Distributed Estimation of Spatially Varying Distributed Parameter System

A thesis submitted in partial fulfilment of requirement for the degree of

Master of Technology

in

Electronics and Communication Engineering

Signal and Image Processing

By

Hareesh Siriki

213EC6260

Under the guidance of

Prof. Upendra Kumar Sahoo

Department of Electronics and Communication Engineering

National Institute of Technology

Rourkela

May 2015

NATIONAL INSTITUTE of TECHNOLOGY

ROURKELA

CERTIFICATE

This is to certify that the work done in the thesis entitled **Distributed Estimation of Spatially Varying Distributed Parameter System** by **Hareesh Siriki** is a record of research work carried out by him under my supervision and guidance during 2014-2015 in partial fulfilment for the award of the degree in Master of Technology in Electronics and Communication Engineering (Signal and Image Processing), National Institute of Technology, Rourkela.

Date: 01/06/2015

Prof. Upendra Kumar Sahoo

Dept. of Electronics and Communication Engineering

National Institute of Technology

Rourkela

NATIONAL INSTITUTE of TECHNOLOGY

ROURKELA

DECLARATION

I certify that:

- The work presented in this thesis is an original content of the research done by myself under the general supervision of my supervisor.
- The project work or any part of it has not been submitted to any other institute for any degree or diploma.
- I have followed the guidelines prescribed by the Institute in writing my thesis.
- I have given due credit to the sources (written material) used by quoting them where I used them and have cited those sources. Also their details are mentioned in the references.

Hareesh Siriki

ACKNOWLEDGEMENT

This research work is one of the significant works that I have done in my life and is made possible because of the unending encouragement and motivation given by so many in every part of my life. It is an immense pleasure to have this opportunity to express my gratitude and regards to them.

First and foremost, I would like to express my profoundest gratitude and sincere thanks to **Prof. Upendra Kumar Sahoo**, in department of Electronics and Communication Engineering for his esteemed supervision, support and guidance during the tenure of my project work. He is an amicable personality under whom it is my privilege to work. I would also like to thank him for providing best facilities in the department.

I would like to express my gratitude and respect to Prof. K.K.Mahaptra, Prof. S.K.Meher, Prof. A.K. Sahoo, Prof. L.P.Roy, Prof. Samit Ari, Prof. S. Maiti, Prof. A.K. Swain, Prof. D.P.Acharya for their guidance and suggestions throughout the M.Tech course. I take this opportunity to express gratitude to all of the Department faculty members for their help and support.

I would like to express my sincere thanks to the Ph.D. scholar Mr. Sanand Kumar and my friends for their help throughout my project work which inspired me a lot. I would like to express my heartfelt wishes to my friends and classmates whose company and support made me feel much better than what I am.

Hareesh Siriki

ABSTRACT

Adaptive filter shows a significant role in the field of digital signal processing and wireless communication. It integrates LMS algorithm in real time situations because of its low computational complexity and simplicity. The adaptive distributed strategy is built on the diffusion cooperation scheme among nodes at different locations that are dispersed over a wide topographical area. Computations have been performed in all the nodes and all the results are shared among them so as to obtain precise parameters of interest.

There are some scenarios where estimation parameters vary over both space and time domains across the network. A set of basis functions i.e. Chebyshev polynomials is used to describe the space-varying nature of the parameters and diffusion least mean-squares strategy is proposed to recover these parameters. The parameters of our concern are assessed for both one dimensional and two dimensional networks. Stability and convergence of the proposed algorithm have been analysed and expressions are derived to predict the behaviour. Network stochastic matrices are used to combine exchanged information between nodes. The results show that the performances of the networks also depend upon the combination matrices. The resulting algorithm is distributed, co-operative and able to respond to the real time changes in environment.

Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction

Wireless network of sensors consists of vast collection of nodes i.e. sensors which are self-powered are distributed over some wide geographical area. The availability of low power sensors generates need for in-network processing. Data communication in future will be done by distributed network of sensors, mobile phones, connection of PCs. These sensors, in a wireless fashion, communicate with other nodes in the sensor network, by collecting information from their counterparts. There are numerous applications in remote sensing in which communication with other nodes plays an important role. So, sensing information, processing it and communicating it with others are the main features in distributed networks. Some applications include battlefield surveillance, medical applications, environmental monitoring, precision agriculture, target localization[1]. It is due to these types of applications that make efficient design and implementation of wireless network of nodes a field of research currently. In all the above cases, the robustness of the currently performing tasks and the improvement of the probability of signal detection can be done in distributed networks by providing spatial diversity which works along temporal dimension. During the estimation of required parameters of interest, communication among nodes and computations consume more power[10]. As the nodes i.e. sensors work with less battery power, it is of utmost necessity that we need to design networks with less communication.

Initially, to estimate a certain parameter of interest, all the nodes in a network collect noisy observations related to it. In centralized processing technique, the parameters of interest are estimated by extracting data i.e. local estimates of all nodes in the network by the central processor and getting fused. So this technique requires a powerful processor and immense amounts of communications between the central processor and nodes in a network. The liberty of the network is restricted and due to the presence of central node[5], a critical point of failure is also added in the network. The main drawback of the centralized technique is it restricts the ability of the nodes to adapt in real time to time varying data. This in turn degrades the tracking performance. Being non-robust to fusion center failure is another drawback of this procedure.

Another way to increase tracking ability is by estimating required parameters of our interest from local observations and by cooperating with neighbors. Distributed processing copes with getting information from data merged at nodes which are spread over a wide geographic region. Every node in the network operates as an individual adaptive filter whose goal is to estimate required parameters of our regard by using local observations. These individual estimates of each node get united with its neighbors to get an estimate from the data at the nodes nearby. Case in point, each node in a network of nodes could gather noisy findings identified with required parameters of interest. The neighbors would then connect with one another in a certain way, as requested by the system topology, to take on at a rough estimate of the parameter. The main objective is to get an assessment that is just as precise as the case that would be acquired if every node had obtained the data over the entire web of nodes[10]. In wireless and sensor networks, distributed estimation algorithms are utilized where

robustness, scalability and low power consumption are preferable. So, these algorithms can have excess levels of communication exchange of information among the nodes.

1.2 Motivation

The availability of less power sensors is the motivation factor behind employing wireless distributed sensor networks. In wired networks, the sensor nodes are connected to a central processing unit through wires where all the processing is performed. Here, we focus on wireless distributed sensor networks in sensing and processing is distributed.

In centralized processing solution, the parameters of interest are estimated by extracting data i.e. local estimates of all nodes in the network by the central processor and getting fused. So this technique requires a powerful processor and immense amounts of communications between the central processor and nodes in a network. The liberty of the network is restricted and due to the presence of central node, a critical point of failure is also added in the network[10]. The main drawback of the centralized technique is it restricts the ability of the nodes to adapt in real time to time varying data. This in turn degrades the tracking performance.

Distributed processing copes with acquiring information from data amalgamated at particular nodes which are dispersed over a broad topographic area. In distributed implementation, all nodes in the network convey information to all the neighbors in the subset and processing is distributed among all the sensor nodes in the network. The information from neighbor sensor nodes is sent to certain nodes which are fused with that from certain nodes.

The result from the particular nodes depends not only on the local data but also on the data from the neighbors of the nodes. The potential of distributed implementations depends on the cooperation strategies which are permitted among sensor nodes.

In incremental cooperation policy[8], data is allowed from one sensor node to successive node. This demands least amounts of interaction among the nodes and requires low power. The collaboration among the nodes is in the form of cyclic pattern. On the other hand, diffusion implementation involves communication of each node with all of its neighbors as demanded by network topology. Compared to incremental solution, this technique requires more amounts of interaction between nodes in a network. This in turn increases consumption of power more than that of incremental solution. These drawbacks can be compensated by one thing i.e. the nodes in the network get access to more information from the neighboring nodes.

At the same time, in diffusion cooperation scheme[9], the communications can be reduced by allowing each sensor node to interact with a subset of its neighbors only. The selection of neighbors can be done according to performance criterion. For the issues like less communications and computational complexity, operation in real-time, need for adaptive implementations, distributed least mean squares algorithms have been proposed for both computations and communications and for the robustness of implementations in LMS. Diffusion cooperation policies provide robustness to node and connection failure, good performance and are amiable to distributed implementations. In some diffusion algorithms used for adaptation over networks, the required parameters that have to be estimated are considered as space-invariant.

1.3 Objective

The main objective of this is to achieve a precise estimate of a required parameter over a wide geographic region under consideration. For this, the sensors dispersed over a topographical region collect data from the information attained from the neighbors and reach a pinpoint estimate using cooperative policies. The sensors do not have any perception of their arrangement or who their neighbors are which a humorous thing is and yet their aim is to converge to an ideal value. In cooperative policies, there is an enormous collaboration among the sensors. The sensor nodes collaborate among themselves in such a pattern that output of nodes not only rely on the data obtained from themselves but also from their neighbors in the proximity of the considering nodes. To equip nodes with adaptive abilities, they have to share their information with others, should have the potential to adapt in real time. Utilizing distributed strategies estimates are updated in real time through local interactions. Estimation and tracking of parameters of our concern over spatial domain are to be performed using cooperation strategies. Achievement of an estimate that is just as precise as the node which had access to information that is derived from data extracted from all the other sensor nodes and estimation of unfamiliar parameters over space where no readings are gathered i.e. interpolation mechanism are the main objectives.

1.4 Thesis Layout

The chapter 2 elucidates about the basics of distributed network, design aspects to be considered for the cooperation strategies i.e. incremental and diffusion schemes and their

formulation. The Chapter 3 describes about the estimation of space-time varying parameters for some phenomena using adaptive centralized and distributed optimization and also Runge's phenomenon. The chapter 4 provides an idea of two-dimensional network with 4-connected and 8-connected neighbors and the deduction of space- varying regression model. The chapter 5 shows the simulation results and discussion on them. The chapter 6 concludes the work and summarizes that has been observed from this.

Chapter 2

Design Aspects for Cooperation Strategies

Wireless sensor networks comprises of set of nodes which are scattered over a wide topographical region. These are self-powered and undergo local computing operation. The nodes in a particular network communicate with all of its neighbors or with a centralized processor in a wireless approach. The technique which copes with acquiring information from the data gathered from the nodes in the whole network is known as distributed processing. The sensor nodes interact with their neighbors according to the network arrangement. The potential of performance of distributed networks depend on cooperation strategies. Here, we consider incremental and diffusion cooperation strategies.

2.1 Distributed network

The main goal is assess an essential parameter of concern from the perceived data. Every node in the network acts as an adaptive filter whose goal is to assess the parameter from the local observations. The assessed readings in different nodes are amalgamated with those of other nodes in order to procure a better assessment. There are different procedures used for merging of estimates from different nodes. Such methods extract power and communication resources of a considerable amount. They might also restrict the tracking potential of the network to acknowledge in real time to the statistical changes in the data. If the distinctive filters greet mainly the local temporal data, their ability to utilize and acknowledge to spatial information across the sensor nodes is reduced.

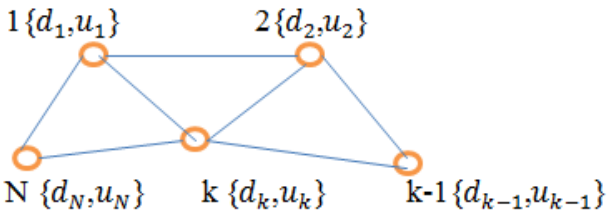

Fig. 1 Distributed Network

Consider a wireless network of N sensor nodes as shown in fig.1 which observe temporal data from various spatial sources with different data profiles. At each and every time instant, every node l receives scalar quantity $d_l(t)$ of a random process. An unknown vector is to be assessed based on the readings gathered at N nodes dispersed over a network. The main aim is to approximate some parameter vector for all nodes in the network. The nodes are considered as connected if they can interact with other nodes in the network. A node is connected to itself on all occasions. The group of nodes which are connected to a particular node l is degree of the node and indicated by n_l. The total number of nodes linked to a node l, including itself is called as neighborhood of node l and is indicated by $\mathcal{N}_l$.

2.2 Incremental Cooperation Strategy

In this method, the information extracted from the data acquired from the nodes is flown one node to its successive node. This involves collaboration among the nodes in the form of cyclic pattern. As the data is passed from one sensor node to its adjacent node, it demands low power and communication resources. The following figure denotes incremental cooperation strategy.

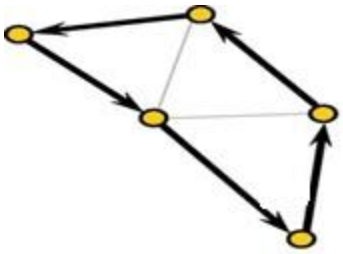

Fig. 2 Incremental Cooperation Scheme

The coloured dots denote nodes dispersed in a network. The data acquired from previous node is passed to the current node as in fig.2. This node performs necessary computations and gives the updated data to its successive node. Each node is denoted by l and it has access to data $\{d_l(t), u_{l,t}\}$ at every iteration t. This data is taken from information from different sources $\{\boldsymbol{d}_l, \boldsymbol{u}_l\}$, where $l = 1,2,\cdots,N$, $\boldsymbol{d}_l$ is a scalar measurement and $\boldsymbol{u}_l$ is Mx1 row regression vector. Regression data and observations are collected in a vector as follows:

$$\boldsymbol{U} \triangleq col\{u_1, u_2, \dots, u_N\} \tag{1.}$$

$$\boldsymbol{d} \triangleq col\{d_1, d_2, \dots, d_N\}\ (N \times 1) \tag{2.}$$

These quantities gather information from all the nodes. The main aim is to assess the Mx1 vector w that solves

$$\min_{w} \quad J(w) \tag{3.}$$

Where $J(w)$ represents the cost function and is identified as MSE which is as follows:

$$J(w) = \mathrm{E}\|\boldsymbol{d} - \boldsymbol{U}w\|^2 \tag{4.}$$

The optimal solution w is determined by utilizing orthogonality condition as follows:

$$E\|\boldsymbol{d} - \boldsymbol{U}w\|^2 = 0 \tag{5.}$$

w^0 is the solution to the normal equations

$$R_{du} = R_u w^0 \quad (6.)$$

Where R_u=E$U^* U$ $(M \times M)$, R_{du} =E$U^* d$=$\sum_{l=1}^{N} R_{du,l}$

R_u is the auto correlation and R_{du} is the cross correlation quantities.

So, the above equations demand access to complete information in the whole network i.e. $\{R_u, R_{du}\}$. Another approach is centralized processing and transfer the information to all the nodes in the network. The drawback of these methods is requirement of more number of computations, excessive amounts of interaction and more power by all the nodes. Another disadvantage is they do not have the potential to track the parameters with changes in the statistical data $\{d_l(t), u_{l,t}\}$ at each and every iteration t.

2.2.1 Steepest Descent Technique

The steepest descent and incremental methods are reviewed before we go for adaptive solution for incremental cooperation technique. The cost function has to be disintegrated into a number of error functions of all the nodes in the network as follows:

$$J(w) = \sum_{l=1}^{N} J_l(w) \quad (7.)$$

Where $J_l(w)$ is the individual error function of each node in the network.

$$J_l(w) \triangleq E|d_l - u_l w|^2 \quad (8.)$$

$$J_l(w) = {\sigma_{d,l}}^2 - R_{u,l} w - w^* R_{du,l} + w^* R_{u,l} w \quad (9.)$$

The different parts in the above equation are as follows:

$$\sigma_{d,l}{}^2 = E|d_l|^2,\ R_{u,l} = Eu_l{}^* u_l, \text{ and } R_{du,l} = Ed_l u_l{}^*$$

The steepest descent form of equation for finding the solution of w^0 is as follows:

$$w_t = w_{t-1} - \mu[\nabla J(w_{t-1})]^*, w_{-1} = initial\ condition \tag{10.}$$

$$w_t = w_{t-1} - \mu \sum_{l=1}^{N} [\nabla J_l(w_{t-1})]^* \tag{11.}$$

$$w_t = w_{t-1} + \mu \sum_{l=1}^{N} \left(R_{du,l} - R_{u,l} w_{t-1}\right) \tag{12.}$$

Where w_t is an assessment of w^0 at time instant t, w_{t-1} is an assessment of w^0 at time instant $t-1$, $\nabla J(w_{t-1})$ denotes gradient of cost function $J(w)$ found at iteration $t-1$ with respect to w, μ is step size which is positive.

For a network arrangement, a cycle is defined such that every sensor node is linked to its immediate neighbour i.e. its adjacent node in the network. The following is the steepest descent implementation in which the local estimate $\psi_l{}^{(t)}$ is of node l.

$$\psi_0{}^{(t)} = w_{t-1} \tag{13.}$$

$$\psi_l{}^{(t)} = \psi_{l-1}{}^{(t)} - \mu_l[\nabla J_l(w_{t-1})]^*,\ \ l = 1,2,\cdots,N \tag{14.}$$

$$w_t = \psi_N{}^{(t)} \tag{15.}$$

$\psi_l{}^{(t)}$ is an estimate at node l at time instant t. $\psi_{l-1}{}^{(t)}$ is an estimate of previous node $l-1$ at iteration t. $\psi_N{}^{(t)}$ is a local assessment at node N. At each iteration t, the global estimate w_{t-1}

for w^0 is considered initially with ${\psi_0}^{(t)} = w_{t-1}$ and iterations are run which after all of them are completed, the local assessment at the last node N coincides with w_t i.e. $w_t = {\psi_N}^{(t)}$.

2.2.2 Incremental Steepest Descent Technique

Each and every node k receives information from its previous node i.e its immediate neighbour but in this procedure each node acquires global information w_{t-1} to analyze $\nabla J_l(w_{t-1})$. Though the above iterative equations involve cooperation among the nodes in the network, they do not satisfy the condition of full distributed processing technique.

To resolve this trouble, we consider incremental gradient methods. In this, each node assesses the gradient $\nabla J_l\left({\psi_{l-1}}^{(t)}\right)$ with respect to ${\psi_{l-1}}^{(t)}$ in contrast with w_{t-1}. Finally, this technique develops into an incremental version. The following are the iterative equations.

$${\psi_0}^{(t)} = w_{t-1} \tag{16.}$$

$${\psi_l}^{(t)} = {\psi_{l-1}}^{(t)} - \mu_l\left[\nabla J_l\left({\psi_{l-1}}^{(t)}\right)\right]^*, \quad l = 1,2,\cdots,N \tag{17.}$$

$$w_t = {\psi_N}^{(t)} \tag{18.}$$

The above equations show that they form a correct distributed method as the cooperation strategy relies on local information available. Obviously, each node interacts with its adjacent neighbor which in turn conserves communication resources and power.

2.2.3 Incremental Adaptive Technique

The operation of incremental technique is shown in fig. 3. To evaluate the gradient $\nabla J_l\left({\psi_{l-1}}^{(t)}\right)$, the incremental technique depends on the perception of second order moments i.e. auto correlation and cross correlation matrices $R_{du,l}$ and $R_{u,l}$. The above equations can be

turned into adaptive form by superseding correlation matrices by their instantaneous estimations which are as follows:

$$R_{du,l} \approx d_l(t)u_{l,t}^{*} \tag{19.}$$

$$R_{u,l} \approx u_{l,t}^{*}u_{l,t} \tag{20.}$$

The above estimations are obtained with the help of statistical data $\{d_l(t), u_{l,t}\}$ at every iteration t. Utilizing these estimations, the steepest descent incremental algorithm is transformed into distributed adaptive incremental algorithm or just incremental least mean squares technique.

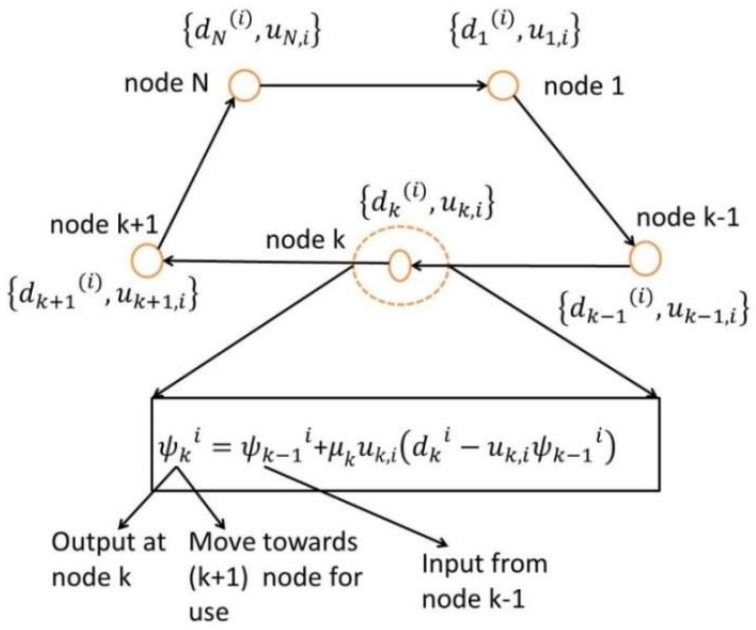

Fig. 3 Operation of Incremental mode

The equations are obtained in the following method:

$$\psi_0{}^{(t)} = w_{t-1} \tag{21.}$$

$$\left[\psi_l{}^{(t)} = \psi_{l-1}{}^{(t)} - \mu_l \nabla J_l\left(\psi_{l-1}{}^{(t)}\right)\right]^*, \;\; l = 1,2,\cdots,N \tag{22.}$$

$$w_t = \psi_N{}^{(t)} \tag{23.}$$

Substituting the equation in the following one gives

$$\left[\psi_l{}^{(t)} = \psi_{l-1}{}^{(t)} - \mu_l \nabla J_l\left(\psi_{l-1}{}^{(t)}\right)\right]^* \tag{24.}$$

$$\left[\psi_l{}^{(t)} = \psi_{l-1}{}^{(t)} - \mu_l \sum_{l=1}^{N}\left(R_{du,l} - R_{u,l} w_{t-1}\right)\right]^* \tag{25.}$$

$$\left[\psi_l{}^{(t)} = \psi_{l-1}{}^{(t)} + \mu_l \sum_{l=1}^{N}\left(d_l(t) u_{l,t}{}^* - u_{l,t}{}^* \psi_{l-1}{}^{(t)}\right)\right]^* \tag{26.}$$

Using the instantaneous estimations, the following equations are obtained.

$$\psi_0{}^{(t)} = w_{t-1} \tag{27.}$$

$$\psi_l{}^{(t)} = \psi_{l-1}{}^{(t)} + \mu_l u_{l,t}{}^*\left(d_l(t) - u_{l,t}\psi_{l-1}{}^{(t)}\right), \;\; l = 1,2,\cdots,N \tag{28.}$$

$$w_t = \psi_N{}^{(i)} \tag{29.}$$

The nodes in the above equations use statistical data $\{d_l(t), u_{l,t}\}$ at each and every iteration t and perform the following operations to estimate the weights.

1) Calculates the error quantity i.e. $e_l(t) = d_l(t) - u_{l,t}\psi_{l-1}{}^{(t)}$
2) Updates the estimate by the equation $\psi_l{}^{(t)} = \psi_{l-1}{}^{t} + \mu_l e_l(t)$
3) Transfers the updated assessment to its neighbouring node l+1

By this we can say that performance of incremental technique is much better than that of steepest descent technique because it utilizes spatial diversity efficiently. The steepest descent technique employs w_{t-1} but incremental technique employs updates $\{\psi_0{}^{(t)}, \psi_1{}^{(t)}, \ldots\ldots, \psi_{N-1}{}^{(t)}\}$.

2.3 Diffusion Cooperation Strategy

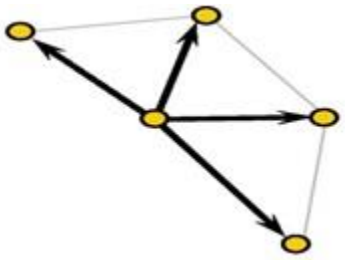

Fig. 4 Diffusion Cooperation Scheme

In this strategy, each node is made to interact with all the nodes in its neighborhood at every iteration as shown in fig.4. In the computations, each node exchanges data with other nodes in its neighborhood and estimates the necessary parameters of our concern. So the data required to estimate the weights depends not only on the data from the node itself but also on the data from the nodes nearby the particular node. At each node, the traded data gets amalgamated and instantly given to a local adaptive filter.

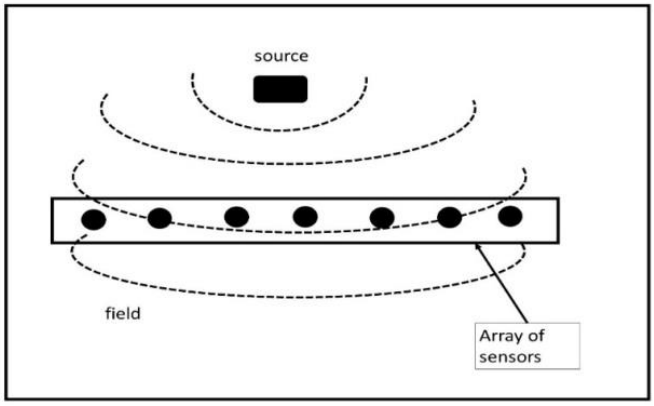

Fig. 5 Diffusion in a phenomenon

We can say that the passing of information among the nodes within the neighborhood of a particular node is in the form of diffusion as in fig.5. The diffused information makes the assessment as a function of not only temporal data but also spatial data of the nodes. This process improves the potential of adaptive filters i.e. nodes to adapt in real time with respect to the changes in the data statistics $\{d_l(t), u_{l,t}\}$. The spatial data alters for all nodes. The analysis of all the nodes is more burdensome than studying about the performance of single adaptive filters because the nodes influence each other.

2.3.1 Diffusion LMS formulation

The main goal is to estimate the necessary parameters of our concern as precise as it can and dispatches the assessment of the vector to all the nodes in the network. Initially, a cooperation strategy has to be chosen if an adaptive estimation solution is to be derived. A sensor node k obtains a group of assessments at any time instant $t-1$. The unbiased estimates are $\{\psi_l{}^{(t-1)}\}_{k\in\mathcal{N}_k}$. The quantity $\mathcal{N}_l$ is the symbol for neighborhood of l which comprises of

group of all nodes connecting to it and including itself. The estimates are erroneous versions of w^0 as follows:

$$\psi_l{}^{(t-1)} = w^0 - \psi_l{}^{(t-1)} \tag{30.}$$

The local estimates of all nodes are amalgamated at node k which gives the following:

$$\phi_l{}^{(t-1)} = f_l(\psi_k{}^{(t-1)}; k \in \mathcal{N}_{l,t-1}) \tag{31.}$$

Where f_l is local fusing function.

Let f_l be considered as some weighted combination like

$$\phi_l{}^{(t-1)} = \sum_{k \in \mathcal{N}_{l,t-1}} c_{lk}\, \psi_k{}^{(t-1)} \tag{32.}$$

Where c_{lk} are coefficients used for combination which is to be calculated i.e. $c_{lk} \geq 0$.

$$\phi_l{}^{(t-1)} = \sum_{k \in \mathcal{N}_{l,t-1}} \frac{c_{lk}}{\sum_{r \in \mathcal{N}_{l,t-1}} c_{lr}} \psi_k \tag{33.}$$

Finally, these coefficients give rise to some coefficient matrices C = $[c_{lk}]$ that bears information about network arrangement. For example, an item c_{lk} which is non-zero denotes that nodes l and k are linked to each other. If the coefficients are defined again as

$$c_{lk} \leftarrow \frac{c_{lk}}{\sum_{r \in \mathcal{N}_{l,t-1}} c_{lr}} \tag{34.}$$

then we get the following as the sum of all elements in row is equal to 1.

$$\sum_k c_{lk} = 1, k \in \mathcal{N}_{l,t-1} \tag{35.}$$

So, this says that C is a stochastic matrix. The combining step helps in amalgamating information from all the nodes in the network into node l. In general, each node has different neighborhood $\mathcal{N}_l$ for node l.

In order to promote adaptivity, the aggregated assessment ${\phi_l}^{(t-1)}$ extracted from all the nodes in the neighborhood of l is merged with the information of the node l such that it can acknowledge the alterations in the surroundings and update it to ${\psi_l}^{(t)}$. So, this solution turns into an adaptive one which attains small erroneous levels in steady state when contrasted with non-cooperative solution. In non-cooperative solution, each node adapts without complete knowledge of data from the nodes themselves and that of aggregation.

The diffusion scheme put forward is of the form as follows:

$$ {\phi_l}^{(t-1)} = f_l({\psi_k}^{(t-1)}; k \in \mathcal{N}_{l,t-1}) \tag{36.} $$

$$ {\psi_l}^{(t)} = {\phi_l}^{(t-1)} + \mu_l {u_{l,t}}^* \left(d_l(t) - u_{l,t} {\phi_l}^{(t-1)}\right), \ \ l = 1,2,\cdots,N \tag{37.} $$

The above network is robust in nature, immune to connection failures and utilizes network connectivity. A linear combiner is used as an aggregating function. The scheme is as follows:

$$ {\phi_l}^{(t-1)} = \sum_{k \in \mathcal{N}_{l,t-1}} c_{lk} \, {\psi_k}^{(t-1)}, \quad {\phi_l}^{(-1)} = 0 \tag{38.} $$

$$ {\psi_l}^{(t)} = {\phi_l}^{(t-1)} + \mu_l {u_{l,i}}^* \left(d_l(t) - u_{l,t} {\phi_l}^{(t-1)}\right) \tag{39.} $$

These equations are used for estimation purposes in the form of linear models.

$$ \boldsymbol{d}_l(t) = \boldsymbol{u}_{l,t} w^o + \boldsymbol{n}_l(t) \tag{40.} $$

Likely choices for combination matrix C[1] are metropolis, nearest neighbor and laplacian rules.

The following example of combination matrix is metropolis rule.

$$c_{lk} = \begin{cases} \dfrac{1}{\max(n_l, n_k)}, & \text{if l and k are connected} \\ 0, & \text{if l and k not connected} \\ 1 - \sum\limits_{\substack{k \in \mathcal{N}_l \\ l}} c_{lk}, & if\ l = k \end{cases}$$

The following rule is of uniform combination.

$$c_{lk} = 1/(n_l), \quad \text{if } l \neq k \text{ are connected}$$

$$c_{lk} = 0, \quad \text{if } l \text{ and } k \text{ not connected}$$

Where, A_d is NxN adjacent matrix which can formed as

$$A_d = \begin{cases} 1, & if\ l\ and\ k\ are\ connected \\ 0, & otherwise \end{cases}$$

If a node is connected to itself then $[A_d]_{ll} = 1$.

Chapter 3

Space-Time Varying Parameters

In many theories on diffusion cooperation schemes, the parameters to be estimated are considered as space invariant. The complete group of nodes in a network detects and evaluates information which is obtained from a physical model represented by non-variable parameters over space. Nevertheless, there are different scenarios in which necessary parameters to be estimated are space varying in the network. The above situation is present in scenarios like tracing of distribution of population, observation of fluid flow in underground porous media, confinement of distributed sources in dynamic systems. The parameters altering in space are obtained by discretizing the coefficients of partial differential equation through spatial sampling.

In the assessment of the parameters of our concern, the few techniques that have been applied to prior studies can be applied here too. The results depend on the utilization of central processing unit and distributed network and in-network processing results is less considered. Distributed algorithms are employed when the network in which nodes are present have no influential amalgamation centre and also when the amount of communication resources and power are not sufficient enough to supply to the complete network. A few number of distributed algorithms have been proposed in the papers including incremental method, consensus methods and diffusion methods. Incremental solutions demand cyclic path between the sensor nodes over the network and therefore vulnerable to connection failures. Consensus solutions necessitate doubly-stochastic combining policies and causes instability in the network

where the scenarios require relentless adaptation and tracking. Diffusion policies have the ability to track the parameters of our concern in a continual pattern and adapt with them in real time. Also, the networks provide stable behavior despite the arrangement of the nodes in the network.

In this, the parameters of our concern are assessed and traced by developing the distributed algorithm with diffusion least mean squares policy and they alter over space[5] and time simultaneously. Initially, this outlook begins with designing the space varying regression model which describes the distinctive nature of the space varying situations. Partial differential equations are the crucial mathematical tools to describe the spatio-temporal processes. This model is derived by discretizing the partial differential equations in describing the dynamic systems with spatially altering features. Here, a set of basis functions have been introduced to indicate the space varying parameters in terms of a set of space-invariant coefficients. Based on this depiction, a diffusion LMS method has been proposed to estimates, tracks and interpolates the parameters of our concern over the network.

3.1 Modeling

In this part, the linear regression model has been introduced which is utilized to explain the systems with space altering features. Here, the model considered is evolution of pressure distribution in an inhomogeneous media[6]. To describe this, a second order one dimensional partial differential equation is used. This has space varying diffusion coefficient and input distribution. This method can extended to partial differential equations of over two-dimensional space in shaping the space varying phenomena.

The PDE [4] that is considered is

$$\frac{\partial f(x,i)}{\partial i} = \frac{\partial}{\partial x}\left(\tau(x)\frac{\partial f(x,i)}{\partial x}\right) + p(x,i) \tag{41.}$$

where,

$(x,i)\ \epsilon\ [0,L]$ x [0,T] indicate variables across space and time,

$f(x,i)$ indicates the system distribution (e.g., pressure or temperature),

$\tau(x)$ is the space-varying coefficient,

$p(x,i)$ is the input distribution

The partial differential equation[4],[12] considered is presumed to fulfil the boundary conditions $f(0,i)$= $f(L,i)$ = 0 for all i ϵ [0,T]. Hence, the distribution at the system at i=0 is given by $f(x,0) = y(x)$ for $x \in [0,L]$. The Finite Difference Method[13] is employed for discretizing the partial differential equation over space and time domains. For N and P which are positive integers, $\Delta x = L/(N+1)$ and $x_l = l\Delta x$ for $l \in \{0,1,2,\ldots.,N+1\}$ and in the similar way $\Delta i = T/P$ and $i_t = t\Delta i$ for $i \in \{0,1,2,\ldots.,P\}$. The sampled values of the pressure distribution $f_l(t) \triangleq f(x_l 0, i_t)$, input $p_l(t) \triangleq p(x_l, i_t)$ and space altering coefficient $\tau(x) \triangleq \tau(x_l)$. Hence, the recursive equation can be obtained from Finite Difference Method[13] as follows:

$$f_l(t) = v_{l,t} g_l{}^o + \Delta i\, p_l(t-1),\ l\,\epsilon\,\{1,\ 2,\ldots..,\ N\} \tag{42.}$$

where the vectors ${g_l}^o$ and $v_{l,t}$ are defined as,

$$ {g_l}^o = [{g_{1,l}}^o, {g_{2,l}}^o, {g_{3,l}}^o]\mathrm{T} \tag{43.} $$

$$ v_{l,i} = [f_{l-1}(t-1), f_l(t-1), f_{l+1}(t-1)] \tag{44.} $$

The entries of ${g_{m,l}}^o$ are:

$$ {g_{1,l}}^o = \frac{v}{4}(\tau_{l-1} + 4\tau_l - \tau_{l+1}) \tag{45.} $$

$$ {g_{2,l}}^o = 1 - 2\tau_l \tag{46.} $$

$$ {g_{3,l}}^o = \frac{v}{4}(-\tau_{l-1} + 4\tau_l + \tau_{l+1}) \tag{47.} $$

where,

$$ a = \Delta i/\Delta x2 $$

$$ l \epsilon \{1,\ 2, \ldots\ldots,\ N\} $$

The above relation (3) is well-defined $l \epsilon \{1,\ 2, \ldots\ldots,\ N\}$. This means that data sampling is not necessary at $x = \{0, L\}$ beacause $f_0(t)$ and $f_{N+1}(t)$ have a close similarity to the boundary conditions $f(0, i)$ and conditions $f(L, i)$. In the observing process i.e. estimation of $\tau(x)$, the nodes in the network collect erroneous information parts of $f(x, i)$ across the network. The error information is represented as follows:

$$ z_l(t) = f_l(t) + n_l(t) \tag{48.} $$

where, $n_l(t)$ is a randomized noise term.

The above equation can be written as,

$$ d_l(t) = u_{l,t}{g_l}^o + n_l(t) \tag{49.} $$

where,

$$d_l(t) = z_l(t) - \Delta i\, p_l(t-1) \tag{50.}$$

The space reliant model can be global to explain systems with more than one spatial dimension. The assumption that $u_{l,t}$ is random because of the probability of erroneous samples is denoted by boldface notation $\boldsymbol{u}_{l,t}$. Therefore, ${g_l}^o$ and $u_{l,t}$ are both vectors of M dimensions. The noise is generally represented by $v_l(t)$ for various forms of errors which include noise in readings. So, the space altering regression model[5] is of the form:

$$\boldsymbol{d}_l(t) = \boldsymbol{u}_{l,t}{g_l}^o + \boldsymbol{v}_l(t) \tag{51.}$$

where, $\boldsymbol{d}_l(t) \in \mathbb{R}$, $\boldsymbol{u}_{l,t} \in \mathbb{R}^{1\times M}$, ${g_l}^o \in \mathbb{R}^{M\times 1}$, $\boldsymbol{v}_l(t) \in \mathbb{R}$. So, the networks which monitor phenomena described by regression models of the form (12) are studied where the main aim is the space altering parameters of concern ${g_l}^o$ are to be assessed $l \in \{1, 2, \ldots\ldots, N\}$. So, a distributed solution of diffusion cooperation strategy in the form of adaptive algorithm is required to estimate the parameters over both space and time. The accessible data $\{\boldsymbol{d}_l(t), \boldsymbol{u}_{l,t}\}$ that is obtained from the information extracted from all the N nodes at each location x_l is to estimate the parameters ${g_l}^o$.

3.2 Optimization Techniques

3.2.1 Centralized Optimization

In centralized optimization, space altering parameters are found by diminishing the following global cost function over the variables g_l:

$$J(g_1, \ldots\ldots, g_N) \triangleq \sum_{l=1}^{N} J_l(g_l) \tag{52.}$$

where,

$$J_l(g_l) \triangleq \mathbb{E}\left|\boldsymbol{d}_l(t) - \boldsymbol{u}_{l,t} g_l\right|^2 \tag{53.}$$

In distributed mechanisms, to find g_l, some steps are essential to transform global cost (13) into a form for decentralized optimization. From (6)-(8) the collaboration is favorable because the g_l of neighbors are associated with each other through space reliant function $\tau(x)$. Individually, if the nodes assess the space altering parameters by reducing $J_l(g_l)$, then at each time instant in order to find the model parameters over locations of space where readings are not taken, they have to send the estimates to an amalgamation center for interpolation. Cooperate strategies are utilized to help the nodes upgrade the assessments and accomplish more precise interpolation.

3.2.2 Adaptive Distributed Optimization

The nodes in the network accomplish their common objective through collaboration in all the distributed enhancement over the systems. So, the objective can be attained by finding the global parameter vector and using it in diminishing the cost function that comprises of the complete set of nodes. The anonymous parameters in (13) are node-reliant. In continuous pace domain, the items of every g_l i.e. $\{g_{1,l}{}^o, \ldots\ldots, g_{M,l}{}^o\}$ are deduced into samples of M anonymous space-altering parameter functions $\{g_1{}^o(\mathrm{x}), \ldots\ldots, g_M{}^o(\mathrm{x})\}$ at position $x = x_l$ as shown in figure 1.

To assess the space altering coefficients by the utilization of distributed optimization, the theory of interpolation is used to find a group of linear expansion coefficients mutual to all the nodes. The m-th unfamiliar space altering parameter function is conveyed as a distinctive linear arrangement of N_b basis functions i.e.

$$g_m{}^o(x) = W_{m,1}b_1(x) + W_{m,2}b_2(x) + \ldots\ldots + W_{m,N_b}b_{N_b}(x) \quad (54.)$$

Where, $W_{m,n}$ are the sole expansion coefficients and $b_n(x)$ are the basis functions. Here the basis functions assumed are Chebyshev polynomials[14],[15]. The following are the expressions for basis functions:

$$b_1(x) = 1 \quad (55.)$$

$$b_2(x) = 2x - 1 \quad (56.)$$

$$b_{n+1}(x) = 2(2x-1)b_n(x) - b_{n-1}(x), \qquad 2 < n < N_b \quad (57.)$$

Based on the specification of applications and guidance of many concerns like less computational complexity, ability to interpolate at points where readings are not collected, efficiency in representation and properties like orthogonality, the selection of basis functions is done. The main advantage in using Chebyshev polynomials is it avoids Runge's phenomenon and also produces worthy results in the above conditions.

3.3 Runge's Phenomenon

Runge's phenomenon results in deviating approximation. It is an issue of oscillations at the edges of an interim that happens when utilizing polynomial interpolation with polynomials of high degree more than an arrangement of equispaced interpolation points. Consider the situation where one yearns to insert through n+1 equispaced points of a function $f(x)$ utilizing the n-degree polynomial $p_n(x)$ that goes through those locations. Normally, one may anticipate that utilizing more points would lead toward a more exact remaking of $f(x)$. Notwithstanding, this specific arrangement of polynomial functions $p_n(x)$ is not ensured to have the property of convergence in uniform fashion. The $p_n(x)$ created in this way might

actually separate far from $f(x)$ as n builds. This regularly happens in a oscillating example that amplifies close to the finishes of the interpolation points. So, error increases as n value increases. Polynomials with high degree are inappropriate to interpolate between equidistant points. Finally, the interpolation error surges with increase in degree of polynomials.

3.4 Deduction of Regression Model

The sampled form of m-th space altering parameter ${g_m}^o(x)$ in (15) at $x = x_l = l\Delta x$ is written as

$$ {g_{m,l}}^o = {W_m}^T b_l \tag{58.} $$

Where,

$$ W_m = [W_{m,1}, W_{m,2}, \ldots., W_{m,N_b}]\mathrm{T} \tag{59.} $$

$$ b_l = [b_{1,l}, b_{2,l}, \ldots., b_{N_b,l}]\mathrm{T} \tag{60.} $$

Sampling the basis functions at locations $x = x_l = l\Delta x$, produces all the elements $b_{n,l}$ i.e.

$$ b_{n,l} = b_n(x_l) = b_n(l \Delta x) \tag{61.} $$

Assembling all the samples of the parameter functions ${g_m}^o(x)$ $m \epsilon \{1, 2, \ldots., M\}$ into a column vector as

$$ {g_l}^o = [{g_{1,l}}^o, {g_{2,l}}^o, \ldots., {g_{M,l}}^o]^T \tag{62.} $$

Utilizing the equation (18), the following is attained:

$$ {g_l}^o = W^o b_l \tag{63.} $$

Where,

$$W^o \triangleq \begin{bmatrix} {W_{1,1}}^o & {W_{1,2}}^o & \cdots & {W_{1,N_b}}^o \\ {W_{2,1}}^o & {W_{2,2}}^o & & {W_{2,N_b}}^o \\ \vdots & & \ddots & \vdots \\ {W_{M,1}}^o & {W_{M,2}}^o & \cdots & {W_{M,N_b}}^o \end{bmatrix}$$

Rearrangement can be done by piling up all the columns of ${W^o}^T$ i.e. w^o= vec(${W^o}^T$). This transforms W^o into an $MN_b \times 1$ column vector. Using this, a block diagonal matrix can be defined as follows:

$$B_l \triangleq I_M \otimes {b_l}^T \tag{64.}$$

So, this makes the size of B_l as $M \times MN_b$. The equation (23) can be rewritten as:

$${g_l}^o = B_l w^o \tag{65.}$$

Substituting the above equation in (12), we obtain the following space varying regression model:

$$\boldsymbol{d}_l(t) = \boldsymbol{u}_{l,t} B_l w^o + \boldsymbol{v}_l(t) \tag{66.}$$

So, the global cost function accordingly, becomes:

$$J(w) = \sum_{l=1}^{N} \mathbb{E}\left|\boldsymbol{d}_l(t) - \boldsymbol{u}_{l,t} B_l w^o\right|^2 \tag{67.}$$

3.5 Estimation using Optimization Techniques

3.5.1 Centralized Optimization

The following statistical conditions have been assumed for the data $\{\boldsymbol{d}_l(t), \boldsymbol{u}_{l,t}, \boldsymbol{v}_l(t)\}$ in the network:

1) $\boldsymbol{d}_l(t)$ and $\boldsymbol{u}_{l,t}$ are zero mean jointly wide sense stationary random processes with second order moments:

$$r_{du,l} = \mathbb{E}\left[\boldsymbol{d}_l(t){\boldsymbol{u}_{l,t}}^T\right] \in \mathbb{R}^{M\times 1} \tag{68.}$$

$$R_{u,l} = \mathbb{E}\left[{\boldsymbol{u}_{l,t}}^T\boldsymbol{u}_{l,t}\right] \in \mathbb{R}^{M\times M} \tag{69.}$$

2) The regression information $\{\boldsymbol{u}_{l,t}\}$ are i.i.d over time, independent over space and covariance matrices $R_{u,l}$ are positive definite $\forall l$.
3) Noise considered here,$\boldsymbol{v}_l(t)$, is i.i.d. over time and independent over space with variance $\{{\sigma_{v,l}}^2\}$.
4) The noise $\boldsymbol{v}_l(t)$ is independent of the regression data $\boldsymbol{u}_{m,j}$ $\forall i,j,l,m$.

The ideal parameter w^o for reducing the global cost function is to make the gradient vector of $J(w)$ equal to zero. Using this, the following equations have been obtained:

$$\left(\sum_{l=1}^{N} R1_{u,l}\right)w^o = \sum_{l=1}^{N} r1_{du,l} \tag{70.}$$

Where, $\{R1_{u,l}, r1_{du,l}\}$ are the second order moments of $\boldsymbol{u}_{l,t}B_l$ and $\boldsymbol{d}_l(t)$:

$$R1_{u,l} \triangleq {B_l}^T R_{u,l}B_l, \quad r1_{du,l} \triangleq {B_l}^T r_{du,l} \tag{71.}$$

After the solution for global parameter is obtained, the space altering parameters can be assessed at locations other than the node locations x_l. The result w^o can be obtained in iteration utilizing steepest descent recursion:

$$\boldsymbol{w}_t = \boldsymbol{w}_{t-1} + \mu\sum_{l=1}^{N}(r1_{du,l} - R1_{u,l}\boldsymbol{w}_{t-1}) \tag{72.}$$

μ is considered as step size which is greater than zero and $\boldsymbol{w}_t$ is the assessment of w^o at t-th iteration. The (33) shows that the powerful central processor necessitates to have the knowledge of the covariance matrices, $R1_{u,l}$ and cross covariance matrices $r1_{du,l}$ at all nodes. Knowing the matrices in advance is not possible and the instantaneous approximations are used. Using these approximations, the recursion equations can be led to centralized LMS scheme for assessment of space altering parameters of all the nodes in the network.

$$\boldsymbol{w}_t = \boldsymbol{w}_{t-1} + \mu \sum_{l=1}^{N} [B_l{}^T \boldsymbol{u}_{l,t}{}^T (\boldsymbol{d}_l(t) - \boldsymbol{u}_{l,t} B_l \boldsymbol{w}_{t-1})] \tag{73.}$$

$$\boldsymbol{g}_{l,t} = B_l \boldsymbol{w}_t \tag{74.}$$

Each sensor node, in every iteration, drives the information $\{\boldsymbol{d}_l(t), \boldsymbol{u}_{l,t}\}$ to the central processing unit for fusion to update the weight in the previous iteration, $\boldsymbol{w}_{t-1}$. The latter equation shows the estimation of space altering parameters using the updated weights $\boldsymbol{w}_t$ at iteration t and basis function matrix at position l i.e. B_l. The second equation also denotes as the interpolation mechanism to approximate the space-altering parameters by utilizing the basis function matrix at each node, $B(x)$.

3.5.2 Diffusion Optimization

The optimization scheme is to decouple the global cost and write it as a group of constrained optimization problems with local weights $\boldsymbol{w}_l$ as:

$$\min_{w_l} \sum_{k \in \mathcal{N}_l} c_{k,l} \mathbb{E} \left| \boldsymbol{d}_k(t) - \boldsymbol{u}_{k,t} B_k w_l \right|^2 \tag{75.}$$

Where, $c_{k,l}$ are the elements of right stochastic matrix $C \in \mathbb{R}^{N\times N}$ which satisfies:

$$c_{k,l} = 0 \text{ if } k \notin \mathcal{N}_l \text{ and } C\mathbb{I} = \mathbb{I} \tag{76.}$$

$\mathbb{I}$ is a column matrix with unit entries.

The limitations enforced all the nodes to agree on an precise result but this lowers the tracking and learning abilities of all the nodes in the network. It is always beneficial for the nodes that sense appropriate data to respond right away to it without delaying by waiting for all of them to get same result. This makes the information to be diffused among the nodes in the network.

The scheme that does not undergo the tracking and learning abilities of all the nodes in real time is diffusion strategy which reduces the global cost solve the optimization problem for $l \epsilon \{1,\ 2, \ldots\ldots,\ N\}$:

$$\min_{w}\ (\sum_{k\in\mathcal{N}_l} c_{k,l}\mathbb{E}\left|\boldsymbol{d}_k(t) - \boldsymbol{u}_{k,t}B_k w\right|^2 + \sum_{k\in\mathcal{N}_l\backslash\{l\}} p_{k,l}\|w-\psi_k\|^2) \tag{77.}$$

Where ψ_k is the available assessment of the global parameter at node k. $\mathcal{N}_l\backslash\{l\}$ is the set of neighborhood of l without including node l and $p_{k,l}$ are some parameters. The minimization of (38) leads to the following recursion equations:

$$\boldsymbol{\phi}_{l,t-1} = \sum_{k\in\mathcal{N}_l} {a_{k,l}}^{(1)}\boldsymbol{w}_{k,t-1} \tag{78.}$$

$$\boldsymbol{\psi}_{l,t} = \boldsymbol{\phi}_{l,t-1} + \mu_l \textstyle\sum_{k\epsilon\mathcal{N}_l}[c_{k,l}{B_k}^T{\boldsymbol{u}_{k,t}}^T(\boldsymbol{d}_k(t) - \boldsymbol{u}_{k,t}B_k\boldsymbol{\phi}_{l,t-1})] \tag{79.}$$

$$\boldsymbol{w}_{l,t} = \sum_{k\epsilon\mathcal{N}_l} {a_{k,l}}^{(2)}\boldsymbol{\psi}_{k,t} \tag{80.}$$

$$\boldsymbol{h}_{l,t} = B_l \boldsymbol{w}_{l,t} \tag{81.}$$

In the algorithm, μ_l denotes the step size at node l which is greater than zero. $\boldsymbol{w}_{l,t-1}$, $\boldsymbol{\psi}_{l,t}$, $\boldsymbol{\phi}_{l,t-1}$ are the intermediary estimates of w^o, $g_{l,t}$ is an intermediary estimate of ${g_l}^o$ and $\{{a_{l,k}}^{(1)}, {a_{l,k}}^{(2)}\}$ are non-negative elements of left stochastic matrices $A1, A2 \in \mathbb{R}^{N\times N}$ which satisfies:

$$ {a_{k,l}}^{(1)} = {a_{k,l}}^{(2)} = 0 \quad \text{if } k \notin \mathcal{N}_l \tag{82.}$$

$$A1^T \mathbb{I} = \mathbb{I}, \quad A2^T \mathbb{I} = \mathbb{I}$$

The following steps describe the operations performed by the above equations.

1) The first step denotes the combination step which fuses $\{w_{k,t-1}\}_{k\in \mathcal{N}_l}$ to create $\boldsymbol{\phi}_{l,t-1}$.
2) The second step denotes the adaptation in which each node l uses the information from the node itself and that of from the neighboring nodes i.e. $\{\boldsymbol{d}_l(t), \boldsymbol{u}_{l,t}\}$ to update $\boldsymbol{\phi}_{l,t-1}$ to an intermediary assessment $\boldsymbol{\psi}_{l,t}$.
3) The third step also denotes the combination step in which intermediary assessments $\{\boldsymbol{\psi}_{l,t}\}_{k\in \mathcal{N}_l}$ are combined to further align the global estimate at locations x_l including neighbors.
4) Finally, the anticipated parameters of our concern are assessed by utilizing the weights $\boldsymbol{w}_{l,t}$.

Using the $N_b M$ global invariant coefficients in w^o, the proposed algorithm of diffusion LMS assesses NM spatially relying variables $\{{h_k}^o\}$. The combination matrices $A1, A2, C$ depend upon different choices. If $A1 = A2 = C = I$, then the diffusion algorithm gets transformed into non-cooperative scheme. Making $C = I$, the adaptation equation (40) changes to a scheme

where each node l utilizes its own information to execute local adaptation. Using the choice, $A1 = I, A2 = A$, where A is some left stochastic matrix makes it discards the first fusion step and it reduces to adaptation followed by combination which makes it Adapt-then-Combine (ATC) diffusion algorithm.

$$\boldsymbol{\psi}_{l,t} = \boldsymbol{w}_{l,t-1} + \mu_l {B_l}^T {\boldsymbol{u}_{l,t}}^T (\boldsymbol{d}_l(t) - \boldsymbol{u}_{l,t} B_l \boldsymbol{w}_{l,t-1}) \tag{83.}$$

$$\boldsymbol{w}_{l,t} = \sum_{k \epsilon \mathcal{N}_l} a_{k,l} \boldsymbol{\psi}_{k,t} \tag{84.}$$

$$\boldsymbol{h}_{l,t} = B_l \boldsymbol{w}_{l,t} \tag{85.}$$

The choice $A1 = A, A2 = I$ discards the second combination equation and algorithm minimizes to combination followed by adaptation which makes it Combine-then-Adapt (CTA) diffusion algorithm.

$$\boldsymbol{\phi}_{l,t-1} = \sum_{k \epsilon \mathcal{N}_l} a_{k,l} \boldsymbol{w}_{k,t-1} \tag{86.}$$

$$\boldsymbol{w}_{l,t} = \boldsymbol{\phi}_{l,t-1} + \mu_l {B_l}^T {\boldsymbol{u}_{l,t}}^T (\boldsymbol{d}_l(t) - \boldsymbol{u}_{l,t} B_l \boldsymbol{\phi}_{l,t-1}) \tag{87.}$$

$$\boldsymbol{h}_{l,t} = B_l \boldsymbol{w}_{l,t} \tag{88.}$$

Both the versions of diffusion algorithms i.e. ATC and CTA have been used for $C = I$. The optimum values of the combination matrices differ from each other which depends upon the arrangement of network and communication between the nodes. The combination rules[1] used for stochastic matrices $\{A1, A2, C\}$ are as follows:

Name	**Combination Rule** ($k \in \mathcal{N}_l, k \neq l$)
Uniform	$a_{k,l} = 1/n_l$
Laplacian	$a_{k,l} = 1/n_{max}$
Maximum Degree	$a_{k,l} = 1/N$
Metropolis	$a_{k,l} = \frac{1}{\max(n_l, n_k)}$
Relative Degree	$a_{k,l} = \frac{n_k}{\sum_{m \in \mathcal{N}_l} n_m}$

Chapter 4

Two-Dimensional Process Estimation

4.1 Climate Models

The practical application of the two dimensional network is climate model. The climate models are designed to foresee the weather. Weather prediction and forecasting is so difficult such that it comprises many calculations. It's in the same way, an architect plans a model for a building, a scientist makes a climate model to predict the behavior of climate. Physics and chemistry of the biosphere, atmosphere are integrated into the climate models to predict the weather at some location or to respond to the question like predicting the next El Nino to occur and to monitor the concentration of greenhouse gases in the atmosphere.

The major task for the climate models is to go into the future and see the weather conditions i.e. quicker than that of the variations in oceans and atmosphere. To perform this, the climate models have to undergo the large number of calculations and assumptions that are used for simplifications. To explore various features of climate, different kinds of models are required. A single aspect of climate can be predicted using one, two and three dimensional models. So, all the three different dimensional models can be applied to the processes in all the realms like oceans, land surface and atmosphere.

Now and again, where difficult procedures change as indicated by a wide mixture of components, it can be best to start to investigate the procedures in just one dimension. For instance, when taking a gander at chemical responses that fluctuate with the physical

conditions through the depth of the air, one methodology is to take a gander at the reactions at every level from the beginning to the highest point of the atmosphere. One-dimensional models were at first utilized for energy related investigations of the atmosphere in the climate models. As more noteworthy certainty has gotten in the way a basic model handles a specific process, the thoughts are then consolidated into more mind boggling two-, three- and four (time)-dimensional representations which consolidate the motion of the atmosphere framework.

The climate can be represented mathematically by the climate models. The models split the oceans, atmosphere, earth into a grid. Here, the values of variables to be anticipated are wind, temperature, humidity, rainfall, pressure, etc. and they are calculated at every grid time at all times to estimate the future values. The climate models have to cope with the features of atmosphere for analysis of the ocean and surface temperatures. At the boundaries, the land surface and oceans are coupled with the atmosphere through interchange of heat, moisture, etc. These models are called as general circulation models (GCMs). Coupling the sea procedures to climatic GCMs is a noteworthy test. The warm limit of the seas is enormous contrasted with the air and can give to, or remove from, the climate, gigantic measures of dormant and warm warmth. Speaking to their warmth stockpiling, and the ingestion of greenhouse gases by the seas, in long haul simulations of atmosphere obliges an entire three-dimensional sea model, which re-enacts even the profound streams. Changes in the force and area of deep-water streams can at last have significant consequences for the air. Previously, changes in the flow of the seas have created major atmospheric reactions.

4.2 Modeling

A two dimensional network with $N \times N$ nodes which are equally spaced over unit square $(x, y) \in [0,1] \times [0,1]$ with spacing Δx and Δy is considered as shown in fig.(2). In these networks, each node can communicate with a set of 4 neighbours and 8 neighbours in fig. 6. As the number of neighbors of a node increases, the data accessed by the node increases. This, in turn, reduces the mean square error.

Fig. 6 Sets of neighbours

The network observers a phenomena $f(x, y)$ defined by Poisson partial differential equation:

$$\frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2} = g(x,y) \tag{89.}$$

Where $g(x, y): [0,1]^2 \rightarrow \mathbb{R}$ is an unfamiliar input function. The partial differential equation is presumed to have satisfied the following boundary conditions:

$$f(x,0) = f(0,y) = f(x,1) = f(1,y) = 0$$

The main aim is to assess the parameters from the nodes which receive noisy readings. The nodes considered here are the internal node locations. In order to discretize the partial differential equation, the finite difference method (FDM)[13] is employed with positioning of Δx and Δy uniformly. The locations of nodes can be determined by $x_{l_1} \triangleq l_1 \Delta x,\ y_{l_2} \triangleq l_2 \Delta y$ and the sample values of the function can be identified by $f_{l_1,l_2} \triangleq f(x_{l_1}, y_{l_2})$ and ${g^o}_{l_1,l_2} \triangleq g(x_{l_1}, y_{l_2})$. The central difference strategy is utilized to approximate the second partial derivatives:

$$\frac{\partial^2 f(x,y,i)}{\partial x^2} \approx \frac{1}{\Delta x^2}[f_{l_1+1,l_2} - 2f_{l_1,l_2} + f_{l_1-1,l_2}] \tag{90.}$$

$$\frac{\partial^2 f(x,y,i)}{\partial y^2} \approx \frac{1}{\Delta y^2}[f_{l_1,l_2+1} - 2f_{l_1,l_2} + f_{l_1,l_2-1}] \tag{91.}$$

The above equations give rise to the following equation:

$${g^o}_{l_1,l_2} = \frac{1}{\Delta x^2}(f_{l_1+1,l_2} + f_{l_1,l_2+1} + f_{l_1-1,l_2} + f_{l_1,l_2-1} - 4f_{l_1,l_2}) \tag{92.}$$

To assess the space altering parameters g_{l_1,l_2}, the information received by every node are erroneous samples as follows:

$$z_{l_1,l_2}(t) = f_{l_1,l_2} + \boldsymbol{n}_{l_1,l_2}(t) \tag{93.}$$

The error samples considered here are zero mean, white and not spatially dependent. Here, the reference signal which is denoted by $\boldsymbol{d}_{l_1,l_2}(t)$ in two dimensional networks is distorted signal of ${g^o}_{l_1,l_2}$. Using the erroneous samples, the reference signal is attained by (124) as follows:

$$\boldsymbol{d}_{l_1,l_2}(t) = \frac{1}{\Delta x^2}(z_{l_1+1,l_2}(t) + z_{l_1,l_2+1}(t) + z_{l_1-1,l_2}(t) + z_{l_1,l_2-1}(t) - 4z_{l_1,l_2}(t)) \tag{94.}$$

Using the equation (126), the two dimensional space altering regression model can be defined as follows:

$$\boldsymbol{d}_{l_1,l_2}(t) = \boldsymbol{u}_{l_1,l_2}(t) g^{o}{}_{l_1,l_2} + \boldsymbol{v}_{l_1,l_2}(t) \tag{95.}$$

If $\boldsymbol{u}_{l_1,l_2}(t) = 1$, the above model turns into deterministic regression model in contrast with random regression model. So, we can say that diffusion scheme can be applied even to the regression models with deterministic information.

In the two dimensional network, the space altering parameters $g^{o}{}_{l_1,l_2}$ can be found by using two one-dimensional Chebyshev polynomials[14],[15] as basis functions. So, $g^{o}{}_{l_1,l_2}$ can be conveyed as:

$$g^{o}{}_{l_1,l_2} = \sum_{n=1}^{N_b} w^{o} p_{n,l_1,l_2} \tag{96.}$$

$$p_{n,l_1,l_2} = b_{n_1,l_1} b_{n_2,l_2} \tag{97.}$$

Where, p_{n,l_1,l_2} is an element of the two dimensional basis group, b_{n_1,l_1} and b_{n_2,l_2} are the one dimensional Chebyshev polynomials used as basis functions in the x and y directions respectively.

Chapter 5

Simulation Results and Discussion

5.1 Simulation results

The performance of the diffusion scheme has been shown from the computer simulations performed. For generating the learning curves, 200 experiments have been performed and averaged. The steady-state performances of the curves are shown for some iterations depending on the analysis of the curves. The measurement data $d_l(t)$ is attained using the data model at each node and the anticipated $M \times 1$ vector to be estimated is set as $w^o = col\{1,1,\ldots.,1\}/\sqrt{M}$ where M is the size of the weights. Here M is considered as 2. For one dimensional case, a network of 7 nodes has been taken into account. In two-dimensional case, the network 7×7 nodes i.e. have been taken. The noise signal applied is white Gaussian noise with zero mean and variance, $\sigma^2 = 0.0001$. The input signal is presumed to be corrupt with white Gaussian noise zero mean. The step size considered here is $\mu = 0.01$ and distance between the nodes is $\Delta x = 0.02$. The quantities MSE, MSD have been found as follows:

$$MSE = \left|d_l(t) - u_{l,t}{\psi_{l-1}}^{(t)}\right|^2 \tag{98.}$$

$$MSD = \left|{\psi_l}^{(t)} - w^o\right|^2 \tag{99.}$$

The MSD values of the network have been obtained as shown in the following figures.

One-dimensional network:

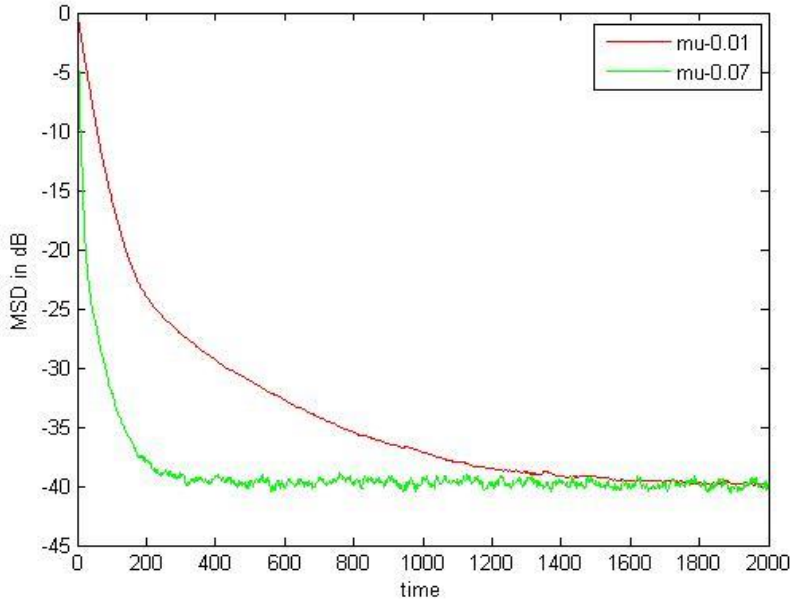

Fig. 7 MSD performance for step size values

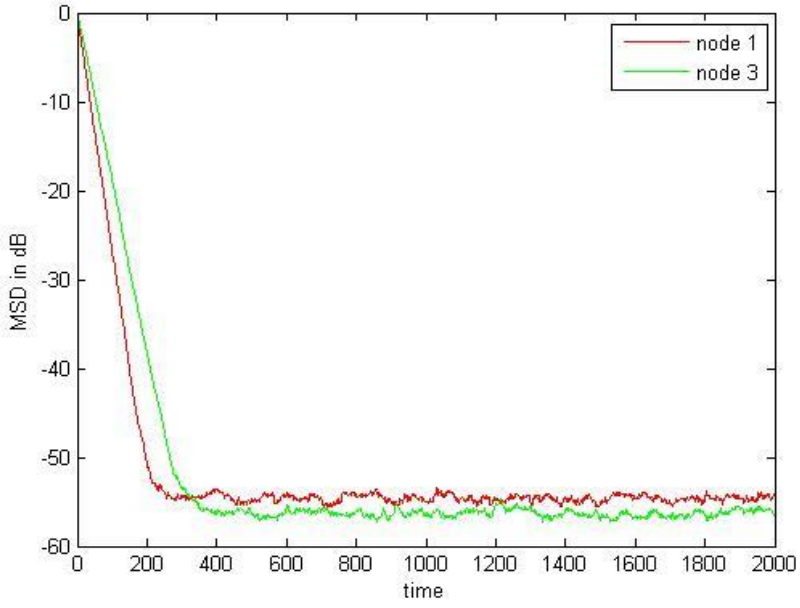

Fig. 8 MSD performance comparison of two nodes

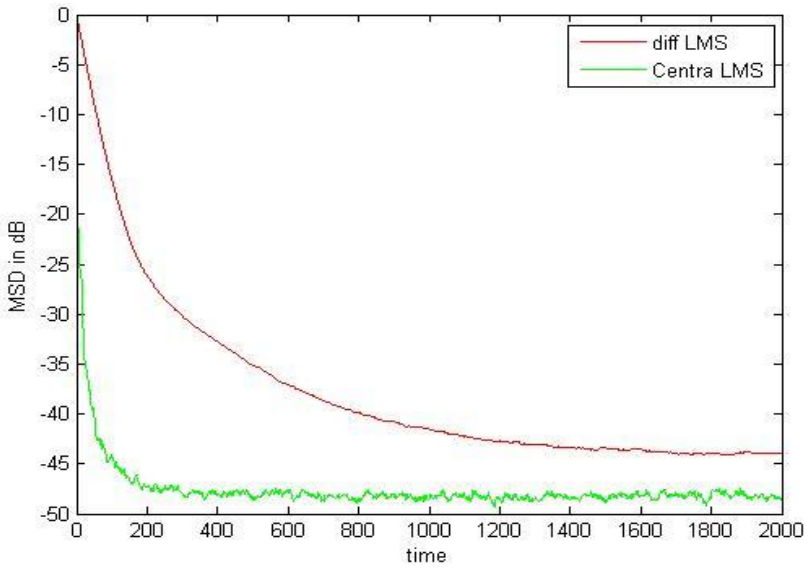

Fig. 9 MSD performance of diffusion and centralized LMS

Two dimensional network:

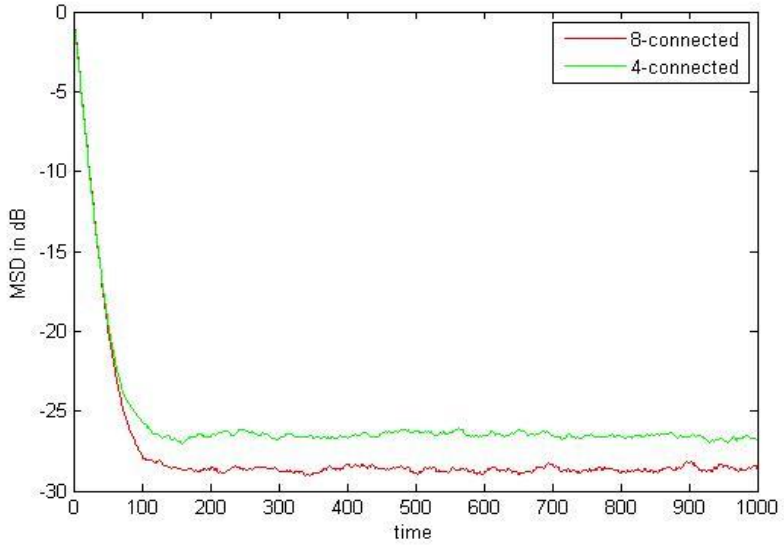

Fig. 10 MSD performance of 4- and 8-connected neighbors

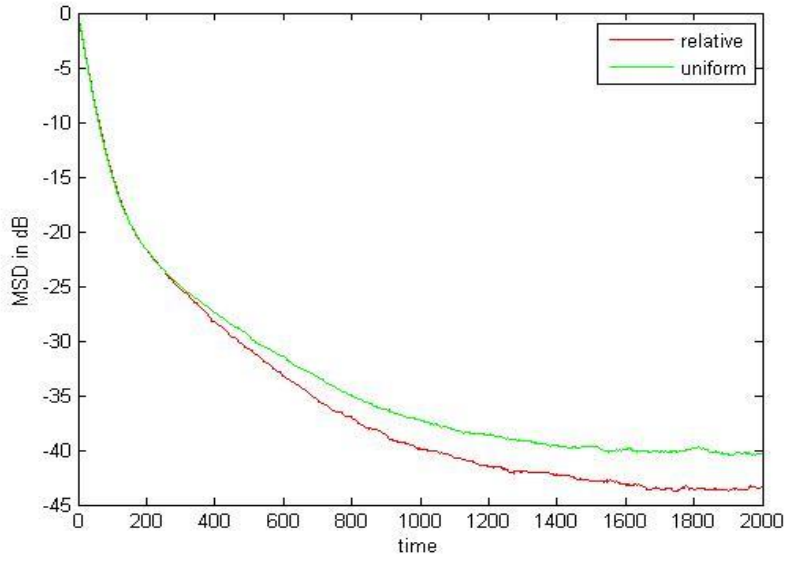

Fig. 11 MSD of relative and uniform combination rules

Chapter 6

Conclusion

6.1 Conclusion

The results of this effort illustrate that performance is enhanced by cooperation scheme from the assessment of the parameters. It shows that the network is stable and improvement of global performance by the utilization of diffusion cooperation scheme. Utilizing the distributed adaptive algorithms and interpolation scheme, the parameters of our concern can be precise with their estimation and tracking with diffusion LMS strategy. The discussed algorithm assesses the parameters both at node positions and locations where no readings are collected. The performance of learning curves depends upon combination matrices that are used to merge information exchanged with the nodes. As the degree of all nodes varies, the stochastic matrices i.e. combination rules also vary. The learning curve with more step size converges faster than the one with less step size. The MSD values decrease with increase in number of neighbours.

Bibliography

[1] F. S. Cattivelli and A. H. Sayed, “Diffusion LMS strategies for distributed estimation,” IEEE Trans. Signal Process., vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[2] J. Chen and A. H. Sayed, “Diffusion adaptation strategies for distributed optimization and learning over networks,” IEEE Trans. Signal Process., vol. 60, no. 8, pp. 4289–4305, Aug. 2012.

[3] A. H. Sayed, “Diffusion adaptation over networks,” in Academic Press Library in Signal Processing, 2013. , R. Chellapa and S. Theodoridis, Eds. New York, NY, USA: Academic, Elsevier, pp. 323–454, 2012.

[4] E. Holmes, M. Lewis, J. Banks, and R. Veit, “Partial differential equations in ecology: Spatial interactions and population dynamics,” Ecology, vol. 75, no. 1, pp. 17–29, Jan. 1994.

[5] A.H.Sayed and Benoit Champagne, “Estimation of Space-Time varying parameters using a Diffusion LMS algorithm”, IEEE Trans. Signal Process. Vol. 62, no.2, Jan. 2014.

[6] N. V. Kampen, “Diffusion in inhomogeneous media,” J. Phys. Chem. Solids, vol. 49, no. 1, pp. 673–677, Jun. 1988.

[7] A. H. Sayed, Fundamentals of Adaptive Filtering. New York: Wiley, 2003.

[8] C. G. Lopes and A. H. Sayed, “Incremental adaptive strategies over distributed networks,” IEEE Trans, Signal Process,, vol. 55, no. 8, pp. 4064–4077, Aug. 2007.

[9] A. H. Sayed and C. G. Lopes, “Adaptive processing over distributed networks,” IEICE Trans. Fund. Electron., Commun. Comput. Sci., vol. E90-A, no. 8, pp. 1504–1510, 2007.

[10] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. Towfic, “Diffusion strategies for adaptation and learning over networks,” IEEE Signal Process. Mag., vol. 30, pp. 155–171, May 2013.

[11] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A survey on sensor networks,” IEEE Commun. Mag., vol. 40, no. 8, pp. 102–114, Aug. 2002.

[12] R.Mattheij, S. Rienstra, and J. ten Thije Boonkkamp, Partial Differential Equations: Modeling, Analysis, Computation. Philadelphia, PA, USA: SIAM, 2005.

[13] J. Thomas, Numerical Partial Differential Equations: Finite Difference Methods. New York, NY, USA: Springer-Verlag, 1995.

[14] J. C. Mason and D. C. Handscomb, “Chebyshev Polynomials” Boca Raton, FL, USA: Chapman & Hall/CRC, 2003.

[15] George T. Gilbert, “Orthogonal Polynomials”, TCU Seminar Lecture Notes, Feb. 2011.