

Peer-to-Peer Estimator over Wireless Sensor Networks [★]

Yuzhe Xu, Carlo Fischione, ^{*} Alberto Speranzon ^{**}

^{*} *ACCESS Linnaeus Center, Electrical Engineering, Royal Institute of Technology, 100-44 Stockholm, Sweden
(e-mail: {yuzhe, carlofi}@kth.se).*

^{**} *United Technologies Research Center, East Hartford, CT, USA
(e-mail: alberto.speranzon@utrc.utc.com).*

Abstract: Distributed estimation of physical variables over wireless sensor networks is important to enable remote control of large scale plants, detection and monitoring of processes, etc. A peer-to-peer estimator, which is a particular implementation of the distributed estimation, performs local estimates in each node by exchanging information among neighbors without any central coordination. However, it is difficult to design a peer-to-peer minimum variance estimators, due to information losses over networks, lack of central coordination in the networks. In this paper we propose a peer-to-peer estimator for tracking a time varying scalar signal corrupted by a non-zero mean independent noise or disturbances. Information losses occurring over the network and the lack of central coordination are considered. Novel theoretical solutions are developed taking advantage of a model of the signal dynamics. The proposed approach simultaneously guarantees the mean estimation error to decrease over time as well as an estimate that is minimum variances. Simulations results are used to validate the performance of the proposed estimator.

Keywords: Distributed estimation, Lipschitz Optimization, Wireless Sensor Networks, Parallel and Distributed Computation, Vector Optimization

1. INTRODUCTION

Because of the growing interests in networked control, detection and monitoring system, wireless sensor networks (WSNs) play a key role providing communication and sensing capabilities for such systems. Distributed estimation of physical variables is a typical task of interest for WSN. Accurate estimates of these variables is a major need for many applications, spanning from traffic control, industrial automation, environment monitoring, to security systems Luo et al. (2006).

However, in general, efficient distributed computation over WSNs is difficult to design, due to the limited communication range and sensing capabilities of the nodes, packet losses, the lack of central coordination, and the measurement noise. Adaptive estimation algorithms must be designed to cope with these adverse conditions, while offering high accuracy. A particular instantiation of distributed estimators is a peer-to-peer estimator where local estimates and measurements are locally combined to produce an estimate. Such scheme can cope with the lack of central coordination naturally as each node takes advantage of data exchange only among neighbors. The challenge of a peer-to-peer estimation is that local processing must be carefully designed to avoid heavy computations and uncontrolled uncertainty propagation throughout the network.

This paper is a natural extension of the peer-to-peer estimator we proposed in Speranzon et al. (2008), which was designed under the assumption of perfect communication and limited knowledge of the environment. In this paper, a novel methodology is developed to design peer-to-peer estimators in the presence of the information losses (or packet losses) and the knowledge of the signal dynamics. The design of the estimator requires the solution of a novel Lipschitz optimization problem that provides us with conditions on adaptive estimator weights so that the expected estimation error converges asymptotically to a neighbor of the origin. We show how this problem has a unique global optimal solution, which can be computed numerically in a peer-to-peer fashion.

The proposed distributed adaptive estimator is designed to work on arbitrary networks and in presence of packet losses. Related work include Xiao et al. (2007); Xiao and Boyd (2004); Luo (2005) where however estimator weights are computed in a centralized way, Xiao et al. (2005); Xiao and Boyd (2004) and Russell et al. (2011) where each node requires the full knowledge of the Laplacian matrix associated to the communication graph. However, in the proposed estimator, nodes need to exchange information with other nodes located up to two-hop distance, as opposed to Cattivelli et al. (2008); Lopes and Sayed (2008), where communication is only with one-hop neighbors. The estimator we propose in this paper is related to recent contributions on low-pass filtering by diffusion or model-based mechanisms. For example, in Speranzon

[★] We acknowledge the support of the European Institute of Technology, Smart Energy Systems, the NoE Hycon2 and EU STREP Hydrobionets.

et al. (2006); Olfati-Saber and Shamma (2005); Hu et al. (2012); Cattivelli and Sayed (2010a,b), each node of the network obtains the average of the initial samples collected by nodes. In Shang et al. (2004); Luo (2005); Farina et al. (2010), distributed filtering using model-based approaches is studied in various wireless network contexts. In particular, distributed Kalman filters and more recently a combination of the diffusion mechanism with distributed Kalman filtering have been proposed, e.g., Kamgarpour and Tomlin (2008); Hu et al. (2012); Liang et al. (2010). Furthermore in Olfati-Saber and Jalalkamali (2012); Kar et al. (2012), a strategy where the estimator works at the same time as the communication update is studied. The distributed estimator that is proposed in this paper features better estimates when compared to similar distributed algorithms presented in the literature, but at the cost of an increased computational and communication complexity.

1.1 Notation

Given a stochastic variable x , $\mathbb{E}x$ denotes its expected value. With $\mathbb{E}_x f(x)$ we mean that the expected value of a function $f(\cdot)$ is taken with respect to the probability density function of the random variable x . We keep explicit the time dependence to remind the reader that the realization is given at time t . With $\|\cdot\|$ we denote the ℓ^2 -norm of a vector or the spectral norm of a matrix. Given a matrix \mathbf{A} , $\ell_m(\mathbf{A})$ and $\ell_M(\mathbf{A})$ denote the minimum and maximum eigenvalue (with respect to the absolute value of their real part), respectively, and its largest singular value is denoted by $\gamma(\mathbf{A})$. Given the matrix \mathbf{B} , $\mathbf{A} \circ \mathbf{B}$ is the Hadamard (element-wise) product between \mathbf{A} and \mathbf{B} . With \mathbf{A}^\dagger we denote the Moore-Penrose pseudo-inverse of the matrix \mathbf{A} , see Horn and Johnson (1985) for details. With $\mathbf{a} \leq \mathbf{b}$ and $\mathbf{a} \succeq \mathbf{b}$ denote the element-wise inequalities. With \mathbf{I} and $\mathbf{1}$ we denote the identity matrix and the vector $(1, \dots, 1)^T$, respectively, whose dimensions are clear from the context. Let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. To keep the notation lighter, the time dependence of the variables and parameters is not explicitly indicated, when this does not create misunderstandings.

2. PROBLEM FORMULATION

Consider a WSN with $N > 1$ sensor nodes placed at random and static positions in space. At every time instant, each sensor in the network takes a noisy measurement $y_i(t)$ of a scalar signal $x(t)$ described by a partially known model. Formally we have that for $t \in \mathbb{N}_0$ and $i = 1, \dots, N$

$$x(t) = ax(t-1) + \delta(t-1), \quad (1a)$$

$$y_i(t) = c_i x(t) + v_i(t), \quad (1b)$$

where $v_i(t)$ is normal distributed with zero mean and variance $\sigma_{v_i}^2$ and $\mathbb{E}v_i(t)v_j(t) = 0$ for all $t \in \mathbb{N}_0$, $i \neq j$ and where $a \in \mathbb{R}$ is known. We assume that $\delta(t-1)$ models an unknown disturbance. We will provide in the following a more detailed description of what we assume known.

We model the network as a graph. In particular we consider a graph, $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \dots, N\}$ is the vertex set and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set. We assume that the vertex and the edge sets are fixed. The

set of neighbors of node $i \in \mathcal{V}$ plus node i is denoted as $\mathcal{N}_i = \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\} \cup \{i\}$. Namely \mathcal{N}_i is the set containing the neighbors that a node i can have, including itself.

Every node broadcasts data packets, so that these packets can be received by any other node in the communication range as described by the edge set. However, over an edge, packets may be dropped because of bad channel conditions or radio interference. Let $\phi_{ij}(t)$, with $i \neq j$, be a binary random variable associated to the packet losses from node i to j at time t Stüber (1996). For $i \neq j$, we assume that the random variables $\phi_{ij}(t)$ are independent with probability mass function:

$$\begin{aligned} \Pr(\phi_{ij}(t) = 1) &= p_{ij}, \\ \Pr(\phi_{ij}(t) = 0) &= q_{ij} = 1 - p_{ij}, \end{aligned}$$

where $p_{ij} \in [0, 1]$ denotes the successful packet reception probability. The packet reception probabilities are assumed to be independent among links, and independent from past packet losses. These assumptions are natural when the coherence time of the wireless channel is small with respect to the typical communication rate of packets over WSNs Stüber (1996).

We assume every node i computes an estimate $\hat{x}_i(t)$ of $x(t)$ by taking a linear combination of its own and of its neighbours' estimates and measurements. Define $\hat{\mathbf{x}}(t) = (\hat{x}_1(t), \dots, \hat{x}_N(t))^T$ and similarly $\mathbf{y}(t) = (y_1(t), \dots, y_N(t))^T$, then each node computes

$$\begin{aligned} \hat{x}_i(t) &= a(\mathbf{k}_i \boldsymbol{\varphi}_{i|t})^T(t) \hat{\mathbf{x}}(t-1) \\ &\quad + (\mathbf{h}_i \boldsymbol{\varphi}_{i|t})^T(t) (\mathbf{y}(t) - a\mathbf{C}\hat{\mathbf{x}}(t-1)), \quad (2) \end{aligned}$$

with $\hat{\mathbf{x}}(0) = \mathbf{y}(0)$, \mathbf{C} is a diagonal matrix whose diagonal is the vector $(c_1, \dots, c_N)^T$ and where

$$(\mathbf{k}_i \boldsymbol{\varphi}_{i|t})(t) = \mathbf{k}_i(t) \circ \boldsymbol{\varphi}_{i|t},$$

with $\mathbf{k}_i^T(t) \in \mathbb{R}^{N \times 1}$, in which the j -th element is the weight coefficient used by node i for information coming from node j at time t , and $\boldsymbol{\varphi}_{i|t} \in \mathbb{R}^{N \times 1}$ denotes the vector of the packet reception process realization of the process $\phi_i(t)$ at time t , as seen from node i with respect to all nodes of the network. Specifically, let the j th element of $\boldsymbol{\varphi}_{i|t}$, with $j \neq i$, be $\varphi_{ij|t}$. Notice that at a given time instant, the j -th component of $\boldsymbol{\varphi}_{i|t}$ is zero if no data packets are received from node j . Let $\mathcal{N}_{\boldsymbol{\varphi}_i} = \{j \in \mathcal{N}_i : \varphi_{ij|t} \neq 0\}$, namely such a set collects the nodes communicating with node i at time t . The number of nodes in the set is $|\mathcal{N}_{\boldsymbol{\varphi}_i}| = \boldsymbol{\varphi}_{i|t}^T \boldsymbol{\varphi}_{i|t}$. The vector $(\mathbf{h}_i \boldsymbol{\varphi}_{i|t})(t) \in \mathbb{R}^{N \times 1}$ and $\mathbf{h}_i^T(t) \in \mathbb{R}^{N \times 1}$ are constructed from the elements $h_{ij}(t)$, similarly to $(\mathbf{k}_i \boldsymbol{\varphi}_{i|t})(t)$.

3. PEER-TO-PEER MINIMUM VARIANCE ESTIMATOR

In this section we extend the weights derivation proposed in Speranzon et al. (2008) to the case of networks with packet losses.

3.1 Distributed variance minimum

Let $\mathbf{e}(t) = (e_1(t), \dots, e_N(t))^T$, with $e_i(t) = x(t) - \hat{x}_i(t)$, $i = 1, \dots, N$, be the vector of the estimation errors.

Assume that $(\mathbf{k}_i \boldsymbol{\varphi}_{i|t})^T \mathbf{1} = 1$, then For each node i , the error dynamics can be obtained by

$$e_i(t) = a(\mathbf{g}_i \boldsymbol{\varphi}_{i|t})^T(t) \mathbf{e}(t-1) + (\mathbf{g}_i \boldsymbol{\varphi}_{i|t})^T(t) \mathbf{1} \delta(t-1) - (\mathbf{h}_i \boldsymbol{\varphi}_{i|t})^T(t) \mathbf{v}(t),$$

with $\mathbf{g}_i(t) = \mathbf{k}_i(t) - \mathbf{C} \mathbf{h}_i(t)$, $\mathbf{v} = (v_1(t), \dots, v_N(t))^T$.

Define $\mathbf{G}(t)$ the matrix with i -th row given by the vector $\mathbf{g}_i(t)$, for $i = 1, \dots, N$. Let $\boldsymbol{\varphi}_{i|t}$ be the matrix where the i -th row is the vector $\boldsymbol{\varphi}_{i|t}^T$. The average estimation error of the estimator (2) is bounded throughout the network provided that a condition on the maximum singular value of the matrix $\mathbf{G}(t) \circ \boldsymbol{\varphi}_{i|t}$ holds:

Proposition 1. Assume that

- (i) $\gamma(\mathbf{G}(t) \circ \boldsymbol{\varphi}_{i|t}) \leq \gamma_{\max} < \min(1, 1/a)$ for all $t \in \mathbb{N}_0$ and for each and every packet loss realization $\boldsymbol{\varphi}_{i|t}$ of $\Phi(t)$.
- (ii) $\delta(t) = d(t) + w(t)$, where $|d(t)| < \Delta$ represents the disturbances and $w(t) \sim \mathcal{N}(0, \sigma_w^2)$ is Gaussian noise for all $t \in \mathbb{N}_0$.

Then the corresponding function of the estimation error, computed with respect to the measurement noise and packet losses, is

$$\lim_{t \rightarrow +\infty} \|\mathbb{E}_{\boldsymbol{\varphi}} \mathbb{E}_{\mathbf{w}, \mathbf{v}} \mathbf{e}(t)\| \leq \frac{\Delta \sqrt{N} \gamma_{\max}}{1 - \gamma_{\max}}. \quad (3)$$

Proof. A proof is given in Xu et al. (2012) online.

Remark 2. Notice that the expected error converges to a neighborhood of the origin exponentially fast, and more precisely with rate $\gamma_{\max} < \min(1, 1/a)$.

Remark 3. Notice that $\delta(t)$ could be sum of the independent Gaussian noise and the disturbance. Or it could represent noise including the model uncertainty.

The previous proposition is useful because gives us a constraint on the weights so that the estimation error is stable. Moreover, we compute the weights so that the estimation error variance is minimized under the stability constraint. Every node computes the weights by solving at each time step the following optimization problem

$$\min_{\substack{\mathbf{g}_i(t), \mathbf{h}_i(t) \\ \psi_i(t)}} \mathbf{g}_i^T(t) \Gamma(t-1) \mathbf{g}_i(t) + \mathbf{h}_i^T(t) \mathbf{Q}(t-1) \mathbf{h}_i(t) \quad (4a)$$

$$\text{s.t.} \quad \left((\mathbf{g}_i(t) + \mathbf{C} \mathbf{h}_i(t))^T \circ \boldsymbol{\varphi}_{i|t} \right) \mathbf{1} = 1 \quad (4b)$$

$$\|\mathbf{g}_i(t) \circ \boldsymbol{\varphi}_{i|t}\|^2 \leq \psi_i(t). \quad (4c)$$

where $\Gamma(t) = (a^2 \mathbf{P}(t) + \sigma_w^2 \mathbf{I}) \circ (\boldsymbol{\varphi}_{i|t} \boldsymbol{\varphi}_{i|t}^T)$ with $\mathbf{P}(t) = \mathbb{E}(\mathbf{e}(t) - \mathbb{E} \mathbf{e}(t))(\mathbf{e}(t) - \mathbb{E} \mathbf{e}(t))^T$, while $\mathbf{Q}(t) = \Sigma(t) \circ (\boldsymbol{\varphi}_{i|t} \boldsymbol{\varphi}_{i|t}^T)$ with Σ which is a diagonal matrix whose diagonal is vector $(\sigma_{v_1}^2, \dots, \sigma_{v_N}^2)^T$. If the σ_w^2 is unknown, let $\Gamma(t) = a^2 \mathbf{P}(t) \circ (\boldsymbol{\varphi}_{i|t} \boldsymbol{\varphi}_{i|t}^T)$.

In this optimization problem, the objective function is the average variance of the estimation error at node i . The first constraint is motivated by assumption $(\mathbf{k}_i \boldsymbol{\varphi}_{i|t})^T = 1$, whereas the last constraint is a consequence of Proposition 1 and the Proposition III.1 in Speranzon et al. (2008). Specifically, the last constraint in problem (4) guarantees that $\gamma(\mathbf{G}(t) \circ \boldsymbol{\varphi}_{i|t}) \leq \gamma_{\max}$, provided that there exists some positive scalars $\psi_i(t)$, $i = 1, \dots, N$, such that

$$S_i(\boldsymbol{\psi}(t)) = \psi_i(t) + \sqrt{\psi_i(t)} \sum_{j \in \Theta_{\boldsymbol{\varphi}_i}} \sqrt{\psi_j(t)} - \gamma_{\max} \leq 0, \quad (5)$$

where $\Theta_{\boldsymbol{\varphi}_i} = \{j \neq i : \mathcal{N}_{\boldsymbol{\varphi}_i} \cap \mathcal{N}_{\boldsymbol{\varphi}_j} \neq \emptyset\} \cup \{\mathcal{N}_{\boldsymbol{\varphi}_i}\}$, which is the collection of communicating nodes located at two hops distance from node i plus communicating neighbours of i at time t .

The optimal solution to problem (4) is obtained in two steps: first, $\psi_i(t)$ is assumed fixed and the problem is solved by applying Lagrange dual theory for the variables $\mathbf{g}_i(t)$ and $\mathbf{h}_i(t)$, thus achieving expressions of $\mathbf{g}_i(t)$ and $\mathbf{h}_i(t)$ as function of $\psi_i(t)$. Finally, these expressions are used in the cost function, which is then minimized in the valuable $\psi_i(t)$. Details follows in the sequel.

For the first step, given a covariance matrix $\mathbf{P}(t-1)$ and a realization $\boldsymbol{\varphi}_{i|t}$ of $\boldsymbol{\varphi}_i(t)$, the weights that solve the optimization problem (4) are

$$\mathbf{g}_i(t) = \frac{((\Gamma(t-1) + \lambda_i(t) \mathbf{I}) \circ \boldsymbol{\varphi}_{i|t} \boldsymbol{\varphi}_{i|t}^T)^\dagger \boldsymbol{\varphi}_{i|t}}{\boldsymbol{\varphi}_{i|t}^T ((\Gamma(t-1) + \lambda_i(t) \mathbf{I}) \circ \boldsymbol{\varphi}_{i|t} \boldsymbol{\varphi}_{i|t}^T)^\dagger + \mathbf{C} \mathbf{Q}^{-1} \mathbf{C}} \boldsymbol{\varphi}_{i|t}, \quad (6)$$

$$\mathbf{h}_i(t) = \frac{\mathbf{Q}^{-1} \mathbf{C} \boldsymbol{\varphi}_{i|t}}{\boldsymbol{\varphi}_{i|t}^T ((\Gamma(t-1) + \lambda_i(t) \mathbf{I}) \circ \boldsymbol{\varphi}_{i|t} \boldsymbol{\varphi}_{i|t}^T)^\dagger + \mathbf{C} \mathbf{Q}^{-1} \mathbf{C}} \boldsymbol{\varphi}_{i|t}, \quad (7)$$

$$\lambda_i(t) = \begin{cases} 0 & \text{if } \left[\boldsymbol{\varphi}_{i|t}^T \mathbf{g}_i^T(t) \mathbf{g}_i(t) \boldsymbol{\varphi}_{i|t} \right]_{\lambda_i(t)=0} \leq \psi_i(t) \\ \lambda_i^*(t) & \text{otherwise} \end{cases}. \quad (8)$$

Here $\lambda_i^*(t)$ is determined by equation

$$\left[\boldsymbol{\varphi}_{i|t}^T \mathbf{g}_i^T(t) \mathbf{g}_i(t) \boldsymbol{\varphi}_{i|t} \right]_{\lambda_i^*(t)} = \psi_i(t). \quad (9)$$

The value of $\lambda_i^*(t)$ is in the interval

$$\left[0, \max \left(0, \frac{1}{\mathbf{1}^T \mathbf{C} \mathbf{Q}^{-1} \mathbf{C} \mathbf{1}} \sqrt{\frac{1}{\psi_i(t)}} - a^2 \ell(\mathbf{P}(t-1)) \right) \right],$$

which can be computed by a simple bisection algorithm Speranzon et al. (2008). The optimal value of $\lambda_i(t)$ determined by Eq.(8) is based on the following Proposition 4.

Proposition 4. If $\Gamma(t-1)$ is positive definite and symmetric, then over $\lambda_i(t) \geq 0$

- (i) $f_1(\lambda_i(t)) = \mathbf{g}_i^T(t) \mathbf{g}_i(t)$ is monotonically decreasing;
- (ii) $f_0(\lambda_i(t)) = \mathbf{g}_i^T(t) \Gamma(t-1) \mathbf{g}_i(t) + \mathbf{h}_i^T(t) \mathbf{Q}(t-1) \mathbf{h}_i(t)$ is monotonically increasing.

Proof. A proof is given in Xu et al. (2012).

3.2 Error Covariance Matrix

The weights $\mathbf{g}_i(t)$ and $\mathbf{h}_i(t)$, whose expression are given in (6) and (7), depend on the thresholds $\psi_i(t)$, through the values $\lambda_i(t)$, and on the error covariance matrix $\mathbf{P}(t-1)$. We dedicate the rest of this subsection on how to compute locally the error covariance matrix $\mathbf{P}(t-1)$, whereas we discuss in the next subsection how to compute the thresholds $\psi_i(t)$ in a distributed way.

Because of the packet loss process, node i requires only the elements of $\mathbf{P}(t-1)$ corresponding to the neighbors that have successfully communicated with node i , namely the matrix $\mathbf{P}(t-1) \circ \boldsymbol{\varphi}_{i|t} \boldsymbol{\varphi}_{i|t}^T$. In case of perfect communication, each node could estimate efficiently the error covariance

matrix from data. In particular, let $\hat{\mathbf{P}}_i(t-1)$ the estimation of the covariance matrix computed by node i , then

$$\hat{\mathbf{P}}_i(t-1) = \frac{1}{t} \sum_{\tau=0}^{t-1} (\hat{\mathbf{e}}_i(\tau) - \hat{\mathbf{m}}_i(\tau)) (\hat{\mathbf{e}}_i(\tau) - \hat{\mathbf{m}}_i(\tau))^T, \quad (10)$$

where

$$\hat{\mathbf{m}}_i(t) = \frac{1}{t} \sum_{\tau=0}^t \hat{\mathbf{e}}_i(\tau),$$

is the sample mean. The vector $\hat{\mathbf{e}}_i(t)$ is the vector of the estimation errors of the neighboring nodes available at node i , which is obtained by a Tichonov regularization approach, as discussed in Speranzon et al. (2008) with a different matrix \mathbf{A} given by

$$\mathbf{A} = \begin{pmatrix} \mathbf{1} & \mathbf{I} \\ \mathbf{C}\mathbf{1} & \mathbf{0} \end{pmatrix}.$$

However, here we need to extend the covariance matrix estimation to the case of packet losses.

When a node j exchanges data with its neighboring node i , after an outage period, node i needs to re-initialize reasonably the j -th row and column of $\hat{\mathbf{P}}_i(t)$ in order to take advantage of the new acquired neighbor. We consider the following re-initialization of elements of the error covariance matrix $\hat{\mathbf{P}}_i(t)$. If at time t a new neighbor of a node is exchanging data, then the diagonal element of the estimate of the error covariance matrix at time $t-1$, corresponding to such a neighbor, is initialized to the maximum element in the diagonal of the error covariance matrix. More precisely, 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 we have

$$\left[\hat{\mathbf{P}}_i(t_2 - 1) \right]_{jj} := \max_k \left[\hat{\mathbf{P}}_i(t_1 - 1) \right]_{kk}$$

and for $\ell \in \mathcal{N}_i$,

$$\left[\hat{\mathbf{P}}_i(t_2 - 1) \right]_{\ell j} := \left[\hat{\mathbf{P}}_i(t_1 - 1) \right]_{j\ell} := 0.$$

This heuristic is motivated by that all nodes are collaborating to build an estimate of $x(t)$, and they are using the same algorithm. Thus 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. Obviously, chances are that the heuristic might overestimate the variance associated to a new neighbor. However, from simulations in Section 5.2 we see that this strategy works well in practice, even in the presence of high packet loss probabilities.

3.3 Computation of the thresholds

We have now to compute the values of $\psi_i(t)$ that solve the optimization problem (4). By substituting (6) and (7) in the cost function of (4), we see that the larger is $\psi_i(t)$, the lower is the cost function. In other words, the larger is $\psi_i(t)$, the lower is the estimation error variance at node i . Since $\psi_i(t)$ must be maximized for $i = 1, \dots, N$, it follows that $\psi_i(t)$, $i = 1, \dots, N$, is given by the solution to the following multi-criterion optimization problem

$$\max_{\boldsymbol{\psi}(t)} \quad \boldsymbol{\psi}(t) \quad (11)$$

$$\text{s.t.} \quad \mathbf{S}(\boldsymbol{\psi}(t)) \preceq 0 \quad (12)$$

$$\boldsymbol{\psi}(t) \succ 0,$$

where $\mathbf{S}(\boldsymbol{\psi}(t)) = (S_1(\boldsymbol{\psi}(t)), \dots, S_N(\boldsymbol{\psi}(t)))^T$. Notice that the cost function is a vector whose components are coupled by the constraints (12). Thus the problem is a multi-criterion optimization problem and each threshold $\psi_i(t)$, $i = 1, \dots, N$, must be optimized simultaneously. This problem could be solved as a Fast-Lipschitz optimization problem Fischione (2011). The solution is given in Speranzon et al. (2008).

4. PERFORMANCE ANALYSIS

The variance of the estimation error is obtained by averaging over the distribution of the measurement noise.

Proposition 5. For any packet loss realization $\varphi_{i|t}$ of $\phi_i(t)$, the optimal value of $\mathbf{k}_i(t)$ and $\mathbf{h}_i(t)$ are such that the error variance at node i satisfies

(i)

$$\begin{aligned} \mathbb{E}_{\mathbf{v}}[(e_i^2 - \mathbb{E} e_i^2)^2 | \phi_i(t) = \varphi_{i|t}] \\ \geq \frac{1}{\boldsymbol{\varphi}_{i|t}^T (\Gamma^\dagger(t-1) + \mathbf{C}\mathbf{Q}^{-1}\mathbf{C}) \boldsymbol{\varphi}_{i|t}}, \end{aligned}$$

(ii)

$$\mathbb{E}_{\mathbf{v}}[(e_i^2 - \mathbb{E} e_i^2)^2 | \phi_i(t) = \varphi_{i|t}] < \frac{1}{\boldsymbol{\varphi}_{i|t}^T \mathbf{C}\mathbf{Q}^{-1}\mathbf{C} \boldsymbol{\varphi}_{i|t}}.$$

Proof. A proof is given in Xu et al. (2012).

5. IMPLEMENTATION AND NUMERICAL RESULTS

5.1 Implementation

In this section we summarize the main steps that a node has to implement according our estimation algorithm:

- (1) The first step of the algorithm is the determination of the optimal thresholds ψ_i^* by using the function `compute_threshold`.
- (2) Once the thresholds have been computed, each node determines the ID of the current communicating nodes. After that, the estimate of the covariance matrix $\hat{\mathbf{P}}$ and the estimate of the mean $\hat{\mathbf{m}}$ are adjusted depending on the communicating nodes.
- (3) The optimal Lagrangian multiplier $\lambda_i^*(t)$ is computed by using the bisection algorithm. Notice that the `bisection` function takes the search interval, the maximum error ϵ and the maximum number of iterations `Max_iter`.
- (4) Then the optimal weights are computed, and the estimate of the signal, $x_i(t)$, is computed. The updated of the covariance matrix and mean vector from estimates and measurements, as described in more detail in Speranzon et al. (2008).

5.2 Numerical Results

Numerical simulations have been carried out to compare the estimator proposed in this paper with some related estimators available from the literature. The section is concluded with a study of the communication and computational complexity of our estimator. We consider the following five estimators:

E_1 : $\mathbf{K}(t) = \mathbf{0}$ and $\mathbf{H}(t) = [h_{ij}(t)]$ with $h_{ij} = 1/|\mathcal{N}_{\varphi_i}|$ if node i and j communicate, and $h_{ij} = 0$ otherwise.

Thus, the updated estimate is the average of the measurements.

E_2 : $\mathbf{K}(t) = [k_{ij}(t)]$, where $k_{ii}(t) = 1/2|\mathcal{N}_{\varphi_i}|$, $k_{ij} = 1/|\mathcal{N}_{\varphi_i}|$ if node i and j communicate, $k_{ij}(t) = 0$ otherwise, whereas $\mathbf{H}(t) = [h_{ij}(t)]$ with $h_{ii} = 1/2|\mathcal{N}_{\varphi_i}|$, and $h_{ij} = 0$ elsewhere. This is the average of the old estimates and node's single measurement.

E_3 : $\mathbf{K}(t) = \mathbf{H}(t)$ with $k_{ij}(t) = 1/2|\mathcal{N}_{\varphi_i}|$ if node i and j communicate, and $k_{ij}(t) = 0$ otherwise. The updated estimate is the average of the old estimates and all local new measurements.

E_4 : $\mathbf{K}(t) = \mathbf{0}$ and $\mathbf{H}(t)$ is calculated by solving optimal problem

$$\begin{aligned} \min_{\mathbf{h}_i(t)} \quad & \mathbf{h}_i^T(t) \mathbf{Q}(t-1) \mathbf{h}_i(t) \\ \text{s.t.} \quad & ((\mathbf{g}_i(t) + \mathbf{C}\mathbf{h}_i(t))^T \circ \varphi_{i|t}) \mathbf{1} = 1. \end{aligned}$$

This is another variance minimum estimator which only uses node's measurements, where the optimal $\mathbf{h}_i^*(t) = \mathbf{Q}^{-1} \mathbf{C} \varphi_{i|t} / (\varphi_{i|t}^T \mathbf{C} \mathbf{Q} (t-1)^{-1} \mathbf{C} \varphi_{i|t})$.

E_p : The estimator proposed in this paper.

5.3 Estimation Accuracy

We take the mean square error (MSE) of the estimates of each node as performance measure. Further, based on MSE we define the relative average mean square error of the estimator E_i , $i = 1, \dots, 4$, as

$$\text{RMSE}_i = \frac{\text{MSE}(E_i) - \text{MSE}(E_p)}{\text{MSE}(E_p)}$$

where $\text{MSE}(E_i)$ is the average mean square error associated to the estimator E_i .

Figure 2 shows the RMSE for all the estimators as a function of the packet loss probability. Notice that E_p , represented as the line at 0, outperforms all other estimators for any considered packet loss probability. The two estimators E_3 and E_4 have performance closer to E_p . This is motivated by that the estimator E_3 takes the average values of all measurements and last step estimations, while E_4 is the optimally weighted the measurements minimizing the variance.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we characterized performance of a distributed peer-to-peer estimation algorithm for tracking a time-varying signal using a wireless sensor network with packet losses. A mathematical framework was proposed to analyze the filter, which run locally in each node of the network. The analysis showed that the filter is stable, and the variance of the estimation error is bounded. Numerical results illustrated the validity of our approach, which outperforms other estimators available from the literature.

Future studies will be devoted to the extension of our design methodology to the case, in which lossy communication links with memory will also be included.

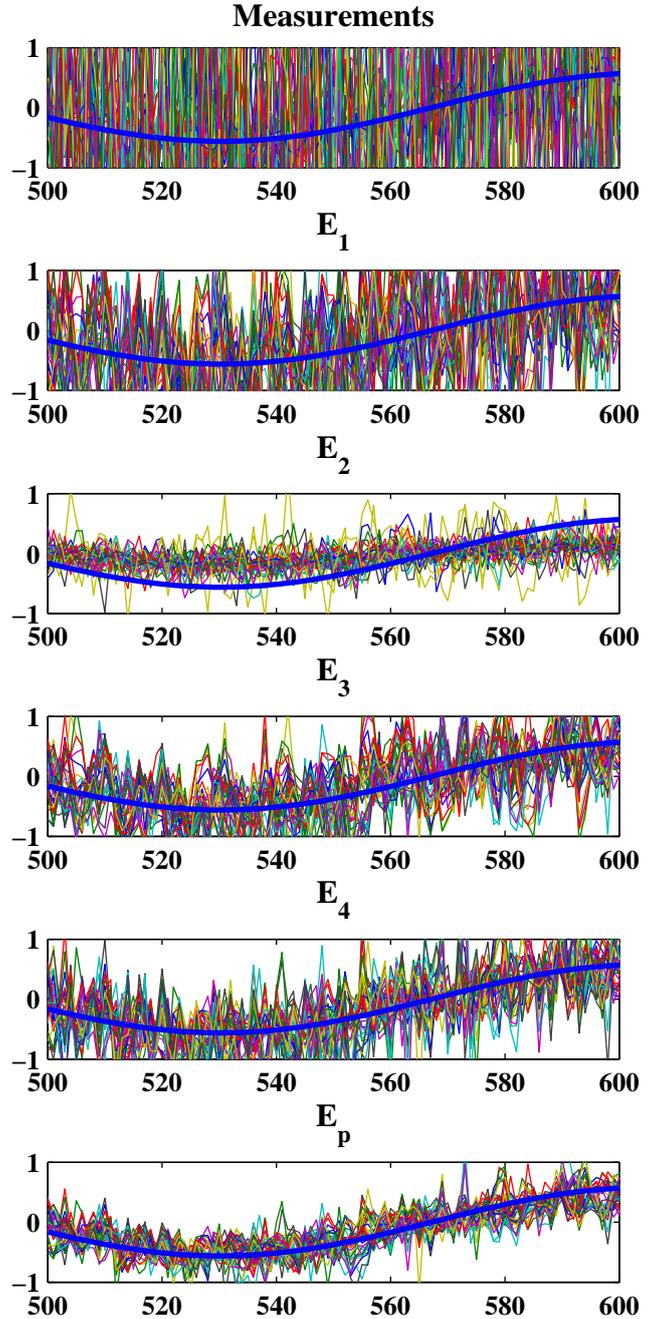


Fig. 1. Realizations of the signal to be tracked, as it is measured by all the $N = 35$ nodes, with a packet loss probability $q = 20\% \pm 5\%$ and $t \in [500, 600]$. Notice that the proposed estimator E_p , visibly outperforms all the other estimators in term of variance. It is possible to show that the proposed estimator presents a small bias when the signal changes more rapidly.

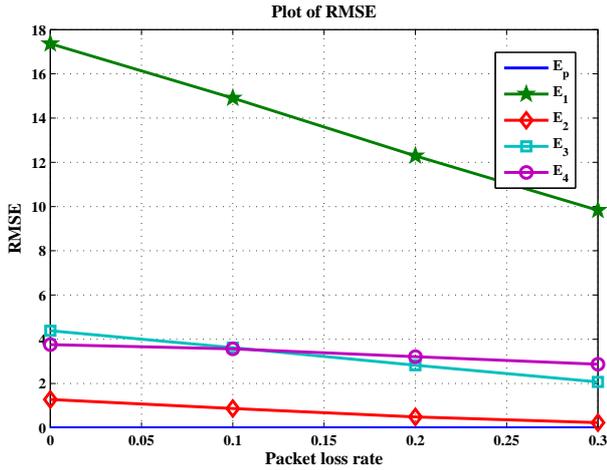


Fig. 2. Relative Mean Square Error (RMSE) performance comparison among estimators for various packet loss probabilities q . The marker (\star) refers to E_1 , (\diamond) refers to E_2 , (\square) refers to E_3 , (\circ) refers to E_4 . The proposed estimator E_p has RMSE = 0. The plots are generated from 100 Monte Carlo simulations. As the variance of the simulation results is very small, we do not show it in the plot. Notice that the proposed estimator E_p has the lowest variance, with a performance improvement going from a minimum of 20% up to 1800%.

ACKNOWLEDGEMENTS

We acknowledge the support of the European Institute of Technology, Smart Energy Systems, the NoE Hycon2 and EU STREP Hydrobionets.

REFERENCES

- Cattivelli, F.S., Lopes, C.G., and Sayed, A.H. (2008). Diffusion recursive least-squares for distributed estimation over adaptive networks. *IEEE Transactions on Signal Processing*, 56, 1865 – 1877.
- Cattivelli, F. and Sayed, A. (2010a). Diffusion lms strategies for distributed estimation. *Signal Processing, IEEE Transactions on*, 58(3), 1035 – 1048. doi: 10.1109/TSP.2009.2033729.
- Cattivelli, F. and Sayed, A. (2010b). Diffusion strategies for distributed kalman filtering and smoothing. *Automatic Control, IEEE Transactions on*, 55(9), 2069 – 2084. doi:10.1109/TAC.2010.2042987.
- Farina, M., Ferrari-Trecate, G., and Scattolini, R. (2010). Distributed moving horizon estimation for linear constrained systems. *Automatic Control, IEEE Transactions on*, 55(11), 2462 – 2475. doi: 10.1109/TAC.2010.2046058.
- Fischione, C. (2011). Fast-lipschitz optimization with wireless sensor networks applications. *Automatic Control, IEEE Transactions on*, 56(10), 2319 – 2331. doi: 10.1109/TAC.2011.2163855.
- Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis*. Cambridge University Press.
- Hu, J., Xie, L., and Zhang, C. (2012). Diffusion kalman filtering based on covariance intersection. *Signal Processing, IEEE Transactions on*, 60(2), 891 – 902. doi: 10.1109/TSP.2011.2175386.
- Kamgarpour, M. and Tomlin, C. (2008). Convergence properties of a decentralized Kalman filter. In *IEEE Conference on Decision and Control*.
- Kar, S., Moura, J., and Ramanan, K. (2012). Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication. *Information Theory, IEEE Transactions on*, PP, 1 – 1.
- Liang, Y., Chen, T., and Pan, Q. (2010). Optimal linear state estimator with multiple packet dropouts. *Automatic Control, IEEE Transactions on*, 55(6), 1428 – 1433. doi:10.1109/TAC.2010.2044263.
- Lopes, C.G. and Sayed, A.H. (2008). Diffusion least-mean squares over adaptive networks: Formulation and performance analysis. *IEEE Transactions on Signal Processing*, 56(7), 3122–3135.
- Luo, Z.Q. (2005). Universal decentralized estimation in a bandwidth-constrained sensor network. *IEEE Transactions on Information Theory*, 4, iv/829 – iv/832.
- Luo, Z.Q., Gatspar, M., Liu, J., and Swami, A. (eds.) (2006). *IEEE Signal Processing Magazine: Special Issue on Distributed Signal Processing in Sensor Networks*. IEEE.
- Olfati-Saber, R. and Jalalkamali, P. (2012). Coupled distributed estimation and control for mobile sensor networks. *Automatic Control, IEEE Transactions on*, PP, 1 – 1.
- Olfati-Saber, R. and Shamma, J.S. (2005). Consensus filters for sensor networks and distributed sensor fusion. In *Proceedings of IEEE Conference on Decision and Control*.
- Russell, W., Klein, D., and Hespanha, J. (2011). Optimal estimation on the graph cycle space. *Signal Processing, IEEE Transactions on*, 59(6), 2834 – 2846. doi: 10.1109/TSP.2011.2117422.
- Shang, Y., Ruml, W., Zhang, Y., and Fromherz, M. (2004). Localization from connectivity in sensor networks. *IEEE Transaction on Parallel and Distributed System, Vol. 15, No. 11*.
- Speranzon, A., Fischione, C., and Johansson, K.H. (2006). Distributed and collaborative estimation over wireless sensor networks. In *IEEE Conference on Decision and Control*.
- Speranzon, A., Fischione, C., Johansson, K.H., and Sangiovanni-Vincentelli, A. (2008). A distributed minimum variance estimator for sensor networks. *IEEE Journal on Selected Areas in Communications, Special Issue on Control and Communication*, 26, 609–621.
- Stüber, G.L. (1996). *Principles of Mobile Communication*. Kluwer Academic Publishers.
- Xiao, L. and Boyd, S. (2004). Fast linear iterations for distributed averaging. *System Control Letters*, 53, 65–78.
- Xiao, L., Boyd, S., and Kim, S.J. (2007). Distributed average consensus with least-mean-square deviation. *Journal of Parallel and Distributed Computing*, 67, 33–46.
- Xiao, L., Boyd, S., and Lall, S. (2005). A scheme for robust distributed sensor fusion based on average consensus. In *Proceedings of IEEE IPSN*.
- Xu, Y., Fischione, C., and Speranzon, A. (2012). Peer-to-peer estimation over wireless sensor network.