2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Decentralized Identification for Errors-in-Variables Systems Based on a Consensus Algorithm

Miloš S. Stanković, Srdjan S. Stanković and Dušan M. Stipanović

Abstract— **In this paper a new consensus based algorithm for decentralized recursive estimation of parameters in linear discrete-time stochastic errors-in-variables MIMO systems is proposed. One starts from a multi-agent setting, in which an agent has access only to a subset of noisy input-output variables. The proposed algorithm consists of two stages. The first stage is based on a combination of local stochastic approximation algorithms for estimating input-output covariance functions based on locally available measurements and a dynamic first order consensus scheme. At the second stage each agent utilizes a stochastic approximation algorithm with expanding truncations for generating all system parameter estimates on the basis of current estimates of the matrices in the modified Yule-Walker equations obtained at the first stage. In the given convergence analysis it is proved that the estimates of the covariance functions and the overall parameter estimates converge almost surely to their true values under appropriate assumptions concerning system properties and the multi-agent network topology.**

I. Introduction

Theory and practice of complex and large scale systems has been focused on decentralized methods for control and estimation, *e.g.*, [1], [2]. Considering a dynamic system as an interconnection of subsystems, which may be either physical entities of mathematical abstractions, it is possible to achieve reduction of dimensionality in control and estimation problems, as well as to increase robustness and reliability (see, *e.g.*, [1]). In many real life situations designers of control systems are faced with inherent *decentralized information structure constraints*, leading to different types of *decentralized design* or *decentralized implementation*.

A lot of attention has been paid recently to *networked control systems*, *multi-agent systems* and *sensor networks*, *e.g.*, [3]. Many efficient methodologies have been theoretically developed and applied in practice. It has been found that one of very successful methodologies for getting efficient compromises between decentralization and global functioning of large scale systems is based on dynamic consensus schemes, *e.g.*, [4]–[8]. Within the framework of decentralized methods for large scale systems, consensus has been applied to state estimation, optimization and control, *e.g.*, [9]–[13].

Miloš S. Stanković who is the corresponding author is with the School of Electrical Engineering, KTH Royal Institute of Technology, 100-44 Stockholm, Sweden; E-mail: milsta@kth.se

Srdjan S. Stanković is with Faculty of Electrical Engineering, University of Belgrade, Serbia. E-mail: stankovic@etf.rs

Dušan M. Stipanović is with the Department of Industrial and Enterprise Systems Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Illinois, USA. E-mail: dusan@illinois.edu

Decentralized identification of dynamic systems has not been yet considered in the literature from a general standpoint (to the best of the authors' knowledge). The main problem is the contradiction between decentralized data acquisition implied by the information structure constraints, on the one side, and the overall interconnectedness between the model variables and functioning of the system as a whole, on the other.

The concept of decentralized identification of a given multivariable large scale system developed in this paper starts from a multi-agent setting in which an agent has access only to a *subset of noisy input and output measurements*, in accordance with the given information structure constraints. The task of every agent is to obtain estimates of the *overall system parameters* in real time, *without recurring to any type of centralized strategy or fusion center*. The contradiction between the global nature of the task and local constraints is resolved by introducing communications between the agents, implementing a *dynamic consensus strategy*.

It will be assumed that the system to be identified is represented by a linear MIMO (multiple input - multiple output) model with the input generated by a MIMO ARMA model [14], [15]. The assumed measurement setting represents a generalization of the basic errors-in-variables identification problem [16], [17]. The starting point is to construct an algorithm in which the agents recursively estimate a set of input - output covariance functions using a dynamic consensus scheme by algorithms of stochastic approximation type, applying the methodology from [9]–[11], and extending the idea presented in [18]. The obtained current covariance estimates are then used within a stochastic approximation scheme with expanding truncations in order to generate recursively estimates of the overall system model parameters using the modified Yule-Walker equations [14], [18], [19].

The paper is organized as follows. Section II contains the problem formulation and the proposed algorithm definition, while Section III is devoted to the convergence analysis.

II. Problem Definition and Algorithm Formulation

Consider a linear dynamic MIMO system described by

$$A(z)y(t) = B(z)u(t), \tag{1}$$

where t is a discrete time instant, z stands for the backward shift operator, $y(t) \in R^n$ and $u(t) \in R^m$ are the input and output vectors, respectively, while $A(z) = I + A_1 z + \cdots + A_{\nu_a} z^{\nu_a}$ and $B(z) = B_1 z + \cdots + B_{\nu_b} z^{\nu_b}$, where

$A_i = \text{diag}\{a_1^{[i]}, \ldots, a_n^{[i]}\}$, $i = 1, \ldots, \nu_a$, and $B_i = [b_{jk}^{([i]}]$, $i = 1, \ldots, \nu_b$, $j = 1, \ldots, n$, $k = 1, \ldots, m$.

The input $u(t)$ is assumed to be a MIMO ARMA process generated by

$$P(z)u(t) = Q(z)\varepsilon(t), \tag{2}$$

where $P(z) = I + P_1 z + \cdots + P_{\nu_p} z^{\nu_p}$ and $Q(z) = I + Q_1 z + \cdots + Q_{\nu_q} z^{\nu_q}$, $P_i = \text{diag}\{p_1^{[i]}, \ldots, p_n^{[i)]}\}$, $i = 1, \ldots, \nu_p$, and $Q_i = [q_{jk}^{[i]}]$, $i = 1, \ldots, \nu_q$, $j = 1, \ldots, m$, $k = 1, \ldots, m$. The sequence $\{\varepsilon(t)\}$ is assumed to be composed of i.i.d. m-dimensional zero mean random vectors satisfying $E\{\varepsilon(t)\varepsilon(t)^T\} = R_\varepsilon$ and $E\{\|\varepsilon(t)\|^4\} < \infty$.

A compact representation of (1) and (2) is given by $G(z)z(t) = S(z)\varepsilon(t)$, where $G(z) = \left[\begin{array}{c:c} A(z) & -B(z) \\ \hdashline 0 & P(z) \end{array}\right] = I + G_1 z + \cdots + G_\nu z^\nu$ and $S(z) = \left[\begin{array}{c} 0 \\ \hdashline Q(z) \end{array}\right] = S_0 + S_1 z + \cdots + S_\nu z^\nu$, while $\nu = \max(\nu_a, \nu_b, \nu_p, \nu_q)$ [14], [15].

Set $Y_j(t-1) = [y_j(t-1) \cdots y_j(t-\nu_a)]^T$, $U_k(t-1) = [u_k(t-1) \cdots u_k(t-\nu_b)]^T$, $\varphi_j(t) = [Y_j(t)^T \vdots U_1(t)^T \cdots U_m(t)^T]^T$, and

$$\theta_j^y = [-a_j^{[1]} \cdots - a_j^{[\nu_a]} \vdots b_{j1}^{([1]} \cdots b_{j1}^{[\nu_b]} \vdots \cdots \vdots b_{jm}^{([1]} \cdots b_{jm}^{[\nu_b]}],$$

$$\theta_k^u = [-p_k^{[1]} \vdots \cdots \vdots - p_k^{[\nu_p]}],$$

$j = 1, \ldots, n$, $k = 1, \ldots, m$. Let, also, $\Phi(t) = \text{diag}\{\varphi_1(t), \ldots, \varphi_n(t)\}$, $\Psi(t) = \text{diag}\{U_1(t), \ldots, U_m(t)\}$, $\theta^y = \text{diag}\{\theta_1^y, \ldots, \theta_n^y\}$, $\theta^u = \text{diag}\{\theta_1^u, \ldots, \theta_m^u\}$. From (1) and (2) we have

$$z(t) = \left[\begin{array}{c} y(t) \\ \hdashline u(t) \end{array}\right] = \theta^T Z(t-1) + \left[\begin{array}{c} 0 \\ \hdashline Q(z)\varepsilon(t) \end{array}\right], \tag{3}$$

where $\theta = \text{diag}\{\theta^y, \theta^u\}$ is the parameter matrix and $Z(t) = \text{diag}\{\Phi(t), \Psi(t)\}$.

Assume that N autonomous agents perform input-output measurements in such a way that every agent has access to a *subset of input/output variables* contaminated by *local additive measurement noises*. Let the index sets $S_u^{(i)} = \{m_1^{(i)}, \ldots, m_{\sigma_u^{(i)}}^{(i)}\} \subseteq \{1, \ldots, n\}$, $S_y^{(i)} = \{l_1^{(i)}, \ldots, l_{\sigma_y^{(i)}}^{(i)}\} \subseteq \{1, \ldots, m\}$, ($\sigma_u^{(i)} \leq m$, $\sigma_y^{(i)} \leq n$) determine the indices of the components of the vectors $u(t)$ and $y(t)$ accessible to the *i*-th agent (the index sets can be overlapping). Accordingly, define the vector $z^{(i)}(t)$ (derived from $z(t)$ in (3)), which contains as its nonzero components the measurements acquired by the i-th agent, as

$$z^{(i)}(t) = \left[\begin{array}{c} y^{(i)}(t) \\ \hdashline u^{(i)}(t) \end{array}\right] = \left[\begin{array}{c} [y(t) + \xi^{(i)}(t)]I_y^{(i)}(t) \\ \hdashline [u(t) + \eta^{(i)}(t)]I_u^{(i)}(t) \end{array}\right], \tag{4}$$

where $\xi^{(i)}(t)$ and $\eta^{(i)}(t)$ are measurement noise vectors, while $I_y^{(i)}(t) = \text{diag}\{I_{y,1}^{(i)}(t), \ldots, I_{y,n}^{(i)}(t)\}$ and $I_u^{(i)}(t) = \text{diag}\{I_{u,1}^{(i)}(t), \ldots, I_{u,m}^{(i)}(t)\}$, where $\{I_{y,j}^{(i)}(t)\}$ and $\{I_{u,k}^{(i)}(t)\}$, $j = 1, \ldots, n$, $k = 1, \ldots, m$, are random binary sequences independent from the system variables such that, for all t, $P\{I_{y,j}^{(i)}(t) = 1\} > 0$ when $j \in S_y^{(i)}$ and $P\{I_{u,k}^{(i)}(t) = 1\} > 0$ when $k \in S_u^{(i)}$; otherwise, $P\{I_{u,j}^{(i)}(t) = 1\} = 0$ and $P\{I_{u,k}^{(i)}(t) = 1\} = 0$. Randomness of $I_y^{(i)}(t)$ and $I_u^{(i)}(t)$ is introduced in order to model intermittent measurements.

The input and output measurement noise vector sequences $\{\eta^{(i)}(t)\} = \{[\eta_1^{(i)}(t) \cdots \eta_m^{(i)}(t)]^T\}$ and $\{\xi^{(i)}(t)\} = \{[\xi_1^{(i)}(t) \cdots \xi_n^{(i)}(t)]^T\}$, $i = 1, \ldots, N$, are assumed to be i.i.d. zero mean sequences mutually independent and independent from $\{\varepsilon(t)\}$, satisfying $E\{\xi^{(i)}(t)\xi^{(i)}(t)^T\} = R_\xi^{(i)}$, $E\{\eta^{(i)}(t)\eta^{(i)}(t)^T\} = R_\eta^{(i)}$, $E\{\|\xi_j^{(i)}(t)\|^4\} < \infty$ and $E\{\|\eta_k^{(i)}(t)\|^4\} < \infty$, for all t.

According to (3), we introduce also $Y_j^{(i)}(t)$, $U_k^{(i)}(t)$, $\varphi_j^{(i)}(t)$, $\Phi^{(i)}(t)$, $\Psi^{(i)}(t)$ and $Z^{(i)}(t)$, $i = 1, \ldots, N$, $j = 1, \ldots, n$, $k = 1, \ldots, m$ as the "noisy" versions of $Y_j(t)$, $U_k(t)$, $\varphi_j(t)$, $\Phi(t)$, $\Psi(t)$ and $Z(t)$, respectively. Similarly, we construct $\tilde{Y}_j^{(i)}(t)$, $\tilde{U}_k^{(i)}(t)$, $\tilde{\varphi}_j^{(i)}(t)$, $\tilde{\Phi}^{(i)}(t)$, $\tilde{\Psi}^{(i)}(t)$ and $\tilde{Z}^{(i)}(t)$ in the same way as $Y_j^{(i)}(t)$, $U_k^{(i)}(t)$, $\varphi_j^{(i)}(t)$, $\Phi^{(i)}(t)$, $\Psi^{(i)}(t)$ and $Z^{(i)}(t)$, $i = 1, \ldots, N$, $j = 1, \ldots, n$, $k = 1, \ldots, m$, but with $\tilde{Y}_j^{(i)}(t) = [y_j^{(i)}(t) \cdots y_j^{(i)}(t-\mu)]^T$ and $\tilde{U}_k^{(i)}(t) = [u_k^{(i)}(t) \cdots u_k^{(i)}(t-\mu)]^T$ instead of $Y_j^{(i)}(t)$, $U_k^{(i)}(t)$, respectively, where μ is a selected integer.

We assume that $A(z)$ and $P(z)$ are stable matrices, so that $\{z(t)\}$ is a stationary and ergodic sequence with $E\{\|z(t)\|^2\} < \infty$, and

$$E\{z(t)z(t-\tau)^T\} = R^{zz}(\tau) = \left[\begin{array}{c} R^{yy}(\tau) \;\; R^{yu}(\tau) \\ R^{uy}(\tau) \;\; R^{uu}(\tau) \end{array}\right] \tag{5}$$

where $R^{yy}(\tau) = [r_{kl}^{yy}(\tau)]_{n\times n} = E\{y(t)y(t-\tau)^T\}$, $R^{yu}(\tau) = [r_{kl}^{yu}(\tau)]_{n\times m} = E\{y(t)u(t-\tau)^T\}$ and $R^{uu}(\tau) = [r_{kl}^{uu}(\tau)]_{m\times m} = E\{u(t)u(t-\tau)^T\}$. Introduce λ such that $\mu \geq \lambda > \nu_q$, so that we obtain, after multiplying (3) with $\tilde{Z}(t-1-\lambda)$ from the right and taking the mathematical expectation, the following modified Yule-Walker equations

$$W = \theta^T G, \tag{6}$$

where $W = E\{z(t)\tilde{Z}(t-1-\lambda)^T\}$ and $G = E\{Z(t-1)\tilde{Z}(t-1-\lambda)^T\}$ [18], [19]. Solving (6) one obtains, in general, $\hat{\theta} = WG^+ + \theta_0(I - GG^+)$, where θ_0 is any matrix with appropriate dimensions; the necessary and sufficient condition for uniquely defining θ from (6) is non-degeneracy of G (A^+ stands for the pseudoinverse of a matrix A), [18].

Based on the exposed formal setting we formulate the problem of distributed identification of the system represented by (1) and (2) as the problem of estimating the parameter matrix θ in (3) in real time *by all the agents* using only *local measurements* described in (4), by allowing *inter-agent communication* of local (partial) identification results (not of measurements themselves), without recurring to any type of fusion center. The formulated problem belongs to system identification problems of errors-in-variables type (see, *e.g.*, [16]); the problem is, however, specific, due to the given information structure constraints.

Estimation of the correlation functions between the measured input-output variables represents a firm starting point

for the formulated identification task, since it can be done locally irrespective of the functioning of the rest of the system (see also, *e.g.*, [20]). By interchanging the local correlation function estimates between the agents using the available communication links one can circumvent the given information structure constraints and get a model of the system as a whole. Once an estimate of input/output correlation functions is available to all the agents, they can locally estimate all the system parameters by using, for example, the above formulated modified Yule-Walker equations (6).

We propose the following distributed recursive algorithm based on stochastic approximation for estimating the unknown parameter matrix θ in (3) by all N agents:

$$\bar{G}^{(i)}(t) = G^{(i)}(t-1)+ \\ + \frac{1}{t}[Z^{(i)}(t-1)\tilde{Z}^{(i)}(t-\lambda-1)^T - \tilde{G}^{(i)}(t-1)], \quad (7)$$

$$\mathrm{vec}\{G^{(i)}(t)\} = \sum_{j=1}^{N} C_{ij}^{G}(t)\,\mathrm{vec}\{\bar{G}^{(j)}(t)\}, \quad (8)$$

$$\bar{W}^{(i)}(t) = W^{(i)}(t-1)+ \\ + \frac{1}{t}[z^{(i)}(t)\tilde{Z}^{(i)}(t-\lambda-1)^T - \tilde{W}^{(i)}(t-1)], \quad (9)$$

$$\mathrm{vec}\{W^{(i)}(t)\} = \sum_{j=1}^{N} C_{ij}^{W}(t)\,\mathrm{vec}\{\bar{W}^{(j)}(t)\}, \quad (10)$$

$$\theta^{(i)}(t) = \bar{\theta}^{(i)}(t-1) I_{[\|\bar{\theta}^{(i)}(t-1)\| \le M(\sigma^{(i)}(t-1))]}, \quad (11)$$

$$\bar{\theta}^{(i)}(t-1) = \theta^{(i)}(t-1)+ \\ + \frac{1}{t}G^{(i)}(t)[W^{(i)}(t)^T - G^{(i)}(t)^T\theta^{(i)}(t-1)], \quad (12)$$

$$\sigma^{(i)}(t) = \sigma^{(i)}(t-1) + I_{[\|\bar{\theta}^{(i)}(t-1)\| > M(\sigma^{(i)}(t-1))]}, \quad (13)$$

where:

- $G^{(i)}(t)$, $W^{(i)}(t)$ and $\theta^{(i)}(t)$ are the current estimates of G, W and θ in the model (3) generated by the i-th agent at the instant t (the initial values $G^{(i)}(0)$, $W^{(i)}(0)$ and $\theta^{(i)}(0)$ are arbitrary and finite almost surely),

- $\tilde{G}^{(i)}(t)$ and $\tilde{W}^{(i)}(t)$ are obtained from $G^{(i)}(t)$ and $W^{(i)}(t)$ by inserting zeros at those places where the elements of $Z^{(i)}(t)\tilde{Z}^{(i)}(t-\lambda)^T$ and $z^{(i)}(t+1)\tilde{Z}^{(i)}(t-\lambda)^T$, respectively, are equal to zero, as a consequence of the fact that the corresponding instantaneous values of the indices $I_{y,j}^{(i)}(.)$ and $I_{u,k}^{(i)}(.)$ are equal to zero (accessibility of measurements, intermittent observations),

- $\mathrm{vec}\{A\}$ denotes the column vector obtained by concatenating all the column vectors of A,

- $I_{[A]}$ denotes the indicator of an ω-set: $I_{[A]} = 1$ if $\omega \in A$ and $I_{[A]} = 0$ if $\omega \notin A$,

- $M(\sigma^{(i)}(t))$ is a monotonically increasing function to be specified later;

- matrices $C_{ij}^{G}(t)$ and $C_{ij}^{W}(t)$, defining communications between the agents, are assumed to be given by $C_{ij}^{G}(t) = J_{ij}(t)\bar{C}_{ij}^{G}(t)$ and $C_{ij}^{W}(t) = J_{ij}(t)\bar{C}_{ij}^{W}(t)$, where $\bar{C}_{ij}^{G}(t)$ and $\bar{C}_{ij}^{W}(t)$ are diagonal matrices with positive elements (see, *e.g.*, [9]–[11]), while $J_{ii}(t) = 1$, $i = 1, \ldots, N$ and $\{J_{ij}(t)\}$, $i, j = 1, \ldots, N$, $i \neq j$, are scalar random binary sequences satisfying $P\{J_{ij}(t) = 1\} > 0$ for $(i, j) \in \mathcal{E} \subset \{1, \ldots N\} \times \{1, \ldots, N\}$, where the set $\mathcal{E}$ specifies the communication structure constraints, *i.e.*, the existing communication links between the agents (pair (i, j) denotes the directed link from the agent j to the agent i); otherwise, $P\{J_{ij}(t) = 1\} = 0$;

- $\sum_{j=1}^{N} C_{ij}^{G}(t) = I$ and $\sum_{j=1}^{N} C_{ij}^{W}(t) = I$ for all t and i.

The algorithm consists formally of two main parts, the first consisting of two consensus based stochastic approximation algorithms for estimating covariance matrices G and W in (6) (algorithms (7) and (8)), and the second representing a recursive algorithm for final estimation of the model parameters θ in (6) using the currently available covariance estimates (algorithm (9)). The algorithm represents an extension to the decentralized multi-agent case of the algorithm proposed in [18] for centralized errors-in-variables identification.

In this way, the whole algorithm extends the idea of combining recursive state and parameter estimation algorithms with the first order consensus scheme (presented in [9]–[11]) to the problem of system identification. Recall that recursions of stochastic approximation type with consensus have been discussed from a general standpoint in [21], proposed for distributed optimization in [4] and applied to parameter estimation in linear static regression models in [11].

It will be adopted that the communication links between the agents with positive communication probabilities are represented by a directed graph $G = \{\mathcal{N}, \mathcal{E}\}$, where $\mathcal{N}$ is the node set, and $\mathcal{E}$ the edge set. Correspondingly, it is possible to compose the overall communication (consensus) matrix $C^{net}(t) = [C_{ij}^{GW}(t)]$, $i, j = 1, \ldots, N$, where $C_{ij}^{GW}(t) = \mathrm{diag}\{C_{ij}^{G}(t), C_{ij}^{W}(t)\}$, which has zero blocks almost surely for $(i, j) \in \bar{\mathcal{E}} = \{1, \ldots N\} \times \{1, \ldots, N\} - \mathcal{E}$. Matrices $\bar{C}_{ij}^{G}(t)$ and $\bar{C}_{ij}^{W}(t)$ in $C_{ij}^{G}(t)$ and $C_{ij}^{W}(t)$ are assumed to be such that $C^{net}(t)$ is a row stochastic matrix for any realization of $J_{ij}(t)$, $i, j = 1, \ldots, N$ (concrete ways of generating realizations of $C^{net}(t)$ corresponding to the realizations of the $N \times N$ binary matrix $J(t) = [J_{ij}(t)]$ are discussed in [10]). Having in mind that (7) and (8) consist of independent recursions for estimating elements of G and W in (6), $C^{net}(t)$ can be decomposed into the corresponding number of $N \times N$ row stochastic consensus matrices.

III. Convergence Analysis

The key arguments for the convergence of the proposed identification algorithm are related to the estimation of the correlation functions in the schemes generating matrices $G^{(i)}(t)$ and $W^{(i)}(t)$ in (7) and (8). Our analysis will be initially concentrated on one arbitrarily selected element of either G or W; the obtained results can be directly generalized to all the elements of G and W.

Assume that $\rho^{(i)}(t)$ represent the current estimate of, *e.g.*, $r_{kl}^{yy}(\tau)$ generated by the i-th agent for some k and l, and τ satisfying (6). Then, according to (7) and (8), we have for

$i = 1, \ldots, N$

$$\bar{\rho}^{(i)}(t) = \rho^{(i)}(t-1) + \tfrac{1}{t} I_k^{(i)}(t-1) I_l^{(i)}(t-\tau) \times [y_k^{(i)}(t-1) y_l^{(i)}(t-\tau-1) - \rho^{(i)}(t-1)], \quad (14)$$

$$\rho^{(i)}(t) = \textstyle\sum_{j=1}^{N} c_{ij}(t) \bar{\rho}^{(j)}(t), \quad (15)$$

where $c_{ij}(t)$ are scalar coefficients of the form $c_{ij}(t) = J_{ij}(t)\bar{c}_{ij}(t)$ satisfying $\sum_{j=1}^{N} c_{ij}(t) = 1$ for all t, according to (7) and (8). Properties of $\bar{c}_{ij}(t)$ follow directly from the properties of either $\bar{C}_{ij}^{G}(t)$ or $\bar{C}_{ij}^{W}(t)$.

For vector $R(t) = [\rho^{(1)}(t) \cdots \rho^{(N)}(t)]^T$ we have

$$R(t) = C(t)R(t-1) + \frac{1}{t} C(t)\Gamma(t) \times [Y^D(t-1)Y(t-\tau-1) - R(t-1)], \quad (16)$$

where $C(t) = [c_{ij}(t)]$, $i, j = 1, \ldots, N$, $\Gamma(t) = \mathrm{diag}\{I_k^{(1)}(t-1) I_l^{(1)}(t-\tau-1), \ldots, I_k^{(N)}(t-1) I_l^{(N)}(t-\tau-1)\}$, $Y^D(t) = \mathrm{diag}\{y_k^{(1)}(t), \ldots, y_k^{(N)}(t)\}$ and $Y(t) = [y_l^{(1)}(t) \cdots y_l^{(N)}(t)]^T$. The consensus matrix $C(t)$ is a random row-stochastic matrix for all t, while $\Gamma(t)$ models the availability of measurements.

We will analyze convergence of (16) under the following assumptions:

A1) The set $\mathcal{N}^* \subset \mathcal{N}$ containing all the nodes of the graph $\mathcal{G}$ which have the indices i corresponding to the condition $P\{I_k^{(i)}(t)\} P\{I_l^{(i)}(t)\} > 0$ is nonempty, and each node in $\mathcal{N}$ is reachable from at least one node from $\mathcal{N}^*$.

A2) For all $(i,j) \in \mathcal{E}$ and all t, $P\{J_{ij}(\bar{t}) = 1\} = 1$ for some $\bar{t} \in [t, t+T_1]$, $T_1 < \infty$.

A3) For all $i \in \mathcal{N}^*$ and all t, $P\{I_k^{(i)}(\bar{t}) I_l^{(i)}(\bar{t}-\tau) = 1\} = 1$ for some $\bar{t} \in [t, t+T_2]$, $T_2 < \infty$.

Remark 1: Assumption A1) ensures that the measurements necessary for generating the estimates are available to every agent. Assumption A2) defines an upper bound on the intervals between two successive inter-agent communications, and Assumption A3) an upper bound on the intervals between two successive measurements.

Lemma 1: Let Assumption A1) be satisfied. Then, $E\{C(t)\}$ is cogredient (amenable by permutation transformations) to

$$\bar{C} = \begin{bmatrix} \bar{C}_1 & \cdots & & 0 \\ 0 & \bar{C}_2 & \cdots & \\ 0 & \cdots & \bar{C}_\kappa & 0 \\ \bar{D}_1 & \cdots & \bar{D}_\kappa & \bar{C}_0 \end{bmatrix}, \quad (17)$$

where $\bar{C}_0$ is an $\beta_0 \times \beta_0$ matrix, $0 \le \beta_0 < N$, $\bar{C}_i$ are *irreducible* $\beta_i \times \beta_i$ matrices satisfying $0 < \beta_i < N$, $\bar{D}_i$ are $\beta_0 \times \beta_i$ matrices, $i = 1, \ldots, \kappa$, and $1 \le \kappa \le |\mathcal{N}^*|$ ($|.|$ denotes cardinality of a given set).

The proof of the lemma can be derived by construction; details can be found in [11]. One can easily see that $E\{J(t)\}$ has the same structure of positive elements as $\bar{C}$, *i.e.*, $E\{C(t)\} \sim E\{J(t)\}$. The number of irreducible components κ in $\bar{C}$ and the dimension of $\bar{C}_0$ depend on the adopted network topology. Consequently, the structure of $\bar{C}$ allows updating of the correlation coefficient estimates for all the nodes.

Lemma 2: Let the l-th power of $\bar{C}$ in (17) be represented as

$$\bar{C}^l = \begin{bmatrix} \mathrm{diag}\{\bar{C}_1^l, \ldots, \bar{C}_k^l\} & 0 \\ \bar{D}^{[l]} & \bar{C}_0^l \end{bmatrix}, \quad (18)$$

where $\bar{D}^{[l]} = \begin{bmatrix} \bar{D}_1^{[l]} & \cdots & \bar{D}_k^{[l]} \end{bmatrix}$. Then:

(a) for each j for which $\bar{C}_j \neq 0$ there exists an integer l_j such that for $l \ge l_j$ the block matrix $\bar{D}_j^{[l]}$ contains at least one row whose all elements are positive;

(b) there exists an integer l' such that for $l \ge l'$ each row of $\bar{D}^{[l]}$ contains at least one entire row of some of the blocks $\bar{D}_j^{[l]}$, $j \in \{1, \ldots, k\}$, whose all elements are positive.

The proof of this lemma can be derived using the results in [11].

Theorem 1: Let the Assumptions A1) – A3) be satisfied. Then

$$\|R(t) - r_{kl}^{yy}(\tau)\mathbf{1}\| = o(t^{-\delta})$$

with probability one, where $0 < \delta < \frac{1}{2}$ and $\mathbf{1} = [1 \cdots 1]_{1\times N}^T$.

Proof: Denoting $\tilde{R}(t) = R(t) - r_{kl}^{yy}(\tau)\mathbf{1}$ and $e(t-1) = Y^D(t-1)Y(t-\tau-1) - r_{kl}^{yy}(\tau)\mathbf{1}$, we obtain from (16)

$$\tilde{R}(t) = C(t)(I - \frac{1}{t}\Gamma(t))\tilde{R}(t-1) + \frac{1}{t} C(t)\Gamma(t) e(t-1). \quad (19)$$

After iterating back to the initial condition, one obtains

$$\tilde{R}(t) = \Phi(t,\tau)\tilde{R}(\tau-1) + \sum_{j=\tau-1}^{t} \frac{1}{j} \Phi(t, j+1) C(j)\Gamma(j) e(j-1), \quad (20)$$

where $\Phi(t,s) = C(t)(I - \frac{1}{t}\Gamma(t)) \cdots C(s)(I - \frac{1}{s}\Gamma(s))$ for $t \ge s$, with $\Phi(t, t+1) = I$.

a) In order to analyze the first term at the right hand side of (20), assume first, for the sake of simplicity, that $E\{C(t)\}$ can be represented by a matrix cogredient to the structure (17), with $\kappa = 1$ and $\bar{C}_0 = 0$, *i.e.*, that $E\{C(t)\}$ is irreducible and also primitive, having in mind that it is aperiodic by positiveness of its diagonal elements (by assumption, $J_{ii}(t) = 1$, $i = 1, \ldots, N$). In this case, by Assumption A2), at almost every realization of the sequence $\{C(t)\}$, for each t large enough there exists such a $\tau_1 > 0$ that $\Pi(t, t-\tau_1) = C(t) \cdots C(t-\tau_1) \succ 0$ ($A \succ 0$ denotes that all the elements of a matrix A are positive). This conclusion follows directly from the fact that primitiveness of $E\{C(t)\}$ implies that there exists such a positive integer l that $(E\{C(t)\})^l \succ 0$, [22]. This essentially structural property of nonnegative matrices is easily extended to the products $\Pi(t, t-\tau_1)$ utilizing elementary properties of products of realizations of matrices $C(t)$ which have positive diagonals and positive elements located at some of the places of positive elements in $E\{C(t)\}$. Coming back to the definition of $\Phi(t,\tau)$, we immediately conclude that on almost every realization of the sequence $\{C(t)\}$, for each t large enough, there exists such a $\tau_2 > 0$ that $\Phi(t, t-\tau_2) = C(t)(I - \frac{1}{t}\Gamma(t)) \cdots C(t-\tau_2)(I - \frac{1}{t-\tau_2}\Gamma(t)) \succ 0$ (matrix $I - \frac{1}{t}\Gamma(t)$ is diagonal with positive elements at the diagonal).

Moreover, Assumption A3) implies that for each t large enough there exists such a $T_2 > 0$ that (at least one) element $\Gamma_{ii}(s) = 1$, where $t \leq s < t - T_2$, for some node i defined by the index set $S_y^{(i)}$. This, in turn, implies that for each t large enough there exists such a $\tau_3 > 0$ that $\Phi(t, t-\tau_3) \succ 0$ and, at the same time that $\Phi(t, t-\tau_3) \prec \Pi(t, t-\tau_3)$ in such a way that $\Phi(t, t-\tau_3)$ is obtained after multiplying at least one element in each row of $\Pi(t, t-\tau_3)$ by the term $1 - \frac{1}{t-\tau_3}$. One such term corresponds to one node having access to measurements; in the case of more nodes having access to measurements, more elements of $\Pi(t, t-\tau_3)$ become multiplied by analogous terms.

In the case of a general structure of $E\{C(t)\}$ which is cogredient to (17) (Lemma 1), we can apply Lemma 2 and conclude that at almost every realization of the sequence $\{C(t)\}$, for each t large enough there exists such a $\tau_4 > 0$ that $\Pi(t, t-\tau_4)$ is cogredient to a matrix similar to $\bar{C}^l$, where $\bar{C}^l$ is defined in (18), in which $\bar{C}_i^l \succ 0$, $i = 1, \ldots, \kappa$ and $\bar{D}_j^{[l]}$, $j \in \{1, \ldots, k\}$ satisfy assertion b) of Lemma 2 (recall that matrices $\bar{C}_i$ are primitive). The last statement follows directly form the fact that Lemma 2 is concerned with structural properties, rather than with particular values of the elements of the matrices in (18), so that the properties of the powers of $E\{C(t)\}$ can be directly transferred to the properties of similar matrices resulting from the products of the realizations of $C(t)$ (as above).

Coming back to the definition of $\Phi(t, \tau)$ in (20), we immediately conclude that at almost every realization of the sequence $\{C(t)\}$, for each t large enough there exists such a $\tau_5 > 0$ that $\Phi(t, t-\tau_5) = C(t)(I - \frac{1}{t}\Gamma(t)) \cdots C(t-\tau_5)(I - \frac{1}{t-\tau_5}\Gamma(t-\tau_5)) \succ 0$. Having in mind general properties of $\Gamma(t)$, we realize that there are at least κ positive diagonal elements of $E\{\Gamma(t)\}$ and that their indices are such that at least one such element corresponds to each irreducible component of $\Phi(t, t-\tau_5)$. Having in mind the important property from assertion b) in Lemma 2 applied to almost all realizations of $\Phi(t, 1)$, one concludes that for t large enough there exists such a $T > 0$ that $\Phi(t, t-T) \prec \Pi(t, t-T)$, in such a way that at least one element in each row of $\Pi(t, t-T)$ becomes multiplied by the term $1 - \frac{1}{t-T}$ in order to generate $\Phi(t, t-T)$.

Consequently, we have that

$$\|\Phi(t,\tau)\|_\infty \leq (1 - \frac{k_s}{sT}) \cdots (1 - \frac{k_1}{T}) \leq \exp(-\frac{1}{T}\sum_{i=1}^{s}\frac{k_i}{i}), \tag{21}$$

where $s = [\frac{t-\tau}{T}]$ and $0 < k_i < 1$, $i = 1, \ldots, s$, so that we conclude that $\lim_{t\to\infty} \|\Phi(t,\tau)\|_\infty = 0$ almost surely for any fixed τ, [23].

b) In order to analyze the second term at the right hand side of (20), we will construct the following state model

$$x(t) = Fx(t-1) + L\varepsilon(t), \quad z(t) = Hx(t), \tag{22}$$

where $F = \begin{bmatrix} -G_1 & I & \cdots & 0 & 0 \\ \vdots & & & & \\ -G_\nu & 0 & \cdots & 0 & I \\ 0 & 0 & \cdots & & 0 \end{bmatrix}$, $L = [S_0^T \cdots S_\nu^T]^T$ and $H = [I \vdots 0 \vdots \cdots \vdots 0]$.

Assuming, for the sake of clarity of presentation, that the required measurements are always available, we have further from (4) that $y_k^{(i)}(t) = h_k x(t) + \xi_k^{(i)}(t)$ where h_k, $k = 1, \ldots, n$, is the corresponding transformation derived from (22) and $\xi_k^{(i)}(t)$ the measurement noise. Consequently,

$$\begin{aligned} Y^D(t-1)Y(t-\tau-1) = h_k x(t-1)x(t-\tau-1)h_l^T\mathbf{1} + \\ +\Xi_k(t)x(t)^T h_k^T + \Xi_l(t-\tau)x(t-\tau)^T h_l^T + \\ +\Xi_k^D(t)\Xi_l(t-\tau), \end{aligned} \tag{23}$$

where $\Xi_k(t) = [\xi_k^{(1)}(t) \cdots \xi_k^{(N)}(t)]^T$ and $\Xi_l^D(t) = \mathrm{diag}\{\xi_l^{(1)}(t) \cdots \xi_l^{(N)}(t)\}$. Furthermore, the term $x(t)x(t-\tau)^T$ can be represented using [18], p. 1237, in the following convenient form following from (22):

$$\begin{aligned} x(t)x(t-\tau)^T = F^{t-\tau}x(\tau)x(0)^T F^{T(t-\tau)} + \\ \sum_{s=0}^{t-\tau-1} F^{t-s}x(s)\varepsilon(s+1)^T L^T F^{T(t-\tau-1-s)} + \\ \sum_{s=0}^{t-\tau-1}\sum_{j=s}^{s+\tau-1} F^{t-1-j}L\varepsilon(j+1)\times \\ \varepsilon(s+1)^T L^T F^{T(t-\tau-1-s)} + \\ \sum_{s=0}^{t-\tau-1} F^{t-\tau-1-s}\varepsilon(\tau+1+s)x(s)^T F^{T(t-\tau-s)} + \\ \sum_{s=0}^{t-\tau-1} F^{t-\tau-1-s}L\varepsilon(\tau+s+1)\times \\ \varepsilon(s+1)^T L^T F^{T(t-\tau-1-s)}. \end{aligned} \tag{24}$$

Using (24) we conclude that $e(t-1)$ consists of several terms involving the measurement noise on one side, and the input noise on the other side, and that all these terms are in the form of martingale difference sequences (except the one that depends on the initial condition, see [18], [23]).

Having in mind that a detailed analysis of all terms requires a lot of space and a very cumbersome notation, take, for example, the most important part of the third term at the right hand side of (24) containing only $\varepsilon(s+1)\varepsilon(s+1)^T$, $s = 0, \ldots, t-\tau-1$. In such a way we obtain the term

$$e_1(t-1) = h_k[\sum_{s=0}^{t-\tau-1} F^{t-s}LS(s+1)L^T F^{T(t-\tau-1-s)}]h_l^T\mathbf{1}, \tag{25}$$

where $S(s+1) = \varepsilon(s+1)\varepsilon(s+1)^T - R_\varepsilon$, which represents a part of $e(t-1)$. Coming back to (20), we introduce $e_1(j-1)$ in the second term from the right hand side and obtain

$$\begin{aligned} \alpha_1(t) = \sum_{s=0}^{t-\tau-1}\sum_{j=\tau+1+s}^{t}\frac{1}{j}\Phi(t, j+1)C(j)\Gamma(j)\mathbf{1} \\ \times h_k F^{j-s}LS(s+1)L^T F^{T(j-\tau-1-s)}h_l^T. \end{aligned} \tag{26}$$

As $\{S(t)\}$ is a zero mean i.i.d. sequence with $E\{||S(t)||^2\} < \infty$ according to the adopted assumptions, we can apply to (26) Theorem B.6.1 from [23], dealing with weighted sums of martingale difference sequences. We obtain that almost surely

$$\begin{aligned} &\|\alpha_1(t)\| = \\ &O(\sum_{s=0}^{t-\tau-1}[\sum_{j=\tau+1+s}^{t}\frac{1}{j^2}\|\Phi(t, j+1)\|^2\|F\|^{2(j-s)}]^{\frac{1}{2}}) = \\ &O(\sum_{s=0}^{t}[\frac{1}{j^2}\exp(-c\sum_{i=1}^{t}\frac{1}{i})]^{\frac{1}{2}}) \to_{t\to\infty} 0, \end{aligned} \tag{27}$$

where $c > 0$, having in mind that $\max_i\{|\lambda_i(F)|\} < 1$ (see [23], Lemma 3.1.1).

We can proceed in a completely analogous way with the analysis of the remaining components of $e(t-1)$.

Coming back to (20), we conclude that on almost all sample paths $\lim_{t\to\infty}\|\tilde{R}(t)\| = 0$.

c) The last part of the proof is devoted to the convergence rate of the recursion (20). In order to prove the assertion of the theorem, divide both sides of (19) by t^δ, $\delta > 0$. Denoting $\tilde{R}(t)/t^\delta$ as $\tilde{R}_\delta(t)$, we obtain for t large enough that

$$\tilde{R}_\delta(t) = C(t)(I-\frac{1}{t}\Gamma_\delta(t))\tilde{R}_\delta(t-1)+\frac{1}{t^{1-\delta}}C(t)\Gamma(t)e(t-1), \quad (28)$$

where $\Gamma_\delta(t) = \Gamma(t) - \delta I$, after using the representation $(\frac{t-1}{t})^\delta = 1 - \frac{\delta}{t} + O(\frac{1}{t^2})$. Using the same arguments as above, one can directly show that $\Phi_\delta(t,s) = C(t)(I-\frac{1}{t}\Gamma_\delta(t))\cdots C(s)(I-\frac{1}{s}\Gamma_\delta(s))$ satisfies almost surely $\|\Phi^\delta(t,s)\| \le \exp(-c_\delta\sum_{i=1}^s\frac{1}{i})$, where $c_\delta > 0$, provided $\delta < 1$. Furthermore, following the same line of thought as above, one comes to the basic relation which holds almost surely for $0 < \delta < \frac{1}{2}$

$$\lim_{t\to\infty}\|\tilde{R}_\delta(t)\| \le \lim_{t\to\infty}\sum_{s=0}^{t}([\frac{1}{j^{2(1-\delta)}}\exp(-c_\delta\sum_{i=1}^{t}\frac{1}{i})]^{\frac{1}{2}} = 0 \quad (29)$$

(see, for example, [23], Lemma 3.1.1). Thus, the result. ■

Following the above line of thought one can easily prove the following general result:

Theorem 2: Let Assumptions A1) - A3) be satisfied and let

A4) $\bigcup_{i=1}^N(S_u^{(i)}\times S_u^{(i)}) = \{1,\dots,m\}\times\{1,\dots,m\}$; $\bigcup_{i=1}^N(S_y^{(i)}\times S_u^{(i)}) = \{1,\dots,n\}\times\{1,\dots,m\}$.

Then, for all $i = 1,\dots,N$,

$$\|G^{(i)}(t) - G(\lambda)\| = o(t^{-\delta}), \|W^{(i)}(t) - W(\lambda)\| = o(t^{-\delta}), \quad (30)$$

almost surely, where $0 < \delta < \frac{1}{2}$.

The proof is obtained in a straightforward way by applying the procedure from Theorem 1 to all the estimated correlation coefficients, elements of matrices G and W. It is to be noted that the cross correlation functions between the outputs need not to be estimated, due to the adopted form of the polynomials $A(z)$ and $P(z)$ in (1) and (2), respectively. In such a way, the number of the output correlation functions needed for identifying the system is reduced at the expense of an increased model order.

This result enables formulating the final convergence result.

Theorem 3: Assume that Assumptions A1)– A4) hold. Then $\theta^{(i)}(t)$ given by (9) converges almost surely, for all $i = 1,\dots,N$, to a limit belonging to the solution set of the Yule-Walker equation (6).

Proof: The proof is based on the general convergence theorem for stochastic approximation algorithms with expanding truncations and, concretely, on the results presented in [18], [23]. Namely, equations (30) obtained as the result of the above analysis, enable choosing the function $M(\sigma^{(i)}(t-1))$ in (9) in such a way that $M(\sigma^{(i)}(t-1)) < t^\gamma$, where $0 < \gamma < \delta$. Following [18], one obtains

$$\sum_{t=1}^{\infty}\frac{1}{t}\|G-G^{(i)}(t)\|\|\theta^{(i)}(t-1)\| \le \sum_{t=1}^{\infty}\frac{1}{t^{1+\delta-\gamma}} < \infty, \quad (31)$$

which implies $\sum_{t=1}^\infty\frac{1}{t}n^{(i)}(t) < \infty$ (a.s.), where $n^{(i)}(t) = G^{(i)}(t)\,[W^{(i)}(t)^T - G^{(i)}(t))^T\theta^{(i)}(t-1)]+GG^T\theta^{(i)}(t-1)-GW^T$ is the equivalent noise term, $i = 1,\dots,N$. Hence the result. ■

REFERENCES

[1] D. D. Šiljak, *Decentralized Control of Complex Systems*. New York: Academic Press, 1991.

[2] S. S. Stanković and D. D. Šiljak, "Robust stabilization of nonlinear interconnected systems by decentralized dynamic output feedback," *Systems and Control Letters*, vol. 58, pp. 271–275, 2009.

[3] C. Canudas, Ed., *Proceedings of the 2nd IFAC Workshop on Estimation and Control on Network Systems, NecSys'10*, September 2010.

[4] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, pp. 803–812, 1986.

[5] A. Fax and R. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Trans. Autom. Control*, vol. 49, pp. 1465–1476, 2004.

[6] A. Jadbabaie, J. Lin, and A. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, pp. 988–1001, 2003.

[7] R. Olfati-Saber, A. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, pp. 215–233, 2007.

[8] W. Ren and R. Beard, "Consensus seeking in multi-agent systems using dynamically changing interaction topologies," *IEEE Trans. Autom. Control*, vol. 50, pp. 655–661, 2005.

[9] S. S. Stanković, M. S. Stanković, and D. M. Stipanović, "Consensus based overlapping decentralized estimator," *IEEE Trans. Autom. Control*, vol. 54, pp. 410–415, 2009.

[10] ——, "Consensus based overlapping decentralized estimation with missing observations and communication faults," *Automatica*, vol. 45, pp. 1397–1406, 2009.

[11] ——, "Decentralized parameter estimation by consensus based stochastic approximation," *IEEE Trans. Autom. Control*, vol. 56, no. 3, pp. 531–543, 2011.

[12] M. S. Stanković, S. S. Stanković, and D. M. Stipanović, "Consensus based multi-agent control structures," in *Proc. IEEE Conf. Decision and Control*, 2008.

[13] M. S. Stanković, "Control and estimation algorithms for multiple-agent systems," Ph.D. dissertation, University of Illinois, Urbana-Champaign, 2009.

[14] T. Söderström and P. Stoica, *System Identification*. Hemel Hempstaed, UK: Prentice Hall International, 1989.

[15] L. Ljung, *System Identification - theory for the user*. englewood Cliffs, N.J.: Prentice Hall International, 1989.

[16] T. Söderström, "Error-in-variables methods in system identification," *Automatica*, vol. 43, pp. 939–958, 2007.

[17] L. Ljung and T. Söderström, *Theory and pracice of recursive identification*. Cambridge, Mass.: MIT Press, 1983.

[18] H. F. Chen, "Recursive identification for errors-in-variables systems," *Automatica*, vol. 43, pp. 1234–1242, 2007.

[19] P. Stoica, "Generalized Yule-Walker equations and testing the orders of multivariate time-series," *Intern. J. on Control*, vol. 37, pp. 541–543, 1987.

[20] M. Masberg, Y. Irshad, and T. Söderström, "A covariance function based approach to networked control system identification," in *Proc. 2nd IFAC Conf. Necsys'10*, 2010, pp. 127–132.

[21] H. J. Kushner and G. Yin, "Stochastic approximation algorithms for parallel and distributed processing," *Stochastics*, vol. 22, pp. 219–250, 1987.

[22] R. A. Horn and C. A. Johnson, *Matrix Analysis*. Cambridge Univ. Press, 1985.

[23] H. F. Chen, *Stochastic approximation and its applications*. Dordrecht, the Netherlands: Kluwer Academic, 2002.