ_1 norms of the input signals for the unknown system and the adaptive filter as the criteria. The kernel parameter will be updated according to the differences. The qualitative results of the proposed method is shown by the computer simulations.**

***Keywords*—Kernel normalized least mean square algorithm; kernel parameter; Gaussian kernel**

I. Introduction

Adaptive filters (AFs) are used in a wide area of applications, e.g., echo or noise cancelling etc [1]. They enable the autonomous learning of an unknown linear system from its input and output signals.

Several attempts for extending AFs for nonlinear problems are considered and proposed as nonlinear AFs (NAFs) [2]–[6]. The kernel AF (KAF) [6] is a recently proposed NAF which is derived by applying the kernel method to linear AFs. For implementing a KAF, the input signal is transformed into a higher characteristics space, and then, an adaptive filter is applied in this space. The adaptive algorithms for the KAFs are derived by slightly modifying the conventional ones for the linear AFs. We can say that the KAFs is one of the linear-in-the-parameters (LIP) nonlinear filters, because the KAF can be seen as a linear filter with the transformed higher order input signal.

When the KAFs are implemented, the selection of the kernel function is important, and the Gaussian kernel is usually selected. The Gaussian kernel has a parameter called the band width and it is known that its value affects the convergence characteristics of the KAFs, e.g., the rate of convergence, the excess MSE, and so forth. This is one of the practical problems of KAFs when they are applied to the actual applications [7]. In many cases, when they are applied to real applications, the kernel parameter is determined experimentally in advance.

One of the solutions for relaxing the difficulty of the selection of the kernel parameter is to use multi-kernel AFs (MKAFs) [8]–[10]. MKAFs are implemented with multiple KAFs with different kernels to improve the performance of KAFs. Namely, multiple Gaussian kernels can be used with different values assigned to the kernel parameters, and this enables the improvement of the convergence characteristics if one or more of those parameters fit the input signal. However, it still remains that near optimum value is to be selected to one of the kernels even when the MKAFs are used.

Besides, several advanced structures of the KAFs are proposed so far. The mixture structures of KAFs [11], [12] is an extension of the mixture structure for the linear AFs [13]. On the other hand, the structures of [3], [5], [14]–[16] are improving the convergence characteristics when the target system consists of the linear and nonlinear components. Although these advanced structures are applied, the selection of the kernel parameter is still an important issue when the Gaussian kernel is used.

In this paper, we consider an adaptive method for adjusting the kernel parameter in order to relax the difficulty of the selection. The proposed method uses the LIP nature of the KAFs instead of altering the structure as mentioned above. For that, we investigate the effect of the kernel parameter on the transformed input signal. From the consideration, we propose an update equation of the kernel parameter. The proposed method is based on the difference between the ℓ_1 norms of the input signals to the KAF and the unknown system. It is shown by using computer simulations that the proposed method can adaptively adjust the kernel parameter and improve the convergence characteristics of the KAFs.

II. Kernel adaptive filters

Here, the kernel adaptive filter is summarized.

A. Kernel adaptive filter as an LIP nonlinear filter

The KAF is an extension of the linear AFs (LAFs) based on the kernel method [6], [7], [17]–[19].

The LAF, the input signal of the unknown system will be directly used as the input signal of the LAF. On the other hand, when a KAF is used, the transformed original input signal represents the input of the KAF. Namely, the input signal of the unknown system, expressed as $\{x(n) \mid n = 0, 1, 2, \ldots\}$ in the following lines will be mapped onto a higher order

978-1-5386-3842-2/17/$31.00 ©2017 IEEE

characteristics space using a mapping function $\psi(\mathbf{x}(n))$, where $\mathbf{x}(n)$ is the vector consists of $x(n)$ and is expressed as

$$\mathbf{x}(n) = [x(n) \quad x(n-1) \quad \cdots x(n-N+1)]^{\mathrm{T}} \tag{1}$$

where N is the length of the vector and it corresponding to the length of the adaptive filter, and the superscript $^{\mathrm{T}}$ shows the transpose of a vector or a matrix.

Then, the unknown system is modeled as a weighted sum of $\psi(\mathbf{x}(m))$ as

$$\bar{w}(n) = h_0\phi(\mathbf{x}(0)) + \cdots + h_{n-1}\phi(\mathbf{x}(n-1)) \tag{2}$$

where $\{h_i \mid i = 0, \ldots, n-1\}$ are the unknown weights. Using kernel adaptive algorithm, the set of $[h_0 \quad h_1 \quad \cdots \quad h_{n-1}]$ are optimized as the coefficient vector of a KAF. The form of (2) is a linearly weighted combination of h_i, and therefore, the KAF can be regarded as one of the LIP nonlinear filters. In Fig. 1, we show the structure of the KAF.

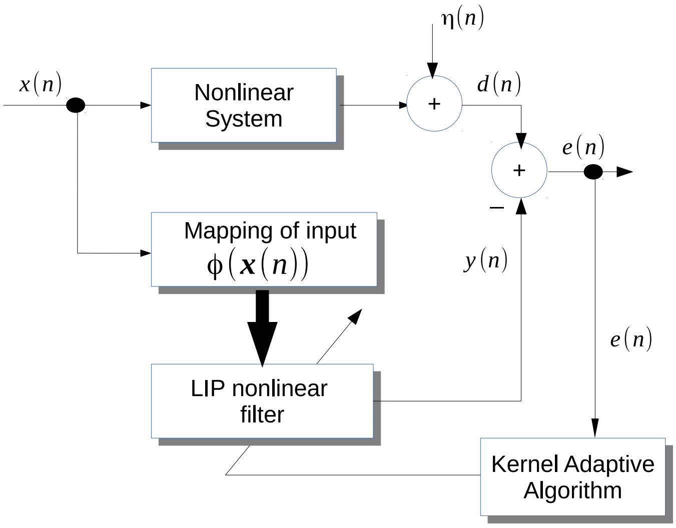

Fig. 1. Typical input and output relation of a kernel adaptive filter as a LIP system

By using the kernel trick [6], we can express $\upsilon(n)$, the output of $\bar{w}(n)$, as

$$\upsilon(n) = \quad h_0\kappa(\mathbf{x}(n), \mathbf{x}(0)) + \cdots + h_n\kappa(\mathbf{x}(n), \mathbf{x}(n-1)) \tag{3}$$

where $\kappa(\cdot, \cdot)$ shows a kernel function. The following form of the kernel function is widely used in kernel adaptive filtering

$$\kappa(\mathbf{x}, \mathbf{y}) = \exp\left(-\zeta||\mathbf{x} - \mathbf{y}||^2\right) \tag{4}$$

where ζ is a parameter of the kernel function called the bandwidth whose value affects the convergence characteristics. This form of the kernel is called the Gaussian kernel.

We define the vectors $\mathcal{S}(n)$ and $\boldsymbol{h}(n)$ as

$$\begin{aligned} \mathcal{S}(n) &= [\kappa(\mathbf{x}(n), \mathbf{x}(0)) \quad \cdots \quad \kappa(\mathbf{x}(n), \mathbf{x}(n-1))]^{\mathrm{T}} \\ \boldsymbol{h}(n) &= [h_0 \quad \ldots \quad h_{n-1}]^{\mathrm{T}} \end{aligned} \tag{5}$$

and, by substituting them into (3), we obtain the following relation

$$\upsilon(n) = \mathcal{S}^{\mathrm{T}}(n)\boldsymbol{h}(n). \tag{6}$$

This equation can be seen as the input-output relation of $\boldsymbol{h}(n)$. We show a typical structure of the kernel filter in Fig. 2. In this figure, M is used to show the number of elements in the dictionary for learning which is described in the next subsection.

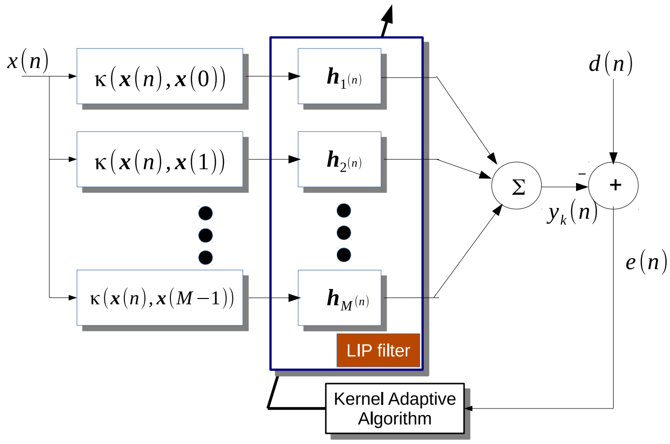

Fig. 2. Transform of input signal by kernel function

B. Kernel Normalized LMS algorithm

A lot of adaptive algorithms for KAFs have been proposed so far [6]. In this paper we use the kernel normalized least mean square (KNLMS) algorithm [7] which is as extension of the linear NLMS algorithm [1] for nonlinear problems.

The algorithm updates all the filter coefficients $\boldsymbol{h}(n)$ at each time n by

$$\boldsymbol{h}(n+1) = \boldsymbol{h}(n) + \eta\frac{e(n)\mathcal{S}(n)}{\epsilon + \mathcal{S}^{\mathrm{T}}(n)\mathcal{S}(n)} \tag{7}$$

$$e(n) = d(n) - \mathcal{S}^{\mathrm{T}}(n)\boldsymbol{h}(n-1) \tag{8}$$

where η is the step size and ϵ is the stabilization parameter.

At each time n, the input signal vector $\mathcal{S}(n)$ is stored in a dictionary $D = \{\mathcal{S}(m)|m = 0, \ldots, n-1\}$. This means that the order of the KAF increases at each time, and required computation becomes larger as time advances. To prevent this unwanted increase of computational load, several sparsification methods of the dictionary, e.g., [7], [20], were proposed. We use the sparsification method from [7]. When a sparsification method is used, the number of the input signal vectors $\mathcal{S}(n)$ stored in the dictionary becomes smaller than $n-1$. We use the variable M $(0 < M < n)$ to show the number of vectors in D. For the detail of the sparsification methods, please refer the references above.

C. The Gaussian kernel and the kernel parameter ζ

It is widely known that, when the kernel is the Gaussian, the value of ζ affects the convergence characteristics of the KAFs, e.g., rate of convergence, the excess MSE, etc [7], [20].

So far, there are no widely used criterion to select the value of the kernel parameter. Instead, suitable value would be selected based on the results of experiments as in [7]. On the other hand, the multi-kernel structure [8] was proposed which uses multiple kernels with different settings for relaxing the difficulty in the selection of the optimum value for the kernel parameter. However, even if the multiple kernels are used, it is not certain that one of the parameters is near the optimal value. Hence, it is desirable to obtain the adaptive method for adjusting ζ.

III. Proposed method

Here, we consider an adaptive method to assign an appropriate value for the kernel parameter. Note that, in the following, we assume that the KNLMS algorithm is used with a single Gaussian kernel.

A. Effect of selection of the kernel parameter ζ

We describe the proposed method that adaptively adjusts ζ of the Gaussian kernel. In order to develop an adjusting method, we need to find a criterion under which we can formulate the problem as an optimization problem.

Under the viewpoint that the KAF is a LIP nonlinear filter, we describe the effect of ζ on the output signal of the KAF. The output signal $v(n)$ of a KAF is given as an inner product as in (6), and the input signal of $\boldsymbol{h}(n)$ is $\mathcal{S}(n)$. Each component of $\mathcal{S}(n)$ is given as

$$\kappa(\mathbf{x}(n), \mathbf{x}(m)) = \exp\left(-\zeta ||\mathbf{x}(n) - \mathbf{x}(m)||^2\right) \quad m = 0, \ldots, n-1 \tag{9}$$

From this equation, we can see the magnitude of each component of $\mathcal{S}(n)$ is determined by the selection of ζ. Hence, the ℓ_1 norm $\|\mathcal{S}(n)\|_1$ or the ℓ_2 norm $\|\mathcal{S}(n)\|_2$ is also depends on ζ.

This means that the magnitude of $v(n)$ also becomes the function of ζ because the following relation holds

$$\|\mathcal{S}(n)\|\|\boldsymbol{h}(n)\| > \|\mathcal{S}(n) \cdot \boldsymbol{h}(n)\| \tag{10}$$

from the Cauchy-Schwarz inequality. When the magnitude of the output signal becomes smaller, then the error signal $e(n) = d(n) - v(n)$ becomes larger under the condition that $d(n)$ is the same. It is reasonable to maintain the magnitude of the output signal $v(n)$ to be as close as the desired signal $d(n)$ for the KAF to be updated.

We should note here that the optimum magnitude of $\mathcal{S}(n)$ may not be required because the adaptive algorithm adapts the filter coefficients $\boldsymbol{h}(n)$. On the other hand, it is well known that the convergence characteristics of the kernel filters, e.g., the rate of convergence, the excess MSE, etc, depend on the selection of ζ. Hence, in the following, we consider a method of adaptively adjusting ζ based on the consideration above.

B. The proposed adjusting method for ζ

Let us develop an adaptive method for adjusting the kernel parameter ζ under assuming that the kernel adaptive filter is a LIP nonlinear filter.

Our objective is to search for the near optimum value of kernel parameter for improving the convergence characteristics of the KAFs. An optimization based on the criterion using the error signal $e(n)$ could be considered. However, this requires the simultaneous optimization of ζ and the coefficients of the KAF, and it may be complicated to develop.

Instead, we consider here a simpler method based on the view of the KAF as an LIP. Namely, we propose to adjust the value of the kernel parameter using the ℓ_1 norm of $\mathcal{S}(n)$. As mentioned in the previous subsection, the ℓ_1 norm becomes a function of ζ. We require this ℓ_1 norm to be close to that of $\mathbf{x}(n)$. Namely, we adjust ζ according to the difference between the ℓ_1 norms of the input signals of the unknown system and the KAF:

$$\varepsilon_\ell(n) = \|\mathbf{x}(n)\|_1 - \|\mathcal{S}(n)\|_1 \tag{11}$$

where $\|\mathbf{x}\|_1$ shows the ℓ_1 norm of the vector $\mathbf{x}$, i.e., $\|\mathbf{x}\|_1 = |x_0| + |x_1| + \cdots + |x_{M-1}|$.

Using the error defined by (11), we update the kernel parameter using the LMS-style equation.

$$\zeta(n+1) = \zeta(n) + \mu_\zeta \varepsilon_\ell(n) \tag{12}$$

where μ_ζ shows the step-size parameter. Note that $\zeta(n)$ shows the kernel parameter ζ at time n because in the proposed method ζ becomes time-variant. Although the length of $\mathbf{x}(n)$ is fixed as N, the length of $\mathcal{S}(n)$ is time-varying.

In actual applications, we might need to limit the value of $\zeta(n)$ in the predefined region $\zeta_{\text{min}} \leq \zeta(n) \leq \zeta_{\text{max}}$ by

$$\begin{cases} \zeta(n+1) = \zeta_{\text{max}} & \text{if } \zeta_{\text{max}} < \zeta(n+1) \\ \zeta(n+1) = \zeta_{\text{min}} & \text{if } \zeta_{\text{min}} > \zeta(n+1) \end{cases} \tag{13}$$

after update of ζ at each time to prevent the divergence of ζ.

IV. Simulation Results

Simulations were performed in order to show the qualitative results and prove the validity of the proposed method.

We compared three KAFs, namely two of them are KNLMS adaptive filters with the fixed ζ, i.e., (i) $\zeta = 0.5$, (ii) $\zeta = 1.5$, and (iii) the proposed method. Note that the values of ζ for the KNLMS were determined from the results of simulations with varying its value under the conditions below.

The Gaussian kernel was employed and the KNLMS algorithm was used to update the filters. For the proposed method, the initial value for $\zeta(n)$ was set as 1.0 regardless of the simulation type. The step-size parameters of the KNLMS filters were set as $\eta = 0.1$. Besides, μ_ζ was set to 0.1 after some trials. The length of adaptive filters N was set to $N = 30$. The zero mean Gaussian noise with 0.001 variance was added to the desired signal. The results of ensemble averages of 1000 independent simulations are shown.

A. Time invariant nonlinear system

We simulated two non-linear models used in [21]. Note that the models were originally proposed in [22] by extending the primary model proposed in [23]. In the first simulation, the equation below was used to generate the signal $d(n)$:

$$\begin{cases} v(n) = \dfrac{v(n-1)}{1 + v^2(n-1)} + u^3(n-1) \\ d(n) = v(n) + \xi(n) \end{cases} \tag{14}$$

where $u(n)$ and $v(n)$ are the input and the output signals respectively. We used a Gaussian process of the zero-mean and the standard deviation $\sigma_u = 1$ to generate $u(n)$.

Figure 3 show that the KNLMS with $\zeta = 0.5$ provides better convergence characteristics than that with $\zeta = 1.5$.

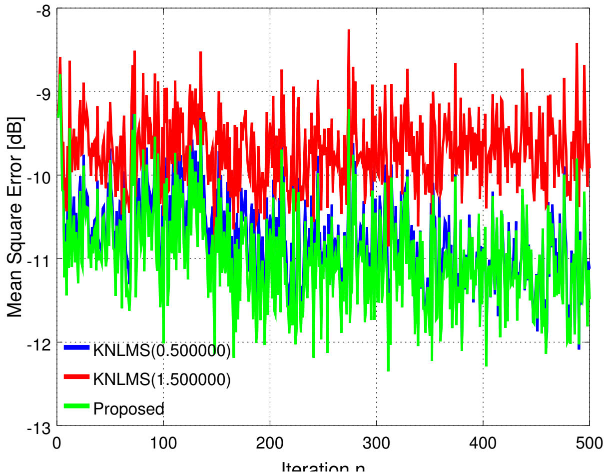

Fig. 3. Comparison of MSE for the system (14).

Next, we simulated the second model in [21]:

$$\phi(v(n)) = \begin{cases} \dfrac{v(n)}{3[0.1 + 0.9v^2(n)]^{1/2}} & \text{for } v(n) \le 0 \\ \dfrac{-v^2(n)[1 - \exp(0.7v(n)]}{3} & \text{for } v(n) > 0 \end{cases}$$

$$d(n) = \phi(v(n)) + \xi(n) \tag{15}$$

where $\xi(n)$ is the additive noise. Besides, $v(n)$ is given as

$$v(n) = \mathbf{a}^{\mathrm{T}}\mathbf{u}(n) - 0.2v(n-1) + 0.35v(n-2) \tag{16}$$

where $\mathbf{a}$ and $\mathbf{u}(n)$ are

$$\mathbf{a} = [1 \;\; 0.5]^{\mathrm{T}} \tag{17}$$

$$\mathbf{u}(n) = [u_1(n) \;\; u_2(n)]^{\mathrm{T}} \tag{18}$$

respectively.

The results are shown in Fig. 4. It can be seen that the convergence characteristics of the proposed method approaches those of the KNLMS with $\zeta = 0.5$.

By comparing the results in Figs 3 and 4, it can be seen that the proposed method achieves the comparative convergence property with the better one of the KNLMS KAFs. Therefore, we can say that the proposed method has an ability to adaptively adjust ζ for these systems.

B. Time varying nonlinear system

In this section the simulation results of the proposed method for the time varying nonlinear systems are shown in order to confirm its tractability.

In this case, we add an linear filter to the models of Sec. IV-A. Namely, the output signal is given as

$$y(n) = a_1 y_n(n) + (1 - a_1) y_l(n) \tag{19}$$

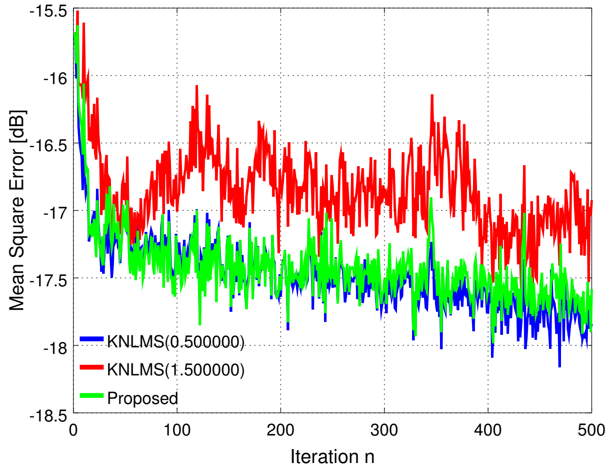

Fig. 4. Comparison of MSE for the system (15).

where $y_l(n)$ and $y_n(n)$ are the output of a linear and a nonlinear filter respectively, and a_1 is the weight to balance the linear and nonlinear components of $y(n)$. As the nonlinear filter, we used the model 1 and 2 of Sec. IV-A, i.e., $y_n(n)$ is obtained from either (14) or (15). On the other hand, as the linear filter, a low pass filter with tap length 21 designed using the Remez algorithm was used. The value of the weight a_1 was set to 1 ($0 < n < 300$), 0.5 ($300 \le n < 500$), 0 ($500 \le n$) to artificially change the system from nonlinear to linear.

Fig. 5 shows the results for the first model, while Fig. 6 shows the results for the second model. By comparing these results, it can be seen that, after the system change at $n = 200$, the KNLMS with $\zeta = 1.5$ provides better convergence characteristics in Fig. 5, and on the other hand, that with $\zeta = 0.5$ in Fig. 6. The proposed method provides almost the same characteristics as the one that achieves the better performance in both cases. These results suggests the validity of the proposed method.

Although there are a lot of points to be considered further to improve the performance, we can say that it is confirmed that the proposed method has an ability to adjust the kernel parameter according to the environments.

V. Conclusion

In this paper, we considered a method for adaptive adjustment of the kernel parameter ζ of the Gaussian kernel used for the KAFs from the view point of LIP nonlinear filters. We investigated the relation of the ℓ_1 norm of the input signal and ζ. From that we proposed a method for adaptive adjustment of ζ based on the difference of ℓ_1 norms. We applied the proposed method to the nonlinear system modeling in computer simulations. The results show that the proposed method enables to adaptively adjust ζ. Although the simulated environments are limited and the method may not obtain the optimum value, the possibility to adjust the kernel parameter was shown in the investigated examples. For the future work, we would develop a theoretical analysis of the

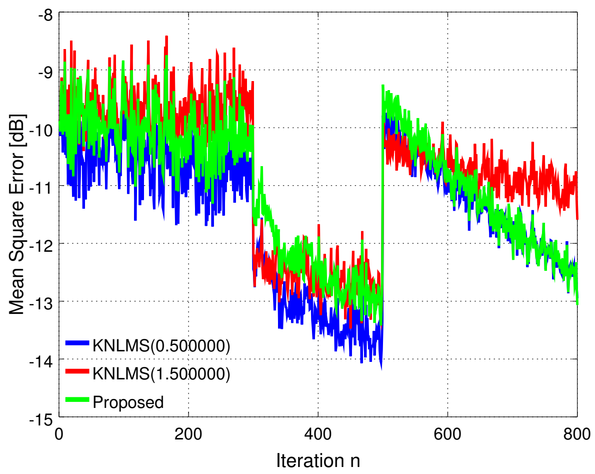

Fig. 5. MSE performance of the investigated algorithms for the time varying system with the nonlinear system defined by the equation (14) and a linear system. The system was changed at $n = 200$, and $n = 500$.

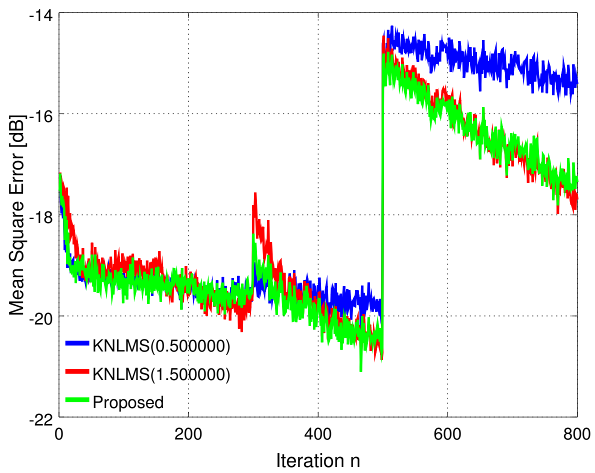

Fig. 6. MSE performance of the investigated algorithms for the time varying system with the nonlinear system defined by the equation (14) and a linear system. The system was changed at $n = 200$, and $n = 500$.

proposed method. Besides, we will consider the effect of the selection of the parameters, i.e., the step size μ_ζ in (12), the upper and the lower bounds for the kernel parameter in (13).

ACKNOWLEDGMENT

The work of Felix Albu, Dinu Coltuc and Henri-George Coanda was supported by a grant of the Romanian National Authority for Scientific research and Innovation, CNCS/CCCDI-UEFISCDI project number PN-III-P2-2.1-PED-2016-0651.

REFERENCES

[1] A. H. Sayed, *Fundamentals of Adaptive Filtering*. John Wiley & Sons, 2003.

[2] V. Mathews, "Adaptive polynomial filters," *Signal Processing Magazine, IEEE*, vol. 8, no. 3, pp. 10–26, July 1991.

[3] L. Azpicueta-Ruiz, M. Zeller, A. Figueiras-Vidal, J. Arenas-Garcia, and W. Kellermann, "Adaptive combination of volterra kernels and its application to nonlinear acoustic echo cancellation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 1, pp. 97–110, Jan 2011.

[4] D. Comminiello, M. Scarpiniti, L. Azpicueta-Ruiz, J. Arenas-Garcia, and A. Uncini, "Functional link adaptive filters for nonlinear acoustic echo cancellation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, no. 7, pp. 1502–1512, July 2013.

[5] G. Sicuranza and A. Carini, "Nonlinear system identification by means of mixtures of linear-in-the-parameters nonlinear filters," in *Proceedings on ISPA2013*, Sept 2013, pp. 337–342.

[6] W. Liu, J. C. Principe, and S. Haykin, *Kernel Adaptive Filtering*. Wiley, 2010.

[7] C. Richard, J. C. M. Bermudez, and P. Honeine, "Online Prediction of Time Series Data With Kernels," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1058–1067, Mar. 2009.

[8] M. Yukawa, "Multikernel Adaptive Filtering," *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4672–4682, Sept 2012.

[9] J. Arenas-Garcia, A. Figueiras-Vidal, and A. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 1078–1090, Mar. 2006.

[10] S. V. Vaerenbergh, L. A. Azpicueta-Ruiz, and D. Comminiello, "A split kernel adaptive filtering architecture for nonlinear acoustic echo cancellation," in *2016 24th European Signal Processing Conference (EUSIPCO)*, Aug 2016, pp. 1768–1772.

[11] R. Pokharel, S. Seth, and J. Principe, "Mixture kernel least mean square," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, Aug 2013, pp. 1–7.

[12] W. Gao, C. Richard, J.-C. Bermudez, and J. Huang, "Convex combinations of kernel adaptive filters," in *Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on*, Sept 2014, pp. 1–5.

[13] S. Kozat, A. Erdogan, A. Singer, and A. Sayed, "Steady-state mse performance analysis of mixture approaches to adaptive filtering," *Signal Processing, IEEE Transactions on*, vol. 58, no. 8, pp. 4050–4063, Aug 2010.

[14] K. Nishikawa and F. Albu, "Consideration on the Performance of Kernel Adaptive Filters for the Mixture of Linear and Non-Linear Environments," in *Proc. APSIPA ASC 2014*, Dec. 2014.

[15] ——, " Implementation method of Kernel Adaptive Filter as an Add-on for a Linear Adaptive Filter ," in *Proc. EUSIPCO 2015*, Sep. 2015.

[16] J. Gil-Cacho, M. Signoretto, T. van Waterschoot, M. Moonen, and S. Jensen, "Nonlinear Acoustic Echo Cancellation Based on a Sliding-Window Leaky Kernel Affine Projection Algorithm," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, no. 9, pp. 1867–1878, Sept 2013.

[17] Y. Engel, S. Mannor, and R. Meir, "The Kernel Recursive Least-Squares Algorithm," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2275–2285, Aug. 2004.

[18] K. Nishikawa, Y. Ogawa, and F. Albu, "Fixed Order Implementation of Kernel RLS-DCD Adaptive Filters," in *Proc. APSIPA ASC 2013*, Oct. 2013.

[19] F. Albu and K. Nishikawa, "The Kernel Proportionate NLMS Algorithm," in *Proc. EUSIPCO 2013*, Marrakech, Morocco, Sep. 2013.

[20] W. Liu, I. Park, and J. C. Principe, "An information theoretic approach of designing sparse kernel adaptive filters." *IEEE Transactions on Neural Networks*, vol. 20, no. 12, pp. 1950–61, Dec. 2009.

[21] W. Gao, J. Chen, C. Richard, and J. Huang, "Online Dictionary Learning for Kernel LMS," *IEEE Transactions on Signal Processing*, vol. 62, no. 11, pp. 2765–2777, Jun. 2014.

[22] W. D. Parreira, J. C. M. Bermudez, C. Richard, and J. Y. Tourneret, "Stochastic behavior analysis of the gaussian kernel least-mean-square algorithm," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2208–2222, May 2012.

[23] J. Voros, "Modeling and identification of wiener systems with two-segment nonlinearities," *IEEE Transactions on Control Systems Technology*, vol. 11, no. 2, pp. 253–257, Mar 2003.