

# ANYTIME OPTIMAL DISTRIBUTED KALMAN FILTERING AND SMOOTHING

Ioannis D. Schizas, Georgios B. Giannakis, Stergios I. Roumeliotis and Alejandro Ribeiro

University of Minnesota, 200 Union Str. SE, Minneapolis, MN 55455, USA

## ABSTRACT

Distributed algorithms are derived for estimation and smoothing of nonstationary dynamical processes based on correlated observations collected by ad hoc wireless sensor networks (WSNs). Specifically, distributed Kalman filtering (KF) and smoothing schemes are constructed for *any-time minimum mean-square error* (MMSE) optimal consensus-based state estimation using WSNs. The novel distributed filtering/smoothing approach is flexible to trade-off estimation delay for MSE reduction, while it exhibits robustness in the presence of communication noise. Numerical examples demonstrate the merits of the proposed approach with respect to existing alternatives.

**Index Terms**— Distributed estimation and tracking, Kalman filtering

## 1. INTRODUCTION

A popular application of WSNs is decentralized tracking of nonstationary dynamic signals using discrete-time samples collected across sensors. Different from WSNs that rely on a fusion center, ad hoc WSNs are robust against fusion center failures and require single hop communications. In this context, consensus-based suboptimum Kalman filtering schemes were developed by [1, 4, 7] for estimation of dynamical state-space processes. These schemes are well motivated for distributed tracking applications but allow only for relatively slow-varying state processes, are inconsistent with the underlying observation model and/or, since they are based on variants of the consensus averaging algorithm of [8], they inherit its noise sensitivity when inter-sensor links are non-ideal.

In this paper, we consider distributed filtering and smoothing of nonstationary random processes. We take advantage of the inherent delay present in the consensus phase of the existing distributed KF approaches in order to build a distributed Kalman smoother (KS) that trades-off estimation delay for estimation quality (MSE). We utilize the alternating-direction method of multipliers [3, 5] to obtain ‘consensus-enriched’ observations across sensors and develop judicious local KS recursions enabling each sensor to form any-time MMSE optimal filtered and smoothed state estimates. Further, the distributed smoother is shown resilient to communication noise [5].

After delineating the problem setup in Section 2, we consider and motivate the problem of smoothing in Section 3.1. Then, we reformulate the centralized KS, and utilize the alternating-direction method of multipliers (Section 3.2) to derive the distributed KS recursions across sensors (Section 3.3). Motivating numerical examples are provided in Section 4 to corroborate our theoretical findings.

Prepared through collaborative participation in the Communication and Networks Consortium sponsored by the U. S. Army Research Lab under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

## 2. PROBLEM STATEMENT

Consider an ad hoc WSN comprising  $J$  sensors, where only single-hop communications are allowed; i.e., the  $j$ -th sensor communicates solely with nodes  $j'$  in its neighborhood  $\mathcal{N}_j \subseteq [1, J]$ . Communication links are assumed symmetric, and the WSN is modeled as an undirected graph whose vertices are the sensors and its edges represent the available links; see Fig. 1. As in [1, 4, 7], the communication graph is assumed to be connected. The WSN is deployed to track a  $p \times 1$  generally nonstationary signal  $\mathbf{s}(t)$  based on sensor observations  $\{x_j(t)\}_{j=1}^J$ , where  $t$  denotes discrete time. The state process obeys the Gauss-Markov model

$$\mathbf{s}(t) = \Phi(t-1)\mathbf{s}(t-1) + \mathbf{w}(t-1) \quad (1)$$

where  $\mathbf{w}(t-1)$  is zero-mean uncorrelated (in time) Gaussian with covariance matrix  $\mathbf{Q}(t-1)$  and  $\mathbf{s}(-1)$  denotes the initial state which is also zero-mean Gaussian with covariance  $\mathbf{C}_{ss}(-1)$ . Each sensor, say the  $j$ -th, observes the time series

$$x_j(t) = \mathbf{h}_j^T(t)\mathbf{s}(t) + n_j(t), \quad j = 1, \dots, J \quad (2)$$

where  $n_j(t)$  is zero-mean white Gaussian with variance  $\sigma_{n_j}^2(t)$ . Matrices  $\Phi(t-1)$  and  $\mathbf{Q}(t-1)$ , as well as  $\sigma_{n_j}^2(t)$  and  $\mathbf{h}_j(t)$  are available at sensor  $j$  and can be acquired from the physics of the problem.

If  $\mathbf{x}(t) := [x_1(t) \dots x_J(t)]^T$  were available at a *central* location, the MMSE optimal estimator of  $\mathbf{s}(t)$  given  $\{\mathbf{x}(t')\}_{t'=0}^t$  is the conditional expectation  $E[\mathbf{s}(t)|\mathbf{x}(t), \dots, \mathbf{x}(0)] := \hat{\mathbf{s}}(t|t)$  which can be recursively obtained using the KF [2, pg. 177]. If one can afford a delay  $i > 0$  in estimating the state, a lower MSE can be attained by forming  $\hat{\mathbf{s}}(t-i|t) := E[\mathbf{s}(t-i)|\mathbf{x}(t), \dots, \mathbf{x}(0)]$  for  $i = 0, 1, \dots, K$  via a fixed-lag Kalman smoother (KS), see e.g., [2, Ch. 8], with  $K$  denoting the maximum lag.

Based on single-hop communications, we wish to derive, *distributed* MMSE optimal KS estimates  $\{\hat{\mathbf{s}}_j(t-i|t; t : t+k)\}_{i=0}^K$  using  $k+1$  local iterates (starting at  $t$  and ending at  $t+k$  as indicated by the arguments after the semicolon) so that: If sensor  $j$  knows  $\Phi(t-1)$ ,  $\mathbf{Q}(t-1)$ ,  $\mathbf{C}_{ss}(-1)$  as well as  $\mathbf{h}_j(t)$  and  $\sigma_{n_j}^2(t)$ , then  $\lim_{k \rightarrow \infty} \hat{\mathbf{s}}_j(t-i|t; t : t+k) = \hat{\mathbf{s}}(t-i|t)$  for  $i = 0, 1, \dots, K$  and  $j = 1, 2, \dots, J$ .

Relative to [1, 4, 7], the distributed KF and KS approaches developed here do not limit  $\mathbf{s}(t)$  to vary slowly, and enjoy well defined MSE optimality as well as resilience to non-ideal links.

## 3. OPTIMAL DISTRIBUTED KALMAN SMOOTHING

The information form of the correction step of the centralized KF for obtaining the filtered estimate  $\hat{\mathbf{s}}(t|t)$  and the covariance matrix

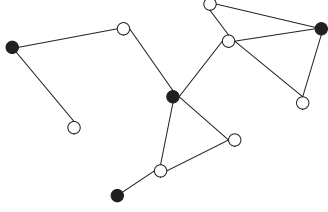


Fig. 1. An ad-hoc wireless sensor network.

$\mathbf{C}(t|t)$  of the estimation error  $\mathbf{s}(t) - \hat{\mathbf{s}}(t|t)$  is (see e.g., [2, pg. 139])

$$\hat{\mathbf{s}}(t|t-1) = \Phi(t-1)\hat{\mathbf{s}}(t-1|t-1) \quad (3)$$

$$\mathbf{C}(t|t-1) = \Phi(t-1)\mathbf{C}(t-1|t-1)\Phi^T(t-1) + \mathbf{Q}(t-1) \quad (4)$$

$$\mathbf{C}(t|t) = [\mathbf{H}^T(t)\mathbf{R}^{-1}(t)\mathbf{H}(t) + \mathbf{C}^{-1}(t|t-1)]^{-1} \quad (5)$$

$$\begin{aligned} \hat{\mathbf{s}}(t|t) &= \hat{\mathbf{s}}(t|t-1) + \mathbf{C}(t|t)\mathbf{H}^T(t)\mathbf{R}^{-1}(t) \\ &\quad \times [\mathbf{x}(t) - \mathbf{H}(t)\hat{\mathbf{s}}(t|t-1)] \end{aligned} \quad (6)$$

where  $\mathbf{C}(t|t-1)$  is the predicted covariance, while  $\mathbf{H}(t) := [\mathbf{h}_1(t) \dots \mathbf{h}_J(t)]^T$ . If sensors had available local estimates  $\hat{\mathbf{s}}_j(t-1|t-1)$  and the corresponding covariance  $\mathbf{C}_j(t-1|t-1)$ , they could run (3) and (4) in a distributed fashion since  $\Phi(t-1)$  and  $\mathbf{Q}(t-1)$  are assumed locally known. However, (5) and (6) can be run only if quantities

$$\mathcal{I}(t) := \mathbf{H}^T(t)\mathbf{R}^{-1}(t)\mathbf{H}(t) = \sum_{j=1}^J \sigma_{n_j}^{-2}(t)\mathbf{h}_j(t)\mathbf{h}_j^T(t) \quad (7)$$

$$\chi(t) := \mathbf{H}^T(t)\mathbf{R}^{-1}(t)\mathbf{x}(t) = \sum_{j=1}^J \sigma_{n_j}^{-2}(t)\mathbf{h}_j(t)x_j(t) \quad (8)$$

could be somehow estimated at each sensor  $j$ . This is possible because as the last equalities in (7) and (8) show,  $\mathcal{I}(t)$  and  $\chi(t)$  can be expressed as averages with the  $j$ -th summand available at sensor  $j$ .

Through  $K+1$  iterations that start at  $t$  and end at  $t+K$ , [1, 7] proposed (in our notation) to form estimates  $\hat{\mathcal{I}}_j(t; t:t+K)$  and  $\hat{\chi}_j(t; t:t+K)$  for  $\mathcal{I}(t)$  and  $\chi(t)$  respectively, using the consensus averaging based algorithms in [6] and [8], respectively; see also [4] where  $K=0$  was adopted. With these estimates plugged into recursions (5) and (6) it is possible to obtain local filtered estimates  $\hat{\mathbf{s}}_j(t|t; t:t+K)$ , that become available at  $t+K$ . Clearly, there is a delay  $K$  in forming these estimates limiting the operation of [1, 7] only to applications with slow-varying  $\mathbf{s}(t)$  and/or fast communications needed to complete  $K \gg 1$  consensus steps. In addition, [1, 7] inherit the noise sensitivity of [6, 8]. More important, the estimates  $\hat{\mathbf{s}}_j(t|t; t:t+K)$  in [1, 4, 7] are not MMSE optimal given the available information in  $\hat{\chi}_j(t; t:t+K)$ , unless  $K \rightarrow \infty$ . This suboptimality renders the distributed KF estimates in [1, 4, 7] inconsistent with the underlying data model.

### 3.1. Smoothing versus Filtering

Instead of filtering advocated by [1, 4, 7], the delay incurred by the  $K$  consensus averaging iterations needed to form  $\hat{\mathbf{s}}_j(t|t; t:t+K)$  prompts us to consider fixed-lag distributed Kalman smoothing. Specifically, our first idea is to seek at time instant  $t+K$ , *local MMSE optimal* smoothed estimates,  $\hat{\mathbf{s}}_j(t|t+i; t+i:t+K)$ , for  $i=1, \dots, K$ , that take advantage of all available data during the interval  $[t, t+K]$  and generally yield a lower MSE than the filtered estimates  $\hat{\mathbf{s}}_j(t|t; t:t+K)$ . Further, we wish to obtain zero delay ( $K=0$ ) filtered estimates, i.e.,  $\hat{\mathbf{s}}_j(t|t; t:t)$ , as well as *any-time MSE optimal* estimates  $\{\hat{\mathbf{s}}_j(t+i-j|t+i; t+i:t+K)\}_{j=0}^K$ , for  $i=0, 1, \dots, K$ , which are not available in the suboptimal alternatives [1, 4, 7].

To this end, we first express the fixed-lag Kalman smoother (KS) as a KF applied to a properly augmented state model. Consider the  $p(K+1) \times 1$  augmented state model [cf. (1)]

$$\begin{aligned} \check{\mathbf{s}}(t) &= \begin{bmatrix} \Phi(t-1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \dots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{I} & \mathbf{0} \end{bmatrix} \check{\mathbf{s}}(t-1) + \begin{bmatrix} \mathbf{w}(t-1) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \\ &:= \check{\Phi}(t-1)\check{\mathbf{s}}(t-1) + \check{\mathbf{w}}(t-1) \end{aligned} \quad (9)$$

where  $\check{\mathbf{s}}(t) := [\mathbf{s}^T(t) \dots \mathbf{s}^T(t-K)]^T$ ,  $\check{\mathbf{w}}(t) := [\mathbf{w}^T(t) \mathbf{0}^T \dots \mathbf{0}^T]^T$  and  $\check{\Phi}(t-1)$  is the matrix multiplying  $\check{\mathbf{s}}(t-1)$  after the first equality in (9). The aggregate observations  $\mathbf{x}(t) := [x_1(t) \dots x_J(t)]^T$  obey

$$\mathbf{x}(t) = \check{\mathbf{H}}(t)\check{\mathbf{s}}(t) + \mathbf{n}(t) \quad (10)$$

where  $\check{\mathbf{H}}(t) := [\mathbf{H}(t), \mathbf{0} \dots, \mathbf{0}]$  and  $\mathbf{n}(t) := [n_1(t) \dots n_J(t)]^T$  has covariance matrix  $\mathbf{R}(t) := \text{diag}(\sigma_{n_1}^2(t), \dots, \sigma_{n_J}^2(t))$ . Note that this state augmentation guarantees that the augmented noise  $\check{\mathbf{w}}(t-1)$  is uncorrelated across time. Then, the centralized KS corresponding to the augmented state model in (9) and the observation model (10) can be implemented via (3)-(6) after replacing  $\Phi(t-1)$  and  $\mathbf{H}(t)$  with  $\check{\Phi}(t-1)$  and  $\check{\mathbf{H}}(t)$ , respectively. The KS recursions are summarized as follows:

$$\hat{\check{\mathbf{s}}}(t|t-1) = \check{\Phi}(t-1)\hat{\check{\mathbf{s}}}(t-1|t-1) \quad (11)$$

$$\begin{aligned} \check{\mathbf{C}}(t|t-1) &= \check{\Phi}(t-1)\check{\mathbf{C}}(t-1|t-1)\check{\Phi}^T(t-1) \\ &\quad + \text{diag}(\mathbf{Q}(t-1), \mathbf{0}, \dots, \mathbf{0}) \end{aligned} \quad (12)$$

$$\check{\mathbf{C}}(t|t) = [\check{\mathbf{H}}^T(t)\mathbf{R}^{-1}(t)\check{\mathbf{H}}(t) + \check{\mathbf{C}}^{-1}(t|t-1)]^{-1} \quad (13)$$

$$\begin{aligned} \hat{\check{\mathbf{s}}}(t|t) &= \hat{\check{\mathbf{s}}}(t|t-1) + \check{\mathbf{C}}(t|t)\check{\mathbf{H}}^T(t)\mathbf{R}^{-1}(t) \\ &\quad \times [\mathbf{x}(t) - \check{\mathbf{H}}(t)\hat{\check{\mathbf{s}}}(t|t-1)] \end{aligned} \quad (14)$$

where  $\check{\mathbf{C}}(t|t)$  denotes the covariance matrix associated with the estimation error of the augmented state  $\check{\mathbf{s}}(t) - \hat{\check{\mathbf{s}}}(t|t)$ , while  $\check{\mathbf{C}}(t|t-1)$  denotes the covariance matrix for the prediction error. Note that the augmented state estimate  $\hat{\check{\mathbf{s}}}(t|t) := [\hat{\mathbf{s}}^T(t|t) \dots \hat{\mathbf{s}}^T(t-K|t)]^T$  obtained from KS, contains both a filtered estimate of the original state  $\mathbf{s}(t)$  as well as smoothed estimates of  $\mathbf{s}(t-k)$ , for  $k=1, \dots, K$ , using all the available data up to time  $t$ .

### 3.2. The Alternating-Direction Method of Multipliers

Next, we will derive a distributed estimation/smoothing algorithm that guarantees *any-time* MSE optimality under ideal links, while being robust in the presence of communication noise. As it becomes apparent from recursions (13)-(14), in order to implement a distributed KS scheme we will need local estimates of  $\check{\chi}(t) := \check{\mathbf{H}}^T(t)\mathbf{R}^{-1}(t)\mathbf{x}(t)$ ,  $\check{\mathcal{I}}(t) := \check{\mathbf{H}}^T(t)\mathbf{R}^{-1}(t)\check{\mathbf{H}}(t)$ . But since  $\check{\mathcal{I}}(t) := \text{diag}(\mathcal{I}(t), \mathbf{0}, \dots, \mathbf{0})$  and  $\check{\chi}(t) := [\chi^T(t), \mathbf{0}, \dots, \mathbf{0}]^T$ , it suffices to devise distributed estimators of  $\mathcal{I}(t)$  and  $\chi(t)$ . This is done after re-expressing  $\chi(t)$  as the optimal solution of the following minimization problem ( $\mathcal{I}(t)$  can be rewritten likewise)

$$\chi(t) := \arg \min_{\chi} \sum_{j=1}^J \|\chi - \mathbf{J}\mathbf{h}_j(t)\sigma_{n_j}^{-2}(t)x_j(t)\|_2^2 \quad (15)$$

where the term  $\mathbf{J}\mathbf{h}_j(t)\sigma_{n_j}^{-2}(t)x_j(t)$  is locally available at sensor  $j$ .

Our goal is to minimize in a distributed fashion the cost in (15), whose optimal solution yields the desired quantity  $\chi(t)$ . Toward

this end, since the summands in (15) are coupled through  $\chi$  we introduce the auxiliary variables  $\{\chi_j\}_{j=1}^J$  to represent the local estimates across sensors. Using the  $\chi_j$ 's we can re-write (15) as a *constrained* separable optimization problem

$$\begin{aligned} \{\chi_j(t)\}_{j=1}^J &:= \arg \min_{\chi_j} \sum_{j=1}^J \|\chi_j - J\mathbf{h}_j(t)\sigma_{n_j}^{-2}(t)x_j(t)\|_2^2 \\ \text{s. to } \chi_j &= \bar{\chi}_b, \quad b \in \mathcal{B}, j \in \mathcal{N}_b \end{aligned} \quad (16)$$

where  $\mathcal{B} \subseteq [1, J]$  is the bridge sensor subset introduced in [5] which is defined by the following pair of conditions: (i)  $\forall j \in [1, J]$  there exists at least one  $b \in \mathcal{B}$  such that  $b \in \mathcal{N}_j$  (the bridge neighbors of sensor  $j$  will be denoted by  $\mathcal{B}_j := \mathcal{N}_j \cap \mathcal{B}$ ); and, (ii)  $\forall b_1 \in \mathcal{B}$  there exists another sensor  $b_2 \in \mathcal{B}$  such that the shortest path between  $b_1$  and  $b_2$  has at most two edges. For the WSN in Fig. 1 the filled circles denote a possible selection for  $\mathcal{B}$ . Each sensor  $b \in \mathcal{B}$  maintains a local vector  $\bar{\chi}_b$  via which consensus among local variables  $\chi_j$  across all sensors is achieved a fortiori. Specifically, we have proved in [5] that the two conditions associated with  $\mathcal{B}$ , and combined with the connectivity of the WSN, provide a necessary and sufficient condition ensuring that the constraint set in (16) implies that  $\chi_j = \chi_{j'}, \forall j', j \in [1, J]$ . This guarantees the equivalence of (15) and (16) in the sense that  $\chi(t) = \chi_j(t), \forall j \in [1, J]$ . Next, consider the augmented Lagrangian of (16) given by

$$\begin{aligned} \mathcal{L}_a[\phi, \bar{\phi}, \mathbf{v}] &= \sum_{j=1}^J \|\chi_j - J\mathbf{h}_j(t)\sigma_{n_j}^{-2}(t)x_j(t)\|_2^2 \\ &+ \sum_{b \in \mathcal{B}} \sum_{j \in \mathcal{N}_b} (\mathbf{v}_j^b)^T (\chi_j - \bar{\chi}_b) + \sum_{b \in \mathcal{B}} \sum_{j \in \mathcal{N}_b} \frac{c_j}{2} \|\chi_j - \bar{\chi}_b\|_2^2 \end{aligned} \quad (17)$$

where  $\phi := \{\chi_j\}_{j=1}^J$ ,  $\bar{\phi} := \{\bar{\chi}_b\}_{b \in \mathcal{B}}$ ,  $\mathbf{v} := \{\mathbf{v}_j^b\}_{j \in [1, J], b \in \mathcal{B}_j}$  comprises the Lagrange multiplier vectors, and  $c_j > 0$  are penalty coefficients corresponding to the constraint  $\chi_j = \bar{\chi}_b, \forall b \in \mathcal{B}$ . Using the alternating-direction method of multipliers [3] to minimize (17), we can mimic the steps in [5, Appendix B] to prove the following proposition

**Proposition 1:** *For the minimization problem defined in (16) at time instant  $t$ , and per iteration index  $k$  consider iterates  $\mathbf{v}_j^b(t; t+k)$ ,  $\hat{\chi}_j(t; t+k)$  and  $\bar{\chi}_b(t; t+k)$  defined by the recursions*

$$\begin{aligned} \mathbf{v}_j^b(t; t+k) &= \mathbf{v}_j^b(t; t+k-1) + c_j [\hat{\chi}_j(t; t+k) \\ &\quad - \bar{\chi}_b(t; t+k)], \quad b \in \mathcal{B}_j \end{aligned} \quad (18)$$

$$\begin{aligned} \hat{\chi}_j(t; t+k+1) &= \hat{\chi}_j(t) - (2 + c_j|\mathcal{B}_j|)^{-1} \sum_{b \in \mathcal{B}_j} [\mathbf{v}_j^b(t; t+k) \\ &\quad - c_j \bar{\chi}_b(t; t+k)] \end{aligned} \quad (19)$$

$$\begin{aligned} \bar{\chi}_b(t; t+k+1) &= \sum_{j \in \mathcal{N}_b} \frac{1}{\sum_{\beta \in \mathcal{N}_b} c_\beta} \left[ \mathbf{v}_j^b(t; t+k) + \right. \\ &\quad \left. c_j \hat{\chi}_j(t; t+k+1) \right], \quad b \in \mathcal{B} \end{aligned} \quad (20)$$

where  $\hat{\chi}_j(t) := (2 + c_j|\mathcal{B}_j|)^{-1} 2J\mathbf{h}_j(t)\sigma_{n_j}^{-2}(t)x_j(t)$  and let the initial values  $\{\mathbf{v}_j^b(t; t-1)\}_{b \in \mathcal{B}_j}$ ,  $\{\hat{\chi}_j(t; t)\}_{j=1}^J$  and  $\{\bar{\chi}_b(t; t)\}_{b \in \mathcal{B}}$  be arbitrary. If sensor links are ideal and the WSN connected, then the local estimates  $\hat{\chi}_j(t; t+k+1)$  and  $\bar{\chi}_b(t; t+k+1)$  converge to  $\chi(t)$  as  $k \rightarrow \infty$ ; i.e.,  $\forall j \in [1, J]$  and  $\forall b \in \mathcal{B}$ ,

$$\lim_{k \rightarrow \infty} \hat{\chi}_j(t; t+k) = \lim_{k \rightarrow \infty} \bar{\chi}_b(t; t+k) = \chi(t).$$

Recall that  $t$  is the time index while  $k$  is the iteration index which

runs within the interval  $[0, K]$  and  $K < \infty$ . Through recursions (18)-(20) all the sensors  $j \in [1, J]$  keep track of the local estimate  $\hat{\chi}_j(t; t+k)$  and the Lagrange multipliers  $\mathbf{v}_j^b(t; t+k)$  for  $b \in \mathcal{B}_j$ . The sensors that also belong to subset  $\mathcal{B}$  keep also track of the consensus enforcing variables  $\bar{\chi}_b(t; t+k)$ . During time instant  $t+k$  sensor  $j$  receives the consensus variable  $\bar{\chi}_b(t; t+k)$  from its bridge neighbors within  $\mathcal{B}_j$ , and updates its Lagrange multipliers  $\{\mathbf{v}_j^b(t; t+k)\}_{b \in \mathcal{B}_j}$  using (18), which are used next to compute  $\hat{\chi}_j(t; t+k+1)$  through (19). After completing this iteration step, sensor  $j$  transmits to each of its bridge neighbors  $b \in \mathcal{B}_j$  the vector  $c_j^{-1}\mathbf{v}_j^b(t; t+k) + \hat{\chi}_j(t; t+k+1)$ . Subsequently, each sensor  $b \in \mathcal{B}$  receives the vectors  $c_j^{-1}\mathbf{v}_j^b(t; t+k) + \hat{\chi}_j(t; t+k+1)$  from all its neighbors  $j \in \mathcal{N}_b$  and proceeds to compute  $\bar{\chi}_b(t; t+k+1)$  using (20). This completes the  $k$ -th iteration and all the sensors in  $\mathcal{B}$  proceed to transmit  $\bar{\chi}_b(t; t+k+1)$  to all their neighbors  $j \in \mathcal{N}_b$  starting the  $(k+1)$ -st iteration.

Using exactly the same approach, a set of recursions similar to (18)-(20) can be obtained, through which sensor  $j$  forms local estimates  $\hat{\mathcal{I}}_j(t; t+k)$  that converge to  $\mathcal{I}(t)$  as  $k \rightarrow \infty$ . In the presence of additive noise, convergence of  $\hat{\mathcal{I}}_j(t; t+k)$  and  $\hat{\chi}_j(t; t+k)$  to  $\mathcal{I}(t)$  and  $\chi(t)$  is guaranteed in the mean sense, while the variance of the noise within  $\hat{\mathcal{I}}_j(t; t+k)$  and  $\hat{\chi}_j(t; t+k)$  converges to a bounded value ensuring their noise-resilience (due to space limitations we omit the details which can be found in [5]).

### 3.3. The Distributed Kalman Smoother (D-KS)

Following steps similar to those in [5, Section V], we can write the local recursions for  $\hat{\chi}_j(t; t+k)$  and  $\hat{\mathcal{I}}_j(t; t+k)$  for  $k = 1, \dots, K$ , in compact form as

$$\begin{aligned} \hat{\chi}_j(t; t+k) &= \mathbf{A}_j(k) \left[ \hat{\chi}^T(t; t+k-1), \right. \\ &\quad \left. \hat{\chi}^T(t; t+k-2) \right]^T \end{aligned} \quad (21)$$

$$\begin{aligned} \hat{\mathcal{I}}_j(t; t+k) &= \mathbf{A}_j(k) \left[ \hat{\mathcal{I}}^T(t; t+k-1), \right. \\ &\quad \left. \hat{\mathcal{I}}^T(t; t+k-2) \right]^T \end{aligned} \quad (22)$$

where  $\hat{\chi}(t; t+k) := [\hat{\chi}_1^T(t; t+k) \dots \hat{\chi}_J^T(t; t+k)]^T$  (similarly for  $\hat{\mathcal{I}}(t; t+k)$ ), while the  $p \times 2Jp$  matrix  $\mathbf{A}_j(k)$  contains  $p \times p$  coefficient submatrices that weigh appropriately only the information received by the bridge neighbors in  $\mathcal{B}_j$ , thus allowing for distributed implementation ( $\mathbf{A}_j(k)$  can be obtained from [5]). The local estimates in (21) and (22) are initialized as follows:

$$\begin{aligned} \hat{\chi}_j(t; t) &= 2J(2 + c_j|\mathcal{B}_j|)^{-1} \mathbf{h}_j(t)\sigma_{n_j}^{-2}(t)x_j(t), \\ \hat{\mathcal{I}}_j(t; t) &= 2J(2 + c_j|\mathcal{B}_j|)^{-1} \mathbf{h}_j(t)\sigma_{n_j}^{-2}(t)\mathbf{h}_j^T(t), \end{aligned}$$

where  $\hat{\chi}_j(t; t-1) = \mathbf{0}$  and  $\hat{\mathcal{I}}_j(t; t-1) = \mathbf{0}$ .

Later on, the estimate  $\hat{\chi}_j(t; t+k)$  available at sensor  $j$  and at time instant  $t+k$  with  $k \in [0, K]$ , will be utilized as an enriched information in the distributed (D-) KS algorithm. In order to do that, it is crucial to determine the data model for  $\hat{\chi}_j(t; t+k)$ ; i.e., how  $\hat{\chi}_j(t; t+k)$  is related with the state vector  $\mathbf{s}(t)$ . Interestingly, it turns out that  $\hat{\chi}_j(t; t+k)$  and  $\hat{\mathcal{I}}_j(t; t+k)$  are linearly related. Specifically, we have proved that (details in [5]):

**Lemma 1:** *Under ideal channel links,  $\hat{\chi}_j(t; t+k)$  and  $\hat{\mathcal{I}}_j(t; t+k)$*

$t+k$  are linearly related; i.e.,

$$\hat{\chi}_j(t; t:t+k) = \hat{\mathcal{I}}_j(t; t:t+k)\mathbf{s}(t) + \hat{\mathbf{n}}_j(t; t:t+k) \quad (23)$$

where  $\hat{\mathbf{n}}_j(t; t:t+k)$  is zero-mean Gaussian with covariance  $\mathcal{R}_j(t; t:t+k)$  defined by the recursion

$$\mathcal{R}_j(t; t:t+k) = \mathbf{A}_j(k)\mathcal{R}(t; t:t+k-1)\mathbf{A}_j^T(k) \quad (24)$$

where  $\mathcal{R}(t; t:t+k)$  is the  $2Jp \times 2Jp$  covariance matrix of the noise vector  $[\hat{\mathbf{n}}_1^T(t; t:t+k), \dots, \hat{\mathbf{n}}_J^T(t; t:t+k), \hat{\mathbf{n}}_1^T(t; t:t+k-1), \dots, \hat{\mathbf{n}}_J^T(t; t:t+k-1)]^T$ . The recursion in (24) is initialized by  $\mathcal{R}(t; t:t) = \text{diag}(\mathcal{R}_1(t; t:t), \dots, \mathcal{R}_J(t; t:t), \mathbf{0}_{Jp \times Jp})$ , where  $\mathcal{R}_j(t; t:t) = (2J)^2(2+c_j|\mathcal{B}_j|)^{-2}\sigma_{n_j}^{-2}(t)\mathbf{h}_j(t)\mathbf{h}_j^T(t)$ . Note that  $\mathcal{R}_j(t; t:t+k)$  is the  $j$ -th  $p \times p$  diagonal block of  $\mathcal{R}(t; t:t+k)$ .

In terms of the augmented state model, (23) can be expressed as

$$\hat{\chi}_j(t; t:t+k) = \hat{\mathcal{I}}_j(t; t:t+k)\hat{\mathbf{s}}(t) + \hat{\mathbf{n}}_j(t; t:t+k) \quad (25)$$

where  $\hat{\mathcal{I}}_j(t; t:t+k) := [\hat{\mathcal{I}}_j(t; t:t+k), \mathbf{0}_{p \times p}, \dots, \mathbf{0}_{p \times p}]$ . Also, note that the noise covariance  $\mathcal{R}_j(t; t:t+k)$  can be computed locally at sensor  $j$  using a set of recursions similar to (18)-(20) (details in [5, Section VI]).

Our key idea is to view  $\hat{\chi}_j(t; t:t+k)$  in (23) as a ‘consensus-enriched’ local observation vector per sensor  $j$ , and rely on it to derive any-time MSE optimal state estimates. Note that besides  $x_j(t)$ , quantity  $\hat{\chi}_j(t; t:t+k)$  includes ‘consensus data’ from neighboring sensors whose number increases as  $k$  increases. Since  $\hat{\chi}_j(t; t:t+k)$  contains more information than  $x_j(t)$ , state estimates based on it will clearly exhibit improved performance. Each sensor utilizes its ‘consensus-enriched’ model introduced in Lemma 1, as well as its parameters  $\hat{\mathcal{I}}_j(t; t:t+k)$  and  $\mathcal{R}_j(t; t:t+k)$  to implement local KS recursions corresponding to the observation model in (25) and the state model in (9). These local KS recursions as well as the distributed algorithms for forming the estimates  $\hat{\chi}_j(t; t:t+k)$  and  $\hat{\mathcal{I}}_j(t; t:t+k)$  constitute the D-KS algorithm whose properties are summarized next:

**Proposition 2:** Consider per sensor  $j$  the local augmented state estimates  $\hat{\mathbf{s}}_j(t+k|t+k; t+k:t+K)$ , with  $k=0, 1, \dots, K$ , obtained at time instant  $t+K$  through the KS recursions

$$\check{\mathbf{C}}_j(t+k|t+k-1; t+k-1:t+K) = \check{\Phi}(t+k-1) \quad (26)$$

$$\check{\mathbf{C}}_j(t+k-1|t+k-1; t+k-1:t+K) \times \check{\Phi}^T(t+k-1) + \text{diag}(\mathbf{Q}(t+k-1), \mathbf{0}, \dots, \mathbf{0})$$

$$\hat{\mathbf{s}}_j(t+k|t+k-1; t+k-1:t+K) = \check{\Phi}(t+k-1) \quad (27)$$

$$\times \hat{\mathbf{s}}_j(t+k-1|t+k-1; t+k-1:t+K) \\ \check{\mathbf{C}}_j(t+k|t+k; t+k:t+K) = \quad (28)$$

$$\left[ \hat{\mathcal{I}}_j^T(t+k; t+k:t+K)\mathcal{R}_j^{-1}(t+k; t+k:t+K) \right. \\ \left. \times \hat{\mathcal{I}}_j(t+k; t+k:t+K) \right. \\ \left. + \check{\mathbf{C}}_j^{-1}(t+k|t+k-1; t+k-1:t+K) \right]^{-1}$$

$$\hat{\mathbf{s}}_j(t+k|t+k; t+k:t+K) = \quad (29)$$

$$\hat{\mathbf{s}}_j(t+k|t+k-1; t+k-1:t+K)$$

$$+ \check{\mathbf{C}}_j(t+k|t+k; t+k:t+K)\hat{\mathcal{I}}^T(t+k; t+k:t+K)$$

$$\times \mathcal{R}_j^{-1}(t+k; t+k:t+K) \times [\hat{\chi}_j(t+k; t+k:t+K)$$

$$- \hat{\mathcal{I}}^T(t+k; t+k:t+K) \times \hat{\mathbf{s}}_j(t+k|t+k-1; t+k-1:t+K)]$$

corresponding to the augmented state model in (9) and the ‘enriched’ observation model in (25). Then, the local estimates  $\hat{\mathbf{s}}_j(t+k|t+k; t+k:t+K)$  are MMSE optimal in the sense that  $\hat{\mathbf{s}}_j(t+k|t+k; t+k:t+K) := E[\hat{\mathbf{s}}(t+k)|\{\hat{\chi}_j(t'; t': t'+K)\}_{t'=0}^{t+k}, \{\hat{\chi}_j(t'; t': t+K)\}_{t'=t+1}^{t+k}]$  which is the MMSE estimator of  $\hat{\mathbf{s}}(t+k)$  given the ‘enriched’ local observations available up to time  $t+K$ . Further, as the number of local iterates  $K \rightarrow \infty$  the local augmented state estimates converge to their centralized counterparts; i.e.  $\forall j \in [1, J], t=0, 1, \dots$ ,

$$\lim_{K \rightarrow \infty} \hat{\mathbf{s}}_j(t|t; t:t+K) = \hat{\mathbf{s}}(t|t), \quad \lim_{K \rightarrow \infty} \check{\mathbf{C}}_j(t|t; t:t+K) = \check{\mathbf{C}}(t|t).$$

**Proof:** MSE optimality of  $\hat{\mathbf{s}}_j(t+k|t+k; t+k:t+K)$  holds because these estimates are derived by local KS recursions that adhere to the state model in (9) and (25). Thus,  $\hat{\mathbf{s}}_j(t+k|t+k; t+k:t+K)$  is the MMSE estimator of  $\hat{\mathbf{s}}(t+k)$  at sensor  $j$  given the available data  $\{\hat{\chi}_j(t'; t': t+K)\}_{t'=0}^{t+k}$  at sensor  $j$  up to time  $t+K$ . Further, recall that under ideal links  $\lim_{K \rightarrow \infty} \hat{\mathcal{I}}_j(t; t:t+K) = \hat{\mathcal{I}}(t)$  and  $\lim_{K \rightarrow \infty} \hat{\chi}_j(t; t:t+K) = \chi(t)$  for  $t=0, 1, 2, \dots$ . Thus, as  $K \rightarrow \infty$  the local recursions of D-KS across all sensors coincide with the centralized KS recursions given in Section 3.  $\square$

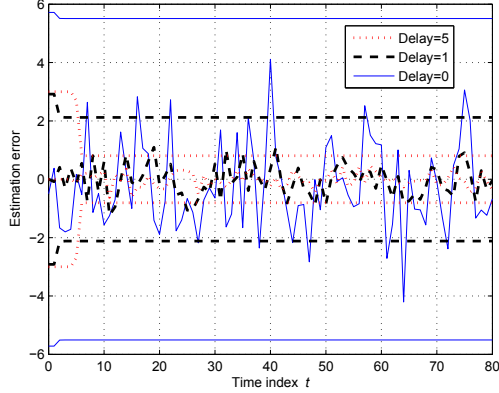
Notice that within the interval  $[t, t+K]$ , D-KS produces a sequence of local MSE optimal state estimates  $\hat{\mathbf{s}}(t|t; t:t+k)$ , for  $k=0, 1, \dots, K$ . Generally, the MSE associated with  $\hat{\mathbf{s}}(t|t; t:t+k)$ , decreases as  $k$  increases since  $\hat{\chi}_j(t; t:t+k)$  improves with  $k$  and provides a better estimate for  $\chi(t)$ . Depending on the delay that can be afforded, D-KS trades-off estimation accuracy (i.e., MSE) for estimation delay. Recall that the motivation for developing D-KS was to utilize the inherent time delay in computing filtered state estimates and form smoothed estimates. Fulfilling this objective, each sensor at time  $t+K$  provides the MSE optimal augmented state estimates  $\hat{\mathbf{s}}_j(t+k|t+k; t+k:t+K)$  with  $k=0, 1, \dots, K$ . Upon recalling that  $\hat{\mathbf{s}}_j(t+K|t+K; t+K:t+K) := [\hat{\mathbf{s}}_j^T(t+K|t+K; t+K:t+K), \dots, \hat{\mathbf{s}}_j^T(t|t+K; t+K:t+K)]$ , the augmented state filtered estimate formed by D-KS at  $t+K$  comprises both a filtered estimate of the original state  $\mathbf{s}(t+K)$  as well as the smoothed estimates  $\{\hat{\mathbf{s}}_j^T(t+k|t+K; t+K:t+K)\}_{k=0}^{K-1}$  for  $\{\mathbf{s}_j(t+k)\}_{k=0}^{K-1}$  respectively, exploiting all the  $\hat{\chi}_j$ ’s available up to time  $t+K$ .

Besides being able to form filtered and smoothed state estimates, D-KS exhibits provable noise robustness and trades-off delay for MSE reduction. The D-KS scheme is tabulated as Algorithm 1.

#### 4. PRELIMINARY SIMULATIONS

Here we test the D-KS algorithm in terms of estimation MSE, and compare it with [7] and [4], abbreviated here SD-KF and OD-KF respectively. We consider a WSN with  $J=60$  sensors. Nodes are randomly placed in the unit square  $[0, 1] \times [0, 1]$  with uniform distribution. We further assume a scalar state process, with  $\Phi(k) = 1$ ,  $Q = 4$ , and initial conditions  $E[\mathbf{s}(-1)] = 0$  and  $\sigma_{ss}^2(-1) = 1$ . Sensor  $j$  acquires at time  $k$  the scalar  $x_j(k)$  for which  $h_j(k) = h_j$  is normally distributed, while  $\sigma_{n_j}^2 = 1.5, \forall j$ . The number of consensus iterations used to estimate  $\mathcal{I}(t)$  and  $\chi(t)$  is  $K=6$ . The penalty coefficients are set to  $c_j = 4/|\mathcal{B}_j|$ .

An interesting property of D-KS is its ability to trade-off time delay for MSE reduction. Specifically, depending on application-dependent delay constraints, all sensors  $j$  at time  $t$  can utilize any of the smoothed estimates  $\hat{s}_j(t-k|t; t:t)$  for  $k=0, \dots, K$ . This is to be contrasted with SD-KF that only provides an estimate for  $s(t-K)$



**Fig. 2.** Estimation error and  $3\sigma$  curves vs. time index  $t$  for variable delays at sensor  $j = 2$ .

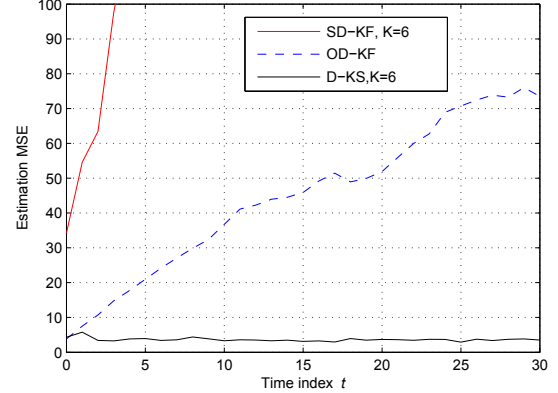
after  $K$  iterations, without using the observations over the interval  $[t - K, t]$ . Fig. 2 depicts the estimation error and corresponding  $3\sigma$  bounds at sensor  $j = 2$ , when estimating at time  $t$  the state  $s(t - k)$  via  $\hat{s}_2(t - k|t; t : t)$ , for  $k = 0, 1$  and  $5$ . Note that longer delays lead to lower MSEs. Further, the estimation error falls within the  $\pm 3\sqrt{C_2(t - k|t; t : t)}$  curves. This is reasonable since the local KS schemes are consistent with the underlying observation model of  $\hat{\chi}_j$ 's.

We now examine the MSE achieved by the filtered estimates provided by SD-KF, OD-KF and D-KS. Recall that SD-KF produces an estimate once every  $K$  consensus steps; thus, in order to have a fair comparison, at time  $t$ , state  $s(t)$  is estimated via  $\hat{s}_j(t - K|t - K; t - K : t)$  at sensor  $j$ . Fig. 3 depicts the MSE as a function of time when reception noise is present in the 'refined' observations  $\hat{\chi}_j(t - k; t - k : t)$  for  $k = 0, 1, \dots, K$ . It can be seen that D-KS tracks the state process through the local filtered estimates  $\hat{s}_j(t|t; t : t)$ , and the MSE reaches steady-state in the presence of noise. The MSE associated with the local estimates provided by SD-KF and OD-KF diverges. This is to be expected since both SD-KF and OD-KF are inconsistent with the true observation model, causing errors to accumulate for the fast varying  $s(t)$ .

## 5. CONCLUSIONS

In this paper we derived distributed estimators of nonstationary random signals. Utilizing the alternating-direction method of multipliers we enabled each sensor to form 'consensus-enriched' observations, which were used to improve considerably the tracking performance of the network. Different from existing suboptimal approaches, an MMSE optimal distributed Kalman smoother was developed that offers any-time optimal state estimates. The novel distributed smoother is flexible to trade-off estimation delay for MSE reduction, while it exhibits noise resilience.<sup>1</sup>

<sup>1</sup>The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies of the Army Research Laboratory or the U. S. Government.



**Fig. 3.** Empirical estimation MSE vs. time index  $t$  for D-KS, SD-KF and OD-KF in the presence of reception noise at sensor  $j = 2$ .

## Algorithm 1 D-KS Algorithm

---

For all  $j \in [1, J]$  initialize  $\check{s}_j(-1|-1) = \mathbf{0}$  and  $\check{C}_j(-1|-1) = \mathbf{C}_{ss}(-1)$ .  
 At time instant  $t + K$   
 For all  $j \in [1, J]$  and  $k = 0, \dots, K$  form estimates  $\hat{\chi}_j(t + k; t + k : t + K)$  and  $\hat{\mathcal{I}}_j(t + k; t + k : t + K)$  using (18)-(20).  
 For all  $j \in [1, J]$  and  $k = 0, \dots, K$  form augmented state estimates  $\check{s}_j(t + k|t + k; t + k : t + K)$  using (26)-(29).

---

## 6. REFERENCES

- [1] M. Alanyali and V. Saligrama, "Distributed Tracking in Multi-Hop Networks with Communication Delays and Packet Losses," in *IEEE Workshop on Stat. Sig. Proc.*, 2005, pp. 1190–1195.
- [2] B. D. Anderson and J. B. Moore, *Optimal Filtering*, Prentice Hall, 1979.
- [3] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena-Scientific, 1999.
- [4] R. Olfati-Saber, "Distributed Kalman Filter with Embedded Consensus Filters," in *Proc. of 44th IEEE Conf. on Dec., and the Eur. Contr. Conf.*, Seville, Spain, 2005, pp. 8179–8184.
- [5] 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*, revised July 2007.
- [6] D. P. Spanos, R. Olfati-Saber, and R. J. Murray, "Distributed Sensor Fusion Using Dynamic Consensus," in *Proc. of the 16th IFAC World Congress*, Prague, Czech, 2005.
- [7] D. P. Spanos, R. Olfati-Saber, and R. J. Murray, "Approximate Distributed Kalman Filtering in Sensor Networks with Quantifiable Performance," in *Proc. of the 4th Intl. Symp. on Info. Proc. in Sensor Networks*, Los Angeles, CA, 2005.
- [8] L. Xiao, S. Boyd, and S. Lall, "A scheme for Robust Distributed Sensor Fusion Based on Average Consensus," in *Proc. of the 4th Intl. Symposium on Info. Proc. in Sensor Networks*, Berkeley, CA, 2005, pp. 63–70.