SPARSITY AWARE MINIMUM ERROR ENTROPY ALGORITHMS

Wentao Ma[1], Hua Qu[1], Jihong Zhao[1], Badong Chen[1], Jose C. Principe[1,2]

[1]School of EIE, Xi'an Jiaotong University, Xi'an, 710049, China

[2] Department of ECE, University of Florida, Gainesville, FL 32611, USA

xjtu.wentaoma@gmail.com, chenbd@mail.xjtu.edu.cn

ABSTRACT

Sparse estimation has received a lot of attention due to its broad applicability. In sparse channel estimation, the parameter vector with sparsity characteristic can be well estimated from noisy measurements through sparse adaptive filters. In previous studies, most works use the mean square error (MSE) based cost to develop sparse filters, which is rational under the assumption of Gaussian distributions. However, Gaussian assumption does not always hold in real-world environments. To address this issue, we incorporate in this work l_1-norm and reweighted l_1-norm into the minimum error entropy (MEE) criterion to develop new sparse adaptive filters, which may perform much better than the MSE based methods especially in non-Gaussian situations, since the error entropy can capture higher-order statistics of the errors. Furthermore, a new approximator of l_0-norm based on the Correntropy Induced Metric (CIM) is also used as a sparsity penalty term (SPT). Simulation results show the excellent performance of the proposed algorithms.

Index Terms—Sparse estimation, minimum error entropy, correntropy induced metric, impulsive noise

1. INTRODUCTION

Sparse channel estimation is to estimate a parameter vector of a channel with most of zero tap under noisy environment, which is in general based on a traditional adaptive filter with a sparsity penalty term (SPT). In recent years, many sparse adaptive filters have been developed for sparse systems identification. Typical examples include sparse least mean square (LMS) [1-4], sparse recursive least square (RLS) [5], and their variations [6-12].

However, there are some limitations of the existing sparse adaptive filters. When data are non-Gaussian (especially when data are disturbed by impulsive noise or containing outliers), they may perform very poorly. The main reason for this is that most of the existing algorithms are developed based on the MSE criterion, which relies heavily on the assumptions of Gaussian distributions. This assumption does not always hold particularly in most practical applications. For instance, different types of artificial noises in electronic devices, atmospheric noises, and lighting spikes in natural phenomena, can be described more accurately using non-Gaussian noise models [13, 14]. When sparse filters are applied in such situations, the performance will become much worse due to the sensitivity to the impulsive noises or outliers [15].

Information theoretic learning (ITL), on the other hand, provides a nice approach for dealing with non-Gaussian signal processing [16,17]. The minimum error entropy (MEE) criterion in ITL was successfully used in adaptive filtering to improve the learning performance in non-Gaussian noises [18-26]. However, to the best of our knowledge, the MEE has not yet been extended to the sparse channel estimation. In this work, we use the MEE instead of the MSE to develop sparse adaptive filters. The new adaptive filters are much more robust against the impulsive non-Gaussian noises.

As an important part, the SPT in sparse filters enable them to fit well the sparse structures of the channel systems. Finding the sparsest solution leads to the l_0-norm minimization, an NP-hard problem. In existing methods, the l_1-norm or reweighted l_1-norm are frequently used as the SPT. As a nice approximator of the l_0-norm, the Correntropy Induced Metric (CIM) can also be used as a sparsity penalty term in sparse estimation [27, 28]. In the present paper, we will incorporate the above mentioned SPTs (l_1-norm, reweighted l_1-norm and CIM) into the sparse aware MEE algorithms.

2. MEE AND CIM

2.1. Minimum Error Entropy Criterion

Consider a channel model, where the input vector $X(n)=[x_n, x_{n-1}, \cdots, x_{n-M+1})]^T$ at time n is sent over an FIR channel with parameter vector $W^*=[w_1^*, w_2^*, \cdots, w_M^*]^T$ (M is the size of the channel memory). Assume that the channel parameters are real-valued, and most of them are zero. The received signal $d(n)$ is then

$$d(n)=W^{*T}X(n)+v(n) \tag{1}$$

where $v(n)$ denotes an interference noise. Let $W(n)=[w_1(n), w_2(n), \cdots, w_M(n)]^T$ be the weight vector of an adaptive filter. The instantaneous error can be calculated as $e(n)=d(n)-y(n)$,where $y(n)=W^T(n)X(n)$ is the filter output. Based on Parzen window approach, the

This work was supported by National Natural Science Foundation of China (61371807, 61372152)

probability density function (PDF) of the error can be estimated as [16,17]

$$\hat{f}_e(\mathrm{e}) = \frac{1}{N}\sum_{i=1}^{N}\kappa_\sigma(\mathrm{e}-\mathrm{e}_i) \tag{2}$$

where $\kappa_\sigma(\cdot)$ denotes a kernel function with bandwidth σ. Gaussian kernel function is one of the most popular kernels, which is given by

$$\kappa_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{x^2}{2\sigma^2}) \tag{3}$$

Renyi's quadratic entropy estimator for a set of error samples can be expressed as [16,17]

$$H_{R2}(\mathrm{e}) = -\log\int \hat{f}_e^2 de = -\log V(\mathrm{e}) \tag{4}$$

$$V(\mathrm{e}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\kappa_{\sqrt{2}\sigma}(e_j - e_i) \tag{5}$$

The argument ($V(\mathrm{e})$) of the 'log' in (4) is called the information potential. Obviously, minimizing the error entropy is equivalent to maximizing the information potential. Thus, the optimization cost for MEE can be

$$J_{MEE}(\mathrm{e}) = \max_W V(\mathrm{e}) \tag{6}$$

A steepest ascent algorithm for estimating the weight vector can be derived as

$$W(n+1) = W(n) + \eta\nabla V(\mathrm{e}) \tag{7}$$

where η denotes a step size, and $\nabla V(\mathrm{e})$ stands for the gradient of the information potential with respect to the weight vector, expressed as

$$\nabla V(\mathrm{e}) = \frac{1}{2N^2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\kappa_{\sqrt{2}\sigma}(e_i-e_j)(e_i-e_j)\left(\frac{\partial y_i}{\partial W}-\frac{\partial y_j}{\partial W}\right) \tag{8}$$

2.2. Correntropy Induced Metric

Given two vectors in a sample space: $X=[x_1,\cdots,x_N]^T$, $Y=[y_1,\cdots,y_N]^T$, the Correntropy Induced Metric (CIM) is defined as [27]

$$CIM(\mathrm{X},Y) = \left(\kappa(0)-\hat{V}(X,Y)\right)^{1/2} \tag{9}$$

where $\hat{\mathrm{V}}(X,Y) = \frac{1}{N}\sum_{i=1}^{N}\kappa_\sigma(x_i-y_i)$ is the estimation of the correntropy between X and Y (see [27] for the definition of correntropy), the kernel is assumed to be a Gaussian kernel with the kernel width σ ,and $\kappa(0)=1/\sigma\sqrt{2\pi}$. The CIM provides a nice approximation for the l_0-norm. Given a vector $X=[x_1,\cdots,x_N]^T$, the l_0-norm can be approximated by [27, 28]

$$\|X\|_0 \sim CIM^2(X,0) = \frac{\kappa(0)}{N}\sum_{i=1}^{N}(1-\exp(-\frac{x_i^2}{2\sigma^2})) \tag{10}$$

With Gaussian kernel, the CIM behaves like an l_2-norm when the two vectors are close, like an l_1-norm outside the l_2-norm zone, and like an l_0-norm as they go farther apart [27, 28]. It has been shown that if $|x_i|>\delta$, $\forall x_i\neq 0$, then as $\sigma\to 0$, one can get a solution arbitrarily close to that of the l_0-norm, where δ is a small positive number [28]. As an approximation of the l_0-norm, the CIM favors sparsity and can be used as a penalty term in sparse channel estimation.

3. SPARSE MEE ALGORITHMS

3.1. Sparse MEE with Zero-Attracting (l_1-norm) Penalty Term (ZAMEE)

To develop a sparse MEE algorithm with zero-attracting (l_1 norm) penalty term, we introduce the cost function

$$\begin{aligned} &J_{ZAMEE}(n) \\ &= -J_{MEE}(n) + \lambda J_{ZA}(n) \\ &= -\frac{1}{L^2}\sum_{i=n-L+1}^{n}\sum_{j=n-L+1}^{n}\kappa_{\sqrt{2}\sigma_1}(e_i-e_j) + \lambda\|\mathrm{W(n)}\|_1 \end{aligned} \tag{11}$$

where $J_{ZA}(\mathrm{n}) = \|\mathrm{W(n)})\|_1$ denotes the l_1-norm of the estimated parameter vector, L is the sliding data length (SDL), and σ_1 is the kernel width in MEE. In (11), the MEE term is robust to impulsive noises, and the ZA penalty term is a sparsity inducing term, and the two terms are balanced by a weight factor $\lambda\geq 0$.

Based on the cost function (11), one can derive the following adaptive algorithm:

$$\begin{aligned} &W(n+1) \\ &= W(n) - \eta\frac{\partial J_{ZAMEE}(n)}{\partial W(n)} \\ &= W(n) - \eta\left[-\frac{1}{2\sigma_1^2L^2}\sum_{i=n-L+1}^{n}\sum_{j=n-L+1}^{n}[(e_i-e_j)]\kappa_{\sqrt{2}\sigma_1}(e_i-e_j)\left[\frac{\partial y_i}{\partial W(\mathrm{n})}-\frac{\partial y_j}{\partial W(\mathrm{n})}\right]+\lambda sign(\mathrm{W(n)})\right] \\ &= W(n) + \frac{\eta}{2\sigma_1^2L^2}\sum_{i=n-L+1}^{n}\sum_{j=n-L+1}^{n}[(e_i-e_j)]\kappa_{\sqrt{2}\sigma_1}(e_i-e_j)\left[X(i)-X(j)\right] - \rho sign(\mathrm{W(n)}) \end{aligned} \tag{12}$$

where $\rho=\eta\lambda$ is the zero-attractor control factor, and $sign(\cdot)$ is a component-wise sign function. The algorithm (12) is referred to as the ZAMEE algorithm.

3.2. Sparse MEE with the Logarithmic Penalty Term

In this part, we derive a sparse MEE algorithm with a logarithmic penalty term which also can generate a zero attractor. The corresponding cost function is given by

$$\begin{aligned} &J_{RZAMEE}(n) \\ &= -J_{MEE}(n) + \lambda J_{RZA}(n) \\ &= -\frac{1}{L^2}\sum_{i=n-L+1}^{n}\sum_{j=n-L+1}^{n}\kappa_{\sqrt{2}\sigma_1}(e_i-e_j) + \lambda\sum_{i=1}^{M}\log(1+|\mathrm{w}_i|/\delta) \end{aligned} \tag{13}$$

where the log-sum penalty $\sum_{i=1}^{M}\log(1+|\mathrm{w}_i|/\delta)$ behaves more similarly to the l_0-norm than the l_1-norm $\|\mathrm{W}\|_1$, and δ is a positive number. Then, a gradient based adaptive algorithm can be easily derived as

$$\begin{aligned} &w_i(n+1) \\ &= w_i(n) - \eta\frac{\partial J_{RZAMEE}(n)}{\partial \mathrm{w}_i(n)} \\ &= \mathrm{w}_i(n) - \eta\left[-\frac{1}{2\sigma_1^2L^2}\sum_{i=n-L+1}^{n}\sum_{j=n-L+1}^{n}[(e_i-e_j)]\kappa_{\sqrt{2}\sigma}(e_i-e_j)\left[X(i)-X(j)\right]+\lambda\frac{\mathrm{sign}(\mathrm{w}_i(n))}{1+\delta'|\mathrm{w}_i(n)|}\right] \\ &= \mathrm{w}_i(n) + \frac{\eta}{2\sigma_1^2L^2}\sum_{i=n-L+1}^{n}\sum_{j=n-L+1}^{n}(e_i-e_j)]\kappa_{\sqrt{2}\sigma_1}(e_i-e_j)\left[X(i)-X(j)\right] - \rho\frac{\mathrm{sign}(\mathrm{w}_i(n))}{1+\delta'|\mathrm{w}_i(n)|} \end{aligned} \tag{14}$$

where $\delta' = \frac{1}{\delta}$. This algorithm is referred to as the RZAMEE algorithm.

3.3. Sparse MEE with CIM Penalty Term

One can also employ the CIM as a sparsity penalty term to develop a sparse MEE algorithm. A new cost function can be defined by

$$\begin{aligned} & J_{CIMMEE}(n) \\ & = -J_{MEE}(n) + \lambda J_{CIM}(n) \\ & = -\frac{1}{L^2}\sum_{i=n-L+1}^{n}\sum_{j=n-L+1}^{n}\kappa_{\sqrt{2}\sigma_1}(e_i - e_j) + \lambda\frac{1}{M\sigma_2\sqrt{2\pi}}\sum_{i=1}^{M}(1-\exp(-\frac{w_i(n)^2}{2\sigma_2^2})) \end{aligned} \tag{15}$$

where σ_2 denotes the kernel width in CIM. The second term (i.e. the CIM) with a smaller kernel width will become a sparsity inducing term. Based on the new cost function of (15), we derive a gradient based adaptive algorithm

$$\begin{aligned} & W(n+1) \\ & = W(n) - \eta\frac{\partial J_{CIMMEE}(n)}{\partial W(n)} \\ & = W(n) - \eta\left[-\frac{1}{2\sigma_1^2 L^2}\sum_{i=n-L+1}^{n}\sum_{j=n-L+1}^{n}[(e_i - e_j)]\kappa_{\sqrt{2}\sigma_1}(e_i - e_j)[X(i) - X(j)] + \lambda\frac{1}{M\sigma_2^3\sqrt{2\pi}}W(n).*\exp(-\frac{W(n)^2}{2\sigma_2^2})\right] \\ & = W(n) + \frac{\eta}{2\sigma_1^2 L^2}\sum_{i=n-L+1}^{n}\sum_{j=n-L+1}^{n}[(e_i - e_j)]\kappa_{\sqrt{2}\sigma_1}(e_i - e_j)[X(i) - X(j)] - \rho\frac{1}{M\sigma_2^3\sqrt{2\pi}}W(n).*\exp(-\frac{W(n)^2}{2\sigma_2^2}) \end{aligned} \tag{16}$$

The above algorithm is referred to as the CIMMEE algorithm. The kernel width σ_2 is a key parameter in the penalty term. A proper kernel width will make CIM be a good approximator for the l_0-norm [27, 28].

4. SIMULATION RESULTS

In this section, we perform simulations on time-varying channel estimation to demonstrate the performance of the proposed sparse aware MEE algorithms (ZAMEE, RZAMEE, and CIMMEE), compared with several other algorithms including least absolute deviation (LAD) [29], MEE, ZALMS, and RZALMS. The parameter vector of the unknown channel is assumed to be

$$W^* = \begin{cases} [0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0] & n \le 2000 \\ [1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0] & 2000 < n \le 3000 \\ [1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1,1,-1] & 3000 < n \end{cases} \tag{17}$$

In (17), the channel memory size M is 20. The channel model has a sparsity of 1/20 during 1 to 2000 iterations, while the sparsity changes to 1/2 when the iteration is from 2000 to 3000, and it is non-sparsity after 3000 iterations. The input signal $\{x(n)\}$ is a white Gaussian random sequence with zero mean and unit variance. All simulation results below are obtained by averaging over 100 independent Monte Carlo runs, and each run performs 5000 iterations.

We employ the alpha-stable distribution [30] as impulsive noise model which has been widely applied in the literature [31-32]. The characteristic function of the alpha-stable distribution is given by

$$f(t) = \exp\{j\delta t - \gamma|t|^{\alpha}[1 + j\beta\,\mathrm{sgn}(t)S(t,\alpha)]\} \tag{18}$$

in which

$$S(t,\alpha) = \begin{cases} \tan\frac{\alpha\pi}{2} & if\ \alpha \ne 1 \\ \frac{2}{\pi}\log|t| & if\ \alpha = 1 \end{cases} \tag{19}$$

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. Such a distribution is called a symmetric alpha-stable ($S\alpha S$) distribution when $\beta = 0$. We define the parameters vector as $V = (\alpha, \beta, \gamma, \delta)$.

First, we investigate the convergence behavior of the proposed methods in impulsive noises, where the noise parameters vector is $V = (1.2, 0, 0.2, 0)$. The SDL is L=20. The step size is set at 0.03 for all algorithms. The kernel widths in MEE and CIM are 2.0 and 0.04, respectively. For all sparse aware algorithms, ρ is set at 0.0001. The parameter δ' for RZALMS and RZAMEE is 10. The average convergence curves in terms of the mean square deviation (MSD) are shown in Fig.1. As one can see from the MSD results, when the channel system is very sparse(before the 2000th iteration), the sparse aware MEE achieve faster convergence rate and better steady-state performance than the other robust algorithms (LAD, MEE), while ZALMS and RZALMS work poorly as they are sensitive to the impulsive noises. Thus, we only consider the MEE, LAD algorithms comparing with the proposed algorithm in next experiment case. In addition, CIMMEE achieves lower MSD than ZAMEE and RZAMEE since the CIM provides a nice approximation for the l_0-norm. After the 2000th iteration, as the number of non-zero taps increases to ten, the performance of the ZAMEE and RZAMEE deteriorates while the CIMMEE maintains the best performance among all the sparse aware filters. After 3000 iterations, the sparse aware MEE algorithms still perform comparable with the MEE even though the system is now completely non-sparse.

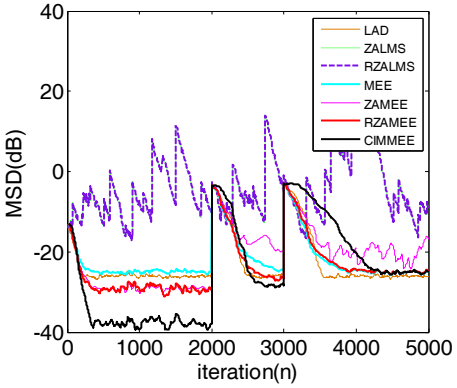

Fig.1. Tracking and steady-state behaviours of 20-order adaptive filters

Second, we conduct the simulation with different γ $(0.2,0.4,0.6,0.8,1)$ and α $(1,1.2,1.3,1.4,1.5,1.6,1.7)$to further demonstrate the performance of the proposed method. In this simulation, we mainly focus on the fully sparse channel system in the first stage of the proposed model. The $\eta = 0.02$ for all algorithms, and other parameter settings are the same as in the previous simulation for all algorithms. The MSD, versus different γ and α, are illustrated in Fig. 2 and Fig. 3 respectively. Evidently, the sparse aware MEE algorithms perform well with the different parameter of the impulsive noise model. Moreover, we see that the CIMMEE achieves much lower MSDs in all the cases. Simulation results confirm that the proposed sparse aware MEE algorithms, especially CIMMEE, can efficiently estimate a sparse channel in impulsive noise environment.

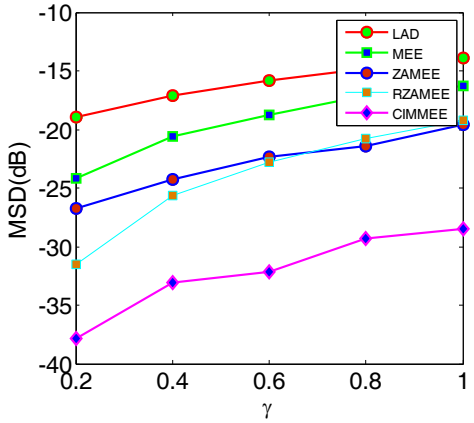

Fig.2. Steady-state MSD versus γ

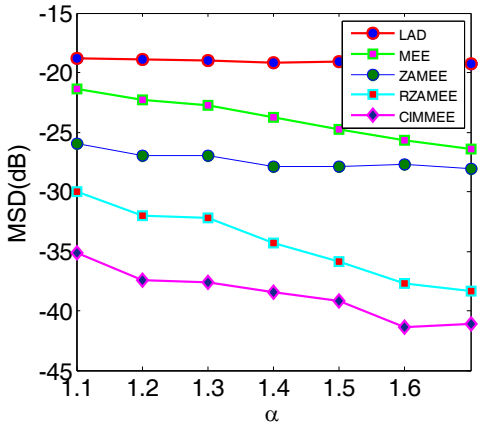

Fig.3. Steady-state MSD versus α

Third, we perform simulations to investigate how the kernel width σ_1 affects the performance, which is an important parameter for the sparse aware MEE. Here, the steady-state MSDs of the CIMMEE with different σ_1 (0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, and 5) and different α (1, 1.2, 1.4, 1.6, 1.8, and 2) are computed. Other parameters are set as: $\gamma = 1$, $\eta = 0.01$, $\rho = 0.0001$, $\sigma_2 = 0.04$ and $\delta' = 10$. The results are given in Fig.4. One can see that the CIMMEE achieves different MSDs with different σ_1 and under different noise distributions. In this example, the lowest MSD will be obtained around $\sigma_1 = 1.5$. From the simulation results we may conclude that the kernel width in MEE has a significant influence on the performance. The selection of the kernel width is critical to the success of the sparse aware MEE, and this will be an interesting research topic in the future study.

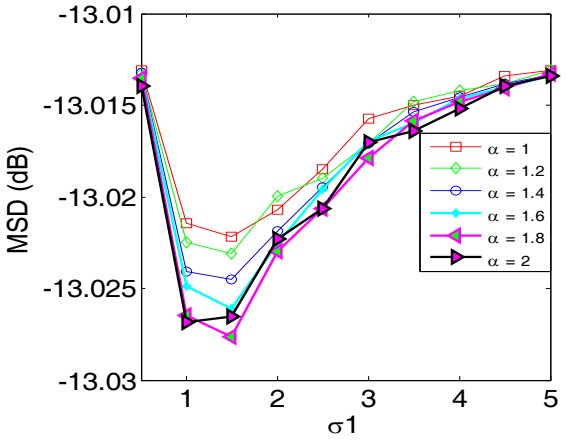

Fig.4. Steady-state MSD of CIMMEE with different kernel size σ_1 for different α.

5. CONCLUSION

In this work, we develop several sparsity aware minimum error entropy (MEE) algorithms, including ZAMEE, RZAMEE, and CIMMEE, which are derived by incorporating different sparsity penalty terms into the MEE criterion. Simulation results of sparse channel estimation show that the proposed methods can achieve excellent performance and outperform most of the existing sparsity-aware algorithms especially when the measurements are disturbed by impulsive non-Gaussian noises.

6. REFERENCE

[1]Chen Y, Gu Y, Hero A O. "Sparse LMS for system identification," *In: ICASSP Conf*, 2009, 3125-3128

[2]Gu Y, Jin J, Mei S, "l_0 norm constraint LMS algorithm for sparse system identification," *IEEE Signal Processing Letters*, vol.16, no.9, pp.774-777, 2009.

[3]Su G, Jin J, Gu Y, et al, "Performance analysis of l_0 norm constraint least mean square algorithm," *IEEE Transactions on Signal Processing*, vol.60, no.5, pp. 2223-2235, 2012.

[4]Shi K, Shi P, "Convergence analysis of sparse LMS algorithms with l_1-norm penalty based on white input signal," *Signal Processing*, vol.90,no.12,pp. 3289-3293,2010.

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

[6]Taheri O, Vorobyov S A., "Sparse channel estimation with lp norm and reweighted l_1-norm penalized least mean squares," *In Conference of ICASSP*, 2011, 2864-2867.

[7]F.Y. Wu, F. Tong. "Gradient optimization p-norm-like constraint LMS algorithm for sparse system estimation," *Signal Processing*, vol.93, no.4, pp. 967-971, 2013.

[8]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.

[9] Aliyu M L, Alkassim M A, Salman M S, "A p-norm variable step-size LMS algorithm for sparse system identification," *Signal, Image and Video Processing*, pp.1-7,2014.

[10]Turan, C., Salman, M.S., "A sparse function controlled vari able step-size LMS algorithm for system identification," *In IEEE Signal Processing and Communications Applications Conference*, 2014, 329-332.

[11]Gui G, Peng W, Adachi F., "Improved adaptive sparse channel estimation based on the least mean square algorithm," *In IEEE Wireless Communications and Networking Conference (WCNC)*, 2013, 3105-3109.

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

[13]K.N.Plataniotis, D. Androutsos, A. N. Venetsanopoulos, "Nonlinear fltering of non-Gaussian noise," *J. Intelligent and Robotic Syst.*, vol. 19, pp. 207-231, 1997.

[14]B. Weng and K. E. Barner, "Nonlinear system identification in impulsive environments," *IEEE Trans. Signal Process*, vol. 53, pp. 2588-2594, 2005.

[15]Golub GH, Van Loan CF, *Matrix computation*, Johns Hopkins University Press, Baltimore, 1983.

[16]J. C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives,* Springer, 2010.

[17]Badong Chen, Yu Zhu, Jinchun Hu, Jose C. Principe, *System Parameter Identification: Information Criteria and Algorithms*, Elsevier, 2013.

[18]D. Erdogmus and J. C. Principe, "Generalized information potential criterion for adaptive system training," *IEEE Trans. Neural Netw.*, vol.13, no.5, pp.1035-1044, 2002.

[19]D. Erdogmus and J. C. Principe, "From linear adaptive filtering to nonlinear information processing," *IEEE Signal Process. Mag.*, vol.23, no.6, pp.15-33, 2006.

[20]D. Erdogmus and J. C. Principe, "An error-entropy minimization for supervised training of nonlinear adaptive systems," *IEEE Trans. Signal Process.*, vol.50, no.7, pp.1780-1786, 2002.

[21]B. Chen, J. Hu, L. Pu, Z. Sun, "Stochastic gradient algorithm under (h, phi)-entropy criterion," *Circuit, Systems and Signal Processing*, vol.26,pp.941-960,2007.

[22]B. Chen, Z. Yuan, N. Zheng, J. C. Principe, "Kernel minimum error entropy algorithm," *Neurocomputing,* vol. 121, pp.160-169, 2013.

[23]B. Chen, J. C. Principe, "Some further results on the minimum error entropy estimation," *Entropy*, vol14, no.5, pp. 966-977, 2012.

[24]B. Chen, Y. Zhu, J. Hu, "Mean-square convergence analysis of ADALINE training with minimum error entropy criterion," *IEEE Transactions on Neural Networks*, vol. 21, no.7, pp.1168-1179, 2010.

[25]Li C, Shen P, Liu Y, et al, "Diffusion information theoretic learning for distributed estimation over network," *IEEE Transactions on Signal Processing*, vol. 61, no.16, pp.4011-4024, 2013.

[26]Han S, Principe J. "A fixed-point minimum error entropy algorithm," *IEEE proceedings of Signal Processing Society Workshop on Machine Learning for Signal Processing,* 2006, 167-172.

[27]Weifeng Liu, Puskal P.Pokharel, Jose C. Principe., "Correntropy: Properties and Applications in Non-Gaussian Signal Processing," *IEEE Trans. Signal Process*, vol.55, no.11, pp.5286-5298, 2007.

[28]Seth S, Príncipe J CC., "Compressed signal reconstruction using the correntropy induced metric," IEEE Conference of ICASSP, 2008. p. 3845-3848.

[29]Papoulis E V, Stathaki T., "A normalized robust mixed-norm adaptive algorithm for system identification," *Signal Process Lett*, vol.11, no.1, pp.5286-5298, 2004.

[30]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.

[31]Georgiadis A T, Mulgrew B., "A family of recursive algorithms for channel identification in alpha-stable noise," *In Fifth Bayona Workshop on Emerging Technologies in Telecommunications*, 1999, 153-157

[32]Wang J, Kuruoglu E E, Zhou T., "Alpha-stable channel capacity," *IEEE Communications Letters* , vol.15, no.10, pp. 1107-1109, 2011.