Sparse Robust Distributed Estimation by Diffusion Adaptation

Sowjanya Modalavalasa, Upendra Kumar Sahoo, Ajit Kumar Sahoo,Ajitesh Yadav
Electronics and Communication Department
National Institute of Technology
Rourkela, Odisha
India, 769008
Email: sowjanya.mar15@gmail.com,(sahooupen,ajitsahoo@nitrkl.ac.in)

***Abstract*—The least squares based cost functions are sensitive to outliers in the measured data. The presence of outliers is considered as impulsive noise. In practical scenarios, the co channel interference, saturation effects, non linearity of the measuring instruments, atmospheric conditions and malfunction of sensors will result in outliers or impulsive noise. The robust function obtained by considering the error as the linear combination of sign preserving basis functions is found to be robust against outliers in the desired data. In many practical applications, the parameter to be estimated can be sparse in nature, i.e. only a few elements are large values and the rest are insignificantly small. In such sparse systems, if the prior information about the sparsity is known, then the known information can be incorporated in the cost function as a regularization function. A robust sparse diffusion algorithm is proposed in this work, which is robust against outliers in the desired data and performs better than the existing algorithms in sparsity underlying systems. Simulations performed for different cases of outliers conditions and sparsity conditions validate that the proposed method outperforms the state of the art methods.**

***Keywords*—Distributed processing, Impulsive noise, Sparsity, Outliers, Robust Sparse Diffusion**

I. Introduction

Wireless sensor networks have wide range of applications ranging from precision agriculture to cognitive radio[1], [2]. Centralized processing involves a fusion center, which processes all the data received from the network and broadcasts back the final estimate. This fusion center based learning results in huge communication burden if the sensors are remotely placed and which in turn results in more power requirements. Distributed signal processing overcomes all these problems, by enabling each sensor with processing capabilities. Each sensor in a distributed network will sense the noisy measurements from a geographic area of interest, processes them and share their estimates and data with their neighborhood nodes. In distributed estimation, there are three major strategies namely incremental [3], diffusion [4] and consensus [2]. The incremental strategy requires a Hamiltonian path for communication between the nodes, which is a NP hard problem. Moreover incremental strategy will fail if there are any link failures. Diffusion strategy outperforms the consensus strategy [5], [2]. Diffusion strategy[6] finds applications in many areas including bacteria foraging[7] ,fish schooling[8], bird flight formulations[9], cognitive radio[10]. Most of the techniques utilizes l_2 or l_1 norm based cost function for optimization. The noise present in the estimation process is assumed to be Gaussian in nature. But in practical scenarios, the noise will be heavy tailed unlike Gaussian due to saturation effects, non linearities, link failures, atmospheric conditions etc. In such scenarios, the noise cannot has to be treated as impulsive noise or outliers. The performance of the algorithms based on least squares cost function will degrade since it is sensitive to outliers. Robust cost functions have been proposed in the literature to handle perturbations in the desired data. In such robust cost functions[11], [12], the error is replace by a score function generated using the error such that it suppresses outliers. The order statistics based methods use mean, median and similar properties of the data for adaptation during the time iterations. In [13], the error is considered as linear combination of sign preserving basis functions, which resulted in a Robust diffusion algorithm. In many practical applications, the parameter of interest can be sparse i.e. only a few elements are large values and the rest are insignificantly small [14], [15], [16]. In such sparse systems, if the prior information about the sparsity is known, then the known information can be incorporated in the cost function as a regularization function. In [14], a sparse based cost function is proposed, where the regularization function utilizes the prior sparse information. Since it is based on least squares criteria, the performance will degrade in presence of outliers.

A novel algorithm is proposed in this work, which incorporates the time varying non linear error and regularization function exploiting sparsity in its cost function. This diffusion based distributed estimation algorithms is robust against outliers in desired data and performs well if the parameter to be estimated is sparse.

II. Problem Formulation

Consider a wireless sensor network having $N = 20$ nodes fully connected as shown in Fig. 1. Each node will sense its noisy data $\mathbf{x}_{k,i}, y_{k,i}$,which are related as

$$y_{k,i} = \mathbf{x}_{k,i}^T \mathbf{w}^0 + v_{k,i}, \tag{1}$$

where $\mathbf{w}^0$ is the parameter to be estimated and $v_{k,i}$ is the noise at k^{th} node in i^{th} iteration. Each and every node in the network will have access to its neighbor nodes, hence they will share the information throughout the network. The objective of this work is to estimate the parameter of interest $\mathbf{w}$ by utilizing all the information collected across the network in a distributed and iterative manner. This can be defined as the cost function as below

$$J(\mathbf{w}) = \arg\min_{\mathbf{w}} \left\|\mathbf{y}(i) - \mathbf{X}^T(i)\,\mathbf{w}\right\|_*, \quad (2)$$

where $\|\|_*$ represents norm and $\mathbf{y}, \mathbf{X}$ define the desired and input data obtained from all the nodes in the network up to i^{th} time. Thus, the (2) is not distributed in nature. To obtain distributed processing the cost function in (2) is redefined as sum of N local cost functions as given below

$$J(\mathbf{w}) = \arg\min_{\mathbf{w}} \sum_{k=1}^{N} \left\|\mathbf{y}_k(i) - \mathbf{X}_k^T(i)\,\mathbf{w}\right\|_*\ , \quad (3)$$

where $\mathbf{y_k}(n), \mathbf{X_k}(n)$ are the desired and input data obtained at k^{th} node up to i^{th} time. Now, the objective of this work is to optimize (3).

III. PROPOSED METHOD

The cost function in usually l_2 or l_1 norm of the error as in [2], [4]. But these algorithms are not robust against outliers. If there are outlier in the desired data, then their performance will degrade. Motivated from [14] [13], the robust cost function exploiting sparsity is defined as below:

$$J(\mathbf{w}) = \arg\min_{\mathbf{w}} \sum_{k=1}^{N} \beta\left(y_k(i) - \mathbf{x}_k^T(i)\,\mathbf{w}\right) + \zeta f(\mathbf{w}), \quad (4)$$

where $\beta()$ is a maximum likelihood function defined as in [17] and ζ is a regularization constant. The regularization function $f(\mathbf{w})$ should be convex and two different regularization functions are considered in this work motivated from [14],[17]. The steepest decent solution of the proposed method is given as

$$\mathbf{w}_{k,i} = \mathbf{w}_{k,i-1} - \mu \nabla_{\mathbf{w}_{k,i-1}}\left(J(\mathbf{w}_{k,i-1})\right), \quad (5)$$

where μ is the step size. From [17][13], the (4) is redefined for adapt and the combine diffusion strategy as below

$$\begin{aligned} \psi_{k,i} &= \mathbf{w}_{k,i-1} + \mu \sum_{p \in N_k} c_{p,k} \mathbf{x}_{k,i}^T h_{k,i}(e_k(i)) \\ &\quad - \zeta \partial_{\mathbf{w}_{k,i}}\left(f(\mathbf{w}_{k,i})\right) \\ \mathbf{w}_{k,i} &= \sum_{p \in N_k} a_{p,k} \psi_{p,i} \end{aligned} \quad (6)$$

where $h_{k,i}(e(i))$ is defined as linear combination of B sign preserving basis functions as given below

$$h_{k,i}(e_k(i)) = \boldsymbol{\alpha}_{k,i}^T \boldsymbol{\phi}_{k,i}, \quad (7)$$

where $\boldsymbol{\alpha}_k$ represents the coefficient vector corresponding to B_k basis functions vector $\boldsymbol{\phi}_k$ defined as below

$$\begin{aligned} \boldsymbol{\alpha}_{k,i} &\triangleq \left[\alpha_{k,i}(1), \alpha_{k,i}(2), ..., \alpha_{k,i}(B)\right]^T \\ \boldsymbol{\phi}_{k,i} &\triangleq \left[\psi_{k,1}(e_k(i)), \psi_{k,2}(e_k(i)), ..., \psi_{k,B}(e_k(i))\right]^T \end{aligned} \quad (8)$$

From [17], the basis functions that are able to suppress the impulse samples are given as below

$$\begin{aligned} \psi_{k,1}(x) &= x, \\ \psi_{k,b}(x) &= \tanh((b-1)x), b = 2, 3, .., B. \end{aligned} \quad (9)$$

A. Zero-Attracting Robust Sparse Diffusion (ZA RSD)

The new cost function obtained by incorporating l_1 regularization function is defined as below

$$J(\mathbf{w}) = \arg\min_{\mathbf{w}} \sum_{k=1}^{N} \beta\left(y_k(i) - \mathbf{x}_k^T(i)\,\mathbf{w}\right) + \|\mathbf{w}\|_1 \quad (10)$$

The steepest decent solution is given as

$$\begin{aligned} \boldsymbol{\psi}_{k,i} &= \mathbf{w}_{k,i-1} + \mu_k \sum_{p \in \mathcal{N}_k} c_{p,k} \mathbf{x}_{k,i}^T h_{k,i}(e_k(i)) \\ &\quad - \zeta_{za}\, sign(\mathbf{w}_{k,i-1}) \\ \mathbf{w}_{k,i} &= \sum_{p \in N_k} a_{p,k} \boldsymbol{\psi}_{p,i} \end{aligned} \quad (11)$$

The regularization term uniformly shrinks all elements of the vector $\mathbf{w}$, hence it is referred as zero attracting (ZA) algorithm.

Algorithm 1 : Robust Sparse Diffusion

Initialization: $\mathbf{w_{k,-1}} = \mathbf{0_{L\times 1}}$, $B = 2, \epsilon = 10^{-6}, \vartheta = 0.9, \Pi \triangleq I - \frac{11^T}{B}$. For each time $i \geq 0$,

Incremental Step(repeat):

for k=1:N

- Calculate the error $e_k(i) = y_k(i) - x_{k,i}^T \mathbf{w}$
- Calculate the error non linearity using below equations:
  - $\psi_{k,b}(i) = \psi_{k,b}(e_k(i)), b = 1, 2, ..., B$
  - $\boldsymbol{\phi}_{k,i} = col\{\psi_{k,1}(i), ..., \psi_{k,B}(i)\}$
  - $\mathbf{R}_{\boldsymbol{\phi}_{k,i}} = \vartheta \mathbf{R}_{\boldsymbol{\phi}_{k,i-1}} + (1-\vartheta)\boldsymbol{\phi}_{k,i}\boldsymbol{\phi}_{k,i}^T$
  - $\psi'_{k,b}(i) = \psi'_b(e(i)), b = 1, 2, ..., B$
  - $\boldsymbol{\phi}_{k,i}{}' = col\left\{\boldsymbol{\phi}_{k,1}{}'(i), ..., \boldsymbol{\phi}_{k,B}{}'(i)\right\}$
  - $\tilde{\boldsymbol{\phi}}_{k,i}{}' = \vartheta \tilde{\boldsymbol{\phi}}'_{k,i-1} + (1-\vartheta)\tilde{\boldsymbol{\phi}}'_{k,i-1}$
  - $\boldsymbol{\theta}_{k,i} = 2\Pi\left(\mathbf{R}_{\boldsymbol{\phi}_{k,i}} \boldsymbol{\alpha}_{k,i} - \tilde{\boldsymbol{\phi}}_{k,i}{}'\right)$
  - $\tau_k(i) = \vartheta \tau_k(i-1) + (1-\vartheta)\frac{\|\mathbf{x}_{k,i}\|^2}{M}$
  - $\breve{\lambda}_k(i) = sigmoid\left[\left(\boldsymbol{\alpha}_{k,i-1}^T \tilde{\boldsymbol{\phi}}_{k,i}{}'\right)\tau_k(i)\right]$
  - $\lambda_k(i) = \breve{\lambda}_k(i)\frac{\min\{\alpha_{k,i-1}(b), 1 \leq b \leq B\}}{\|\boldsymbol{\alpha}_{k,i}\|_\infty + \varepsilon}$
  - $\boldsymbol{\alpha}_{k,i} = \boldsymbol{\alpha}_{k,i-1} - \lambda_k(i)\boldsymbol{\theta}_{k,i}$
- Update the non linear error vector as $h_k(i) = \boldsymbol{\alpha}_{k,i}^T \boldsymbol{\phi}_{k,i}$
- Weight update for ZA RS-LMS:

$$\begin{aligned} \boldsymbol{\psi}_{k,i} &= \mathbf{w}_{k,i-1} + \mu \sum_{p \in \mathcal{N}_k} c_{p,k} \mathbf{x}_{k,i}^T h_{k,i}(e_k(i)) \\ &\quad - \zeta_{za}\, sign(\mathbf{w}_{k,i-1}) \end{aligned}$$

- Weight update for RZA RS-LMS:

$$\begin{aligned} \boldsymbol{\psi}_{k,i} &= \mathbf{w}_{k,i-1} + \mu \sum_{p \in \mathcal{N}_k} c_{p,k} \mathbf{x}_{k,i}^T h_{k,i}(e_k(i)) \\ &- \zeta_{rza} diag\left\{\frac{1}{\varepsilon + |w_1|}, \frac{1}{\varepsilon + |w_2|}, .., \frac{1}{\varepsilon + |w_M|}\right\} sign(\mathbf{w}_{k,i-1}) \end{aligned}$$

end

Diffusion step(repeat):

$$\mathbf{w}_{k,i} = \sum_{p \in N_k} a_{p,k} \boldsymbol{\psi}_{p,i}$$

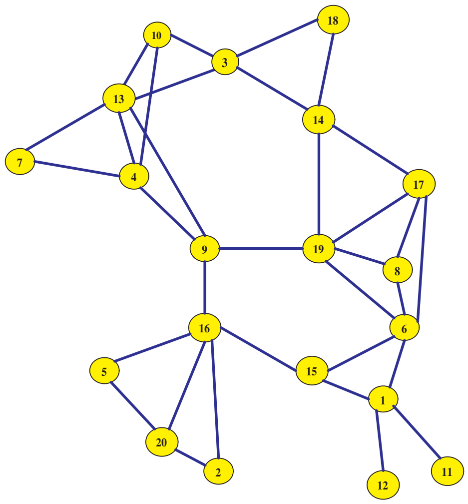

Fig. 1: 20 nodes network

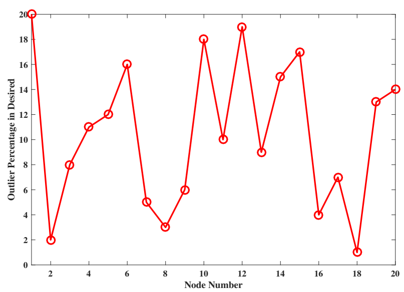

Fig. 2: Outlier percentages in the desired data at each node in the network

B. Reweighted Zero-Attracting Robust Sparse LMS (RZA RSD)

The best measure of sparsity is the l_0 norm. Hence, incorporating the surrogate approximation of l_0 norm as the regularization parameter, the cost function is defined as below

$$J(\mathbf{w}) = \arg\min_{\mathbf{w}} \sum_{k=1}^{N} \beta\left(y_k(i) - \mathbf{x}_k^T(i)\,\mathbf{w}\right) + \zeta_{rza} \sum_{l=1}^{M} \frac{|w_l|}{\varepsilon + |w_l|} \tag{12}$$

The steepest decent solution results in the reweighted zero attracting (RZA) algorithm defined as below

$$\begin{aligned}\boldsymbol{\psi}_{k,i} &= \mathbf{w}_{k,i-1} + \mu \sum_{p\in\mathcal{N}_k} c_{p,k}\mathbf{x}_{k,i}^T h_{k,i}\left(e_k(i)\right) \\ &\quad -\zeta_{rza} diag\left\{\frac{1}{\varepsilon+|w_1|}, \frac{1}{\varepsilon+|w_2|}, \ldots\ldots, \frac{1}{\varepsilon+|w_M|}\right\} sign(\mathbf{w}_{k,i-1}) \\ \mathbf{w}_{k,i} &= \sum_{p\in N_k} a_{p,k}\boldsymbol{\psi}_{p,i}\end{aligned} \tag{13}$$

where ϵ is a very small number ($\epsilon = 0.1$ in this work) and M is the length of the parameter vector $\mathbf{w}$. The ZA and RZA robust sparse diffusion algorithms are depicted in Algorithm 1.

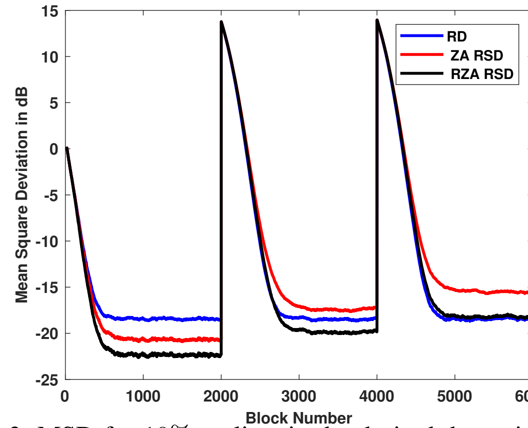

Fig. 3: MSD for 10% outliers in the desired data with $k = 50$

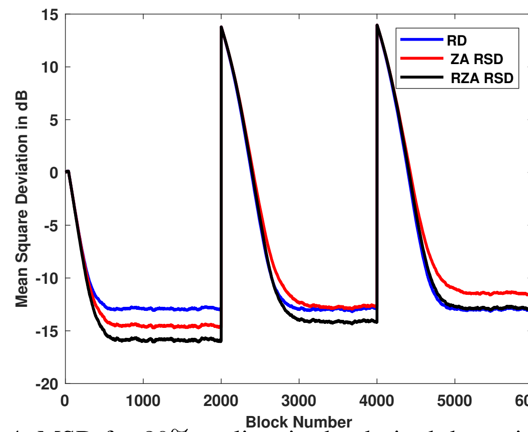

Fig. 4: MSD for 30% outliers in the desired data with $k = 50$

IV. Simulation Results and Discussion

The proposed algorithms are validated for different cases of outliers distribution and sparsity ratios. A 20 nodes network depicted in Fig. 1. is considered for all simulation based experiments in this work. The input generated is Gaussian with 0.1 variance and zero mean. The parameter of interest or the parameter to be estimated $\mathbf{w}$ is considered to be of size $M = 50$. The desired data is generated using the relation in (1). The noise present in the desired data is a combination of Gaussian and impulsive noise or outliers and it is generated using $\epsilon-$ contamination model, whose probability density function is given as

$$f_\nu = (1-\epsilon)\,\mathrm{N}\left(0, \bar{\sigma}_\nu^2\right) + \epsilon\mathrm{N}\left(0, k\bar{\sigma}_\nu^2\right), \tag{14}$$

where ϵ is the contamination ratio. If 10% outliers are present, then $\epsilon = 0.1$. $\bar{\sigma}_\epsilon^2$ is the variance and $k >> 1$ is the scale. The system will revert back to Gaussian noise case if $\epsilon = 0$. The outliers percentage is considered as 10% and 30% for low and high contamination scenarios in the network. The corresponding mean square deviation (MSD) figures are depicted in Fig. 3 and Fig. 4 for 10% and 30% outliers respectively. For the first 2000 iterations , the vector $\mathbf{w^0}$ is assumed to have only one non zero element such that the sparsity ratio is 1/50. For the next 2000 iterations, 25 elements are considered to have non zero elements and the rest to be zeros such that the sparsity ratio is 25/50. For the last

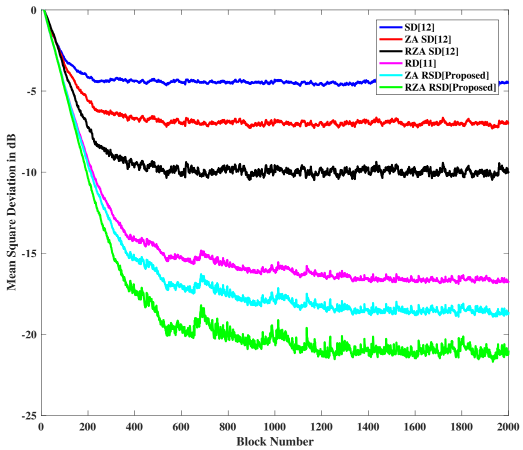

Fig. 5: MSD for different outliers percentages in the desired data with $k = 50$

2000 iteration, all the elements are considered as non zeros resulting in non sparse system. The regularization constants are empirically chosen as $\zeta_{ZA} = 0.001, \zeta_{RZA} = 0.00025$ for both the cases. The step size is considered as $\mu = 0.08$ for all the nodes in the network. All the simulation results depicted are obtained by averaging over 100 random independent experiments. The proposed algorithm is compared with sparse diffusion LMS (SD) proposed in [15] in Fig. 5 for different percentages of outliers across the network as shown in Fig. 2. The proposed algorithm outperforms the SD, ZA SD, RZA SD[15] algorithms from the literature. When the sparsity ratio is $1/50$, the ZA RSD and RZA RSD perform better than RSD[13]. If the sparsity ratio is $25/50$ then the performance of ZA RSD degrades compared to the other two counterparts. If the system is made completely non sparse, the RSD and RZA RSD will have similar performance. Hence the proposed algorithms will perform better in presence of sparsity and will perform as good as the unregularized algorithm if the system is non sparse.

V. Conclusion

A new algorithm has been proposed which improves the performance of Robust diffusion by exploiting sparsity thereby introducing a sparsity based regularization term in the cost function. The algorithm is validated for distributed estimation problem for different percentages of outliers in the desired data and for different sparsity ratio scenarios. Simulation results show that the proposed algorithms ZA RSD and RZA RSD outperforms the RSD, SD, ZA SD, RZA SD in presence of sparsity in the parameter of interest. The RZA RSD algorithm will perform equivalent to RSD if the system is completely non-sparse.

References

[1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *icassp*, vol. 1. Citeseer, 2001, pp. 2033–2036.

[2] A. H. Sayed *et al.*, "Adaptation, learning, and optimization over networks," *Foundations and Trends® in Machine Learning*, vol. 7, no. 4-5, pp. 311–801, 2014.

[3] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064–4077, 2007.

[4] F. S. Cattivelli and A. H. Sayed, "Diffusion lms strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, 2009.

[5] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6217–6234, 2012.

[6] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, 2008.

[7] J. Chen, X. Zhao, and A. H. Sayed, "Bacterial motility via diffusion adaptation," in *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*. IEEE, 2010, pp. 1930–1934.

[8] S.-Y. Tu and A. H. Sayed, "Tracking behavior of mobile adaptive networks," in *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*. IEEE, 2010, pp. 698–702.

[9] F. S. Cattivelli and A. H. Sayed, "Modeling bird flight formations using diffusion adaptation," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2038–2051, 2011.

[10] ——, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 1917–1932, 2011.

[11] S. Modalavalasa, U. K. Sahoo, A. K. Sahoo, and S. Kumar, "Diffusion minimum generalized rank norm over distributed adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal and Information Processing over Networks*, 2019.

[12] S. Kumar, U. K. Sahoo, A. K. Sahoo, and D. P. Acharya, "Diffusion minimum-wilcoxon-norm over distributed adaptive networks: Formulation and performance analysis," *Digital Signal Processing*, vol. 51, pp. 156–169, 2016.

[13] S. Al-Sayed, A. M. Zoubir, and A. H. Sayed, "Robust distributed estimation by networked agents," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 3909–3921, 2017.

[14] P. Di Lorenzo and A. H. Sayed, "Sparse distributed learning based on diffusion adaptation," *IEEE Transactions on signal processing*, vol. 61, no. 6, pp. 1419–1433, 2012.

[15] Y. Liu, C. Li, and Z. Zhang, "Diffusion sparse least-mean squares over networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4480–4485, 2012.

[16] G. Mateos, J. A. Bazerque, and G. B. Giannakis, "Distributed sparse linear regression," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5262–5276, 2010.

[17] S. Al-Sayed, A. M. Zoubir, and A. H. Sayed, "Robust adaptation in impulsive noise," *IEEE Transactions on Signal Processing*, vol. 64, no. 11, pp. 2851–2865, 2016.