Digital Signal Processing 51 (2016) 156–169

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Diffusion minimum-Wilcoxon-norm over distributed adaptive networks: Formulation and performance analysis

Sananda Kumar *, Upendra Kumar Sahoo, Ajit Kumar Sahoo, Debiprasad Priyabrata Acharya

Department of ECE, National Institute of Technology, Rourkela-769008, India

ARTICLE INFO

Article history:
Available online 4 February 2016

Keywords:
Diffusion strategy
Sensor network
Wilcoxon norm
Asymptotic linearity

ABSTRACT

This paper deals with the development of robust diffusion strategy for wireless sensor networks using minimum-Wilcoxon-norm. The Wilcoxon norm based robust estimation now-a-days has drawn the attention of the signal processing community for its scale equivariant property and simplicity. Exhaustive mathematical analysis has been presented to obtain the proposed diffusion minimum-Wilcoxon-norm (dMWN) algorithm. Steady state performance analysis of the algorithm has also been carried out to show the stability of the proposed method. Asymptotic linearity of rank test [1] is used for the convergence analysis of the proposed method. Extensive simulations have been given to demonstrate the efficacy of the proposed algorithm.

1. Introduction

Wireless sensor networks (WSN) find extensive application in modern world ranging from agriculture to tracking and smart house [2,3]. These applications rely on estimation of some parameters of interest from the spatio-temporal data measured by sensor node spread throughout the environment. Ad-hoc scenario and finite power constraints of WSNs have forced to design distributed algorithms based on data sharing among the neighboring sensor nodes for estimation of the parameter of interest. Different distributed strategies have been addressed in the existing literature. The major strategies are: (i) incremental [4–6], (ii) consensus [7,8] and (iii) diffusion type [9–12]. Among these the diffusion strategy is superior and robust compared to other two strategies in terms of performance, implementation and flexibility to node/link failure [12–14]. The diffusion strategy uses cost function optimization approach to estimate the parameters. All adaptive estimation techniques can be extended to adapt the diffusion strategy for distributed parametric estimation problem. For example the diffusion LMS and RLS algorithms have been reported to estimate the parameters using the distributed strategy. The diffusion techniques have also been developed for distributed space-time varying parameter estimation in [15]; non-stationary data and imperfect information exchange scenario [16]; distributed tracking the state of a dynamic system [17]; distributed detection [18]; distributed dictionary learning [19] and distributed pareto optimization [20]. Also this strategy has been used for modeling bird flight formation [21].

However, all these techniques are based on least squared error cost function which is sensitive to outliers present in the measured data. In addition to the additive white Gaussian noise (AWGN), impulsive random noise generated from atmosphere [22], co-channel interference, node failure, and presence of the nonlinearities and saturation effects in the practical sensor also affect the measured data. Hence, the distribution of the noise can no more be considered to be Gaussian. Estimation in such adverse situations can be done using any one of (i) model expansion; (ii) recursive estimation of the probability density function (PDF); and (iii) robust statistics based approach. Since in most of the signal processing algorithm the parameter is estimated recursively hence model expansion is not suitable. Because for expansion of the model order all the measured data should be available at the processor of the sensor node, which is not possible for a distributed ad-hoc network. Recursive estimation of PDF of the noise require more computation. Apart from this since the data are present distributively throughout the environment, the estimation requires more communication overhead. Under such a scenario the third approach seems to be more feasible for the WSN environment.

Robust statistics based algorithms are broadly classified into three types: M-, L- and R-type estimators. M-estimators like Huber's loss function depend on a predefined parameter which needs to be fine tuned in order to get good estimation accuracy. This parameter depends on the deviation of the noise distribution from the normal distribution. In general this method is not scale equiv-

* Corresponding author.
E-mail address: sanandanit@gmail.com (S. Kumar).

http://dx.doi.org/10.1016/j.dsp.2016.02.001

ariant. In order to make it scale equivariant another parameter such as median absolute deviation (MAD) needs to be estimated and to be incorporated in the estimation process. This requires more computations. M-estimators are divided into Monotone M-estimator and redescending M-estimators. In monotone M-estimator monotone function of the error is being used. This monotone function being a convex function, the optimum parameter can be achieved using gradient method. However, the main drawback is that in presence of heavy tailed noise or strong outliers its performance is poor [23,24]. Where as redescending M-estimator uses a non-convex function of the error and hence strongly reduces the effect of outliers. Due to the presence of the redescending function, the gradient based solution may fall into the local minima. By incorporating the MAD or other scale estimator to get the solution near to the global minima and then shifting to the redescending estimator superior performance is achieved. The L-estimators are themselves scale equivariant and do not depend on any predefined parameter. But their performance is inferior to that of the M- and R-estimators. R-estimator is called rank based estimator. In this type of estimator, the cost function is designed using the score value, which depends on the rank order and magnitude of the error value in one block [25]. This estimator is also scale equivariant and it's performance is superior than other robust estimators. This estimator has many other advantages too [26].

Due to the above stated reason, the Wilcoxon norm [1,27] has drawn the attention of the signal processing community to design efficient robust algorithms. The Wilcoxon norm has been used in [28–30] for robust learning machine and robust system identification. This norm has also been used in [31,32] for distributed robust parameter estimation. In this paper, in order to address the concerns of different traditional least square techniques discussed above in a distributed scenario with the limitation of incremental mode of cooperation, a robust diffusion minimum Wilcoxon norm is proposed. Besides this we also carry out the convergence analysis of the proposed algorithm which could have done in many of the reported work.

This paper not only deals with the development of the algorithm but also a strong performance comparison is presented. The Wilcoxon norm depends on the rank order and magnitude of the error value in a given block. Hence rank based norm analysis deviates from the least square error analysis by a large factor (Illustrated extensively in Section 2). Performance analysis of the proposed algorithm is done using the asymptotic linearity of rank test [1] and energy conservation principle [33]. Moreover, this type of performance approach differs from the statistical literature based performance measure. In statistical literature the main performance measures of robust techniques are: finite breakdown point and the bounded influence functional of the robust norm. Unlike statistical method our proposed technique deals with the iterative formulation of the algorithm and then the performance curve is evaluated similar to the energy conservation principle [33].

Exhaustive simulations have been performed for varieties of the environmental condition and it is observed that the proposed algorithm is robust against outliers in the desired data. It is also found that the theoretical predicted steady state matches strongly with the simulation steady state result.

This paper uses the following denotation. Bold capital and bold small letters are used to denote the matrix and vector respectively. Small and capital letters are used to represent scalar and functions respectively. The superscript $[.]^T$ represents the transpose of a matrix or a vector. The trace of a matrix is denoted by $Tr(.)$, expectation is denoted by $E(.)$, $\otimes$ denotes Kronecker product. The notation $col\{\dots\}$ stands for a vector obtained by stacking the specified vectors. Similarly, we use $\mathrm{diag}\{\dots\}$ to denote the block diagonal matrix consisting of the specified vectors or matrices.

The rest of the paper is organized as follows. Problem is formulated in Section 2. Section 3 deals with the development of the proposed distributed algorithm. Mean stability analysis of the algorithm is given in Section 4. Steady-state analysis of the proposed method is provided in Section 5. Simulation results and discussions are presented in Section 6. Finally the conclusion of the proposed work is given in Section 7.

2. Problem formulation

In WSN, sensors are spread throughout the environment so that the spatial events present in the environment can be captured. Suppose N number of sensor nodes are present in the environment from time T_0 to time T_1 and the input and the measured data are related by a linear model. Let $\mathbf{X}(n)$ and $\mathbf{y}(n)$ are the entire spatio-temporal input and desired data from time T_0 to T_1, given as

$$\begin{aligned} \mathbf{X}(n) &= \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_N \end{bmatrix} \qquad \Re^{p\times nN} \\ \mathbf{y}(n) &= \begin{bmatrix} \mathbf{y}_1^T & \mathbf{y}_2^T & \cdots & \mathbf{y}_N^T \end{bmatrix}^T \qquad \Re^{nN\times 1} \end{aligned} \tag{1}$$

where $\mathbf{X}_k$ and $\mathbf{y}_k$ are the local input and desired data at any node $k : 1 \le k \le N$ which is given by $\mathbf{X}_k = \begin{bmatrix} \mathbf{X}_{k,1} & \mathbf{X}_{k,2} & \cdots & \mathbf{X}_{k,n} \end{bmatrix}$ and $\mathbf{y}_k = \begin{bmatrix} y_{k,1} & y_{k,2} & \cdots & y_{k,n} \end{bmatrix}^T$ where, $\mathbf{X}_{k,q} \in \Re^p$ for $q = 1, 2, \dots, n$ and term p represents the filter order. The relation between input and desired data is given below

$$\mathbf{y}(n) = \mathbf{X}^T(n)\mathbf{w}^o + \mathbf{v}(n) \tag{2}$$

where, $\mathbf{v}(n)$ is the white noise present in the measured data. The objective is to estimate the parameter $\mathbf{w}^o$ from the desired data $\mathbf{y}(n)$ and the corresponding input data $\mathbf{X}(n)$. The objective can be formulated as an optimization problem as given in (3)

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \left\| \mathbf{y}(n) - \mathbf{X}^T(n)\mathbf{w} \right\|_* \tag{3}$$

When the noise $\mathbf{v}(n)$ is white and Gaussian the optimization problem in (3) can be changed to $\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \left\| \mathbf{y}(n) - \mathbf{X}^T(n)\mathbf{w} \right\|_2^2$. The least square error cost function is sensitive to the outliers present in the measured and input data. It has been already discussed, the measured data are often corrupted by the outliers. Hence, least square cost function does not converge to the optimum value. This triggers to the use of Wilcoxon norm cost function for estimation of the parameter.

2.1. The Wilcoxon norm

The Wilcoxon norm, a pseudo norm, [1,27] is usually defined by a score function $\varphi(u)$, $u \in [0, 1]$. The score function is a nondecreasing function, having two basic properties i.e. (i) $\int_0^1 \varphi(u)\, du = 0$, (ii) $\int_0^1 \varphi^2(u)\, du < \infty$.

The Wilcoxon norm of any vector $\mathbf{s} \in \Re^L$ is given as

$$\|\mathbf{s}\|_w = \sum_{i=1}^{L} \varphi\left(\mathrm{R}(s_i)/(L+1)\right) s_i \tag{4}$$

where $\mathrm{R}(s_i)$ is the rank order of element s_i in the vector $\mathbf{s}$ which means, if all the elements in the vector $\mathbf{s}$ are arranged in the increasing order, then the rank order of s_i is the position of s_i from the minimum one. The score function $\varphi(u)$ is given by $\varphi(u) = \sqrt{12}\,(u - 0.5)$.