An Adaptive Kernel Width Update Method of Correntropy for Channel Estimation

Weihua Wang, Jihong Zhao, Hua Qu, Badong Chen
School of Electronic and Information Engineering,
Xi'an Jiaotong University
Xi'an, China
chenbd@mail.xjtu.edu.cn

Jose C. Principe
Department of Electrical and Computer Engineering
University of Florida
Gainesville, USA
principe@cnel.ufl.edu

***Abstract*—The maximum correntropy criterion (MCC) is a robust adaptation criterion, which has been successfully applied in various signal processing applications. The kernel width plays an important role in MCC adaptive filtering and has significant influence on convergence rate and steady-state performance. In this paper, we develop an adaptive kernel width update method for correntropy. The proposed algorithm may achieve faster convergence speed and lower steady-state excess mean square error (EMSE) than the original MCC algorithm. More importantly, the new method involves no extra free parameters and is computationally very simple.**

Keywords—correntropy; MCC; adaptive kernel width; channel estimation

I. INTRODUCTION

In recent years, information theoretic learning (ITL) [1,2] methods have been shown to be efficient approaches in non-Gaussian signal processing since they capture the whole distribution rather than a specific moment of the data. The minimum error entropy (MEE) [3-5] and the maximum correntropy criterion (MCC) [6-8] are two popular optimization criteria in ITL. The MEE cost function is computationally much more expensive than a LMS-type algorithm. The MCC is a robust optimality criterion for non-Gaussian signal processing, and computationally very simple. Thus, the MCC is more suitable for practical implementation. Recently, the MCC has been successfully applied to various signal processing and machine learning problems [9-11].

The kernel width of correntropy is an important free parameter, which determines the performance of the MCC learning. How to choose a proper kernel width is thus a crucial problem in correntropy applications. Silverman's rule [12] is one of the most widely used methods for kernel selection, but in general it is only effective for probability density function (PDF) estimation problems. Other PDF estimation based methods include the statistical method [13] and cross-validation techniques [14]. In order to optimize the kernel width in general ITL learning, an adaptive update algorithm by minimizing a Kullback-Leibler (KL) divergence was derived in [15]. A fixed-point update rule with no free parameters was further developed in [16]. However, all the methods mentioned above are not directly related to the goodness criteria of an adaptive filtering algorithm, which are usually measured by convergence speed or accuracy. An adaptive algorithm was developed [17] for the selection of kernel width, which utilizes an appropriate kernel width under Gaussian condition as a standard to search a proper kernel width for other conditions, and updates the kernel width at every iteration based on the shape of the error distribution. However, estimating the shape of the error distribution is in general involved and inaccurate.

In this work, a new adaptive algorithm for kernel width update is derived. This method updates the kernel width at every iteration based on the prediction error. Specifically, the kernel width (squared) equals the sum of a predetermined kernel width (squared) and the prediction error (squared). The new algorithm may achieve excellent performance while keeping the simplicity and robustness of the original MCC algorithm.

The rest of the paper is organized as follows. In section II, using correntropy as a cost function in adaptive filters is reviewed. In section III, an adaptive kernel width update method of correntropy is developed. In section IV, simulation results of time-varying multipath Rayleigh channel estimation are presented to illustrate the desirable performance of the proposed algorithm. Finally, conclusion is given in section V.

II. MAXIMUM CORRENTROPY CRITERION

Correntropy is a nonlinear similarity measure between two random variables X and Y in kernel spaces [6]. In practice, the data distribution is usually unknown, and only a finite number of samples $\{(x_i, y_i)\}_{i=1}^N$ for X and Y are available. In this case, it is common to use sample mean estimator for the expectation operator, and the correntropy can be estimated as:

$$\hat{\mathrm{V}}_{N,\sigma}(X,Y) = \frac{1}{N}\sum_{i=1}^{N}\kappa_\sigma(x_i, y_i) \tag{1}$$

In this work, without mentioned otherwise, the kernel function in correntropy is a Gaussian kernel, defined by

$$\kappa_\sigma(x,y) = \exp(-\frac{\|x-y\|^2}{\sigma^2}) \tag{2}$$

Like MSE criterion, correntropy can be used as a cost function in adaptive filters [9]. The objective of the filter is to optimize the cost function in such a way that the filter output

resembles as closely as possible to the desired signal. Let's define our cost function to be the correntropy between the desired signal d_i and the filter output y_i. We will use a normalized Gaussian kernel to compute correntropy.

$$J_n = \frac{1}{N}\sum_{i=n-N+1}^{n} \exp(\frac{-|d_i - y_i|^2}{\sigma^2}) \tag{3}$$

where $y_n = \mathbf{W}_n^{\mathrm{T}}\mathbf{X}_n$, and $\mathbf{W}_n$ is the filter weight at nth iteration.

Bayesian estimation under maximum correntropy criterion has been studied in [18], it has been shown that maximum correntropy (MC) estimation is a smoothed maximum a posteriori (MAP) estimation, including the MAP and the minimum mean square error (MMSE) estimation as the extreme case.

Similarly with MSE criterion, we can use an iterative gradient ascent approach to search the optimal solution. Therefore,

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \frac{\mu}{\sigma^2}\cdot\frac{1}{N}\sum_{i=n-N+1}^{n}\left[\exp(\frac{-|e_i|^2}{\sigma^2})e_i\mathbf{X}_i\right] \tag{4}$$

Inspired by the stochastic gradient, the weight update equation using the correntropy cost function has therefore been reduced to this simple form.

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \frac{\mu}{\sigma^2}\exp\left(\frac{-|e_n|^2}{\sigma^2}\right)e_n\mathbf{X}_n \tag{5}$$

in which $e_n = d_n - \mathbf{W}_n^{\mathrm{T}}\mathbf{X}_n$ is the prediction error.

The steady-state excess mean square error (EMSE) of the adaptive filtering under the maximum correntropy criterion in the presence of Gaussian and non-Gaussian noise environment has been studied [19].

The kernel width is the window within which the similarity of the two random variables is computed. A very large kernel width will yield a similarity measure close to MSE value. A small value of the order of the error is useful in exploiting the advantageous properties of correntropy, but will lead to a slower convergence speed when the error is large initially. Thus the kernel width of correntropy is a very important parameter in the applications of the MCC algorithm.

III. ADAPTIVE KERNEL WIDTH

In the original MCC algorithm, the kernel width is fixed throughout the whole process, but the error in an adaptive system decreases gradually as the system weights continuously update to a steady-state. A fixed kernel width may affect the adaptation dynamics and would be a compromise between fast learning initially, and fast learning near the optimum point. Therefore, adaptively update the kernel width according to the error can potentially improve the overall convergence rate.

In this section, an adaptive kernel width update method of correntropy is derived to improve the overall convergence rate while keeping the robustness and steady-state behavior of the original MCC algorithm. The basic idea of the proposed algorithm is to adaptively update the kernel width according to the prediction error. The new algorithm selects the kernel width which makes the prediction error decrease along the largest slope, which makes sure the algorithm converges most quickly at every iteration.

The weight update expression of adaptive kernel width method is given by

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \frac{\mu}{\sigma_n^{\,2}}\exp\left(\frac{-|e_n|^2}{\sigma_n^{\,2}}\right)e_n\mathbf{X}_n \tag{6}$$

where σ_n is the adaptive kernel width, which is calculated according to the prediction error in nth iteration.

When we got the prediction error e_n at nth iteration, and then calculate the kernel width that makes error decrease along the largest slope

$$\max_{\sigma_n} J_n' = \frac{1}{\sigma_n^2}\exp\left(\frac{-|e_n|^2}{\sigma_n^2}\right)e_n \tag{7}$$

It is obvious that the adaptive kernel width σ_n at nth iteration is

$$\sigma_n^2 = e_n^{\,2} \tag{8}$$

The largest slope at nth iteration is

$$\max J_n' = \exp(-1)/e_n \tag{9}$$

It is should be noticed that if we adopt adaptive kernel width as $\sigma_n^2 = e_n^{\,2}$, then the exponential part of J' will be a constant, and this algorithm is no longer a correntropy based algorithm. More importantly, the algorithm will not converge since the maximum of J' growing larger as the error decreasing.

In order to deal with this problem, we derived a switch method in a previous work [20]. Here we develop a new method to tackle with this problem by modifying the kernel width update rule:

$$\sigma_n^2 = e_n^{\,2} + \sigma^2 \tag{10}$$

where σ denotes a predetermined kernel width which can be calculated by Silverman's rule [12] or other methods. The predetermined kernel width is usually relatively smaller than the initial error, so it has little influence on the convergence speed at initial stage. Since the kernel width is adaptively updated, the developed algorithm is denoted as AMCC.

The J' determines the update gain of the weight vector, and controls the convergence rate and steady-state misjudgment of the algorithm. The curves of J' versus |e| with different kernel widths are shown in Fig. 1.

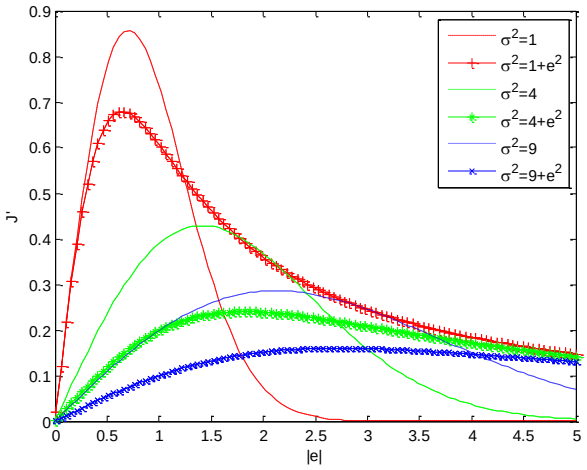

Fig.1. The curves of J' versus |e|

The J' curves with adaptive kernel width $\sigma_n^2 = e_n^{\ 2} + \sigma^2$ are above the curves with corresponding fixed kernel width when the error is large, this means that the proposed algorithm converges faster than original MCC at the initial stage. As the error decreases to zero, The J' curves with adaptive kernel width are below the curves with corresponding fixed kernel width, which means that the new algorithm may converge to a lower steady-state excess mean square error (EMSE) than the original MCC algorithm.

The proposed AMCC algorithm converges faster than the original MCC while achieving a lower steady-state EMSE. More importantly, the AMCC algorithm achieves the superior performance without involving any extra free parameters, and it keeps the simplicity and robust property of the original MCC algorithm. For a better understanding, the AMCC algorithm is summarized as Tab. I.

TABLE I. AMCC ALGORITHM

Switch kernel width maximum correntropy criterion

Initialization

set step-size μ **and predetermined kernel width** σ

$\mathbf{W}_0 = 0$, $e_1 = d_1$

Computation

While $\{\mathbf{X}_n, d_n\}$ **available do**

% compute the error

$e_n = d_n - \mathbf{W}_n^{\mathrm{T}} \mathbf{X}_n$

% update the kernel width

$\sigma_n^2 = e_n^{\ 2} + \sigma^2$

% update the weight

$$\mathbf{W}_{n+1} = \mathbf{W}_n + \frac{\mu}{\sigma_n^{\ 2}} \exp\left(\frac{-|e_n|^2}{\sigma_n^{\ 2}}\right) e_n \mathbf{X}_n$$

end while

IV. SIMULATION RESULTS

We now present simulation results to demonstrate the performance of the proposed AMCC algorithm. In the following, we compare the performance of the MCC, sign algorithm (SA) [21], adaptive kernel width selection algorithm derived in [17] (denoted as ASMCC) and the AMCC algorithm for channel estimation in multipath communication systems.

Adaptive channel estimation is a widely used technique for multipath communication systems. Among numerous adaptive techniques that exist in the open literature, pilot based algorithms are widely used for its simplicity. Pilot based channel estimation schemes have been illustrated in [22-24], all those methods were performed under the assumption of additive white Gaussian noise (AWGN). However, impulsive noises often exist in practical communication systems. In these cases, most of the algorithms would deteriorate rapidly. A correntropy-based robust algorithm can estimate the multipath channel under impulsive noise.

The overall channel estimation framework of the AMCC algorithm is shown in Fig. 2. The QPSK signals are transmitted via an unknown and time-varying multipath Rayleigh channels, and encountered with impulsive environment noise. We tried to estimate the multipath channel coefficients by correntropy-based adaptive filter. Firstly, the kernel width is updated according to the current prediction error, and then the weight of the adaptive filter is updated based on the current prediction error and the updated kernel width.

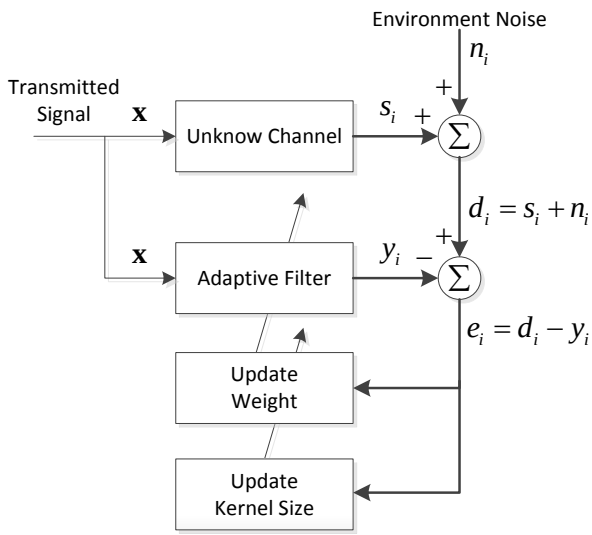

Fig.2. Framework of the AMCC algorithm

Assume the length of the multipath channel is 8, so there will be 8 unknown and time-varying Rayleigh channel coefficients need to be estimated. There 2000 QPSK training samples transmitted via the unknown and time-varying multipath Rayleigh channels, and corrupted by alpha-stable noise, the channels changed in the middle of the training sequence.

The alpha-stable noise model is adopted to simulate the impulsive environment noise,

$$f(t)=\exp\left\{ j\delta t-\gamma|t|^{\alpha}\,[1+j\beta\,\mathrm{sign}(t)S(t,\alpha)]\right\} \tag{11}$$

in which

$$S(t,\alpha)=\begin{cases}\tan\dfrac{\alpha\pi}{2} & if\ \alpha\neq 1\\ \dfrac{2}{\pi}\log|t| & if\ \alpha=1\end{cases} \tag{12}$$

where $\alpha\in(0,2]$ is the characteristic exponent, which measures the tail heaviness of the distribution. $-\infty<\delta<+\infty$ is the location parameter, $\beta\in[-1,1]$ is the symmetry parameter, and $\gamma>0$ is the dispersion parameter, measures the dispersion of the distribution, which plays a similar role to the variance of Gaussian distribution. In our simulations, the noise model n_i is defined as $N=(\alpha,\beta,\gamma,\delta)$. Parameters in the alpha-stable noise model are set as $N=(1.2,0,0.1,0)$.

As analyzed in section III, the AMCC algorithm converges faster than the original MCC while achieving a lower steady-state EMSE. So we first compare the convergence curves of those algorithms, in term of mean square deviation (MSD) of the real channel coefficients and the coefficients of the adaptive filter.

The step size of MCC and AMCC are set to 0.4, and the predetermined kernel width for MCC and AMCC are set to 2. We adjust the parameters of ASMCC and SA to meet the same level of steady-state MSD to compare its convergence rate with the MCC and AMCC. 50 Monte-Carlo simulations are run for the different date with different noises, the simulation results are shown in Fig. 3.

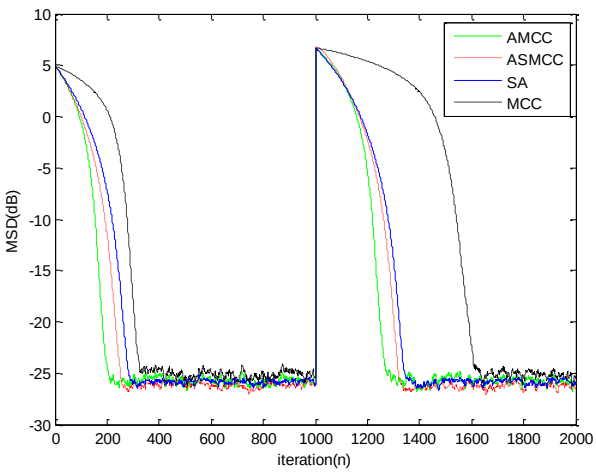

Fig.3. Convergence curves of AMCC, ASMCC, SA and MCC in an alpha-stable noise environment

It is obvious that the AMCC achieved a superior performance than other algorithms. There is a significant improvement in term of convergence rate compare AMCC with the original MCC, and the steady-state EMSE of AMCC is slightly smaller than the original MCC. The AMCC algorithm also converges faster than ASMCC and SA, while achieving the same level steady-state EMSE and it is computationally simpler than ASMCC.

In the following part, the effect of the predetermined kernel width σ is demonstrated. The step size is set to 0.3, 50 Monte-Carlo simulations are run for the different training data and impulsive noise with different predetermined kernel widths. The simulation results are shown in Fig. 4.

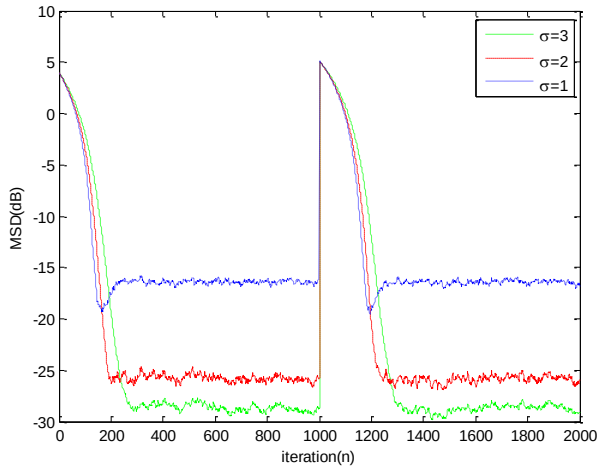

Fig.4. Convergence curves of AMCC with different predetermined kernel widths in an alpha-stable noise environment

As we can see in Fig.4, the AMCC algorithm with smaller predetermined kernel widths converge faster, but it has a larger steady-state EMSE, and if the predetermined kernel width is too small, the algorithm would achieve a small steady-state EMSE at first and then converge to a larger steady-state EMSE. This problem is existed in the original MCC as well and even worse. Because the slope of correntropy growing dramatically as the error decrease, this in turn will increase the update gain of the weight vector, and lead the prediction error grow larger.

V. CONCLUSION

In this paper, we developed an adaptive algorithm to update the kernel width. The proposed AMCC algorithm has a significant improvement on convergence rate while achieving a slightly lower steady-state EMSE than the original MCC algorithm. The algorithm involves no extra free parameters and keeps the simplicity and robustness of the original MCC algorithm. Simulation results demonstrate that the new algorithm may achieve an excellent performance in channel estimation under impulsive noises.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 61372152, No. 61372092), the National High Technology Research and Development Program of China (863 Program) (No.2014AA01A701), and the Key Project of major national science and technology on new generation of broadband wireless mobile communication network (No. 2013ZX0302010-003).

REFERENCES

[1] J.C. Principe. Information theoretic learning: Rényi's entropy and kernel perspectives. Springer, 2010.

[2] B. Chen, Y. Zhu, J. Hu, J.C. Principe. System parameter identification: information criteria and algorithms. Newnes, 2013.

[3] D. Erdogmus and J.C. Principe, "Generalized information potential criterion for adaptive system training," Neural Networks, IEEE Transactions on. 13.5(2002). pp.1035-1044.

[4] B. Chen, J. Hu, L. Pu, Z. Sun . "Stochastic gradient algorithm under (h, φ)-entropy criterion." Circuits, Systems & Signal Processing 26.6 (2007): pp.941-960.

[5] B. Chen, P. Zhu, J.C. Principe. "Survival information potential: a new criterion for adaptive system training." Signal Processing, IEEE Transactions on, 60.3 (2012). pp.1184-1194.

[6] W. Liu, P.P. Pokharel, J.C. Principe. "Correntropy: A localized similarity measure." Neural Networks, 2006. IJCNN'06. International Joint Conference on. IEEE, 2006. pp.4919-4924.

[7] I. Santamaría, P.P. Pokharel, J.C. Príncipe. "Generalized correlation function: definition, properties, and application to blind equalization." Signal Processing, IEEE Transactions on, 54.6 (2006). pp.2187-2197.

[8] W. Liu, P.P. Pokharel, J.C. Príncipe "Correntropy: properties and applications in non-Gaussian signal processing." Signal Processing, IEEE Transactions on, 55.11 (2007): pp.5286-5298.

[9] A. Singh, J.C. Principe. "Using correntropy as a cost function in linear adaptive filters." Neural Networks, 2009. IJCNN 2009. International Joint Conference on. IEEE, 2009. pp.2950-2955.

[10] S. Zhao, B. Chen, J.C. Principe. "Kernel adaptive filtering with maximum correntropy criterion." Neural Networks,2011. IJCNN 2011. International Joint Conference on. IEEE, 2011. pp.2012-2017

[11] H. QU, W. MA, J. ZHAO, T. WANG. "Prediction method for network traffic based on maximum correntropy criterion." *China Communications* 10.1 (2013): pp.135-142.

[12] B.W. Silverman, Density estimation for statistics and data analysis, vol. 3, Chapman and hall London, 1986.

[13] M.C. Jones, J.S. Marron, and S.J. Sheather, "A brief survey of bandwidth selection for density estimation," Journal of the American Statistical Association, 91.433(1996): pp. 401–407.

[14] A.W. Bowman, "An alternative method of cross-validation for the smoothing of density estimates,"Biometrika, 71.2(1984): pp. 353–360.

[15] A. Singh, J.C. Príncipe. "Information theoretic learning with adaptive kernels." Signal Processing 91.2 (2011):pp.203-213.

[16] A.R.C. Paiva and J.C. Principe, "A fixed point update for kernel width adaptation in information theoretic criteria." Machine Learning for Signal Processing (MLSP), 2010 IEEE International Workshop on. IEEE, 2010, pp. 262–265.

[17] S. Zhao, B. Chen, J.C. Principe. "An adaptive kernel width update for correntropy." Neural Networks (IJCNN), The 2012 International Joint Conference on. IEEE, 2012, pp. 1-5.

[18] B. Chen, J.C. Principe. "Maximum correntropy estimation is a smoothed MAP estimation." IEEE Signal Processing Letters, 19.8 (2012), pp.491-494.

[19] B. Chen, L. Xing, J. Liang, N. Zheng, J.C. Principe. "Steady-state Mean-square Error Analysis for Adaptive Filtering under the Maximum Correntropy Criterion." IEEE Signal Processing Letters, 21.7 (2014), pp.880-884.

[20] W. Wang, J. Zhao, H. Qu, B. Chen, J.C. Principe. "A switch kernel width method of correntropy for channel estimation." Neural Networks (IJCNN), The 2015 International Joint Conference on. IEEE, July 12-17, 2015 Killarney, Ireland.

[21] V.J. Mathews, S.H. Cho. "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm. " Acoustics, Speech and Signal Processing, IEEE Transactions on, 35.4 (1987): 450-454.

[22] Y. Li, N. Seshadri, S. Ariyavisitakul. "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels."Selected Areas in Communications, IEEE Journal on 17.3 (1999): pp. 461-471.

[23] M.M. Rana, M.K. Hosain. "Adaptive channel estimation techniques for MIMO OFDM systems." International Journal of Advanced Computer Science and Applications 1.6 (2010): pp.134-138.

[24] M. Hossain, S.M. Farhad, and M.T. Riasat. "Performance analysis of RLS and VSS-LMS channel estimation techniques for 4G MIMO OFDM systems." Computer and Information Technology (ICCIT), 2012 15th International Conference on. IEEE, 2012, pp.267-270.