A Norm Constraint Lorentzian Algorithm Under Alpha-stable Measurement Noise

Xinqi Huang*, Yingsong Li* and Felix Albu†
* College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China
E-mail: huangxinqi@hrbeu.edu.cn, liyingsong@ieee.org
† Department of Electronics, Valahia University of Targoviste, Targoviste, 130082, Romania
E-mail: felix.albu@valahia.ro

***Abstract*—An l_0-norm constraint Lorentzian (L_0-CL) algorithm is proposed for adaptive sparse system identification to combat impulsive noise. The L_0-CL algorithm is derived via exerting an l_0-norm penalty on the coefficients in the cost function, which is equivalent to add a zero-attractor in the iterations. The zero-attractor attracts the coefficients to zero during the iterations. By the way, the L_0-CL algorithm can achieve lower mean square error (MSE) for estimating the sparse systems. The simulation results presented in this paper demonstrate that the proposed algorithm has superior performance in both convergence rate and steady-state behavior by identifying the sparse systems in the impulsive noise environment.**

***Index Terms*—Sparse adaptive filtering, lorentzian norm, l_0-norm, impulsive noise.**

I. Introduction

Among the wide range of adaptive filtering (AF) applications like echo cancellation, system identification, noise cancellation and channel estimation, more and more AF algorithms have been developed over past decades [1], [2]. The least mean square (LMS) algorithm, normalized form (normalized LMS, NLMS) and their variants are popular methods for many applications owing to their low implementations [3-11]. However, their performance may be deteriorated in non-Gaussian environments [12]. To solve the mentioned problem above, the sign algorithm (SA), which is derived from the cost function of the absolute value of the error, shows robust performance against impulsive noise [13]. Since the SA algorithm has a slow convergence rate, the affine projection sign (APS) algorithm has proposed to offer both the faster convergence and better robustness [14].

Furthermore, the maximum correntropy criterion (MCC) algorithm and its variants have been presented within the AF framework for non-Gaussian environments [15-18]. Though the MCC algorithm can give a resistant in impulsive noise environments, it cannot take advantage of the sparsity of sparse natural channels that only small percentage of the impulse response coefficients have significant magnitudes while the rest are zero or very small. The zero attracting based sparse AF algorithm has been proposed based on the MCC algorithm, which is derived by incorporating an l_1-norm penalty into the cost function of the MCC. As a result, the zero attractor appears in its iteration and it is known as zero attracting (ZA) MCC (ZA-MCC) algorithm [19]. Simultaneously, a reweighted ZA-MCC (RZA-LMS) algorithm has been promoted to further improve the filtering performance [19].

Motivated by Lorentzian adaptive filtering (LAF) [20], [21] and l_0-norm constraint LMS (l_0-LMS) [22] algorithm, a new Lorentzian algorithm with l_0-norm constraint is proposed to identify a sparse system in the impulsive noise environment. The idea is to introduce a penalty that exploits sparsity in the cost function and to use a gradient descent search on the cost function. The proposed method will accelerate the convergence of near-zero coefficients for identifying sparse systems. Simulation examples show that the L_0-CL algorithm performs well for sparse system identifications

II. Review of the LAF algorithm

In this paper, the desired signal from an unknown system $\mathbf{w}_o \in \mathbf{R}^{N\times 1}$ is $d(n) = \mathbf{u}^T(n)\,\mathbf{w}_o + v(n)$, where $\mathbf{u}(n) = [u(n), u(n-1), \ldots, u(n-N+1)]^T$ is the input signal vector. The variable $v(n)$ is the measurement noise and $(\cdot)^T$ is the transpose operator. Let $\mathbf{w}(n)$ be an estimation of $\mathbf{w}_o$ at iteration n, and $\mathbf{w}(n) = [w_1(n), w_2(n), \ldots, w_N(n)]^T$. The estimation error is defined as

$$e(n) = d(n) - \mathbf{u}^T(n)\,\mathbf{w}_o. \tag{1}$$

The input matrix is defined as $\mathbf{U}(n) = [\mathbf{u}(n), \mathbf{u}(n-1), \ldots, \mathbf{u}(n-M+1)]$ and the estimated error vector can be expressed as $\mathbf{e}(n) = [e(n), e(n-1), \ldots, e(n-M+1)]^T$. Then, the update equation of LAF [19] can be presented as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu_{\mathrm{LAF}}\mathbf{U}(n)\mathbf{G}(n)\mathbf{e}(n), \tag{2}$$

where μ_{LAF} is the step-size of adaptation and $\mathbf{G}(n) = \mathrm{diag}(g_n(1), g_n(2), \ldots, g_n(M))$ is an $M \times M$ diagonal matrix with

$$g_n(k) = \frac{\gamma^2}{\gamma^2 + e(n-k+1)}, k = 1, 2, \ldots, M \tag{3}$$

in which $\gamma > 0$ is a small constant. $\mathbf{G}(n)$ provides the weight factor $g_n(k)$ which decreases as the magnitude of the error increases. Denote $\mathbf{R}_u$ as the covariance matrix of the input matrix $\mathbf{U}(n)$ and $\lambda_{\max}$ as its maximum eigenvalue. The convergence condition for the LAF is

$$0 < \mu_{\mathrm{LAF}} < \frac{2}{M\lambda_{\max}} \tag{4}$$

III. The proposed L_0-CL algorithm

The studies on compressed sensing reveal that the sparsest solution can be obtained using l_0-norm, which shows that a penalty of AF based on l_0-norm can be integrated into the cost function when the unknown system is sparse. Thus, the cost function for the proposed is expressed as

$$J(n) = \|\mathbf{e}(n)\|_{LL2,\gamma} + \alpha\|\mathbf{w}(n)\|_0, \tag{5}$$

where $\|\cdot\|_0$ denotes l_0-norm which exploits the sparsity information of the unknown system, $\|\cdot\|_{LL2,\gamma}$ is Lorentzian norm given by

$$\|\mathbf{x}\|_{LL2,\gamma} = \sum_{i=1}^{M} \log\left[1 + \left(\frac{x_i}{\gamma}\right)^2\right], \tag{6}$$

and $\alpha > 0$ is a parameter to balance the penalty and the estimation error. Using a gradient descent search on (5), the update equation can be obtained as follow:

$$\begin{aligned}\mathbf{w}(n+1) &= \mathbf{w}(n) - \mu' \frac{\partial J(n)}{\partial \mathbf{w}(n)} \\ &= \mathbf{w}(n) + \frac{\mu'}{\gamma^2}\mathbf{U}(n)\mathbf{G}(n)\mathbf{e}(n) - \mu'\alpha \frac{\partial \|\mathbf{w}(n)\|_0}{\partial \mathbf{w}(n)}.\end{aligned} \tag{7}$$

A popular approximation for l_0-norm constraint [23] is defined as

$$\|\mathbf{w}(n)\|_0 \approx \sum_{j=1}^{N} \left(1 - \exp\left(-\beta\left|w_i(n)\right|\right)\right) \tag{8}$$

where β is the parameter to balance the expansion and the l_0-norm constraint. Then, the corresponding derivatives of $f_\beta(w_i(n))$ are expressed as

$$f_\beta(w_i(n)) \triangleq \frac{d\|\mathbf{w}(n)\|_0}{dw_i(n)}. \tag{9}$$

Then, we set

$$f_\beta(w_i(n)) \approx \beta sign\left(w_i(n)\right)\exp\left(-\beta\left|w_i(n)\right|\right). \tag{10}$$

Thus, the gradient of $\|\mathbf{w}(n)\|_0$ with respect to $\mathbf{w}(n)$ can be given by

$$\begin{aligned}\frac{\partial\|\mathbf{w}(n)\|_0}{\partial \mathbf{w}(n)} &= \mathbf{f}_\beta\left(\mathbf{w}(n)\right) \\ &\triangleq \left[f_\beta(w_0(n)), f_\beta(w_i(n)), \cdots, f_\beta(w_{N-1}(n))\right]^T.\end{aligned} \tag{11}$$

The update equation is presented as

$$\begin{aligned}\mathbf{w}(n+1) &= \mathbf{w}(n) + \frac{\mu'}{\gamma^2}\mathbf{U}(n)\mathbf{G}(n)\mathbf{e}(n) - \mu'\alpha\mathbf{f}_\beta\left(\mathbf{w}(n)\right) \\ &= \mathbf{w}(n) + \mu\mathbf{U}(n)\mathbf{G}(n)\mathbf{e}(n) - \lambda\mathbf{f}_\beta\left(\mathbf{w}(n)\right),\end{aligned} \tag{12}$$

where $\mu = \mu'/\gamma^2$ is considered as the step size and $\lambda = \mu'\alpha$ is a parameter to control the zero attractor. Comparing the L_0-CL algorithm in (12) to the LAF in (2), the L_0-CL algorithm has an additional term $-\lambda\mathbf{f}_\beta\left(\mathbf{w}(n)\right)$. The convergence condition of the L_0-CL algorithm is provided herein.

Theorem 1: With a zero-mean input $u(n)$, the L_0-CL algorithm produces stable $\mathbf{w}(n)$ which approaches to $\mathbf{w}_o$ at infinity if the step-size μ satisfies equation (4).

Proof: we define the weight error vector as follows:

$$\tilde{\mathbf{w}}(n) = \mathbf{w}_o - \mathbf{w}(n). \tag{13}$$

From the iteration equation (12) and the definition in (13), we have

$$\begin{aligned}\tilde{\mathbf{w}}(n+1) = &\left(\mathbf{I} - \mu\mathbf{U}(n)\mathbf{G}(n)\mathbf{U}^T(n)\right)\tilde{\mathbf{w}}(n) \\ &- \mu\mathbf{U}(n)\mathbf{G}(n)\mathbf{v}(n) + \lambda\mathbf{f}_\beta\left(\mathbf{w}(n)\right),\end{aligned} \tag{14}$$

where $\mathbf{v}(n) = \left[v(n), v(n-1), \ldots, v(n-M+1)\right]^T$.

Taking expectations on both sides of (14), yields

$$\begin{aligned}E\left[\tilde{\mathbf{w}}(n+1)\right] = &\left(\mathbf{I} - \mu E\left[\mathbf{U}(n)\mathbf{G}(n)\mathbf{U}^T(n)\right]\right)E\left[\tilde{\mathbf{w}}(n)\right] \\ &+ \lambda E\left[\mathbf{f}_\beta\left(\mathbf{w}(n)\right)\right],\end{aligned} \tag{15}$$

where $E\left[\mathbf{v}(n)\right]$ is zero because $\mathbf{v}(n)$ is zero mean and is assumed to be statistically independent of $\mathbf{U}(n)\mathbf{G}(n)$. From the definition of $\mathbf{f}_\beta\left(\mathbf{w}(n)\right)$, it can be clearly seen that there is a limitation of E[f(w(n))]. Therefore, the condition for guaranteeing the convergence is that the maximal eigenvalue of $\left(\mathbf{I} - \mu E\left[\mathbf{U}(n)\mathbf{G}(n)\mathbf{U}^T(n)\right]\right)$ is less than 1. Considering $0 < g_n(k) < 1, k = 1, 2, ..., M$ and assuming that $\mathbf{G}(n)$ and $\left[\mathbf{U}(n)\mathbf{U}^T(n)\right]$ are statistically independent, we have

$$\begin{aligned}&\lambda_{\max}(E\left[\mathbf{U}(n)\mathbf{G}(n)\mathbf{U}^T(n)\right]) \\ &= \lambda_{\max}(E\left[\mathbf{U}^T(n)\mathbf{U}(n)\mathbf{G}(n)\right]) \\ &= \lambda_{\max}(E\left[\mathbf{U}^T(n)\mathbf{U}(n)\right]E\left[\mathbf{G}(n)\right]) \\ &< \lambda_{\max}(E\left[\mathbf{U}^T(n)\mathbf{U}(n)\right]) \\ &= M\lambda_{\max}.\end{aligned} \tag{16}$$

Thus, the condition for guaranteeing the convergence is

$$0 < \mu < \frac{2}{M\lambda_{\max}}. \tag{17}$$

We can find that the convergence condition for the L_0-CL algorithm and LAF algorithm is same.

IV. Simulation results

Several experiments are established to evaluate the behaviors of the proposed L_0-CL algorithm. The unknown channel is a one-group channel whose active coefficients are set in [257,288]. In all simulation experiments, the length of the channel (N) is set to be 1024. Two input signals, white Gaussian noise (WGN) and colored noise, are used to investigate the performance of the L_0-CL algorithm. The colored noise is obtained from WGN filtering through a first-order system with a pole at 0.8. For each simulation, all the results are obtained by taking the ensemble average of the mean square deviation over 20 independent Monte Carlo runs. To illustrate the robust performance of the proposed algorithms, the alpha-stable noise is chosen to construct the non-Gaussian environment which may occur in practice. The characteristic function of the alpha-stable process is expressed as

$$\varphi(t) = \exp\left\{j\delta t - \eta|t|^\theta\left[1 + j\xi\mathrm{sign}(t)\,\omega(t,\theta)\right]\right\} \tag{18}$$

in which

$$\omega(t,\theta)=\begin{cases}\tan\frac{\pi\theta}{2} & \text{if}\theta\neq 1\\ \frac{2}{\pi}\log|t| & \text{if}\theta=1\end{cases}, \tag{19}$$

where $0<\theta\leq 2$ is the characteristic index which determines the strength of the alpha-stable distribution impulse, $-1\leq\xi\leq 1$ is the symmetry, $-\infty<\delta<+\infty$ is the location factor which controls the location of the alpha-stable distribution center and $\eta>0$ is the dispersion factor. Herein, $\theta=1.2$, $\xi=0$, $\eta=1$, and $\delta=0$ are selected to construct the impulsive noise. The power of the input signal is set to be 1 and the impulsive noise is added to the background of the unknown channel at a signal-to-noise ratio (SNR) of 10 dB in the following experiments. For all the experiments, M is set to be 8 and the scale parameter γ is set as follows

$$\gamma=\frac{d_{2/3}(n)-d_{1/3}(n)}{2}, \tag{20}$$

where $d_q(n)$ denotes the q-th quantile of $\mathbf{d}(n)$ ($\mathbf{d}(n)=[d(n),d(n-1),\ldots,d(n-M+1)]^T$). The performance metric chosen to evaluate the proposed algorithm was the normalized misalignment (in dB) of the coefficient vector, given by $10\log\left(\|\mathbf{w}(n)-\mathbf{w}_o\|_2^2/\|\mathbf{w}_o\|_2^2\right)$.

A. Performance of the L$_0$-CL algorithm with different parameters

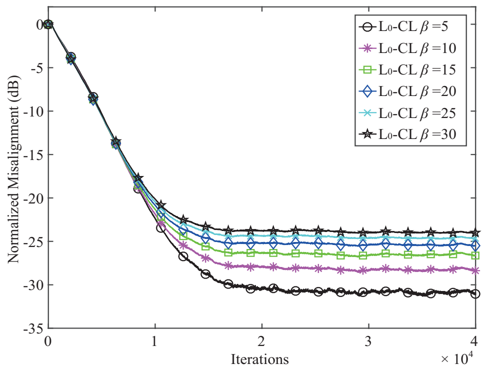

Fig. 1. Effects of β on L$_0$-CL algorithm.

The experiments are constructed to investigate the effects of the parameters β and λ on the performance of the L$_0$-CL algorithm. Since β and λ are key parameters for the proposed L$_0$-CL algorithm. Different β and λ are chosen to examine the L$_0$-CL algorithm with the colored input signal. The step size of the proposed L$_0$-CL algorithm is set 0.0005.

In Fig. 1, the results are obtained by varying β from 5 to 30 and keeping $\lambda=6\times10^{-7}$ as a constant, whereas in Fig. 2, the results were obtained by varying λ, and $\beta=5$. The simulation result presented in Fig. 1 indicates that the performance of the L$_0$-CL algorithm decreases as β increases. In Fig. 2, the steady-state misalignment improves with the increase of λ, and decreases when λ is greater than 6×10^{-7}.

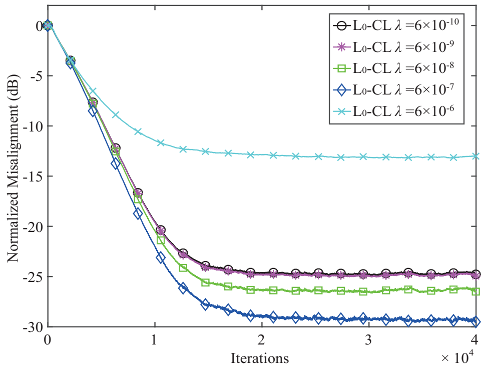

Fig. 2. Effects of λ on L$_0$-CL algorithm.

B. Performance of the L$_0$-CL algorithm with different input signals

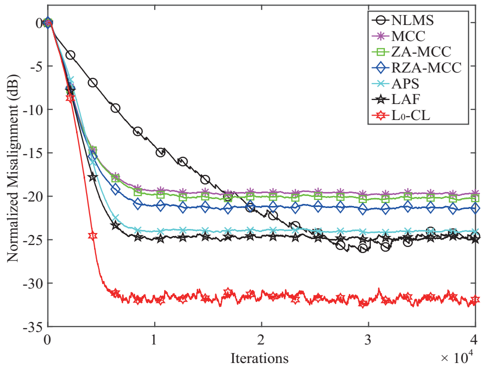

Fig. 3. Performance of the L$_0$-CL algorithm with WGN input signal.

According to the previous discussion, the parameters β is set to be 5 and $\lambda=6\times10^{-7}$ to get lower steady-state misalignment. In this subsection, the performance of the L$_0$-CL algorithm is compared with the NLMS, MCC, ZA-MCC, RZA-MCC, APS and LAF. To guarantee the same initial convergence rate, $\mu_{\mathrm{NLMS}}=0.2$, $\mu_{\mathrm{MCC}}=0.0016$, $\mu_{\mathrm{ZA-MCC}}=0.0017$, $\mu_{\mathrm{RZA-MCC}}=0.0016$, $\mu_{\mathrm{APS}}=0.005$, $\mu_{\mathrm{LAF}}=0.0005$ and $\mu_{\mathrm{L_0-CL}}=0.0016$ are selected for WGN signal. And, $\mu_{\mathrm{NLMS}}=0.4$, $\mu_{\mathrm{MCC}}=0.0015$, $\mu_{\mathrm{ZA-MCC}}=0.0016$, $\mu_{\mathrm{RZA-MCC}}=0.0015$, $\mu_{\mathrm{APS}}=0.0045$, $\mu_{\mathrm{LAF}}=0.0005$ and $\mu_{\mathrm{L_0-CL}}=0.0005$ are selected for colored signal. Further, the kernel size is chosen as 0.9 for MCC and its variant algorithms, ρ is 610-7 for ZA-MCC and RZA-MCC algorithms, ε is 30 for the RZA-MCC algorithm.

The performance comparisons of the L$_0$-CL algorithm under different input signals for unknown channel analysis are presented in Figs. 3 and 4, respectively. Fig. 3 and Fig. 4 show that the performance of the LAF algorithm outperforms all the other algorithms except the L$_0$-CL algorithm in WGN signal, but gets worse performance than the APS algorithm for the colored input signal. The proposed L$_0$-CL algorithm achieves the lowest normalized misalignment compared with the mentioned algorithms in two different input signals for identifying the sparse systems.

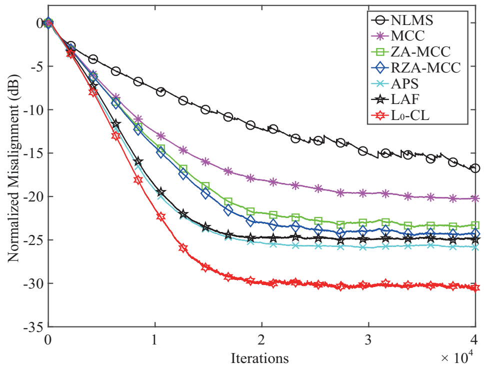

Fig. 4. Performance of the L$_0$-CL algorithm with colored input signal.

Next, to get the tracking performance of the L$_0$-CL algorithm, we consider a two-group channel with active coefficients distributed in [257,272] and [769,784]. The input signal is chosen as colored noise, and the set-up of this experiment is the same as considered earlier. The tracking performance of the proposed L$_0$-CL algorithm is investigated by one-group and two-group channels mentioned above. Simulation results are described in Fig. 5. It is found that the proposed L$_0$-CL algorithm outperforms others by producing lesser steady-state misalignment and the faster convergence rate for tracking the sudden changed sparse systems.

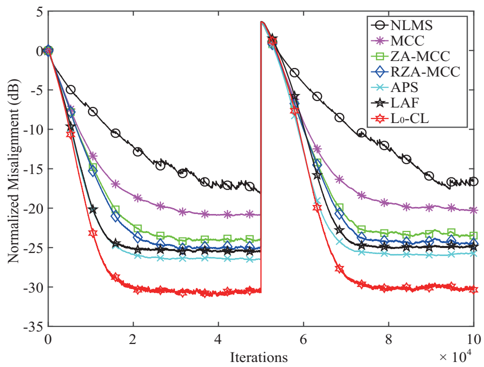

Fig. 5. Tracking performance of the L$_0$-CL algorithm with colored input signal.

Conclusion

In order to improve the performance of sparse system identification in impulse environment, the L$_0$-CL algorithm is proposed to exploit sparsity by introducing l_0-norm to the cost function as an additional constraint. The proposed method can accelerate the convergence of near-zero coefficients in the impulse response of sparse systems. Simulation results demonstrate that the L$_0$-CL algorithm accelerates the identification of speed, and outperforms the mentioned algorithms in term of the misalignment. The effects of parameters for the L$_0$-CL algorithm are also shown in the experiments.

Acknowledgement

This work was supported the Natural Science Foundation of Heilongjiang Province, China (Grant No. F2017004).

References

[1] W. F. Schreiber, "Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions," Proc. IEEE, vol. 83, no. 6, pp. 958-981, Jun. 1995.
[2] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," IEEE Trans. on Speech Audio Process, vol. 8, no. 5, pp. 508-518, Sep. 2000.
[3] B. Widrow, and S. D. Stearns, Adaptive signal processing, Prentice Hall, New Jersey, USA, 1985.
[4] S. C. Douglas, and T. H. -Y. Meng, "Normalized data nonlinearities for LMS adaptation," IEEE Trans. on Signal Process., vol. 42, no. 6, pp. 1352-1365, June 1994.
[5] J. Homer, I. Mareels, R.R. Bitmead, B. Wahlberg, A. Gustafsson, "LMS estimation via structural detection," IEEE Trans. on Signal Process., vol. 46, no. 10, pp. 2651-2663, Oct. 1998.
[6] Y. Chen, Y. Gu, A. O. Hero, "Sparse LMS for system identification," IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 3125-3128, Taipei, Taiwan, Apr. 2009.
[7] Y. Li, Y. Wang, T. Jiang, "Norm-adaption penalized least mean square/fourth algorithm for sparse channel estimation," Signal Processing, vol. 128, no. 2016, pp. 243-251, Apr. 2016.
[8] Y. Li, Y. Gu, K. Tang, "Parallel NLMS filters with stochastic active taps and step-sizes for sparse system identification," IEEE Trans. on Signal Process., vol. 46, no. 10, pp. 2651-2663, Oct. 1998.
[9] M. Godavarti, A. O. Hero, "Partial update LMS algorithms," IEEE Trans. Signal Process., vol. 53, no. 7, pp. 2382-2399, 2005.
[10] D.L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," IEEE Trans. Speech Audio Process., vol. 8, no. 5, pp. 508-518, Sep. 2000.
[11] Y. Li, Y. Wang, T. Jiang, "Sparse-aware set-membership NLMS algorithms and their application for sparse channel estimation and echo cancelation," AEU-Int. J. Electron. Commun., vol. 70, no.7, pp. 895-902, 2017.
[12] S. Pei, C. Tseng, "Least mean p-power error criterion for adaptive FIR filter," IEEE Journal on Selected Areas in Communications, vol. 12, no. 9, pp. 1540-1547, Dec. 1994.
[13] V. J. Mathews and S. H. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," IEEE Trans. Acoust., Speech, Signal Process., vol. 35, no. 4, pp. 450-454, Apr. 1987.
[14] T. Shao, Y. R. Zheng, and J. Benesty, "An affine projection sign algorithm robust against impulsive interferences," IEEE Signal Process. Lett., vol. 17, no. 4, pp. 327-330, Apr. 2010.
[15] Singh, A. and J. C. Principe, "Using correntropy as a cost function in linear adaptive filters," Int. Joint Conf. on Neural Networks, pp. 2950-2955, Atlanta, GA, USA, June 2009.
[16] W. Shi, Y. Li, Y. Wang, "Noise-free maximum correntropy criterion algorithm in non-Gaussian environment," IEEE Trans. Circuits and Systems II: Express Briefs, doi:10.1109/TCSII.2019.2914511.
[17] Y. Li, Z. Jiang, W. Shi, X. Han, B. Chen, "Blocked maximum correntropy criterion algorithm for cluster-sparse system identifications," IEEE Trans. Circuits and Systems II: Express Briefs, doi: 10.1109/TCSII.2019.2891654.
[18] W. Liu, P. P. Pokharel, Jose C. Principe, "Correntropy: properties and applications in non-Gaussian signal processing," IEEE Trans. on Signal Process., vol. 55, no. 11, pp. 5286-5298, Nov. 2007.
[19] Ma, W., H. Qu, G. Gui, L. Xu, J. Zhao, B. Chen, "Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments," Journal of the Franklin Institute, vol. 352, no. 7, pp. 2708-2727, Jul. 2015.
[20] R. E. Carrillo, K. E. Barner, "Lorentzian based iterative hard thresholding for compressed sensing," Proc. IEEE International Conference on Acoustic Speech and Signal Processing, Prague, Czech Republic, pp.3664-3667, May 2011.
[21] R. L. Das, M. Narwaria, "Lorentzian based adaptive filters for impulsive noise environments," IEEE Trans. Circuits and Systems I: Regular Papers, vol.64, no.6, pp.1529-1539, Jun. 2017.
[22] Y. Gu, J. Jin, S. Mei, "l_0-norm constraint LMS algorithm for sparse system identification," IEEE Signal Process. Lett., vol. 16, no. 9, pp. 774-777, Sept. 2009.
[23] P. Huber, Robust Statistics, Wiley, New York, NY, USA, 1st edition, 1981.