Sparsity-Aware Recursive Maximum Correntropy Criteria Adaptive Filtering Algorithm

Wentao Ma[1], Badong Chen[2], Jiandong Duan[1], Hongyun Wei[3], and Guan Gui[4]
1. Department of Electronic Engineering, Xi'an University of Technology, Xi'an, China
2. School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China
3. Department of Math and Natural Sciences, Akita International University, Akita, Japan
4. College of Telecommunications and Information Engineering, NUPT, Nanjing 210003, China
Email: xjtu.wentaoma@gmail.com

Abstract—**To address sparse channel estimation problem in non-Gaussian impulsive noise environment, a recursive maximum correntropy criteria (RMCC) algorithm using sparse constraint is proposed to combat impulsive-inducing instability. Specifically, the recursive algorithm on the correntrioy with a forgetting factor of error at iteration is to solve steady-state error for improving the maximum correntropy criteria (MCC) based algorithms. Considering an unknown sparse channel, the simple and efficient zero-attracting is employed in the RMCC algorithm to exploit sparsity as well as to mitigate the impulsive noise simultaneously. Numerical simulations are given to show that the proposed algorithm is robust while providing robust steady-state estimation performance.**

Keywords—Recursive correnyropy criterion algorithm; l_1 norm constraint; sparse parameter estimation; non-Gaussian impulsive noise.

I. INTRODUCTION

In recent years, more and more sparse adaptive filtering algorithms have been developed to estimate sparse systems, such as multipath channel estimation, underwater acoustic (UWA) channel probing as well as echo cancellation [1-2]. Two kinds of adaptive filtering algorithms, least mean square (LMS) and recursive least squares (RLS), are very simple and hence they can be easily implemented in real systems. LMS-type sparse adaptive filtering algorithms [3-5] and its variations [6-8] have been developed to address the aforementioned problems. To the best of our knowledge, however, the development of RLS-based algorithms is very limited.

Sparsity-aware regularized least square is presented in [9], where the RLS cost function is regularized by adding a weighted ℓ_1 -norm sparse constraint. RLS algorithm with convex regularization is developed, which the regularizing term is defined as a general convex function of the system estimate [10]. Furthermore, ℓ_1-norm and reweighted ℓ_1-norm based RLS are developed for sparse system identification in [11-12]. Under the impulsive noise environments, however, these methods results in poor performance due to the fact that the mean square error (MSE) criterion assumes the Gaussian distributed error. Indeed, many physical experiments have confirmed that impulsive noises are often occurred in many systems such as man-made low frequency atmospheric noise systems and underwater acoustic systems [13-15].

Recent years, sparse channel estimation methods have been investigated intensely. One of the efforts is searching for more robust cost function because it is well know that MSE based cost function is sensitive to impulsive noise for the square operation. A number of p-norm based sparse channel parameter estimation methods are proposed to alleviate the negative effect of impulsive noise [16-17]. These methods, such as sparse LMP [16] and sparse RLP [17], adopt the gradient decent and recursive form to obtain the optimal parameter. Minimum error entropy (MEE) [18] and maximum correntropy criterion (MCC) [19] are two robust cost functions which can be derived in information theoretic learning [20]. Sparsity-aware MEE and MCC algorithms [21-22] are developed to estimate the channel parameter issue under impulsive noise environment.

In this paper, we propose zero-attracting recursive MCC algorithm (ZA-RMCC) in a manner alike to the approach as outlined in [11]. Recursive MCC algorithm and its kernel version are proposed in [23-24], which perform robustly. Simulation results are given to verify the proposed algorithm.

The rest of the paper is organized as follows. In Section II, the correntropy with forgetting factor is briefly reviewed. In Section III, sparse RMCC algorithm, i.e., ZARMCC, is derived. In Section IV, simulation results are provided to verify the proposed algorithms. Finally, conclusion is given in section V.

II. REVIEW OF CORRENTROPY

Definition 1. The correntropy between any two random variables X and Y is a generalized similarity measure defined as

$$V(X,Y) = E\left(\kappa_\sigma\left(X-Y\right)\right) = \iint_{x,y} \kappa_\sigma\left(X-Y\right) f_{XY}(x,y)dxdy, \quad (1)$$

where $E(\cdot)$ denotes the mathematical expectation, κ_σ is a kernel function with kernel size σ , $f_{XY}(x,y)$ is the PDF of X and Y. The formula can be extended to arbitrary dimensions. In practice, the given sample data of random variable is finite and the PDF is unknown, leading to the sample estimator of Correntropy by the parzen window estimation as definition 2. Here, the finite samples $\{(x_i, y_i)\}_{i=1}^N$ of the variables X and Y are considered.

Definition 2. Give two arbitrary random variables *X* and *Y*, the estimator of correntropy can be defined as

$$\hat{V}(X,Y)=\frac{1}{N}\sum_{i=1}^{N}\kappa_\sigma(x_i-y_i)=\frac{1}{N}\sum_{i=1}^{N}\kappa_\sigma(e_i), \tag{2}$$

where $e_i=x_i-y_i$. The Correntropy as a similarity function is a measure of how similar two random variables are, within a small neighborhood determined by the kernel size σ. As a cost function, it has been used to develop adaptive filter [25-27]. Based on the above definition and the weight average, the weight or namely forgetting factor is introduced, a new estimator with forgetting factor of correntropy is defined in [23].

Definition 3: The new estimator with forgetting factor of correntropy, namely correntropy-FF $C_{ff}(X,Y)$ as

$$C_{ff}(X,Y)=\sum_{i=1}^{n}\lambda^{n-i}k_\sigma(x_i-y_i),(0<\lambda<1), \tag{3}$$

where λ is the forgetting factor in the correntropy-FF, which is a small positive constant value very close to zero.

III. Sparse Recursive Maximum Correntropy Criteria Algorithms

A. System descriptions

Sparse system parameter model is considered as follows. The input vector $X(n)=[x_n,x_{n-1},\cdots,x_{n-M+1}]^T$ is sent over the FIR system with parameter vector $W_o=[w_{o,1},w_{o,2},\cdots,w_{o,\mathrm{M}}]^T$ with sparse form, where M is the size of the channel memory. It is assumed that the channel parameters are real-valued, and most of them are zero. The received signal $d(n)$ is modeled as

$$d(n)=W_o^TX(n)+v(n), \tag{4}$$

where $v(n)$ is the additive background noise that is independent of $X(n)$.

B. Recursive maximum correntropy criteria algorithm with l1 norm constraints

We define the new cost function employing the Correntropy -FF (using Gaussian kernel) and the l1 norm as

$$Cost(n)=\frac{1}{\sqrt{2\pi}\sigma}\sum_{i=1}^{n}\lambda^{n-i}\exp\left(-\frac{(d_i-W^T(n)X(i))^2}{2\sigma^2}\right)+\rho g(n) \tag{5}$$

where $X(i)$ is the input data, $W(n)=[w_1(n),\cdots,w_M(n)]^T$ is estimation weights at the *n*-th time instant. The vector $g(n)=\|W(n)\|_1$ is the ℓ_1-norm of current tab estimate as a sparsity term, ρ is the regularization parameter to balance the updating MCC and sparse penalty of the *n*-th updated channel estimator $W(n)$. The definition of ℓ_1-norm of the weight vector is

$$\|W(n)\|_1=\sum_{k=1}^{M}|w_k(n)|. \tag{6}$$

From above analysis, the optimal parameter W_{opt} can make the cost function maximum. That is,

$$W_{opt}=\arg\max_w Cost(n). \tag{7}$$

Here, the gradient of *Cost*(*n*) can be computed to obtain the optimal solution of cost function, and let it equal to zero.

$$\partial Cost(n)/\partial W=0. \tag{8}$$

By combining (5) and (8), we have

$$\begin{aligned}&\frac{1}{\sqrt{2\pi}\sigma^3}\sum_{i=0}^{n}\lambda^{n-i}\exp\left(-{(d(i)-W(n)^TX(i))^2}\big/{2\sigma^2}\right)\\&\cdot\left(d(i)-W(n)^TX(i)\right)X(i)=-\rho g'(n),\end{aligned} \tag{9}$$

where $(\cdot)^T$ is transpose operator, and $g'(n)=\mathrm{sign}(W(n))$. We employ the idea of the recursive as the RLS to obtain W_o. From (9), we have

$$\begin{aligned}&\sum_{i=0}^{n}\lambda^{n-i}\exp\left(-{(d(i)-W(n)^TX(i))^2}\big/{2\sigma^2}\right)X(i)X^T(i)W(n)\\&=\sum_{i=0}^{n}\lambda^{n-i}\exp\left(-{(d(i)-W(n)^TX(i))^2}\big/{2\sigma^2}\right)X(i)d(i)+\rho g'(\mathrm{n}).\end{aligned} \tag{10}$$

Both sides of the equation (9) have similar structure, we define $\Phi(n)$ and $\Psi(n)$ as formula (11) respectively.

$$\begin{aligned}\Psi(n)&=\sum_{i=0}^{n}\lambda^{n-i}\exp\left(-{(d(i)-W(n)^TX(i))^2}\big/{2\sigma^2}\right)u(i)u^T(i),\\\Phi(n)&=\sum_{i=0}^{n}\lambda^{n-i}\exp\left(-{(d(i)-W(n)^TX(i))^2}\big/{2\sigma^2}\right)u(i)d(i).\end{aligned} \tag{11}$$

One can easily see that an extra exponential factor is included in $\Phi(n)$ and $\Psi(n)$, which is introduced by the correntropy-FF. According to (11), Eq. (10) can be rewritten in matrix form as

$$\begin{aligned}&\Psi(n)W(n)=\Phi(n)+\rho g'(\mathrm{n})\Rightarrow\\&W(n)=\Psi^{-1}(\mathrm{n})(\Phi(n)+\rho g'(\mathrm{n}))=\Psi^{-1}(\mathrm{n})r(\mathrm{n}),\end{aligned} \tag{12}$$

where $r(\mathrm{n})=\Phi(n)+\rho g'(\mathrm{n})$. To avoid the difficulty of computing the inverse of $\Psi(n)$, we represent the $\Psi(n)$ recursively as

$$\begin{aligned}\Psi(n)=\lambda&\sum_{i=0}^{n-1}\lambda^{n-i-1}\exp\left(-\frac{(d(i)-W(n)^Tu(i))^2}{2\sigma^2}\right)X(i)X^T(i)\\&+\exp\left(-\frac{(d(n)-W(n)^Tu(n))^2}{2\sigma^2}\right)X(n)X^T(n).\end{aligned} \tag{13}$$

As the learning iterations increase, the $(n-1)$-th weight coefficient closes to n-th, i.e., $W(n-1) \to W(n)$, $n \to \infty$. Under this condition, we get the recursive form of $\Psi(n)$ as

$$\Psi(n) = \lambda \Psi(n-1) + \omega X(n) \omega X^T(n), \tag{14}$$

where

$$\omega = \sqrt{\exp(-(d(n) - W(n)^T X(n))^2 / 2\sigma^2)} \,. \tag{15}$$

Similarly, the recursive form of $\Phi(n)$ is induced as

$$\Phi(n) = \lambda \Phi(n-1) + \omega^2 X(n) d(n), \tag{16}$$

Combining (12) and (16), yields

$$\begin{aligned} r(\mathrm{n}) &= \Phi(n) + \rho g'(\mathrm{n}) \\ &= \lambda \Phi(n-1) + \omega^2 X(n) d(n) + \rho g'(\mathrm{n}) \\ &= \lambda \Phi(n-1) + \lambda \rho g'(\mathrm{n-1}) \\ &\qquad + \omega^2 X(n) d(n) + \rho g'(\mathrm{n}) - \lambda \rho g'(\mathrm{n-1}) \\ &= \lambda r(\mathrm{n-1}) + \omega^2 X(n) d(n) + \rho g'(\mathrm{n}) - \lambda \rho g'(\mathrm{n-1}). \end{aligned} \tag{17}$$

To this end, we assume that the sign of the weight values do not change significantly in a single time step. Hence, we approximate (17) by

$$r(n) = \lambda r(n-1) + \omega^2 X(n) d(n) + \rho(1-\lambda) g'(n-1). \tag{18}$$

Now, we give a matrix form of equation (14), first some new symbols can be defined as:

$$A = \Psi(\mathrm{n}), B^{-1} = \lambda \Psi(n-1), C = \omega u(n), D = I,$$

where I is the unit matrix. Thus, the matrix form of equation (14) becomes

$$A = B^{-1} + C D^{-1} C^T.$$

Using the matrix inversion theorem and considering the equation (14), we get the inverse of $\Psi(n)$ as

$$\Psi^{-1}(n) = \lambda^{-1} \Psi^{-1}(n-1) - \frac{\omega^2 \lambda^{-1} \Psi^{-1}(n-1) X(n) X^T(n) \Psi^{-1}(n-1)}{\lambda + \omega^2 X^T(n) \Psi^{-1}(n-1) X(n)}. \tag{19}$$

For simple description of equation (18), denoting $\Omega(n)$ and $\mathrm{K}(n)$ as follows

$$\Omega(n) = \Psi^{-1}(n) \,, \mathrm{K}(n) = \frac{\Omega(n-1) X(n)}{\lambda + \omega^2 X^T(n) \Omega(n\text{-}1) X(n)}. \tag{20}$$

From the above definition, we get the new form of (19) as

$$\Omega(n) = \lambda^{-1} [\Omega(n-1) - \omega^2 \mathrm{K}(n) X^T(n) \Omega(n-1)], \tag{21}$$

where $K(n)$ and $\Omega(n)$ are similar to the definition in RLS, and they can be defined as the extended kalman gain vector. After some tedious computation, we get

$$\begin{aligned} W(n) &= \Psi^{-1}(n) r(n) \\ &= \Omega(n)[\lambda r(n-1) + \omega^2 X(n) d(n) + \rho(1-\lambda) g'(\mathrm{n-1})] \\ &= \{\lambda^{-1}[\Omega(n-1) - \omega^2 \mathrm{K}(n) X^T(n) \Omega(n-1)]\} \\ &\quad \{[\lambda r(n-1) + \omega^2 X(n) d(n) + \rho(1-\lambda) g'(n-1)]\} \\ &= W(n-1) + \omega^2 \mathrm{K}(n)[d(n) - X^T(\mathrm{n}) \mathrm{W}(n-1)] \\ &\quad + \rho \lambda^{-1}(1-\lambda)[1 - \omega^2 \mathrm{K}(n) X^T(n)] \Omega(n-1) g'(n-1) \\ &= W(n-1) + \omega^2 \mathrm{K}(n) \mathrm{e}(n) \\ &\quad + \rho \lambda^{-1}(1-\lambda)[I - \omega^2 \mathrm{K}(n) X^T(n)] \Omega(n-1) g'(n-1) \end{aligned} \tag{22}$$

where $e(n) = d(n) - X^T(n) W(n-1)$. Now, the recursive MCC with l_1 norm penalty algorithm is developed, and we denote it as ZARMCC. We summarized this algorithm in Tab. 1.

Table 1. ZARMCC algorithm

Initialization: W(0)=0, $\Omega(0) = \delta^{-1} I$, ρ, λ, σ
For n=1,2,$\cdots$ Do
$y(n) = W^T(\mathrm{n}-1) X(n)$
$e(n) = d(n) - y(n)$
$\mathrm{K}(\mathrm{n}) = \dfrac{\Omega(\mathrm{n\text{-}1}) X(n)}{\lambda + \omega^2 X^T(n) \Omega(n-1) X(n)}$
$W(n) = W(n-1) + \omega^2 \mathrm{K}(n) \mathrm{e}(n) + \rho \lambda^{-1}(1-\lambda)[I - \omega^2 \mathrm{K}(n) X^T(\mathrm{n})] \Omega(n-1) g'(\mathrm{n-1})$
$\Omega(n) = \lambda^{-1}[\Omega(\mathrm{n}-1) - \omega^2 \mathrm{K}(n) X^T(\mathrm{n}) \Omega(\mathrm{n\text{-}1})]$
End

Remark : On the one hand, when $\omega^2 = 1$, the ZARMCC will be reduced to ZARLS[11]. On the other hand, when $\rho = 0$, it will be reduced to the RMCC algorithm[23].

IV. NUMERICAL SIMULATION

In this section, we evaluate the proposed ZARMCC algorithm with respect to RLS, ZARLS, MCC, ZALMP, ZAMCC, and ZAMEE algorithms. The sparse channel estimation problems are conducted under the impulsive noise environments. Simulation curves are obtained by averaging over 100 independent Monte Carlo (MC) runs and 10000 iterations are run for each MC. The convergence is evaluated by mean square deviation (MSD) which is calculated by

$$\mathrm{MSD}(n) = E\left\{ \left\| W_o - W(n) \right\|^2 \right\}, \tag{23}$$

In the following simulation, the α-stable distribution is considered to model the impulsive noise，which provides a good model for such heavy-tailed noises [28]. The characteristic function of the α -stable process is given by

$$f(t) = \exp\left\{ j\delta t - \gamma |t|^{\alpha} [1 + j\beta \operatorname{sgn}(t) S(t,\alpha)] \right\}, \tag{24}$$

in which

$$S(t,\alpha) = \begin{cases} \tan\dfrac{\alpha\pi}{2} & if\ \alpha \neq 1 \\ \dfrac{2}{\pi}\log|t| & if\ \alpha = 1 \end{cases} \tag{25}$$

where $\alpha \in (0,2]$ is the characteristic factor, $-\infty < \delta < +\infty$ is the location parameter, $\beta \in [-1,1]$ is the symmetry parameter, and $\gamma > 0$ is the dispersion parameter. The characteristic factor α measures the tail heaviness of the distribution. The smaller α is, the heavier the tail is. The parameters vector of the noise model is defined as $V = (\alpha, \beta, \gamma, \delta)$.

The input signal is assumed to be a white Gaussian process with zero mean and unit variance. Assume that the channel memory size is $M = 20$, and the impulsive response of the sparse channel model is illustrated in Fig.1.

TAB. I.SIMULATION PARAMETERS FOR ALL MENTIONED ALGORITHMS

Parameter / Algorithm	μ	ρ	L(Window length)	λ	δ	σ
ZALMP	1e-2	1e-4	N/A	N/A	N/A	N/A
ZAMEE	1e-2	1e-4	20	N/A	N/A	1
ZAMCC	1e-2	1e-4	N/A	N/A	N/A	1
ZARMCC	N/A	1e-4	N/A	0.99	1e-5	1
ZARLS	N/A	1e-4	N/A	0.99	1e-5	N/A
RMCC	N/A	N/A	N/A	0.99	1e-5	1
MCC	1e-2	N/A	N/A	N/A	N/A	1
RLS	N/A	N/A	N/A	0.99	1e-5	N/A

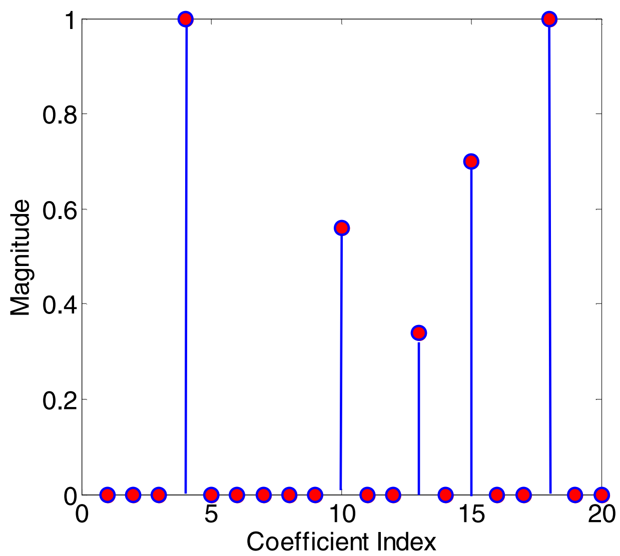

Fig.1. 20-length sparse channel supporting by 5 nonzero coefficients.

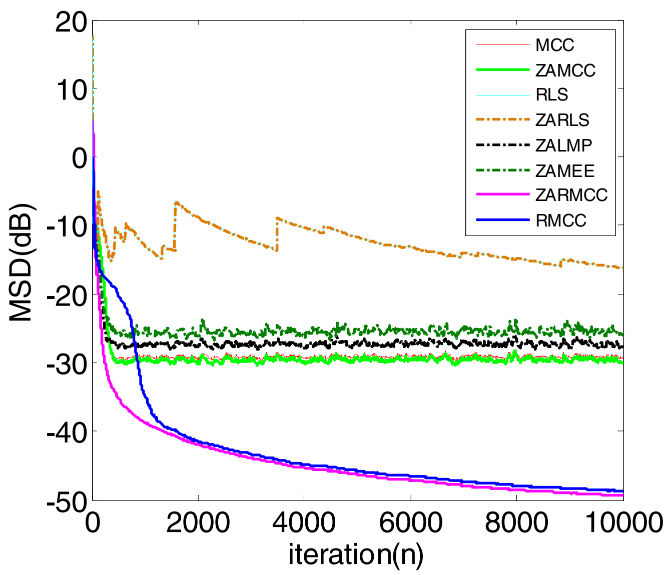

Fig. 2. Tracking and steady-state behaviours.

Experiment 1: Performance evaluation of existing methods

First, we show a comparison among all the proposed algorithms in term of MSD. All parameters values of the mentioned algorithms are listed in Tab.1, which values are set by scanning for the best results. The noise parameter vector is $V = (1.2, 0, 0.2, 0)$. One can find that sparse type algorithms achieve faster convergence rate and better steady-state performance than the standard algorithms (LMP, MEE and MCC). In addition, Fig. 2 also shows that proposed ZARMCC achieves lower MSD than other ZA version algorithms, which illustrate the advantage of the recursive MCC. It is worth notice that both RLS and ZARLS are unstable which is caused by impulsive noises. In the subsequent experiments, unstable RLS-type algorithms are omitted.

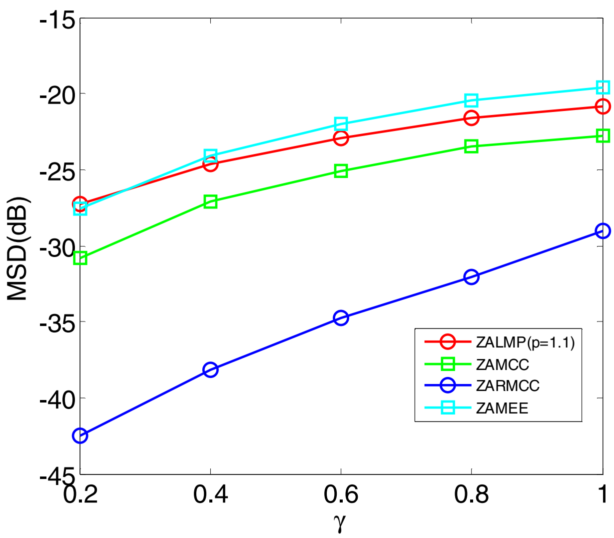

Fig. 3. Performance evaluation versus parameter γ.

In the second example, we show the robust property of the proposed algorithm in terms of the dispersion parameter γ. Note that the parameters are set as the same as the first simulation for all mentioned algorithms above. The Steady-state MSD (ssMSD) is calculated by the mean value of the last 1000 iterations. Fig.3 illustrates the steady-state MSD versus different γ (0.2 0.4 0.6 0.8 1), where the parameter α is fixed at 1.2. This result confirms that the ZARMCC achieve a better accuracy than other algorithms.

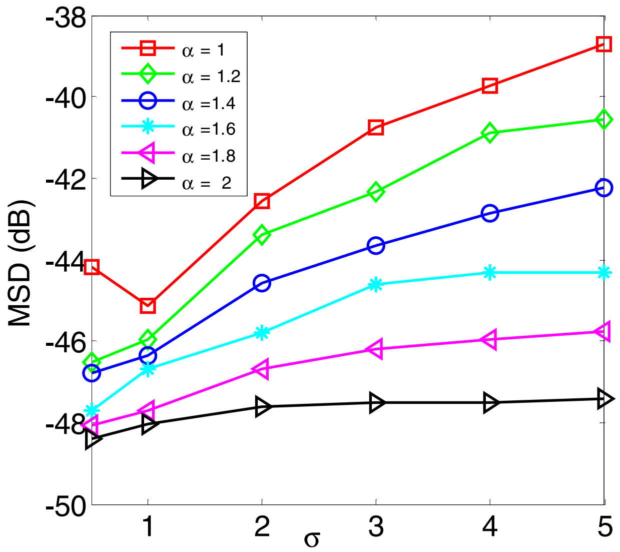

Fig.4. Steady-state MSD of ZARMCC for variations in σ and α.

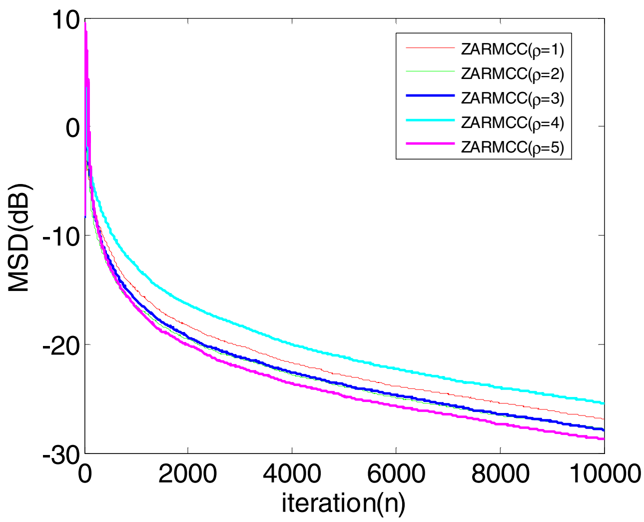

Fig. 5. Performance evaluation versus parameter ρ.

Experiment 2: ZARMCC against different parameter

Firstly, we investigate the MSD of the kernel size (0.5, 1, 2, 3, 4, and 5) for different α (1, 1.2, 1.4, 1.5, 1.6, 1.8, and 2). The sensitivity of the ZARMCC for a range of above parameter values within the algorithm are shown in Fig.4. We observe that for different σ, a more impulsive noise (smaller α) leads to a larger MSD. Further, the lowest MSD is obtained when α is 2, which means that the noise is reduced to Gaussian noise.Secondly, we evaluate the performance of the ZARMCC in view of different regularization parameters ρ. Performance curves of ZARMCC is depicted in Fig.5. One can observe that MSD performance is near optimal with $\rho = 5$.According to this result, we should choose the suitable regularization parameter for the proposed ZARMCC to get the best performance.

V. Conclusion

In this paper, we have proposed RMCC algorithms with ℓ_1 norm constrained, namely ZARMCC, which can be applied in sparse system identification under impulsive noise environment. The closed form of this algorithm is developed by regularizing the combination of the correntropy with forgetting factor and the ℓ_1 norm sparse constraint. Simulation results demonstrate that the proposed method can achieve robust steady-state error when the sparse system is under impulsive noise environments.

Acknowledgment

This work was supported in part by Japan Society for the Promotion of Science (JSPS) research grants (No. 15K06072, No. 26889050) as well as the National Natural Science Foundation of China grants (No. 61401069, No. 61271240).

References

[1] N.R. Yousef, A.H. Sayed, N., Khajehnouri., "Detection of fading overlapping multipath components," *Signal Processing,* vol.86, no.9, pp.2407-2425,2006.

[2] A.C. Singer, J.K. Nelson, S.S. Kozat.,"Signal processing for underwater acoustic communications," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 90-96, 2009.

[3] Chen Y, Gu Y, Hero A O.,"Sparse LMS for system identification," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3125-3128, 2009.

[4] Su G, Jin J, Gu Y, et al., "Performance analysis of lp-norm constraint least meansquare algorithm," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2223-2235, 2012

[5] Taheri O, Vorobyov S A.,"Sparse channel estimation with lp-norm and reweighted l1-norm penalized least mean squares," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2864-2867,2011.

[6] Gui G, Mehbodniya A, Adachi F.,"Least mean square/fourth algorithm for adaptive sparse channel estimation," *IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pp. 296-300, 2013.

[7] Salman M S.,"Sparse leaky-LMS algorithm for system identification and its convergence analysis," *International Journal of Adaptive Control and Signal Processing*,vol.28,no.10, pp.1065-1072, 2014.

[8] Jahromi M N S, Salman M S, Hocanin A. , "Convergence analysis of the zero-attracting variable step-size LMS algorithm for sparse system identification," *Signal, Image and Video Processing*, pp. 1-4, 2013.

[9] Babadi B, Kalouptsidis N, Tarokh V., "SPARLS: The sparse RLS algorithm," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4013-4025, 2010.

[10] Eksioglu E M., "RLS algorithm with convex regularization," *IEEE Signal Processing Letters*, vol. 18, no. 8, pp. 470-473,2011.

[11] Eksioglu E M., "RLS adaptive filtering with sparsity regularization," *IEEE 10th International Conference on. Information Sciences Signal Processing and their Applications (ISSPA)*, pp. 550-553, 2010.

[12] Eksioglu E M., "Sparsity regularised recursive least squares adaptive filtering". *IET Signal Processing* , vol. 5, no. 5, pp. 480-487, 2011.

[13] B. Mandelbrot and I. W. Van Ness, "Fractional brownian motions, fractional noises, and applications," *SIAM Review*, vol. 10, pp. 422-437, 1968.
[14] S. S. Pillai and M. Harisankar, "Simulated performance of a DS spread spectrum system in impulsive atmospheric noise," *IEEE Transactions on Electromagnetic Compability*, vol. 29, pp. 80-82, 1987.
[15] M. Bouvet and S. C. Schwartz," Comparison of adaptive and robust receivers for signal detection in ambient underwater noise," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, pp. 621-626, 1989.
[16] Wentao Ma, Badong Chen, Hua Qu, Jihong, Zhao, "Sparse least mean *p*-power algorithms for channel estimation in the presence of impulsive noise," *Signal image and video processing*, DOI : 10.1007/s11760-015-0757-5 .
[17] Pelekanakis K, Chitre M., "Adaptive sparse channel estimation under symmetric alpha-stable noise," *IEEE Transactions on Wireless Communications*, vol.13,no.6,pp. 3183-3195, 2014.
[18] Erdogmus D, Principe J C., "Convergence properties and data efficiency of the minimum error entropy criterion in adaline training," *IEEE Transactions on Signal Processing* , pp. 51, no. 7, pp. 1966-1978, 2003.
[19] Weifeng Liu, Puskal P. Pokharel, Jose C. Principe.,"Correntropy: Properties and Applications in Non-Gaussian Signal Processing,". *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286-5298, 2007.
[20] Principe J C. Information theoretic learning: Renyi's entropy and kernel perspectives. *Springer Science & Business Media*, 2010..
[21] Wentao Ma, Hua Qu, Jihong Zhao, Badong Chen, Jose C. Principe. ,"Sparsity aware minimum error entropy algorithms," *The 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, pp.19-24, 2015.
[22] Wentao Ma, Hua Qu, Guan Gui, Jihong Zhao, Li Xu, Badong Chen, "Maximum Correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments," *Journal of The Franklin Institute*, 2015,352 (7):2708-2727.
[23] Ma W, Qu H, Zhao J., "Estimator with forgetting factor of correntropy and recursive algorithm for traffic network prediction," *25th IEEE Chinese Control and Decision Conference (CCDC),* pp. 490-494, 2013.
[24] Wu Z, Shi J, Zhang X, et al, "Kernel recursive maximum correntropy," Signal Processing, 2015 (DOI: 10.1016/j.sigpro.2015.04.024).
[25] Badong Chen, Lei Xing, Junli Liang, Nanning Zheng, Jose C. Principe, "Steady-state mean-square error analysis for adaptive filtering under the maximum correntropy criterion,"*IEEE Signal Processing Letters*, vol. 21, no. 7, pp. 880-884, 2014
[26] Singh A, Principe J C. , "Using correntropy as a cost function in linear adaptive filters," *IEEE International Joint Conference on Neural Networks, (IJCNN)* pp. 2950-2955, 2009.
[27] Zhao S, Chen B, Principe J C., "Kernel adaptive filtering with maximum correntropy criterion," *The 2011 International Joint Conference on Neural Networks (IJCNN)*, *IEEE*, pp. 2012-2017, 2011.
[28] Shao M, Nikias C L.,"Signal processing with fractional lower order moments: stable processes and their applications," *Proceedings of the IEEE*, vol. 81,no. 7,pp. 986-1010,1993.