

# Model Based Peer-to-Peer Estimator over Wireless Sensor Networks with Lossy Channels

Y. Xu<sup>a</sup>, C. Fischione<sup>a</sup>, A. Speranzon<sup>b</sup>

<sup>a</sup>*ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden*

<sup>b</sup>*United Technologies Research Center, East Hartford, CT, USA*

---

## Abstract

In this paper, the design methods and fundamental performance analysis of an adaptive peer-to-peer estimator are established for networks exhibiting message losses. Based on a signal state model, estimates are locally computed at each node of the network by adaptively filtering neighboring nodes' estimates and measurements communicated over lossy channels. The computation is based on a distributed optimization approach that guarantees the stability of the estimator while minimizing the estimation error variance. The fundamental performance limitations of the estimator are established based on the variance of the estimation error in relation to the message loss process. Numerical simulations validate the theoretical analysis and illustrate the performance with respect to other estimators.

*Key words:* Distributed Estimation; Distributed Optimization; Wireless Sensor Networks.

---

## 1 Introduction

Monitoring physical variables is a typical task that can be efficiently performed by wireless sensor networks (WSNs). Accurate estimation of these variables is necessary in many applications, spanning from traffic monitoring and control, industrial automation, environment monitoring, to security systems [1]. However, nodes of WSNs are typically characterized by both limited sensing, for which measurements are noisy, and communication capabilities, for which message losses may be not insignificant. Estimation algorithms must be designed to cope with these adverse conditions, while offering high accuracy of the estimates.

There are two main estimation strategies for WSNs. A traditional approach consists in letting nodes sense the environment and then report data to a central unit, which extracts the desired physical variables and sends the estimate to each local node. This approach has strong limitations: large amount of communication resources

(radio power, bandwidth, routing, etc.) have to be managed for the transmission of information from nodes to the central unit and vice versa. This mode of operation reduces the nodes' lifetime as most of the energy consumption is due to communication and message forwarding to the central location. An alternative approach, which we investigate in this paper, enables each node to produce locally accurate estimates taking advantage, through communication, of neighboring nodes' local estimations and measurements. The challenge of such a distributed estimation is that local processing must be carefully characterized to avoid uncontrolled propagation of the estimation errors throughout the network and, at the same time, to guarantee good estimation performance.

The estimator presented in this paper is related to contributions on low-pass filtering by diffusion mechanisms, e.g., [1–8], where each node of the network obtains the average of the initial samples collected by nodes. In [9,10] the authors study a distributed average computation of a time-varying signal, when the signal is affected by a zero-mean noise. Distributed filtering using model-based approaches is studied in various wireless network contexts, e.g., [11–15]. In particular, distributed Kalman filters and more recently a combination of the diffusion mechanism with distributed Kalman filtering have been proposed, e.g., [5,16,17]. In [18], a strategy where the

---

*Email addresses:* yuzhe@kth.se (Y. Xu),  
carlofi@kth.se (C. Fischione),  
alberto.speranzon@utrc.utc.com (A. Speranzon).

<sup>1</sup> The work by C. Fischione and Y. Xu was funded by the Swedish Research Council and the EU Projects Hydrobionets and Hycon2.

estimator works at the same time as the communication update is studied. In [19], two iterative linear distributed estimators are proposed. The first one considers an incremental mode of cooperation, while the second considers a diffusive iterative mode, but both modes are based on the measurements only and no model of the signal is assumed. In [20], a theoretical framework for coupled distributed estimation and motion control of mobile sensor networks for collaborative target tracking is proposed. In [21] a distributed  $H_\infty$  state estimation for discrete time varying nonlinear systems is studied. In this estimator, sensor nodes have knowledge of the fixed network topology. In [22] the estimation performance is studied on centralized Kalman filtering in fading wireless channels suffering message losses, where power control is considered to alleviate the fading. In [23], non-Gaussian measurement noise is considered, where the error entropy is minimized by the estimator.

In this paper, we propose a distributed approach to estimate a time-varying multi-dimensional signal affected by unknown additive disturbances. This is in contrast to approaches where statistical models of disturbances are assumed [13,24,25], or to approaches that are focused on averaging initial samples [2,9,26–28], or to methods where the estimation variables are static parameters [12,18,29–31], or to those where only local measurements are diffused over the network [32]. Compared to [2,7,9,13,26,28], we do not use the consensus algorithms in the estimator, and compared to [9], we do not use the Laplacian matrix associated to the communication graph to design the estimator. Our estimation parameters are computed through distributed algorithms that adapt to the network topology and message losses, in contrast to [12], which rely on centralized fusion center, or to [9,10,33], where the computation of estimator need the full knowledge of the communication graph. This paper is a natural extension of the distributed estimators proposed in [34,35], which were designed for scalar signals under the assumption of perfect communication, namely no message losses among nodes. We extend and substantially generalize [34,35] as follows: First, we consider multi-dimensional system. Second, the estimator is designed to be robust to message losses, leading to a new optimal estimator structure. Third, new analytical results are provided to demonstrate the performance achieved by the new estimator in the presence of message losses.

The remainder of the paper is organized as follows. In Section 2 we define the estimator structure and the message loss model. In Section 3 we characterize the estimator design method when message losses are present and propose an estimation algorithm. In Section 4 we establish bounds on the estimation error variance of the proposed algorithm. In Section 5, we report the computational complexity of the proposed algorithm. Numerical simulations are reported in Section 6 to illustrate the performance of the proposed algorithm compared to

others in literatures. Conclusions are drawn in Section 7.

### 1.1 Notation

Let  $\mathbb{E}x$  and  $\text{Cov}x$  denote the expectation and the covariance matrix of stochastic variable  $x$ , respectively, whereas  $\mathbb{E}_y[x(y)]$  and  $\text{Cov}_y[x(y)]$  denote the expectation and the covariance matrix taken with respect to the probability density function of  $y$ , respectively.  $\|\cdot\|$  is the spectral norm of a matrix, or  $\ell^2$ -norm of a vector, while  $\|\cdot\|_F$  is Frobenius norm of a matrix. For matrix  $A$ , let  $\ell_m(A)$  and  $\ell_M(A)$  denote the minimum and maximum eigenvalue (with respect to the absolute value of the real part), respectively.  $\text{tr} A$  is the trace of  $A$ . Let  $\circ$  and  $\otimes$  denote the Hadamard (element-wise) product and the Kronecker product for matrices, respectively. Let  $^\dagger$  be the Moore-Penrose pseudo-inverse for matrices [36]. For matrix  $A$ ,  $A \geq 0$  ( $A \leq 0$ ) is equivalent to  $A$  ( $-A$ ) being positive semidefinite. Let  $\mathbf{I}$  and  $\mathbf{1}$  be the identity matrix and the vector  $(1, \dots, 1)^T$ , respectively, whose dimensions may be indicated by a subscript. Let  $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ .

## 2 Problem Formulation

Suppose a multi-dimensional signal needs to be estimated by a WSN. The signal is modeled by a linear system with unknown additive disturbance

$$x(t+1) = Ax(t) + w(t), \quad (1)$$

for  $t \in \mathbb{N}_0$ , where  $x(t) \in \mathbb{R}^n$  is system state at time  $t$ , and  $A \in \mathbb{R}^{n \times n}$  represents its dynamics.  $w(t) \in \mathbb{R}^n$  models the disturbance in the state, whose  $\ell^2$ -norm is bounded by  $\Delta$  for all the time [37,38]. Note that the unknown disturbance is commonly addressed by  $H_\infty$  filtering [21,39,37,38]. Suppose we place  $N$  sensor nodes at static positions in the space. These nodes can communicate to their neighbors. At each time, node  $i$  takes a noisy measurement of the system as

$$y_i(t) = C_i x(t) + v_i(t), \quad (2)$$

where  $y_i(t) \in \mathbb{R}^{m_i}$  is the measurement of  $x(t)$ , and  $C_i \in \mathbb{R}^{m_i \times n}$ .  $v_i(t) \in \mathbb{R}^{m_i}$  has a Gaussian distribution with zero mean and diagonal covariance  $R_i$ , and  $\mathbb{E}[v_i(t)v_j^T(t)] = \mathbf{0}$  for all  $t \in \mathbb{N}_0$ ,  $i \neq j$ . We assume that node  $i$  knows  $A$  and  $C_i$ .

The communication network is modelled by a graph, in which  $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V} = \{1, \dots, N\}$  is the vertex set and  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  is the edge set. The set of neighbors of node  $i \in \mathcal{V}$  plus node  $i$  is denoted as  $\mathcal{N}_i = \{j \in \mathcal{V} : (i, j) \in \mathcal{E}\} \cup \{i\}$ . Namely  $\mathcal{N}_i$  is the set containing the neighbors that a node  $i$  can have, including itself. Let  $N_i$  be the cardinality of set  $\mathcal{N}_i$ .

In each time interval, every node takes the measurement of signal (1) before broadcasting its previous estimates

and measurements to its neighbors. After receiving messages from neighbors, each node computes the latest estimate of the signal. However, the wireless communication may be lossy because of bad channel conditions, caused by radio interference and/or transmission conflicts (hidden nodes or exposed nodes). Let  $\varphi_{i_j|t}$  for  $(i, j) \in \mathcal{E}$  denote a binary random variable, modelling the message loss on the edge  $(i, j)$  at time  $t$  from node  $j$  to  $i$ . We assume that  $\varphi_{i_j|t}$  has the probability mass function:

$$\Pr(\varphi_{i_j|t} = 1) = p_{ij}, \Pr(\varphi_{i_j|t} = 0) = 1 - p_{ij} = q_{ij}, \quad (3)$$

where  $p_{ij} \in (0, 1)$  denotes the successful message reception probability, and consequently  $q_{ij} \in (0, 1)$  is the message loss probability. The message loss process is assumed to be independent among links, and independent from past message losses. These assumptions are natural when the coherence time of the wireless channel is small with respect to the typical communication rate of messages over WSNs [40].

**Remark 2.1** Notice that when  $A = 1$ , and the communication is perfect, namely no message losses, the system model described by (1) is the particular case studied in [35]. When  $A = a$ , with  $a \in \mathbb{R}$  and the communication is perfect we recover the case studied in [34].

We also assume that every node  $i$  updates its estimate  $\hat{x}_i(t)$  of  $x(t)$  by taking a linear combination of its own and of its neighbors estimates and measurements:

$$\hat{x}_i(t) = \sum_{i_j \in \mathcal{N}_i} \varphi_{i_j|t} (G_{i_j}(t)A\hat{x}_{i_j}(t-1) + H_{i_j}(t)y_{i_j}(t)), \quad (4)$$

with  $\hat{x}_i(0) = \mathbb{E}x(0)$ ,  $G_{i_j}(t) \in \mathbb{R}^{n \times n}$ , and  $H_{i_j}(t) \in \mathbb{R}^{n \times m_j}$  for all  $i, j$ . Let  $\mathcal{N}_{\varphi_{i|t}} = \{j \in \mathcal{N}_i : \varphi_{i_j|t} \neq 0\}$ , namely such a set collects the nodes communicating to node  $i$  at time  $t$ . The cardinality of the set is denoted by  $N_{\varphi_{i|t}} \geq 1$ . Fig. 1 illustrates an example of a network.

At node  $i$ , let  $\hat{\mathbf{x}}_i(t) = (\hat{x}_{i_1}^T(t), \dots, \hat{x}_{i_{N_i}}^T(t))^T$  and  $\mathbf{y}_i(t) = (y_{i_1}^T(t), \dots, y_{i_{N_i}}^T(t))^T$  collect the estimations and measurements of the neighbors, where we use  $i_1, \dots, i_{N_i}$  to denote the nodes that belongs to  $\mathcal{N}_i$ . Thus, (4) can be rewritten in matrix form as

$$\hat{\mathbf{x}}_i(t) = \mathbf{G}_{\varphi_{i|t}} \mathbf{A}_i \hat{\mathbf{x}}_i(t-1) + \mathbf{H}_{\varphi_{i|t}} \mathbf{y}_i(t), \quad (5)$$

$$\text{where } \mathbf{G}_{\varphi_{i|t}} = \mathbf{G}_i(t) \circ (\varphi_{i|t}^T \otimes \mathbf{1}_n \mathbf{1}_n^T), \quad (6)$$

$$\mathbf{G}_i(t) = (G_{i_1}(t), \dots, G_{i_{N_i}}(t)), \quad (7)$$

$$\varphi_{i|t} = (\varphi_{i_1|t}, \dots, \varphi_{i_{N_i}|t})^T, \quad (8)$$

while  $\mathbf{H}_{\varphi_{i|t}}$  and  $\mathbf{H}_i(t)$  are constructed from the elements  $H_{i_j}(t)$ , similarly to  $\mathbf{G}_{\varphi_{i|t}}$  and  $\mathbf{G}_i(t)$ , respectively. Moreover,  $\mathbf{C}_i$  is a block diagonal matrix whose diagonal is

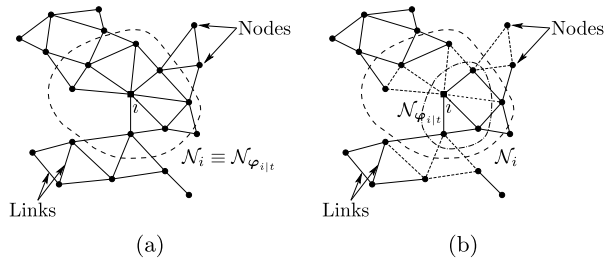


Fig. 1. An example of network with indicated the nodes and active links (solid line). In (a) the set  $\mathcal{N}_i \equiv \mathcal{N}_{\varphi_{i|t}}$  is shown for the case of no message losses (dashed curve). The node  $i$  is the square one. For no message losses,  $\mathcal{N}_i$  coincides with  $\mathcal{N}_{\varphi_{i|t}}$ . In (b) some communication links are not active because of message losses (dashed lines). Notice that the set  $\mathcal{N}_{\varphi_{i|t}}$  (dash-dotted curve) and  $\mathcal{N}_i$  (dashed curve) do not coincides anymore. The set  $\mathcal{N}_i$  remains the same (as it does not depend on the message losses).

$C_{i_1}, \dots, C_{i_{N_i}}$ , while  $\mathbf{A}_i \in \mathbb{R}^{nN_i \times nN_i}$  is a block diagonal matrix whose diagonal blocks are  $A$ s.

In this paper, we derive an estimator only based on local information and computations. In this scenario, a centralized estimator, such as Kalman filtering, cannot be applied, since it needs the global covariance matrices, which can not be collected on time due to message losses. Nevertheless, for performance comparisons with the distributed estimator presented in this paper, in Section 6 we suppose there exists such a Kalman filtering.

### 3 Distributed Minimum Variance Estimator

In this section we derive the design method for the estimator (4) introduced in Section 2. We will show that the optimal estimation coefficients that provide minimum trace of covariance of estimation errors can be computed by using semidefinite programming, while avoiding uncontrolled propagation of the estimation errors. In addition, we characterize these coefficients, which will be very useful for the performance analysis in Section 4. We also show that sub-optimal estimation coefficients can be computed in a closed form, which will be instrumental in Section 4.

Let  $e_i(t) \in \mathbb{R}^n$  be the estimation error at time  $t$  in node  $i$ . Construct  $\mathbf{e}_i(t) = (e_{i_1}^T(t), \dots, e_{i_{N_i}}^T(t))^T$  which is an  $nN_i$  vector collecting the estimation errors of the neighbors of  $i$ . We define  $\boldsymbol{\nu}_{\varphi_{i|t}} = \varphi_{i|t} \otimes \mathbf{I}_n$  and impose

$$(\mathbf{G}_{\varphi_{i|t}} + \mathbf{H}_{\varphi_{i|t}} \mathbf{C}_i) \boldsymbol{\nu}_{\varphi_{i|t}} = \mathbf{I}_n, \quad (9)$$

to remove the explicit dependence on the signal at the times  $t-1$ ,  $t-2$ , and so on. This allows us to write the dynamics of  $e_i(t)$  as

$$e_i(t) = \mathbf{G}_{\varphi_{i|t}} (\mathbf{A}_i \mathbf{e}_i(t-1) + \mathbf{w}_i(t-1)) - \mathbf{H}_{\varphi_{i|t}} \mathbf{v}_i(t), \quad (10)$$

where

$$\mathbf{v}_i(t) = (v_{i_1}^T(t), \dots, v_{i_{N_i}}^T(t))^T, \mathbf{w}_i(t) = \boldsymbol{\nu}_i w(t). \quad (11)$$

Let  $\mathbf{e}(t) = (e_1^T(t), \dots, e_N^T(t))^T$ . The following theorem provides a bound on the average estimation error.

**Theorem 3.1** *Consider the dynamics of  $e_i(t)$  in (10) following (9), and suppose  $\|w(t)\| \leq \Delta$  for all  $t \in \mathbb{N}_0$ . If  $\|\mathbf{G}_{\varphi_{i|t}}\| \leq \gamma/\sqrt{N_{\varphi_{i|t}}}$  with nonnegative scalar  $\gamma \leq 1/\|A\|$  for every message loss realization  $\varphi_{i|t}$ , then there exists  $0 < \gamma' < \gamma$  such that the estimation error of (4) is bounded as*

$$\lim_{t \rightarrow +\infty} \|\mathbb{E}_{\varphi} \mathbf{E}_{\mathbf{v}} \mathbf{e}(t)\| \leq \frac{\gamma' \sqrt{N}}{1 - \gamma' \|A\|} \Delta. \quad (12)$$

Before proving Theorem 3.1, we need the following technical results.

**Lemma 3.1** *Consider matrices  $\mathbf{U} = [U_1, U_2]$  and  $\mathbf{V} = [V_1, \mathbf{0}]$ , in which nonzero matrices  $U_1 \in \mathbb{R}^{n \times p}$ ,  $U_2 \in \mathbb{R}^{n \times q}$ ,  $V_1 \in \mathbb{R}^{n \times p}$ , and zero matrix  $\mathbf{0} \in \mathbb{R}^{n \times q}$ , where  $n \leq \min(p, q)$ . Then,*

$$\|\mathbf{U} + \mathbf{V}\| < \|\mathbf{U}\| + \|\mathbf{V}\|.$$

*Proof:*

$$\begin{aligned} \|\mathbf{U} + \mathbf{V}\|^2 &= \ell_M((U_1 + V_1)(U_1 + V_1)^T + U_2 U_2^T) \\ &= \ell_M(U_1 U_1^T + U_2 U_2^T + V_1 V_1^T + V_1 U_1^T + U_1 V_1^T) \\ &\leq \|\mathbf{U}\|^2 + \|\mathbf{V}\|^2 + 2\ell_M(U_1 V_1^T) \\ &\leq \|\mathbf{U}\|^2 + \|\mathbf{V}\|^2 + 2\|U_1 V_1^T\|, \end{aligned} \quad (13)$$

where the last inequality is due to that for a matrix, the largest singular value is larger or equal to the largest eigenvalue. Moreover,

$$\begin{aligned} \|U_1 V_1^T\| &\leq \|U_1\| \|V_1\| = \ell_M(U_1 U_1^T) \|V_1\| \\ &< \ell_M(U_1 U_1^T + U_2 U_2^T) \|V_1\| = \|\mathbf{U}\| \|\mathbf{V}\|. \end{aligned} \quad (14)$$

Then, by inserting (14) into (13), we have

$$\|\mathbf{U} + \mathbf{V}\|^2 < \|\mathbf{U}\|^2 + \|\mathbf{V}\|^2 + 2\|\mathbf{U}\| \|\mathbf{V}\| = (\|\mathbf{U}\| + \|\mathbf{V}\|)^2,$$

which completes the proof.  $\square$

We are now ready to prove Theorem 3.1.

*Proof:* Let  $V_i(t) = \|\mathbb{E}_{\varphi} \mathbf{E}_{\mathbf{v}} e_i(t)\|$  and  $V(t) = \|\mathbb{E}_{\varphi} \mathbf{E}_{\mathbf{v}} \mathbf{e}(t)\|$ . After some algebraic manipulations on

(10), we have

$$\begin{aligned} V_i(t) &\leq \left\| \mathbb{E}_{\varphi} \mathbf{G}_{\varphi_{i|t}} \sqrt{N_{\varphi_{i|t}}} \right\| \|A\| \max_{j \in \mathcal{N}_{\varphi_{i|t}}} V_j(t-1) \\ &\quad + \left\| \mathbb{E}_{\varphi} \mathbf{G}_{\varphi_{i|t}} \sqrt{N_{\varphi_{i|t}}} \right\| \Delta. \end{aligned} \quad (15)$$

Now note that

$$\mathbb{E}_{\varphi} \mathbf{G}_{\varphi_{i|t}} \sqrt{N_{\varphi_{i|t}}} = \sum_{\varphi_{i|t}} \left( \mathbf{G}_{\varphi_{i|t}} \sqrt{N_{\varphi_{i|t}}} \right) \Pr(\varphi_{i|t}), \quad (16)$$

which follows from that different realizations of  $\varphi_{i|t}$  are mutually exclusive. The expectation is given by the sum of a finite number of combinations of possible message loss realizations, so that it follows

$$\begin{aligned} \left\| \mathbb{E}_{\varphi} \mathbf{G}_{\varphi_{i|t}} \sqrt{N_{\varphi_{i|t}}} \right\| &= \left\| \sum_{\varphi_{i|t}} \left( \mathbf{G}_{\varphi_{i|t}} \sqrt{N_{\varphi_{i|t}}} \right) \Pr(\varphi_{i|t}) \right\| \\ &< \sum_{\varphi_{i|t}} \left\| \mathbf{G}_{\varphi_{i|t}} \right\| \sqrt{N_{\varphi_{i|t}}} \Pr(\varphi_{i|t}) \leq \sum_{\varphi_{i|t}} \gamma \Pr(\varphi_{i|t}) = \gamma, \end{aligned} \quad (17)$$

where the strict inequality follows from Lemma 3.1 and that the submatrix in  $\mathbf{G}_{\varphi_{i|t}}$  corresponding to the message losses will be zero. Therefore, (17) is a strict inequality, and from (15), for all  $i \in \mathcal{N}$ , we have

$$V_i(t) < \gamma \|A\| \max_{j \in \mathcal{N}_{\varphi_{i|t}}} V_j(t-1) + \gamma \Delta, \quad (18)$$

Thus, there exists  $0 < \gamma' < \gamma$  such that at any time  $t$

$$\max_{k \in \mathcal{V}} V_k(t) \leq \gamma' \|A\| \max_{j \in \mathcal{V}} V_j(t-1) + \gamma' \Delta. \quad (19)$$

By recursively applying (19),

$$V_i(t) \leq \max_{k \in \mathcal{V}} V_k(t) \leq \gamma'^t \|A\|^t \max_{j \in \mathcal{V}} V_j(0) + \gamma' \frac{1 - \gamma'^t \|A\|^t}{1 - \gamma'^t \|A\|^t} \Delta,$$

whereby, since  $\gamma' < \gamma \leq 1/\|A\|$ , we have

$$V_i(t) \leq \frac{\gamma'}{1 - \gamma' \|A\|} \Delta \Rightarrow V(t) \leq \frac{\gamma' \sqrt{N}}{1 - \gamma' \|A\|} \Delta, \quad (20)$$

which completes the proof.  $\square$

Based on Theorem 3.1, we could obtain the desired average estimation error by properly choosing  $\gamma$ . We remark that if  $\mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  is full column rank, then  $\|\mathbf{G}_{\varphi_{i|t}}\|$  is only bounded by the assumption in Theorem 3.1. However, if  $\mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  is not full column rank, due to (9),  $\|\mathbf{G}_{\varphi_{i|t}}\|$  is bounded by the following lemma:

**Lemma 3.2** *Suppose (9) hold. If  $\mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  is not full column rank, then  $\|\mathbf{G}_{\varphi_{i|t}}\| \geq 1/\sqrt{N_{\varphi_{i|t}}}$ .*

*Proof:* From (9), we have

$$\mathbf{G}_{\varphi_{i|t}} \boldsymbol{\nu}_{\varphi_{i|t}} = \mathbf{I}_n - \mathbf{H}_{\varphi_{i|t}} \mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}.$$

Moreover, matrix  $\mathbf{H}_{\varphi_{i|t}} \mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  is not full rank, since  $\mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  is not full column rank. By Rank-nullity Theorem,  $\mathbf{H}_{\varphi_{i|t}} \mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  must have at least one eigenvalue at zero. Thus,  $\mathbf{I}_n - \mathbf{H}_{\varphi_{i|t}} \mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  have at least one eigenvalue at one, so does  $\mathbf{G}_{\varphi_{i|t}} \boldsymbol{\nu}_{\varphi_{i|t}}$ . Therefore, we have

$$1 \leq \|\mathbf{G}_{\varphi_{i|t}} \boldsymbol{\nu}_{\varphi_{i|t}}\| \leq \|\mathbf{G}_{\varphi_{i|t}}\| \|\boldsymbol{\nu}_{\varphi_{i|t}}\| = \|\mathbf{G}_{\varphi_{i|t}}\| \sqrt{N_{\varphi_{i|t}}},$$

which completes the proof.  $\square$

We remark that due to the unknown disturbance  $w(t)$ ,  $\text{rank}(C_i) = n$  is a necessary condition for linear observers following [41,42]. In [43], this condition is relaxed however, by identifying the various matrices, the relaxed condition results in  $\text{rank}(C_i) \geq n$ , which can be realized only for  $\text{rank}(C_i) = n$  in our setting. Thus, if  $C_i$  is full rank, observability is guaranteed and, moreover, we have that  $\mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  is full column rank independently of the message loss process. Furthermore, considering Theorem 3.1 and Lemma 3.2, if  $\|A\| \leq 1$ , clearly, the condition on  $C_i$  full rank can be relaxed.

We can now compute the estimation weights  $\mathbf{G}_{\varphi_{i|t}}$  and  $\mathbf{H}_{\varphi_{i|t}}$  under the stability constraint imposed by the condition on  $\mathbf{G}_{\varphi_{i|t}}$  of Theorem 3.1. At each time step, every node  $i \in \mathcal{N}$  computes its estimation weights by solving the following optimization problem that minimizes the trace of the covariance matrix  $\text{Cov}_{\mathbf{v}} e_i(t)$ :

$$\min_{\mathbf{G}_{\varphi_{i|t}}, \mathbf{H}_{\varphi_{i|t}}} \text{tr}(\text{Cov}_{\mathbf{v}} e_i(t)) \quad (21a)$$

$$\text{s.t. } \left( \mathbf{G}_{\varphi_{i|t}} + \mathbf{H}_{\varphi_{i|t}} \mathbf{C}_i \right) \boldsymbol{\nu}_{\varphi_{i|t}} = \mathbf{I}_n \quad (21b)$$

$$\left\| \mathbf{G}_{\varphi_{i|t}} \right\| \leq \gamma / \sqrt{N_{\varphi_{i|t}}}, \quad (21c)$$

where objective function is

$$\begin{aligned} \text{Cov}_{\mathbf{v}} e_i(t) &= \mathbb{E}_{\mathbf{v}}[(e_i(t) - \mathbb{E}_{\mathbf{v}} e_i(t))(e_i(t) - \mathbb{E}_{\mathbf{v}} e_i(t))^T] \\ &= \mathbf{G}_{\varphi_{i|t}} \mathbf{P}_{\varphi_{i|t}} \mathbf{G}_{\varphi_{i|t}} + \mathbf{H}_{\varphi_{i|t}} \mathbf{Q}_{\varphi_{i|t}} \mathbf{H}_{\varphi_{i|t}}, \end{aligned} \quad (22)$$

and matrix  $\mathbf{P}_{\varphi_{i|t}}$  and  $\mathbf{Q}_{\varphi_{i|t}}$  are defined as

$$\mathbf{P}_{\varphi_{i|t}} = (\mathbf{A}_i \text{Cov}_{\mathbf{v}} e_i(t-1) \mathbf{A}_i^T) \circ (\boldsymbol{\varphi}_{i|t} \boldsymbol{\varphi}_{i|t}^T \otimes \mathbf{1}_n \mathbf{1}_n^T) \quad (23)$$

$$\mathbf{Q}_{\varphi_{i|t}} = \mathbb{E}_{\mathbf{v}}[\mathbf{v}_i(t) \mathbf{v}_i^T(t)] \circ (\boldsymbol{\varphi}_{i|t} \boldsymbol{\varphi}_{i|t}^T \otimes \mathbf{1}_n \mathbf{1}_n^T). \quad (24)$$

We remark that in optimization problem (21), if  $\mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  is full column rank, we can choose any  $\gamma \leq 1/\|A\|$  according to the requirement of the estimation error bound given in Theorem 3.1. Otherwise, if  $\mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  is not full

column rank (when  $\|A\| \leq 1$ ), we let  $1 \leq \gamma \leq 1/\|A\|$  to bound the estimation error given in Theorem 3.1. Note that constraint (21b) ensures (9), whereas constraint (21c), together with (21b), ensures that the expectation of the estimation error is bounded [compare with Theorem 3.1]. We remark that both  $\mathbf{P}_{\varphi_{i|t}}$  and  $\mathbf{Q}_{\varphi_{i|t}}$  are computed locally at node  $i$ . The calculation of  $\mathbf{P}_{\varphi_{i|t}}$  is described in Appendix A, which is only an estimation. Problem (21) can be efficiently solved at every node  $i$  by semidefinite programming [44]. Note that the characterization of the solution is very useful when analyzing the estimate properties as we will show in Section 4. The following proposition formally characterize the solution.

**Proposition 3.1** *There exists positive semidefinite matrix  $\mathbf{M}_i$ , such that the optimal solutions of optimization problem (21),  $\mathbf{G}_{\varphi_{i|t}}^*$  and  $\mathbf{H}_{\varphi_{i|t}}^*$ , can be written as*

$$\mathbf{G}_{\varphi_{i|t}}^* = \mathbf{S}_{\varphi_{i|t}}^\dagger \boldsymbol{\nu}_{\varphi_{i|t}}^T (\mathbf{P}_{\varphi_{i|t}} + \mathbf{M}_i)^\dagger, \quad (25a)$$

$$\mathbf{H}_{\varphi_{i|t}}^* = \mathbf{S}_{\varphi_{i|t}}^\dagger \boldsymbol{\nu}_{\varphi_{i|t}}^T \mathbf{C}_i^T \mathbf{Q}_{\varphi_{i|t}}^\dagger, \quad (25b)$$

where

$$\mathbf{S}_{\varphi_{i|t}} = \boldsymbol{\nu}_{\varphi_{i|t}}^T ((\mathbf{P}_{\varphi_{i|t}} + \mathbf{M}_i)^\dagger + \mathbf{C}_i^T \mathbf{Q}_{\varphi_{i|t}}^\dagger \mathbf{C}_i) \boldsymbol{\nu}_{\varphi_{i|t}}. \quad (26)$$

Moreover, if  $\mathbf{M}_i \neq \mathbf{0}$ , then  $\left\| \mathbf{G}_{\varphi_{i|t}}^* \right\| = \gamma / \sqrt{N_{\varphi_{i|t}}}$ .

*Proof:* Note that problem (21) is feasible, since there always exist feasible solutions. When  $\mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  is full column rank, consider solutions  $\mathbf{G}_{\varphi_{i|t}} = \mathbf{0}$  and  $\mathbf{H}_{\varphi_{i|t}} = (\mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}})^\dagger$ , which are feasible to problem (21). Otherwise if  $\mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  is not full column rank, let  $\mathbf{H}_{\varphi_{i|t}} = (\mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}})^\dagger$ , and  $\mathbf{G}_{\varphi_{i|t}} = (\mathbf{I}_n - \mathbf{H}_{\varphi_{i|t}} \mathbf{C}_i) \boldsymbol{\nu}_{\varphi_{i|t}} / N_{\varphi_{i|t}} \otimes \boldsymbol{\varphi}_{i|t}$ . Then,  $\mathbf{H}_{\varphi_{i|t}} \mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  is a diagonal matrix whose diagonal entries are either 0 or 1. Therefore,  $\left\| \mathbf{G}_{\varphi_{i|t}} \right\| = 1 / \sqrt{N_{\varphi_{i|t}}}$ . Solutions  $\mathbf{G}_{\varphi_{i|t}}$  and  $\mathbf{H}_{\varphi_{i|t}}$  are feasible to problem (21). Thus, we've proved problem (21) is feasible. Furthermore, to avoid notation complications, in the rest of the proof, we use characters without subscripts for notation, for example, we use  $\mathbf{G}$  and  $\mathbf{H}$  instead of  $\mathbf{G}_{\varphi_{i|t}}$  and  $\mathbf{H}_{\varphi_{i|t}}$ , respectively.

Since problem (21) is convex and feasible, Slater's condition holds and optimal solution  $\mathbf{G}^*$  and  $\mathbf{H}^*$  exists. Therefore we can use Lagrangian duality to characterize the optimal solution. We start by transforming constraints (21b) and (21c) into more tractable constraints: Firstly, constraint (21b) can be equivalently transformed into constraint:

$$(\mathbf{G} + \mathbf{H}\mathbf{C})\boldsymbol{\nu}u_k = u_k, \quad k = 1, 2, \dots, n, \quad (27)$$

where  $u_k \in \mathbb{R}^n$  is a vector which has 1 in  $k$ -th entry and 0 otherwise. Furthermore, let arbitrarily choose unit vec-

tors (namely, vector having norm 1)  $z_l$  for  $l = 1, \dots, L$ , with  $L \geq 1$  arbitrarily chosen, and consider new constraint (which is a relaxed constraint (21c))

$$z_l^T \mathbf{G} \mathbf{G}^T z_l \leq \gamma^2 / N_{\varphi_{i|t}}, \quad l = 1, \dots, L. \quad (28)$$

Clearly, by increasing  $L$  to infinity, constraint (28) is equivalent to constraint (21c).

It follows that the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for optimality. Consider the Lagrangian multipliers  $\boldsymbol{\mu} = [\mu_1, \dots, \mu_n]$  for (27) and  $\boldsymbol{\zeta} = [\zeta_1, \dots, \zeta_L]$  for (28), then we have

$$\begin{aligned} & \mathcal{L}(\mathbf{G}, \mathbf{H}, \boldsymbol{\mu}, \boldsymbol{\zeta}) \\ = & \text{tr}(\mathbf{G} \mathbf{P} \mathbf{G}^T + \mathbf{H} \mathbf{Q} \mathbf{H}^T) + \sum_{k=1}^n \mu_k^T ((\mathbf{G} + \mathbf{H} \mathbf{C}) \boldsymbol{\nu} u_k - u_k) \\ & + \sum_{l=1}^L \zeta_l (\text{tr}(\mathbf{G} \mathbf{G}^T M_l) - \gamma^2 / N_{\varphi_{i|t}}), \end{aligned} \quad (29)$$

where  $M_l = z_l z_l^T$ , and  $\text{tr}(\mathbf{G} \mathbf{G}^T M_l) = z_l^T \mathbf{G} \mathbf{G}^T z_l$ . Thus, KKT conditions are

$$(\mathbf{G} + \mathbf{H} \mathbf{C}) \boldsymbol{\nu} u_k = u_k, \quad k = 1, \dots, n \quad (30)$$

$$\text{tr}(\mathbf{G} \mathbf{G}^T M_l) - \gamma^2 / N_{\varphi_{i|t}} \leq 0, \quad l = 1, \dots, L \quad (31)$$

$$\zeta_l (\text{tr}(\mathbf{G} \mathbf{G}^T M_l) - \gamma^2 / N_{\varphi_{i|t}}) = 0, \zeta_l \geq 0, \forall l \quad (32)$$

$$2\mathbf{G}(\mathbf{P} + \mathbf{M}_i) + \sum_{k=1}^n \mu_k u_k^T \boldsymbol{\nu}^T = 0 \quad (33)$$

$$2\mathbf{H} \mathbf{Q} + \sum_{k=1}^n \mu_k u_k^T \boldsymbol{\nu}^T \mathbf{C}^T = 0, \quad (34)$$

where  $\mathbf{M}_i = \sum_{l=1}^L \zeta_l M_l$ . The last two KKT conditions follow from

$$\nabla_{\mathbf{G}} \mathcal{L}(\mathbf{G}, \mathbf{H}, \boldsymbol{\mu}, \boldsymbol{\zeta}) = 0, \quad \nabla_{\mathbf{H}} \mathcal{L}(\mathbf{G}, \mathbf{H}, \boldsymbol{\mu}, \boldsymbol{\zeta}) = 0. \quad (35)$$

KKT conditions (33) and (34) imply that

$$\mathbf{G} = -\frac{1}{2} \sum_{k=1}^n \mu_k u_k^T \boldsymbol{\nu}^T (\mathbf{P} + \mathbf{M}_i)^\dagger, \quad (36a)$$

$$\mathbf{H} = -\frac{1}{2} \sum_{k=1}^n \mu_k u_k^T \boldsymbol{\nu}^T \mathbf{C}^T \mathbf{Q}^\dagger. \quad (36b)$$

By inserting (36) into (27), we have

$$-\frac{1}{2} \sum_{k=1}^n \mu_k u_k^T \mathbf{S} u_q = u_q, \quad q = 1, \dots, n, \quad (37)$$

where  $\mathbf{S}$  is  $\mathbf{S}_{\varphi_{i|t}}$  defined by (26). Since  $u_k^T \mathbf{S} u_q = [\mathbf{S}]_{kq}$ , from (37) we have the following equations

$$\mu_1 [\mathbf{S}]_{11} + \mu_2 [\mathbf{S}]_{21} + \dots + \mu_n [\mathbf{S}]_{n1} = -2u_1, \quad (38)$$

\vdots

$$\mu_1 [\mathbf{S}]_{1n} + \mu_2 [\mathbf{S}]_{2n} + \dots + \mu_n [\mathbf{S}]_{nn} = -2u_n, \quad (39)$$

which can be written in a compact form  $\boldsymbol{\mu} \mathbf{S} = -2\mathbf{I}_n$ . This gives  $\boldsymbol{\mu} = -2\mathbf{S}^\dagger$ . Note that after simple manipulations, we have  $\sum_{k=1}^n \mu_k u_k^T = \boldsymbol{\mu}$ , which implies that

$$\sum_{k=1}^n \mu_k u_k^T \boldsymbol{\nu}^T = \boldsymbol{\mu} \boldsymbol{\nu}^T = -2\mathbf{S}^\dagger \boldsymbol{\nu}^T. \quad (40)$$

By inserting (40) into (36), we have (25) by simple algebra. Moreover, since all  $\zeta_l \geq 0$  and  $M_l = z_l z_l^T$  are positive semidefinite positive,  $\mathbf{M}_i$  is also positive semidefinite. Finally, as indicated by the KKT condition (32), either  $\zeta_l = 0$  or (28) holds at equality, which is equivalent to  $\|\mathbf{G}_{\varphi_{i|t}}^*\| = \gamma / \sqrt{N_{\varphi_{i|t}}}$ . It completes the proof.  $\square$

It is worth seeking computable bounds on the optimal value of problem (21). This allows us to efficiently characterize the performance gains of the estimator obtained by the semidefinite programming. Hence, let us consider a tightened version of problem (21), which allows us to derive an upper-bound on the optimal value. In particular, instead of the (21c), we consider the following constraint:

$$\text{tr}(\mathbf{G}_{\varphi_{i|t}} \mathbf{G}_{\varphi_{i|t}}^T) \leq \gamma_{\max}^2 / N_{\varphi_{i|t}}. \quad (41)$$

Note that the optimal value of the tightened problem can be computed in closed-form. This result is established in the following proposition:

**Proposition 3.2** *Consider optimization problem*

$$\min_{\mathbf{G}_{\varphi_{i|t}}, \mathbf{H}_{\varphi_{i|t}}} \text{tr}(\text{Cov}_{\mathbf{v}} e_i(t)) \quad (42a)$$

$$\text{s.t. } (\mathbf{G}_{\varphi_{i|t}} + \mathbf{H}_{\varphi_{i|t}} \mathbf{C}_i) \boldsymbol{\nu}_{\varphi_{i|t}} = \mathbf{I}_n \quad (42b)$$

$$\text{tr}(\mathbf{G}_{\varphi_{i|t}} \mathbf{G}_{\varphi_{i|t}}^T) \leq \gamma_{\max}^2 / N_{\varphi_{i|t}}. \quad (42c)$$

Then, the optimal solutions are

$$\mathbf{G}_{\varphi_{i|t}}^* = \tilde{\mathbf{S}}_{\varphi_{i|t}}^\dagger \boldsymbol{\nu}_{\varphi_{i|t}}^T (\mathbf{P}_{\varphi_{i|t}} + \lambda_i^* \mathbf{I})^\dagger, \quad (43a)$$

$$\mathbf{H}_{\varphi_{i|t}}^* = \tilde{\mathbf{S}}_{\varphi_{i|t}}^\dagger \boldsymbol{\nu}_{\varphi_{i|t}}^T \mathbf{C}_i^T \mathbf{Q}_{\varphi_{i|t}}^\dagger, \quad (43b)$$

where  $\tilde{\mathbf{S}}_{\varphi_{i|t}} = \boldsymbol{\nu}_{\varphi_{i|t}}^T ((\mathbf{P}_{\varphi_{i|t}} + \lambda_i^* \mathbf{I})^\dagger + \mathbf{C}_i^T \mathbf{Q}_{\varphi_{i|t}}^\dagger \mathbf{C}_i) \boldsymbol{\nu}_{\varphi_{i|t}}$ , and  $\lambda_i^*$  is either 0, or the solution to the constraint (42c)

at equality. Moreover,  $\lambda_i^*$  is located within the interval

$$\left[0, \frac{\sqrt{n}N_{\varphi_{i|t}}}{\gamma} \text{tr} \boldsymbol{\vartheta}_{\varphi_{i|t}}^\dagger - \ell_m(\mathbf{P}_{\varphi_{i|t}})\right], \quad (44)$$

in which  $\boldsymbol{\vartheta}_{\varphi_{i|t}} = \boldsymbol{\nu}_{\varphi_{i|t}}^T \mathbf{C}_i^T \mathbf{Q}_{\varphi_{i|t}}^\dagger \mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$ .

We need two intermediate result to prove this proposition.

**Lemma 3.3** [45] *Let  $B, D \in \mathbb{R}^{l \times l}$  be positive semidefinite matrices with finite trace norm. Then*

$$\text{tr}(BD) \leq \text{tr} B \text{tr} D. \quad (45)$$

**Lemma 3.4** *Let  $B \in \mathbb{R}^{l \times l}$  be positive semidefinite matrices with finite and nonzero trace norm, and let  $Z \in \mathbb{R}^{l \times m}$ . Then*

$$\|BZ\|_F^2 \leq \ell_M^2(B) \|Z\|_F^2 = \ell_m^{-2}(B^\dagger) \|Z\|_F^2 \quad (46)$$

*Proof:* Let  $Z = (z_1, \dots, z_m)$ , we have

$$\begin{aligned} \|BZ\|_F^2 &= \sum_{k=1}^m \|Bz_k\|^2 \leq \sum_{k=1}^m \|B\|^2 \|z_k\|^2 = \|B\|^2 \|Z\|_F^2 \\ &= \ell_M^2(B) \|Z\|_F^2 = \ell_m^{-2}(B^\dagger) \|Z\|_F^2, \end{aligned} \quad (47)$$

which completes the proof.  $\square$

Now we are in the position to prove Proposition 3.2.

*Proof:* By following similar steps as those used in the proof of Proposition 3.1, we can find  $\mathbf{G}_{\varphi_{i|t}}^*$  and  $\mathbf{H}_{\varphi_{i|t}}^*$  as (43), where  $\lambda_i^*$  is the optimal solution of the Lagrangian multiplier corresponding to constraint (42c). Indicated by the KKT conditions, either  $\lambda_i^* = 0$ , or constraint (42c) holds at equality. In the rest of the proof, we use the same notations used in proof of Proposition 3.1 for notational convenience.

Now consider  $\mathbf{G}^*$  from (43), we have

$$\begin{aligned} \text{tr}(\mathbf{G}^* \mathbf{G}^{*T}) &= \text{tr}(\tilde{\mathbf{S}}^{\dagger 2} \boldsymbol{\nu}^T (\mathbf{P} + \lambda_i^* \mathbf{I})^{-2} \boldsymbol{\nu}) \\ &\leq \text{tr}(\tilde{\mathbf{S}}^{\dagger 2}) \text{tr}(\boldsymbol{\nu}^T (\mathbf{P} + \lambda_i^* \mathbf{I})^{-2} \boldsymbol{\nu}) \end{aligned} \quad (48a)$$

$$\begin{aligned} &= \text{tr}(\tilde{\mathbf{S}}^{\dagger 2}) \|\mathbf{P} + \lambda_i^* \mathbf{I}\|_F^{-2} \leq \text{tr}^2 \boldsymbol{\vartheta}^\dagger \|\mathbf{P} + \lambda_i^* \mathbf{I}\|_F^{-2} \\ &\leq \text{tr}^2 \boldsymbol{\vartheta}^\dagger \ell_m^{-2}(\mathbf{P} + \lambda_i^* \mathbf{I}) n N_{\varphi_{i|t}}, \end{aligned} \quad (48b)$$

where inequality (48a) comes from Lemma 3.3, and inequality (48b) comes from Lemma 3.4 together with fact that  $\|\boldsymbol{\nu}_{\varphi_{i|t}}\|_F^2 = n N_{\varphi_{i|t}}$ . By enforcing

$$n N_{\varphi_{i|t}} \text{tr}^2 \boldsymbol{\vartheta}^\dagger \ell_m^{-2}(\mathbf{P} + \lambda_i^* \mathbf{I}) \leq \gamma^2 / N_{\varphi_{i|t}}, \quad (49)$$

---

**Algorithm 1** Estimation for node  $i$

---

1.  $\hat{\mathbf{P}}_{\varphi_{i|1}} = \mathbf{P}_0, \hat{x}_i(0) = \mathbb{E}x(0)$
  2. **while true do**
  3. Broadcast  $(\hat{x}_i(t-1), y_i(t))$  to neighbors
  4. Receive  $(\hat{x}_j(t-1), y_j(t))$  from node  $j \in \mathcal{N}_i$
  5. Identify node  $j$  for  $C_j$  and  $R_j$
  6. Find  $\hat{\mathbf{P}}_{\varphi_{i|t}}$  as in Appendix A.
  7. Find  $\mathbf{G}_{\varphi_{i|t}}^*, \mathbf{H}_{\varphi_{i|t}}^*$  by solving problem (21)
  8.  $\hat{x}_i(t) = \mathbf{G}_{\varphi_{i|t}}^* \mathbf{A} \hat{x}_i(t-1) + \mathbf{H}_{\varphi_{i|t}}^* \mathbf{y}_i(t)$
  9. Sense the system, produce  $y_i(t+1)$ .
  10. **end while**
- 

we obtain  $\lambda_i^* \geq \frac{\sqrt{n}N_{\varphi_{i|t}}}{\gamma} \text{tr} \boldsymbol{\vartheta}^\dagger - \ell_m \mathbf{P}$ , (50)

which indicates that for all these values of  $\lambda_i^*$ , constraint (42c) holds at inequality. This implies that  $\lambda^*$  must be in the interval given by (44), which completes the proof.  $\square$

Proposition 3.1 gives an interval within which simple search algorithms, such as a bisection algorithm, can be used to find  $\lambda_i^*$  numerically. Note that the solution (43) is not the same as solution (25), since constraint (42c) and (21c) are not equivalent. However, if constraint (42c) holds then (21c) holds as well.

The whole process of the estimation is summarized in Algorithm 1. When node  $i$  makes the measurements  $y_i(t)$  at time  $t$ , it broadcasts them together with the previous estimates  $\hat{x}_i(t-1)$ . Meanwhile, node  $i$  receives others' estimates and measurements. Node  $i$  estimates the estimation error  $\hat{\epsilon}_i(t)$  and error covariance matrix  $\hat{\mathbf{P}}_{\varphi_{i|t}}$  as in Appendix A. Then by using semidefinite programming, node  $i$  can find the optimal weights  $\mathbf{G}_{\varphi_{i|t}}^*$  and  $\mathbf{H}_{\varphi_{i|t}}^*$  to find the estimates  $\hat{x}_i(t)$  at time  $t$ . The nodes repeats this process along time.

We now turn our attention to the analysis of the fundamental performance limitations of our estimator.

## 4 Performance Analysis

In this section we characterize the performance of our estimator. We highlight the dependence of the estimator's performance on the message losses. We investigate the variance of the estimation error.

Recall that we use  $\text{Cov}_{\mathbf{v}} e_i(t)$  denote the covariance matrix with respect to the distribution of the noise  $\mathbf{v}$ . Now, we have the following corollary:

**Corollary 4.1** *Consider the optimal solutions  $\mathbf{G}_{\varphi_{i|t}}^*$  and  $\mathbf{H}_{\varphi_{i|t}}^*$  of problem (21). Then the covariance matrix*

of estimation error satisfies

$$\text{Cov}_{\mathbf{v}} e_i(t) \leq \left( \boldsymbol{\nu}_{\varphi_{i|t}}^T \mathbf{C}_i^T \mathbf{Q}_{\varphi_{i|t}}^\dagger \mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}} \right)^\dagger. \quad (51)$$

*Proof:* Note that in the following proof, we remove the subscripts for notational convenience. By inserting (25) into (22), we have

$$\begin{aligned} & \text{Cov}_{\mathbf{v}} e_i(t) \\ &= \mathbf{S}^\dagger \boldsymbol{\nu}^T \left( (\mathbf{P} + \mathbf{M}_i)^\dagger \mathbf{P} (\mathbf{P} + \mathbf{M}_i)^\dagger + \mathbf{C}^T \mathbf{Q}^\dagger \mathbf{C} \right) \boldsymbol{\nu} \mathbf{S}^\dagger \\ &\leq \mathbf{S}^\dagger \boldsymbol{\nu}^T \left( (\mathbf{P} + \mathbf{M}_i)^\dagger + \mathbf{C}^T \mathbf{Q}^\dagger \mathbf{C} \right) \boldsymbol{\nu} \mathbf{S}^\dagger \\ &= \mathbf{S}^\dagger \mathbf{S} \mathbf{S}^\dagger = \left( \boldsymbol{\nu}^T \left( (\mathbf{P} + \mathbf{M}_i)^\dagger + \mathbf{C}^T \mathbf{Q}^\dagger \mathbf{C} \right) \boldsymbol{\nu} \right)^\dagger \\ &\leq \left( \boldsymbol{\nu}^T \mathbf{C}^T \mathbf{Q}^\dagger \mathbf{C} \boldsymbol{\nu} \right)^\dagger, \end{aligned} \quad (52)$$

The first inequality come from the fact that both matrices  $\mathbf{P}$  and  $\mathbf{M}_i$  are positive semidefinite. It completes the proof.  $\square$

Note that the previous result provides a rather conservative bound, since we do not use any knowledge about the covariance matrix  $\mathbf{P}_{\varphi_{i|t}}$ . Proposition 3.2 helps us to improve the bound in the following corollary. Recall that we've defined  $\boldsymbol{\vartheta}_{\varphi_{i|t}} = \boldsymbol{\nu}_{\varphi_{i|t}}^T \mathbf{C}_i^T \mathbf{Q}_{\varphi_{i|t}}^\dagger \mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}}$  in Proposition 3.2.

**Corollary 4.2** *Consider the optimal solutions  $\mathbf{G}_{\varphi_{i|t}}^*$  and  $\mathbf{H}_{\varphi_{i|t}}^*$  of problem (21). Then, the covariance matrix of estimation error satisfies*

$$\text{Cov}_{\mathbf{v}} e_i(t) \leq \left( \boldsymbol{\psi}_{\varphi_{i|t}} + \boldsymbol{\vartheta}_{\varphi_{i|t}} \right)^\dagger, \quad (53)$$

$$\text{where } \boldsymbol{\psi}_{\varphi_{i|t}} = \frac{\boldsymbol{\nu}_{\varphi_{i|t}}^T \boldsymbol{\nu}_{\varphi_{i|t}}}{\|A\|_F^2 \sum_{j \in \mathcal{N}_{\varphi_{i|t}}} \text{tr } \boldsymbol{\vartheta}_{\varphi_{j|t-1}}^\dagger + \frac{\sqrt{n} N_{\varphi_{i|t}}}{\gamma} \text{tr } \boldsymbol{\vartheta}_{\varphi_{i|t}}^\dagger}.$$

*Proof:* Since problem (42) is the tightened version of problem (21), thus we can prove this corollary by showing that the resulting covariance matrix of problem (42) satisfies (53). Thus, we consider the optimal solutions  $\mathbf{G}_{\varphi_{i|t}}^*$  and  $\mathbf{H}_{\varphi_{i|t}}^*$  of problem (42) instead of (21).

By following similar steps as those used in the proof of Corollary 4.1, we obtain that

$$\begin{aligned} & \text{Cov}_{\mathbf{v}} e_i(t) \\ &\leq \left( \boldsymbol{\nu}_{\varphi_{i|t}}^T \left( (\mathbf{P}_{\varphi_{i|t}} + \lambda_i^* \mathbf{I})^\dagger + \mathbf{C}_i^T \mathbf{Q}_{\varphi_{i|t}}^\dagger \mathbf{C}_i \right) \boldsymbol{\nu}_{\varphi_{i|t}} \right)^\dagger. \end{aligned} \quad (54)$$

Now note that

$$\begin{aligned} \text{tr } \mathbf{P}_{\varphi_{i|t}} &= \text{tr} \left( \mathbf{A}_i \text{Cov}_{\mathbf{v}} \mathbf{e}_i(t-1) \mathbf{A}_i^T \right) \circ \left( \varphi_{i|t} \varphi_{i|t}^T \otimes \mathbf{1}_n \mathbf{1}_n^T \right) \\ &= \sum_{j \in \mathcal{N}_i} \text{tr} \left( A \text{Cov}_{\mathbf{v}} e_j(t-1) A^T \right) \leq \sum_{j \in \mathcal{N}_i} \text{tr} \left( A \boldsymbol{\vartheta}_{\varphi_{j|t-1}}^\dagger A^T \right) \\ &\leq \sum_{j \in \mathcal{N}_i} \text{tr} (A^T A) \text{tr } \boldsymbol{\vartheta}_{\varphi_{j|t-1}}^\dagger = \sum_{j \in \mathcal{N}_i} \|A\|_F^2 \text{tr } \boldsymbol{\vartheta}_{\varphi_{j|t-1}}^\dagger, \end{aligned} \quad (55)$$

where the first inequality comes from Corollary 4.1, whereas the second inequality comes from Lemma 3.4. Since  $\mathbf{P}_{\varphi_{i|t}}$  is positive semidefinite, then its maximum eigenvalue is smaller than its trace. Thus, we have

$$\begin{aligned} \ell_M \left( \mathbf{P}_{\varphi_{i|t}} + \lambda_i^* \mathbf{I} \right) &\leq \text{tr } \mathbf{P}_{\varphi_{i|t}} + \lambda_i^* \\ &\leq \sum_{j \in \mathcal{N}_{\varphi_{i|t}}} \|A\|_F^2 \text{tr } \boldsymbol{\vartheta}_{\varphi_{j|t-1}}^\dagger + \frac{\sqrt{n} N_{\varphi_{i|t}}}{\gamma} \text{tr } \boldsymbol{\vartheta}_{\varphi_{i|t}}^\dagger, \end{aligned} \quad (56)$$

where the second inequality comes from (44) in Proposition 3.2. Now we can conclude that

$$\begin{aligned} & \boldsymbol{\nu}_{\varphi_{i|t}}^T \left( \mathbf{P}_{\varphi_{i|t}} + \lambda_i^* \mathbf{I} \right)^\dagger \boldsymbol{\nu}_{\varphi_{i|t}} \\ &\geq \ell_M^{-1} \left( \mathbf{P}_{\varphi_{i|t}} + \lambda_i^* \mathbf{I} \right) \boldsymbol{\nu}_{\varphi_{i|t}}^T \boldsymbol{\nu}_{\varphi_{i|t}} \\ &= \frac{\boldsymbol{\nu}_{\varphi_{i|t}}^T \boldsymbol{\nu}_{\varphi_{i|t}}}{\|A\|_F^2 \sum_{j \in \mathcal{N}_{\varphi_{i|t}}} \text{tr } \boldsymbol{\vartheta}_{\varphi_{j|t-1}}^\dagger + \frac{\sqrt{n} N_{\varphi_{i|t}}}{\gamma} \text{tr } \boldsymbol{\vartheta}_{\varphi_{i|t}}^\dagger}, \end{aligned} \quad (57)$$

which together with (54) completes the proof.  $\square$

The previous corollary guarantees that the estimation error at time  $t$  and in any node  $i$ , is always upper bounded by a term that depends on the message loss process  $\varphi_{i|t}$ . By using the corollary, we can compute the worst case performance with respect to the message loss process. With this goal in mind, we need two intermediate technical results. Now suppose there exists  $\sigma > 0$  such that

$$\mathbf{C}_i^T \mathbf{R}_i^{-1} \mathbf{C}_i \geq \sigma^{-1} \mathbf{I} \quad (58)$$

for any node  $i$ , thus we have the following lemma:

**Lemma 4.1** *Consider the optimal solutions  $\mathbf{G}_{\varphi_{i|t}}^*$  and  $\mathbf{H}_{\varphi_{i|t}}^*$  of problem (21). Then, for any message loss realization  $\varphi_{i|t}$ ,*

$$\ell_M \left( \mathbf{P}_{\varphi_{i|t}} + \lambda_i^* \mathbf{I} \right) \leq n \sigma \left( \|A\|_F^2 N_i + \frac{\sqrt{n}}{\gamma} \right). \quad (59)$$

*Proof:* From (58), we have

$$\boldsymbol{\vartheta}_{\varphi_{i|t}}^\dagger = \left( \boldsymbol{\nu}_{\varphi_{i|t}}^T \mathbf{C}_i^T \mathbf{Q}_{\varphi_{i|t}}^\dagger \mathbf{C}_i \boldsymbol{\nu}_{\varphi_{i|t}} \right)^\dagger \leq \frac{\sigma}{N_{\varphi_{i|t}}} \mathbf{I}, \quad (60)$$



which implies that the  $j$ -th  $n$  by  $n$  diagonal block of  $\mathbf{P}_{\varphi_{i|t}}$  are less than  $\|A\|^2 \sigma / N_{\varphi_{j|t-1}} \mathbf{I}$ . Assume that each diagonal element of a covariance matrix is the largest value along its row. Hence, it follows that

$$\ell_M \mathbf{P}_{\varphi_{i|t}} \leq \|A\|^2 \sum_{j \in \mathcal{N}_{\varphi_{i|t}}} \frac{n\sigma}{N_{\varphi_{j|t-1}}} \leq \|A\|^2 n\sigma N_i, \quad (61)$$

since  $N_{\varphi_{j|t-1}} \geq 1$ . Together with (44), we have

$$\ell_M (\mathbf{P}_{\varphi_{i|t}} + \lambda_i^* \mathbf{I}) \leq n\sigma N_i \|A\|^2 + \frac{\sqrt{n} N_{\varphi_{i|t}}}{\gamma} \text{tr} \boldsymbol{\vartheta}_{\varphi_{i|t}}^\dagger. \quad (62)$$

The proof is completed by inserting (60) into (62).  $\square$

**Lemma 4.2** For any message loss realization  $\varphi_{i|t}$ ,

$$\mathbb{E}_{\varphi} [\varphi_{i|t}^T \varphi_{i|t}]^{-1} = \sum_{k=0}^{N_i-1} \frac{\chi(k)}{k+1}, \quad (63)$$

where

$$\chi(k) = \sum_{l=1}^{\binom{N_i-1}{k}} \left( \prod_{n=1}^k q_{is(n)} \cdot \prod_{m=k+1}^{N_i-1} p_{is(m)} \right), \quad (64)$$

and the function  $s: \{1, 2, \dots, N_i - 1\} \rightarrow \{1, 2, \dots, N_i - 1\}$  is a permutation. Namely the  $k$ -th coefficient of the polynomial (63) is the sum of  $\binom{N_i-1}{k}$  terms in which there are  $k$  factors  $q_{ij}$  and  $N_i - 1 - k$  factors  $p_{ir}$  with  $j \neq r$ .

*Proof:* Random variable  $\varphi_{i|t}^T \varphi_{i|t} = 1 + \sum_{j=1, j \neq i}^{N_i} \varphi_{ij|t}$  is given by the sum of  $N_i - 1$  independent Bernoulli random variables having different parameter. Then, we have [46]

$$\mathbb{E}_{\varphi} [\varphi_{i|t}^T \varphi_{i|t}]^{-1} = \int_0^1 g_1(z) dz, \quad (65)$$

where  $g_1(z)$  is defined as

$$\begin{aligned} g_1(z) &= \mathbb{E} z^{\varphi_{i|t}^T \varphi_{i|t} - 1} = \prod_{\substack{j=1 \\ j \neq i}}^{N_i} \mathbb{E} z^{\varphi_{ij|t}} = \prod_{\substack{j=1 \\ j \neq i}}^{N_i} (q_{ij} + p_{ij} z) \\ &= \chi_0 + \chi_1 z + \dots + \chi_{N_i-1} z^{N_i-1} \end{aligned} \quad (66)$$

where the last equality is achieved by developing the product of terms  $q_{ij} + p_{ij} z$  in a polynomial in the general form. After tedious manipulations, we see that the coefficients of the polynomial are given by (64). By using  $g_1(z)$  in the integral (65), we obtain the result.  $\square$

We are now ready to give the following important result.

**Proposition 4.1** Consider the optimal solutions  $\mathbf{G}_{\varphi_{i|t}}^*$  and  $\mathbf{H}_{\varphi_{i|t}}^*$  of problem (21). Then for any message loss realization  $\varphi_{i|t}$ ,

$$\begin{aligned} &\mathbb{E}_{\varphi} \text{Cov}_{\mathbf{v}} e_i(t) \\ &\leq \frac{n\sigma \mathbf{I}}{\left(\|A\|^2 N_i + \frac{\sqrt{n}}{\gamma}\right)^{-1} + n} \sum_{k=0}^{N_i-1} \frac{\chi(k)}{k+1}. \end{aligned} \quad (67)$$

*Proof:* We use the previous two lemmas to prove this proposition. By considering (54) together with (58), we have

$$\begin{aligned} &\mathbb{E}_{\varphi} \text{Cov}_{\mathbf{v}} e_i(t) \\ &\leq \mathbb{E}_{\varphi} \left( \boldsymbol{\nu}_{\varphi_{i|t}}^T (\mathbf{P}_{\varphi_{i|t}} + \lambda_i^* \mathbf{I})^\dagger \boldsymbol{\nu}_{\varphi_{i|t}} + \sigma^{-1} \boldsymbol{\nu}_{\varphi_{i|t}}^T \boldsymbol{\nu}_{\varphi_{i|t}} \right)^\dagger. \end{aligned} \quad (68)$$

Moreover, from Lemma 4.1, we have

$$\begin{aligned} &\boldsymbol{\nu}_{\varphi_{i|t}}^T (\mathbf{P}_{\varphi_{i|t}} + \lambda_i^* \mathbf{I})^\dagger \boldsymbol{\nu}_{\varphi_{i|t}} \geq \ell_M^{-1} \left( \mathbf{P}_{\varphi_{i|t}} + \lambda_i^* \mathbf{I} \right) \boldsymbol{\nu}_{\varphi_{i|t}}^T \boldsymbol{\nu}_{\varphi_{i|t}} \\ &\geq \frac{\boldsymbol{\nu}_{\varphi_{i|t}}^T \boldsymbol{\nu}_{\varphi_{i|t}}}{n\sigma \left(\|A\|^2 N_i + \frac{\sqrt{n}}{\gamma}\right)}. \end{aligned} \quad (69)$$

Thus, we have

$$\mathbb{E}_{\varphi} \text{Cov}_{\mathbf{v}} e_i(t) \leq \frac{\mathbb{E}_{\varphi} [\varphi_{i|t}^T \varphi_{i|t}]^{-1} n\sigma \mathbf{I}}{\left(\|A\|^2 N_i + \frac{\sqrt{n}}{\gamma}\right)^{-1} + n}. \quad (70)$$

The proposition follows by invoking Lemma 4.2.  $\square$

Observe that the estimation error variance given by Proposition 4.1 depends on the message loss probabilities  $q_{ij}$ , on the maximum number of neighbors for each node  $N_i$ , on the dimension of the state  $n$ , and on the largest singular value of the matrix  $A$ . Note that in (67), the first factor  $\frac{n\sigma \mathbf{I}}{\left(\|A\|^2 N_i + \frac{\sqrt{n}}{\gamma}\right)^{-1} + n}$  is always smaller than  $\sigma \mathbf{I}$ , whereas the second factor clearly depends on the value attained by the various  $q_{ij}$ . We consider here the simple case when  $q_{ij} = q$  for all  $i, j$ , which gives

$$\sum_{k=0}^{N_i-1} \frac{\chi(k)}{k+1} = \frac{1 - q^{N_i}}{(1 - q)N_i}. \quad (71)$$

In Fig. 2, we have plotted such a function for various values of  $q$  and  $N_i$ . The function decreases very quickly as the maximum number of neighbors of a node increases, for all values of  $q$  (notice that we have considered that  $q = 0.3$  at most, namely a message loss probability of 30%). This is rather intuitive, since as the number of neighbors increases message losses have less impact on the estimation and thus better performance are

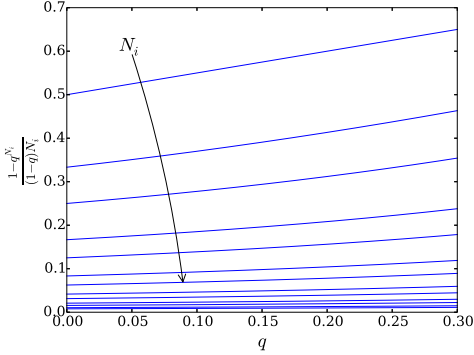


Fig. 2. Second factor of (67) as function of  $q$  for increasing values of  $N_i$  ranging from 2 to 100. The factor is always less than 0.8. The smallest values are achieved when  $q$  is small and  $N_i$  is large. This is explained by that the message loss probability has a decreasing negative effect when the number of neighbors of a node increases, which translates into a smaller value of the second factor of (67).

achieved. In particular, even when the first factor in (67) is very close to 1, and the number of neighbors is greater than 2, we have that the product of the two coefficients does not exceed 0.65 with a message loss of  $q = 0.3$ . It is only a 30% higher than the case when no message losses are present.

## 5 Computational Complexity Analysis

In this section, we characterize the computational complexity of the estimator we have proposed. Besides matrix/vector multiplication to update the current estimate, the computational complexity is mainly given by two components: the semidefinite programming and the estimation of matrix  $\mathbf{P}_{\varphi_{i|t}}$ .

Generally, the estimator requires matrix multiplications of size  $nN_{\varphi_{i|t}}$  at each node  $i$  at time  $t$ . Thus the complexity is  $\mathcal{O}(n^3 N_{\varphi_{i|t}}^3)$  using Gauss-Jordan elimination [47]. To simplify the analysis, we use this most classical algorithm to do the multiplication and the complexity results in  $\mathcal{O}(n^3 N_{\varphi_{i|t}}^3)$  for the matrices multiplication.

A worst case complexity result to solve the semidefinite programming problem (21) is the following [48]: Let  $c$  be the number of constraints. Then, the worst-case number of interior-point algorithm iterations to compute an  $\epsilon$ -solution of problem (21), meaning its objective value is at most  $\epsilon$  above the minimal one, where positive  $\epsilon \ll 1$ , is bounded by  $\mathcal{O}(\sqrt{n+c} \log 1/\epsilon)$ .

The computation for the matrix  $\mathbf{P}_{\varphi_{i|t}}$ , which is described in Appendix A, consists of those for quadratically constrained least-square problem and for matrix multiplication. The quadratically constrained least-square prob-

lem is used to find a value of  $\rho_i$  by the Generalized Cross Validation as in [35]. Note that when the static topology is initialized, we can use a  $N_i$ -by- $N_i$  look-up table to speed-up the computation of a quadratically constrained least-square problem (see [49]). Then, the computation is generally  $\mathcal{O}(N_i \log N_i)$  to create the table and  $\mathcal{O}(\log N_i)$  to pop the value.

Combining these three components, the total computational complexity can be found as

$$\mathcal{O}\left(n^3 N_i^3 \left(\sqrt{n+c} \log \frac{1}{\epsilon} + \log N_i\right)\right), \quad (72)$$

where  $c = 2$ . Note that in general  $\log N_i$  is much smaller than  $\sqrt{n+c} \log 1/\epsilon$ . In following section, we will present the numerical performance of the proposed estimator.

## 6 Numerical Experiments

In this section, the numerical results and communication cost of the proposed estimator are illustrated. Numerical simulations have been carried out to compare the estimator proposed in this paper with some related estimators available from the literature. We take the mean square error of the estimates of each node as performance measure. The mean square error (MSE) has been averaged over all nodes of the network. We consider the following estimators for comparison purposes. Note that for all the estimators the communication channels in simulations suffer from the messages losses.

$E_H$ : Let  $\mathbf{H}_i(t)$  be the solution of the linear least square problem at node  $i$ , which is equivalent to following optimization problem

$$\min_{\mathbf{h}_i(t)} \text{tr} \{ \mathbf{H}_{\varphi_{i|t}} \mathbf{Q}_{\varphi_{i|t}} \mathbf{H}_{\varphi_{i|t}}^T \} \quad (73a)$$

$$\text{s.t. } \mathbf{H}_{\varphi_{i|t}} \mathbf{C}_i \mathbf{v}_i = \mathbf{I}. \quad (73b)$$

This is a measurement diffusion method that is close to the strategy in literatures as [31].

$E_{AKF}$ : Let  $G_{ij}(t) = \mathbf{I}/N_{\varphi_{i|t}}$ , if node  $i$  and  $j$  communicate, and  $G_{ij}(t) = \mathbf{0}$  otherwise, while each node runs local Kalman filter based on its measurements only. The updated estimate is the average of the old estimates, which is close to the averaging strategy used in [1].

$E_{DKF}$ : Distributed Kalman filter, which was introduced in [20].

$E_{CKF}$ : Centralized Kalman filter, which gathers all the measurements and obtain the estimates by Kalman filter. Recall that the centralized Kalman filter always performs better in MSE than the distributed Kalman filter when  $w(t)$  is zero mean Gaussian noise.

$E_{H_\infty}$ :  $H_\infty$  filter, which gathers all the measurements, and could performs better in MSE than the centralized Kalman filter when  $w(t)$  is biased [39].

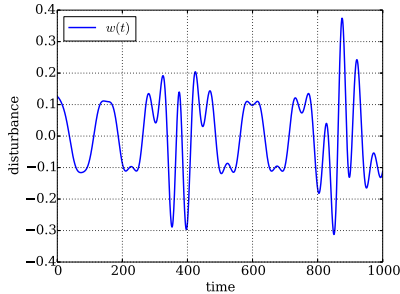


Fig. 3. Illustrations of the nonlinear disturbance from time 0 to 1000 with zero mean.

$E_P$ : The estimator proposed in this paper.

The system used in simulation is defined as

$$A = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.6 \end{pmatrix}, \quad (74)$$

while  $C_i$  is generated randomly as a 2-by-2 or 1-by-2 matrices with probability 0.5, 0.5 respectively. Furthermore, we consider three types of disturbance: the nonlinear disturbance as illustrated in Fig. 3, the zeros mean Gaussian noise, and the non-zero mean Gaussian noise. In the simulation, we let message loss rate  $q_{ij} = q$  for all  $i$  and  $j$ , and vary from 0 to 0.3.

Fig. 4 shows the MSE for all the estimators as a function of the message loss probability. Recall also that in all the simulations related to the proposed estimator  $E_P$ , the error covariance matrix is estimated locally at each node as described in Appendix A. As shown in Fig. 4,  $E_P$  outperforms all other estimators for any considered message loss probability in both nonlinear disturbance and non-zero mean Gaussian noise. Fig. 5 illustrates the first component of the signal, the estimates, and the corresponding square of error for  $E_{DKF}$ ,  $E_{CKF}$ ,  $E_{H\infty}$  and  $E_P$  with nonlinear disturbance and Gaussian noise, where the estimates come from an arbitrary node in the WSN. Note that for all estimators, the statistical parameters of additive disturbance are unknown, meaning that their knowledge is not assumed to derive the estimators. However, in the simulations, the true variances of the additive noises are used in  $E_{AKF}$ ,  $E_{DKF}$  and  $E_{CKF}$ .

Finally, the average computational time is approximately  $1 \sim 2$  seconds (MATLAB program on a common laptop) for the nodes at each iteration in  $E_P$ , which could be less than 0.01 second by C program with the same hardware. Moreover, the communication cost of  $E_P$  is slightly higher if compared with the estimators  $E_H$  and  $E_{AKF}$ , where nodes need exchange extra messages (the local estimates) among its neighbors. Informally, in  $E_P$  each node  $i$  needs 8 bytes more per iteration to broadcast, where 8 bytes accounts for extra measurements in

float-point (double). Suppose there are 100 nodes in a wireless sensor network, thus  $E_P$  needs  $8N = 800$  bytes more to broadcast, which is less than 1 kB.

## 7 Conclusions and Future Work

In this paper, we characterized the performance of a decentralized peer-to-peer algorithm for estimating a time-varying and multi-dimensional signal using a wireless sensor network. Measurements are corrupted by additive unknown and bounded disturbances, and communication occurs over channels with message losses. A mathematical framework was proposed to analyze the estimator, which runs locally in each node of the network. The performance analysis was carried out for networks with both symmetrical and asymmetrical message losses over the communication channels. We investigated how the estimation quality depends on message loss process, the network size and the average number of neighboring nodes. The theoretical analysis showed that the estimator converges, and the variance of the estimation error is bounded even in the presence of severe message loss probabilities. Numerical results illustrated the validity and the benefits of our approach, which outperforms other estimators available from the literature.

Future studies will be devoted to the extension of our design methodology to non-linear cases, while allowing to minimize the estimation variance over a future time horizon. Lossy communication links with memory will be considered. We also plan to consider the case of distributed control.

## References

- [1] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.
- [2] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, "Communication constraints in the average consensus problem," *Automatica*, vol. 44, no. 3, 2008.
- [3] F. Cattivelli and A. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, 2010.
- [4] F. Cattivelli and A. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 2069–2084, 2010.
- [5] J. Hu, L. Xie, and C. Zhang, "Diffusion Kalman filtering based on covariance intersection," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 891–902, 2012.
- [6] I. Matei and J. S. Baras, "Consensus-based linear distributed filtering," *Automatica*, vol. 48, no. 8, pp. 1776 – 1782, 2012.
- [7] P. Millán, L. Orihuela, C. Vivas, and F. R. Rubio, "Distributed consensus-based estimation considering network induced delays and dropouts," *Automatica*, vol. 48, no. 10, pp. 2726 – 2729, 2012.

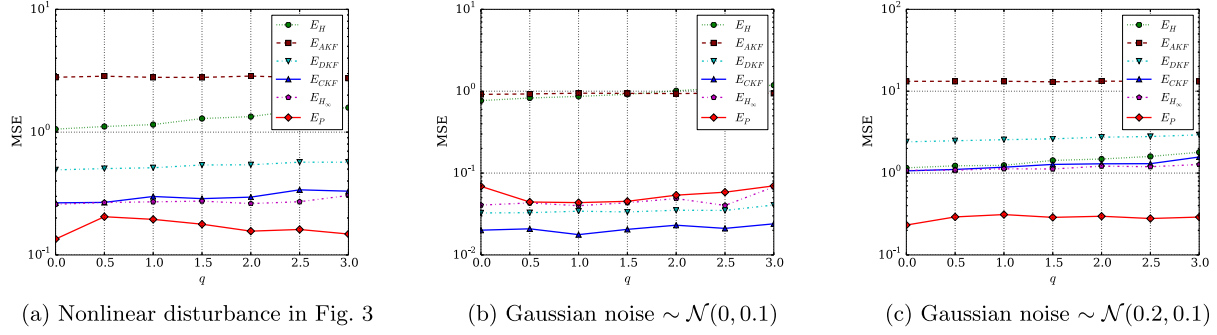


Fig. 4. Mean Square Error (MSE) performance comparison among estimators for various message loss probabilities  $q$ . Fig. (a) presents MSE vs.  $q$ , when nonlinear disturbance in Fig. 3 is considered, while Fig. (b) and (c) present MSE vs.  $q$  when Gaussian noise  $\mathcal{N}(0, 0.1)$  and  $\mathcal{N}(0.2, 0.1)$  are considered, respectively.

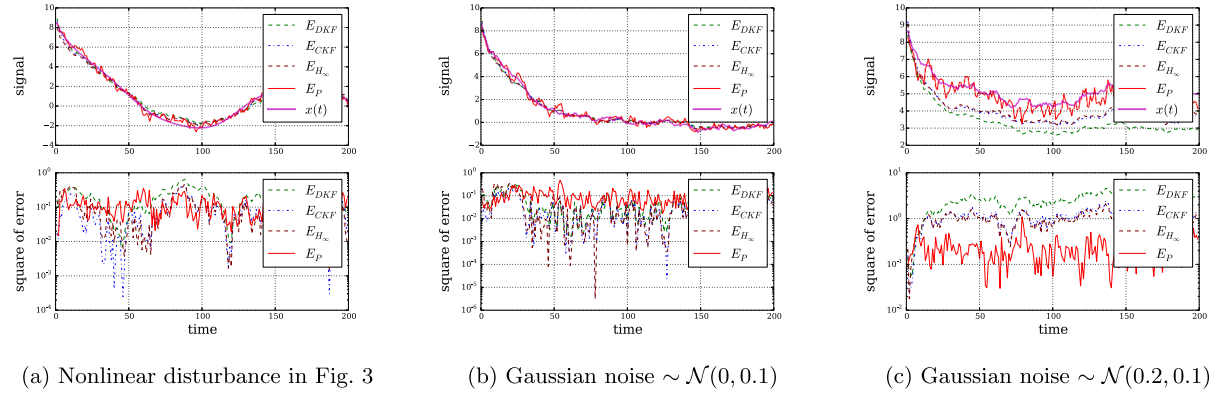


Fig. 5. Illustration of the first component of the signals, the estimates and the corresponding square of errors of  $E_{DKF}$ ,  $E_{CKF}$  and  $E_P$ . Fig. (a) presents the illustration when nonlinear disturbance in Fig. 3 is considered, while Fig. (b) and (c) present the illustration when Gaussian noise  $\mathcal{N}(0, 0.1)$  and  $\mathcal{N}(0.2, 0.1)$  are considered, respectively.

[8] I. D. Schizas, G. B. Giannakis, S. I. Roumeliotis, and A. Ribeiro, "Consensus in ad hoc WSNs with noisy links - part II: Distributed estimation and smoothing of random signals," *IEEE Transactions on Signal Processing*, vol. 56, pp. 1650–1666, 2008.

[9] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *System Control Letters*, vol. 53, pp. 65–78, 2004.

[10] L. Xiao, S. Boyd, and S. J. Kim, "Distributed average consensus with least-mean-square deviation," *Journal of Parallel and Distributed Computing*, vol. 67, pp. 33–46, 2007.

[11] P. Hovareshti, J. S. Baras, and V. Gupta, "Average consensus over small world networks: A probabilistic framework," in *Proceedings of IEEE Conference on Decision and Control*, 2008.

[12] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth-constrained sensor network," *IEEE Transactions on Information Theory*, vol. 4, pp. 829–832, 2005.

[13] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," in *Proceedings of IEEE Conference on Decision and Control*, 2007.

[14] Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz, "Localization from connectivity in sensor networks," *IEEE Transaction on Parallel and Distributed System*, Vol. 15, No. 11, 2004.

[15] L. Shi, *Resource optimization for networked estimator with guaranteed estimation quality*. PhD thesis, Caltech, 2008.

[16] Y. Liang, T. Chen, and Q. Pan, "Optimal linear state estimator with multiple packet dropouts," *IEEE Transactions on Automatic Control*, vol. 55, no. 6, pp. 1428–1433, 2010.

[17] J. Sijs and M. Lazar, "State fusion with unknown correlation: Ellipsoidal intersection," *Automatica*, vol. 48, no. 8, pp. 1874–1878, 2012.

[18] S. Kar, J. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Transactions On Information Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.

[19] P. Pasqualetti, R. Carli, and F. Bullo, "Distributed estimation via iterative projections with application to power network monitoring," *Automatica*, vol. 48, no. 5, pp. 747–758, 2012.

[20] R. Olfati-Saber and P. Jalalkamali, "Coupled distributed estimation and control for mobile sensor networks," *IEEE Transactions on Automatic Control*, vol. 57, no. 10, pp. 2609–2614, 2012.

[21] D. Ding, Z. Wang, H. Dong, and H. Shu, "Distributed H-infinity state estimation with stochastic parameters and

- nonlinearities through sensor networks: The finite-horizon case,” *Automatica*, vol. 48, no. 8, pp. 1575–1585, 2012.
- [22] Q. E. Quevedo, A. Ahlen, A. S. Leong, and S. Dey, “On kalman filtering over fading wireless channels with controlled transmission powers,” *Automatica*, vol. 48, no. 7, pp. 1306–1316, 2012.
- [23] C. Li, P. Shen, Y. Liu, and Z. Zhang, “Diffusion information theoretic learning for distributed estimation over network,” *IEEE Transactions on Signal Processing*, vol. 61, pp. 4011–4024, Aug 2013.
- [24] A. Chiuso and L. Schenato, “Information fusion strategies and performance bounds in packet-drop networks,” *Automatica*, vol. 47, no. 7, pp. 1304 – 1316, 2011.
- [25] W.-A. Zhang, G. Feng, and L. Yu, “Multi-rate distributed fusion estimation for sensor networks with packet losses,” *Automatica*, vol. 48, no. 9, pp. 2016 – 2028, 2012.
- [26] R. Olfati-Saber and R. M. Murray, “Consensus problems in networks of agents with switching topology and time-delays,” *IEEE Transactions on Automatic Control*, vol. 49, pp. 1520–1533, 2004.
- [27] R. Carli, G. Como, P. Frasca, and F. Garin, “Distributed averaging on digital erasure networks,” *Automatica*, vol. 47, no. 1, pp. 115–121, 2011.
- [28] R. Carli, P. Frasca, F. Fagnani, and S. Zampieri, “Gossip consensus algorithms via quantized communication,” *Automatica*, vol. 46, no. 1, pp. 70–80, 2010.
- [29] C. Liao and P. Barooah, “Distributed clock skew and offset estimation from relative measurements in mobile networks with markovian switching topology,” *Automatica*, vol. 49, no. 10, pp. 3015 – 3022, 2013.
- [30] V. Delouille, R. Neelamani, and R. G. Baraniuk, “Robust distributed estimation using the embedded subgraphs algorithm,” *IEEE Transactions on Signal Processing*, vol. 54, pp. 2998–3010, August 2006.
- [31] C. G. Lopes and A. H. Sayed, “Diffusion least-mean squares over adaptive networks: Formulation and performance analysis,” *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3135, 2008.
- [32] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, “Diffusion recursive least-squares for distributed estimation over adaptive networks,” *IEEE Transactions on Signal Processing*, vol. 56, pp. 1865–1877, 2008.
- [33] W. Russell, D. Klein, and J. Hespanha, “Optimal estimation on the graph cycle space,” *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2834–2846, 2011.
- [34] Y. Xu, C. Fischione, and A. Speranzon, “Model based peer-to-peer estimation on wireless sensor networks,” *In Proceedings of IFAC Workshop on Distributed Estimation and Control in Networked Systems (IFAC NecSys)*, 2012.
- [35] A. Speranzon, C. Fischione, K. H. Johansson, and A. Sangiovanni-Vincentelli, “A distributed minimum variance estimator for sensor networks,” *IEEE Journal on Selected Areas in Communications, Special Issue on Control and Communication*, vol. 26, pp. 609–621, 2008.
- [36] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
- [37] W. Li and Y. Jia, “H-infinity filtering for a class of nonlinear discrete-time systems based on unscented transform,” *Signal Processing*, vol. 90, pp. 3301–3307, Dec. 2010.
- [38] J. Zhang, Y. Xia, and P. Shi, “Parameter-dependent robust H infinity filtering for uncertain discrete-time systems,” *Automatica*, vol. 45, pp. 560–565, Feb. 2009.
- [39] D. S. U. Simon, *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*. John Wiley & Sons, 2006.
- [40] G. L. Stüber, *Principles of Mobile Communication*. Kluwer Academic Publishers, 1996.
- [41] G. Basile and G. Marro, “On the observability of linear, time-invariant systems with unknown inputs,” *Journal of Optimization Theory and Applications*, vol. 3, no. 6, pp. 410–415, 1969.
- [42] P. Kudva, N. Viswanadham, and A. Ramakrishna, “Observers for linear systems with unknown inputs,” *IEEE Transactions on Automatic Control*, vol. AC-25, no. 1, pp. 113–115, 1980.
- [43] T. Floquet and J. P. Barbot, “An observability form for linear systems with unknown inputs,” *International Journal of control*, pp. 132–139, 2006.
- [44] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.
- [45] E. Lieb and W. Thirring, *Studies in Mathematical Physics, Essays in Honor of Valentine Bartmann*. Princeton, NJ, USA: Princeton University Press, 1976.
- [46] M. T. Chao and W. E. Strawderman, “Negative moments of positive random variables,” *Journal of the American Statistical Association*, vol. 67, no. 388, 1972.
- [47] G. Golub and A. van Loan, *Matrix Computations*. Johns Hopkins U. Press, 1996.
- [48] P. Biswas, T. chen Liang, T. chung Wang, and Y. Ye, “Semidefinite programming based algorithms for sensor network localization,” *ACM Transactions on Sensor Networks*, vol. 2, 2006.
- [49] D. Knuth, *The Art of Computer Programming, Volume 3: Sorting and Searching*. Addison-Wesley, 1997.

## Appendix

### A Distributed Estimation of $\mathbf{P}_{\varphi_{i|t}}$

In contrast to [7,13], we do not use the consensus-based algorithms to obtain the matrix  $\mathbf{P}_{\varphi_{i|t}}$ . We compute its estimate,  $\hat{\mathbf{P}}_{\varphi_{i|t}}$ , locally by solving a regularized problem as in [35]. Moreover, we need to extend the matrix estimation by taking message losses into account. Let  $j \in \mathcal{N}_i$  and assume that for  $t \in (t_1, t_2)$ ,  $j \notin \mathcal{N}_{\varphi_{i|t_1}}$ , and that  $j \in \mathcal{N}_{\varphi_{i|t_2}}$ . Then heuristically we set

$$\left[ \hat{\mathbf{P}}_{\varphi_{i|t_2}} \right]_{jj} = \max_k \left[ \hat{\mathbf{P}}_{\varphi_{i|t_1}} \right]_{kk}, \left[ \hat{\mathbf{P}}_{\varphi_{i|t_2}} \right]_{lj} = \left[ \hat{\mathbf{P}}_{\varphi_{i|t_1}} \right]_{jl} = 0,$$

for  $l \in \mathcal{N}_i$ , while the other entries are kept the same as the previous  $\hat{\mathbf{P}}_{\varphi_{i|t_2}}$ . It is motivated by that the maximum variance of the estimation error that a neighbor of a node is affected by must not be larger than the worst variance of the estimation error of other neighbors. Because all nodes are collaborating to build the state estimation by the same algorithm.