Sparse System Identification in The Presence of Noisy Input Signal Using Biased Compensator Minimum Error Entropy Algorithm

Rajni Yadav, C.S Rai

University School of Information, Communication and Technology

Guru Gobind Singh Indraprastha University.
Sector - 16C, Dwarka
Delhi - 110078, India
rajni@mait.ac.in

Abstract: - Sparse systems are identified effectively by correntropy induced minimum error entropy (CIM-MEE) algorithm in the presence of non- Gaussian noise. But this does not take into account the noisy input signal. This paper presents a new approach for sparse system identification having input signal corrupted by white Gaussian noise. The noisy input signal produces a bias during estimation. The proposed scheme incorporates a bias compensator to overcome this bias by adding a constraint in objective function of CIM-MEE algorithm. Simulations carried out in MATLAB confirm better performance of proposed Biased Compensator CIM-MEE (BC-CIM-MEE) algorithm for noisy input signal in the presence of impulsive measurement noise.

Key-Words: - Minimum error entropy, Sparse system identification, Bias, Non-Gaussian

Received: November 7, 2018. Revised: January 14, 2019. Accepted: February 2, 2019.
Published: February 23, 2019

1 Introduction

Most of adaptive algorithms are based on minimum mean square error (MSE) criterion due to ease of implementation [1]. However, from a statistical point of view, MSE only takes into account the second order statistics and is therefore only optimal in the case of Gaussian signals [2]. But these algorithms perform poorly in non -Gaussian environment [2]. Several researchers have developed adaptive algorithm which are robust against non- Gaussian environment [3-5]. Recently researches have been done in field of information theory for developing the adaptive algorithms which perform well in non-Gaussian environment [6-7]. Some of criterions are entropy, mutual information and dispersion of information [7]. These criterions give rise to various learning algorithms supervised or unsupervised. Minimum error entropy (MEE) is one such principle in information theoretical learning and provides variants of supervised learning algorithms [8]. The basic idea of MEE is to extract from data as much information as possible about the unknown systems by minimizing the entropy of error between unknown system output and estimated output. This improves the estimation performance of system. In information theory, entropies are used to measure average information quantitatively.

In MEE algorithm, the characteristics of system are not considered in prior. Hence the performance of sparse system identification can be further effectively improved by providing the system prior information. Several adaptive algorithms are being developed based on least absolutely shrinkage and selection operator (LASSO) [9] and latest research in Compressive Sensing (CS) [10] which takes in account the prior information of system. These adaptive filtering algorithms have been developed by incorporating a penalty term acquainted with information regarding the sparseness of system into the objective function of the standard adaptive algorithms [11-17]. The most commonly proposed sparsity constraints in above sparsity aware adaptive algorithms are the l1-norm and reweighted l1-norm, which generate a zero-attractor in the iterations of filtering algorithms [11-17]. Recently several researches are done on correntropy induced metric for sparsity constraint. CIM based constraint better approximates zero attraction penalty to consider system dominant coefficients than l1norm and reweighted l1 –norm [3], [18-20]. Zongze Wu 1, Siyuan Peng et.al have developed CIM-MEE algorithm which perform well when the system is sparse [21]. The above algorithm performs well in sparse system identification in the presence of

impulsive measurement noise but input is considered non-noisy. This does not take into account the bias caused by noisy input signal. B. Kang, J. Yoo, and P. Park have proposed the concept of bias compensator to consider the noisy input signal. The bias compensator results in an unbiased estimation behavior for noisy input signal. The proposed BC-CIM-MEE algorithm adds a biased compensator in the objective function of CIM-MEE algorithm. Simulation confirms that the proposed algorithm performs better than CIM-MEE, zero attracting MEE (ZA-MEE) algorithms in the presence of noisy input.
The rest of the paper is organized as follows. In Section 2, reviews the MEE criterion, and CIM-MEE algorithm. In Section 3, we derive the proposed BC- CIM-MEE algorithms. In Section 4, simulation results are carried out to excel the performance of the proposed algorithm. Finally, In Section 5, conclusion is drawn.

2 Review of Minimum Error Entropy

Consider **x**(k), as input to unknown system and adaptive filter , $\hat{\boldsymbol{w}}$(k) is coefficient vector of adaptive filter we can express output **y**(k)of the adaptive filter as:

$$\boldsymbol{y}(k) = \hat{\boldsymbol{w}}^{T}(k)\,\mathbf{x}(k) \qquad (1)$$

where **x**(k) = [x(k) x(k-1) x(k-2) ………..x(k-N+1)]^T is input vector with variance σ_x^2 and $\hat{\boldsymbol{w}}^{T}$(k) is estimated coefficient vector of N-dimension and d(k) is reference output of unknown system and given as:

$$\mathbf{d}(k) = \boldsymbol{w}_0^{T}\,\mathbf{x}(k) + \mathbf{z}(k) \qquad (2)$$

where **z**(k) is measurement noise with zero mean and variance σ_z^2 and $\mathbf{w_0}$ is unknown system coefficient vector of M-dimensional which we want to estimate. So, we can write the instantaneous estimation error e(k) of adaptive filter as

$$\mathbf{e}(k) = \mathbf{d}(k) - \boldsymbol{y}(k) \qquad (3)$$

Let g(e) is probability density function of e(k). So Renyi's entropy of order two of error is given as:

$$\mathrm{H}_{\mathrm{R}_2} = -\log \int g_e^2(e)\, de \qquad (4)$$

Since all error samples are not available in practically. So we will estimate the value of g(e) from available error samples using Parzen window.
So we can write

$$\hat{g}_e(e) = \frac{1}{M}\sum_{i=1}^{M} f_\sigma(e - e(i)) \qquad (5)$$

Where, M is number of available samples and f_σ is kernel function having bandwidth σ. The most popular kernel used is Gaussian kernel defined as:

$$f_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-x^2}{2\sigma^2}} \qquad (6)$$

So we can write estimated Renyi's entropy $\widehat{\mathrm{H}}_{\mathrm{R}_2}$ as:

$$\begin{aligned}\widehat{\mathrm{H}}_{\mathrm{R}_2} &= -\log \int \hat{g}_e^2(e)\, de \\ &= -\log \frac{1}{M^2} \int \Big[\sum_{i=1}^{M} f_\sigma\,(e - e(i))\Big]^2 de \\ &= -\log \frac{1}{M^2} \sum_{i=1}^{M}\sum_{j=1}^{M} \int f_\sigma\,(e - e(i)\, f_\sigma\,(e - e(j)\, de \\ &= -\log \frac{1}{M^2} \sum_{i=1}^{M}\sum_{j=1}^{M} f_{\sigma\sqrt{2}}(e(i) - e(j)\,) \\ &= -\log\ V(e) \end{aligned} \qquad (7)$$

Where $V(e) = \frac{1}{M^2} \sum_{i=1}^{N}\sum_{j=1}^{N} f_{\sigma\sqrt{2}}(e(i) - e(j)\,)$, is defined as information potential .Thus, it is clear that the objective of MEE principle is to minimize error entropy which is equivalent to maximizing information potential.
Thus we will maximize V(e).
Hence objective function of MEE is

$$G_{MEE} = \underbrace{argmax}_{w \in \boldsymbol{R}}\ V(e) \qquad (8)$$

By method of gradient ascent, the weight update equation of MEE algorithm becomes:

$$\hat{w}(k+1) = \hat{w}(k) + \mu\, \nabla V(e(k)) \qquad (9)$$

$$\begin{aligned}\nabla V(e) &= \frac{\partial V(e(k))}{\partial \hat{w}(k)} \\ &= \frac{\partial}{\partial \hat{w}(k)}\Big[\frac{1}{M^2} \sum_{i=1}^{M}\sum_{j=1}^{M} f_{\sigma\sqrt{2}}(e(i) - e(j)\,)\Big] \\ &= \frac{1}{2M^2}\Big[\sum_{i=1}^{M}\sum_{j=1}^{M} f_{\sigma\sqrt{2}}(e(i) - e(j)\,)\big(e(i) - e(j)\big)\big(x(i) - x(j)\big)\Big]\end{aligned} \qquad (10)$$

Where $x(i)$ *and* $x(j)$ are inputs to the system at time i and j, respectively

2.1 Correntropy Induced Metric MEE (CIM-MEE)

The correntropy is a nonlinear measure of the similarity between two random variables A=[a_1, a_2, …, a_L] and B = [b_1, b_2, …, b_L] in kernel space defined as:

$$V(A,B) = E[f_\sigma(A,B)] = \int f_\sigma(a, b)\, dF_{A,B}(a, b) \qquad (11)$$

Where $F_{A,B}(a, b)$ is joint distribution function of random variable A& B and ξ is error between random variable a and b.

Practically, joint distribution function $F_{A,B}(a, b)$ of random variable A& B is unknown for calculating correntropy measure. So estimation of correntropy is calculated from finite number of samples of random variable defined as:

$$\hat{V}(A,B) = \frac{1}{L}\sum_{i=1}^{L} f_\sigma(a_i, b_i) \qquad (12)$$

Where M is number of elements of random variables A or B.

The CIM is a nonlinear function defined as:

$$\mathrm{CIM}(A,B) = \sqrt{(f(0) - \hat{V}(A,B))} \quad \text{and}$$

$$\mathrm{CIM}^2(A,B) = (f(0) - \hat{V}(A,B)) \qquad (13)$$

where $f(0) = \frac{1}{\sqrt{2\pi}\sigma}$

The CIM constraint leads to a better approximation to l0-norm than other previous smoothing continuous approximations [3]. CIM based sparsity constraint is defined as:

$$\|A\|_0 \sim \mathrm{CIM}^2(A,0) = \frac{\kappa(0)}{L}\sum_{i=1}^{L}\left(1 - \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{a_i^2}{2\sigma^2}}\right) \qquad (14)$$

From equation (14), it is clear that for any $a_i \neq 0$ and $|a_i| > \sigma$,the solution reaches very close to l0-norm for $\sigma \to 0$. Thus CIM provides a good approximation to sparsity constraint and can be used for development of sparsity aware adaptive algorithms for sparse system identification.

So the objective function of CIM-MEE becomes:

$$G_{CIM-MEE} = \underbrace{argmax}_{\hat{w}\in\boldsymbol{R}} (V(e) - \gamma\, \mathrm{CIM}^2(\hat{w},0)) \qquad (15)$$

Where γ is a proportionality constant associated with the zero attractor.

So By gradient ascent method, the weight update equation of CIM-MEE becomes:

$$\hat{w}(k+1) = \hat{w}(k) + \left[\frac{\mu}{2M^2}\left\{\sum_{i=1}^{M}\sum_{j=1}^{M} f_{\sigma_1\sqrt{2}}(e(i) - e(j))\big(e(i) - e(j)\big)(x(i) - x(j))\right\} - \rho\left\{\frac{\hat{w}(k)}{N\sqrt{2\pi}\sigma_2^3} e^{-\frac{\hat{w}^2(k)}{2\sigma_2^2}}\right\}\right] \qquad (16)$$

Where $\rho = \mu\gamma$

And σ_1 is kernel width of MEE criterion and σ_2 is kernel width of CIM.

So σ_2 should be carefully selected so that CIM approximate l0-norm appropriately.

3 Proposed BC-CIM-MEE

Consider the input signal x(k) is corrupted by white noise v(k) as

$$\mathbf{u}(k) = \mathbf{x}(k) + \mathbf{v}(k) \qquad (17)$$

and $q(k) = d(k) - \hat{\boldsymbol{w}}^{\mathbf{T}}(k)\,\mathbf{u}(k)$

$$= \boldsymbol{w}_0^T(k)\,\mathbf{u}(k) + \mathbf{z}(k) - \hat{\boldsymbol{w}}^{\mathbf{T}}(k)\,\mathbf{u}(k)$$

$$= \boldsymbol{w}_0^T(k)\,(\mathbf{x}(k)+\mathbf{v}(k)) + \mathbf{z}(k) - \hat{\boldsymbol{w}}^{\mathbf{T}}(k)\,(\mathbf{x}(k)+\mathbf{v}(k))$$

$$= (\boldsymbol{w}_0^T(k) - \hat{\boldsymbol{w}}^{\mathbf{T}}(k))\,\mathbf{x}(k) + \mathbf{z}(k) + (\boldsymbol{w}_0^T(k) - \hat{\boldsymbol{w}}^{\mathbf{T}}(k))\mathbf{v}(k)$$

$$= \Delta\hat{\boldsymbol{w}}^T(k)\,\mathbf{x}(k) + \mathbf{z}(k) + \Delta\hat{\boldsymbol{w}}^{\mathbf{T}}(k)\,\mathbf{v}(k)$$

$$= \mathbf{e_p(k)} + \mathbf{z}(k) + \Delta\hat{\boldsymbol{w}}^{\mathbf{T}}(k)\,\mathbf{v}(k) \qquad (18)$$

Where $\mathbf{e_p(k)}$ = $\Delta\hat{\boldsymbol{w}}^T(k)\,\mathbf{x}(k)$ is the priori error.

To overcome the effect of bias generated by noisy input, the update equation of proposed algorithm becomes:

$$\hat{w}(k+1) = \hat{w}(k) + \left[\frac{\mu}{2M^2}\left\{\sum_{i=1}^{M}\sum_{j=1}^{M} f_{\sigma_1\sqrt{2}}(q(i) - q(j))\big(q(i) - q(j)\big)(u(i) - u(j))\right\} - \rho\left\{\frac{\hat{w}(k)}{N\sqrt{2\pi\sigma_2^2}} e^{-\frac{\hat{w}^2(k)}{2\sigma_2^2}}\right\}\right] + C(k) \qquad (19)$$

Where C(k) is bias compensator vector to compensate bias produced by noisy input.

Consider weight error vector $\Delta\hat{w}(k) = w_0 - \hat{w}(k)$

So equation (19) becomes:

$$\Delta\hat{w}(k+1) = \Delta\hat{w}(k) - \left[\left[\frac{\mu}{2M^2}\left\{\sum_{i=1}^{M}\sum_{j=1}^{M} f_{\sigma_1\sqrt{2}}(q(i) - q(j))\big(q(i) - q(j)\big)(u(i) - u(j))\right\} + \rho\left\{\frac{\hat{w}(k)}{N\sqrt{2\pi\sigma_2^2}} e^{-\frac{\hat{w}^2(k)}{2\sigma_2^2}}\right\}\right]\right] - C(k) \qquad (20)$$

The term C(k) is related to bias compensator so, the third term in equation(20) should not be considered in derivation. By omitting third term in right side of equation (20), the update equation becomes:

$$\Delta\hat{w}(k+1) = \Delta\hat{w}(k) - \left[\frac{\mu}{2M^2}\left\{\sum_{i=1}^{M}\sum_{j=1}^{M} f_{\sigma_1\sqrt{2}}(q(i) - q(j))\big(q(i) - q(j)\big)(u(i) - u(j))\right\}\right] - C(k) \qquad (21)$$

As $\hat{\boldsymbol{w}}(k)$ approaches w_0 in steady state , $k\to\infty$

So $\Delta\hat{\boldsymbol{w}}(k) \to 0$

$$E[\Delta\hat{\boldsymbol{w}}(k+1) \mid u(k)] = E[\Delta\hat{\boldsymbol{w}}(k) \mid u(k)] = 0$$

$$E[C(k) \mid u(k)] = -E\left[\left\{\frac{\mu}{2M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} f_{\sigma_1\sqrt{2}}(q(i) - q(j))\big(q(i) - q(j)\big)(u(i) - u(j))\right\} \mid u(k)\right] \qquad (22)$$

Using sliding data length of size M, we can write equation (22) as:

$= -\,E\left[\ \left\{\frac{\mu}{2M^2}\sum_{i=k-M+1}^{k}\sum_{j=k-M+1}^{k} f_{\sigma_1\sqrt{2}}\left(\left(e_p(i)+z(i)+\Delta\hat{\boldsymbol{w}}^T(i)v(i)\right)-\left(e_p(j)+z(j)+\Delta\hat{\boldsymbol{w}}^T(j)v(j)\right)\right)\left(q(i)-q(j)\right)\left(u(i)-u(j)\right\}\ \mid u(\mathrm{k})\right]$ (23)

$= -\,E\left[\left\{\frac{\mu}{2M^2}\sum_{i=k-M+1}^{k}\sum_{j=k-M+1}^{k} f_{\sigma_1\sqrt{2}}\left((e_p(i)-e_p(j)+z(i)-z(j)+(\Delta\hat{\boldsymbol{w}}^T(i)v(i)-\Delta\hat{\boldsymbol{w}}^T(j)v(j))\right)\right)\left(q(i)-q(j)\right)\left(u(i)-u(j)\right\}\mid u(\mathrm{k})\right]$ (24)

We have considered following assumptions for the derivation of the bias compensator in the proposed algorithm.

Assumption 1: The measurement noise $z(k)$ is considered α-stable distribution noise with zero mean and input noise $v(k)$ is white Gaussian noise having zero mean.

Assumption 2: The signals $v(k),\ z(k)\ and\ x(k)$ and $\hat{w}(k)$ are statistically independent.

Assumption 3: The non-linear function of the estimation error f($z(k)$), v(k) and q(k) are statistically independent.

Taking Taylor series expansion of $f_{\sigma_1\sqrt{2}}\left((e_p(i)-e_p(j)+z(i)-z(j)+(\Delta\hat{\boldsymbol{w}}^T(i)v(i)-\Delta\hat{\boldsymbol{w}}^T(j)v(j)))\right)$with respect to $\left[\left(e_p(i)-e_p(j)\right)+\left(\Delta\hat{\boldsymbol{w}}^T(i)v(i)-\Delta\hat{\boldsymbol{w}}^T(j)v(j)\right)\right]$ around $\left(z(i)-z(j)\right)$

We can write as:

$$f_{\sigma_1\sqrt{2}}\left((e_p(i)-e_p(j)+z(i)-z(j)+(\Delta\hat{w}^T(i)v(i)-\Delta\hat{w}^T(j)v(j)))\right)= f_{\sigma_1\sqrt{2}}\left(z(i)-z(j)\right)+ f'_{\sigma_1\sqrt{2}}\left(z(i)-z(j)\right)\left[\left(e_p(i)-e_p(j)\right)+\left(\Delta\hat{\boldsymbol{w}}^T(i)v(i)-\Delta\hat{\boldsymbol{w}}^T(j)v(j)\right)\right] + O\left(\left[\left(e_p(i)-e_p(j)\right)+\left(\Delta\hat{\boldsymbol{w}}^T(i)v(i)-\Delta\hat{\boldsymbol{w}}^T(j)v(j)\right)\right]^2\right) \quad (25)$$

$= -\,E\left[\ \left\{\frac{\mu}{2M^2}\sum_{i=k-M+1}^{k}\sum_{j=k-M+1}^{k}\left(f_{\sigma_1\sqrt{2}}\left(z(i)-z(j)\right)+ f'_{\sigma_1\sqrt{2}}\left(z(i)-z(j)\right)\left[\left(e_p(i)-e_p(j)\right)+\left(\Delta\hat{\boldsymbol{w}}^T(i)v(i)-\Delta\hat{\boldsymbol{w}}^T(j)v(j)\right)\right]+\ O\left[\left(e_p(i)-e_p(j)\right)+\left(\Delta\hat{\boldsymbol{w}}^T(i)v(i)-\Delta\hat{\boldsymbol{w}}^T(j)v(j)\right)\right]^3\right)\left(q(i)-q(j)\right)\left(u(i)-u(j)\right)\right\}\ \mid u(\mathrm{k})\right]$ (26)

$= -\ \frac{\mu}{2M^2}\sum_{i=k-M+1}^{k}\sum_{j=k-M+1}^{k} E\left(f_{\sigma_1\sqrt{2}}\left(z(i)-z(j)\right)(q(i)-q(j))\,(u(i)-u(j))\mid u(\mathrm{k})\right)\ +\ E\left(f'_{\sigma_1\sqrt{2}}\left(z(i)-z(j)\right)\left[\left(e_p(i)-e_p(j)\right)+\left(\Delta\hat{\boldsymbol{w}}^T(i)v(i)-\Delta\hat{\boldsymbol{w}}^T(j)v(j)\right)\right](q(i)-q(j))\,(u(i)-u(j))\mid u(\mathrm{k})\right) + E\left(O\left[\left(e_p(i)-e_p(j)\right)+\left(\Delta\hat{\boldsymbol{w}}^T(i)v(i)-\Delta\hat{\boldsymbol{w}}^T(j)v(j)\right)\right]^2(q(i)-q(j))\,(u(i)-u(j))\mid u(\mathrm{k})\right)$ (27)

When k → ∞ , $e_p(k) \to 0$ and using above assumptions, we can write as

$$E\left(f_{\sigma_1\sqrt{2}}\left(z(i)-z(j)\right)(q(i)-q(j))\,(u(i)-u(j))\mid u(\mathrm{k})\right) = E\left[f_{\sigma_1\sqrt{2}}\left(z(i)-z(j)\right)\right]E[(q(i)-q(j))\,(u(i)-u(j))] \quad (28)$$

And, $E[(q(i)-q(j))\,(u(i)-u(j))] = E\left[\left\{\left(e_p(i)+z(i)+\Delta\hat{w}^T(i)v(i)\right)-\left(e_p(j)+z(j)+\ \Delta\hat{w}^T(j)v(j)\right)\right\}(x(i)+v(i)-x(j)-v(j))\right]$ (29)

$= E\left[\left\{\left(z(i)+\Delta\hat{w}^T(i)v(i)\right)-\left(z(j)+\ \Delta\hat{w}^T(j)v(j)\right)\right\}(x(i)+v(i)-x(j)-v(j))\right]$ (30)

$= E\left[\left\{\left(z(i)-z(j)\right)+\left(\Delta\hat{w}^T(i)v(i)-\Delta\hat{w}^T(j)v(j)\right)\right\}(x(i)+v(i)-x(j)-v(j))\right]$ (31)

$= \ \sigma_v^2\left(\Delta\hat{w}^T(i)-\Delta\hat{w}^T(j)\right)$

$= \ -\,\sigma_v^2\left(\hat{w}^T(i)-\hat{w}^T(j)\right)$ (32)

Where $\sigma_v^2 = E\,[v(i)v(i)]$

So $E\ \left(f'_{\sigma_1\sqrt{2}}\left(z(i)-z(j)\right)\left[\left(e_p(i)-e_p(j)\right)+\ \left(\Delta\hat{w}^T(i)v(i)-\Delta\hat{w}^T(j)v(j)\right)\right](q(i)-q(j))\,(u(i)-u(j))\mid u(\mathrm{k})\right) = E\ \left(f'_{\sigma_1\sqrt{2}}\left(z(i)-z(j)\right)\left[\ \left(\Delta\hat{w}^T(i)v(i)-\Delta\hat{w}^T(j)v(j)\right)\right](q(i)-q(j))\,(u(i)-u(j))\mid u(\mathrm{k})\right)\ = 0$ (33)

And,

$E\left(O\left[\left(e_p(i)-e_p(j)\right)+\left(\Delta\hat{\boldsymbol{w}}^T(i)v(i)-\Delta\hat{\boldsymbol{w}}^T(j)v(j)\right)\right]^2(q(i)-q(j))\,(u(i)-u(j))\mid u(\mathrm{k})\right)\ = E\left(O\left[\ \left(\Delta\hat{\boldsymbol{w}}^T(i)v(i)-\Delta\hat{\boldsymbol{w}}^T(j)v(j)\right)\right]^2\ (q(i)-q(j))\,(u(i)-u(j))\mid u(\mathrm{k})\right)=0$ (34)

Combining equation (32), equation (33) and equation (34), we can write the value of biased compensator as:

So $C(k) = \left[\ \sigma_v^2\ \frac{\mu}{2M^2}\sum_{i=k-M+1}^{k}\sum_{j=k-M+1}^{k} f_{\sigma_1\sqrt{2}}\left(z(i)-z(j)\right)\left(\ \hat{\boldsymbol{w}}(i)-\ \hat{\boldsymbol{w}}(j)\ \right)\ \right]$ (35)

So the update equation of proposed BC-CIM-MEE becomes:

$\hat{\boldsymbol{w}}(\mathrm{k}+1) = \left[\hat{\boldsymbol{w}}(\mathrm{k})\ +\ \sigma_v^2\ \frac{\mu}{2M^2}\ \sum_{i=k-M+1}^{k}\sum_{j=k-M+1}^{k} f_{\sigma_1\sqrt{2}}\left(z(i)-z(j)\right)\left(\ \hat{\boldsymbol{w}}(i)-\ \hat{\boldsymbol{w}}(j)\ \right)\ \right]+$

$$\frac{\mu}{2M^2}\ \{\sum_{i=i=k-M+1}^{k}\sum_{j=i=k-M+1}^{k} f_{\sigma_1\sqrt{2}}(q(i)-q(j))(q(i)-q(j))(u(i)-u(j)\} - \rho\left\{\frac{\hat{w}(k)}{N\sqrt{2\pi}\sigma_2^3}\, e^{-\frac{\hat{w}^2(k)}{2\sigma_2^2}}\right\} \quad (36)$$

So the proposed BC-CIM-MEE algorithm provides better estimation characteristics in the presence of both noisy input and impulsive measurement noise. When input signal is not noisy that is input noise variance is zero, the proposed algorithm behaves like CIM-MEE algorithm.

Generally the input noise variance is not available, so it should be estimated. In past SungEun Jo, and Sang Woo Kim have proposed a method to estimate input noise variance [23].

4 Simulation Results

In this section we evaluate the performance of proposed BC-CIM-MEE algorithm in sparse system identification for different sparsity level. We have taken the normalized mean square deviation (NMSD) error as a criterion to evaluate the performance of sparse system identification which can be written as:

$$\text{NMSD}= 10^* \log_{10} \text{E}\left[\frac{\|w_0-\hat{w}(k)\|^2}{\|w_0\|^2}\right] \quad (37)$$

Both input signal and input noise are considered white Gaussian sequence with zero mean. We have considered the alpha stable noise as impulsive noise for evaluating the performance of the proposed algorithm.

The characteristics function of alpha stable noise is given as:

$$\text{g(t)}= \exp\{j\delta(t)-\gamma|t|^{\alpha}\,[1+j\,\beta\, sgn(t)S(t,\alpha)\} \quad (38)$$

Where

$$S(t,\alpha)=\begin{cases}\tan\left(\frac{\alpha\pi}{2}\right)\ if\ \alpha\ \neq 1\\ \frac{2}{\pi}log|t|\ if\ \alpha=1\end{cases} \quad (39)$$

Where α ∈ (0,2] is the characteristic factor, $\beta \in$ (-1, 1) is the symmetry parameter and $-\infty<\delta<+\infty$ is location parameter, $\gamma>0$ is dispersion parameter. The characteristics factor α measures the weight of tail of impulsive noise. It is inversely proportional to tail of impulsive noise i.e. larger the value of α, smaller the weight of tail of impulsive noise. γ is same as variance as in Gaussian distribution. We can define the impulsive noise model v(k) as V$(\alpha,\beta,\gamma,\delta)$. We have taken measurement noise as V(1.3,0,0.2,0).

The Simulation are averaged over fifty Monte Karle runs.We have performed seven experiments to evaluate the performance of sparse identification. We have taken sytem length N equal to 128.

In the first experiment, we have examined the performance of BC-CIM-MEE algorithm and have compared with MEE, CIM-MEE for different sparsity rates.We have taken different saprsity rates , K as $\frac{1}{32}$, $\frac{1}{16}$ and $\frac{1}{8}$.

Where sparsity rate,

$$\text{K} = \frac{number\ of\ non\ zero\ coefficients}{total\ number\ of\ coefficients} \quad (40)$$

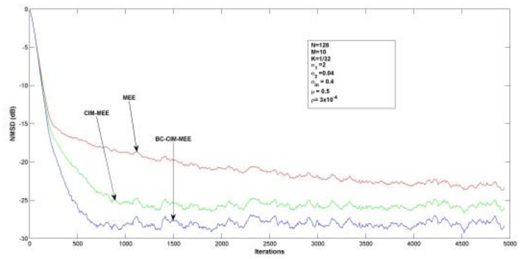
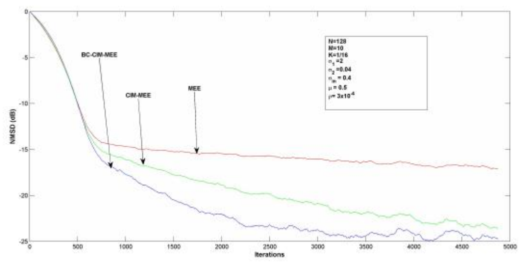


Fig.1 Estimation Performance in term of NMSD for sparsity level , K = $\frac{1}{32}$



(38)**Fig.2 Estimation Performance in term of NMSD for sparsity level , K = $\frac{1}{16}$**

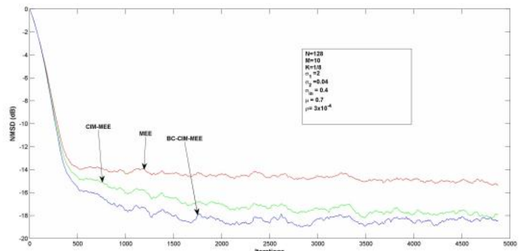


Fig.3 Estimation Performance in term of NMSD for sparsity level , K = $\frac{1}{8}$

Figures (1), (2) and (3) show that the proposed algorithm perform better than MEE and CIM- MEE algorithm in every case of sparsity level.

In second experiment we investigate the effect of step size on the estimation performance of system. We have considered μ={0.2 0.5,0.7,0.8 ,1}. From Figure 4, it is clear that as step size increases,convergence speed increases but NMSD error also increases ,so it should be chosen appropriately to improve the performance of system . So we have chosen μ =0.5 in all simulations.

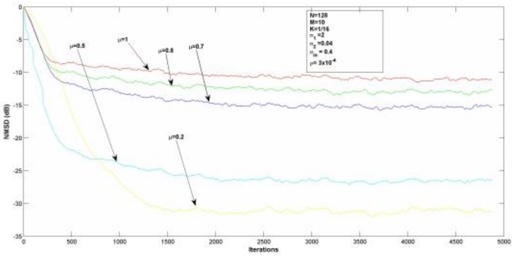


Fig.4 Effect of step size, μ on the estimation performance of proposed algorithm

In the next experiment we investigate the effect of σ_{in} on the estimation performance of system. We take the values of σ_{in} = {0.2 0.4 0.6 0.8 } and all other parameters are set same as in previous simulation.

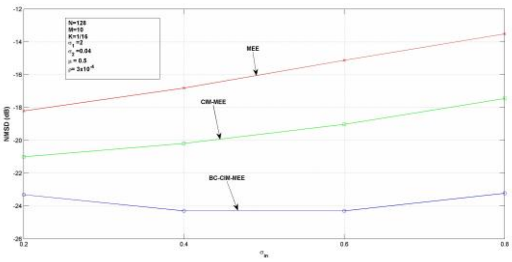


Fig.5 Effect of input noise variance, σ_{in} on the estimation performance of proposed algorithm

From figure 5, it is clear that estimation performance of proposed algorithm is consistant to value of σ_{in} . However, in MEE and CIM-MEE, the steady state error increases with increase in the value of σ_{in} .

In next simulation, we observe the effect of MEE kernal width σ_1 on the estimation performance of the system. We have taken σ_1 = {1,2 ,3 ,4 ,5 ,6,7}. From Figure 6, it is clear that the performance of proposed algorithm is varied according to the value of σ_1 . However, it shows better performance for σ_1 =2. So we have taken σ_1 =2 in all other simulations.

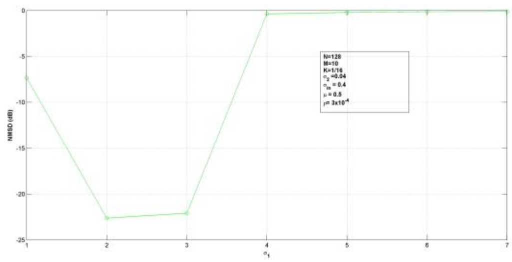


Fig.6 Effect of MEE kernal width, σ_1 on the estimation performance of proposed algorithm

In next experiment, we examine the effect of CIM kernal width , σ_2 on the performance of proposed algorithm. As Figure 7 shows that the estimation performance of proposed algorithm is varied according to σ_2. So it must be chosen carefully so that it approximates l_0-norm. The proposed algorithm has lowest NMSD for σ_2 = 0.04. So we have taken σ_2 = 0.04 in all simulations.

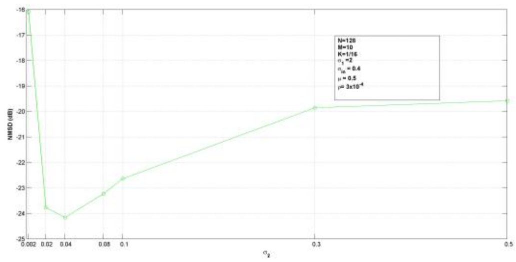


Fig.7 Effect of σ_2 on the estimation performance of proposed algorithm

In the next experiment, we examine the combined effect of step size, μ and input noise variance, σ_{in} on the estimation performance of the proposed algorithm. We have taken all other parameters same as in above experiment.

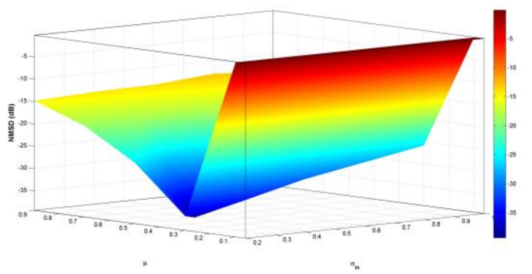


Fig.8 Combined Effect of step size, μ and input noise variance, σ_{in} on the estimation performance of proposed algorithm

From Figure8, it is clear that when step size, μ and input noise variance, σ_{in} increase, the steady state error (NMSD) increases. For μ =0.5 and σ_{in} = 0.4, the proposed algorithm has least NMSD error. So we have taken μ =0.5 and σ_{in} = 0.4 in simulations.

In the last, we examine the combined effect of input noise variance, σ_{in} and MEE kernal width ,σ_1 on the estimation performance of system. From figure 9, it is clear that proposed algorithm outperforms for σ_1 =2 and σ_{in} =0.4.

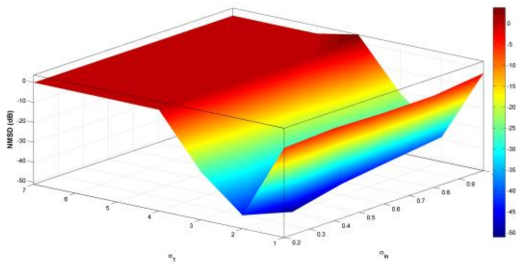


Fig.9 Combined Effect of input noise variance, σ_{in} and and MEE kernal width ,σ_1 on the estimation performance of proposed algorithm

5 Conclusion

This paper presents BC-CIM-MEE algorithm ,which outperforms in the presence of noise input against impilsive measurement noise in sparse system identification. The proposed algorithm is unbiased against noisy input. We also observe the effect of step size μ , input noise variance σ_{in} , MEE kernal width σ_1 and CIM kernal width σ_2 on the estimation performance of proposed algorithm. The proposed algorithm performs better than MEE and CIM-MEE algorithms in all sparsity rates.

References

[1]. A.H. Sayed, Fundamentals of Adaptive Filtering, Wiley-IEEE Press, 2003

[2]. L. Cheng, K. Wang, M. F. Ren and X. Y. Xu,, Comprehensive Analysis of Multipath Estimation Algorithms in the Framework of Information Theoretic Learning, *IEEE Access*, vol. 6, pp. 5521-5530,2018. https://doi.org /10.1109/ACCESS.2018.2793896

[3]. Wentao Ma, Hua Qu, Guan Gui, Li Xu, Jihong Zhao, Badong Chen, Maximum correntropy criterion based sparse adaptive filtering algorithms for robust channel estimation under non-Gaussian environments, Journal of the Franklin Institute, Volume 352, Issue 7,2015, pp. 2708-2727, https://doi.org/10.1016/j.jfranklin.2015.03.039

[4].C. D. Richmond, Adaptive array processing in non-Gaussian environments, *Proceedings of 8th Workshop on Statistical Signal and Array Processing, Corfu, 1996,* pp. 562-565. https://doi.org/10.1109/SSAP.1996.534939

[5]. Yahong Rosa Zheng, Vítor H. Nascimento, Two variable step-size adaptive algorithms for non-Gaussian interference environment using fractionally lower-order moment minimization, Digital Signal Processing, Volume 23, Issue 3, 2013,pp.831-844. https://doi.org/10.1016/j.dsp.2012.12.019

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

[7]. Badong Chen, Yu Zhu, Jinchun Hu, Jose C. Principe, Chapter 2 - Information Measures, System Parameter Identification, Elsevier, 2013, pp. 13-28. https://doi.org/10.1016/B978-0-12-404574-3.00002-6

[8]. Badong Chen, Yu Zhu, Jinchun Hu, Jose C. Principe, Chapter 4 - System Identification Under Minimum Error Entropy Criteria, System Parameter Identification, Elsevier, 2013, pp. 61-165.

https://doi.org/10.1016/B978-0-12-404574-3.00004-X

[9].R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B. Methodological,vol. 58, no. 1, pp. 267–288, 1996.

[10]. D. L. Donoho, Compressed sensing, IEEE Transactions on Information Theory, vol. 52, no. 4, 1289-1306, 2006.

[11]. Radhika, S & Sivabalan, A., ZA-APA with zero attractor controller selection criterion for sparse system identification, Signal, Image and Video Processing, September 2017

[12]. Chen, Y. Gu, and A. Hero., Sparse LMS for system identification, *In Proc. IEEE Int. Conf. Acoust. Speech and Signal Processing*, pages 3125–3128, April 2009.

[13]. Guan Gui, Wei Peng, Li Xu, Beiyi Liu and Fumiyuk Adachi, Variable-step-size based sparse adaptive filtering algorithm for channel estimation in broadband wireless communication systems,

EURASIP Journal on Wireless Communications and Networking, 2014.

[14].Yingsong Li , Yanyan Wang, Tao Jiang, Sparse-aware set-membership NLMS algorithms and their application for sparse channel estimation and echo cancelation, *Elsevier, Int. J. Electron. Commun. (AEÜ)* 70 (2016) 895–902

[15]. B. Babadi, N. Kalouptsidis, and V. Tarokh, SPARLS: The sparse RLS algorithm, *IEEE Trans. Signal Process.*, 58(8): 4013–4025, 2010

[16]. Ran Meng, R. C. de Lamare and V. H. Nascimento, Sparsity-aware affine projection adaptive algorithms for system identification, *Sensor Signal Processing for Defence (SSPD 2011)*, London, 2011, pp.1-5.

[17]. Omid Taheri, Sergiy A. Vorobyov, Reweighted l1-norm penalized LMS for sparse channel estimation and its analysis,Signal Processing, Volume 104, 2014, Pages 70-79. https://doi.org/10.1016/j.sigpro.2014.03.048

[18]. Y. Wang, Y. Li, F. Albu and R. Yang, Sparse Channel Estimation Using Correntropy Induced Metric Criterion Based SM-NLMS Algorithm, *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, San Francisco, CA, 2017, pp. 1-6. https://doi.org/10.1109/WCNC.2017.7925628

[19]. Seth, S & Principe, J 2008, Compressed signal reconstruction using the correntropy induced metric, *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008,pp.* 3845-3848. https://doi.org/ 10.1109/ICASSP.2008.4518492

[20]. Jin, Zhan & Li, Yingsong & Wang, Yanya, An Enhanced Set-Membership PNLMS Algorithm with a Correntropy Induced Metric Constraint for Acoustic Channel Estimation, Entropy, 19. 281. https://doi.org/10.3390/e19060281

[21]. Wu, Zongze and Peng, Siyuan and Ma, Wentao and Chen, Badong and Principe, Jose C., Minimum Error Entropy Algorithms with Sparsity Penalty Constraints, Entropy, 17(5), pp. 3419-343, 2015.

[22]. B. Kang, J. Yoo, P. Park, Bias-compensated normalised lms algorithm with noisy input, IEEE Electronics Letters 49 (8) , pp. 538–539, 2013.

[23]. S. Jo, S. W. Kim, Consistent normalized least mean square filtering with noisy data matrix, IEEE Transactions on Signal Processing 53 (6), 2112–2123, 2005.