Robust Estimation in Distributed Wireless Sensor Network

Sananda Kumar, Ajit Kumar Sahoo, Debiprasad Priyabrata Acharya
Electronics and Communication Department
National Institute of Technology
Rourkela, Odisha
India, 769008
Email: sananda4u@gmail.com, {ajitsahoo, dpacharya@nitrkl.ac.in}

***Abstract*—Practical environment is always endowed with high impulsive noise in addition to Gaussian noise. Traditional methods like least square method are very limited in their performance for estimating the parameter of interest. To address such problem a robust distributed block diffusion Huber algorithm is proposed in this paper.**

Keywords*—*Diffusion strategy, Sensor network, Huber loss function

I. Introduction

Distributed algorithms are employed in large geographic area in order to reduce the burden of centralized processor and minimize the cost of communication, thus booming the lifetime of the sensor networks. These distributed sensors are well equipped with computation ability, memory, processor and trans receiver antennae [1]. Based on the mode of communication among these sensor nodes they are broadly classified into incremental, and diffusion type. In large geographic area, dynamically determining of proper Hamiltonian path is not feasible, and for failure of a single sensor node the whole network fails to achieve the desired estimation [2]. For, such reason the diffusion mode of cooperation is given emphasized in this paper. In diffusion mode of cooperation [2]–[5] each of the sensor node communicates among it's neighborhood for estimation of the global parameter [3] following two basic steps (1) adapt (2) combine. Based on the steps followed they are referred as ATC (Adapt Then Combine) or CTA (Combine Then Adapt).

In addition to Gaussian noise, the data in practical environment are always corrupted by impulsive noise [6]. These impulsive random noises are generated in the environment due to electromagnetic interference, co-channel interference, node failure, presence of non-linearities, saturation effect of sensor nodes and are in general termed as outliers. These outliers have a heavy tailed noise distribution for which the traditional least square methods become much sensitive and they fail to estimate the global parameter of interest.

So, to address such issues a robust algorithm is proposed to work in a distributed environment using block diffusion mode of cooperation among the sensor nodes. Implementation of diffusion algorithm in block method [7], [8] reduces communication cost among the nodes by block size L. Robust algorithms are broadly classified into three categories [9] a) L-estimator b) M-estimator c) R-estimator. The Huber a M-estimator type, basically refered as maximum likelihood type estimator under *least favorable* distribution. A Huber adaptive filter [10]–[12] minimizes a Huber's objective function which is a hybrid of both l_1 and l_2 norm. This adaptive filters down weights the outliers present in data as l_1 norm and behaves like a least square for rest of the signal.

978-1-4673-6540-6/15/$31.00 ©2015 IEEE

Capital letters are used for matrices and small letters for vectors and scalars. All vectors are column vectors except for regression vectors, which are row vector throughout. L denotes for block length of data and M denotes for filter order. The superscript $(.)^T$ represents the transpose of a matrix or a vector. The notation $col\{...\}$ stands for a vector obtained by stacking the specified vectors. Similarly, we use $diag\{...\}$ to denote the (block) diagonal matrix consisting of the specified vectors or matrices. The trace of a matrix is denoted by $Tr(.)$, expectation is denoted by $E[.]$ and $\otimes$ denotes Kronecker product.

II. Problem Formulation

Consider N number of sensor nodes in a network. The objective of the network is to estimate the global parameter using the sensor node data. Each node k collects a scalar measurement $y_k(i)$ and regression data vector $x_{k,i}$. Both the input regressor and desired data in block format is given as

$$X_{k,i} = col\left\{x_{k,(i-1)L+1}, x_{k,(i-1)L+2}, \ldots\ldots, x_{k,iL}\right\} \quad (1)$$

$$y_{k,i} = col\left\{y_k\,(i-1)\,L+1, y_k\,(i-2)\,L+2, \ldots\ldots, y_k(iL)\right\} \quad (2)$$

with dimensions $L \times M$ and $L \times 1$ respectively over successive time instant $i \geq 0$ and for L denotes the block size. With the linear regression model the unknown global parameter w^0 of $M \times 1$ across all the nodes in the network is given by relation

$$y_{k,i} = X_{k,i} w^0 + v_{k,i} \quad (3)$$

where $v_{k,i}$ is the associated model noise. For the noise to be Gaussian noise with zero mean and variance $\sigma^2_{v,k}$, the parameter w^0 can be estimated by solving the minimization problem using least square method as linear function of the nodes in the network.

$$\min_{w} \sum_{k=1}^{N} E|y_{k,i} - X_{k,i}w|^2 \tag{4}$$

For the presence of outliers in the desired data, the noise can no further considered to be Gaussian for which traditional least square methods will fail to estimate the parameter. So, there is a need of a robust algorithm to overcome these outliers and to better estimate the global parameter. A Huber algorithm is a robust M-estimator that minimizes the Huber loss function.

$$J^{Huber} = \sum_{i=1}^{L} \rho\left({r_i}/{\sigma_i}\right) \tag{5}$$

where r_i is the residual error and σ_i is the corresponding scale factor and L is block length. $\rho(.)$ is a real-valued function that is even and nondecreasing for positive residuals, and $\rho(0) = 0$. For the Huber M-estimator the ρ function is given by

$$\rho(r) = \begin{cases} \frac{1}{2}r^2 & for\ |r| \le \gamma \\ \gamma|r| - \frac{1}{2}\gamma^2 & for\ |r| > \gamma \end{cases} \tag{6}$$

where γ is the threshold value. The ρ function is not strictly convex function. The derivative $\frac{\partial\rho}{\partial r}$ is an odd function defined as score function $\varphi(r)$ is given as

$$\varphi(r) = \begin{cases} r & for\ |r| \le \gamma \\ \gamma.sgn(r) & for\ |r| > \gamma \end{cases} \tag{7}$$

The weight function $\left[{\varphi(r)}/{r}\right]$ for Huber M-estimator is given by

$$q(r) = {\varphi(r)}/{r} = \begin{cases} 1 & for\ |r| \le \gamma \\ {\gamma.sgn(r)}/{r} & for\ |r| > \gamma \end{cases} \tag{8}$$

Incorporating the diffusion mode of cooperation into the nodes, we can write the local cost function at any node as

$$J_k^{loc} = \sum_{l\in\aleph_k} c_{l,k} J_l^{Huber} \tag{9}$$

where, $c_{l,k}$ is the element-wise entry of data diffusion matrix C. By completion-of-squares argument method [3], the local Huber cost function can be defined as

$$J_k^{loc} = \left\|w - w_k^{loc}\right\|_{\Gamma_k}^2 + residue \tag{10}$$

for $\Gamma_k = \sum_{l\in\aleph_k} c_{l,k}R_l$ and R_l is the autocorrelation matrix of node l. For ρ value of $\gamma|r| - \frac{1}{2}\gamma^2$, we take $|r| = sgn(r).r \approx r.\tanh(kr)$ for $k \gg 1$. Upon Taylor series expansion of hyperbolic tangent and excluding higher order terms this can be easily analysed similar to least square method to obtain the above equation (10). The global cost function defined as linear sum of individual local cost function of each node

$$J^G = \sum_{l=1}^{N} J_l^{loc} \tag{11}$$

So, now the global cost function using (9) and (10) can be expressed as

$$J^{G'} = \sum_{l\in\aleph_k} c_{l,k} J_l^{Huber} + \sum_{l\neq k}^{N} \left\|w - w_l^{loc}\right\|_{\Gamma_l}^2 \tag{12}$$

The presence of the term w_l^{loc} tells for every node should have access to optimal local weight estimate, which again restricts the idea of distributed network. So, w_l^{loc} is replaced with an intermediate weight estimate that is available at node l as ψ_l^{loc}. So, now the fully distributed cost function is defined as

$$J^{dist} = \sum_{l\in\aleph_k} c_{l,k} J_l^{Huber} + \sum_{l\neq k}^{N} b_{l,k}\|w - \psi_l\|^2 \tag{13}$$

for the covariance matrix taken as $\Gamma_l = b_{l,k}I_M$.

III. Steepest Descent approach

The global parameter can be estimated iteratively using gradient approach of steepest descent method as

$$w_{k,i} = w_{k,i-1} - \mu_k\left[\nabla_{w_{k,i-1}}\left(J^{dist}(w_{k,i-1})\right)\right] \tag{14}$$

with the gradient of Huber cost function is defined as $\frac{\partial J_k^{Huber}}{\partial w} = -X_{k,i}^T.\varphi_{k,i}(r)$. Replacing this gradient value to the equation (14) the overall diffusion method can be derived [3], which is a fused form of both CTA and ATC strategies, and is given as

$$\phi_{k,i-1} = \sum_{l\in\aleph_k} a_{1,lk} w_{l,i-1} \tag{15}$$

$$\psi_{k,i} = \phi_{k,i-1} + \mu_k' \sum_{l\in\aleph_k} c_{l,k}.X_{l,i}^T\varphi_{l,i}(r) \tag{16}$$

$$w_{k,i} = \sum_{l\in\aleph_k} a_{2,lk}\Psi_{l,i} \tag{17}$$

for μ_k' is small positive block step-size and $a_{1,lk}, c_{lk}, a_{2,lk}$ are non-negative entries of the $N \times N$ matrices with (A_1, C, A_2) respectively. The coefficients $a_{1,lk}, c_{lk}, a_{2,lk}$ are zero, wherever l is not connected the node k is $l \notin \aleph_k$, where $\aleph_k$ denotes the neighborhood of node k. For ATC implementation $A_1 = I$ taken and for CTA implementation $A_2 = I$ is taken and for noncooperation implementation all the coefficients are taken to be I.

The matrices (A_1, C, A_2) are either left or right stochastic i.e.

$$\mathrm{A}_1^T 1_N = 1_N, \mathrm{A}_2^T 1_N = 1_N, C1_N = 1_N \tag{18}$$

Following assumptions are assumed for mean stability analysis of the algorithm: (i) input is independent and identically distributed Gaussian distribution; (ii) any two sensor nodes spatial correlation of data is zero; (iii) the outliers present in the system are independent of the input data and also independent of the noise present in the system; (iv) the system noise is independent and identically distributed with Gaussian distribution; (v) the probability density function of noise plus outliers data is continuous;. The initial assumption states to consider the input regressor $X_{k,i}$ independent of $w_{l,j}$ for all l and for $j \le i-1$, reasonably valid for sufficiently small step-sizes.

Defining the error vectors by subtracting from the optimal global weight w^0 we have

$$\begin{aligned} \tilde{\phi}_{k,i-1} &= w^0 - \phi_{k,i-1} \\ \tilde{\Psi}_{k,i} &= w^0 - \Psi_{k,i} \\ \tilde{w}_{k,i} &= w^0 - w_{k,i} \end{aligned} \quad (19)$$

The adaptive equation (16) now with the error vector is now represented as

$$\tilde{\psi}_{k,i} = \tilde{\phi}_{k,i-1} - \mu'_k \sum_{l\in\aleph_k} c_{l,k}.X^T_{l,i}\varphi_{l,i}(r) \quad (20)$$

Globalising all the vectors and matrices we have

$$\begin{aligned} \tilde{\psi}_i &\triangleq col\left\{\tilde{\Psi}_1, \tilde{\psi}_2, ..., \tilde{\psi}_N\right\} \\ \tilde{w}_i &\triangleq col\left\{\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_N\right\} \\ \tilde{\phi}_i &\triangleq col\left\{\tilde{\phi}_1, \tilde{\phi}_2, ..., \tilde{\phi}_N\right\} \\ \varphi_i(r) &= col\left\{\varphi_{1,i}(r), \cdots, \varphi_{N,i}(r)\right\} \\ \mu' &\triangleq diag\left\{\mu'_1 I_M, \mu'_2 I_M, ..., \mu'_N I_M\right\} \\ X^T_i &= diag\left(X^T_{1,i}, \cdots, X^T_{N,i}\right) \end{aligned} \quad (21)$$

So, equation (20) is rewritten in global format as

$$\tilde{\psi}_i = \tilde{\phi}_{i-1} - \mu' C.X^T_i\varphi_i(r) \quad (22)$$

Multiplying A_2 both sides and equating to (17), and taking their expectation we get

$$E[\tilde{w}_i] = A_2 E\left[\tilde{\phi}_{i-1}\right] - \mu' A_2 C.E[H_i] \quad (23)$$

where, $H_i = X^T_i\varphi_i(r)$. Solving for $E[H_i]$ which is exactly an expectation over $\{v_i, X, \eta_g\}$ [10] written as $E_{\{X_i,v_i,\eta_g\}}[H_i]$ for η_g denoting Gaussian noise sequence with variance σ^2_g.

$$\begin{aligned} E_{\{X,v,\eta_g\}}[H_i] &= E_{\{X,v,\eta_g\}}\left[X^T_i\varphi_i(r)\right] = E_{\{v\}}[H_1] \\ &= E_{\{v\}}\left[E_{\{X,\eta_g\}}\left[X^T_i\varphi_i(r)\right]|v\right] \end{aligned} \quad (24)$$

which gives $H_i \approx A_H\left(\sigma^2_{e,k}\right) D_{\alpha,k} R^G_x \tilde{\phi}_{i-1}$. The parameter are defined[10] as $A_H\left(\sigma^2_{e,k}\right) = \int_{-\infty}^{\infty} \frac{\varphi(e)}{\sqrt{2\pi}\sigma_e}\exp\left(-\frac{e^2}{2\sigma^2_e}\right)de$, $\alpha_i = \int_0^{\infty}\exp(\beta\varepsilon_i)\left(g_i\left(\tilde{\beta}\right)\right)^{-3/2}d\beta$, where, $g_i\left(\tilde{\beta}\right) = \left(1 + 2\tilde{\beta}R_{x,ij}\right)$ and $\sigma^2_{e,k} = E\left[v^T_{i,k}R_{x,k}v_{i,k}\right] + \sigma^2_g$. The auto-correlation matrix is given as $R_{x,k} = E\left[X^T_{k,i}.X_{k,i}\right]$. Equation (23) is now rewritten as

$$E[\tilde{w}_i] = A_2 E\left[\tilde{\phi}_{i-1}\right] - \mu A_2 C.A_H\left(\sigma^2_{e,k}\right) D_{\alpha,k} R^G_x E\left[\tilde{\phi}_{i-1}\right] \quad (25)$$

with these new globalising matrix

$$\begin{aligned} D_{\alpha,k} &= diag\left\{\alpha_{1,k}, \cdots, \alpha_{m,k}\right\} \\ D^G_\alpha &= diag\left\{A_H\left(\sigma^2_{e,1}\right) D_{\alpha,1}, \cdots, A_H\left(\sigma^2_{e,N}\right) D_{\alpha,N}\right\} \\ R^G_x &= diag\left\{R_{x,1}, \cdots, R_{x,N}\right\} \end{aligned} \quad (26)$$

we rewrite (25) as

$$E[\tilde{w}_i] = A_2\left[I - \mu' C.D^G_\alpha R^G_x\right] A_1 E[\tilde{w}_{i-1}] \quad (27)$$

$$E[\tilde{w}_i] = A_2 S A_1 E[\tilde{w}_{i-1}] \quad (28)$$

IV. Condition for Mean Stability

For convergence to occur in mean the coefficient of the matrix S must be stable, i.e. $\rho(S) < 1$. The matrix A^T_1 and A^T_2 are right stochastic matrices, so the matrix S is stable whenever $\left[I - \mu' C.D^G_\alpha R^G_x\right]$ is stable. With this the upper bound of block-step-size μ' to guarantee the convergence of $E(\tilde{w}_i)$ to steady state value must satisfy the condition

$$\left|I - \mu' C D^G_\alpha R^G_x\right| < 1 \quad (29)$$

$$0 < \mu' < \frac{2}{C.D^G_\alpha R^G_x} \quad (30)$$

which can be further written for step-size as maximum eigen value of data at each node

$$0 < \mu_k < \frac{2}{L.\lambda_{\max}\left(\sum\limits_{l=1}^{N} c_{l,k}\left(R_{x,l} A_H\left(\sigma^2_{e,l}\right) D_{\alpha,l}\right)\right)} \quad (31)$$

Huber's M-estimator robust algorithm (15)-(17) is compared with the traditional block diffusion algorithm defined in [7], given as

$$\begin{aligned} \phi_{k,i-1} &= \sum_{l\in\aleph_k} a_{1,lk} w_{l,i-1} \\ \psi_{k,i} &= \phi_{k,i-1} + \mu'_k \sum_{l\in\aleph_k} c_{lk}.X^T_{l,i}\left(y_{l,i} - X_{l,i}.\phi_{k,i-1}\right) \\ w_{k,i} &= \sum_{l\in\aleph_k} a_{2,lk}\psi_{l,i} \end{aligned} \quad (32)$$

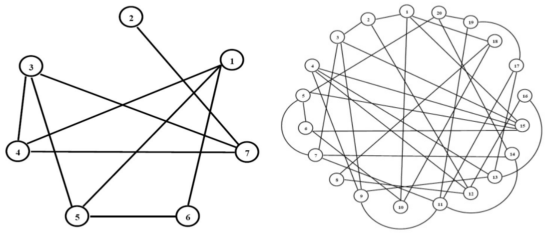
V. Simulation and Results

The size of unknown global vector w^0 is chosen to be $m = 4$ and the weights are considered to be normalized as equals to $[1/2, 1/2, 1/2, 1/2]$. The regressor $X_{k,i}$ is chosen to be Gaussian with zero mean and unit variance. A 30dB of measurement noise is added to each node. The block size L is chosen to be 20, number of samples taken is 4000 and the block step size μ' is selected to be 0.01. The threshold γ is taken to be 1 for implementing Robust Huber algorithm.

Combination of noise and outliers in the desired data can be modeled [13] as $(1-p)\frac{1}{\sqrt{2\pi\sigma^2_v}}\exp\left(-\left(v^2/2\sigma^2_v\right)\right) + p\frac{1}{\sqrt{2\pi\sigma'^2_v}}\exp\left(-\left(v^2/2\sigma'^2_v\right)\right)$ where, p is the probability of occurrence of the outliers in the desired data which can be obtained from the percentage of the outliers values,σ^2_v is the variance of the noise and σ'^2_v is the ratio of variance of the outliers to noise variance. For simulation outliers variance taken as 25. Implementing robust block diffusion algorithm all the coefficients A_1, C, A_2 are chosen as $A_1 = I$, $C =$metropolis and $A_2 =$relative degree. Similarly for noncooperation mode of diffusion adaptation all the coefficients are chosen as identity matrix (I).

Huber's robust block diffusion algorithm is simulated on two different types of network shown in Fig. 1 and their simulation plots for Mean Square Deviation (MSD) is plotted for 10% and 50% of outliers taking 50 independent experiments. Overall MSD is computed by taking average over all the experiments and all the nodes in the network.

Fig. 2 shows the overall MSD plots and the node-wise MSD plots for 10% and 50% of outliers for network structure



(a) (b)

Fig. 1: A network showing (a)7 sensor node (b)20 sensor node.

of 1a. The plots clearly shows the Huber's robust block diffusion algorithm outperforms the traditional diffusion method.

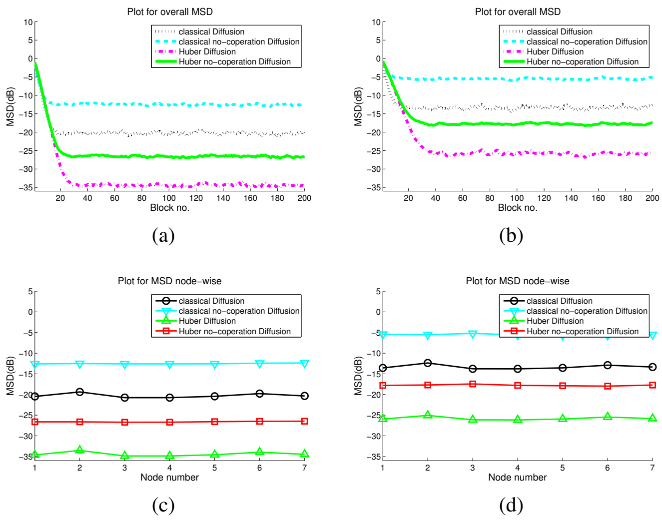


Fig. 2: Overall MSD plots for outliers of (a)10% (b)50% and Node-wise MSD plots for outliers of (c)10% (d)50% of the network Fig. 1a.

Fig. 3 shows the overall MSD plots and the node-wise MSD plots for 10% and 50% of outliers for network structure of 1b. Similar to the 7 node network this 20 node network demonstrates better steady state of the Huber's robust block diffusion algorithm compared to the traditional diffusion method.

VI. Conclusion

From the simulation results it is clear that the traditional method is completely outperformed by the robust Huber's diffusion algorithm in the presence of outliers. Condition for mean stability of Huber's block diffusion is derived. In both the network condition Huber's method shows a better steady state performance.

References

[1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01).*, vol. 4. IEEE, pp. 2033–2036, 2001.

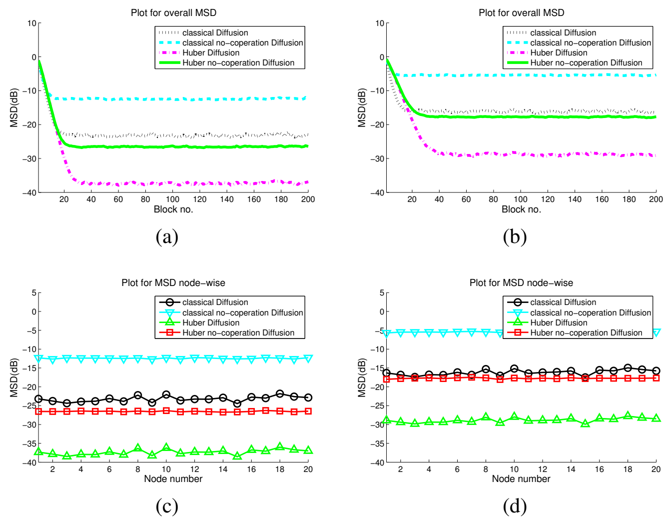


Fig. 3: Overall MSD plots for outliers of (a)10% (b)50% and Node-wise MSD plots for outliers of (c)10% (d)50% of the network Fig. 1b.

[2] 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.

[3] 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, 2010.

[4] C. G. Lopes and A. H. Sayed, "Diffusion Least-Mean Squares Over Adaptive Networks." in *ICASSP (3)*, 2007, pp. 917–920.

[5] C. G. Lopes, "Diffusion adaptive networks with changing topologies." IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008, 2008.

[6] K. Furutsu and T. Ishida, "On the theory of amplitude distribution of impulsive random noise and its application to the atmospheric noise," *Journal of the Radio Research Laboratories (Japan)*, vol. 7, no. 32, 1960.

[7] S. Kumar, A. K. Sahoo, and D. P. Acharya, "Block diffusion adaptation over distributed adaptive networks under imperfect data transmission," in *2014 Annual IEEE India Conference (INDICON)*, Dec 2014, pp. 1–5.

[8] A. Khalili, M. A. Tinati, and A. Rastegarnia, "An incremental block LMS algorithm for distributed adaptive estimation," in *2010 IEEE International Conference on Communication Systems (ICCS)*. IEEE, 2010, pp. 493–496.

[9] P. J. Huber, *Robust statistics*. Wiley, New York, 1981.

[10] S. C. Chan and Y. Zhou, "On the Performance Analysis of a Class of Transform-domain NLMS Algorithms with Gaussian Inputs and Mixture Gaussian Additive Noise Environment," *Journal Signal Processing Systems*, vol. 64, no. 3, pp. 429–445, Sep. 2011.

[11] H. Dai and N. K. Sinha, "A robust off-line output error method for system identification," *IEEE Transactions on Industrial Electronics*, vol. 39, no. 4, pp. 285–292, Aug 1992.

[12] P. Petrus, "Robust Huber adaptive filter," *IEEE Transactions on Signal Processing*, vol. 47, no. 4, pp. 1129–1133, Apr 1999.

[13] T. Hettmansperger and J. McKean, *Robust Nonparametric Statistical Methods*, ser. Kendall's Library of Statistics: An Arnold Publication No. 5. Arnold, 1998.

[14] V. J. Mathews and S. H. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 4, pp. 450–454, 1987.

[15] A. H. Sayed, *Fundamentals of adaptive filtering*. John Wiley & Sons, 2003.