# Stochastic consensus over noisy networks with Markovian and arbitrary switches

ARTICLE *in* AUTOMATICA · OCTOBER 2010

Impact Factor: 3.02 · DOI: 10.1016/j.automatica.2010.06.016 · Source: DBLP



**4 AUTHORS**, INCLUDING:

Minyi Huang
Carleton University


Swayandipta Dey
University of Connecticut

Automatica 46 (2010) 1571–1583

Contents lists available at ScienceDirect

Automatica

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

# Stochastic consensus over noisy networks with Markovian and arbitrary switches☆

Minyi Huang[a,*], Subhrakanti Dey[b], Girish N. Nair[b], Jonathan H. Manton[b]

[a] School of Mathematics and Statistics, Carleton University, Ottawa, ON, K1S 5B6, Canada

[b] Department of Electrical and Electronic Engineering, University of Melbourne, Victoria 3010, Australia

ARTICLE INFO

*Article history:*
Received 5 November 2008
Received in revised form
3 December 2009
Accepted 25 May 2010
Available online 22 July 2010

*Keywords:*
Consensus
Measurement noises
Markovian lossy channels
Stochastic approximation
Quantized data
Packet losses

ABSTRACT

This paper considers stochastic consensus problems over lossy wireless networks. We first propose a measurement model with a random link gain, additive noise, and Markovian lossy signal reception, which captures uncertain operational conditions of practical networks. For consensus seeking, we apply stochastic approximation and derive a Markovian mode dependent recursive algorithm. Mean square and almost sure (i.e., probability one) convergence analysis is developed via a state space decomposition approach when the coefficient matrix in the algorithm satisfies a zero row and column sum condition. Subsequently, we consider a model with arbitrary random switching and a common stochastic Lyapunov function technique is used to prove convergence. Finally, our method is applied to models with heterogeneous quantizers and packet losses, and convergence results are proved.



## 1. Introduction

In distributed multi-agent systems, consensus problems have become one of most focussed research areas due to their wide application backgrounds; see the survey in Olfati-Saber, Fax, and Murray (2007) and Ren, Beard, and Atkins (2005). While most of past research has concentrated on deterministic models (see Jadbabaie, Lin, and Morse (2003), Olfati-Saber and Murray (2004) and Ren and Beard (2005), and references therein), recently, there is a considerable growth of interest in stochastic models addressing various uncertainty factors involved in the inter-agent information exchange. For instance, the communication link between the agents may be available only at random times, and random graphs are suitable for network connectivity modelling (Hatano & Mesbahi, 2005; Tahbaz-Salehi & Jadbabaie, 2008). Another important aspect of consensus models is random noises (Acemoglu, Nedić, & Ozdaglar, 2008; Aysal & Barner, 2009; Carli, Fagnani, Frasca, Taylor, & Zampieri, 2007; Ren, Beard, & Kingston, 2005; Schizas, Ribeiro, & Giannakis, 2008; Xiao, Boyd, & Kim, 2007). This is particularly important when the agents exchange their state information over communication channels (Schizas et al., 2008). For noisy modelling for flocking, formation and rendezvous, the reader is referred to Barooah and Hespanha (2007), Cucker and Mordecki (2008) and Martínez (2007).

This paper considers consensus problems over unreliable networks. We aim to develop a unified modelling and analytic framework addressing uncertainty aspects including measurement noises, random link gains, random signal losses, and quantization errors.

We begin with the signal reception modelling, where the random link gain results from analog channels. The analog signal transmission is motivated by specific sensor network applications. In recent years, a promising scheme for distributed detection/estimation in sensor networks has emerged based on analog forwarding, where measurements of the sensors are transmitted directly (possibly scaled) to a fusion center without any coding, which is motivated by optimality results on uncoded transmissions in point-to-point links (Gastpar, Rimoldi, & Vetterli, 2003; Goblick, 1965). It was shown in Gastpar and Vetterli (2003) that for a Gaussian sensor network, where multiple sensors measure a random scalar Gaussian field in noise and forward their noisy measurements to a fusion center for reconstruction of the source, the analog forwarding scheme is asymptotically optimal and approaches the minimum distortion achievable at the rate of

☆ This work was partially supported by the Natural Sciences and Engineering Research Council of Canada and the Australian Research Council. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Hideaki Ishii under the direction of Editor Ian R. Petersen.

* Corresponding author. Tel.: +1 613 5202600x8022; fax: +1 613 5203536.
*E-mail addresses:* mhuang@math.carleton.ca (M. Huang), sdey@unimelb.edu.au (S. Dey), gnair@unimelb.edu.au (G.N. Nair), jmanton@unimelb.edu.au (J.H. Manton).

0005-1098/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2010.06.016

$1/M$, where $M$ is the number of sensors. Also, the simplicity and low delay properties of uncoded analog forwarding make it an attractive choice for large distributed sensor networks.

During signal exchange of the sensor nodes, an important uncertainty feature is signal losses. This may be caused by the temporary extreme deterioration of the link quality, for instance, due to blocking objects travelling between the transmitting and receiving nodes, or simply by a temporary fault of the transmitter or receiver. For random connectivity modelling, there has existed a fair amount of work adopting random graphs where the communication of a pair of agents fails as independent processes; see, e.g., Hatano and Mesbahi (2005). Tahbaz-Salehi and Jadbabaie (2008) considered averaging with a sequence of independent coefficient matrices, which indicates that the overall network topology evolves independently; such modelling was also adopted for linear synchronization (Wu, 2006). Fagnani and Zampieri (2009) studied average consensus with independent packet dropping and analyzed the effect of the loss probability on performance. However, in practical systems, the loss of connectivity usually occurs with correlations, and for random graph based consensus models, this correlation effect has received relatively little attention. In this paper, we will introduce a Markovian modelling of the occurrence of the signal (or packet) losses, so that the temporal correlation properties of the channel functionality may be captured. Matei, Martins, and Baras (2008) considered a consensus problem with Markovian switching, but no measurement noise was involved; under a joint connectivity assumption they established almost sure convergence via exploiting the linear dynamics governing the evolution of the mean square consensus error.

Compared to measurement noises and random link gains, quantization is also a major source of signal distortion when high data rates are not available. Indeed, in sensor network deployment, due to limited on-board battery, sensors can only afford relatively low data rates. In models with quantization, asymptotic analysis of consensus algorithms is in general challenging and has attracted significant research attention. Carli, Fagnani, Speranzon, and Zampieri (2008) considered logarithmic quantization and developed convergence analysis for average consensus after assuming certain statistical properties of the quantization errors. In an average-consensus setting, Aysal, Coates, and Rabbat (2007, 2008) introduced probabilistic quantization for eliminating bias of the quantization errors, and showed that probabilistic quantization is equivalent to dithering. The authors have proven that their algorithm achieves almost sure convergence. However, their analysis relies on a key assumption that all the nodes use the same set of quantizers so that the quantized state space, as a lattice, contains points of consensus states. For heterogeneous quantizers, the above approach in general fails since the quantized state space may not contain any consensus state, and the iterates may persistently oscillate without converging. Yildiz and Scaglione (2008) analyzed data rate limited consensus models via coding, but assumed that quantization noises are temporally and spatially uncorrelated and that each node knows the network topology. Kashyap, Basar, and Srikant (2007) developed randomized algorithms to achieve nearly average consensus where each node takes values from a set of integers.

In Huang and Manton (2008, 2009, 2010), consensus problems were considered when agents obtain noisy measurements of the states of neighbors, and a stochastic approximation approach was applied to obtain mean square and almost sure convergence in models with fixed network topologies or with independent communications failures (Huang & Manton, 2008). General stochastic gradient based algorithms were introduced in Tsitsiklis, Bertsekas, and Athans (1986) for consensus problems arising in distributed function optimization. Stankovic, Stankovic, and Stipanovic (2007) considered decentralized parameter estimation by combining stochastic approximation of individual nodes with a consensus rule.

In this paper, for developing a unified analytic framework, we first introduce noisy measurements through uncoded analog forwarding to their neighbors via slow fading channels. We assume perfect phase synchronization such that the receiver obtains a scaled (by the fading envelope (amplitude) only) version of the transmitted data in noise when the link functions properly; see Fig. 1. Under this analog channel modelling, we first develop stochastic approximation type algorithms for consensus seeking over noisy networks with Markovian signal losses. This modelling leads to a consensus algorithm with Markovian switches. Compared to the independent communication failure considered in Huang and Manton (2008), the temporal correlation properties of the network switches make the convergence analysis more difficult since the method of viewing the coefficient matrix for averaging as a constant matrix subject to independent perturbations is no longer applicable. Our analysis will depend on more involved Lyapunov energy estimates. In particular, when only a joint connectivity condition is assumed for the noisy network, some special care must be taken to show a persistent decay of the energy. Next, we consider a model with arbitrary switches, for which our method for convergence analysis is to identify a suitable common stochastic Lyapunov function. The interested reader is referred to Olfati-Saber and Murray (2004) on the use of a common Lyapunov function (defined via the so-called disagreement function) in a deterministic setting. Finally, we apply our algorithm to a model with heterogeneous quantizers and packet losses. Convergence is obtained by combining probabilistic quantization Aysal et al. (2007, 2008) with a decreasing step size.

The organization of the paper is as follows. Section 2 describes the lossy signal exchange model. The stochastic approximation algorithm is introduced in Section 3. Convergence analysis is developed in Sections 4 and 5 for models with Markovian and arbitrary switches, respectively. Section 6 applies stochastic approximation to models with quantized data and packet losses. Section 7 presents simulation results and Section 8 concludes the paper.

### 1.1. Notation

The index of an agent will often be used as a superscript, but not an exponent, of various random variables. Throughout the paper we use $C$, $C_0$, $C_1$, etc. to denote generic positive constants whose value may change from place to place. Below we provide a list of the basic notation used in the paper.

$G$: the network topology as a directed graph.
$\mathcal{N}$: the nodes in $G$.
$\mathcal{E}$: the edges in $G$.
$\mathcal{E}_f$: the failure-prone edges in $G$.
$A_i$: the $i$th agent or node.
$\mathcal{N}_i$: the neighbors of $A_i$.
$x_t^i$: the state at node $i$.
$x_t$: the vector of individual states.
$I_t^{ki}$: the channel state on edge $(k, i)$.
$I_t$: the overall channel state.
$g_t^{ik}$: the analog channel gain on edge $(k, i)$.
$g_t$: the vector of individual channel gains.
$w_t^{ik}$: the measurement noise occurring at node $i$.
$y_t^{ik}$: the signal received at $A_i$ from $A_k$.
$G_t$: the network topology at time $t$.
$G^{(k)}$: the values that $G_t$ may take.
$\mathcal{N}_i^{(k)}$: the neighbors of $A_i$ within $G^{(k)}$.
$\mathcal{N}_{it}$: the neighbors of $A_i$ within $G_t$.
$B^{(k)}$: the stochastic approximation coefficient matrix when $G_t$ appears as $G^{(k)}$.
$a_t$: the step size of stochastic approximation.
$r_{i,k}$: the quantization level at node $i$.
$Q_i(t)$: the output of the probabilistic quantizer at node $i$.

## 2. Information exchange over unreliable networks

### 2.1. Preliminaries for network modelling

We begin by introducing some standard graph modelling of the network topology. A digraph $G = (\mathscr{N}, \mathscr{E})$ consists of a set of nodes $\mathscr{N} = \{1, \ldots, n\}$ and a set of directed edges $\mathscr{E} \subset \mathscr{N} \times \mathscr{N}$. A directed edge will simply be called an edge. An edge from node $i$ to node $j$ is denoted as an ordered pair $(i, j)$, where $i \neq j$. A directed path (from $i_1$ to $i_l$) consists of a sequence of nodes $i_1, i_2, \ldots, i_l$, $l \geq 2$, such that $(i_k, i_{k+1}) \in \mathscr{E}$ for $k = 1, \ldots, l-1$. The digraph $G$ is said to be strongly connected if there exists a directed path from each node to any other node. A directed tree is a digraph where each node, except the root, has exactly one parent node. The digraph $G$ is said to contain a spanning tree $G_s = (\mathscr{N}_s, \mathscr{E}_s)$ if $G_s$ is a directed tree such that $\mathscr{N}_s = \mathscr{N}$ and $\mathscr{E}_s \subset \mathscr{E}$. A strongly connected digraph always contains a spanning tree. The two names, agent and node, will be used interchangeably. The agent $A_k$ (resp., node $k$) is a neighbor of $A_i$ (resp., node $i$) if $(k, i) \in \mathscr{E}$, where $k \neq i$. Denote the neighbor set $\mathscr{N}_i = \{k | (k, i) \in \mathscr{E}\} \subset \mathscr{N}$.

### 2.2. Lossy signal reception at individual links

We use $G = (\mathscr{N}, \mathscr{E})$ to model the maximal set of communication links when there is no communication failure (or signal loss). Let $\mathscr{E}_f \subset \mathscr{E}$ denote the set of links that are failure-prone. When $\mathscr{E}_f = \emptyset$ (the empty set), the associated model has a fixed network topology. To avoid triviality, it is assumed that $\mathscr{E}_f \neq \emptyset$. The underlying probability space is denoted by $(\Omega, \mathscr{F}, P)$, where $\Omega$ is the sample space, $\mathscr{F}$ is the $\sigma$-algebra consisting of all events, and $P$ is the probability measure. The link state associated with an edge $(k, i) \in \mathscr{E}_f$ is modelled by a Markov chain $I_t^{ki}$ with state space $\{0, 1\}$ and stationary transition probabilities, where $t \geq 0$. The values 1 and 0, respectively, denote the normal and loss states. The value of $I_t^{ki}$ indicates whether or not node $i$ will successfully receive a measurement from node $k$ at time $t$. Note that if $(k, i) \notin \mathscr{E}_f$, $I_t^{ki}$ is not introduced. Compared with independent loss process modelling, the Markov chain based modelling may give a more realistic characterization of the temporal correlation property of the evolution of the link status. We note that our signal loss modelling may be extended to undirected graphs by using a Markov chain to describe the loss state of a bidirectional failure-prone link.

For agent $A_i$, denote its state at time $t$ by $x_t^i \in \mathbb{R}$, where $t \in \mathbb{Z}^+ = \{0, 1, 2, \ldots\}$. We assume each $A_i$ knows its own state $x_t^i$ exactly. Denote the signal output model

$$\bar{y}_t^{ik} = g_t^{ik} x_t^k + w_t^{ik}, \quad k \in \mathscr{N}_i \neq \emptyset, \tag{1}$$

where $g_t^{ik}$ is a random link gain and $w_t^{ik} \in \mathbb{R}$ is the additive noise. We use (1) to describe the attempted signal transmission from $A_k$ to $A_i$. Concerning each node's information on the channel, neither $A_i$ nor $A_k$ is required to know the value of $g_t^{ik}$. Instead, $A_i$ only knows the mean of $g_k^{ik}$. In other words, the node only has statistical information on the link gain.

If either $(k, i) \in \mathscr{E} \backslash \mathscr{E}_f$ (i.e., it is a lossless link) or $(k, i) \in \mathscr{E}_f$ but the channel operates in a normal condition, i.e., $I_t^{ki} = 1$, the received signal at $A_i$ is

$$y_t^{ik} = \bar{y}_t^{ik}. \tag{2}$$

See Fig. 1 for illustration. If $(k, i) \in \mathscr{E}_f$ and a signal loss occurs, i.e., $I_t^{ki} = 0$, we make the convention that $A_i$ receives

$$y_t^{ik} \equiv 0. \tag{3}$$

Similar loss models have been studied in distributed filtering problems; see, e.g., Huang and Dey (2007), Sinopoli et al. (2004) and Smith and Seiler (2003).

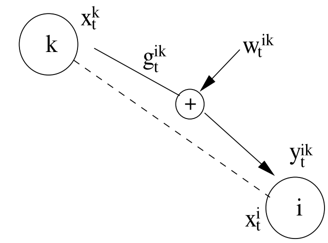



**Fig. 1.** Measurement with link gain $g_t^{ik}$ and additive noise $w_t^{ik}$.

The generic noisy and lossy signal reception model (1)–(3) may be used to describe analog signal transmission. But it is also applicable to certain digital channel based systems. The related detail will be presented in Section 6.

It should be noted that in Eq. (1), $g_t^{ik}$ and $w_t^{ik}$ are defined at all times for all $(k, i) \in \mathscr{E}$. In certain models, $I_t^{ki} = 0$, where $(k, i) \in \mathscr{E}_f$, may mean that $g_t^{ik}$ or $w_t^{ik}$, or both are not physically realized, for instance due to a temporary disorder of either the transmitter or the receiver. In such scenarios, we still keep them as dummy random variables, and their use gives a more unified model specification.

For the Markov chains $I_t^{ki}$, we may use a fixed ordering of all $(k, i) \in \mathscr{E}_f$ to list $I_t^{ki}$ into a vector process $I_t$. By suitable relabeling, the state space of $I_t$ may be denoted by $S_I = \{1, \ldots, K_0\}$, where $K_0 = 2^{|\mathscr{E}_f|}$. If $I_t = k \in S_I$, the real-time network topology, consisting of functioning links at time $t$, may be determined accordingly. Let the network topologies corresponding to the states in $S_I$ be denoted by $G^{(1)}, \ldots, G^{(K_0)}$, each being a subgraph of $G$. Without loss of generality, we assume $G^{(1)} = G$, which corresponds to $I_t = 1$ and means all links are functioning. The network at time $t$ is given as a digraph $G_t = (\mathscr{N}, \mathscr{E}_t)$, where $G_t$ takes one value from $\mathbf{G} = \{G^{(1)}, \ldots, G^{(K_0)}\}$ determined by $I_t$.

For each $t \in \mathbb{Z}^+$, the set of noises $\{w_t^{ik}, i \in \mathscr{N}$ and $k \in \mathscr{N}_i \neq \emptyset\}$ is listed into a vector $\mathbf{w}_t$ in which the position of $w_t^{ik}$ depends only on $(i, k)$ and does not change with $t$. Similarly, the random vector $g_t$ is defined by listing $g_t^{ik}$ by a fixed ordering of all $(i, k)$. Define the state vector

$$x_t = [x_t^1, \ldots, x_t^n]^T, \quad t \geq 0.$$

## 3. The stochastic algorithm

We will describe the algorithm by individual nodes to indicate the local implementation. Let $\mathscr{N}_i^{(k)}$ denote the neighbor set of node $i$ within $G^{(k)}$. We form a matrix $B^{(k)} = (b_{ij}^{(k)})_{1 \leq i,j \leq n}$ as follows.

Case 1. If $\mathscr{N}_i^{(k)} \neq \emptyset$, define

$$\begin{cases} b_{ij}^{(k)} > 0, & \text{if } j \in \mathscr{N}_i^{(k)}, \\ b_{ij}^{(k)} = 0, & \text{if } j \notin \mathscr{N}_i^{(k)} \cup \{i\}, \\ b_{ii}^{(k)} = -\sum_{j \in \mathscr{N}_i^{(k)}} b_{ij}^{(k)}. \end{cases} \tag{4}$$

Case 2. If $\mathscr{N}_i^{(k)} = \emptyset$, define

$$b_{ij}^{(k)} \equiv 0, \quad \text{for all } j \in \mathscr{N}. \tag{5}$$

By (4)–(5), each row sum of $B^{(k)}$ is 0. For agent $i$, denote its neighbor set by $\mathscr{N}_{it}$ when the instantaneous network topology is $G_t$. For $(j, i) \in \mathscr{E}$, we assume that the associated channel link gain $g_t^{ij}$ has a constant mean $\lambda^{ij} \neq 0$ for all $t \geq 0$. The state of agent $i$ is updated by the rule

$$x_{t+1}^i = [1 + a_t b_{ii}(t)] x_t^i + a_t \sum_{j \in \mathscr{N}_{it}} b_{ij}(t) (y_t^{ij} / \lambda^{ij}), \tag{6}$$

where the coefficients $b_{ij}(t)$ are determined by $G_t$ and $a_t > 0$ is the step size at time $t$. If $G_t = G^{(k)}$, the coefficients $b_{ij}(t)$ are obtained from $B^{(k)}$, i.e., $B_t = (b_{ij}(t))_{1\le i,j\le n} = B^{(k)}$. We adopt the convention: $\sum_{k\in\emptyset} = 0$ regardless of the summand. If $\mathcal{N}_{it} = \emptyset$, (6) is interpreted as $x^i_{t+1} = x^i_t$.

**Definition 1** (*Weak Consensus*)**.** The agents are said to reach weak consensus if $E|x_t|^2 < \infty$ for all $t \ge 0$, and $\lim_{t\to\infty} E|x^i_t - x^j_t|^2 = 0$ for all $i, j \in \mathcal{N}$. □

**Definition 2** (*Mean Square Consensus*)**.** The agents are said to reach mean square consensus if $E|x_t|^2 < \infty$ for all $t \ge 0$, and there exists a random variable $x^*$ such that $\lim_{t\to\infty} E|x^i_t - x^*|^2 = 0$ for all $i \in \mathcal{N}$. □

**Definition 3** (*Strong Consensus*)**.** The agents are said to reach strong consensus if there exists a random variable $x^*$ such that $\lim_{t\to\infty} x^i_t = x^*$, a.s., for all $i \in \mathcal{N}$. □

Note that algorithm (6) is based on the assumption that if the channel from $A_k$ to $A_i$ fails, $A_i$ assigns no weight to $A_k$. When no nodes have the ability to distinguish a neighbor's noisy state from a background noise during a signal loss, a weight might be assigned to a pure noise term. This scenario may be formulated as a leader-following problem by adding an artificial leader node $A_0$ with a fixed zero state. An edge appears from $A_0$ to $A_i$ if and only if a signal loss occurs along $(k, i) \in \mathcal{E}_f$. It is of interest to identify conditions for convergence. In fact, for the case of i.i.d. losses, under mild noise conditions we may use the method of perturbed Lyapunov analysis in Huang and Manton (2010) to show that all the individual states will converge to zero, which is the state of the leader.

### 3.1. Assumptions

(A1) The digraph $G = (\mathcal{N}, \mathcal{E})$ is strongly connected. □

Denote the $\sigma$-algebra

$$\mathcal{F}_t = \sigma(x_0, \mathbf{w}_0, \dots, \mathbf{w}_t, g_0, \dots, g_t, I_0, \dots, I_{t+1}), \tag{7}$$

(i.e., the set of all events induced by these random variables) for $t \ge 0$. Then $\mathcal{F}_t \subset \mathcal{F}_{t+1}$. Define $\mathcal{F}_{-1} \triangleq \sigma(I_0)$.

(A2) The sequence $\{\mathbf{w}_t, t \in \mathbb{Z}^+\}$ satisfies the conditions: (i) $E[\mathbf{w}_t|\mathcal{F}_{t-1}] = 0$ for $t \ge 0$, and (ii) $\sup_{t\ge 0} E|\mathbf{w}_t|^2 < \infty$. In addition, $E|x_0|^2 < \infty$. □

Since $\mathbf{w}_t$ is adapted to $\mathcal{F}_t$, (A2) implies that $\{\mathbf{w}_t, t \in \mathbb{Z}^+\}$ is a sequence of martingale differences (see definition in Hall and Heyde (1980) and Stout (1974)) with bounded second moments. The following assumption with independent noises holds as a special case of (A2).

(A2°) The noises $\{w^{ik}_t, t \in \mathbb{Z}^+, i \in \mathcal{N}, k \in \mathcal{N}_i \neq \emptyset\}$ are independent w.r.t. $i, k, t$ and also independent of $x_0$ and the processes $\{I_t, t \ge 0\}$, $\{g_t, t \ge 0\}$. Each $w^{ik}_t$ has zero mean and variance $Q^{ik}_t$. In addition, $E|x_0|^2 < \infty$ and $\sup_{i,k,t} Q^{ik}_t < \infty$. □

(A3) The link gains $g^{ik}_t$ are mutually independent (w.r.t. $i, k, t$). Each $g_t$ is independent of $\{x_0, g_l, 0 \le l \le t-1, \mathbf{w}_k, I_k, 0 \le k \le t\}$. Furthermore, $Eg^{ik}_t = \lambda^{ik} \neq 0$, where $\lambda^{ik}$ does not depend on $t$, and $\sup_{i,k,t} E|g^{ik}_t|^2 < \infty$. □

**Remark.** Note that although in general fading channels are modelled as complex channels, due to the fading channel being a slow fading channel, the phase can be estimated and canceled, therefore the link amplitude gains $g^{ik}_t$ are positive and the link power gains are given by $|g^{ik}_t|^2$. □

(A4) The process $I_t$ is an ergodic Markov chain with stationary transition probability matrix $(p_{ij})_{1\le i,j\le K_0}$, and

$$P(I_{t+1} = j|I_t = i, I_0, \dots, I_{t-1}, \mathbf{w}_0, \dots, \mathbf{w}_t, g_0, \dots, g_t) = p_{ij}.$$

Moreover, $\min_{1\le i\le K_0} p_{i1} > 0$. □

**Remark.** If the Markov chains $I^{ik}_t$ are independent, $I_t$ is also a Markov chain. If, in addition, $P(I^{ik}_{t+1} = 1|I^{ik}_t = s) > 0$ for all $(i, k)$ regardless of $s$ being 0 or 1, the condition $\min_{1\le i\le K_0} p_{i1} > 0$ in (A4) is satisfied. □

(A5) For each $G^{(k)} \in \mathbf{G}$, $1 \le k \le K_0$, the associated matrix $B^{(k)}$ has zero row and column sums (ZRCS). □

(A6) (i) $a_t > 0$ for $t \ge 0$, and (ii) $\sum_{t=0}^{\infty} a_t = \infty$, $\sum_{t=0}^{\infty} a_t^2 < \infty$. □

### 3.2. Discussions on the ZRCS condition

In an average-consensus setting, the ZRCS condition for the coefficient matrices $B^{(1)}, \dots, B^{(K_0)}$ is quite standard. More specifically, in a deterministic average-consensus model

$$x_{t+1} = \overline{A}_t x_t,$$

where $\overline{A}_t$ has all row sums equal to one, the state average $(1/n)\sum_{i=1}^{n} x^i_t$ is an invariant if and only if all column sums of $\overline{A}_t$ are equal to one. In fact, $\overline{A}_t$ may even be allowed to have negative entries (Xiao et al., 2007). The reader is referred to Olfati-Saber and Murray (2004) for the notion of balanced graphs which preserve the initial state average as an invariant during averaging. Under (A5), $I + a_t B_t$ always has all row and column sums equal to one.

Although our current formulation will not lead to average-consensus due to the additive noise, it is possible to achieve approximate average-consensus when certain conditions are satisfied in terms of the noise level and the step size sequence, and this will be of practical interest.

## 4. Consensus results with Markovian switches

### 4.1. The regime dependent recursion

In algorithm (6), the right hand side depends on $\mathcal{N}_{it}$. To facilitate further analysis, we introduce a transformation so that it may be expressed in terms of $I_t$ instead of $\mathcal{N}_{it}$. Notice that the evolution of the network topology is completely characterized by $I_t$. We have the following relation

$$\begin{aligned} w^i_t &\triangleq \sum_{j\in\mathcal{N}_{it}} b_{ij}(t) w^{ij}_t (\lambda^{ij})^{-1} \\ &= \sum_{k=1}^{K_0} 1_{(I_t=k)} \sum_{j\in\mathcal{N}^{(k)}_i} b^{(k)}_{ij} w^{ij}_t (\lambda^{ij})^{-1} \\ &= \sum_{k=1}^{K_0} 1_{(I_t=k)} \sum_{j\in\mathcal{N}_i} b^{(k)}_{ij} w^{ij}_t (\lambda^{ij})^{-1}, \end{aligned} \tag{8}$$

where (8) holds since $\mathcal{N}^{(k)}_i \subset \mathcal{N}_i$ and $b^{(k)}_{ij} = 0$ for $j \in \mathcal{N}_j \backslash \mathcal{N}^{(k)}_j$. Define $w^{(I_t)}_t = [w^1_t, \dots, w^n_t]^T$. Under (A2)–(A3), we may use (8) and the fact that $I_t$ is adapted to $\mathcal{F}_{t-1}$ to obtain a very useful property

$$E[w^{(I_t)}_t|\mathcal{F}_{t-1}] = 0. \tag{9}$$

We further write

$$\delta b_{ij}(t) = \sum_{k=1}^{K_0} 1_{(I_t=k)} b^{(k)}_{ij} (g^{ij}_t/\lambda^{ij} - 1) \tag{10}$$

if $j \in \mathcal{N}_i$, and $\delta b_{ij}(t) = 0$ otherwise. Define the $I_t$ dependent matrix $\Delta B^{(I_t)} = (\delta b_{ij}(t))_{1\le i,j\le n}$.

We may further write (6) in the vector form

$$x_{t+1} = x_t + a_t B^{(I_t)} x_t + a_t \Delta B^{(I_t)} x_t + a_t w^{(I_t)}_t, \quad t \ge 0, \tag{11}$$

where $B^{(I_t)}$ and $w_t^{(I_t)}$ are determined from $G_t$. Since $x_t$ depends on $(x_0, \mathbf{w}_0, I_0, g_0, \dots, \mathbf{w}_{t-1}, I_{t-1}, g_{t-1})$, it follows from (A3) that $E[\Delta B^{(I_t)} x_t] = 0$. Thus the random channel gain contributes to the unbiased perturbation term $\Delta B^{(I_t)} x_t$ in (11).

Owing to the ZRCS condition for $B^{(k)}$, a state space decomposition technique may be applied for convergence analysis when the network topology randomly switches. This decomposition approach has been developed in models with fixed topologies containing a spanning tree (Huang & Manton, 2008, 2010); but for the models considered there, due to fixed topologies, the decomposition method is feasible without the ZRCS condition.

### 4.2. Change of coordinates and convergence

Let $1_n$ be a column vector with all $n$ entries equal to 1. By using Gram–Schmidt orthonormalization (Bellman, 1997), we may construct an orthogonal matrix of the form

$$\Phi = \left[\left(1/\sqrt{n}\right) 1_n, \phi\right], \tag{12}$$

where $\phi$ is an $n \times (n-1)$ matrix. Hence $\Phi^T \Phi = I$. The inverse of $\Phi$ may be represented in the form

$$\Phi^{-1} = \Phi^T = \begin{bmatrix} \left(1/\sqrt{n}\right) 1_n^T \\ \phi^T \end{bmatrix}. \tag{13}$$

We introduce the transformation

$$z_t = \Phi^{-1} x_t. \tag{14}$$

Denote $z_t = [z_t^1, \tilde{z}_t^T]^T$, and $v_t = [v_t^1, \tilde{v}_t^T]^T = \Phi^{-1} w_t^{(I_t)}$, where $z_t^1$ and $v_t^1$, are the first component in $z_t$ and $v_t$, respectively.

**Lemma 4.** *Suppose $B^{(k)}$ is defined by (4)–(5) and satisfies* (A5), *and $\Phi$ is given by (12). We have the assertions.*

(i) *For each $B^{(k)}$, we have*

$$\Phi^{-1} B^{(k)} \Phi = \begin{bmatrix} 0 & \\ & \tilde{B}^{(k)} \end{bmatrix}, \tag{15}$$

*where $\tilde{B}^{(k)}$ is an $(n-1) \times (n-1)$ matrix.*

(ii) *The matrix $(\tilde{B}^{(k)})^T + \tilde{B}^{(k)} \le 0$.*

(iii) *There exists a fixed constant $c_B > 0$ such that $y^T[\tilde{B}^{(k)} + (\tilde{B}^{(k)})^T]y \le -c_B|y|^2$ for all $y \in \mathbb{R}^{n-1}$ if the associated digraph $G^{(k)}$ is strongly connected.*

**Proof.** (i) By using the ZRCS property of $B^{(k)}$, we may verify the relation (15) directly and in fact, $\tilde{B}^{(k)} = \phi^T B^{(k)} \phi$, where $\phi$ is given in (12).

(ii) Following the proof of Theorems 7 and 8 in Olfati-Saber and Murray (2004), we define the adjacency matrix $A^{(k)} = (a_{ij})_{n\times n}$ for $G^{(k)}$ such that $a_{ij} = b_{ij}^{(k)}$ for $i \neq j$, and $a_{ii} = 0$ for all $i$. Then $-B^{(k)}$ may be identified as a Laplacian for $G^{(k)}$ (w.r.t. the adjacency matrix $A^{(k)}$). Subsequently, by Theorem 7 in Olfati-Saber and Murray (2004), $-B^{(k)} - (B^{(k)})^T$ may be interpreted as the Laplacian of a weighted undirected graph $\hat{G}$. Then (ii) follows easily.

(iii) Again, following Olfati-Saber and Murray (2004), when $G^{(k)}$ is strongly connected, $-B^{(k)} - (B^{(k)})^T$ may be interpreted as the Laplacian of a weighted undirected and connected graph $\hat{G}$, so that the nonnegative definite matrix $-B^{(k)} - (B^{(k)})^T$ has its null space equal to span$\{1_n\}$, which implies (iii). □

An alternative method for proving Lemma 4(ii)–(iii) is to interpret $B^{(k)} + (B^{(k)})^T$ as the generator of a continuous time Markov chain, which is ergodic when $G^{(k)}$ is strongly connected.

By (15), we may write

$$\Phi^{-1} B^{(I_t)} \Phi = \begin{bmatrix} 0 & \\ & \tilde{B}^{(I_t)} \end{bmatrix}. \tag{16}$$

We also denote $\tilde{B}_t = \tilde{B}^{(I_t)}$.

By (13) and (16), Eq. (11) may be rewritten as

$$z_{t+1}^1 = z_t^1 + a_t \left[\left(1/\sqrt{n}\right) 1_n^T \Delta B^{(I_t)} \Phi\right] z_t + a_t v_t^1, \tag{17}$$

$$\tilde{z}_{t+1} = \tilde{z}_t + a_t \tilde{B}_t \tilde{z}_t + a_t [\phi^T \Delta B^{(I_t)} \Phi] z_t + a_t \tilde{v}_t, \tag{18}$$

where $t \ge 0$. Notice that $v_t^1$ and $\tilde{v}_t$ both depend on $I_t$.

**Lemma 5** (Tsitsiklis et al., 1986)**.** *Suppose the two sequences of nonnegative random variables $\{\xi_t,\ t \ge 0\}$ and $\{\xi_t',\ t \ge 0\}$ are both adapted to the increasing sequence of $\sigma$-algebras $\{\mathscr{G}_t,\ t \ge 0\}$, and for $t \ge 0$,*

$$E[\xi_{t+1}|\mathscr{G}_t] \le \xi_t + \xi_t', \quad \sum_{t=0}^{\infty} E\xi_t' < \infty.$$

*Then $\xi_t$ converges a.s. to a random variable $\xi_\infty$.* □

**Theorem 6.** *Under* (A1)–(A6), *algorithm (11) ensures both mean square and strong consensus.*

**Proof.** See Appendix. □

We give some discussion on the relation between the individual limit states $x_\infty^1 = x_\infty^2 = \cdots = x_\infty^n$ and the initial state average. Denote $\mathrm{Ave}(x_t) = (1/n) 1_n^T x_t$. By (11) it follows that

$$\mathrm{Ave}(x_{t+1}) = \mathrm{Ave}(x_t) + a_t (1/n) 1_n^T \Delta B^{(I_t)} x_t + a_t (1/n) 1_n^T w_t^{(I_t)}.$$

Hence for each $k$,

$$x_\infty^k - \mathrm{Ave}(x_0) = (1/n) \sum_{t=0}^{\infty} [a_t 1_n^T \Delta B^{(I_t)} x_t + a_t 1_n^T w_t^{(I_t)}], \tag{19}$$

where the right hand side converges in mean square. One may reduce the deviation of $x_\infty^k$ from $\mathrm{Ave}(x_0)$ by using a small step size sequence, but this may also decrease the convergence rate. Thus, it is practically important to have a good trade-off between controlling this deviation and convergence rate.

### 4.3. Relaxation of the connectivity condition

A further relaxation of the condition for the Markov chain $I_t$ and also network connectivity is possible. In general, when $I_t$ is only ergodic and there are no instances of strongly connected real-time network topologies, unlike what is shown in the proof of Theorem 6 there is no guarantee that after one step the energy function $E|\tilde{z}_t|^2$, up to some higher order perturbation, will decay by the rate $(1 - c_0 a_t)$ for some fixed $c_0 > 0$. In this situation, a useful strategy is to compare the energy function between $l_0(t+1)$ and $l_0 t$ for an appropriately chosen $l_0 \ge 1$. This requires the so-called joint connectivity condition and a better behaved sequence of step sizes than merely assuming (A6).

In this subsection, we will reuse some notation previously introduced and they in general take new values. This should cause no risk of confusion. Denote the collection of digraphs

$$\mathbf{G} = \{G^{(1)}, \dots, G^{(K_0')}\}, \tag{20}$$

where $K_0' \ge 1$ indicates the number of network topologies which may occur. Denote $G^{(i)} = (\mathscr{N}, \mathscr{E}^{(i)})$. Now we take

$$G = (\mathscr{N}, \mathscr{E}) \triangleq (\mathscr{N}, \cup_{i=1}^{K_0'} \mathscr{E}^{(i)}) \tag{21}$$

as the union graph of $G^{(1)}, \dots, G^{(K_0')}$.

Suppose now the Markov chain $I_t$ has the state space $\{1, \dots, K_0'\}$. If $I_t = i$, the real-time network topology is determined as $G_t = G^{(i)}$. Let the matrix $B^{(i)}$ still be constructed by the rule (4)–(5) when the associated graph is $G^{(i)}$.

As our convention, the link gain $g_t^{jk}$ and noise $w_t^{jk}$ are always defined as long as $(k, j) \in \cup_{i=1}^{K_0'} \mathscr{E}^{(i)}$. When $(k, j) \notin G_t$, $g_t^{jk}$ and $w_t^{jk}$ are introduced as dummy random variables. The vectors $\mathbf{w}_t$ and $g_t$ are formed accordingly as in Section 2.

We rewrite the associated algorithm

$$x_{t+1} = x_t + a_t B^{(I_t)} x_t + a_t \Delta B^{(I_t)} x_t + a_t w_t^{(I_t)}, \quad t \ge 0, \tag{22}$$

where the determination of $\Delta B^{(I_t)}$ and $w_t^{(I_t)}$ is in parallel to that in (11). Now we make the assumption:

(A4′) $\{I_t, \ t \ \ge \ 0\}$ is an ergodic Markov chain with stationary transition probability matrix $(p'_{ij})_{1 \le i,j \le K_0'}$, and

$$P(I_{t+1} = j | I_t = i, I_0, \ldots, I_{t-1}, \mathbf{w}_0, \ldots, \mathbf{w}_t, g_0, \ldots, g_t) = p'_{ij}. \quad \square$$

**Theorem 7.** *Assume*

(i) (A1) *holds with* $G$ *defined by* (21) *(joint connectivity);*
(ii) (A2)–(A3), (A4′) *and* (A5) *hold;*
(iii) *in addition to* (A6), *there exist* $0 < \alpha_1 < \alpha_2$ *such that*

$$\alpha_1 a_t \le a_{t+1} \le \alpha_2 a_t, \quad t \ge 0. \tag{23}$$

*Then algorithm* (22) *ensures mean square and strong consensus.*

**Proof.** See Appendix. $\square$

## 5. Models with arbitrary switches

In this section, we consider arbitrary random switches. In contrast to Sections 2 and 3, here we do not start from the statistical modelling of individual communication links although the coefficient matrix $B_t$ used below may be interpreted via an associated digraph $G_t = (\mathscr{N}, \mathscr{E}_t)$.

Let the algorithm be given as

$$x_{t+1} = x_t + a_t B_t x_t + a_t w_t, \quad t \ge 0, \tag{24}$$

where $x_t$ is in $\mathbb{R}^n$, $w_t$ is the additive noise and $B_t$ takes values from $\{B^{(1)}, \ldots, B^{(k_0)}\}$ for some $k_0 \ge 1$. The matrix $B_t = (b_{ij}(t))$ has zero row sums and nonnegative off-diagonal entries. It is not required to satisfy the ZRCS condition.

Define the class of symmetric matrices: $\mathscr{D} = \{D | D \ge 0, \mathrm{Null}(D) = 1_n\}$. Then $\mathscr{D}$ is a cone, i.e., for any $D_1 \in \mathscr{D}$ and $D_2 \in \mathscr{D}$, we have $\alpha D_1 \in \mathscr{D}$ and $\alpha D_1 + \beta D_2 \in \mathscr{D}$ for all $\alpha > 0$ and $\beta > 0$.

(A7) There exists $Q \in \mathscr{D}$ such that the cone conditions

$$-\{(B^{(i)})^T Q + Q B^{(i)}\} \in \mathscr{D}, \quad i = 1, \ldots, k_0, \tag{25}$$

are simultaneously satisfied by $B^{(i)}, \ i = 1, \ldots, k_0$. $\square$

For a fixed $B^{(i)}$, the following lemma gives a characterization of condition (25) in terms of an associated network topology when the rule (4)–(5) is used for the construction of $B^{(i)}$.

**Lemma 8.** *Suppose* (i) $B = (b_{ij})_{1 \le i,j \le n}$ *has zero row sums and nonnegative off-diagonal entries, and* (ii) $B$ *is obtained from a digraph* $\overline{G}$ *such that* $b_{ij} > 0, j \ne i$, *if and only if* $(j, i)$ *is an edge in* $\overline{G}$. *Then given any* $D \in \mathscr{D}$, *the equation*

$$B^T Q + Q B = -D \tag{26}$$

*has a unique solution* $Q \in \mathscr{D}$ *if and only if* $\overline{G}$ *contains a spanning tree. The solution* $Q$, *if existing, has the representation* $Q = \int_0^\infty e^{B^T t} D e^{Bt} dt$.

**Proof.** See Appendix. $\square$

### 5.1. An example

Let $G^{(k)}, k = 1, 2, 3$, be digraphs with the same set of nodes $\mathscr{N} = \{1, 2, 3\}$. Suppose $G^{(1)}$ has the edges $\mathscr{E} = \{(1, 2), (1, 3), (2, 1), (3, 2)\}$. We take the weight matrix $B^{(1)}$ and $D \in \mathscr{D}$ as follows

$$B^{(1)} = \begin{bmatrix} -1 & 1 & 0 \\ 0.5 & -1 & 0.5 \\ 1 & 0 & -1 \end{bmatrix}, \qquad D = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

For equation $(B^{(1)})^T Q + Q B^{(1)} = -D$, the solution is

$$Q = \int_0^\infty e^{(B^{(1)})^T t} D e^{B^{(1)} t} dt = \frac{1}{30} \begin{bmatrix} 17 & -8 & -9 \\ -8 & 22 & -14 \\ -9 & -14 & 23 \end{bmatrix}. \tag{27}$$

The eigenvalues of $B^{(1)}$ are $0, -1.5 \pm 0.5\mathbf{i}$, and the eigenvalues of $Q$ are 0, 0.847741 and 1.218925.

Let $G^{(2)}$ have the edges $\mathscr{E} = \{(1, 2), (1, 3), (2, 1)\}$, and $G^{(3)}$ have the edges $\mathscr{E} = \{(1, 2), (2, 1), (3, 2)\}$. Let the corresponding weight matrices be given by

$$B^{(2)} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}, \qquad B^{(3)} = \begin{bmatrix} -1 & 1 & 0 \\ 0.5 & -1 & 0.5 \\ 0 & 0 & 0 \end{bmatrix}.$$

For $Q$ given by (27), it may be verified that

$$-\{(B^{(k)})^T Q + Q B^{(k)}\} \in \mathscr{D}$$

for $k = 2, 3$. Thus, for $B^{(k)}$, $k = 1, 2, 3$, (A7) is satisfied.

### 5.2. Convergence result

Denote $\mathscr{F}_t = \sigma(x_0, w_k, B_k, k \le t)$.

**Theorem 9.** *Assume* (i) (A6)–(A7) *hold, and* (ii) $E|x_0|^2 < \infty$, $\{w_t, t \ge 0\}$ *is a sequence of martingale differences w.r.t. the* $\sigma$-*algebras* $\mathscr{F}_t$ *and* $\sup_{t \ge 0} E|w_t|^2 < \infty$. *Then algorithm* (24) *ensures mean square consensus.*

**Proof.** Letting $Q$ be given by (A7), we have

$$\begin{aligned} x_{t+1}^T Q x_{t+1} &= (x_t + a_t B_t x_t + a_t w_t)^T Q (x_t + a_t B_t x_t + a_t w_t) \\ &= x_t^T Q x_t + a_t x_t^T (B_t^T Q + Q B_t) x_t + a_t (w_t^T Q x_t \\ &\quad + x_t^T Q w_t) + a_t^2 x_t^T B_t^T Q B_t x_t + a_t^2 w_t^T Q w_t. \end{aligned} \tag{28}$$

Given $D_1$ and $D_2$, both in $\mathscr{D}$, since they have the same null space $\mathrm{span}\{1_n\}$, by elementary linear algebra it may be shown that there exist two constants $0 < c_1 \le c_2$ such that

$$c_1 D_1 \le D_2 \le c_2 D_1. \tag{29}$$

See Huang and Manton (2009) for similar estimates. Since (A7) holds and $B_t$ takes values from a finite set, we may find $c_3 > 0$ such that

$$B_t^T Q + Q B_t \le -c_3 Q \tag{30}$$

for all $t$. Furthermore, we may find $c_4 > 0$ such that

$$B_t^T Q B_t \le c_4 Q. \tag{31}$$

Combining (28) with (30) and (31), we may find a sufficiently large $T_1$ such that for all $t \ge T_1$ we have

$$E[x_{t+1}^T Q x_{t+1} | \mathscr{F}_t] \le (1 - \tau a_t) x_t^T Q x_t + C a_t^2 E[|w_t|^2 | \mathscr{F}_t]$$

for some $\tau > 0$ and $C > 0$, which further implies

$$E x_{t+1}^T Q x_{t+1} \le (1 - \tau a_t) E x_t^T Q x_t + C a_t^2 E|w_t|^2.$$

Then by (A6), we obtain $\lim_{t \to \infty} E x_t^T Q x_t = 0$, and using the method in Huang and Manton (2009, Section 5), we further obtain weak consensus.

Next, we show mean square convergence of $x_t$. Our method is to show that $x_t$ is a fundamental sequence under the norm $\|x_t\| = (E|x_t|^2)^{1/2}$. Let $\varepsilon > 0$ be any given small constant. Define

$\Pi_{k,i} = \prod_{j=i+1}^{k}(I + a_j B_j)$, where $k > i$ and $I + a_{i+1}B_{i+1}$ is the most right term in the successive matrix product. We denote $\Pi_{k,k} = I$. Select $t_0 > 0$. For $t \geq t_0$, we have

$$x_{t+1} = \Pi_{t,t_0-1}x_{t_0} + \sum_{k=t_0}^{t} \Pi_{t,k}a_k w_k. \tag{32}$$

By weak consensus, there exists $T_0 \geq 0$ such that for all $t, t_0$ satisfying $t \geq t_0 \geq T_0$, we have $E|x_{t_0}^1 - x_{t_0}^k|^2 \leq \varepsilon$, where $2 \leq k \leq n$. Without loss of generality we may assume that $T_0$ is sufficiently large such that each $I + a_j B_j$ is a nonnegative matrix and hence a stochastic matrix for all $j \geq T_0$. Then for all $t \geq t_0 \geq T_0$, $\Pi_{t,t_0-1}$ is a stochastic matrix for any given sample $\omega \in \Omega$. We have

$$\begin{aligned} \Pi_{t,t_0-1}x_{t_0} - x_{t_0} &= [\Pi_{t,t_0-1} - I][x_{t_0}^1 1_n + x_{t_0} - x_{t_0}^1 1_n] \\ &= [\Pi_{t,t_0-1} - I][x_{t_0} - x_{t_0}^1 1_n], \end{aligned} \tag{33}$$

where $x_{t_0}^1$ is the first component in $x_{t_0}$. By (33), we have

$$E|\Pi_{t,t_0-1}x_{t_0} - x_{t_0}|^2 \leq n^3\varepsilon,$$

for $t \geq t_0 \geq T_0$. Then by (32) it is straightforward to find a fixed constant $C_0$ independent of $(t, t_0)$ such that

$$E|x_{t+1} - x_{t_0}|^2 \leq n^3\varepsilon + C_0 \sum_{k=t_0}^{t} a_k^2.$$

Since $\sum_{k=0}^{\infty} a_k^2 < \infty$, there exists a sufficiently large $T_1$ such that for all $t_0 \geq T_1$, we have $C_0 \sum_{k=t_0}^{t} a_k^2 \leq \varepsilon$. Hence, for all $t \geq t_0 \geq T_0 \vee T_1$, we have $E|x_{t+1} - x_{t_0}|^2 \leq (n^3 + 1)\varepsilon$. Since $\varepsilon > 0$ is arbitrary, there exists a random variable $x_\infty$ such that $\lim_{t\to\infty} E|x_t - x_\infty|^2 = 0$, which combined with weak consensus implies mean square consensus. □

## 6. Application to networks with quantized data and packet losses

In this section, we consider models with quantized data and packet losses. In particular, we will apply probabilistic quantization, which is effective in eliminating bias when the agents exchange state information by rate-limited digital communication channels. This approach is recently introduced into consensus problems in Aysal et al. (2007), and it is also applied in sensor networks for data fusion with quantized information (Krasnopeev, Xiao, & Luo, 2005; Xiao, Cui, Luo, & Goldsmith, 2006). We assume that packets are transmitted from each node to its neighbors via Markovian lossy channels.

Let the network topology be described by an *undirected* graph $G$. For easing the distributed construction of the weight matrix for averaging, we restrict our attention to undirected graphs. At time $t$, denote the state of node $i$ by $x_t^i$. Let $\mathscr{E}_f$ be the set of failure-prone links. At time $t$, the instantaneous network topology is denoted by $G_t$, as a subgraph of $G$, and the neighbor set of node $i$ is denoted by $\mathscr{N}_{it}$. When $(k, i) \in \mathscr{E}_f$, denote the channel state between node $k$ and $i$ by a Markov chain $I_t^{ki}$ taking values from $\{0, 1\}$, where $k \in \mathscr{N}_i$. The transition probability matrix of $I_t^{ki}$ is $(p_{lm}^{ki})_{1\leq l,m\leq 2}$, where $P(I_{t+1}^{ki} = 0|I_t^{ki} = 0) = p_{11}^{ki}$. For $k \in \mathscr{N}_i$, it is in $\mathscr{N}_{it}$ if and only if $I_t^{ki} = 1$. If $I_t^{ki} = 0$, a packet loss occurs. Again, we form the process $I_t$ by stacking all $I_t^{ki}$, where $(k, i) \in \mathscr{E}_f$.

At node $i$, suppose the real line $\mathbb{R}$ is partitioned by the quantization levels $r_{i,k}$, where $k \in \mathbb{Z}$ (the set of all integers), and $r_{i,k} < r_{i,k+1}$ for all $k$. Since the set of quantization levels differ from node to node, this gives heterogeneous quantizers.

### 6.1. A heuristic description of probabilistic quantizers

To better convey the idea of generating unbiased quantization errors, we give a heuristic description of probabilistic quantization first. If node $i$ observes $x_t^i \in (r_{i,k}, r_{i,k+1}]$, the "randomized" output $Q_i(t)$ of the quantizer is a random variable taking $r_{i,k}$ and $r_{i,k+1}$, respectively, with probabilities

$$p_{r_{i,k}} = (r_{i,k+1} - x_t^i)/(r_{i,k+1} - r_{i,k}), \tag{34}$$

$$p_{r_{i,k+1}} = (x_t^i - r_{i,k})/(r_{i,k+1} - r_{i,k}). \tag{35}$$

At time $t$, if $(k, i) \in \mathscr{E}\backslash\mathscr{E}_f$, or $(k, i) \in \mathscr{E}_f$ but $I_t^{ki} = 1$, node $i$ obtains the data

$$y_t^{ik} = Q_k(t) = x_t^k + [Q_k(t) - x_t^k].$$

We write

$$y_t^{ik} = x_t^k + w_t^k, \quad k \in \mathscr{N}_{it}, \tag{36}$$

where $w_t^k \triangleq Q_k(t) - x_t^k$ is the quantization error. If $I_t^{ki} = 0$ for $(k, i) \in \mathscr{E}_f$, then $y_t^{ik} \equiv 0$. Now (36) may be viewed as a special case of (1)–(3) if we take $g_t^{ik} \triangleq 1$, $w_t^{ik} \triangleq w_t^k$.

Corresponding to (6), we apply the algorithm:

$$x_{t+1}^i = [1 + a_t b_{ii}(t)]x_t^i + a_t \sum_{k\in\mathscr{N}_{it}} b_{ik}(t)y_t^{ik}, \quad t \geq 0. \tag{37}$$

The weight matrix $B_t = (b_{ij}(t))$ is now constructed using the Metropolis weights as follows

$$b_{ij}(t) = \begin{cases} \dfrac{1}{\max\{d_i(t), d_j(t)\} + 1} & \text{if } (i,j) \in \mathscr{E}_t, \\ -\sum\limits_{k\in\mathscr{N}_{it}} b_{ik}(t) & \text{if } j = i, \\ 0 & \text{otherwise}, \end{cases} \tag{38}$$

where $d_i(t)$ is the number of neighbors of node $i$. So $B_t$ takes values from a finite set and satisfies the ZRCS condition.

### 6.2. Specification of the stochastic recursion

Below we focus on the central issue of how to characterize $Q_i(t)$, so that the quantization error has desired properties for ensuring convergence of algorithm (37). To have a rigorous specification of the probabilistic quantizer, we need to ensure that $(x_t, I_t, Q(t))$ corresponds to a well defined random process, where $Q(t) = [Q_1(t), \ldots, Q_n(t)]$.

From the point of view of implementation, it will be convenient to introduce the following new random variables $Z_t^i$. Once $x_t^i \in [r_{i,k}, r_{i,k+1})$ is observed at node $i$, a random variable $Z_t^i$ is generated, uniformly distributed on $[r_{i,k}, r_{i,k+1})$. Let $F_{[r_{i,k},r_{i,k+1})}(y)$, $y \in \mathbb{R}$, denote the distribution function of the uniform distribution on $[r_{i,k}, r_{i,k+1})$. Denote $Z_t = [Z_t^1, \ldots, Z_t^n]^T$. Let $H_t$ denote an event in the $\sigma$-algebra $\sigma(x_k, I_k, Z_{k-1}, k \leq t)$. We specify the conditional probability distribution by the following rule

$$\begin{aligned} P\Big(&Z_t^i \leq z^i, I_{t+1}^{ji} = s'_{ji}, \text{ all } i \in \mathscr{N}, \text{ all } j \in \mathscr{N}_i \,|H_t, x_t^i \in [r_{i,k}, r_{i,k+1}), \\ &I_t^{ji} = s_{ji}, \text{ all } i \in \mathscr{N}, \text{ all } j \in \mathscr{N}_i\Big) \\ &= \Bigg(\prod_{i\in\mathscr{N}, j\in\mathscr{N}_i} p_{s_{ji},s'_{ji}}^{ji}\Bigg) \prod_{i\in\mathscr{N}} F_{[r_{i,k},r_{i,k+1})}(z^i), \end{aligned}$$

when $P(H_t \cap \{x_t^i \in [r_{i,k}, r_{i,k+1}), I_t^{ji} = s_{ji}, \text{ all } i \in \mathscr{N}, j \in \mathscr{N}_i\}) > 0$. In the above, $s_{ji}$ and $s'_{ji}$ take values from $\{0, 1\}$. Then it is

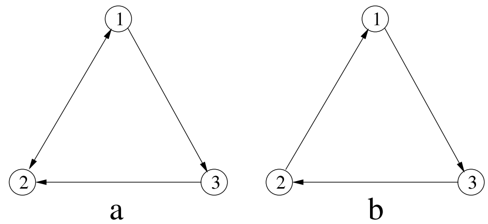



**Fig. 2.** The network topology.

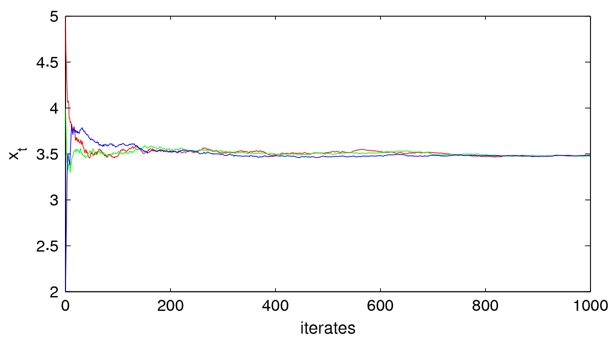



**Fig. 3.** Convergence with analog channels and Markovian switches.

straightforward to determine the joint distribution of $x_k, I_{k+1}, Z_k$ for $k \le t$. Next, we set $Q_t^i$ as follows. If $x_t^i \in [r_{i,k}, r_{i,k+1})$ and $Z_t^i \le x_t^i$, we define $Q_t^i = r_{i,k+1}$, and otherwise, $Q_t^i = r_{i,k}$. After generating $Q_t^i$ in this manner, we use (37) to generate $x_{t+1}$ as a function of $x_t, I_t, Q_t^i, i \in \mathcal{N}$. So the joint distribution of $x_k, I_k, Z_{k-1}, k \le t+1$, is determined. Thus, by starting with $x_0, I_0$, we may construct a well defined process $(x_t, I_t, Z_t, Q_t)$ in a suitable probability space. Subsequently, we may show that $\{w_t^k, t \ge 0\}$ forms a sequence of martingale differences, which is desired for eliminating the bias of the quantization error.

Define the quantization resolution parameter for node $i$ by

$$\Delta_i = \sup_{-\infty<k<\infty} |r_{i,k+1} - r_{i,k}|.$$

We make the following assumptions.

(A8) For each $i \in \mathcal{N}$, $\lim_{k\to-\infty} r_{i,k} = -\infty$ and $\lim_{k\to\infty} r_{i,k} = \infty$. Moreover, $\max_{i\in\mathcal{N}} \Delta_i < \infty$. □

(A9) The Markov chains $I_t^{ki}$, $(k,i) \in \mathcal{E}_f$, are independent and $\min_{s\in\{0,1\}} P(I_{t+1}^{ki} = 1 | I_t^{ki} = s) > 0$ for all $(k, i)$. □

We state the following corollary on convergence.

**Corollary 10.** *Suppose $G$ is a connected undirected graph, $E|x_0|^2 < \infty$, and* (A6), (A8)–(A9) *hold. Then algorithm* (37) *achieves both mean square and strong consensus.*

**Proof.** Denote $w_t = [w_t^1, \dots, w_t^n]^T$ for the quantization errors. Under (A8), $\{w_t, t \ge 0\}$ forms a sequence of martingale differences w.r.t. the $\sigma$-algebras $\mathcal{F}_t = \sigma(x_0, w_0, \dots, w_t, I_0, \dots, I_{t+1})$ and has bounded second moments. Hence the corollary follows from Theorem 6. □

## 7. Simulations

### 7.1. Simulation with analog channels

Let the network topology with the maximal set of communication links be denoted by $G \triangleq G^{(1)}$ as shown in Fig. 2(a). The edge $(1, 2)$ in $G^{(1)}$ is failure-prone, and the Markov chain $I_t^{12}$ has the transition probability matrix

$$P = (p_{ij})_{1\le i,j\le 2} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix},$$

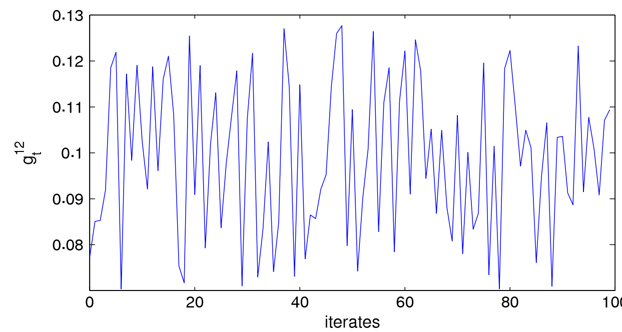



**Fig. 4.** The channel gain $g_t^{12}$.

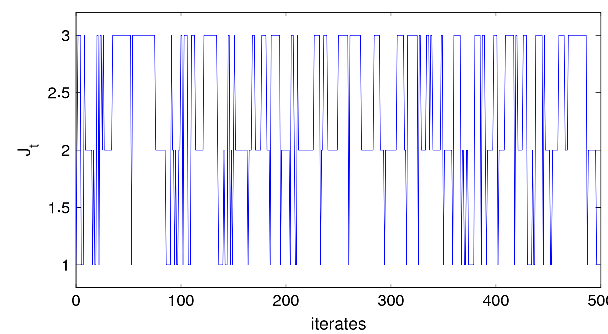



**Fig. 5.** Switches of the coefficient matrices.

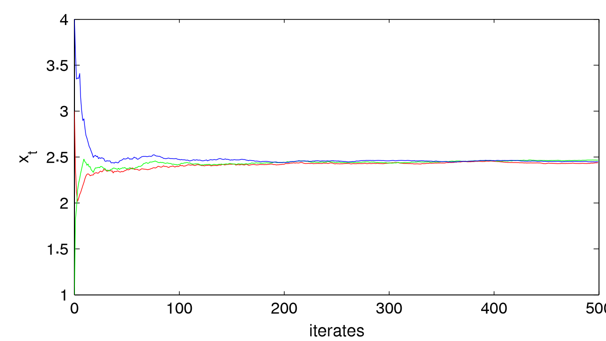



**Fig. 6.** Convergence of the 3 trajectories with arbitrary switching.

where $p_{11} = P(I_{t+1}^{12} = 0 | I_t^{12} = 0)$ and $p_{22} = P(I_{t+1}^{12} = 1 | I_t^{12} = 1)$, etc. The choice of these parameters in $P$ suggests relatively slow switches between the failure and recovery of the link. When $I_t^{12} = 0$, let the resulting graph be denoted by $G^{(2)}$; see Fig. 2(b). In the output equation

$$\bar{y}^{ik} = g_t^{ik} x_t^k + w^{ik},$$

all link gains $g_t^{ik}$ are i.i.d. and uniformly distributed on the interval $[0.07, 0.13]$. All i.i.d. Gaussian noises $w_t^{ik}$ have zero mean and variance $\sigma_w^2 = 10^{-4}$.

For $G^{(1)}$ and $G^{(2)}$, the associated coefficient matrices for averaging are, respectively, given by

$$B^{(1)} = \begin{bmatrix} -1 & 1 & 0 \\ 0.5 & -1 & 0.5 \\ 0.5 & 0 & -0.5 \end{bmatrix}, \qquad B^{(2)} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix},$$

which satisfy the ZRCS condition. By applying algorithm (6), the convergence behavior of $x_t$ is shown in Fig. 3, where the three trajectories converge to approximately 3.51. The associated channel gain process $g_t^{12}$ is displayed in Fig. 4 for the first 100 steps.

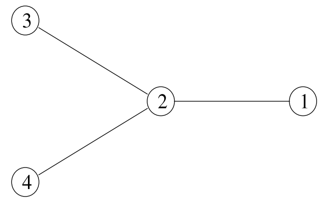



**Fig. 7.** The undirected graph for the maximal set of communication links.

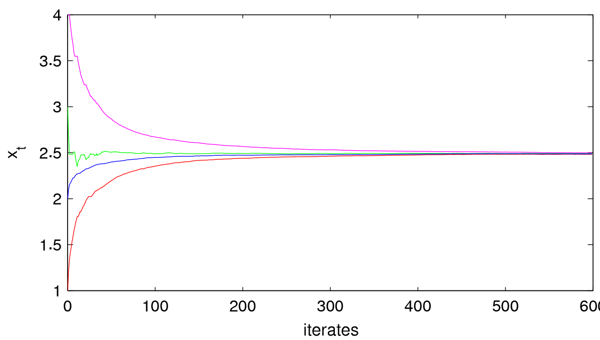



**Fig. 8.** Convergence with quantized data and decreasing step sizes.

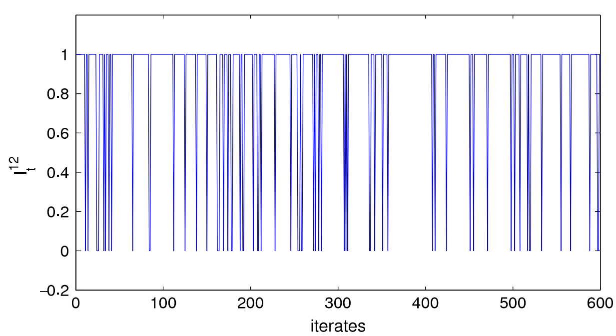



**Fig. 9.** The packet loss process $I_t^{12}$ along the edge $(1, 2)$.

## 7.2. Simulation with arbitrary switches

We take $B^{(1)}$, $B^{(2)}$ and $B^{(3)}$ given in Section 5.1. Algorithm (24) is applied where $w_t$ is an i.i.d. Gaussian noise process with covariance matrix $0.04I_{3\times 3}$, where $I_{3\times 3}$ is the identity matrix. The initial state vector $x_0 = [3, 1, 4]^T$. The step size $a_t = 1/(t+2)^{0.85}$ for $t \geq 0$. In the simulation, for a simple generation of the arbitrary switches of $B_t$, they are mimicked by a sample path of a Markov chain $J_t$ taking values from $\{1, 2, 3\}$ so that $B_t = B^{(J_t)}$. The trajectory of $J_t$ is displayed in Fig. 5. The convergence of $x_t$ is shown in Fig. 6.

## 7.3. Simulation with quantized data and packet losses

The network topology is modelled as the undirected graph $G$ shown in Fig. 7. Each link in $G$ is subject to packet losses. The loss processes $I_t^{ik}$ are modelled by 3 independent Markov processes, with initial states $I_0^{12} = 1$, $I_0^{23} = 1$ and $I_0^{24} = 0$, respectively. Each of the 3 Markov chains has the transition probability matrix

$$P = \begin{bmatrix} 0.2 & 0.8 \\ 0.1 & 0.9 \end{bmatrix}. \tag{39}$$

So $P(I_{t+1}^{12} = 0 | I_t^{12} = 0) = 0.2$, etc. We take $x_0 = [1, 3, 2, 4]^T$.

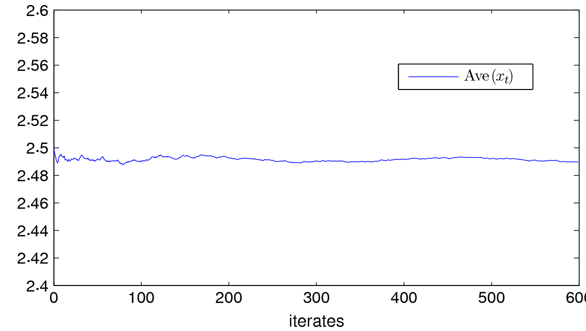



**Fig. 10.** The average state $\mathrm{Ave}(x_t) \triangleq (1/4)\sum_{i=1}^{4} x_t^i$.

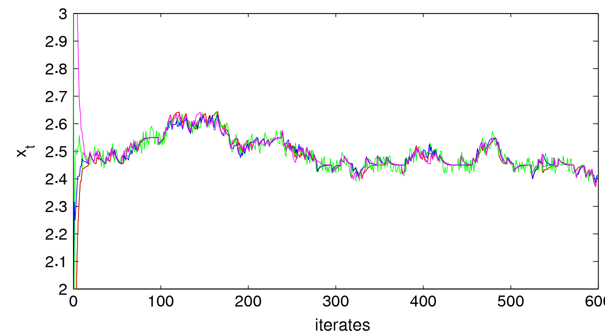



**Fig. 11.** Divergence of $x_t$ for the 4 agents with probabilistic quantizers and fixed step sizes.

For implementing probabilistic quantizers, the two sets of quantization levels are

$$\mathbf{L}_1 = \{(k+1/2)\Delta : k \in \mathbb{Z}\}, \qquad \mathbf{L}_2 = \{k\Delta : k \in \mathbb{Z}\}, \tag{40}$$

where the constant $\Delta > 0$ is the quantization resolution parameter and $\mathbb{Z}$ denotes all integers. When $\mathbf{L}_1$ is used, we have the partition $\mathbb{R} = \cup_{-\infty<k<\infty}[(k-1/2)\Delta,\ (k+1/2)\Delta)$. In parallel, when $\mathbf{L}_2$ is used, $\mathbb{R} = \cup_{-\infty<k<\infty}[k\Delta,\ (k+1)\Delta)$.

In the simulation, we take $\Delta = 0.1$. We use $\mathbf{L}_1$ at nodes 1 and 2, and $\mathbf{L}_2$ at nodes 3 and 4. Fig. 8 shows a convergence behavior when the step size $a_t = 1/(t+4)^{0.65}$ is used. The packet loss process $I_t^{12}$ is shown in Fig. 9. Fig. 10 shows that the state average is maintained within a small neighborhood of the initial state average $(1+3+2+4)/4 = 2.5$.

Finally, for comparison, Fig. 11 shows divergence of $x_t$ when the same set of probabilistic quantizers are applied with the constant step size $a_t \equiv 1$, and in this case, the resulting coefficient matrix $I + a_t B_t = I + B_t$ reduces to the standard Metropolis weights. This divergence behavior is in sharp contrast to the convergence behavior observed in Aysal et al. (2008), where a fixed weight coefficient matrix and homogenous quantizers, which have the same set of quantization levels, are used at all nodes.

## 7.4. A large random network

Let $n = 100$ nodes be independently and uniformly distributed in a unit square; see Fig. 12. Each node has a sensing radii of 0.25, which further determines the underlying network topology as an undirected graph $G$. Each node selects $\mathbf{L}_1$ and $\mathbf{L}_2$ in (40) with equal probability, where $\Delta = 0.1$, and then retains the selected quantizer.

Suppose each link in $G$ is described by a Markov chain with transition probability matrix $P$ given in (39), and these Markov chains are stationary and independent.

In the simulation, the initial states $x^i(0)$ are i.i.d. Gaussian $N(\mu, \sigma^2)$ with $\mu = 6$ and $\sigma^2 = 4$. In the initialization, the

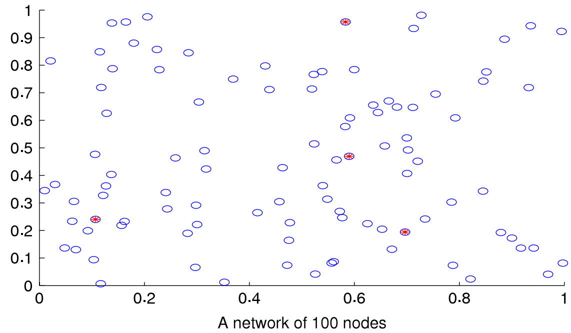



**Fig. 12.** The randomly distributed nodes with a sensing radii of 0.25.

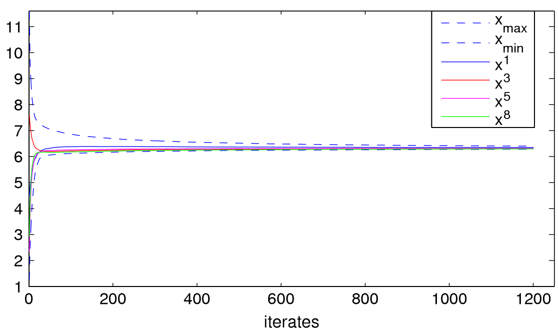



**Fig. 13.** The trajectory of 4 representative agents and the upper/lower envelope of the 100 state trajectories.

minimum and maximum values of $x^i(0)$ are respectively, 1.244132 and 11.573608. The initial state average is $\mathrm{Ave}(x(0)) = (1/100)\sum_{i=1}^{100} x^i(0) = 6.323485$. We run algorithm (37) for 1200 iterates with $a_t = 2(t+31)^{-0.65}$. At time $t$, denote the empirical mean square error (with respect to $\mathrm{Ave}(x(0))$) by

$$\mathrm{MSE}(t) = \frac{1}{N}\sum_{i=1}^{N}[x^i(t) - \mathrm{Ave}(x(0))]^2,$$

which measures the deviation of $x^i(t)$'s from the initial state average. We have $\mathrm{MSE}(0) = 4.898526$. Denote $x_{\min}(t) = \min_i x^i(t)$ and $x_{\max}(t) = \max_i x^i(t)$, which, respectively, give the lower and upper envelopes of all agents' trajectories. Fig. 13 shows the trajectories of $x_{\min}, x_{\max}, x^1, x^3, x^5$ and $x^8$. These four representative nodes are marked by "$*$" and displayed in Fig. 12. At the terminal time $T = 1200$, $x_{\min}(1200) = 6.283269$, $x_{\max}(1200) = 6.401205$, $\mathrm{MSE}(1200) = 9.8 \times 10^{-4}$.

A large number of repeats of the simulation show similar convergence behavior. The MSE at $T = 1200$ has noticeable variability, but it is generally at the order of $10^{-3}$ or even smaller. Also, it is observed that $\mathrm{MSE}(T)$ is relatively insensitive to the change of $\Delta$ from 0.1 to 0.025, 0.05 or 0.2.

## 8. Conclusions

This paper considers stochastic consensus problems where agents exchange state information via lossy analog or digital communication channels. Stochastic approximation type algorithms are applied to obtain mean square and almost sure convergence. We also consider consensus models with arbitrary switches, and the convergence analysis is based on a common stochastic Lyapunov function. Finally, we apply the algorithm to models with quantized data and packet losses by combining probabilistic quantization with a decreasing step size. For future work it will be of interest to consider more general connectivity conditions in switching networks.

## Acknowledgements

We thank the referees for their comments and for bringing to our attention several latest references appearing after our first submission, and thank an anonymous referee for suggesting the use of the joint connectivity condition.

## Appendix

**Proof of Theorem 6.** Let $z_t^1$ and $\tilde{z}_t$ be given by (17)–(18). Below we prove that there exists a mean square integrable random variable $z_\infty^1$ such that

$$\lim_{t\to\infty} E|z_t^1 - z_\infty^1|^2 = 0, \qquad \lim_{t\to\infty} E|\tilde{z}_t|^2 = 0.$$

Taking squares of the Euclidean norm on both sides of (18) gives

$$\begin{aligned}|\tilde{z}_{t+1}|^2 = {} & |\tilde{z}_t + a_t\tilde{B}_t\tilde{z}_t|^2 + a_t^2|[\phi^T\Delta B^{(I_t)}\Phi]z_t|^2 + a_t^2|\tilde{v}_t|^2 \\ & + 2a_t(\tilde{z}_t + a_t\tilde{B}_t\tilde{z}_t)^T[\phi^T\Delta B^{(I_t)}\Phi]z_t \quad (\triangleq 2a_tY_1) \\ & + 2a_t(\tilde{z}_t + a_t\tilde{B}_t\tilde{z}_t)^T\tilde{v}_t \quad (\triangleq 2a_tY_2) \\ & + 2a_t^2([\phi^T\Delta B^{(I_t)}\Phi]z_t)^T\tilde{v}_t \quad (\triangleq 2a_t^2Y_3). \end{aligned} \tag{A.1}$$

Denote the $\sigma$-algebra

$$\mathscr{F}_t' = \sigma(x_0, \mathbf{w}_0, \dots, \mathbf{w}_{t-1}, g_0, \dots, g_{t-1}, I_0, \dots, I_{t-1}).$$

In view of (17)–(18), it is evident that $z_t^1$ and $\tilde{z}_t$ are adapted to $\mathscr{F}_t'$. Since (A3) ensures that $g_t^{ij}/\lambda^{ij} - 1$ contained in $\Delta B^{(I_t)}$ has zero mean and is independent of $(\mathbf{w}_k, I_k, z_k, 0 \le k \le t)$, we may apply (10) to show that

$$E[Y_1|\mathscr{F}_t] = (\tilde{z}_t + a_t\tilde{B}_t\tilde{z}_t)^T\phi^T E[\Delta B^{(I_t)}|\mathscr{F}_t]\Phi z_t = 0.$$

Since $\mathscr{F}_t' \subset \mathscr{F}_t$, it follows that

$$E[Y_1|\mathscr{F}_t'] = E\{E[Y_1|\mathscr{F}_t]|\mathscr{F}_t'\} = 0. \tag{A.2}$$

Similarly, we apply (A2) to show that

$$E[Y_2|\mathscr{F}_t'] = E[Y_3|\mathscr{F}_t'] = 0. \tag{A.3}$$

By (A.1)–(A.3), it is straightforward to show that

$$\begin{aligned} E[|\tilde{z}_{t+1}|^2|\mathscr{F}_t'] \le {} & |\tilde{z}_t|^2 + a_t\tilde{z}_t^T E[(\tilde{B}_t^T + \tilde{B}_t)|\mathscr{F}_t']\tilde{z}_t \\ & + C_1a_t^2(E[|\mathbf{w}_t|^2|\mathscr{F}_t'] + |z_t^1|^2 + |\tilde{z}_t|^2), \end{aligned} \tag{A.4}$$

where $C_1$ is a constant independent of $t$ and we have used the fact that $|\tilde{v}_t|^2 \le C|\mathbf{w}_t|^2$ for some constant $C$.

Since under (A4), $I_t$ will take a value $1 \in \{1, \dots, K_0\}$ (so that $G^{(I_t)}$ is strongly connected) with a positive probability irrespective of $I_{t-1}$, it follows from Lemma 4(ii)–(iii) that

$$E[\tilde{B}_t^T + \tilde{B}_t|\mathscr{F}_t'] \le -c_0 I, \tag{A.5}$$

for some constant $c_0 > 0$.

Denote $s_t = E|z_t^1|^2$ and $V_t = E|\tilde{z}_t|^2$. Then (A.4) gives

$$V_{t+1} \le (1 - c_0a_t)V_t + C_1a_t^2(1 + s_t + V_t), \tag{A.6}$$

for all $t \ge 0$. And furthermore, (17) gives

$$s_{t+1} \le s_t + C_2a_t^2(1 + s_t + V_t), \tag{A.7}$$

for some $C_2 > 0$. Then by (A.6), (A.7) and Lemma 12 in Huang and Manton (2010), we obtain

$$\lim_{t\to\infty} s_t = s_\infty, \qquad \lim_{t\to\infty} V_t = 0, \tag{A.8}$$

for some finite value $s_\infty$. Hence, $\tilde{z}_t$ converges to 0 in mean square. And mean square convergence of $z_t^1$ to a limit $z_\infty^1$ follows readily from (A.8) and (17).

We proceed to prove the almost sure convergence of $z_t^1$ and $\tilde{z}_t$. Denote $\xi_t = |\tilde{z}_t|^2$. By (A.4)–(A.5), it follows that

$$E[\xi_{t+1}|\mathscr{F}_t'] \le (1-c_0a_t)\xi_t + Ca_t^2(E[|\mathbf{w}_t|^2|\mathscr{F}_t'] + |z_t^1|^2 + |\tilde{z}_t|^2)$$

for some $C>0$ and all $t\ge 0$. By Lemma 5 and

$$\sum_{t=0}^{\infty} a_t^2(E|\mathbf{w}_t|^2 + s_t + V_t) < \infty,$$

it follows that $\xi_t = |\tilde{z}_t|^2$ converges a.s. But on the other hand, it has been shown that $\tilde{z}_t$ converges to 0 in mean square. Hence, both $\xi_t$ and $\tilde{z}_t$ necessarily converge to 0 a.s.

Next, by (A2) and (9), $\{v_t^1, t \ge 0\}$ is a martingale difference sequence w.r.t. the $\sigma$-algebras $\mathscr{F}_t$ defined by (7). Also, $\{1_n^T\Delta B^{(I_t)}\Phi z_t, t\ge 0\}$ is a martingale difference sequence w.r.t. the $\sigma$-algebras $\mathscr{F}_t'' = \sigma(x_0, \mathbf{w}_0, \ldots, \mathbf{w}_t, g_0, \ldots, g_t, I_0, \ldots, I_t)$. Since $\sup_{t\ge 0}(E|v_t^1|^2 + E|1_n^T\Delta B^{(I_t)}\Phi z_t|^2) < \infty$ due to (A3) and (A.8), the a.s. convergence of $z_t^1$ follows from (17) and the martingale convergence theorem (Hall & Heyde, 1980; Stout, 1974).

Finally, by the relation $x_t = \Phi z_t = \left(1/\sqrt{n}\right) 1_n z_t^1 + \phi\tilde{z}_t$ and $(z_t^1, \tilde{z}_t) \to (z_\infty^1, 0)$, as $t\to\infty$, both in mean square and a.s., mean square and strong consensus follows. □

Before proving Theorem 7, we give a technical lemma.

**Lemma 11.** *Let $l_0$ be fixed and $t \ge l_0$, and assume (23). If each value in $\mathbf{G}$ defined by (20) appears in the sequence $B_{t-l_0}, \ldots, B_t$ as least once, then there exists $c>0$ such that*

$$\sum_{k=t-l_0}^{t} a_k(\tilde{B}_k + \tilde{B}_k^T) \le -ca_tI,$$

*for all $t \ge l_0$, where $\tilde{B}_k$ is determined by Lemma 4, i.e., $\Phi^{-1}B_k\Phi = Diag[0,\ \tilde{B}_k]$.*

**Proof.** Denote

$$H = \sum_{k=t-l_0}^{t}(a_k/a_t)(B_k + B_k^T) \triangleq \sum_{k=t-l_0}^{t}\gamma_k(B_k+B_k^T).$$

Then $H$ is the Laplacian of a strongly connected digraph. For given coefficients $(\gamma_{t-l_0}, \ldots, \gamma_t)$, by Lemma 4, we have

$$\sum_{k=t-l_0}^{t}\gamma_k(\tilde{B}_k + \tilde{B}_k^T) \le -c'I$$

for some $c'>0$. Since $\gamma_\bullet \triangleq (\gamma_{t-l_0}, \ldots, \gamma_t)$ is from a compact set by (23), there exists $c>0$ independent of $\gamma_\bullet$ such that

$$\sum_{k=t-l_0}^{t}\gamma_k(\tilde{B}_k + \tilde{B}_k^T) \le -cI.$$

Finally, $c$ may be taken to be independent of the particular values of $B_{t-l_0}, \ldots, B_t$ since there are only a finite number of such sequences (not distinguished by the starting time $t-l_0$) such that each value in $\mathbf{G}$ appears at least once. □

**Proof of Theorem 7.** Let $z_t^1$ and $\tilde{z}_t$ be given by (17)–(18). Denote $s_t = E|z_t^1|^2$ and $V_t = E|\tilde{z}_t|^2$. Following the method in proving Theorem 6, we can first show that

$$V_{t+1} \le V_t + Ca_t^2(1+s_t+V_t), \tag{A.9}$$

which differs from (A.6) by removing $1-c_0a_t$. Note that $C$ is a generic constant. In parallel, we may show that

$$s_t \le s_{t+1} \le s_t + Ca_t^2(1+s_t+V_t). \tag{A.10}$$

By using (A.9)–(A.10) and adapting the proof of Lemma 12 in Huang and Manton (2010), we obtain $V_{t+1} \le (1+Ca_t^2)\max_{0\le i\le t} V_i + Ca_t^2$, which implies

$$\max_{0\le i\le t+1} V_i \le (1+Ca_t^2)\max_{0\le i\le t} V_i + Ca_t^2. \tag{A.11}$$

By iterating (A.11), it may be shown that $\sup_{i\ge 0} V_i < \infty$, which further implies that $\sup_{i\ge 0} s_i < \infty$.

By (18), we obtain

$$\begin{aligned}
E|\tilde{z}_{t+1}|^2 &\le E|(I+a_t\tilde{B}_t)\tilde{z}_t|^2 + Ca_t^2(1+s_t+V_t)\\
&\le E|(I+a_t\tilde{B}_t)\{(I+a_{t-1}\tilde{B}_{t-1})\tilde{z}_{t-1}\\
&\quad + a_{t-1}[\phi^T\Delta B^{(I_{t-1})}]z_{t-1} + a_{t-1}\tilde{v}_t\}|^2\\
&\quad + Ca_t^2(1+s_t+V_t).
\end{aligned} \tag{A.12}$$

Let

$$\xi_1 = (I+a_t\tilde{B}_t)(I+a_{t-1}\tilde{B}_{t-1})\tilde{z}_{t-1},$$
$$\xi_2 = (I+a_t\tilde{B}_t)a_{t-1}[\phi^T\Delta B^{(I_{t-1})}]z_{t-1},$$
$$\xi_3 = (I+a_t\tilde{B}_t)a_{t-1}\tilde{v}_t.$$

We may use conditioning to show that the cross terms

$E\xi_1^T\xi_2 = E\xi_2^T\xi_3 = E\xi_3^T\xi_1 = 0.$

Then it follows from (A.12) that

$$\begin{aligned}
E|\tilde{z}_{t+1}|^2 &\le E|(I+a_t\tilde{B}_t)(I+a_{t-1}\tilde{B}_{t-1})\tilde{z}_{t-1}|^2\\
&\quad + Ca_{t-1}^2(1+s_{t-1}+V_{t-1}) + Ca_t^2(1+s_t+V_t).
\end{aligned}$$

Repeating this and using (23) and $\sup_{t\ge 0}(s_t+V_t)<\infty$, we obtain,

$$E|\tilde{z}_{t+1}|^2 \le E|(I+a_t\tilde{B}_t)\cdots(I+a_{t-l_0}\tilde{B}_{t-l_0})\tilde{z}_{t-l_0}|^2 + Ca_t^2, \tag{A.13}$$

where $l_0>0$ is any fixed integer. It follows from (A.13) that

$$\begin{aligned}
E|\tilde{z}_{t+1}|^2 &\le E|\tilde{z}_{t-l_0}|^2 + E\left[\tilde{z}_{t-l_0}^T\sum_{k=t-l_0}^{t}a_k(\tilde{B}_k+\tilde{B}_k^T)\tilde{z}_{t-l_0}\right] + Ca_t^2\\
&\quad + Ca_t^2E|\tilde{z}_{t-l_0}|^2\\
&\le E|\tilde{z}_{t-l_0}|^2 + E\left[\tilde{z}_{t-l_0}^T\sum_{k=t-l_0}^{t}a_k(\tilde{B}_k+\tilde{B}_k^T)\tilde{z}_{t-l_0}\right]\\
&\quad + Ca_t^2,
\end{aligned} \tag{A.14}$$

where the second inequality follows from $\sup_{t\ge 0}E|\tilde{z}_t|^2<\infty$.

We choose a sufficiently large $l_0$ such that regardless of the state at time $t-l_0$, the Markov chain $I_t$ will visit all its states from $t-l_0$ to $t$ with probability at least $\epsilon_0>0$, where $\epsilon_0$ does not depend on $t$. Then using conditioning and Lemma 11, we can show that for some $\delta_0>0$,

$$E\left[\tilde{z}_{t-l_0}^T\sum_{k=t-l_0}^{t}a_k(\tilde{B}_k+\tilde{B}_k^T)\tilde{z}_{t-l_0}\right] \le -\delta_0a_tE|\tilde{z}_{t-l_0}|^2. \tag{A.15}$$

By (A.14), (A.15) and (23), it follows that for some $c_0>0$,

$$V_{t+l_0+1} \le (1-c_0a_t)V_t + Ca_t^2.$$

Then by (A6), $\lim_{t\to\infty}V_t = 0$. By (A.10) we further obtain $\lim_{t\to\infty}s_t = s_\infty$ for some finite $s_\infty$. Then mean square consensus follows. Again, by a martingale convergence argument we obtain almost sure convergence of the algorithm. □

**Proof of Lemma 8.** Since $G$ contains a spanning tree, $B$ has the eigenvalue 0 and another $n-1$ eigenvalues with strictly negative real parts (Ren & Beard, 2005) and there exists a real matrix $\Phi \triangleq (1_n, \phi_{n\times n-1})$, where $\phi_{n\times(n-1)}$ is an $n\times(n-1)$ matrix, such that

$$\Phi^{-1}B\Phi = \begin{pmatrix} 0 & 0 \\ 0 & \tilde{B}_{n-1}\end{pmatrix}, \tag{A.16}$$

where the $(n-1)\times(n-1)$ matrix $\tilde{B}_{n-1}$ is Hurwitz (Huang & Manton, 2010). The rest part for showing the existence and uniqueness of a solution with the integral representation follows the same method as in proving Theorem 5 in Huang and Manton (2007).

To show necessity, we construct the deterministic consensus algorithm $z_{t+1} = (I+\varepsilon_0 B)z_t$, where $z_t = [z_t^1,\ldots,z_t^n]^T$ and $t\geq 0$. Take a sufficiently small $\varepsilon_0 > 0$ such that $I+\varepsilon_0 B$ is a stochastic matrix with positive diagonal entries and $z_{t+1}^T Q z_{t+1} \leq (1-c_0)z_t^T Q z_t$ for some $c_0>0$, which implies $\lim_{t\to\infty}\max_{i,j}|z_t^i - z_t^j| \to 0$. Next, by $z_{t+k} = (I+\varepsilon_0 B)^k z_t$, we may show $\lim_{t\to\infty}\sup_{k\geq 1}|z_{t+k} - z_t| \to 0$, so that $z_t$ converges. Hence, we conclude that consensus is achieved. By Ren and Beard (2005), $\overline{G}$ necessarily contains a spanning tree. □

## References


Acemoglu, D., Nedić, A., & Ozdaglar, A. (2008). Convergence of rule-of-thumb learning rules in social networks. In *Proc. the 47th IEEE conference on decision and control, Cancun, Mexico, December* (pp. 1714–1720).

Aysal, T. C., & Barner, K. E. (2009). On the convergence of perturbed non-stationary consensus algorithms. In *Proc. IEEE infocom, Rio de Janeiro, Brazil, April* (pp. 2132–2140).

Aysal, T. C., Coates, M. J., & Rabbat, M. G. (2007). Distributed average consensus using probabilistic quantization. In *IEEE statistical signal processing workshop. Madison, Wisconsin, August* (pp. 640–644).

Aysal, T. C., Coates, M. J., & Rabbat, M. G. (2008). Distributed average consensus with dithered quantization. *IEEE Transactions on Signal Processing*, *56*(10), 4905–4918.

Barooah, P., & Hespanha, J. P. (2007). Estimation on graphs from relative measurements: distributed algorithms and fundamental limits. *IEEE Control Systems Magazine*, *27*, 57–74.

Bellman, R. (1997). *Introduction to matrix analysis* (2nd ed.). Philadelphia: SIAM.

Carli, R., Fagnani, F., Frasca, P., Taylor, T., & Zampieri, S. (2007). Average consensus on networks with transmission noise or quantization. In *Proc. European control conf, Kos, Greece, July* (pp. 1852–1857).

Carli, R., Fagnani, F., Speranzon, A., & Zampieri, S. (2008). Communication constraints in the average consensus problem. *Automatica*, *44*, 671–684.

Cucker, F., & Mordecki, E. (2008). Flocking in noisy environments. *Journal de Mathématiques Pures et Appliquées*, *89*(3), 278–296.

Fagnani, F., & Zampieri, S. (2009). Average consensus with packet drop communication. *SIAM Journal on Control and Optimization*, *48*(1), 102–133.

Gastpar, M., Rimoldi, B., & Vetterli, M. (2003). To code, or not to code: lossy source-channel communication revisited. *IEEE Transactions on Information Theory*, *49*(5), 1147–1158.

Gastpar, M., & Vetterli, M. (2003). Source-channel communication in sensor networks. In F. Zhao, & L. Guibas (Eds.), *Springer lecture notes in computer science*: *Vol. 2634* (pp. 162–177). Berlin: Springer-Verlag.

Goblick, T. J. (1965). Theoretical limitations on the transmission of data from analog sources. *IEEE Transactions on Information Theory*, *11*(4), 558–567.

Hall, P., & Heyde, C. C. (1980). *Martingale limit theory and its application*. New York: Academic Press.

Hatano, Y., & Mesbahi, M. (2005). Agreement in random networks. *IEEE Transactions on Automatic Control*, *51*(11), 1867–1872.

Huang, M., & Dey, S. (2007). Stability of Kalman filtering with Markovian packet losses. *Automatica*, *43*, 598–607.

Huang, M., & Manton, J. H. (2007). Stochastic approximation for consensus seeking: mean square and almost sure convergence. In *Proc. 46th IEEE CDC conference, New Orleans, LA, December* (pp. 306–311).

Huang, M., & Manton, J. H. (2008). Stochastic consensus seeking with measurement noise: convergence and asymptotic normality. In *Proc. American control conf, Seattle, WA, June* (pp. 1337–1342).

Huang, M., & Manton, J. H. (2009). Coordination and consensus of networked agents with noisy measurements: stochastic algorithms and asymptotic behavior. *SIAM Journal on Control and Optimization*, *48*(1), 134–161.

Huang, M., & Manton, J. H. (2010). Stochastic consensus seeking with noisy and directed inter-agent communication: fixed and randomly varying topologies. *IEEE Transactions on Automatic Control*, *55*(1), 235–241.

Jadbabaie, A., Lin, J., & Morse, A. S. (2003). Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, *48*, 988–1000.

Kashyap, A., Basar, T., & Srikant, R. (2007). Quantized consensus. *Automatica*, *43*, 1192–1203.

Krasnopeev, A., Xiao, J.-J., & Luo, Z.-Q. (2005). Minimum energy decentralized estimation in a wireless sensor network with correlated sensor noises. *EURASIP Journal on Wireless Communications and Networking*, *2005*(4), 473–482.

Martínez, S. (2007). Practical rendezvous through modified circumcenter algorithms. In *Proc. 46th IEEE conference on decision and control, New Orleans, LA, December* (pp. 2369–2374).

Matei, I., Martins, N., & Baras, J. S. (2008). Almost sure convergence to consensus in Markovian random graphs. In *Proc. 47th IEEE conference on decision and control, Cancun, Mexico, December* (pp. 3535–3540).

Olfati-Saber, R., Fax, J. A., & Murray, R. M. (2007). Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, *95*(1), 215–233.

Olfati-Saber, R., & Murray, R. M. (2004). Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, *49*, 1520–1533.

Ren, W., & Beard, R. W. (2005). Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, *50*(5), 655–661.

Ren, W., Beard, R. W., & Kingston, D. B. (2005). Multi-agent Kalman consensus with relative uncertainty. In *Proc. American control conf, Portland, OR, June* (pp. 1865–1870).

Ren, W., Beard, R. W., & Atkins, E. M. (2005). A survey of consensus problems in multi-agent coordination. In *Proc. American control conference, Portland, OR, June* (pp. 1859–1864).

Schizas, I. D., Ribeiro, A., & Giannakis, G. B. (2008). Consensus in ad hoc WSNs with noisy links—part I: distributed estimation of deterministic signals. *IEEE Transactions on Signal Processing*, *56*, 350–364.

Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K., Jordan, M. I., & Sastry, S. S. (2004). Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, *49*, 1453–1464.

Smith, S. C., & Seiler, P. (2003). Estimation with lossy measurements: jump estimators for jump systems. *IEEE Transactions on Automatic Control*, *48*, 2163–2171.

Stankovic, S. S., Stankovic, M. S., & Stipanovic, D. M. (2007). Decentralized parameter estimation by consensus based stochastic approximation. In *Proc. 46th IEEE conference on decision and control, New Orleans, LA, December* (pp. 1535–1540).

Stout, W. F. (1974). *Almost sure convergence*. New York: Academic Press.

Tahbaz-Salehi, A., & Jadbabaie, A. (2008). A necessary and sufficient condition for consensus over random networks. *IEEE Transactions on Automatic Control*, *53*(3), 791–795.

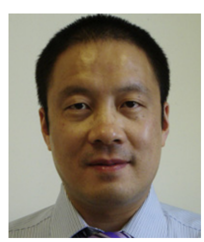
Tsitsiklis, J. N., Bertsekas, D. P., & Athans, M. (1986). Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, *31*(9), 803–812.

Wu, C. W. (2006). Synchronization and convergence of linear dynamics in random directed networks. *IEEE Transactions on Automatic Control*, *51*(7), 1207–1210.

Xiao, L., Boyd, S., & Kim, S.-J. (2007). Distributed average consensus with least-mean-square deviation. *Journal of Parallel and Distributed Computing*, *67*, 33–46.

Xiao, J.-J., Cui, S., Luo, Z.-Q., & Goldsmith, A. J. (2006). Power scheduling of universal decentralized estimation in sensor networks. *IEEE Transactions on Signal Processing*, *54*(2), 413–422.

Yildiz, M. E., & Scaglione, A. (2008). Coding with side information for rate-constrained consensus. *IEEE Transactions on Signal Processing*, *56*(8), 3753–3764.


**Minyi Huang** received the B.Sc. degree from Shandong University, Jinan, Shandong, China, in 1995, the M.Sc. degree from the Institute of Systems Science, Chinese Academy of Sciences, Beijing, China, in 1998, and the Ph.D. degree from the Department of Electrical and Computer Engineering, McGill University, Montreal, Canada, in 2003, all in the area of systems and control.

From February 2004 to March 2006, he was a Research Fellow with the Department of Electrical and Electronic Engineering, the University of Melbourne, Victoria, Australia. From April 2006 to June 2007, he was a Research Fellow with the Department of Information Engineering, Research School of Information Sciences and Engineering, the Australian National University, Canberra, Australia. He joined Carleton University, Ottawa, ON, Canada, in July 2007, where he is an Assistant Professor in the School of Mathematics and Statistics. His research interests include stochastic control and game theory, multiagent stochastic systems, stochastic algorithms, and wireless networks.



**Subhrakanti Dey** was born in Calcutta, India, in 1968. He received the B.Tech. and M.Tech. degrees from the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, India, in 1991 and 1993, respectively, and the Ph.D. degree from the Department of Systems Engineering, Research School of Information Sciences and Engineering, Australian National University, Canberra, Australia, in 1996.

He has been with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Australia, since February 2000, where he is currently a full Professor. From September 1995 to September 1997 and September 1998 to February 2000, he was a postdoctoral Research Fellow with the Department of Systems Engineering, Australian National University. From September 1997 to September 1998, he was a postdoctoral Research Associate with the Institute for Systems Research, University

of Maryland, College Park. His current research interests include networked control systems, wireless communications and networks, signal processing for sensor networks, and stochastic and adaptive estimation and control. Dr. Dey currently serves on the Editorial Board of the IEEE Transactions on Signal Processing and Elsevier Systems and Control Letters. He was also an Associate Editor for the IEEE Transactions on Automatic Control during 2005–2007. He is a Senior Member of IEEE.

**Girish N. Nair** was born in Petaling Jaya, Malaysia. He obtained a B.E. (Electrical, 1st class honours) in 1994, B.Sc. (Mathematics) in 1995, and Ph.D. (Electrical Engineering) in 2000, on scholarships from the Australian government and the University of Melbourne. He is currently an associate professor in the Department of Electrical and Electronic Engineering at the University of Melbourne and has also held visiting positions at the University of Padova, Italy and Boston University, USA. He has received several prizes, including a SIAM Outstanding Paper Prize in 2006 and the Best Theory Paper Prize at the UKACC International Conference on Control, Cambridge University, 2000. His research interests lie in communications, information theory and control and he serves as an associate editor for the SIAM Journal on Control and Optimization.

**Jonathan H. Manton** received his Bachelor of Science (Mathematics) and Bachelor of Engineering (Electrical) degrees in 1995 and his Ph.D. degree in 1998, all from the University of Melbourne, Australia. From 1998 to 2004, he was with the Department of Electrical and Electronic Engineering at the University of Melbourne. During that time, he held a Postdoctoral Research Fellowship then subsequently a Queen Elizabeth II Fellowship, both from the Australian Research Council. In 2005 he became a full Professor in the Department of Information Engineering, Research School of Information Sciences and Engineering (RSISE) at the Australian National University. From July 2006 till May 2008, he was on secondment to the Australian Research Council as Executive Director, Mathematics, Information and Communication Sciences. Currently, he holds a distinguished Chair at the University of Melbourne with the title Future Generation Professor. He is also an Adjunct Professor in the Mathematical Sciences Institute at the Australian National University. Professor Manton's traditional research interests range from pure mathematics (e.g. commutative algebra, algebraic geometry, differential geometry) to engineering (e.g. signal processing, wireless communications). Recently though, led by a desire to participate in the convergence of the life sciences and the mathematical sciences, he has commenced learning neuroscience. Professor Manton also has extensive experience in software development.