Signal Processing 94 (2014) 373–385

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Enhanced incremental LMS with norm constraints for distributed in-network estimation

Ying Liu [a,*], Wallace K.S. Tang [b]

[a] Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, P.R. China
[b] Department of Electronic Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 30 July 2012
Received in revised form
30 June 2013
Accepted 1 July 2013
Available online 11 July 2013

Keywords:
Distributed estimation
Incremental least-mean-square
Network
Norm constraint
Sparse

ABSTRACT

This paper addresses the problem of distributed in-network estimation for a vector of interest, which is sparse in nature. To exploit the underlying sparsity of the considered vector, the ℓ_1 and ℓ_0 norms are incorporated into the quadratic cost function of the standard distributed incremental least-mean-square (DILMS) algorithm, and some sparse DILMS (Sp-DILMS) algorithms are proposed correspondingly. The performances of the proposed Sp-DILMS algorithms in the mean and mean-square derivation are analyzed. Mathematical analyses show that the Sp-DILMS outperforms the DILMS, if a suitable intensity of the zero-point attractor is selected. Considering that such intensity may not be easily determined in real cases, a new adaptive strategy is designed for its selection. Its effectiveness is verified by both theoretical analysis and numerical simulations. Even though the criterion for intensity selection is derived from the case that the observations are white and Gaussian, simulation results show that it still provides an empirical good choice if the observations are correlated regression vectors.

1. Introduction

Distributed in-network estimation has recently aroused considerable interest and been applied to many real-world applications, such as signal and information processing, target tracking and localization, precision agriculture and environmental monitoring [1]. Considering a set of nodes distributed over a geographic region, distributed estimation manages the data collected from multiple nodes to extract the information of interest [2–10]. Unlike the centralized processing, which relies on a powerful centralized fusion center to collect all the nodes' information, distributed estimation allows each node to exchange information with a subset of its neighbors for local computation. The information of interest, for which usually expressed in a vector form, is then estimated by exploiting the spatial diversity of geographically distributed nodes as well as the temporal diversity over the network. Since no data-fusion center is needed and every node shares a certain computational burden, distributed estimation is more robust, scalable and thus more desirable for practical applications.

In the last decade, various distributed algorithms have been proposed, such as incremental LMS [2–4], incremental affine projection [5], diffusion LMS [6–9], diffusion Kalman filtering and smoothing [10]. Although these algorithms provide good solutions for general cases, the prior characteristics of the vector of interest are ignored, and each component is estimated in the same manner. So, it is believed that there are rooms for performance improvement if some distinct properties of the unknown vectors are taken into account.

On the other hand, sparsity commonly exists in nature. For example, in the context of target counting and localization, the position vector of the unknown point target is usually sparse because the number of grids utilized to

* Corresponding author. Tel.: +86 57187951577.
E-mail addresses: yingliu@zju.edu.cn (Y. Liu), eekstang@cityu.edu.hk (W.K.S. Tang).

0165-1684/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2013.07.001

represent the target's position is much more comparing to that of the target [11]. Besides, some communication channels are also shown to be sparse [12–14]. Recent advances in compressed sensing have shown that exploiting the sparsity of the vector of interest contributes to improving the performance of estimation in both convergence and accuracy [15–22]. Unfortunately, these works did not consider the designs of distributed solutions over networks. To tackle with the problem, we have proposed some sparse estimation algorithms based on *diffusion cooperative protocol* using the regularized estimation [23]. By penalizing the cost function of the standard diffusion LMS with some norm constraints, a zero-point attractor is incorporated in the adaptive process so as to enforce the sparsity of the vector of interest. In this paper, we further extend the concept of sparse estimation into the *incremental cooperative protocol* [2–4], and some sparse distributed incremental LMS (Sp-DILMS) are proposed. Furthermore, an adaptive rule for selecting the intensity of the sparse LMS is developed and its effectiveness is verified with simulations.

The rest of the paper is organized as follows. In Section 2, the problem of sparse vector estimation is formulated by penalizing the cost function of the standard DILMS with different norm constraints, and some Sp-DILMS algorithms are developed. Their performances under the measures of mean stability and mean-square errors are analyzed in Section 3. In Section 4, an adaptive criterion for selecting the intensity of the sparse LMS is proposed. The effectiveness of the proposed algorithms is then verified with numerical simulations in Section 5. Finally, conclusions are given in Section 6.

2. Problem formulation and methodology

In this section, we formulate the problem of signal estimation with a distributed incremental cooperative manner by taking the sparsity of unknown vector into account. In what follows, we let the boldface letters denote the random quantities and normal font denote the non-random (deterministic) quantities. For example, $\boldsymbol{d}_k(i)$ is a random observation quantity and $d_k(i)$ is a realization or measurement for it. Capital letters denote matrices and small letters denote vectors or scalars. The notation $(\cdot)^T$ represents the transpose of a vector or a matrix with real data.

2.1. Sparse distributed incremental LMS

Consider a network consisting of N nodes distributed over a geographic region. At every time instant i, each node k takes a scalar measurement $d_k(i) \in \Re$ and an M-dimensional row vector $u_{k,i} \in \Re^M$, where both are real, forming a random process $\{\boldsymbol{d}_k, \boldsymbol{u}_k\}$. The measurement $d_k(i)$ and the observations $u_{k,i}$ are assumed to follow a linear model given by

$$d_k(i) = u_{k,i} w^o + v_k(i), \tag{1}$$

where $w^o = [w_1^o, w_2^o, \dots, w_M^o]^T \in \Re^M$ is an M-dimensional sparse vector, $v_k(i)$ denotes a noise sequence with variance $\sigma_{v,k}^2$. Our target is to find a good estimate for w^o by taking its sparseness into account, based on the data $\{\boldsymbol{d}_k, \boldsymbol{u}_k\}$. In this paper, the DILMS proposed in [2–4] is focused and its cooperation strategy is depicted in Fig. 1. In DILMS, each node k only cooperates with one of its immediate neighbors $(k-1)$ in a cyclic pattern. It demands fewer communications and consumes relatively lower energy comparing to other cooperation protocols. It is also worth to point out that, although the incremental cooperative rule requires a Hamiltonian cycle through which signal estimates are sequentially circulated from one node to another, it does not mean that the nodes have to distribute with a cyclic topology. For more details, readers can refer to [2–4].

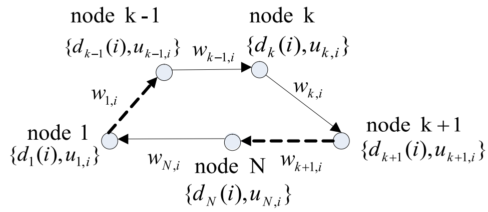

Fig. 1. Distributed incremental cooperation structure over the networks.

For each node k, this estimation problem can be tackled by minimizing the following ℓ_p−norm penalized cost function with regard to its neighbor node $(k-1)$

$$\min_w J_k(w_{k-1}) = E\|\boldsymbol{d}_k - \boldsymbol{u}_k w_{k-1}\|_2^2 + \gamma_k \xi_p(w_{k-1}), \tag{2}$$

where w_{k-1} denotes the estimate for w^o at node $(k-1)$, $\xi_p(w_{k-1}) \triangleq \|w_{k-1}\|_p$ denotes the ℓ_p−norm with regard to w_{k-1}, and $\gamma_k \ge 0$ is the regularization parameter to balance the penalty of the norm constraint and the estimation error. When $\gamma_k = 0$, (2) equals to the cost function of the conventional DILMS.

It is remarked that the optimization is performed locally at each node based on (2). Although global optimization with cost function

$$\min_w J(w) = \sum_{k=1}^{N} J_k(w), \tag{3}$$

may be desirable, it requires a much higher communication cost as all the other nodes' information should be available.

Following the procedure in the standard DILMS [2–4], we can derive the steepest-descent algorithm with an incremental implementation such that the recursion with the ℓ_p−norm is determined by

$$w_{k,i} = w_{k-1,i} - \mu_k [\nabla_w J_k(w_{k-1,i})], \tag{4}$$

where $\mu_k > 0$ is a step-size parameter to control the convergence rate of the estimate, and $\nabla_w J_k(w_{k-1,i})$ denotes the partial gradient $\nabla J_k(\cdot)$ with respect to the local estimate $w_{k-1,i}$ evaluated at node $(k-1)$, which is given by[1]

$$\nabla_w J_k(w_{k-1,i}) = R_{u,k} w_{k-1,i} - R_{du,k} + \gamma_k \partial \xi_p(w_{k-1,i}), \tag{5}$$

[1] In fact, there is a factor of 2 multiplying with the first term of the right-hand side of (5) when the data are real. Here, we ignore it as it can be absorbed into the step-size μ_k in (4).