Distributed Change Detection Based on a Consensus Algorithm

Srdjan S. Stanković, Nemanja Ilić, Miloš S. Stanković, and Karl Henrik Johansson

***Abstract*—In this paper a novel distributed recursive algorithm is proposed for real time change detection using sensor networks. The algorithm is based on a combination of geometric moving average control charts generating local statistics and a global consensus strategy; it does not require any fusion center, so that the final decision is made by testing the state of any node in the network with respect to a given common threshold. The mean-square error with respect to the centralized solution defined by a weighted sum of the local statistics is analyzed in the case of constant asymmetric consensus matrices with constant and time varying forgetting factors in the underlying recursions, assuming spatially and temporally correlated data. These results are consistently extended to the case of time varying random consensus matrices, encompassing asymmetric gossip schemes, lossy networks and intermittent measurements, proving that the algorithm can be an efficient tool for practice. The given simulation results illustrate the main characteristics of the proposed algorithm, including the consensus matrix design, the mean square error with respect to the centralized solution as a function of the forgetting factor, the obtained detection quality expressed using deflection and estimation of the instant of parameter change.**

***Index Terms*—Consensus, distributed detection, geometric moving average control charts, real time change detection, sensor networks.**

I. Introduction

DISTRIBUTED sensor systems have received much attention recently, having in mind the low cost of miniature sensor technologies and their increased capacity to collect, analyze, and transmit environmental data. In a typical sensor system, dispersed wireless sensor nodes gather information about the properties or the occurrence of an event of interest, process this information locally and exchange the obtained results among themselves to fulfill a specific purpose. The abstract framework consisting of a collection of geographically dispersed sensor nodes along with a central entity aimed at decision making, termed *fusion center*, is commonly referred to as *distributed* or *decentralized detection* [1]–[4]. There are many levels in which the sensed data can be shared and processed among nodes, e.g., signal level, feature level and decision level. At each of these levels, the information content is, in principle, reduced, and this, in turn, reduces the required amount of data to be communicated between nodes. The decentralized nature of distributed systems is to be contrasted with a *centralized system*, in which the fusion center has access to the full collection of raw observations. The task of the fusion center is often reduced to a classical hypothesis testing problem where the information received from the sensor nodes is viewed as an *observation vector* [1], [2]. However, the detection systems based on fusion center are prone to failures of the fusion center itself and to communication bottlenecks that render it inoperative.

One of possible specific tasks of a distributed detection system can be to detect abrupt *changes* in the environment, requiring fast real time data processing [5]–[7]. In many applications it is desirable to eliminate the need for a fusion center, i.e., to have a possibility to make a global decision by testing the decision variables *in real time at any node* in a given sensor network.

Consensus techniques have been studied for many years, starting from the early 1980s, when important results were obtained in the areas of distributed asynchronous iterations in parallel computation and distributed optimization (e.g., [8]–[13]). There have been some recent attempts to apply consensus techniques to the distributed detection problem under the assumption that the dynamic agreement process starts *after all data had been collected*, implying inapplicability to real time change detection problems [14]–[16]. The problem of distributed detection based on a consensus scheme (called "diffusion") has been treated in [17] supposing a specific parametrization of the measurement model. In [18]–[20], algorithms for distributed state and parameter estimation have been proposed by combining local overlapping decentralized estimation schemes with a dynamic consensus algorithm. Analogous algorithms for distributed detection based on "running consensus" have been proposed and discussed in [21], [22], treating only the case of symmetric consensus gain matrices, like in [23]. An analysis of such algorithms based on the large deviations theory has been presented in [24].

In this paper a new algorithm is proposed for *distributed change detection* while monitoring the environment using a sensor network. It is assumed that all the nodes of the network can locally generate decision variables by recursive schemes belonging to the *geometric moving average control charts* [5], which allow fast tracking of signal properties. Assuming that each node sends its decision variable to a *selected small number of neighboring nodes* (in accordance with a given network topology defined in the form of a directed graph), a dynamic *consensus based scheme* is realized providing, under

Manuscript received March 01, 2011; revised July 02, 2011; accepted August 18, 2011. Date of publication September 15, 2011; date of current version November 16, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Biao Chen. A shorter version of this work was presented at the IFAC Workshop on Distributed Estimation and Control in Networked Systems 2010.

S. S. Stanković and N. Ilić are with the Faculty of Electrical Engineering, University of Belgrade, 11000 Belgrade, Serbia (e-mail: stankovic@etf.rs; nemiliexp@yahoo.com).

M. S. Stanković and K. H. Johansson are with the School of Electrical Engineering, Royal Institute of Technology, 100-44 Stockholm, Sweden (e-mail: milsta@kth.se; kallej@kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2168219

general conditions, nearly equal behavior of all the nodes. Consequently, *any node* in the network can be selected for testing its decision variable w.r.t. a prespecified *common threshold*, distributing the global decision throughout the network and eliminating the need for a fusion center. The algorithm construction follows methodologically [19], [25], and can be considered as a generalization of the algorithm discussed in [21], [22], [24]. The results presented in the paper have been partially presented in [26] and [27].

The theoretical analysis of the proposed algorithm is primarily concerned with the relation between the set of decision variables it generates and the *centralized reference decision variable* defined as a *weighted sum* of the decision variables obtained by the local recursions implemented independently (compare with [21] and [22]). It is assumed that these weights are selected by using any *a priori* given optimality criterion, as well as that the available data are both *spatially and temporally correlated*.

The cases of constant and randomly time varying communication gains (consensus matrices) are analyzed separately. For constant matrices, it is proved that the mean-square error between the decision variables generated by the proposed algorithm and the reference decision variable is bounded by $K(1-\alpha)^2$, where $K<\infty$ and $0<\alpha<1$ is the forgetting factor of the recursive algorithm generating the decision variables, which influences its memory length, and, thus, its tracking properties [5], [28]. In the design phase, for any *a priori* given set of weights in the reference centralized scheme, the consensus matrix is obtained by solving a linear programming problem. In the case of time varying forgetting factors tending to one when t tends to infinity, it is proved, by using the stochastic approximation arguments, that the mean square error converges to zero. In the case of randomly time varying consensus gains, which encompasses asymmetric gossip algorithms and random communication faults, an algorithm for the design of consensus matrices is presented. It is proved that the mean-square error is bounded by $K(1-\alpha)$ for constant forgetting factors, and that it converges to zero for time varying forgetting factors tending to one, but at a rate inferior to the one corresponding to the case of constant consensus matrices. The important case of *intermittent measurements* is also discussed (see [29]).

Simulation results are given as an illustration of the characteristic properties of the proposed change detection algorithm. They provide an insight into the consensus matrix design, mean square error as a function of the forgetting factor, detection quality expressed by using *deflection* [30], as well as the quality of estimation of the moment of parameter change.

The outline of the paper is as follows. In Section II the novel distributed change detection scheme based on a consensus algorithm is presented. In Section III the error analysis with respect to the centralized scheme is given, assuming constant consensus gains for constant and time varying forgetting factors of the local recursive schemes. Section IV treats the case of random consensus gains, while Section V deals with illustrative numerical examples.

In the paper, we use the following notation: $[\mathbf{x}]_i$ denotes the ith component of a vector $\mathbf{x}$, $\mathbf{A}^T$ is the transpose of a matrix $\mathbf{A}$, $\mathbf{1}=[1,\ldots,1]^T$, $||\mathbf{A}||$ denotes the spectral norm of a matrix $\mathbf{A}$, $||\mathbf{A}||_\infty$ denotes the infinity norm defined as $||\mathbf{A}||_\infty=\max_i\sum_j|a_{ij}|$, $\lambda_i(\mathbf{A})$ denotes the ith eigenvalue of a square matrix $\mathbf{A}$, $y=O(x)$ stands for $y\leq Kx$, where $K<\infty$, $y(t)=o(x(t))$ stands for $\lim_{t\to\infty}\frac{y(t)}{x(t)}=0$, $\mathbf{A}\succeq 0$ denotes a matrix with nonnegative elements, $\mathbf{A}\geq 0$ a positive-semidefinite matrix, $\mathrm{vec}\{\mathbf{A}\}$ denotes a vector obtained by concatenating the columns of a matrix $\mathbf{A}$, $\mathbf{A}\otimes\mathbf{B}$ denotes the Kronecker's product, $E\{.\}$ denotes the mathematical expectation, $E_H\{.\}$ the mathematical expectation under hypothesis $\mathbf{H}$ and $\mathrm{var}_H\{.\}$ variance under hypothesis $\mathbf{H}$.

II. Distributed Change Detection Algorithm

Consider a sensor network containing n nodes, where each node collects locally available measurements and generates at each discrete time instant t a scalar quantity $x_i(t)$, $i=1,\ldots,n$, as a result of local signal processing. We shall assume that the whole generated random vector $\mathbf{x}(t)=[x_1(t),\ldots,x_n(t)]^T$ reflects the environment in such a way that from $t=0$ until $t=t_0$ $E\{\mathbf{x}(t)\}=\boldsymbol{\theta}^0=\mathbf{0}$ (hypothesis $\mathbf{H_0}$), and from $t=t_0$ $E\{\mathbf{x}(t)\}=\boldsymbol{\theta}^1=[\theta_1^1,\ldots,\theta_n^1]^T$, where $\theta_i^1\geq 0$, $i=1,\ldots,n$, and $\theta_j^1>0$ for some $j\in\{1,\ldots,n\}$ (hypothesis $\mathbf{H_1}$). It will be also assumed that under each of the hypotheses its covariance satisfies $E\{(\mathbf{x}(t)-E\{\mathbf{x}(t)\})(\mathbf{x}(t+\tau)-E\{\mathbf{x}(t+\tau)\})^T\}=\boldsymbol{\rho}(\tau)$. The purpose of the whole network is the online detection of a change in the observed environment, manifested as a *change of the mean* of the vector $\mathbf{x}(t)$.

According to the paradigm of distributed detection, we shall assume first that each sensor is able to generate its *local decision function* $s_i^L(t)$ autonomously, using the methodology of *geometric moving average control charts* (see [5] for a detailed presentation in the context of change detection), i.e.,

$$s_i^L(t+1)=\alpha s_i^L(t)+(1-\alpha)x_i(t+1),\qquad s_i^L(0)=0 \quad (1)$$

where $0<\alpha<1$ is the *forgetting factor* of the algorithm, influencing its tracking properties. The effective memory length of the algorithm increases when α increases; the algorithm is faster for smaller values of α, but its noise immunity becomes deteriorated. Allowing α to be time varying, it is possible to achieve a better adaptivity to signal characteristics. When $\alpha(t)\to_{t\to\infty}1$, the algorithm essentially looses its tracking capabilities, but allows convergence of $s_i^L(t)$ to a steady state value in a certain sense (see the discussion below, where this case is considered separately, and, also, [28] for a general treatment of the problem). In a completely distributed system, consisting of n independent recursions (1), the local change detection procedure is based on testing $s_i^L(t)$ with respect to an appropriately chosen threshold $d_i^L>0$, so that a change is detected when $s_i^L(t)$ exceeds d_i^L, $i=1,\ldots,n$ (see, e.g., [5] and [31]). Obviously, the detection time is, in general, delayed with respect to the real instant of change. This delay depends on the tracking capabilities of (1), and can be analyzed using the methodology from, e.g., [5].

Instead of having n separate decisions, a global decision (related to the observed phenomenon as a whole) can be done by introducing a *fusion center*, where the *centralized decision function* $s_c(t)$ is formed as a weighted sum of the local decision functions, i.e., $s_c(t)=\mathbf{w}^T\mathbf{s}^L(t)$, where $\mathbf{w}=[w_1,\ldots,w_n]^T$, $w_i\geq 0$, $i=1,\ldots,n$, is the *weight vector*, satisfying, for convenience, the condition $\sum_{i=1}^n w_i=1$, while $\mathbf{s}^L(t)=[s_1^L(t),\ldots,s_n^L(t)]^T$. The detection procedure is now based on testing $s_c(t)$ with respect to an appropriately chosen

threshold $d_c > 0$. Notice that the resulting system still belongs, according to [1] and [2], to the class of distributed detection systems, having in mind that the calculation of $x_i(t)$ and $s_i^L(t)$ is distributed. The choice of the weight vector $\mathbf{w}$ and the threshold d_c can be based on different *a priori* selected criteria; this subject is out of the scope of this paper. The centralized decision function can be written in the form of a recursion

$$s_c(t+1) = \alpha s_c(t) + (1-\alpha)\mathbf{w}^T\mathbf{x}(t+1), \quad s_c(0) = 0. \quad (2)$$

Remark 1: The algorithm (2) with a fusion center can be considered as a representative of a large class of online change detection procedures, which can be obtained for different concrete choices of the vectors $\mathbf{x}(t)$ and $\mathbf{w}$. As an illustration, consider the measurement model

$$\boldsymbol{\chi}(t) = \boldsymbol{\mu} + \boldsymbol{\epsilon}(t) \quad (3)$$

where $\boldsymbol{\mu}$ is a constant n-vector and $\{\boldsymbol{\epsilon}(t)\}$ an i.i.d. sequence with $\boldsymbol{\epsilon}(t) \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a positive-definite matrix. Let the problem be to detect the change from $\boldsymbol{\mu} = \boldsymbol{\mu}^0 = \mathbf{0}$ (hypothesis $\mathbf{H_0}$) to $\boldsymbol{\mu} = \boldsymbol{\mu}^1 \neq \mathbf{0}$ (hypothesis $\mathbf{H_1}$). According to the general detection principles (see, e.g., [5], [24], and [31]), we can calculate the log likelihood ratio for the set containing $\boldsymbol{\chi}(t),\ t = 1, \ldots, N_1$, and obtain the following expression:

$$L(N_1) = \sum_{t=1}^{N_1} \log \frac{p_{\boldsymbol{\mu}^1}(\boldsymbol{\chi}(t))}{p_{\boldsymbol{\mu}^0}(\boldsymbol{\chi}(t))} = \boldsymbol{\mu}^{1T}\boldsymbol{\Sigma}^{-1} \sum_{t=1}^{N_1} \left(\boldsymbol{\chi}(t) - \frac{1}{2}\boldsymbol{\mu}^1\right). \quad (4)$$

A geometric moving average control chart for online detection can be derived from (4) following [5] and [31]. It can be put in the form (2), with

$$\mathbf{w}^T = \kappa^{-1}[|[\boldsymbol{\mu}^{1T}\boldsymbol{\Sigma}^{-1}]_1|, \ldots, |[\boldsymbol{\mu}^{1T}\boldsymbol{\Sigma}^{-1}]_n|]^T,$$
$$\mathbf{x}(t) = [\chi_1(t)\mathrm{sgn}([\boldsymbol{\mu}^{1T}\boldsymbol{\Sigma}^{-1}]_1), \ldots, \chi_n(t)\mathrm{sgn}([\boldsymbol{\mu}^{1T}\boldsymbol{\Sigma}^{-1}]_n)]^T$$

and $\kappa = \sum_{i=1}^n |[\boldsymbol{\mu}^{1T}\boldsymbol{\Sigma}^{-1}]_i|$. Obviously, under $\mathbf{H_1}$, $E\{\mathbf{x}(t)\} = \boldsymbol{\theta}^1 = [\mu_1^1 \mathrm{sgn}([\boldsymbol{\mu}^{1T}\boldsymbol{\Sigma}^{-1}]_1), \ \ldots, \mu_n^1 \ \mathrm{sgn}([\boldsymbol{\mu}^{1T} \ \boldsymbol{\Sigma}^{-1}]_n)]^T$ with nonnegative components. The obtained recursion differs from the one derived in, e.g., [31], by the introduction of the normalization factor κ. The generated decision variable is tested w.r.t. the threshold defined as $d_c = \frac{1}{2}\boldsymbol{\mu}^{1T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^1\kappa^{-1} = \frac{1}{2}\boldsymbol{\theta}^{1T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}^1\kappa^{-1}$.

If we assume that $\boldsymbol{\mu} = \mathbf{0}$ in (3) and that $\boldsymbol{\Sigma} = \mathrm{diag}\{\sigma_1, \ldots, \sigma_n\}$, where σ_i changes abruptly at $t = t_0$ from one constant value to the other, it is possible to make an algorithm for detecting the change in variance in the form (2) by generating locally $x_i(t) = \chi_i(t)^2$.

Supposing that the parameter jump is unknown, the application of the generalized likelihood ratio methodology to the same measurement model leads also to a recursion belonging to the general form (2), with specific choices of the vectors $\mathbf{x}(t)$ and $\mathbf{w}$ for detecting both the mean and the variance changes in (3) (see [31] and [32]). ■

The aim of this paper is to propose an online *distributed change detection* algorithm which *does not require any fusion center*, and in which the output of any preselected node can be used as a representative of the whole network and be tested w.r.t. one prespecified *common threshold*. The basic assumption for such an algorithm is that each node of the network communicates with its neighborhood in accordance with an $n \times n$ time varying *consensus matrix* $\mathbf{C}(t) = [c_{ij}(t)]$, satisfying $c_{ij}(t) \geq 0$, $i \neq j$ and $c_{ii}(t) > 0,\ i, j = 1, \ldots, n$, which formally represents the weighted adjacency matrix for the underlying time varying graph representing the network ($c_{ij}(t)$ is the communication gain from node j to node i). We shall assume, additionally, that $\mathbf{C}(t)$ is *row stochastic* for all t [33]. Consequently, we propose in this paper the following distributed algorithm for generating decision functions at all the nodes of the network:

$$\bar{s}_i(t+1) = \alpha s_i(t) + (1-\alpha)x_i(t+1), \quad s_i(0) = 0,$$
$$s_i(t+1) = \sum_{j=1}^n c_{ij}(t)\bar{s}_j(t+1); \quad (5)$$

$i = 1, \ldots, n$. Notice that the set of neighbors of the ith node (containing those indexes j for which $c_{ij}(t) > 0$) can be small, and the elements $c_{ij}(t)$ themselves random (see the analysis below). Denoting $\mathbf{s}(t) = [s_1(t), \ldots, s_n(t)]^T$, we have the following compact representation of the proposed algorithm

$$\mathbf{s}(t+1) = \alpha\mathbf{C}(t)\mathbf{s}(t) + (1-\alpha)\mathbf{C}(t)\mathbf{x}(t+1), \quad \mathbf{s}(0) = 0. \quad (6)$$

The algorithm (5), (6) is derived from the *consensus* based state and parameter estimation algorithms proposed in [18]–[20]; it is also similar to the "running consensus" detection algorithm based on time averaging proposed in [21], [22], and [24]. The consensus matrix $\mathbf{C}(t)$ in (6) performs "convexification" of the neighboring states for each node and, therefore, aims at enforcing (under appropriate conditions) consensus between all the nodes in the network. The basic underlying idea is to achieve $s_i(t) \approx s_j(t),\ i, j = 1, \ldots, n,\ i \neq j$, so that change detection can be done by testing any state $s_i(t),\ i = 1, \ldots, n$, with respect to a given *common threshold*, thus eliminating the need for a fusion center. This threshold can be chosen to be equal to the threshold d_c for the centralized strategy (2), provided (6) gives a good approximation of $s_c(t)$ in (2). Looking from another standpoint, it is to be expected that the introduction of the consensus scheme in a completely decentralized structure consisting of n independent local detectors (1) has as a consequence an increase in the overall detection quality, having in mind ensemble averaging and "denoising" intrinsic to the consensus based estimation schemes (see, e.g., [18], [19], and [25]). One of the important consequences is that the algorithm is highly robust with respect to variations of local noise characteristics. Denoising aspects will be covered in more detail in Section V.

The theoretical part of the paper, contained in Sections III and IV, will be focused on the relationship between the proposed algorithm and the centralized decision strategy (2) taken as a reference. More specifically, we shall analyze the distance between the states of the proposed algorithm (6) and the centralized scheme (2) as a function of the forgetting factor α. A more detailed analysis of the estimation of change time, false alarm and detection probabilities, etc. are out of the scope of the paper. These issues are addressed through the experimental results in Section V.

As a prerequisite for the further analysis, we define the error between the states of (6) and (2) as

$$\mathbf{e}(t) = \mathbf{s}(t) - \mathbf{1}s_c(t). \quad (7)$$

Iterating (6) and (2) back to the zero initial conditions, we get

$$\mathbf{s}(t) = (1-\alpha)\sum_{i=0}^{t-1}\alpha^i \boldsymbol{\varphi}(t-1, t-i-1)\mathbf{x}(t-i) \quad (8)$$

where $\boldsymbol{\varphi}(i,j) = \mathbf{C}(i)\cdots\mathbf{C}(j)$, $i \geq j$, and

$$s_c(t) = (1-\alpha)\sum_{i=0}^{t-1}\alpha^i \mathbf{w}^T\mathbf{x}(t-i), \quad (9)$$

from which we obtain

$$\mathbf{e}(t) = (1-\alpha)\sum_{i=0}^{t-1}\alpha^i[\boldsymbol{\varphi}(t-1,t-i-1) - \mathbf{1}\mathbf{w}^T]\mathbf{x}(t-i) \quad (10)$$

(compare with [21] and [22], where the special case when $\mathbf{w} = \frac{\mathbf{1}}{n}$ is treated, in conjunction with symmetric consensus matrices).

III. Error Analysis: Constant Consensus Matrices

When $\mathbf{C}(t) = \mathbf{C}$, we start from the following assumptions:

A1) $\mathbf{C}$ has the eigenvalue 1 with algebraic multiplicity 1;

A2) $\lim_{i\to\infty}\mathbf{C}^i = \mathbf{1}\mathbf{w}^T$.

The first assumption is related to the topology of the underlying multi-agent network, implying that the directed graph associated with $\mathbf{C}$ has a spanning tree and that $\mathbf{C}^i$ converges to a nonnegative row stochastic matrix with equal rows when i tends to infinity, e.g., [9], [13]. Assumption A2) establishes a formal connection between the algorithm (6) and the reference scheme (2). For an *a priori* given $\mathbf{C}$, $\mathbf{w}$ follows from the relation

$$\mathbf{w}^T\mathbf{C} = \mathbf{w}^T, \quad (11)$$

known from the theory of stationary Markov chains, e.g., [34]. For an *a priori* given $\mathbf{w}$, the consensus matrix $\mathbf{C}$ can be obtained by solving the linear programming problem associated with (11) under communication structure constraints based on setting preselected elements of $\mathbf{C}$ to zero (indication that there are no communication links between the corresponding nodes). An illustrative example is given in Section V.

A. Constant Forgetting Factor

When both α and $\mathbf{C}$ in (2) and (6) are constant, one obtains using A2) that $s_c(t) = \mathbf{w}^T\mathbf{s}(t)$, and, therefore, that $\mathbf{e}(t) = (\mathbf{I} - \mathbf{1}\mathbf{w}^T)\mathbf{s}(t)$. Also, in this case $\boldsymbol{\varphi}(i,j) = \mathbf{C}^{i-j+1}$, so that (10) and (8) give

$$\mathbf{e}(t) = (1-\alpha)\sum_{i=0}^{t-1}\alpha^i\tilde{\mathbf{C}}^{i+1}\mathbf{x}(t-i) \quad (12)$$

where $\tilde{\mathbf{C}} = \mathbf{C} - \mathbf{1}\mathbf{w}^T$, having in mind that, under A2), we have $(\mathbf{C} - \mathbf{1}\mathbf{w}^T)^i = \mathbf{C}^i - \mathbf{1}\mathbf{w}^T$.

We first realize that $\mathbf{s}(t)$ as an estimator of $\mathbf{1}s_c(t)$ is unbiased under $\mathbf{H_0}$ and, in general, biased under $\mathbf{H_1}$, since we have from (12) that

$$\mathbf{m}_e(t) = E\{\mathbf{e}(t)\} = (1-\alpha)\sum_{i=0}^{t-1}\alpha^i\tilde{\mathbf{C}}^{i+1}\boldsymbol{\theta}^1; \quad (13)$$

notice that $E\{\mathbf{e}(t)\} = 0$ when $\theta_i^1 = \theta_j^1$, $i,j = 1,\ldots,n$, having in mind that $\tilde{\mathbf{C}}\boldsymbol{\theta}^1 = \mathbf{0}$ for $\boldsymbol{\theta}^1 = \mu'\mathbf{1}$, where μ' is a given scalar. The bias is, obviously, smaller when α is closer to one. Namely, in the steady state, we have $\lim_{t\to\infty}E\{\mathbf{s}(t)\} = (1-\alpha)(\mathbf{I} - \alpha\mathbf{C})^{-1}\mathbf{C}\boldsymbol{\theta}^1$, and, consequently,

$$\lim_{t\to\infty}E\{\mathbf{s}(t)\} \approx [(1-\alpha)(\mathbf{I} + \alpha\mathbf{C} + \cdots + \alpha^\nu\mathbf{C}^\nu)\mathbf{C} + \alpha^{\nu+1}\mathbf{1}\mathbf{w}^T]\boldsymbol{\theta}^1 \quad (14)$$

for some ν large enough. For α close to 1, the first term in the brackets is obviously small, and can be neglected, and the second term is approximately equal to $\lim_{t\to\infty}E\{s_c(t)\} = \mathbf{w}^T\boldsymbol{\theta}^1$.

The focus of the analysis is placed on the mean square error matrix $\mathbf{Q}(t) = E\{\mathbf{e}(t)\mathbf{e}(t)^T\}$. Using (12) and (13), one readily obtains

$$\mathbf{Q}(t) = (1-\alpha)^2\boldsymbol{\Phi}(t)^T\mathbf{R}(t)\boldsymbol{\Phi}(t) \quad (15)$$

where $\boldsymbol{\Phi}(t) = [\alpha^{t-1}\tilde{\mathbf{C}}^t\vdots\alpha^{t-2}\tilde{\mathbf{C}}^{t-1}\vdots\cdots\vdots\alpha^0\tilde{\mathbf{C}}]^T$, $\mathbf{R}(t) = E\{\mathbf{X}(t)\mathbf{X}(t)^T\}$ and $\mathbf{X}(t) = [\mathbf{x}(1)^T\cdots\mathbf{x}(t)^T]^T$.

Theorem 1: Let assumptions A1) and A2) hold, together with:

A3) $\sum_{\tau=0}^{\infty}\|\boldsymbol{\rho}(\tau)\| \leq K < \infty$,

where $\boldsymbol{\rho}(\tau) = E\{(\mathbf{x}(t) - E\{\mathbf{x}(t)\})(\mathbf{x}(t+\tau) - E\{\mathbf{x}(t+\tau)\})^T\}$. Then, under both $\mathbf{H_0}$ and $\mathbf{H_1}$,

$$\max_{i,j}Q_{ij}(t) = O((1-\alpha)^2) \quad (16)$$

for all t, where $Q_{ij}(t)$ are the elements of $\mathbf{Q}(t)$ in (15).

Proof: Consider an arbitrary deterministic n-dimensional vector $\mathbf{y}$ and analyze the quadratic form $\mathbf{y}^T\mathbf{Q}(t)\mathbf{y} = \mathbf{y}^T[\mathbf{Q}_1(t) + \mathbf{Q}_2(t)]\mathbf{y}$, in which

$$\mathbf{Q}_1(t) = (1-\alpha)^2\boldsymbol{\Phi}(t)^T\tilde{\mathbf{R}}(t)\boldsymbol{\Phi}(t) \quad (17)$$

and

$$\mathbf{Q}_2(t) = (1-\alpha)^2\boldsymbol{\Phi}(t)^T\mathbf{m}_X(t)\mathbf{m}_X(t)^T\boldsymbol{\Phi}(t) \quad (18)$$

where $\tilde{\mathbf{R}}(t) = \mathbf{R}(t) - \mathbf{m}_X(t)\mathbf{m}_X(t)^T$ and $\mathbf{m}_X(t) = E\{\mathbf{X}(t)\}$.

Starting the analysis with $\mathbf{y}^T\mathbf{Q}_1(t)\mathbf{y}$, we conclude that

$$\tilde{\mathbf{R}}(t) = \begin{bmatrix} \boldsymbol{\rho}(0) & \boldsymbol{\rho}(1) & \cdots & \boldsymbol{\rho}(t-1) \\ \boldsymbol{\rho}(1)^T & \boldsymbol{\rho}(0) & \cdots & \boldsymbol{\rho}(t-2) \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\rho}(t-1)^T & \boldsymbol{\rho}(t-2)^T & \cdots & \boldsymbol{\rho}(0) \end{bmatrix} \quad (19)$$

where $\boldsymbol{\rho}(\tau)$, $\tau = 0,\ldots,t-1$, are $n\times n$ block matrices reflecting spatial and temporal correlations.

In order to analyze $\tilde{\mathbf{R}}(t)$ in this general case, we introduce a special matrix norm in the following way. Let $\mathbf{A} = [\mathbf{A}_{ij}]$, $i,j = 1,\ldots,t$, be a $tn\times tn$ matrix composed of $n\times n$ blocks $\mathbf{A}_{ij}$. We define a norm $\|\mathbf{A}\|_*$ as $\|\mathbf{A}\|_* = \|\mathbf{A}_{ij}^*\|_\infty$, where $\mathbf{A}_{ij}^* = [\|\mathbf{A}_{ij}\|]$ is an $t\times t$ matrix composed of the spectral norms $\|\mathbf{A}_{ij}\|$ of the blocks [33]. The function $\|\mathbf{A}\|_*$ of a given matrix $\mathbf{A}$ represents indeed a matrix norm, having in mind that we have, according to the Conlisk observation presented in [35], the required property that in the case of the infinity norm $\|\mathbf{A}\|_\infty \leq \|\mathbf{B}\|_\infty$ for two $n\times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ satisfying $\mathbf{A} \succeq 0$, $\mathbf{B} \succeq 0$ and $\mathbf{A} \preceq \mathbf{B}$ (see [35] for more details).

Consequently, we construct a $t\times t$ matrix $\tilde{\mathbf{R}}(t)^*=[\tilde{R}_{ij}(t)^*]$, with the elements $\tilde{R}_{ij}(t)^*=\|\boldsymbol{\rho}(i-j)\|$, $i,j=1,\ldots,t$, so that $\|\tilde{\mathbf{R}}(t)\|_*=\|\tilde{\mathbf{R}}(t)^*\|_\infty$. By Assumption A3) we have now that

$$\lim_{t\to\infty}\max_i\{\lambda_i(\tilde{\mathbf{R}}(t))\}\leq\lim_{t\to\infty}\|\tilde{\mathbf{R}}(t)\|_*\leq K<\infty. \tag{20}$$

Coming back to (17), we realize further that the expression $\mathbf{y}^T\boldsymbol{\Phi}(t)^T\boldsymbol{\Phi}(t)\mathbf{y}$ is in the form of a sum of terms containing $\mathbf{y}^T\tilde{\mathbf{C}}^i\tilde{\mathbf{C}}^{iT}\mathbf{y}$, $i=1,\ldots,t$. By assumptions A1) and A2) it follows that $\mathbf{C}$ and $\mathbf{1}\mathbf{w}^T$ have the same eigenvectors and that $\mathbf{C}$ has the same eigenvalues as $\tilde{\mathbf{C}}$, except for the eigenvalue 1 of $\mathbf{C}$, which is replaced by the eigenvalue 0 of $\tilde{\mathbf{C}}$. Having in mind that $c_{ii}>0$, $i=1,\ldots,n$, by assumption, it follows that the modules of all the eigenvalues of $\tilde{\mathbf{C}}$ are strictly less than one [13]. Therefore, we have $\|\mathbf{y}^T\tilde{\mathbf{C}}^i\tilde{\mathbf{C}}^{iT}\mathbf{y}\|\leq k'\lambda_M^{2i}\|\mathbf{y}\|^2$, where $k'<\infty$, $i=1,2,\ldots$ and $\lambda_M=\max_i\{|\lambda_i(\tilde{\mathbf{C}})|\}<1$, so that

$$\begin{aligned}\mathbf{y}^T\boldsymbol{\Phi}(t)^T\tilde{\mathbf{R}}(t)\boldsymbol{\Phi}(t)\mathbf{y}&\leq k''K\sum_{i=0}^{\infty}\alpha^{2i}\lambda_M^{2(i+1)}\|\mathbf{y}\|^2\\&\leq k''K\frac{\lambda_M^2}{1-\lambda_M^2}\|\mathbf{y}\|^2\\&\leq K_1\|\mathbf{y}\|^2\end{aligned} \tag{21}$$

where $k'',K_1<\infty$ do not depend on α.

Analyzing $\mathbf{Q}_2(t)$ in (18) we find that $\mathbf{Q}_2(t)=0$ under $\mathbf{H_0}$ and

$$\|\boldsymbol{\Phi}(t)^T\mathbf{m}_X(t)\|^2\leq(\sum_{i=0}^{\infty}\alpha^i\|\tilde{\mathbf{C}}^{i+1}\|\|\boldsymbol{\theta}^1\|)^2\leq K_2<\infty \tag{22}$$

under $\mathbf{H_1}$.

Consequently, by choosing $\mathbf{y}=\mathbf{u}_i$, where $\mathbf{u}_i$ denotes the n-vector of zeros with only the ith entry equal to one, one obtains that $Q_{ii}(t)=O((1-\alpha)^2)$ for all t, $i=1,\ldots,n$. Furthermore, $|Q_{ij}(t)|\leq\max_i Q_{ii}(t)$ for $i\neq j$, having in mind elementary properties of positive-semidefinite matrices. Thus, the result. ∎

The meaning of the obtained result becomes clearer after realizing that

$$E\{s_c(t)^2\}\leq K(1-\alpha)^2\sum_{i=0}^{\infty}\alpha^{2i}=O(1-\alpha) \tag{23}$$

under $\mathbf{H_0}$ and $E\{s_c(t)^2\}=O(1)$ under $\mathbf{H_1}$, having in mind that $E\{(\sum_{i=0}^{\infty}\alpha^i\mathbf{w}^T\boldsymbol{\theta}^1)^2\}=O((1-\alpha)^{-2})$.

Remark 2: The above given analysis shows that the mean square error between the decision functions generated by the proposed algorithm and the centralized decision function tends to zero when α approaches one. However, this does not imply that high values of α are always the most suitable, as such values of α do not allow fast change detection. In practice, a careful trade off has to be done. Illustrative examples obtained by simulation, including an analysis of the change detection delay, are presented in Section V. ∎

B. Variable Forgetting Factor

Assuming that α is time varying, tending to 1 when t tends to infinity, we obtain from (2) and (6) recursions performing essentially time averaging. They are, therefore, not directly suitable for change detection (see [21] and [22]). They can, instead, be used for testing if the hypotheses $\mathbf{H_0}$ and $\mathbf{H_1}$ hold for all values of t.

Theorem 2: Let in (2) and (6) the forgetting factor be in the form $\alpha(t+1)=1-\gamma(t+1)$, and let the assumptions A1), A2) and A3) be satisfied, together with:

A4) $\gamma(t)$ is a nonincreasing sequence satisfying $\gamma(t)>0$, $\lim_{t\to\infty}\gamma(t)=0$; $\sum_{t=1}^{\infty}\gamma(t)=\infty$.

Then, for both $\mathbf{H_0}$ and $\mathbf{H_1}$, $\|\mathbf{Q}(t)\|=o(1)$.

Proof: Starting from (6) and (8) one obtains (10). Consequently, in the case of time varying gains,

$$\mathbf{e}(t)=\sum_{i=0}^{t-1}\pi(t,t+1-i)\tilde{\mathbf{C}}^{i+1}\gamma(t-i)\mathbf{x}(t-i), \tag{24}$$

where $\pi(i,j)=1$ for $i<j$ and $\pi(i,j)=\alpha(i)\cdots\alpha(j)$ for $i\geq j$.

Similarly as in (17),

$$\mathbf{y}^T\mathbf{Q}_1(t)\mathbf{y}=\mathbf{y}^T\boldsymbol{\Psi}(t)^T\tilde{\mathbf{R}}(t)\boldsymbol{\Psi}(t)\mathbf{y} \tag{25}$$

where $\boldsymbol{\Psi}(t)=[\pi(t,2)\tilde{\mathbf{C}}^t\gamma(1)\vdots\pi(t,3)\tilde{\mathbf{C}}^{t-1}\gamma(2)\vdots\cdots\vdots\tilde{\mathbf{C}}\gamma(t)]$. Proceeding like in the proof of Theorem 1, we obtain

$$\mathbf{y}^T\mathbf{Q}_1(t)\mathbf{y}\leq k''K\sum_{i=0}^{t-1}\pi(t,t+1-i)^2\lambda_M^{2(i+1)}\gamma(t-i)^2\|\mathbf{y}\|^2. \tag{26}$$

Using standard results from the theory of stochastic approximation, we conclude that $\pi(t,j)\leq c_0\exp\{-c_1\sum_{i=j}^{t}\gamma(i)\}$, where c_0 and c_1 are positive-constants [36]. Therefore, it is possible to apply the Kronecker's lemma (see, e.g., [36]), and to conclude using A4) that

$$\lim_{t\to\infty}\sum_{i=0}^{t}\pi(t,t+1-i)^2\lambda_M^{2(i+1)}\gamma(t-i)^2=0. \tag{27}$$

An analogous reasoning can be applied to the term $\mathbf{y}^T\mathbf{Q}_2(t)\mathbf{y}$ in (18). ∎

The result of Theorem 2 can be applied, obviously, to the special case when $\gamma(t)=\frac{1}{t}$, which was treated in [21], [22], under the additional assumption that $w_i=\frac{1}{n}$, $i=1,\ldots,n$, that the consensus matrix is symmetric and that the sequences $\{x_i(t)\}, i=1,\ldots,n$, are mutually independent and i.i.d..

Corollary 1: Under the assumptions of Theorem 2 and with $\gamma(t)=\frac{1}{t}$ we have $\|\mathbf{Q}(t)\|=O(t^{-2})$.

Proof: For $\gamma(t)=\frac{1}{t}$ we have from (24) that

$$\mathbf{e}(t)=\frac{1}{t}\sum_{i=0}^{t-1}\tilde{\mathbf{C}}^{i+1}\mathbf{x}(t-i), \tag{28}$$

and the result immediately follows after applying Theorems 1 and 2. ∎

Notice that

$$s_c(t)=\frac{1}{t}\sum_{i=0}^{t-1}\mathbf{w}^T\mathbf{x}(t-i), \tag{29}$$

so that $E\{s_c(t)\}=0$ for $\mathbf{H_0}$ and $E\{s_c(t)\}=O(1)$ for $\mathbf{H_1}$.

Notice that, in general, both algorithms (2) and (6) can be considered as stochastic approximation algorithms, e.g., [36], so that the corresponding results from the literature may be applied. Stochastic approximation algorithms with consensus, representing a generalization of (6) to the regression problem, have

been analyzed in [20] and [25], starting from the basic results presented in [8].

Remark 3: The proposed distributed detection scheme can efficiently work also in the cases when some nodes do not have access to measurements, i.e., when $x_i(t) \equiv 0$ for some indexes i. Let $\mathbf{\Gamma}$ be a diagonal matrix containing at the diagonal 0's for the indexes for which $x_i(t) \equiv 0$ and 1's for the remaining ones. Then, in (6) we have $\mathbf{\Gamma x}(t)$ instead of $\mathbf{x}(t)$, and the assumption A2) becomes consequently reformulated as $(\lim_{i\to\infty} \mathbf{C}^i - \mathbf{1w}^T)\mathbf{\Gamma} = 0$. In the case when A2) initially holds, i.e., when the ideal centralized scheme is still taken as a reference, we have an increase of bias under hypothesis $\mathbf{H_1}$ due to nonavailability of measurements $\Delta\mathbf{m}_e$ satisfying $\|\Delta\mathbf{m}_e\|_\infty \leq (1-\alpha)\sum_{i=1}^{\infty} \alpha^i \|\mathbf{C}^{i+1}\|_\infty \|(\mathbf{\Gamma} - \mathbf{I})\|_\infty \|\boldsymbol{\theta}^1\|_\infty = \|\boldsymbol{\theta}^1\|_\infty$. ■

IV. Error Analysis: Random Consensus Matrices

The results from the previous section will be generalized here to time varying random consensus matrices. This case is of substantial importance from the point of view of applications of the proposed algorithm in real sensor networks, since it can often appear to be too restrictive and energy consuming to implement all possible communications between the nodes simultaneously at all discrete time instants. It is to be emphasized that the presented results will cover the randomized gossip algorithms connected to *asymmetric consensus matrices*, including the case of one directed communication at a time (only the case of symmetric consensus matrices requiring pairwise communications has been treated in the literature, e.g., [21]–[23]), as well the influence of communication outages.

We shall assume in this section that the sequence $\{\mathbf{C}(t)\}$, $t = 0, 1, \ldots$ is a sequence of i.i.d. random matrices, independent of the sequence $\{\mathbf{x}(t)\}$, such that $\mathbf{C}(t)$ is realized at each discrete time instant t as $\mathbf{C}^{(k)}$ with probability $p_k,\ k = 1, \ldots, N$, $N < \infty$, $\sum_{k=1}^{N} p_k = 1$, so that we have

$$E\{\mathbf{C}(t)\} = \bar{\mathbf{C}} = \sum_{k=1}^{N} \mathbf{C}^{(k)} p_k. \tag{30}$$

The realization matrices $\mathbf{C}^{(k)} = [c_{ij}^{(k)}],\ k = 1, \ldots, N,\ i, j = 1, \ldots, n$, will be assumed to be constant nonnegative row stochastic matrices, satisfying $c_{ii}^{(k)} > 0,\ i = 1, \ldots, n$.

This setting obviously encompasses the asynchronous asymmetric gossip algorithm with one message at a time: if the node j communicates to the node i, the corresponding realization has the form $\mathbf{C}^{(k)} = \mathbf{I} + \mathbf{C}^{[i,j]}$, where $\mathbf{C}^{[i,j]} = [c_{kl}^{[i,j]}],\ k, l = 1, \ldots, n$, is an $n \times n$ matrix which contains zeros everywhere except the (i,j)th element, where it contains γ_{ij} and the (i,i)th place where it contains $-\gamma_{ij}, 0 < \gamma_{ij} < 1$. Various types of synchronous asymmetric gossip algorithms can also be represented in this way by constructing the corresponding realizations $\mathbf{C}^{(k)}$ containing more nonzero off-diagonal elements located at appropriate places. Communication faults can obviously be modelled analogously, by forming realizations $\mathbf{C}^{(k)}$ in accordance with the faults (see, e.g., [19]). We are here not concerned with concrete protocols for generating realizations $\mathbf{C}^{(k)}$: our convergence analysis is applicable to any preselected technical setting satisfying the adopted general model.

We shall analyze algorithm (6) starting from the following assumptions:

B1) $\bar{\mathbf{C}}$ has the eigenvalue 1 with algebraic multiplicity 1;

B2) $\lim_{i\to\infty} \bar{\mathbf{C}}^i = \mathbf{1w}^T$.

Assumptions B1) and B2) are analogous to assumptions A1) and A2): $\mathbf{C}$ in A1) and A2) is now replaced by $\bar{\mathbf{C}}$. Assumption B1) deals with the communication structure constraints, while assumption B2) implies that the realization matrices $\mathbf{C}^{(k)}$ of $\mathbf{C}(t)$, the corresponding probabilities p_k and the weight vector $\mathbf{w}$ of (2) satisfy the relation

$$\mathbf{w}^T \bar{\mathbf{C}} = \mathbf{w}^T \sum_{k=1}^{N} \mathbf{C}^{(k)} p_k = \mathbf{w}^T. \tag{31}$$

For the given sets $\mathbf{C}^{(k)}$ and $p_k,\ k = 1, \ldots, n$, (31) can be solved for $\mathbf{w}$ in the same way as (11). When $\mathbf{w}$ is *a priori* given, (31) becomes nonlinear in the unknown variables $\mathbf{C}^{(k)}$ and p_k. Having now more degrees of freedom than in (11), a practical procedure for solving (31) can consist of adopting one set of parameters (probabilities p_k, for example) and solving the linear programming problem for the remaining set of parameters (parameters in $\mathbf{C}^{(k)}$), or *vice versa* [37]. Notice that in the case of the asynchronous randomized gossip algorithm with one communication at a time, $\mathbf{C}^{(k)}$ is characterized by only one scalar parameter; in general, $\mathbf{C}^{(k)}$ is characterized by more parameters satisfying the given constraints (see Section V for an example).

A. Constant Forgetting Factor

For the mean of $\mathbf{e}(t)$, we obtain now directly from (10) that

$$\mathbf{m}_e(t) = E\{\mathbf{e}(t)\} = (1-\alpha)\sum_{i=0}^{t-1} \alpha^i (\bar{\mathbf{C}}^{i+1} - \mathbf{1w}^T)\mathbf{m} \tag{32}$$

where $\mathbf{m} = \mathrm{E}\{\mathbf{x}(t)\}$, which is analogous to (13).

The mean square error matrix $\mathbf{Q}(t) = E\{\mathbf{e}(t)\mathbf{e}(t)^T\}$ is decomposed as

$$\mathbf{Q}(t) = \mathbf{Q}_3(t) + \mathbf{Q}_4(t) \tag{33}$$

where

$$\mathbf{Q}_3(t) = E\{E_x\{\mathbf{e}(t)\mathbf{e}(t)^T\} - E_x\{\mathbf{e}(t)\}E_x\{\mathbf{e}(t)\}^T\} \tag{34}$$

and

$$\mathbf{Q}_4(t) = E\{E_x\{\mathbf{e}(t)\}E_x\{\mathbf{e}(t)\}^T\}\}, \tag{35}$$

with $E_x\{.\}$ denoting the conditional expectation given the σ-algebra generated by $\mathbf{C}(0), \ldots, \mathbf{C}(t-1)$.

We obtain, in analogy with (15), that

$$\mathbf{Q}_3(t) = (1-\alpha)^2 E\{\tilde{\mathbf{\Phi}}(t)^T \tilde{\mathbf{R}}(t) \tilde{\mathbf{\Phi}}(t)\} \tag{36}$$

where $\tilde{\mathbf{\Phi}}(t) = [\alpha^{t-1}(\boldsymbol{\varphi}(t-1,0) - \mathbf{1w}^T) \vdots\ \alpha^{t-2}(\boldsymbol{\varphi}(t-1,1) - \mathbf{1w}^T) \vdots \cdots \vdots\ \alpha^0(\boldsymbol{\varphi}(t-1,t-1) - \mathbf{1w}^T)]^T$ according to (8) and

$$\mathbf{Q}_4(t) = (1-\alpha)^2 E\{\tilde{\mathbf{\Phi}}(t)^T \mathbf{m}_X(t) \mathbf{m}_X(t)^T \tilde{\mathbf{\Phi}}(t)\}\}. \tag{37}$$

Theorem 3: Let assumptions B1), B2) and A3) hold. Then, in the case of random consensus matrices,

$$\max_{i,j} Q_{ij}(t) = O(1-\alpha) \tag{38}$$

for all t and both $\mathbf{H_0}$ and $\mathbf{H_1}$, where $Q_{ij}(t)$ are the elements of $\mathbf{Q}(t)$, defined by (33), (36) and (37).

Proof: Proceeding like in the proof of Theorem 1, define the quadratic form $\mathbf{y}^T\mathbf{Q}_3(t)\mathbf{y}$, where $\mathbf{Q}_3(t)$ is given by (36). As a consequence of the independence between $\{\mathbf{x}(t)\}$ and $\{\mathbf{C}(t)\}$, we use (20) directly and realize that we are concerned here with the expression

$$E\{\tilde{\mathbf{\Phi}}(t)^T\tilde{\mathbf{\Phi}}(t)\} = \sum_{j=0}^{t-1}\mathbf{D}(t,j)\alpha^{2(t-j-1)} \tag{39}$$

where $\mathbf{D}(t,j) = E\{(\boldsymbol{\varphi}(t-1,j)-\mathbf{1}\mathbf{w}^T)(\boldsymbol{\varphi}(t-1,j)-\mathbf{1}\mathbf{w}^T)^T\}$, $j=0,\ldots,t-1$; furthermore, $\mathbf{D}(t,j) = E\{\boldsymbol{\varphi}(t-1,j)\boldsymbol{\varphi}(t-1,j)^T\}-\bar{\mathbf{C}}^{t-j}\mathbf{w}\mathbf{1}^T-\mathbf{1}\mathbf{w}^T\bar{\mathbf{C}}^{(t-j)T}+\mathbf{1}\mathbf{w}^T\mathbf{w}\mathbf{1}^T$. By adding and subtracting the term $\bar{\mathbf{C}}^{t-j}\bar{\mathbf{C}}^{(t-j)T}$, one obtains

$$\begin{aligned}\mathbf{D}(t,j) &= E\{\boldsymbol{\varphi}(t-1,j)\boldsymbol{\varphi}(t-1,j)^T\}\\ &\quad-\bar{\mathbf{C}}^{t-j}\bar{\mathbf{C}}^{(t-j)T}+(\bar{\mathbf{C}}^{t-j}-1\mathbf{w}^T)(\bar{\mathbf{C}}^{t-j}-1\mathbf{w}^T)^T.\end{aligned} \tag{40}$$

Furthermore, $E\{\boldsymbol{\varphi}(t,j)\boldsymbol{\varphi}(t,j)^T\} = E\{\mathbf{C}(t)E\{\mathbf{C}(t-1)\cdots E\{\mathbf{C}(j)\mathbf{C}(j)^T\}\cdots\mathbf{C}(t-1)^T\}\mathbf{C}(t)^T\} = E\{\boldsymbol{\varphi}(t-j,0)\boldsymbol{\varphi}(t-j,0)^T\}$, having in mind time invariance of the distribution of $\mathbf{C}(t)$.

Define the recursion

$$\mathbf{S}(t+1) = E\{\mathbf{C}(t)\mathbf{S}(t)\mathbf{C}(t)^T\}, \tag{41}$$

with $\mathbf{S}(0)=\mathbf{I}$; obviously, $E\{\boldsymbol{\varphi}(t,j)\boldsymbol{\varphi}(t,j)^T\} = \mathbf{S}(t-j+1)$.

Define also $\mathbf{F}(t+1) = \bar{\mathbf{C}}\mathbf{F}(t)\bar{\mathbf{C}}^T$, with $\mathbf{F}(0)=\mathbf{I}$, as well as $\mathbf{G}(t+1) = (\bar{\mathbf{C}}-\mathbf{1}\mathbf{w}^T)\mathbf{G}(t)(\bar{\mathbf{C}}-\mathbf{1}\mathbf{w}^T)^T$, with $\mathbf{G}(0)=\mathbf{I}$, so that

$$\mathbf{D}(t,j) = \mathbf{S}(t-j)-\mathbf{F}(t-j)+\mathbf{G}(t-j). \tag{42}$$

From (41), we have further

$$\mathrm{vec}\{\mathbf{S}(t+1)\} = E\{\mathbf{C}(t)\otimes\mathbf{C}(t)\}\mathrm{vec}\{\mathbf{S}(t)\}. \tag{43}$$

Similarly,

$$\mathrm{vec}\{\mathbf{F}(t+1)\} = (\bar{\mathbf{C}}\otimes\bar{\mathbf{C}})\mathrm{vec}\{\mathbf{F}(t)\} \tag{44}$$

and

$$\mathrm{vec}\{\mathbf{G}(t+1)\} = ((\bar{\mathbf{C}}-\mathbf{1}\mathbf{w}^T)\otimes(\bar{\mathbf{C}}-\mathbf{1}\mathbf{w}^T))\mathrm{vec}\{\mathbf{G}(t)\}. \tag{45}$$

Consequently,

$$\begin{aligned}\mathrm{vec}\{\mathbf{D}(t,j)\} &= [E\{\mathbf{C}(t)\otimes\mathbf{C}(t)\}^{(t-j)}-(\bar{\mathbf{C}}\otimes\bar{\mathbf{C}})^{(t-j)}\\ &\quad+((\bar{\mathbf{C}}-\mathbf{1}\mathbf{w}^T)\otimes(\bar{\mathbf{C}}-\mathbf{1}\mathbf{w}^T))^{(t-j)}]\mathrm{vec}\{\mathbf{I}\}.\end{aligned} \tag{46}$$

The third term in the brackets at the right-hand side of (46) can be directly analyzed using the arguments of the proof of Theorem 1, having in mind B1) and B2). It can be, therefore, directly concluded that its elements tend to zero exponentially as $|t-j|\to\infty$.

The second term in the brackets has an eigenvalue at 1 of algebraic multiplicity 1, according to B1), and the remaining eigenvalues tend to zero as $|t-j|\to\infty$ according to the assumed properties of the diagonal terms in $\mathbf{C}^{(k)}$, $k=1,\ldots,N$.

The first term in the brackets has also the property that it has an eigenvalue at 1 of algebraic multiplicity 1, which can be proved using the fact that this property holds if and only if the graph associated with $E\{\mathbf{C}(t)\otimes\mathbf{C}(t)\}$ has a spanning tree [13]. Namely, the graph associated with $E\{\mathbf{C}(t)\otimes\mathbf{C}(t)\}$ is composed of n digraphs having the same structure as the graph associated with $\bar{\mathbf{C}}$; these digraphs are interconnected in such a way that the ith digraph is connected with the jth digraph if the ith node is connected with the jth node according to the graph associated with $\bar{\mathbf{C}}$, $i,j=1,\ldots,n$. These interconnections in the graph associated with $E\{\mathbf{C}(t)\otimes\mathbf{C}(t)\}$ connect at least all n nodes of the ith digraph with all n nodes of the jth digraph on the one-to-one basis (according to an assumed ordering), having in mind that the realizations $\mathbf{C}^{(k)}$ have positive-diagonal elements by assumption. Additional interconnection edges also exist in accordance with the structure of $\mathbf{C}^{(k)}$. Consequently, it is easy to conclude that the graph associated with $E\{\mathbf{C}(t)\otimes\mathbf{C}(t)\}$ also possesses a spanning tree, and, therefore, has the eigenvalue at 1 with algebraic multiplicity 1. Modules of all the remaining eigenvalues are strictly less than 1, having in mind that the diagonal elements of $E\{\mathbf{C}(t)\otimes\mathbf{C}(t)\}$ are positive.

Notice that in the limit, when $|t-j|\to\infty$, we have that $E\{\mathbf{C}(t)\otimes\mathbf{C}(t)\}^{(t-j)}\to\mathbf{1}\boldsymbol{\sigma}_1^T$ and $(\bar{\mathbf{C}}\otimes\bar{\mathbf{C}})^{(t-j)}\to\mathbf{1}\boldsymbol{\sigma}_2^T$, where $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ are nonnegative column vectors.

Therefore, coming back to (40), we obtain readily that $\|\mathbf{D}(t,j)\|\le d<\infty$ for all t and j.

Consequently,

$$\mathbf{y}^T\mathbf{Q}_3(t)\mathbf{y}\le(1-\alpha)^2M_1\sum_{i=0}^{\infty}\alpha^{2i}\|\mathbf{y}\|^2 = O(1-\alpha), \tag{47}$$

for all t, where $M_1\le\infty$.

The term $\mathbf{y}^T\mathbf{Q}_4(t)\mathbf{y}$ can be analyzed analogously, starting from (35). Using the fact that

$$\begin{aligned}&\frac{1}{2}E\{\tilde{\mathbf{\Phi}}(t)^T\mathbf{m}_X(t)\mathbf{m}_X(t)^T\tilde{\mathbf{\Phi}}(t)\}\le\\ &\alpha^{2(t-1)}E\{(\boldsymbol{\varphi}(t-1,0)-\mathbf{1}\mathbf{w}^T)\mathbf{m}\mathbf{m}^T(\boldsymbol{\varphi}(t-1,0)-\mathbf{1}\mathbf{w}^T)^T\}\\ &+\cdots+\\ &E\{(\boldsymbol{\varphi}(t-1,t-1)-\mathbf{1}\mathbf{w}^T)\mathbf{m}\mathbf{m}^T(\boldsymbol{\varphi}(t-1,t-1)-\mathbf{1}\mathbf{w}^T)^T\}\end{aligned} \tag{48}$$

we obtain, on the basis on the results related to $\mathbf{Q}_3(t)$, that under $\mathbf{H_1}$

$$\mathbf{y}^T\mathbf{Q}_4(t)\mathbf{y}\le(1-\alpha)^2M_2\sum_{i=0}^{\infty}\alpha^{2i}\|\mathbf{y}\|^2 = O(1-\alpha), \tag{49}$$

for all t, where $M_2<\infty$. Therefore, $\mathbf{y}^T\mathbf{Q}(t)\mathbf{y} = O(1-\alpha)$. Hence, the result. ∎

It is important to notice at this point that the result of Theorem 3, when compared to the result of Theorem 1, shows that randomness of the network communications can cause an increase of the mean square error with respect to the constant consensus matrix case, as it could be expected. It is also intuitively clear that its tracking capabilities can be deteriorated. Illustrations of the efficiency of the resulting detector, which is still practically very satisfactory, will be given in the next section.

B. Variable Forgetting Factor

The analysis of the algorithm in the case of random consensus matrices and variable forgetting factor follows method-

ologically the proof of Theorem 2 in Section III, using the results obtained in Theorem 3.

Theorem 4: Let in (2) and (6) the forgetting factor be in the form $\alpha(t+1) = 1-\gamma(t+1)$, and let the assumptions B1), B2) and A3) be satisfied, together with:

B3) $\gamma(t)$ is a nonincreasing sequence satisfying $\gamma(t) > 0$, $\sum_{t=1}^{\infty}\gamma(t) = \infty$ and $\sum_{t=1}^{\infty}\gamma(t)^2 < \infty$.

Then, $\|\mathbf{Q}(t)\| = o(1)$ for both $\mathbf{H_0}$ and $\mathbf{H_1}$.

Proof: In the case of time varying forgetting factor, (24) becomes

$$\mathbf{e}(t) = \sum_{i=0}^{t-1}\pi(t,t+1-i)[\boldsymbol{\varphi}(t-1,t-i-1)-\mathbf{1}\mathbf{w}^T]\gamma(t-i)\mathbf{x}(t-i). \tag{50}$$

Proceeding like in the proof of Theorem 2, one obtains, after using the results of Theorem 3, that

$$\mathbf{y}^T\mathbf{Q}_3(t)\mathbf{y} \le \|\mathbf{y}\|^2 M_1 \sum_{i=0}^{\infty}\pi(t,t+1-i)^2\gamma(t-i)^2. \tag{51}$$

Having in mind that we have assumed by B3) that $\sum_{t=1}^{\infty}\gamma(t)^2 < \infty$, we can apply the Kronecker's lemma like in (26) and conclude that the expression at the right-hand side of (51) tends to zero when $t \to \infty$. Analogously, one can show that $\mathbf{y}^T\mathbf{Q}_4(t)\mathbf{y} = o(1)$. ∎

Corollary 2: Under the assumptions of Theorem 4 and with $\gamma(t) = \frac{1}{t}$ we have $\|\mathbf{Q}(t)\| = O(t^{-1})$.

Proof: In this case, we have from (50)

$$\mathbf{e}(t) = \frac{1}{t}\sum_{i=0}^{t-1}[\boldsymbol{\varphi}(t-1,t-i-1) - \mathbf{1}\mathbf{w}^T]\mathbf{x}(t-i), \tag{52}$$

and the result immediately follows after applying the methodology of Theorems 3 and 4. ∎

It is important to notice that in the case of time varying forgetting factor we still have convergence of the mean square error to zero, after assuming that $\sum_{t=1}^{\infty}\gamma(t)^2 < \infty$. Uncertainty introduced by randomness of the consensus matrix influences the rate of convergence, as it can be seen in Corollary 2 (compare with Corollary 1).

Remark 4: In the general case of intermittent measurements, we have $\boldsymbol{\Gamma}(t)\mathbf{x}(t)$ in (6) instead of $\mathbf{x}(t)$, where the sequence $\{\boldsymbol{\Gamma}(t)\}$ can be modelled as an i.i.d. random sequence in which $\bar{\boldsymbol{\Gamma}}(t)$ is a diagonal binary matrix, such that $E\{\boldsymbol{\Gamma}(t)\} = \bar{\boldsymbol{\Gamma}} = \mathrm{diag}\{q_1,\ldots,q_n\}$, where $0 \le q_i \le 1$ represent the probabilities of getting measurements [19], [25]. Taking the corresponding centralized scheme with intermittent measurements as a reference, we have, instead of B2), the condition $(\lim_{i\to\infty}\bar{\mathbf{C}}^i - \mathbf{1}\mathbf{w}^T)\bar{\boldsymbol{\Gamma}} = 0$, which is equivalent to B2) when $\bar{\boldsymbol{\Gamma}}$ is nonsingular. In the case when we take the ideal centralized scheme with no missing measurements as a reference, we have an increase of the bias due to intermittent measurements $\Delta\mathbf{m}_e$ satisfying $\|\Delta\mathbf{m}_e\|_\infty \le (1-\alpha)\sum_{i=1}^{\infty}\alpha^i\|\bar{\mathbf{C}}^{i+1}\|_\infty\|(\bar{\boldsymbol{\Gamma}}-\mathbf{I})\|_\infty\|\mathbf{m}\|_\infty = \max_i(1-q_i)\|\mathbf{m}\|_\infty$ (compare with Remark 1). ∎

V. Numerical Examples

Design of Consensus Matrices. In the case when the weight vector $\mathbf{w}$ in (2) is *a priori* selected according to Section II, the design of the communication gains is based on either A2) or B2). Assume that we have $n = 10$ nodes and that $\mathbf{w}^T = [15.72\ \ 13.02\ 20.74\ 11.29\ 5.31\ \ 6.17\ 12.44\ 5.59\ 3.62\ 6.10] \times 10^{-2}$. Assuming that the consensus matrix is constant, we can solve (11) after fixing zero elements of $\mathbf{C}$ at nonsymmetric randomly selected places, and we obtain

$$\mathbf{C} = \begin{bmatrix} 58.57 & 0.16 & 0 & 8.44 & 3.70 & 9.90 & 0 & 0 & 0 & 19.23 \\ 0 & 87.27 & 0 & 0 & 0 & 9.90 & 0 & 2.83 & 0 & 0 \\ 0 & 0.16 & 58.03 & 8.44 & 0 & 0 & 14.13 & 0 & 0 & 19.23 \\ 0 & 0.16 & 0 & 72.98 & 0 & 9.90 & 14.13 & 2.83 & 0 & 0 \\ 0 & 0 & 5.34 & 8.44 & 14.12 & 9.90 & 0 & 0 & 62.19 & 0 \\ 7.32 & 0.16 & 5.34 & 0 & 3.70 & 80.64 & 0 & 2.83 & 0 & 0 \\ 0 & 0 & 5.34 & 0 & 0 & 9.90 & 3.33 & 0 & 62.19 & 19.23 \\ 7.32 & 0.16 & 0 & 0 & 3.70 & 9.90 & 14.13 & 45.55 & 0 & 19.23 \\ 7.32 & 0 & 5.34 & 8.44 & 0 & 0 & 14.13 & 0 & 45.52 & 19.23 \\ 7.32 & 0 & 5.34 & 8.44 & 0 & 9.90 & 14.13 & 2.83 & 0 & 52.03 \end{bmatrix} \times 10^{-2}.$$

The case of random consensus matrices is more complex. In order to demonstrate the methodology, consider a fully connected sensor network with $n = 3$ nodes, with the asymmetric asynchronous gossip algorithm with one communication at a time. Then, there are $N = 6$ possible realization matrices of $\mathbf{C}(t)$: $\mathbf{C}^{(1)} = \mathbf{I}+\mathbf{C}^{[1,2]}$, $\mathbf{C}^{(2)} = \mathbf{I}+\mathbf{C}^{[1,3]}$, $\mathbf{C}^{(3)} = \mathbf{I}+\mathbf{C}^{[2,1]}$, $\mathbf{C}^{(4)} = \mathbf{I}+\mathbf{C}^{[2,3]}$, $\mathbf{C}^{(5)} = \mathbf{I}+\mathbf{C}^{[3,1]}$, and $\mathbf{C}^{(6)} = \mathbf{I}+\mathbf{C}^{[3,2]}$ (see the introduction of Section IV). Consequently, one obtains that

$$\bar{\mathbf{C}} = \sum_{k=1}^{N}\mathbf{C}^{(k)}p_k = \begin{bmatrix} 1-\gamma_{12}p_1-\gamma_{13}p_2 & \gamma_{12}p_1 & \gamma_{13}p_2 \\ \gamma_{21}p_3 & 1-\gamma_{21}p_3-\gamma_{23}p_4 & \gamma_{23}p_4 \\ \gamma_{31}p_5 & \gamma_{32}p_6 & 1-\gamma_{31}p_5-\gamma_{32}p_6 \end{bmatrix}. \tag{53}$$

We have now two main practical options for getting a solution to the problem (31):

a) to adopt values of the probabilities p_k (e.g., $p_k = \frac{1}{N}$, $k = 1,\ldots,N$) and to solve (31) for the remaining set of parameters;

b) to adopt values of the elements of $\mathbf{C}^{(k)}$, i.e., the set of parameters γ_{ij} (e.g., $\gamma_{ij} = 0.5$), and to solve (31) for the probabilities p_k.

Coming back to the above network with $n = 10$, we obtain for the gossip algorithm with one communication at a time and using the methodology b) with $\gamma_{ij} = 0.5$, that

$$\bar{\mathbf{C}} = \begin{bmatrix} 95.70 & 0.02 & 0 & 0.88 & 0.38 & 1.03 & 0 & 0 & 0 & 2.00 \\ 0 & 98.68 & 0 & 0 & 0 & 1.03 & 0 & 0.29 & 0 & 0 \\ 0 & 0.02 & 95.65 & 0.88 & 0 & 0 & 1.47 & 0 & 0 & 2.00 \\ 0 & 0.02 & 0 & 97.20 & 0 & 1.03 & 1.47 & 0.29 & 0 & 0 \\ 0 & 0 & 0.55 & 0.88 & 91.09 & 1.03 & 0 & 0 & 6.45 & 0 \\ 0.76 & 0.02 & 0.55 & 0 & 0.38 & 97.99 & 0 & 0.29 & 0 & 0 \\ 0 & 0 & 0.55 & 0 & 0 & 1.03 & 89.97 & 0 & 6.45 & 2.00 \\ 0.76 & 0.02 & 0 & 0 & 0.38 & 1.03 & 1.47 & 94.35 & 0 & 2.00 \\ 0.76 & 0 & 0.55 & 0.88 & 0 & 0 & 1.47 & 0 & 94.35 & 2.00 \\ 0.76 & 0 & 0.55 & 0.88 & 0 & 1.03 & 1.47 & 0.29 & 0 & 95.02 \end{bmatrix} \times 10^{-2}.$$

One can easily verify that $\lim_{i\to\infty}\bar{\mathbf{C}}^i = \mathbf{1}\mathbf{w}^T$ and that the sum of all nondiagonal elements is equal to 0.5.

In both presented examples the columns of $\bar{\mathbf{C}}$ have equal elements (excluding the diagonal ones); this has been adopted as

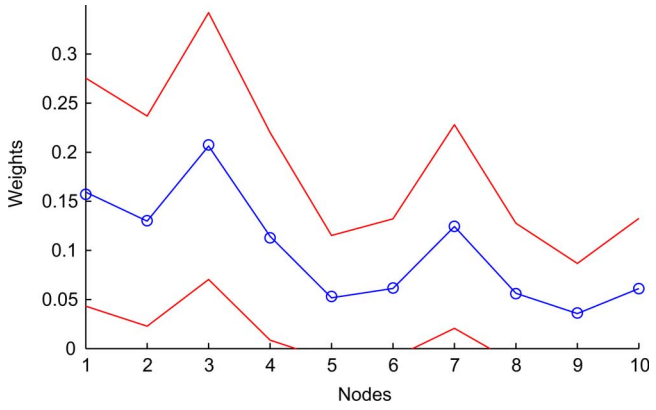

Fig. 1. Estimated mean values $\pm$ one standard deviation of one row of $\mathbf{C}(1000)\cdots\mathbf{C}(1)$; components of the weight vector w are represented by circles.

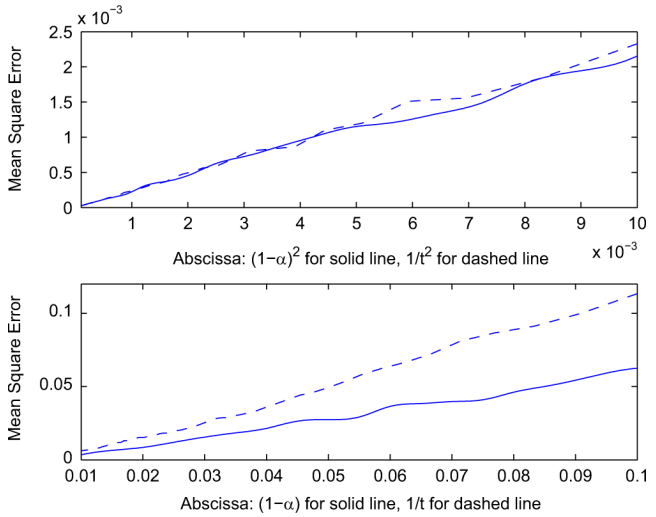

Fig. 2. Mean square error: constant consensus matrices (top), random consensus matrices (bottom); constant α (solid line), $\alpha(t) = 1 - \frac{1}{t}$ (dashed line).

an additional constraint in the linear programming problem, for the sake of practical convenience.

As an illustration of the convergence properties of the algorithm, the products $\boldsymbol{\varphi}(1000, 1) = \mathbf{C}(1000)\cdots\mathbf{C}(1)$ have been calculated using 5000 Monte Carlo runs. Fig. 1 shows that the obtained mean value of any of the rows of $\boldsymbol{\varphi}(1000, 1)$ matches the weight vector $\mathbf{w}$; the standard deviation decreases when the number of agents communicating simultaneously increases.

Distance From the Centralized Statistics. In order to get an insight into the relationship between the performance of the proposed algorithm and the performance of the centralized scheme (2), a sensor network with $n = 10$ has been simulated, adopting model (3); a diagonal covariance matrix has been assumed, with diagonal elements σ_i^2, $i = 1,\ldots,10$, randomly taken from the interval $[0.5, 1.5]$. Consensus matrices for both constant and time varying cases have been designed as described above. In Fig. 2 the mean square error between the decision function generated by the centralized algorithm (2) and those generated by different versions the proposed algorithm (6) is calculated using 1000 Monte Carlo runs as a function of: 1) $(1-\alpha)^2$ at $t = 1000$ (according to Theorem 1), 2) t^{-2} (according to Theorem 2, with $\alpha(t) = 1 - \frac{1}{t}$), 3) $(1-\alpha)$ at $t = 1000$ (according to Theorem 3), and 4) t^{-1} (according to Theorem 4, with $\alpha(t) = 1 - \frac{1}{t}$). The results of the theorems are clearly justified, since all the obtained curves are approximately linear.

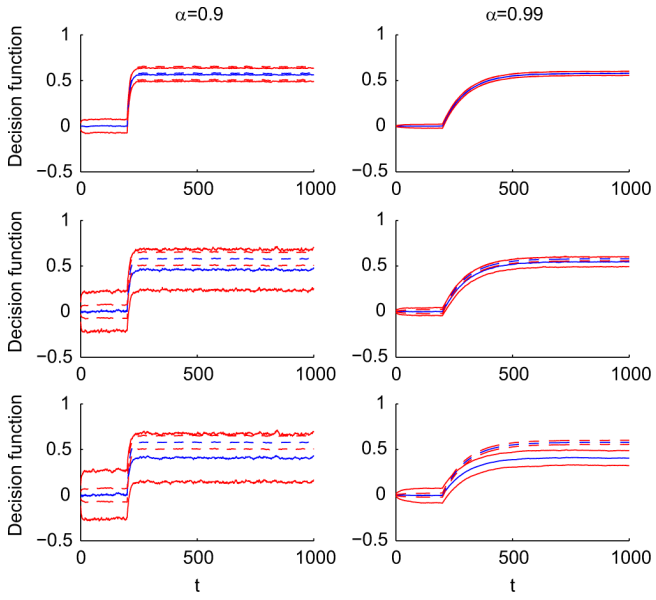

Fig. 3. Means $\pm$ one standard deviation for decision functions of one node: centralized strategy (dashed lines), proposed algorithm (solid lines); constant consensus matrices (top), random consensus matrices (middle), no consensus (bottom).

Detection Quality. As an illustration of the typical performance of the proposed detector, in Fig. 3 the mean $\pm$ one standard deviation is presented for decision functions generated by one node of the network described in the previous example, using: a) constant consensus matrix, b) random consensus matrix (gossip algorithm with one communication at a time) and c) zero communication gains (completely decentralized scheme); the centralized decision function (2) is represented by dashed lines (1000 realizations have been used). The mean vector $\boldsymbol{\mu}$ has been changed at $t = 200$ from $\boldsymbol{\mu} = \boldsymbol{\mu}^0 = \mathbf{0}$ to $\boldsymbol{\mu} = \boldsymbol{\mu}^1$, where μ_i^1, $i = 1,\ldots,10$, have been randomly taken from the interval $(0, 1]$ (see Section II). The proposed algorithm is very close to the centralized scheme in the case of the constant consensus matrix. Random consensus matrices introduce an additional uncertainty; however, the algorithm still remains efficient after using an appropriate threshold. The completely decentralized scheme using independent local detectors provides the worst quality, as expected. Fig. 4 shows the mean $\pm$ one standard deviation for decision functions analogous to those from Fig. 3 for the case of intermittent measurements; the values of the probabilities of getting measurements are set to $q_i = 0.5$, $i = 1,\ldots,n$. It can be seen that the performance of the algorithm is satisfactory even in this case of significant percentage of missing measurements, with an expected increase of the detection delay.

Detection quality is expressed numerically in Table I using steady state *deflections* (output signal-to-noise ratios), calculated for all the nodes according to

$$\Delta_i(t) = \frac{(E_{H_1}\{s_i(t)\} - E_{H_0}\{s_i(t)\})^2}{\mathrm{var}_{H_0}\{s_i(t)\}} \tag{54}$$

(see [30]). It is possible to observe that the proposed algorithm provides a great overall improvement of detection quality with

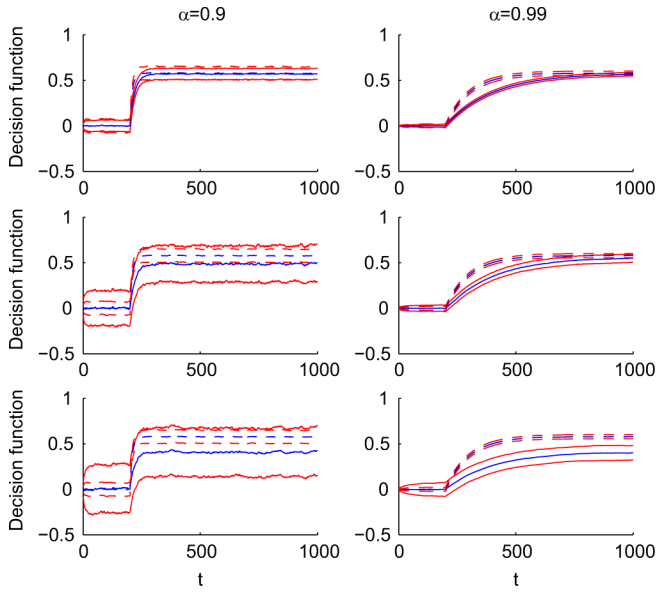

Fig. 4. Means ± one standard deviation for decision functions of one node for 50% measurements missing: centralized strategy (dashed lines), proposed algorithm (solid lines); constant consensus matrices (top), random consensus matrices (middle), no consensus (bottom).

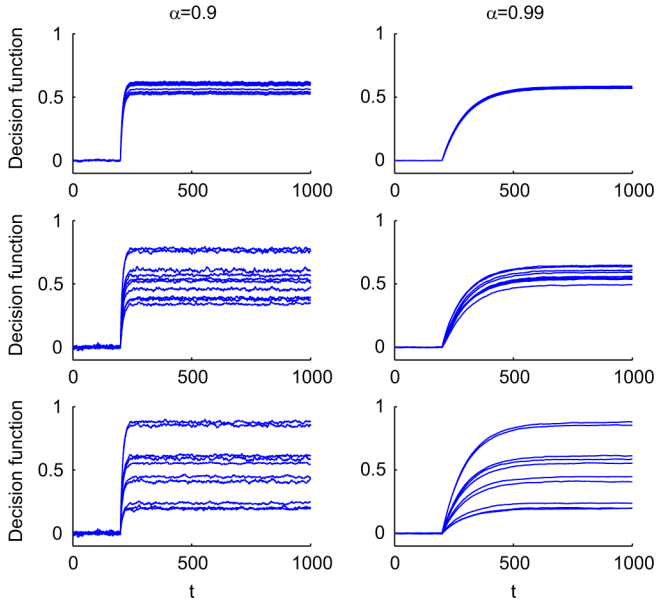

Fig. 5. Means of decision functions for all the nodes: constant consensus matrices (top), random consensus matrices (middle), no consensus (bottom).

TABLE I
DEFLECTIONS FOR DIFFERENT NODES IN THE NETWORK

$\alpha = 0.9$			$\alpha = 0.99$		
C	**C**(t)	local	**C**	**C**(t)	local
39.487	14.354	15.093	575.39	182.95	165.09
35.867	12.308	12.941	575.14	162.14	124.07
35.019	12.17	12.443	599.77	183.48	130.96
33.027	8.4947	7.4379	591.77	158.09	79.155
59.148	4.7765	2.6779	654.6	162.61	25.018
27.073	5.3166	4.1768	532.4	124.58	40.959
42.288	8.9375	5.9423	581.42	194.69	65.75
34.623	3.9882	1.6699	600.16	150.35	15.455
46.46	3.4091	0.78296	596.47	170.82	9.2257
51.288	6.6413	1.4032	623.84	204.71	15.382
average deflection					
40.428	8.0396	6.4568	593.10	169.44	67.106
centralized detector					
61.107			652.93		

respect to the decentralized case: average deflections are larger. Deflections are larger for *all the nodes* of the network when α is close to one. In the case of constant consensus matrices detection quality is close to the one for the centralized scheme. Observe that deflections vary from node to node, having in mind large dispersion in the local means and variances. In order to show that we have such an agreement between the nodes which enables selection of one common detection threshold (as exposed in Section II), we give in Fig. 5 the corresponding families of the decision function means, computed for all the nodes. The desired behavior is obvious, especially for the higher value of α.

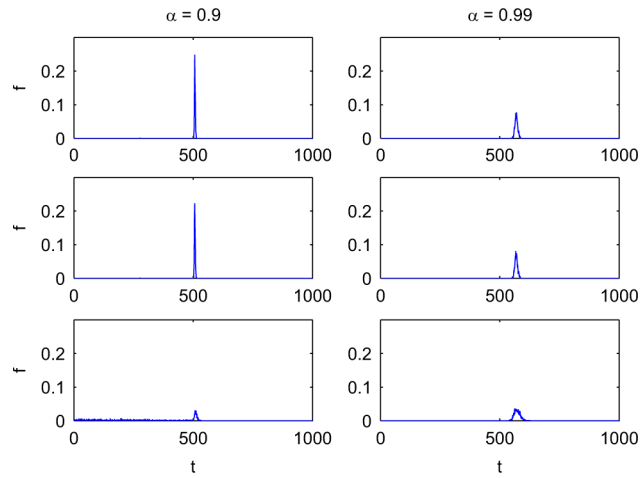

Fig. 6. Histograms of change detection instants: centralized decision function (top), constant consensus matrices (middle) and random consensus matrices (bottom).

Estimation of the instant of change. Theoretical treatment of the estimation of the instant of change (like in, e.g., [5]) has been out of the scope of the paper. However, numerous simulations have been undertaken in order to clarify properties of the proposed algorithm in this respect. Adopting the first crossing of the threshold $\frac{1}{2}\boldsymbol{\theta}^{1T}\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}^{1}\kappa^{-1}$ (for the model (3)) as the estimate of the instant of change, in Fig. 6 histograms (using 1000 realizations) giving relative frequencies of obtaining concrete time instants are represented for: a) the centralized decision function (top), b) constant consensus gains (middle) and c) asymmetric asynchronous gossip algorithm (bottom); the real moment of change was $t = 500$. It can be seen that the false alarms are considerable only in the case of $\alpha = 0.9$ and random consensus gains. Stopping rules based on testing at several consecutive time instants can lead to more robust solutions [31].

Further, efficiency of the algorithm is analyzed in the case of random consensus gains by calculating both the mean delay

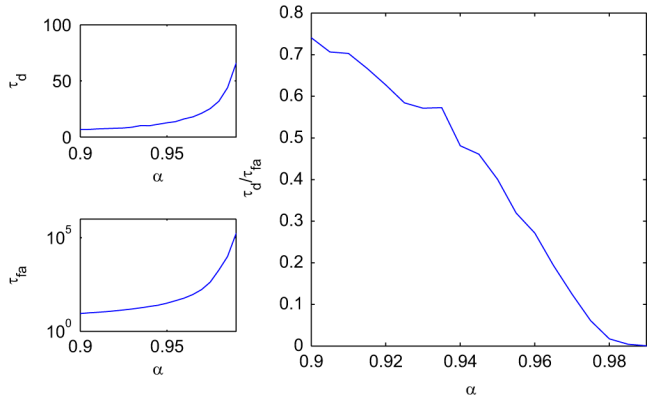

Fig. 7. Mean delay in change detection (left, top), mean time between false alarms (left, bottom) and their ratio (right) for one node in the case of random consensus gains.

in change time estimation τ_d and the mean time between false alarms τ_{fa}. Fig. 7 shows τ_d, τ_{fa} and $\frac{\tau_d}{\tau_{fa}}$ as functions of α (left), for one selected node. In practice, one has to find a good compromise between τ_d and τ_{fa}.

VI. Conclusion

In this paper a novel method for distributed change detection in sensor networks based on a consensus algorithm is proposed. The method is based on a combination of local geometric moving average control charts and a first order linear consensus scheme containing either constant or time varying random communication gains. The algorithm does not require any fusion center: under appropriate conditions, the state of any node in the network can be tested in real time with respect to a common threshold. A detailed analysis of the algorithm is done separately for constant and random consensus gains, assuming spatially and temporally correlated data. The analysis is focused on the mean square error between the state of the global centralized detector and the states of the nodes in the network generated by the proposed algorithm. It is proved, in the case of constant forgetting factor α in the underlying recursions, that the mean square error satisfies $O((1-\alpha)^2)$ for constant consensus gains, and $O(1-\alpha)$ for time varying random consensus gains, encompassing the asynchronous asymmetric "gossip" algorithms and communication faults. It is also proved that, in the case of time varying forgetting factors tending to 1 when t tends to infinity, the mean square error tends to zero. In the special case when the forgetting factor behaves like $1-\frac{1}{t}$, the mean square error is asymptotically bounded by $O(t^{-2})$ in the case of constant consensus gains, and by $O(t^{-1})$ in the case of random consensus gains. The case of missing measurements has also been discussed. Numerical examples cover several characteristic aspects of the algorithm and its applications, including the design of consensus matrices and an analysis of the estimate of the instant of change. Detection quality expressed using deflection demonstrates that the proposed method can represent an efficient tool for practice.

Further development of the adopted approach to distributed change detection can immediately lead to consensus based recursive detection schemes derived from the generalized likelihood ratio methodology (see, e.g., [31]), in the case of unknown jumps of either the mean value or the variance of the observed process. The proposed algorithm can also be directly applied in decentralized fault detection and isolation (FDI) schemes at the stage of residual evaluation.

References

[1] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer, 1996.
[2] R. Vishwanathan and P. Varshney, "Distributed detection with multiple sensors: Part I—Fundamentals," *Proc. IEEE*, vol. 85, pp. 54–63, 1997.
[3] J. N. Tsitsiklis, "Decentralized detection," *Adv. Statist. Signal Process.*, vol. 2, pp. 297–344, 1993.
[4] J. F. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, 2003.
[5] M. Basseville and L. V. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[6] V. V. Veeravalli, "Decentralized quickest change detection," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1657–1665, 2001.
[7] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge, U.K.: Cambridge Univ. Press, 2008.
[8] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, 1986.
[9] R. Olfati-Saber, A. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, 2007.
[10] A. Fax and R. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1465–1476, 2004.
[11] A. Jadbabaie, J. Lin, and A. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control6*, vol. 48, pp. 988–1001, 2003.
[12] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Trans. Autom. Control*, vol. 50, no. 2, pp. 169–182, 2005.
[13] W. Ren and R. Beard, "Consensus seeking in multi-agent systems using dynamically changing interaction topologies," *IEEE Trans. Autom. Control*, vol. 50, no. 5, pp. 655–661, 2005.
[14] E. Franco, R. Olfati-Saber, T. Parisini, and M. M. Polycarpou, "Distributed fault diagnosis using sensor networks and consensus based filters," in *Proc. 45th IEEE CDC Conf.*, 2006, pp. 386–391.
[15] S. Aldosari and J. M. F. Moura, "Distributed detection in sensor networks: Connecting graphs and small world networks," in *Proc. 39th Asilomar Conf. Signals, Syst., Comput.*, 2005, pp. 230–234.
[16] S. Kar and J. M. F. Moura, "Consensus based detection in sensor networks: Topology design under practical constraints," presented at the Workshop Inf. Theory Sensor Netw., Santa Fe, NM, Jun. 2007.
[17] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, 2011.
[18] S. S. Stanković, M. S. Stanković, and D. M. Stipanović, "Consensus based overlapping decentralized estimator," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 410–415, 2009.
[19] S. S. Stanković, M. S. Stanković, and D. M. Stipanović, "Consensus based overlapping decentralized estimation with missing observations and communication faults," *Automatica*, vol. 45, pp. 1397–1406, 2009.
[20] S. S. Stanković, M. S. Stanković, and D. M. Stipanović, "Decentralized parameter estimation by consensus based stochastic approximation," in *Proc. 46th IEEE Conf. Decision Control*, 2007, pp. 1535–1540.
[21] P. Braca, S. Marano, and V. Matta, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, 2008.
[22] P. Braca, S. Marano, V. Matta, and P. Willett, "Asymptotic optimality of running consensus in testing binary hypotheses," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, 2010.
[23] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.
[24] D. Bajović, D. Jakovetić, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Distributed detection over time-varying networks: Large deviations analysis," in *Proc. Allerton Conf.*, 2010, pp. 302–309.

[25] S. S. Stanković, M. S. Stanković, and D. M. Stipanović, "Decentralized parameter estimation by consensus based stochastic approximation," *IEEE Trans. Autom. Control*, vol. 56, no. 3, pp. 531–543, Mar. 2011.
[26] C. Canudas, Ed., in *Proc. 2nd IFAC Workshop on Estimation and Control on Network Systems (NecSys)*, Sep. 2010, IFACPapersOnLine.net.
[27] S. S. Stanković, N. Ilić, M. S. Stanković, and K. H. Johansson, "Distributed change detection based on a randomized consensus algorithm," in *Proc. 5th Europ. Conf. Circuits Syst. Commun.*, 2010, pp. 51–54.
[28] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.
[29] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poola, M. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, 2004.
[30] B. Picinbono, "On deflection as a performance criterion for detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 3, pp. 1072–1081, 1995.
[31] S. X. Ding, *Model Based Fault Diagnosis Techniques—Design Schemes, Algorithms and Tools*. New York: Springer-Verlag, 2008.
[32] N. Ilić, S. Stanković, M. Stanković, and K. H. Johansson, "Consensus based distributed change detection algorithm using GLR methodology," in *Proc. 19th Mediterranean Conf. Control Autom.*, 2011, pp. 1170–1175.
[33] R. A. Horn and C. A. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[34] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
[35] I. F. Pearce, "Matrices with dominating diagonal blocks," *J. Econom. Theory*, vol. 9, pp. 159–170, 1974.
[36] H. F. Chen, *Stochastic Approximation and its Applications*. Dordrecht, The Netherlands: Kluwer Academic, 2002.
[37] N. Ilić and S. Stanković, "Communication gains design in a consensus based distributed change detection algorithm," in *Proc. 8th Eur. Workshop Adv. Control Diagnosis*, 2010, pp. 51–54.

Srdjan S. Stanković received the Dipl.Ing., Mgr.Sc., and Ph.D. degrees from the Faculty of Electrical Engineering, University of Belgrade, Yugoslavia, in 1968, 1972 and 1975, respectively.

He was with the Institute for Nuclear Sciences, Vinća, Belgrade, Yugoslavia, from 1968 to 1972. Since 1973, he has been with the Faculty of Electrical Engineering, University of Belgrade, where he is currently Emeritus Professor of Automatic Control and Head of the Department for Signals and Systems. He held visiting positions at the Eindhoven University of Technology, Eindhoven, The Netherlands, and at Santa Clara University, Santa Clara, CA. He is currently President of the National Council for Higher Education of the Republic of Serbia, as well as President of the Serbian Society for Electronics, Communications, Computers, Control and Nuclear Engineering. He has published numerous scientific papers in diverse fields, including estimation and identification, adaptive systems, digital signal processing, processing of medical images, decentralized control and neural networks. He has also been the leader of numerous scientific projects for government and industry.

Nemanja Ilić received the Bachelor's and Master's degrees from the Faculty of Electrical Engineering at the University of Belgrade, Serbia, in 2007 and 2009, respectively. He is currently working towards the Ph.D. degree at the same faculty.

He published several papers at international conferences. His research interests currently include distributed estimation, detection, and tracking.

Miloš S. Stanković received the Bachelor's and Master's degrees from the School of Electrical Engineering at the University of Belgrade, Serbia, in 2002 and 2006, respectively, and the Ph.D. degree in systems and entrepreneurial engineering from the University of Illinois at Urbana-Champaign (UIUC) in 2009.

He was a Research and Teaching Assistant in the Control and Decision Group of the Coordinated Science Laboratory at the UIUC from 2006 to 2009. In 2009, he joined the Royal Institute of Technology (KTH), Stockholm, Sweden, as a Postdoctoral Researcher in the Automatic Control Laboratory and the ACCESS Linnaeus Centre. His research interests include decentralized control, estimation, detection and system identification, mobile sensor networks, networked control systems, dynamic game theory, extremum seeking control, stochastic and distributed optimization and machine learning.

Karl Henrik Johansson received M.Sc. and Ph.D. degrees in electrical engineering from Lund University, Lund Sweden.

He is Director of the ACCESS Linnaeus Centre and Professor at the School of Electrical Engineering, Royal Institute of Technology, Sweden. He is a Wallenberg Scholar and has held a Senior Researcher Position with the Swedish Research Council. He has held visiting positions at the University of Berkeley, Berkeley, CA, from 1998 to 2000 and the California Institute of Technology, Pasadena, from 2006 to 2007. His research interests are in networked control systems, hybrid and embedded control, and control applications in automotive, automation and communication systems.

Dr. Johansson is the Chair of the IFAC Technical Committee on Networked Systems. He has served on the Executive Committees of several European research projects in the area of networked embedded systems. He is on the Editorial Board of *IET Control Theory and Applications* and the *International Journal of Robust and Nonlinear Control*, and previously of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and *Automatica*. He was awarded an Individual Grant for the Advancement of Research Leaders from the Swedish Foundation for Strategic Research in 2005. He received the triennial Young Author Prize from IFAC in 1996 and the Peccei Award from the International Institute of System Analysis, Austria, in 1993. He received Young Researcher Awards from Scania in 1996 and from Ericsson in 1998 and 1999.