Consensus in Ad Hoc WSNs With Noisy Links—Part II: Distributed Estimation and Smoothing of Random Signals

Ioannis D. Schizas, *Student Member, IEEE*, Georgios B. Giannakis, *Fellow, IEEE*, Stergios I. Roumeliotis, *Member, IEEE*, and Alejandro Ribeiro, *Student Member, IEEE*

***Abstract*—Distributed algorithms are developed for optimal estimation of stationary random signals and smoothing of (even nonstationary) dynamical processes based on generally correlated observations collected by ad hoc wireless sensor networks (WSNs). Maximum *a posteriori* (MAP) and linear minimum mean-square error (LMMSE) schemes, well appreciated for centralized estimation, are shown possible to reformulate for distributed operation through the iterative (alternating-direction) method of multipliers. Sensors communicate with single-hop neighbors their individual estimates as well as multipliers measuring how far local estimates are from consensus. When iterations reach consensus, the resultant distributed (D) MAP and LMMSE estimators converge to their centralized counterparts when inter-sensor communication links are ideal. The D-MAP estimators do not require the desired estimator to be expressible in closed form, the D-LMMSE ones are provably robust to communication or quantization noise and both are particularly simple to implement when the data model is linear-Gaussian. For decentralized tracking applications, distributed Kalman filtering and smoothing algorithms are derived for *any-time MMSE optimal* consensus-based state estimation using WSNs. Analysis and corroborating numerical examples demonstrate the merits of the novel distributed estimators.**

***Index Terms*—Distributed estimation, Kalman smoother, nonlinear optimization, wireless sensor networks (WSNs).**

I. Introduction

In a companion paper [17], we introduced optimal distributed estimators for *deterministic* signals based on *uncorrelated* observations collected by ad hoc wireless sensor networks (WSNs). Different from WSNs that rely on a fusion center, ad hoc WSNs are robust against fusion center failures and require only single-hop communications among neighboring sensors that aim to consent on local estimates formed over a (possibly large) geographical area. Compared to alternative consensus-based distributed algorithms [5], [22]–[24], the ones in [17] offer i) optimal distributed best linear unbiased estimators (D-BLUE) and distributed maximum likelihood estimators (D-MLE) based on the alternating-direction method of Lagrange multipliers; ii) guaranteed convergence to their centralized counterparts when inter-sensor links are ideal; and iii) provable resilience to communication and quantization noise which causes the estimation variance of [22], [23] to grow unbounded.

The present paper significantly broadens the scope of [17] to encompass distributed estimation of *random* stationary and nonstationary signals based on generally *correlated* sensor data using ad hoc WSNs. In this context, distributed WSN-based estimation of stationary Markov random field (MRF) models was pursued in [9]. Distributed LMMSE estimation for MRFs was considered in [8] when sensor observations obey a linear Gaussian model. As far as nonstationary signals are concerned, consensus-based suboptimum (in the MSE sense) Kalman filtering (KF) schemes were developed by [1], [13], [19] 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. Since these schemes are based on variants of the consensus averaging algorithm of [22], they inherit its noise sensitivity when intersensor links are nonideal. An alternative distributed KF scheme was developed in [2], where each sensor forms a weighted average of state estimates received from its neighbors, and incorporates them in the filtering process. However, the covariance information needed per sensor is global, i.e., requires information from all sensors in the network.

The contributions and organization of this paper are as follows: i) distributed maximum *a posteriori* (D-MAP) and linear minimum mean-square error (D-LMMSE) optimal algorithms are derived based on the alternating-direction method of Lagrange multipliers for estimating stationary random signals (Sections III and IV); ii) the D-MAP scheme does not require the corresponding centralized estimator to be available in a known closed form, and the D-LMMSE one requires only availability of second-order statistics of the signal and sensor data which are in both cases allowed to be correlated; iii) robustness to additive quantization and communication noise (Section V); and iv) any-time MMSE optimal distributed consensus-based Kalman filtering and smoothing algorithms capable of tracking

Manuscript received March 7, 2007; revised July 23, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Aleksandar Dogandzic. Work in this paper was supported by the USDoD ARO Grant No. W911NF-05-1-0283; and also through collaborative participation in the C&N Consortium sponsored by the U.S. ARL under the CTA 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. 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. Part of the paper was presented in the Fortieth Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, October 29–November 1, 2006; the 2007 IEEE Workshop on Signal Processing Advances Wireless Communication (SPAWC) Conference, Helsinki, Finland, June 17–20, 2007; and the 2007 SSP Workshop, Madison, WI, August 26–29, 2007.

I. D. Schizas, G. B. Giannakis, and A. Ribeiro are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: schizas@ece.umn.edu; georgios@ece.umn.edu; aribeiro@ece.umn.edu).

S. I. Roumeliotis is with the Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: stergios@cs.umn.edu).

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

Digital Object Identifier 10.1109/TSP.2007.908943

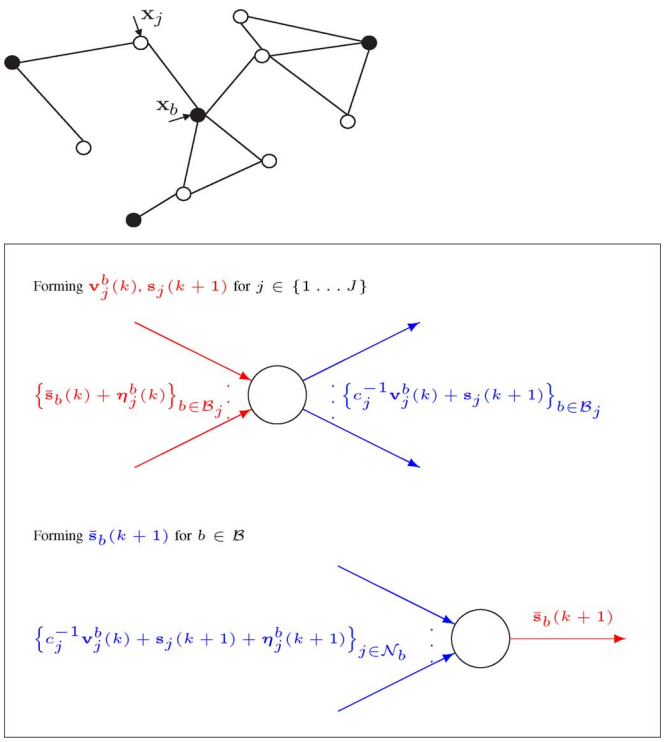

Fig. 1. (Top) Ad hoc wireless sensor network. (Bottom) Implementation of the D-MAP estimation algorithm.

even fast varying processes, trading off delay for MSE reduction and remaining resilient to noise (Section VI). We conclude the paper in Section VII.

II. Problem Formulation and Preliminaries

Consider an ad hoc WSN comprising J sensors, where only single-hop communications are allowed; i.e., the jth sensor communicates solely with nodes j' in its neighborhood $\mathcal{N}_j \subseteq [1, J]$, and $j \in \mathcal{N}_j$. Communication links are symmetric, and the WSN is modelled as an undirected graph whose vertices are the sensors and its edges represent the available links; see Fig. 1 (top). The connectivity information is summarized in the so called adjacency matrix $\mathbf{E} \in \mathbf{r}^{J \times J}$ for which $(\mathbf{E})_{jj'} = 1$ if $j' \in \mathcal{N}_j$, while $(\mathbf{E})_{jj'} = 0$ if $j' \notin \mathcal{N}_j$. Since $j' \in \mathcal{N}_j$ if and only if $j \in \mathcal{N}_{j'}$, it holds that $\mathbf{E} = \mathbf{E}^T$ (^{T} stands for transposition).

The WSN is deployed to i) either estimate a $p \times 1$ stationary signal vector $\mathbf{s}$ using sensor data $\{\mathbf{x}_j \in \mathbf{r}^{L_j \times 1}\}_{j=1}^J$, or ii) track a generally nonstationary $\mathbf{s}(t)$ based on sensor observations $\{\mathbf{x}_j(t) \in \mathbf{r}^{L_j \times 1}\}_{j=1}^J$, where t denotes discrete time. Without loss of generality (wlog) both $\mathbf{s}$, $\mathbf{s}(t)$ and $\{\mathbf{x}_j, \mathbf{x}_j(t)\}_{j=1}^J$ are assumed zero mean. If these vector signals have a nonzero mean and the nonzero mean is known, what follows applies to the corresponding mean-compensated signals. Further, we can compensate for the mean by adding it back to the obtained estimates. Starting with the stationary case i) and depending on the available *a priori* information sensors have about $\mathbf{s}$, we distinguish between two setups.

In the first one, the probability density function (pdf) $p(\mathbf{s})$ is assumed known to all sensors. Conditioned on $\mathbf{s}$, $\{\mathbf{x}_j\}_{j=1}^J$ are further assumed independent with pdfs $p_j(\mathbf{x}_j|\mathbf{s})$ known $\forall j = 1, \dots, J$. Under these assumptions, the MAP estimator (after using Bayes' rule and the natural logarithm $\ln$) can be written as

$$\hat{\mathbf{s}}_{\text{map}} := \arg \min_{\mathbf{s} \in \mathbb{R}^{p \times 1}} -\sum_{j=1}^{J} \left(\ln [p_j(\mathbf{x}_j|\mathbf{s})] + J^{-1} \ln [p(\mathbf{s})] \right). \quad (1)$$

The second scenario arises when the observations adhere to an arbitrary data model, but contrary to (1) only the first- and second-order statistics of $\mathbf{s}$ and $\mathbf{x}_j$ are known; i.e., sensors know only (cross-) covariance matrices $\mathbf{C}_{xx} := E[\mathbf{x}\mathbf{x}^T]$, $\mathbf{C}_{sx} := E[\mathbf{s}\mathbf{x}^T]$ and $\mathbf{C}_{ss} := E[\mathbf{s}\mathbf{s}^T]$, where $\mathbf{x} := [\mathbf{x}_1^T \dots \mathbf{x}_J^T]^T$ contains all the $L := \sum_{j=1}^J L_j$ sensor observations. Different from [5], [17], [24], where sensor data are assumed uncorrelated, $\mathbf{C}_{xx}$ does not have to be block diagonal. Sensor j has available $\mathbf{C}_{x_j} := [\mathbf{C}_{x_j x_1} \dots \mathbf{C}_{x_j x_J}]$, $\mathbf{C}_{sx_j}$, and $\mathbf{C}_{ss}$. These matrices can be acquired either from the physics of the problem, or, during a training phase. Notwithstanding, each sensor does not have to know the entire matrices $\mathbf{C}_{xx}$ and $\mathbf{C}_{sx}$ but only a part of them containing $1/J$ of the total covariance information. When $\mathbf{x}$, $\mathbf{C}_{xx}$, and $\mathbf{C}_{sx}$ are available at a central location, it is possible to form the LMMSE estimator as, see e.g., [12]

$$\hat{\mathbf{s}}_{\text{lmmse}} := \mathbf{C}_{sx}\mathbf{C}_{xx}^{-1}\mathbf{x}. \quad (2)$$

We will develop *distributed* iterative algorithms based on communications with one-hop neighbors that generate (local) time iterates $\mathbf{s}_j(k)$ so that:

- **s1)** if the jth sensor knows only $p_j(\mathbf{x}_j|\mathbf{s})$ and $p(\mathbf{s})$, the local iterates converge as $k \to \infty$ to the global (i.e., centralized) MAP estimator, i.e., $\lim_{k\to\infty} \mathbf{s}_j(k) = \hat{\mathbf{s}}_{\text{map}}$, where $\hat{\mathbf{s}}_{\text{map}}$ is given by (1);
- **s2)** if $\mathbf{C}_{x_j}$ and $\mathbf{C}_{sx_j}$ are known at the jth sensor and $\mathbf{C}_{xx}$ is full rank, then $\lim_{k\to\infty} \mathbf{s}_j(k) = \hat{\mathbf{s}}_{\text{lmmse}}$.

The MAP estimator in s1) is of particular importance for estimating stationary signals in generally nonlinear data models since it is optimal in the sense of minimizing the "hit-or-miss" Bayes risk (see, e.g., [12, p. 372]). The LMMSE estimator on the other hand is MSE optimal within the class of linear estimators; but its separate treatment is well motivated because it remains applicable even if the conditional independence required in (1) does not hold and only $\mathbf{C}_{sx_j}$ and $\mathbf{C}_{x_j x_j}$ are known per sensor j. Clearly, if $\{\mathbf{x}_j\}_{j=1}^J$ and $\mathbf{s}$ are *jointly Gaussian*, then $\hat{\mathbf{s}}_{\text{map}} = \hat{\mathbf{s}}_{\text{lmmse}}$, and consequently scenarios s1) and s2) coincide.

In a third (dynamic) scenario, all sensors wish to track a $p \times 1$ common generally nonstationary state process that obeys the Gauss–Markov model

$$\mathbf{s}(t) = \boldsymbol{\Phi}(t-1)\mathbf{s}(t-1) + \mathbf{w}(t-1) \quad (3)$$

where $\mathbf{w}(t-1)$ is zero-mean Gaussian with covariance $\mathbf{Q}(t-1)$ and $\mathbf{s}(-1)$ denotes the initial state which is also zero-mean Gaussian with covariance $\mathbf{C}_{ss}(-1)$. Matrices $\boldsymbol{\Phi}(t-1)$ and $\mathbf{Q}(t-1)$ are assumed known to all sensors. Each sensor, say the jth, observes the time series

$$x_j(t) = \mathbf{h}_j^T(t)\mathbf{s}(t) + n_j(t), \quad j = 1, \dots, J \quad (4)$$

where $n_j(t)$ is zero-mean Gaussian with variance $\sigma^2_{n_j}(t)$. Both $\sigma^2_{n_j}(t)$ and $\mathbf{h}_j(t)$ are available at sensor j.

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 Kalman filter (KF) [3]. 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., [3, Ch. 8]. Based again on single-hop communications, we will derive, *distributed* MSE 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 semi-colon) so that:

s3) If sensor j knows $\mathbf{\Phi}(t-1)$, $\mathbf{Q}(t-1)$, $\mathbf{C}_{ss}(-1)$ as well as $\mathbf{h}_j(t)$ and $\sigma^2_{n_j}(t)$, then $\lim_{k\to\infty} \hat{\mathbf{s}}_j(t-i|t; t: t+k) = \hat{\mathbf{s}}(t-i|t)$ for $i = 0, 1, \dots, K$ and $j = 1, 2, \dots, J$.

Relative to [13], [19] and [1], 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 noise resilience. If at iteration k sensor j transmits[1] to sensor j' the vector $\mathbf{t}_j^{j'}(k) \in \mathbf{r}^{p\times 1}$ the received noisy vector is $\mathbf{r}_{j'}^{j}(k) = \mathbf{t}_j^{j'}(k) + \boldsymbol{\eta}_{j'}^{j}(k)$, where $\boldsymbol{\eta}_{j'}^{j}(k)$ denotes zero-mean additive noise at sensor j'. Vector $\boldsymbol{\eta}_{j'}^{j}(k)$ is assumed uncorrelated across sensors and time, with $\mathbf{C}_{\eta_{j'}\eta_{j'}} := \mathrm{E}[\boldsymbol{\eta}_{j'}^{j}(k)(\boldsymbol{\eta}_{j'}^{j}(k))^T]$. This noise model encompasses: **n1)** analog communication where $\boldsymbol{\eta}_{j'}^{j}(k)$ is zero-mean white Gaussian and **n2)** digital communication with $\boldsymbol{\eta}_{j'}^{j}(k)$ uniformly distributed over $[-Q_j/2^{m_j}, Q_j/2^{m_j}]$, with $\mathbf{C}_{\eta_j\eta_j} = 2^{-2m_j}(Q_j^2/3)\mathbf{I}_p$, where m_j denotes the number of bits used for quantizing each of the entries in $\mathbf{t}_j^{j'}(k)$, and $\mathbf{I}_p$ is the $p \times p$ identity matrix. We further assume that:

a1) the communication graph is connected; i.e., there exists a path connecting any two sensors;

a2) the pdfs $\{p_j(\mathbf{x}_j|\mathbf{s})\}_{j=1}^{J}$ and $p(\mathbf{s})$ are (strictly) log-concave with respect to (wrt) $\mathbf{s}$, while their sum of logarithms in (1) is a strictly concave function.

As in [5], [13], [14], [17], [19], [22], and [24] network connectivity in a1) ensures that distributed estimators have access to all observation vectors, while log-concavity in a2) guarantees uniqueness even for the centralized MAP estimator and is satisfied by a number of unimodal pdfs encountered in practice. Note that for a2) to be satisfied not all individual pdf summands have to be strictly log-concave.

III. Distributed MAP Estimation

In this section, we derive a distributed (D) MAP estimator for s1) under a1) and a2). To this end, since the summands in (1) are coupled through $\mathbf{s}$ we introduce an auxiliary variable $\mathbf{s}_j$ to represent the local estimate of $\mathbf{s}$ at sensor j. Using $\mathbf{s}_j$ we can rewrite (1) as a *constrained* optimization problem

$$\{\hat{\mathbf{s}}_j\}_{j=1}^{J} := \arg\min_{\mathbf{s}_j} -\sum_{j=1}^{J} \left(\ln[p_j(\mathbf{x}_j|\mathbf{s}_j)] + J^{-1}\ln[p(\mathbf{s}_j)]\right), \quad \text{s.t.} \quad \bar{\mathbf{s}}_b = \mathbf{s}_j, \quad \forall b \in \mathcal{B}, j \in \mathcal{N}_b \tag{5}$$

where the set $\mathcal{B} \subseteq [1, J]$ denotes the bridge sensor subset introduced in [17, Definition 1], and $b \in \mathcal{N}_b$. Note that for a specific $b \in \mathcal{B}$ the equality constraint $\bar{\mathbf{s}}_b = \mathbf{s}_j$ in (5) is imposed for all those sensors j in the neighborhood of b; i.e., all $j \in \mathcal{N}_b$. Recall that $\mathcal{B}$ is a bridge sensor subset if and only if a) $\forall\, j \in [1, J]$ there exists at least one $b \in \mathcal{B}$ so that $b \in \mathcal{N}_j$; and b) if j_1 and j_2 are single-hop neighboring sensors, there must exist a bridge sensor b so that $b \in \mathcal{N}_{j_1} \cap \mathcal{N}_{j_2}$. Each sensor $b \in \mathcal{B}$ maintains a local vector $\bar{\mathbf{s}}_b$ via which consensus among local variables $\mathbf{s}_j$ across all sensors is achieved *a fortiori*, i.e., $\mathbf{s}_j = \mathbf{s}_{j'}\ \forall\, j', j$. The bridge sensor set can be found during the start-up phase using e.g., the distributed algorithm in [21].

For the WSN in Fig. 1 (top) the filled circles denote bridge sensors. Henceforth, the set of bridge neighbors of the jth sensor will be denoted as $\mathcal{B}_j := \mathcal{N}_j \cap \mathcal{B}$, and its cardinality by $|\mathcal{B}_j|$ for $j = 1, \dots, J$. We have proved in [17] that consensus can be achieved across all J sensors, if consensus is reached only among a subset of them, namely $\mathcal{B}$. Conditions a) and b) provide a necessary and sufficient condition for the equivalence of (1) and (5) in the sense that $\hat{\mathbf{s}}_j = \hat{\mathbf{s}}_{\mathrm{map}}\ \forall\, j \in [1, J]$.

A. The D-MAP Algorithm

Consider the augmented Lagrangian of (5), given by

$$\mathcal{L}[\boldsymbol{s}, \bar{\mathbf{s}}, \mathbf{v}] = -\sum_{j=1}^{J} \left(\ln[p_j(\mathbf{x}_j|\mathbf{s}_j)] + J^{-1}\ln[p(\mathbf{s}_j)]\right) + \sum_{b\in\mathcal{B}}\sum_{j\in\mathcal{N}_b} \left(\mathbf{v}_j^b\right)^T(\mathbf{s}_j - \bar{\mathbf{s}}_b) + \sum_{b\in\mathcal{B}}\sum_{j\in\mathcal{N}_b} \frac{c_j}{2}\|\mathbf{s}_j - \bar{\mathbf{s}}_b\|_2^2 \tag{6}$$

where $\mathbf{s} := \{\mathbf{s}_j\}_{j=1}^{J}$, $\bar{\mathbf{s}} := \{\bar{\mathbf{s}}_b\}_{b\in\mathcal{B}}$, $\mathbf{v} := \{\mathbf{v}_j^b\}_{j\in[1,J]}^{b\in\mathcal{B}_j}$ comprises the Lagrange multiplier vectors, and $c_j > 0$ are penalty coefficients corresponding to the constraint $\mathbf{s}_j = \bar{\mathbf{s}}_b\ \forall b \in \mathcal{B}$. The penalty terms $(c_j/2)\|\mathbf{s}_j - \bar{\mathbf{s}}_b\|_2^2$ ensure strict convexity of the local minimization problems and thus guarantee globally convergent local recursions even when individual sensor pdfs are not strictly convex. Using the alternating direction method of multipliers [6] to minimize (6), and allowing for additive noise in sensor exchanges, we can mimick the steps in [17, App. B] to prove the following proposition.

Proposition 1: Per time index k consider iterates $\mathbf{v}_j^b(k), \mathbf{s}_j(k)$ and $\bar{\mathbf{s}}_b(k)$ defined by the recursions

$$\mathbf{v}_j^b(k) = \mathbf{v}_j^b(k-1) + c_j\left[\mathbf{s}_j(k) - \left(\bar{\mathbf{s}}_b(k) + \boldsymbol{\eta}_j^b(k)\right)\right] \tag{7}$$

$$\mathbf{s}_j(k+1) = \arg\min_{\mathbf{s}_j}\Bigg[-\ln\left[p_j(\mathbf{x}_j|\mathbf{s}_j)p^{1/J}(\mathbf{s}_j)\right] + \sum_{b\in\mathcal{B}_j}\left(\mathbf{v}_j^b(k)\right)^T\left[\mathbf{s}_j - \left(\bar{\mathbf{s}}_b(k) + \boldsymbol{\eta}_j^b(k)\right)\right] + \frac{c_j}{2}\sum_{b\in\mathcal{B}_j}\left\|\mathbf{s}_j - \left(\bar{\mathbf{s}}_b(k) + \boldsymbol{\eta}_j^b(k)\right)\right\|_2^2\Bigg] \tag{8}$$

[1]Subscripts specify the sensor at which variables are "controlled" (e.g., computed at and/or transmitted to neighbor sensors), while superscripts denote the sensor at which the variable is communicated to.

$$\bar{\mathbf{s}}_b(k+1)=\sum_{j\in\mathcal{N}_b}\frac{1}{\sum_{\beta\in\mathcal{N}_b}c_\beta}\left[\mathbf{v}_j^b(k)+c_j\left(\mathbf{s}_j(k+1)+\bar{\boldsymbol{\eta}}_b^j(k+1)\right)\right],\qquad b\in\mathcal{B}_j \tag{9}$$

and let the initial values $\{\mathbf{v}_j^b(-1)\}_{b\in\mathcal{B}_j}$, $\{\mathbf{s}_j(0)\}_{j=1}^J$ and $\{\bar{\mathbf{s}}_b(0)\}_{b\in\mathcal{B}}$ be arbitrary. Under a1), a2), and assuming ideal communication links i.e., $\boldsymbol{\eta}_j^b(k)=\mathbf{0}$ and $\bar{\boldsymbol{\eta}}_b^j(k)=\mathbf{0}$, the local estimates $\mathbf{s}_j(k)$ converge to the centralized MAP estimator as $k\to\infty$ and the WSN reaches consensus; i.e., $\lim_{k\to\infty}\mathbf{s}_j(k)=\lim_{k\to\infty}\bar{\mathbf{s}}_b(k)=\hat{\mathbf{s}}_{\text{map}},\forall j=1,\ldots,J,b\in\mathcal{B}$.

Recursions (7)–(9) constitute the D-MAP algorithm. During the kth iteration, sensor j receives the noisy consensus variables $\bar{\mathbf{s}}_b(k)+\boldsymbol{\eta}_j^b(k)$ from all its bridge neighbors in $b\in\mathcal{B}_j$. Based on these consensus variables, it updates through (7) the Lagrange multipliers $\{\mathbf{v}_j^b(k)\}_{b\in\mathcal{B}_j}$, which are then utilized to compute $\mathbf{s}_j(k+1)$ via (8). Then, sensor j transmits to each of its bridge neighbors $b\in\mathcal{B}_j$ the vector $c_j^{-1}\mathbf{v}_j^b(k)+\mathbf{s}_j(k+1)$. Each bridge sensor $b\in\mathcal{B}$ receives the vectors $c_j^{-1}\mathbf{v}_j^b(k)+\mathbf{s}_j(k+1)+\bar{\boldsymbol{\eta}}_b^j(k+1)$ from all its neighbors $j\in\mathcal{N}_b$ and averages them after scaling with c_j to obtain $\bar{\mathbf{s}}_b(k+1)$ while suppressing noise [cf. (9)]. Notice, that bridge sensor b acquires $\{c_j\}_{j\in\mathcal{N}_b}$ from their neighbors during a start-up period of the WSN. This completes the kth iteration, after which all sensors in $\mathcal{B}$ transmit $\bar{\mathbf{s}}_b(k+1)$ to all their neighbors $j\in\mathcal{N}_b$, which can proceed to the $(k+1)$st iteration; see also Fig. 1 (bottom).

The local minimization problem in (8) is strictly convex, due to a2) and the strict convexity of the Euclidean norm. Thus, the optimal solution of (8) is unique and can be obtained accurately after solving, e.g., using Newton's method, the nonlinear equation

$$\begin{aligned}&-\left[p_j\left(\mathbf{x}_j|\mathbf{s}_j(k+1)\right)\right]^{-1}\nabla_{\mathbf{s}_j}p_j\left(\mathbf{x}_j|\mathbf{s}_j(k+1)\right)\\&-J^{-1}\left[p\left(\mathbf{s}_j(k+1)\right)\right]^{-1}\nabla_{\mathbf{s}_j}p\left(\mathbf{s}_j(k+1)\right)+\sum_{b\in\mathcal{B}_j}\mathbf{v}_j^b(k)\\&-c_j\sum_{b\in\mathcal{B}_j}\left[\bar{\mathbf{s}}_b(k)+\boldsymbol{\eta}_j^b(k)\right]+c_j|\mathcal{B}_j|\mathbf{s}_j(k+1)=\mathbf{0}.\end{aligned} \tag{10}$$

Resemblance of (7)–(9) with a stochastic gradient algorithm, explains why additive noise causes $\mathbf{s}_j(k)$ to fluctuate around the optimal solution $\hat{\mathbf{s}}_{\text{map}}$ with the magnitude of fluctuations being proportional to the noise variance. As with D-MLE in [17], this implies that $\mathbf{s}_j(k)$ in the presence of noise is guaranteed to be within a ball around $\hat{\mathbf{s}}_{\text{map}}$ with high probability. Noise robustness of D-MAP will be confirmed also by simulations.

Remark 1: The distributed schemes in [5], [8], [16], [22], and [24] require knowledge of the desired centralized estimator in closed form. Similar to D-MLE in [17], the D-MAP algorithm in (7)–(9) does not require a closed-form expression for the MAP estimator. The edge of D-MAP over D-MLE is twofold: i) similar to all Bayesian approaches D-MAP facilitates incorporation of *a priori* information about the unknown $\mathbf{s}$; and ii) the conditional independence in (1) allows for correlated sensor data (not possible in [17] and [24]). Furthermore, the D-MAP formulation subsumes distributed estimation of Markov random fields [9] if $\mathbf{s}$ is formed so that its jth entry is the sample of the field measured by sensor j. Although field estimation is beyond the scope of this paper, the D-MAP approach here allows for distributed field estimation without imposing the Markovianity assumption in [9].

Remark 2: In case of a bridge sensor failure, D-MAP incurs performance loss, but remains operational after the neighbors of the failed bridge sensor modify their local recursions accordingly. Specifically, if bridge sensor $b'\in\mathcal{B}$ fails, then some of the nodes in $\mathcal{N}_{b'}$ can be converted to bridges as needed, in order for the new bridge sensor set, call it $\mathcal{B}'$, to satisfy the properties of $\mathcal{B}$. This conversion can be accommodated using the algorithm in [21]. Then, all sensors in $\mathcal{N}_{b'}$ can modify their local recursions (7)–(9) by adding the corresponding terms associated with the new bridges in $\mathcal{N}_{b'}$, and removing the ones corresponding to b'. D-MAP will converge to the MAP estimate given in (1) after excluding the term associated with sensor b'. The same approach can also be followed in the following distributed algorithms.

B. Linear-Gaussian and Quantized Observations

Consider as a special case the popular linear-Gaussian model $\{\mathbf{x}_j=\mathbf{h}_j\mathbf{s}+\mathbf{n}_j\}_{j=1}^J$, where $\mathbf{h}_j$ is deterministic, $\mathbf{s}$ is zero-mean Gaussian, and $\mathbf{n}_j$ is zero-mean Gaussian with covariance $\mathbf{C}_{n_jn_j}$ and uncorrelated across sensors. Clearly $p_j(\mathbf{x}_j|\mathbf{s}_j)$ is Gaussian with mean $\mathbf{H}_j\mathbf{s}_j$ and covariance matrix $\mathbf{C}_{n_jn_j}\ \forall j$. After making the necessary substitutions in (10), the optimal solution of (8) is now available in closed form as

$$\mathbf{s}_j(k+1)=\left[\mathbf{H}_j^T\mathbf{C}_{n_jn_j}^{-1}\mathbf{H}_j+c_j|\mathcal{B}_j|\mathbf{I}_p+(J\mathbf{C}_{ss})^{-1}\right]^{-1}\times\left[\mathbf{H}_j^T\mathbf{C}_{n_jn_j}^{-1}\mathbf{x}_j+\sum_{b\in\mathcal{B}_j}\left[c_j\left(\bar{\mathbf{s}}_b(k)+\boldsymbol{\eta}_j^b(k)\right)-\mathbf{v}_j^b(k)\right]\right].$$

Because the matrix inverse does not depend on k and can be evaluated off-line, it is clear that D-MAP is simple to implement for the important class of linear-Gaussian models.

The next special case entails a nonlinear data model, where due to limited sensing capabilities sensors have to rely on a binary quantized version $\mathbf{b}_j$ of $\mathbf{x}_j$ to estimate $\mathbf{s}$; see also [18]. Also, note that such harsh quantization allows for employment of powerful error control codes, where many redundant bits can be used to mitigate nonideal channel effects. Each local quantizer, say that of the jth sensor, splits the measurement space $\mathbf{r}^{L_j}$ into Q_j convex regions $\{\mathcal{R}_{jq}\}_{q=1}^{Q_j}$. Vector quantization of $\mathbf{x}_j$ yields $\mathbf{b}_j(q)=\mathbf{1}\{\mathbf{x}_j\in\mathcal{R}_{jq}\}$, where $\mathbf{1}$ denotes the indicator function; i.e., vector $\mathbf{b}_j(q)$ has binary 0/1 entries (1 if and only if $\mathbf{x}_j$ falls in the quantization region $\mathcal{R}_{jq}$). The MAP estimator can be determined as in (1), with $p_j(\mathbf{x}_j|\mathbf{s}_j)$ substituted by $\Pr_j(\mathbf{b}_j|\mathbf{s}_j)=\sum_{q=1}^{Q_j}\Pr_j(\mathbf{b}_j(q)|\mathbf{s}_j)\delta(\mathbf{b}_j-\mathbf{b}_j(q))$, where $\delta(\cdot)$ denotes Kronecker's delta, while $\Pr_j(\mathbf{b}_j(q)|\mathbf{s}_j):=\int_{\mathbf{x}_j\in\mathcal{R}_{jq}}p_j(\mathbf{x}_j|\mathbf{s}_j)d\mathbf{x}_j$. Being the integral of a log-concave function over the convex set $\mathcal{R}_{jq}$, $\Pr_j(\mathbf{b}_j(q)|\mathbf{s}_j)$ is log concave. Thus, the D-MAP algorithm is still applicable despite the fact that the centralized MAP estimator is not expressible in closed form. The local estimate $\mathbf{s}_j(k+1)$ is determined from (8) after replacing $p_j(\mathbf{x}_j|\mathbf{s}_j)$ with $\Pr_j(\mathbf{b}_j|\mathbf{s}_j)$.

Example: Consider the D-MAP estimator with $J=60$ sensors, each quantizing $\mathbf{x}_j$ as described earlier. Nodes are randomly placed in the unit square [0,1] × [0,1] with uniform distribution. In this and all simulations of this paper the WSN is generated using the geometric random graph model [10], where two sensors are able to communicate if their Euclidean distance is less than 1/4. Sensor j acquires $L_j=5$ observations and $\mathbf{s}$ has

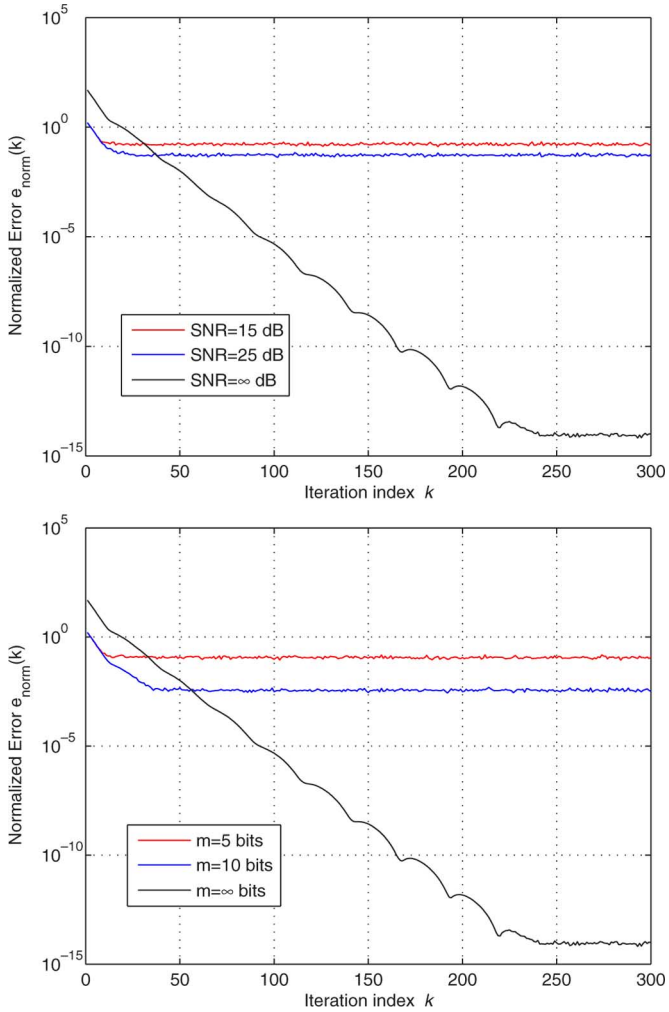

Fig. 2. Normalized error $e_{\rm norm}(k)$ versus iteration index k for D-MAP in presence of (top) reception noise with SNR $=$ 15, 25, and ∞ dB, and (bottom) quantization noise using $m = 5$, 10 and ∞ number of bits.

dimensionality $p = 2$ with $\mathbf{C}_{ss} = \text{diag}(0.2, 0.3)$. The entries of $\mathbf{H}_j$ are random uniformly distributed over [−0.5,0.5] and noise vectors $\{\mathbf{n}_j\}_{j=1}^J$ are zero-mean Gaussian with $\mathbf{C}_{n_j n_j} = 0.2\mathbf{I}_{p\times p}$. The quantizer at the jth sensor has $Q_j = 32$ regions $\mathcal{R}_{j1} = \{\mathbf{x}_{j,1} > 0\} \cup \ldots \cup \{\mathbf{x}_{j,5} > 0\}, \ldots, \mathcal{R}_{j32} = \{\mathbf{x}_{j,1} < 0\} \cup \ldots \cup \{\mathbf{x}_{j,5} < 0\}\ \forall\ j = 1, \ldots, J$. The penalty coefficients are set to $c_j = 4/|\mathcal{B}_j|$. The normalized error $e_{\rm norm}(k) := \sum_{j=1}^J \|\mathbf{s}_j(k) - \hat{\mathbf{s}}_{\rm map}\|_2 / \|\hat{\mathbf{s}}_{\rm map}\|_2$ is used as figure of merit.

Fig. 2 (top)–(bottom) indicates that $e_{\rm norm}(k) \to 0$ as $k \to \infty$ under ideal links, corroborating Proposition 1. In the presence of reception or quantization noise, we average $e_{\rm norm}(k)$ over 50 independent realizations of D-MAP estimates. Fig. 2 (top) depicts $e_{\rm norm}(k)$ with $\mathbf{C}_{\eta_j\eta_j} = \sigma^2\mathbf{I}_p$, where σ^2 is selected so that SNR $:= 10\log_{10}[\hat{\mathbf{s}}_{ML}/(p\sigma^2)] = 15$ dB and SNR $= 25$ dB. Similar behavior is observed in Fig. 2 (bottom) when vectors $\{c_j^{-1}\mathbf{v}_j^b(k) + \mathbf{s}_j(k+1)\}_{b\in\mathcal{B}_j}$ and $\{\bar{\mathbf{s}}_b(k)\}_{b\in\mathcal{B}}$ are quantized before transmission to their neighbors.

IV. Distributed LMMSE Estimation

Here, we confine distributed estimators to be linear in the data [cf. s2)] but allow correlation among sensors and do not require conditional independence as in the D-MAP estimation setup s1).

A. The D-LMMSE Algorithm

Although $\hat{\mathbf{s}}_{\rm lmmse}$ is well known in the centralized form (2), the equivalent expression given in the next lemma lends itself to a distributed LMMSE estimation algorithm (see Appendix A for the proof).

Lemma 1: With $\mathbf{y}$ representing an auxiliary vector, the LMMSE in (2) can be equivalently expressed as

$$(\hat{\mathbf{s}}_{\rm lmmse}, \hat{\mathbf{y}}) = \arg \min_{\mathbf{s}\in\mathbb{R}^{p\times 1}, \mathbf{y}\in\mathbb{R}^{L\times 1}} \sum_{j=1}^J \left\|\mathbf{s} - J\mathbf{C}_{sx_j}(\mathbf{y})_j\right\|_2^2, \quad \text{s.t.} \quad \mathbf{C}_{xx}\mathbf{y} = \mathbf{x} \tag{11}$$

where $\mathbf{y} := [(\mathbf{y})_1^T \ldots (\mathbf{y})_J^T]^T$ with $(\mathbf{y})_j \in \mathbf{R}^{L_j\times 1}$ denoting the jth block sub-vector of $\mathbf{y}$, and $\hat{\mathbf{y}} = \mathbf{C}_{xx}^{-1}\mathbf{x}$.

Different from D-MAP and distributed estimators for deterministic signals in [17] the constraint $\mathbf{C}_{xx}\mathbf{y} = \mathbf{x}$ entails also sensor correlations and motivates besides $\mathbf{s}_j$ a local estimate[2] $\mathbf{y}_j$ of $\mathbf{y}$ per sensor j. Using the bridge sensor set $\mathcal{B}$ along with the variables $\bar{\mathbf{s}}_b$ and $\bar{\mathbf{y}}_b$, the following separable formulation of (11) can be obtained along the lines of the proof in [17, App. A].

Lemma 2: The optimization problem in (11) is equivalent to

$$\begin{aligned}\{\hat{\mathbf{s}}_j, \hat{\mathbf{y}}_j\}_{j=1}^J := \arg\min \sum_{j=1}^J \left\|\mathbf{s}_j - J\mathbf{C}_{sx_j}(\mathbf{y}_j)_j\right\|_2^2 \\ \text{s.t. } \bar{\mathbf{s}}_b = \mathbf{s}_j, \bar{\mathbf{y}}_b = \mathbf{y}_j,\ b \in \mathcal{B},\ j \in \mathcal{N}_b \\ \text{and} \quad \mathbf{C}_{x_j}\mathbf{y}_j = \mathbf{x}_j,\ j \in [1, J]\end{aligned} \tag{12}$$

in the sense that $\hat{\mathbf{s}}_j = \hat{\mathbf{s}}_{\rm lmmse}$, and $\hat{\mathbf{y}}_j = \hat{\mathbf{y}}$; while $\mathbf{y}_j := [(\mathbf{y}_j)_1^T \ldots (\mathbf{y}_j)_J^T]^T$ with $(\mathbf{y}_j)_i \in \mathbb{R}^{L_i\times 1}$.

If $\{\mathbf{v}_j^b, \mathbf{w}_j^b\}_{b\in\mathcal{B}_j}$ are the Lagrange multipliers associated with the constraints $\mathbf{s}_j = \bar{\mathbf{s}}_b$ and $\mathbf{y}_j = \bar{\mathbf{y}}_b$ respectively, the augmented Lagrangian corresponding to (12) is

$$\begin{aligned}&\mathcal{L}[\boldsymbol{s}, \bar{\mathbf{s}}, \boldsymbol{y}, \bar{\mathbf{y}}, \mathbf{v}, \mathbf{w}] \\ &= \sum_{j=1}^J \left\|\mathbf{s}_j - J\mathbf{C}_{sx_j}(\mathbf{y}_j)_j\right\|_2^2 \\ &\quad + \sum_{b\in\mathcal{B}}\sum_{j\in\mathcal{N}_b} \left[\left(\mathbf{v}_j^b\right)^T(\mathbf{s}_j - \bar{\mathbf{s}}_b) + \left(\mathbf{w}_j^b\right)^T(\mathbf{y}_j - \bar{\mathbf{y}}_b)\right] \\ &\quad + \sum_{b\in\mathcal{B}}\sum_{j\in\mathcal{N}_b} \left[\frac{c_j}{2}\|\mathbf{s}_j - \bar{\mathbf{s}}_b\|_2^2 + \frac{d_j}{2}\|\mathbf{y}_j - \bar{\mathbf{y}}_b\|_2^2\right]\end{aligned}$$

where $\boldsymbol{y} := \{\mathbf{y}_j\}_{j\in[1,J]}$, $\bar{\mathbf{y}} := \{\bar{\mathbf{y}}_b\}_{b\in\mathcal{B}}$ and $\mathbf{w} := \{\mathbf{w}_j^b\}_{j\in[1,J]}^{b\in\mathcal{B}_j}$.

The main result pertaining to the distributed solution of (12) is proved in Appendix B and can be summarized as follows.

Proposition 2: For each sensor j, let iterates $\mathbf{v}_j^b(k), \mathbf{w}_j^b(k)$, $\mathbf{s}_{y,j}(k)$ and $\bar{\mathbf{s}}_{y,b}(k)$ be defined by the recursions

$$\begin{aligned}&\left[\left(\mathbf{v}_j^b(k)\right)^T, \left(\mathbf{w}_j^b(k)\right)^T\right] \\ &\quad = \left[\left(\mathbf{v}_j^b(k-1)\right)^T, \left(\mathbf{w}_j^b(k-1)\right)^T\right] \\ &\qquad + \left[c_j\left(\mathbf{s}_j(k) - \bar{\mathbf{s}}_b(k)\right)^T, d_j\left(\mathbf{y}_j(k) - \bar{\mathbf{y}}_b(k)\right)^T\right]\end{aligned} \tag{13}$$

$$\begin{aligned}&\mathbf{s}_{y,j}(k+1) \\ &\quad := \left[\mathbf{s}_j^T(k+1), \mathbf{y}_j^T(k+1)\right]^T \\ &\quad = \hat{\mathbf{x}}_j - \mathbf{F}_j^{-1}(\mathbf{I}_{L+p} - \mathbf{G}_j)\boldsymbol{\zeta}_j(k), \quad j \in [1, J]\end{aligned} \tag{14}$$

[2]Note that $\mathbf{y}_j$ is not the same as $(\mathbf{y})_j$.

$$\begin{aligned}
&\bar{\mathbf{s}}_{y,b}(k+1)\\
&\quad:=\left[\bar{\mathbf{s}}_b^T(k+1),\bar{\mathbf{y}}_b^T(k+1)\right]^T\\
&\quad=\sum_{j\in\mathcal{N}_b}\operatorname{diag}\left(\left(\sum_{\beta\in\mathcal{N}_b}c_\beta\right)^{-1}\mathbf{I}_p,\left(\sum_{\beta\in\mathcal{N}_b}d_\beta\right)^{-1}\mathbf{I}_L\right)\\
&\qquad\times\left(\left[\left(\mathbf{v}_j^b(k)\right)^T,\left(\mathbf{w}_j^b(k)\right)^T\right]^T\right.\\
&\qquad\left.+\operatorname{diag}(c_j\mathbf{I}_p,d_j\mathbf{I}_L)\mathbf{s}_{y,j}(k+1)\right) \qquad (15)
\end{aligned}$$

where $b\in\mathcal{B}$ and quantities $\hat{\mathbf{x}}_j$, $\mathbf{F}_j$, $\mathbf{G}_j$ and $\boldsymbol{\zeta}_j(k)$ involve only local quantities and are defined as

$$\mathbf{F}_j:=2\begin{bmatrix}(1+0.5c_j|\mathcal{B}_j|)\mathbf{I}_p & -J\bar{\mathbf{C}}_{sx_j}^T\\ -J\bar{\mathbf{C}}_{sx_j} & J^2\bar{\mathbf{C}}_{sx_j}\bar{\mathbf{C}}_{sx_j}^T+0.5d_j|\mathcal{B}_j|\mathbf{I}_L\end{bmatrix} \qquad (16)$$

$$\boldsymbol{\zeta}_j(k):=\left[\sum_{b\in\mathcal{B}_j}\left(\mathbf{v}_j^b(k)-c_j\bar{\mathbf{s}}_b(k)\right)^T\sum_{b\in\mathcal{B}_j}\left(\mathbf{w}_j^b(k)-d_j\bar{\mathbf{y}}_b(k)\right)^T\right]^T \qquad (17)$$

$$\hat{\mathbf{x}}_j:=\mathbf{F}_j^{-1}\bar{\mathbf{C}}_{x_j}^T\left(\bar{\mathbf{C}}_{x_j}\mathbf{F}_j^{-1}\bar{\mathbf{C}}_{x_j}^T\right)^{-1}\mathbf{x}_j, \qquad (18)$$

$$\mathbf{G}_j:=\bar{\mathbf{C}}_{x_j}^T\left(\bar{\mathbf{C}}_{x_j}\mathbf{F}_j^{-1}\bar{\mathbf{C}}_{x_j}^T\right)^{-1}\bar{\mathbf{C}}_{x_j}\mathbf{F}_j^{-1} \qquad (19)$$

with $\bar{\mathbf{C}}_{sx_j}:=[\mathbf{0}_{p\times L_1}\ldots\mathbf{C}_{sx_j}\ldots\mathbf{0}_{p\times L_J}]^T$ and $\bar{\mathbf{C}}_{x_j}:=[\mathbf{0}_{L_j\times p}\mathbf{C}_{x_j}]$, while $\{c_j,d_j>0\}_{j=1}^J$. If links are ideal, iterates $\mathbf{s}_{y,j}(k)$ and $\bar{\mathbf{s}}_{y,b}(k)$ converge under a1) to the LMMSE estimator as $k\to\infty$; i.e., for all $j\in[1,J]$ and $b\in\mathcal{B}$, it holds

$$\lim_{k\to\infty}\left[\mathbf{s}_j(k)^T,\mathbf{y}_j(k)^T\right]^T=\lim_{k\to\infty}\bar{\mathbf{s}}_{y,b}(k)=\left[\hat{\mathbf{s}}_{\text{lmmse}}^T,\hat{\mathbf{y}}^T\right]^T.$$

Recursive updates (13)–(15) constitute the D-LMMSE estimator whereby every sensor j keeps track of the local estimate $\mathbf{s}_j(k)$ along with $\mathbf{y}_j(k)$ and the Lagrange multipliers $\{\mathbf{v}_j^b(k),\mathbf{w}_j^b(k)\}_{b\in\mathcal{B}_j}$. Sensors $b\in\mathcal{B}$ update also the consensus variables $\bar{\mathbf{s}}_b(k)$ and $\bar{\mathbf{y}}_b(k)$. During the kth iteration, sensor j receives $\bar{\mathbf{s}}_{y,b}(k)$ from all its bridge neighbors in $\mathcal{B}_j$. Based on these, it updates its Lagrange multipliers $\{\mathbf{v}_j^b(k),\mathbf{w}_j^b(k)\}_{j\in\mathcal{B}_j}$ using (13), which are then used next to compute $\mathbf{s}_{y,j}(k+1)$ via (14). After completing this iteration step, sensor j transmits to each of its neighbors $b\in\mathcal{B}_j$ the vectors $c_j^{-1}\mathbf{v}_j^b(k)+\mathbf{s}_j(k+1)$ and $d_j^{-1}\mathbf{w}_j^b(k)+\mathbf{y}_j(k+1)$. Each sensor $b\in\mathcal{B}$ receives these vectors from all its neighbors $j\in\mathcal{N}_b$, multiplies them by c_j and d_j respectively, and proceeds to compute $\bar{\mathbf{s}}_{y,b}(k+1)$ via (15). This completes the kth iteration, after which all sensors in $\mathcal{B}$ proceed to transmit $\bar{\mathbf{s}}_b(k+1)$ and $\bar{\mathbf{y}}_b(k+1)$ to their neighbors $j\in\mathcal{N}_b$.

Because matrix $\mathbf{F}_j$ is symmetric, positive definite and does not depend on k, $\mathbf{F}_j^{-1}$ can be found off-line. Hence, the complexity for computing $\mathbf{s}_{y,j}(k+1)$ is $\mathcal{O}((L+p)^2)$. Actually, for the typical case $L>p$ one can take advantage of the special structure of $\mathbf{F}_j$ and utilize the linear system solving procedure described in, e.g., [7, p. 512] to reduce complexity down to $\mathcal{O}((L+p)p)$. Furthermore, the communication process involves transmission of $\mathcal{O}(L+p)$ scalars per iteration which is expected since the correlation information in $\mathbf{C}_{xx}$ and $\mathbf{C}_{sx}$ is scattered across the network.

Remark 3: Relative to [8], the D-LMMSE algorithm neither requires linearity nor Gaussianity in the data model. When the goal in [8] is to have all sensors consent to the LMMSE estimator of a common signal vector (as here), the scheme in [8] applies as long as each sensor can exchange information with all other sensors in the network. However, D-LMMSE guarantees convergence to $\hat{\mathbf{s}}_{\text{lmmse}}$ via single-hop links. Notice also that D-LMMSE inverts $\mathbf{C}_{xx}$ in a distributed fashion, through the local variables $\mathbf{y}_j(k)$ which is essential since every sensor needs to know only a portion of $\mathbf{C}_{xx}$. A fair comparison of [8] with the present D-LMMSE algorithm does not appear possible, since the former requires full connectivity among sensors so that each sensor can consent to the LMMSE estimator of $\mathbf{s}$. Another issue is the fact that [8] relies on inverse covariance matrices which facilitates estimation but leaves open the question of whether possible and how costly acquiring this information is. D-LMMSE recursions appear to have similar structure with the D-BLUE ones in [17]. However, D-LMMSE is more general in scope because it can handle correlated sensor data and guarantees convergence to the MSE optimal linear estimators regardless of the underlying data model.

Algorithm 1 D-LMMSE Estimation

Initialize $\{\mathbf{s}_{y,j}(0)\}_{j=1}^J$, $\{\bar{\mathbf{s}}_{y,b}(0)\}_{b\in\mathcal{B}}$ and $\{\mathbf{v}_j^b(0),\mathbf{w}_j^b(0)\}_{j\in[1,J]}^{b\in\mathcal{B}_j}$ to zero. All bridge sensors $b\in\mathcal{B}$, acquire $\{c_j\}_{j\in\mathcal{N}_b}$.

for $k=0,1,\ldots$ **do**

Bridge sensor $b\in\mathcal{B}$: transmit $\bar{\mathbf{s}}_b(k),\bar{\mathbf{y}}_b(k)$ to its neighbors in $\mathcal{N}_b$

All $j\in[1,J]$: update $\{\mathbf{v}_j^b(k),\mathbf{w}_j^b(k)\}_{b\in\mathcal{B}_j}$ using (13).

All $j\in[1,J]$: update $\mathbf{y}_j(k+1)$ and $\mathbf{s}_j(k+1)$ using (14).

All $j\in[1,J]$: transmit $c_j^{-1}\mathbf{v}_j^b(k)+\mathbf{s}_j(k+1)$ and $d_j^{-1}\mathbf{w}_j^b(k)+\mathbf{y}_j(k+1)$ to each $b\in\mathcal{B}_j$

Bridge sensors $b\in\mathcal{B}$: compute $\bar{\mathbf{s}}_b(k+1)$ and $\bar{\mathbf{y}}_b(k+1)$ through (15).

end for

B. D-LMMSE Estimation for the Linear Data Model

For arbitrary data models, D-LMMSE estimation requires $\mathbf{C}_{sx_j}$ and $\mathbf{C}_{x_j}:=[\mathbf{C}_{x_jx_1}\ldots\mathbf{C}_{x_jx_J}]$ at sensor j. For the linear model $\mathbf{x}_j=\mathbf{H}_j\mathbf{s}+\mathbf{n}_j$, however, D-LMMSE estimation is possible so long as sensor j has available only its local parameters $\mathbf{H}_j$, $\mathbf{C}_{n_jn_j}$ and $\mathbf{C}_{ss}$. Indeed, for the linear model (2) reduces to

$$\hat{\mathbf{s}}_{\text{lmmse}}=\left(\mathbf{C}_{ss}^{-1}+\sum_{j=1}^J\mathbf{H}_j^T\mathbf{C}_{n_jn_j}^{-1}\mathbf{H}_j\right)^{-1}\sum_{j=1}^J\mathbf{H}_j^T\mathbf{C}_{n_jn_j}^{-1}\mathbf{x}_j. \qquad (20)$$

Arguing as in [17, App. A and C] the LMMSE estimator in (20) can be equivalently written as

$$\begin{aligned}
\{\hat{\mathbf{s}}_j\}_{j=1}^J:=\arg\min_{\mathbf{s}_j}&\sum_{j=1}^J\left\|\left(J^{-1}\mathbf{C}_{ss}^{-1}+\mathbf{H}_j^T\mathbf{C}_{n_jn_j}^{-1}\mathbf{H}_j\right)^{1/2}\mathbf{s}_j\right.\\
&\left.-\left(J^{-1}\mathbf{C}_{ss}^{-1}+\mathbf{H}_j^T\mathbf{C}_{n_jn_j}^{-1}\mathbf{H}_j\right)^{-1/2}\mathbf{H}_j^T\mathbf{C}_{n_jn_j}^{-1}\mathbf{x}_j\right\|_2^2\\
\text{s.t. }\bar{\mathbf{s}}_b=&\mathbf{s}_j,\quad b\in\mathcal{B},j\in\mathcal{N}_b \qquad (21)
\end{aligned}$$

where $\hat{\mathbf{s}}_1=\ldots=\hat{\mathbf{s}}_J=\hat{\mathbf{s}}_{\text{lmmse}}$. Forming the augmented Lagrangian corresponding to (21) and applying the alternating-di-

rection method of multipliers as in Appendix B, we obtain the following.

Proposition 3: For each sensor $j \in [1, J]$ let iterates $\mathbf{v}_j^b(k), \mathbf{s}_j(k+1)$ and $\bar{\mathbf{s}}_b(k+1)$ be defined by the recursions

$$\mathbf{v}_j^b(k) = \mathbf{v}_j^b(k-1) + c_j \left[\mathbf{s}_j(k) - \bar{\mathbf{s}}_b(k)\right], \quad b \in \mathcal{B}_j \tag{22}$$

$$\mathbf{s}_j(k+1) = 2\mathbf{B}_j^{-1}\mathbf{H}_j^T\mathbf{C}_{n_jn_j}^{-1}\mathbf{x}_j - \mathbf{B}_j^{-1}\sum_{b\in\mathcal{B}_j}\left[\mathbf{v}_j^b(k) - c_j\bar{\mathbf{s}}_b(k)\right] \tag{23}$$

$$\bar{\mathbf{s}}_b(k+1) = \sum_{j\in\mathcal{N}_b}\frac{1}{\sum_{\beta\in\mathcal{N}_b}c_\beta}\left[\mathbf{v}_j^b(k) + c_j\mathbf{s}_j(k+1)\right], \quad b \in \mathcal{B} \tag{24}$$

where $\mathbf{B}_j := 2(J^{-1}\mathbf{C}_{ss}^{-1} + \mathbf{H}_j^T\mathbf{C}_{n_jn_j}^{-1}\mathbf{H}_j) + c_j|\mathcal{B}_j|\mathbf{I}_p$. Under ideal links and a1), $\mathbf{s}_j(k)$ and $\bar{\mathbf{s}}_b(k)$ converge to the centralized LMMSE estimator in (20) as $k \to \infty$; i.e., $\lim_{k\to\infty}\mathbf{s}_j(k) = \lim_{k\to\infty}\bar{\mathbf{s}}_b(k) = \hat{\mathbf{s}}_{\text{lmmse}}$.

Relative to Proposition 2 which requires $\mathbf{C}_{x_j}$ at sensor j, the D-LMMSE estimator for linear data models summarized in Proposition 3 requires only $\mathbf{H}_j, \mathbf{C}_{n_jn_j}$ and $\mathbf{C}_{ss}$ which are available at sensor j.

V. Noise Robust D-LMMSE

In this section we derive a provably noise resilient D-LMMSE algorithm and analyze its convergence by studying the $\mathbf{s}_j(k+1)$ and $\mathbf{y}_j(k+1)$ updates.

A. Multiplier-Free D-LMMSE

As a first step, we eliminate the Lagrange multipliers $\mathbf{v}_j^b, \mathbf{w}_j^b$ and initialize properly recursions (13)–(15) to rewrite them as described next (see Appendix C for the proof).

Lemma 3: Initialize (13)–(15) with $\{\mathbf{v}_j^b(-1) = \mathbf{0}, \mathbf{w}_j^b(-1) = \mathbf{0}\}_{j\in[1,J]}^{b\in\mathcal{B}_j}$, $\{\bar{\mathbf{s}}_{y,b}(-1) = \mathbf{0}\}_{b\in\mathcal{B}}$ and $\{\mathbf{s}_{y,j}(0) = \hat{\mathbf{x}}_j\}_{j=1}^J$, with $\hat{\mathbf{x}}_j$ as in (19). The local iterates $\mathbf{s}_{y,j}(k+1)$ in (14) and the consensus enforcing variables $\bar{\mathbf{s}}_{y,b}(k) := [\bar{\mathbf{s}}_b^T(k), \bar{\mathbf{y}}_b^T(k)]^T$ in (15) can be rewritten for $k \geq -1$ as

$$\begin{aligned}\mathbf{s}_{y,j}(k+1) = &\left[\mathbf{I}_{L+p} - \mathbf{F}_j^{-1}(\mathbf{I}_{L+p} - \mathbf{G}_j)\mathbf{D}_j^1\right]\mathbf{s}_{y,j}(k) \\ &+ \mathbf{F}_j^{-1}(\mathbf{I}_{L+p} - \mathbf{G}_j)\mathbf{D}_j^2 \\ &\times \sum_{b\in\mathcal{B}_j}\left[2\bar{\mathbf{s}}_{y,b}(k) - \bar{\mathbf{s}}_{y,b}(k-1)\right]\end{aligned} \tag{25}$$

$$\bar{\mathbf{s}}_{y,b}(k+1) = \sum_{j\in\mathcal{N}_b}\mathbf{D}_j^2\mathbf{D}_b\mathbf{s}_{y,j}(k+1), \quad b \in \mathcal{B} \tag{26}$$

where $\mathbf{D}_j^1 := \text{diag}(c_j|\mathcal{B}_j|\mathbf{I}_p, d_j|\mathcal{B}_j|\mathbf{I}_L)$, $\mathbf{D}_j^2 := |\mathcal{B}_j|^{-1}\mathbf{D}_j^1$ and $\mathbf{D}_b := \text{diag}((\sum_{\beta\in\mathcal{N}_b}c_\beta)^{-1}\mathbf{I}_p, (\sum_{\beta\in\mathcal{N}_b}d_\beta)^{-1}\mathbf{I}_L)$.

Devoid of $\mathbf{v}_j^b(k)$ and $\mathbf{w}_j^b(k)$, recursions (25) and (26) are simpler than (13)–(15). Through (26) the consensus variables $\bar{\mathbf{s}}_{y,b}(k+1)$ are expressed as a weighted average of the neighborhood estimates $\{\mathbf{s}_{y,j}(k+1)\}_{j\in\mathcal{N}_b}$, while the constant terms $\{\hat{\mathbf{x}}_j\}_{j\in\mathcal{N}_b}$ are used to initialize $\mathbf{s}_{y,j}(0)$. As suggested by Lemma 3, the kth iteration starts with all sensors receiving $2\bar{\mathbf{s}}_{y,b}(k) - \bar{\mathbf{s}}_{y,b}(k-1)$ from their bridge-neighbors to compute $\mathbf{s}_{y,j}(k+1)$ via (25). Then, the bridge-sensors $b \in \mathcal{B}$ receive from their neighbors $\{\mathbf{s}_{y,j}(k+1)\}_{j\in\mathcal{N}_b}$ to update $\bar{\mathbf{s}}_{y,b}(k+1)$ using (26), and finally form $2\bar{\mathbf{s}}_{y,b}(k+1) - \bar{\mathbf{s}}_{y,b}(k)$ that they transmit to their neighbors to start the $(k+1)$st iteration. It follows from (25) and (26) that $\mathbf{s}_{y,j}(k+1)$ is updated using the consensus variables $\bar{\mathbf{s}}_{y,b}(k)$ and $\bar{\mathbf{s}}_{y,b}(k-1)$ for $b \in \mathcal{B}_j$, which are formed using the local estimates of sensors in the set $\{\mathcal{N}_b\}_{b\in\mathcal{B}_j}$, that contains all sensors within a distance of up to two hops from sensor j.

B. Differences-Based Noise Resilient D-LMMSE Estimation

Recursions (25) and (26) constitute an intermediate step based on which we build next a distributed noise-robust algorithm for D-LMMSE estimation. As in [17], we replace $\mathbf{s}_{y,j}(k)$ by the local variable $\boldsymbol{\phi}_j(k)$. We will show that the mean of successive differences of $\boldsymbol{\phi}_j(k)$ converges (in the mean when the noise is zero-mean) to the LMMSE estimates; i.e., $\lim_{k\to\infty}E[\boldsymbol{\phi}_j(k+1) - \boldsymbol{\phi}_j(k)] = [\hat{\mathbf{s}}_{\text{lmmse}}^T, \hat{\mathbf{y}}^T]^T$, while the covariance matrix of this difference remains bounded. Intuitively, noise terms that propagate from $\boldsymbol{\phi}_j(k)$ to $\boldsymbol{\phi}_j(k+1)$ cancel when considering the difference $\boldsymbol{\phi}_j(k+1) - \boldsymbol{\phi}_j(k)$, thus achieving the desired noise resilience. The following lemma is the counterpart of Lemma 3 for noise-resilient operation.

Lemma 4: If $\boldsymbol{\phi}_j(0) = \hat{\mathbf{x}}_j$ and $\boldsymbol{\phi}_j(-1) = \boldsymbol{\phi}_j(-2) = \mathbf{0}$, the second-order recursions

$$\begin{aligned}\boldsymbol{\phi}_j(k+1) = &\hat{\mathbf{x}}_j + \left(\mathbf{I}_{L+p} - \mathbf{F}_j^{-1}(\mathbf{I}_{L+p} - \mathbf{G}_j)\mathbf{D}_j^1\right)\boldsymbol{\phi}_j(k) \\ &+ \mathbf{F}_j^{-1}(\mathbf{I}_{L+p} - \mathbf{G}_j)\mathbf{D}_j^2\sum_{b\in\mathcal{B}_j}\left[2\bar{\boldsymbol{\phi}}_b(k) - \bar{\boldsymbol{\phi}}_b(k-1)\right]\end{aligned} \tag{27}$$

$$\bar{\boldsymbol{\phi}}_b(k+1) = \sum_{j\in\mathcal{N}_b}\mathbf{D}_j^2\mathbf{D}_b\boldsymbol{\phi}_j(k+1), \; b \in \mathcal{B}, \; k \geq -1 \tag{28}$$

yield iterates $\boldsymbol{\phi}_j(k)$ and $\bar{\boldsymbol{\phi}}_b(k)$ whose differences $\delta\boldsymbol{\phi}_j(k) := \boldsymbol{\phi}_j(k) - \boldsymbol{\phi}_j(k-1)$ and $\delta\bar{\boldsymbol{\phi}}_b(k) := \bar{\boldsymbol{\phi}}_b(k) - \bar{\boldsymbol{\phi}}_b(k-1)$ equal the iterates $\mathbf{s}_{y,j}(k)$ and $\bar{\mathbf{s}}_{y,b}(k)$ produced by (25) and (26), respectively.

Proof: Lemma 4 holds true for $k = -1, 0$, since from $\boldsymbol{\phi}_j(0) = \hat{\mathbf{x}}_j$ and $\boldsymbol{\phi}_j(-1) = \boldsymbol{\phi}_j(-2) = \mathbf{0}$ we find that $\mathbf{s}_j(-1) = \boldsymbol{\phi}_j(-1) - \boldsymbol{\phi}_j(-2) = \mathbf{0}$ and $\mathbf{s}_j(0) = \boldsymbol{\phi}_j(0) - \boldsymbol{\phi}_j(-1) = \mathbf{0} = \hat{\mathbf{x}}_j$. For $k \geq 1$, Lemma 4 follows by induction after subtracting the recursion for $\boldsymbol{\phi}_j(k)$ from the one for $\boldsymbol{\phi}_j(k+1)$. □

Using Lemma 4 and Proposition 2 we deduce that as $k \to \infty$, $\delta\boldsymbol{\phi}_j(k)$ and $\delta\bar{\boldsymbol{\phi}}_b(k)$ converge to $[\hat{\mathbf{s}}_{\text{lmmse}}^T, \hat{\mathbf{y}}^T]^T$ under ideal links. In the presence of noise, let $\bar{\boldsymbol{\psi}}_b(k) := 2\bar{\boldsymbol{\phi}}_b(k) - \bar{\boldsymbol{\phi}}_b(k-1)$ and $\boldsymbol{\psi}_j(k+1) := 2\boldsymbol{\phi}_j(k+1) - \boldsymbol{\phi}_j(k)$. Note that $\bar{\boldsymbol{\psi}}_b(k)$ equals the bth summand in (27), and replace (28) by

$$\bar{\boldsymbol{\psi}}_b(k+1) = \sum_{j\in\mathcal{N}_b}\mathbf{D}_j^2\mathbf{D}_b\boldsymbol{\psi}_j(k+1), \quad b \in \mathcal{B}. \tag{29}$$

Notice that (27)–(28) as well as (27) and (29) produce the same sequence of $\boldsymbol{\phi}_j(k)$ under ideal channels. However, when sensors communicate to their neighbors the vectors $\boldsymbol{\psi}_j(k+1)$ and $\bar{\boldsymbol{\psi}}_b(k)$ over nonideal links the noise present per iteration is smaller than when exchanging $\boldsymbol{\phi}_j(k), \bar{\boldsymbol{\phi}}_b(k)$ and $\boldsymbol{\phi}_j(k-1), \bar{\boldsymbol{\phi}}_b(k-1)$ separately. The steps involved in implementing locally (27) and (29) are i) all sensors j receive $\bar{\boldsymbol{\psi}}_b(k) + \boldsymbol{\eta}_j^b(k)$ from $b \in \mathcal{B}_j$ to form a (noisy) iterate $\boldsymbol{\phi}_j(k+1)$; and ii) bridge sensors receive $\boldsymbol{\psi}_j(k+1) + \bar{\boldsymbol{\eta}}_b^j(k+1)$ from $j \in \mathcal{N}_b$ to form the (noisy)

iterate $\bar{\boldsymbol{\psi}}_b(k+1)$. The noisy versions of (27) and (29) are

$$\begin{aligned}\boldsymbol{\phi}_j(k+1) =& \hat{\mathbf{x}}_j + \left(\mathbf{I}_{L+p} - \mathbf{F}_j^{-1}(\mathbf{I}_{L+p} - \mathbf{G}_j)\mathbf{D}_j^1\right)\boldsymbol{\phi}_j(k) \\ &+ \mathbf{F}_j^{-1}(\mathbf{I}_{L+p} - \mathbf{G}_j)\mathbf{D}_j^2 \sum_{b\in\mathcal{B}_j} \bar{\boldsymbol{\psi}}_b(k) \\ &+ \mathbf{F}_j^{-1}(\mathbf{I}_{L+p} - \mathbf{G}_j)\mathbf{D}_j^2 \sum_{b\in\mathcal{B}_j\setminus\{j\}} \boldsymbol{\eta}_j^b(k) \qquad (30)\end{aligned}$$

$$\bar{\boldsymbol{\psi}}_b(k+1) = \sum_{j\in\mathcal{N}_b} \mathbf{D}_j^2\mathbf{D}_b\boldsymbol{\psi}_j(k+1) + \sum_{j\in\mathcal{N}_b, j\neq b} \mathbf{D}_j^2\mathbf{D}_b\bar{\boldsymbol{\eta}}_b^j(k+1). \qquad (31)$$

Note that j is excluded from the second sum in (30) because if the jth sensor belongs also to $\mathcal{B}$ then it maintains a noise-free version of $\bar{\boldsymbol{\psi}}_j(k)$. Similarly, the bth bridge sensor has available locally $\boldsymbol{\psi}_b(k+1)$; thus, b is excluded from the second summation in (31), The local recursions (30) and (31) form the noise resilient (R) D-LMMSE algorithm. Both RD-LMMSE here and the RD-BLUE in [17] involve linear updating of state variables $\boldsymbol{\psi}_j(k)$ and $\bar{\boldsymbol{\psi}}_b(k)$ and follow the same communication steps. Although RD-LMMSE recursions appear similar to those in RD-BLUE, the matrices involved in RD-LMMSE have different structure necessitating separate convergence analysis. Pertinent to the ensuing RD-LMMSE convergence analysis is the global RD-LMMSE recursion formed by stacking (30) for $j = 1,\ldots,J$. To this end, let us define the matrices $\mathbf{A}_1 := \mathbf{F}_G\text{diag}(\mathbf{D}_1^1\ldots\mathbf{D}_J^1) - 2\mathbf{F}_G\mathbf{W}_E$ and $\mathbf{A}_2 := \mathbf{F}_G\mathbf{W}_E$ with $\mathbf{F}_G := \text{diag}(\mathbf{F}_1^{-1}(\mathbf{I}-\mathbf{G}_1)\ldots\mathbf{F}_J^{-1}(\mathbf{I}-\mathbf{G}_J))$ and

$$\begin{aligned}\mathbf{W}_E := \text{diag}\left(\mathbf{D}_1^2,\ldots,\mathbf{D}_J^2\right)\cdot\sum_{b\in\mathcal{B}}(\mathbf{I}_J\otimes\mathbf{D}_b)(\mathbf{e}_b\otimes\mathbf{I}_{L+p}) \\ \times(\mathbf{e}_b\otimes\mathbf{I}_{L+p})^T\text{diag}\left(\mathbf{D}_1^2,\ldots,\mathbf{D}_J^2\right) \qquad (32)\end{aligned}$$

where $\mathbf{e}_b$ denotes the bth column of the adjacency matrix $\mathbf{E}$ and $\otimes$ is the Kronecker product. Upon substituting (31) into (30), and stacking $\{\boldsymbol{\phi}_j(k)\}_{j=1}^J$ from (30) in $\boldsymbol{\phi}(k) := [\boldsymbol{\phi}_1^T(k)\ldots\boldsymbol{\phi}_J^T(k)]^T$, we show in Appendix D that

$$\boldsymbol{\phi}(k+1) = \hat{\mathbf{x}} + \boldsymbol{\phi}(k) - \mathbf{A}_1\boldsymbol{\phi}(k) - \mathbf{A}_2\boldsymbol{\phi}(k-1) + \bar{\boldsymbol{\eta}}(k) + \bar{\boldsymbol{\eta}}_b(k) \qquad (33)$$

where $\hat{\mathbf{x}} := [\hat{\mathbf{x}}_1^T,\ldots,\hat{\mathbf{x}}_J^T]^T$, and the noise vectors $\bar{\boldsymbol{\eta}}(k) := [\bar{\boldsymbol{\eta}}_1^T(k)\ldots\bar{\boldsymbol{\eta}}_J^T(k)]^T$ and $\bar{\boldsymbol{\eta}}_b(k) := [\bar{\boldsymbol{\eta}}_{b,1}^T(k)\ldots\bar{\boldsymbol{\eta}}_{b,J}^T(k)]^T$ have entries

$$\bar{\boldsymbol{\eta}}_j(k) := \mathbf{F}_j^{-1}(\mathbf{I}-\mathbf{G}_j)\mathbf{D}_j^2\sum_{b\in\mathcal{B}_j\setminus\{j\}}\boldsymbol{\eta}_j^b(k) \qquad (34)$$

$$\begin{aligned}\bar{\boldsymbol{\eta}}_{b,j}(k) :=& \mathbf{F}_j^{-1}(\mathbf{I}-\mathbf{G}_j)\mathbf{D}_j^2 \\ &\times\sum_{b\in\mathcal{B}_j}\sum_{j'\in\mathcal{N}_b, j'\neq b}\mathbf{D}_{j'}^2\mathbf{D}_b\bar{\boldsymbol{\eta}}_b^{j'}(k). \qquad (35)\end{aligned}$$

Observe the differences in matrices $\mathbf{A}_1,\mathbf{A}_2,\mathbf{W}_E$, as well as in $\bar{\boldsymbol{\eta}}(k)$ and $\bar{\boldsymbol{\eta}}_b(k)$ wrt the corresponding ones in RD-BLUE [17]. Before proceeding with the convergence analysis, we will find the covariance matrices for the noise vectors in (33) which are essential for the subsequent derivations. It follows from (34)–(35) that the covariance matrices of $\bar{\boldsymbol{\eta}}(k)$ and $\bar{\boldsymbol{\eta}}_b(k)$ are given respectively by

$$\mathbf{C}_{\bar{\eta}\bar{\eta}} = \mathbf{F}_G\mathbf{D}^2\boldsymbol{C}_1\mathbf{D}^2\mathbf{F}_G^T, \quad \mathbf{C}_{\bar{\eta}_b\bar{\eta}_b} = \mathbf{F}_G\mathbf{D}^2\boldsymbol{C}_2\mathbf{D}^2\mathbf{F}_G^T \qquad (36)$$

where $\mathbf{D}^2 := \text{diag}(\mathbf{D}_1^2,\ldots,\mathbf{D}_J^2)$. The $J(L+p)\times J(L+p)$ matrix $\boldsymbol{C}_1$ is diagonal with diagonal blocks $[\boldsymbol{C}_1]_{jj} = \sum_{b\in\mathcal{B}_j\setminus\{j\}}\mathbf{C}_{\eta_j\eta_j} = |\mathcal{B}_j\setminus\{j\}|\mathbf{C}_{\eta_j\eta_j}$, while $\boldsymbol{C}_2$ matrix consists of the $(L+p)\times(L+p)$ submatrices $[\boldsymbol{C}_2]_{jj'}$ for which

$$[\boldsymbol{C}_2]_{jj'} = \sum_{b\in\mathcal{B}_j\cap\mathcal{B}_{j'}}\sum_{j''\in\mathcal{N}_b\setminus\{b\}}\mathbf{D}_{j''}^2\mathbf{D}_b\mathbf{C}_{\bar{\eta}_b\bar{\eta}_b}\mathbf{D}_b\mathbf{D}_{j''}^2. \qquad (37)$$

C. Convergence Results

We want to show that $E[\delta\boldsymbol{\phi}_j(k)]$ converges to $[\hat{\mathbf{s}}_{\text{lmmse}}^T,\hat{\mathbf{y}}^T]^T$ as $k\to\infty$, while $E[\delta\boldsymbol{\phi}_j(k)(\delta\boldsymbol{\phi}_j(k))^T]$ remains bounded. The RD-LMMSE global recursion has the same structure as that of RD-BLUE, but matrices $\mathbf{A},\mathbf{A}_1,\mathbf{A}_2$, and $\bar{\boldsymbol{\eta}}(k),\bar{\boldsymbol{\eta}}_b(k)$ are different. In fact, we can express $\delta\boldsymbol{\phi}(k)$ as (cf. [17, App. G])

$$\begin{aligned}\delta\bar{\boldsymbol{\phi}}(k+1) :=& \begin{bmatrix}\delta\boldsymbol{\phi}(k+1)\\ \delta\boldsymbol{\phi}(k)\end{bmatrix} \\ =& \mathbf{A}^{k-1}\left(\begin{bmatrix}\mathbf{I}-\mathbf{A}_1\\ \mathbf{I}\end{bmatrix} - \mathbf{A}\begin{bmatrix}\mathbf{A}_1\\ -\mathbf{I}_{J(L+p)}\end{bmatrix}\right)\hat{\mathbf{x}} \\ &+ \begin{bmatrix}\mathbf{I}_{J(p+L)} & \mathbf{I}_{J(p+L)}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\bar{\boldsymbol{\eta}}(k)\\ \bar{\boldsymbol{\eta}}_b(k)\end{bmatrix} \\ &- \sum_{n=1}^{k}\mathbf{A}^{n-1}\begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_1\\ -\mathbf{I}_{J(p+L)} & -\mathbf{I}_{J(p+L)}\end{bmatrix} \\ &\times\begin{bmatrix}\bar{\boldsymbol{\eta}}(k-n)\\ \bar{\boldsymbol{\eta}}_b(k-n)\end{bmatrix} \qquad (38)\end{aligned}$$

where $\mathbf{A}$ is formed by the $J(L+p)\times J(L+p)$ submatrices $[\mathbf{A}]_{11} = \mathbf{I}_{Jp} - \mathbf{A}_1$, $[\mathbf{A}]_{12} = -\mathbf{A}_2$, $[\mathbf{A}]_{21} = \mathbf{I}_{Jp}$ and $[\mathbf{A}]_{22} = \mathbf{0}_{Jp}$, while $\boldsymbol{\phi}(0) := \hat{\mathbf{x}}$ and $\boldsymbol{\phi}(-1) := \mathbf{0}$. The mean of $\delta\bar{\boldsymbol{\phi}}(k)$ is equal to the first term in (38). The per iteration noise covariance matrix $\mathbf{C}_{\bar{\eta}}(k+1) := E[(\delta\bar{\boldsymbol{\phi}}(k+1) - E[\delta\bar{\boldsymbol{\phi}}(k+1)])(\delta\bar{\boldsymbol{\phi}}(k+1) - E[\delta\bar{\boldsymbol{\phi}}(k+1)])^T]$ can be evaluated using [17, eq. (44)] after making the necessary substitutions. The following proposition summarizes the asymptotic behavior of the RD-LMMSE estimator (see Appendix D for the proof).

Proposition 4: The RD-LMMSE recursions (30)–(31) reach consensus in the mean sense, i.e.,

$$\lim_{k\to\infty} E\left[\delta\boldsymbol{\phi}_j(k)\right] = \left[\hat{\mathbf{s}}_{\text{lmmse}}^T,\hat{\mathbf{y}}^T\right]^T \qquad (39)$$

while the noise covariance matrix converges to

$$\begin{aligned}\lim_{k\to\infty}\mathbf{C}_\eta(k) = \bar{\mathbf{C}}_{\bar{\eta}\bar{\eta}_b} + \sum_{i=L+p+1}^{2J(L+p)}\sum_{i'=L+p+1}^{2J(p+L)}\frac{\mathbf{u}_{A,i}\mathbf{u}_{A,i'}^T}{1-\lambda_{A,i}\lambda_{A,i'}}\mathbf{v}_{A,i}^T \\ \times\begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_1\\ -\mathbf{I} & -\mathbf{I}\end{bmatrix}\mathbf{C}_{\bar{\eta}\bar{\eta}_b}\begin{bmatrix}\mathbf{A}_1^T & -\mathbf{I}\\ \mathbf{A}_1^T & -\mathbf{I}\end{bmatrix}\mathbf{v}_{A,i'} \qquad (40)\end{aligned}$$

where $\lambda_{A,i},\mathbf{u}_{A,i},\mathbf{v}_{A,i}$ denote the ith largest eigenvalue of $\mathbf{A}$ and the corresponding left and right eigenvectors, respectively, for $i = 1,\ldots,J(L+p)$, while $\bar{\mathbf{C}}_{\bar{\eta}\bar{\eta}_b} := \text{diag}(\mathbf{C}_{\bar{\eta}\bar{\eta}} + \mathbf{C}_{\bar{\eta}_b\bar{\eta}_b})$ and $\mathbf{C}_{\bar{\eta}\bar{\eta}_b} := \text{diag}(\mathbf{C}_{\bar{\eta}\bar{\eta}},\mathbf{C}_{\bar{\eta}_b\bar{\eta}_b})$.

Remark 4: The bounded covariance asserted by Proposition 4 has to be contrasted with the unbounded noise that is inherent to consensus averaging [23]. Notice also that Proposition 4 holds universally for general data models, allowing for and exploiting arbitrary correlation patterns among sensor data (not possible

with the RD-BLUE in [17]). Further, the noise-robust framework and the corresponding analysis carries over to the simplified D-LMMSE algorithm for linear-Gaussian models.

D. Simulations

Here we test the convergence of D-LMMSE and RD-LMMSE along with their noise resilience properties in the presence of either reception or quantization noise. Consider $J = 30$ sensors with observations obeying the linear model in Section III-B. Fig. 3 (top) depicts the normalized error $e_{\text{norm}}(k)$ versus iteration index k for different SNR values. The penalty coefficients for both D-LMMSE and RD-LMMSE are set to $c_j = d_j = 4$. Analytical guidelines for selecting c_j, d_j seem to be difficult to derive if not impossible without global information. However, we have observed that increasing c_j, d_j up to a point usually improves the convergence speed. Notice that under ideal channel links both D-LMMSE and RD-LMMSE iterates coincide as asserted by Lemma 4, and $e_{\text{norm}}(k) \to 0$ as $k \to \infty$ as per Proposition 2. In the presence of reception noise, we average $e_{\text{norm}}(k)$ over 50 independent D-LMMSE and RD-LMMSE estimates. As expected by Proposition 4, $e_{\text{norm}}(k)$ obtained from RD-LMMSE exhibits error floor confirming that the noise covariance converges to a matrix with bounded entries. Relative to RD-LMMSE, the D-LMMSE algorithm exhibits noise resilience at the expense of higher steady-state variance. Clearly, as the SNR increases the steady-state error for both D-LMMSE and RD-LMMSE decreases. RD-LMMSE and D-LMMSE exhibit the same behavior also in Fig. 3 which depicts the ensemble average of $e_{\text{norm}}(k)$ over 50 independent realizations, for a variable number of quantization bits (common across all sensors).

VI. Optimal Distributed Kalman Filtering and Smoothing

In this section we consider distributed estimation and smoothing under the dynamical setup **s3)**. For future use, recall the information form of the correction step of the centralized KF for obtaining $\mathbf{C}(t|t)$ and $\hat{\mathbf{s}}(t|t)$, see e.g., [3, pp. 40 and 139]

$$\hat{\mathbf{s}}(t|t-1) = \mathbf{\Phi}(t-1)\hat{\mathbf{s}}(t-1|t-1) \tag{41}$$

$$\mathbf{C}(t|t-1) = \mathbf{\Phi}(t-1)\mathbf{C}(t-1|t-1)\mathbf{\Phi}^T(t-1) + \mathbf{Q}(t-1) \tag{42}$$

$$\mathbf{C}(t|t) = \left[\mathbf{H}^T(t)\mathbf{R}^{-1}(t)\mathbf{H}(t) + \mathbf{C}^{-1}(t|t-1)\right]^{-1} \tag{43}$$

$$\hat{\mathbf{s}}(t|t) = \hat{\mathbf{s}}(t|t-1) + \mathbf{C}(t|t)\mathbf{H}^T(t)\mathbf{R}^{-1}(t) \times \left[\mathbf{x}(t) - \mathbf{H}(t)\hat{\mathbf{s}}(t|t-1)\right] \tag{44}$$

where $\mathbf{C}(t|t)$ denotes the filtered covariance matrix of the estimation error $\mathbf{s}(t) - \hat{\mathbf{s}}(t|t)$ and likewise for the predicted covariance $\mathbf{C}(t|t-1)$, while $\mathbf{H}(t) := [\mathbf{h}_1(t) \ldots \mathbf{h}_J(t)]^T$. 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 (41) and (42) in a distributed fashion since $\mathbf{\Phi}(t-1)$ and $\mathbf{Q}(t-1)$ are assumed locally known. However, (43) and (44) can be run only if quantities

$$\mathcal{I}(t) := \mathbf{H}^T(t)\mathbf{R}^{-1}(t)\mathbf{H}(t) = \sum_{j=1}^{J} \sigma_{n_j}^{-2}(t)\mathbf{h}_j(t)\mathbf{h}_j^T(t) \tag{45}$$

$$\boldsymbol{\chi}(t) := \mathbf{H}^T(t)\mathbf{R}^{-1}(t)\mathbf{x}(t) = \sum_{j=1}^{J} \sigma_{n_j}^{-2}(t)\mathbf{h}_j(t)x_j(t) \tag{46}$$

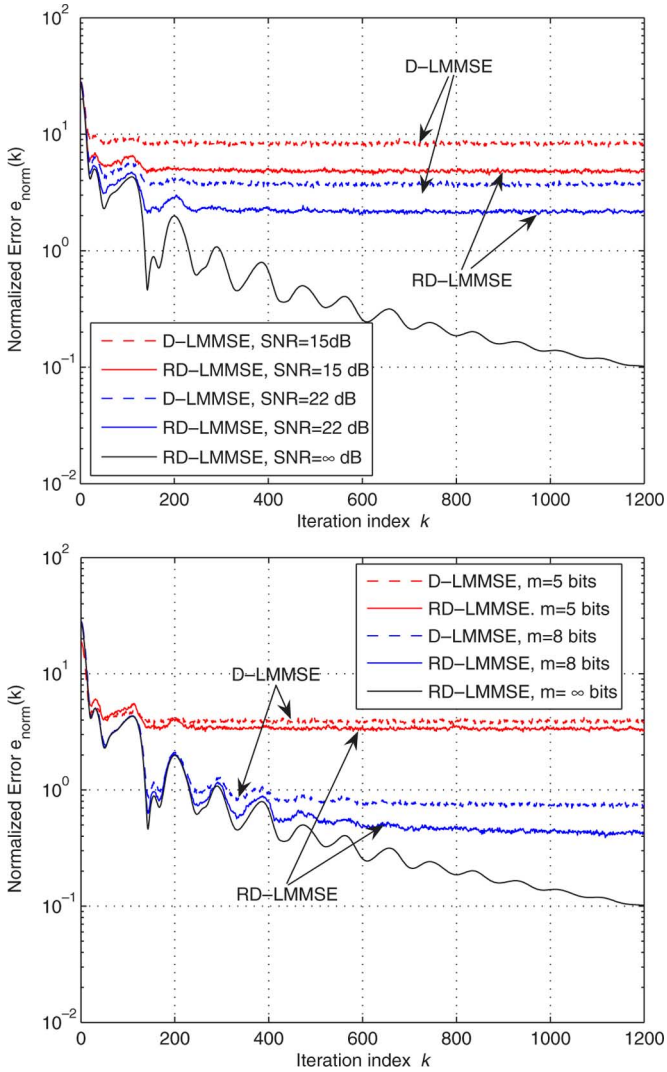

Fig. 3. Normalized error $e_{\text{norm}}(k)$ versus iteration index k for D-LMMSE and RD-LMMSE in the presence of (top) reception noise with SNR $= 15$, 22, and ∞ dB, and (bottom) quantization noise using $m = 5$, 10, and ∞ number of bits.

could be somehow estimated at each sensor j. This is possible because as the last equalities in (45) and (46) show, $\mathcal{I}(t)$ and $\boldsymbol{\chi}(t)$ can be expressed as averages with the jth summand available at sensor j.

Through $K+1$ iterations that start at t and end at $t+K$, [1], [19] proposed (in our notation) to form estimates $\hat{\mathcal{I}}(t; t:t+K)$ and $\hat{\boldsymbol{\chi}}(t; t:t+K)$ using the consensus averaging based algorithms in [15], [20], and [24], respectively; see also [13], where $K = 0$ was adopted since only one iteration can be afforded during the interval $[t, t+1)$. With these estimates plugged into recursions (43) and (44), 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], [19] only to applications with slow varying $\mathbf{s}(t)$ and/or fast communications needed to complete $K \gg$ consensus steps between t and $t+1$. In addition, [1] and [19] inherit the noise sensitivity of [20] and [24]. More important, the estimates $\hat{\mathbf{s}}_j(t|t; t:t+K)$ in [1], [13], and [19] are not MSE optimal given the available information in $\hat{\boldsymbol{\chi}}(t; t:t+K)$, unless $K \to \infty$. This suboptimality renders the D-KF estimates in [1],

[13], [19] inconsistent with the underlying data model, which in turn is known to yield tracking errors violating the 3σ uncertainty bounds [4, p. 233]. In the distributed approach of [2] each sensor performs local KF after substituting its local state predictions in (41) and (44) with a weighted average of the neighboring sensors predictions. Though, each sensor should also update appropriately the covariance matrices of the resulting prediction error. However, this update requires global statistical information from all other sensors (cf. [2, eq. (11) and (15)]).

A. Smoothing Versus Filtering

Instead of filtering advocated by [1], [2], [13], and [19], 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 MSE optimal* smoothed estimates, $\hat{\mathbf{s}}_j(t|t+i;t+i:t+K)$ for $i=0,\ldots,K$, 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 $\{\hat{\mathbf{s}}_j(t+i-j|t+i;t+i:t+K)\}_{j=0}^{K}$ which are not available in the alternatives [1], [2], [13], [19].

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. (3)]

$$\breve{\mathbf{s}}(t)=\begin{bmatrix}\boldsymbol{\Phi}(t-1) & \mathbf{0} & \ldots & \mathbf{0}\\ \mathbf{I} & \mathbf{0} & \ldots & \mathbf{0}\\ \vdots & \ldots & \ddots & \vdots\\ \mathbf{0} & \ldots & \mathbf{I} & \mathbf{0}\end{bmatrix}\breve{\mathbf{s}}(t-1)+\begin{bmatrix}\mathbf{w}(t-1)\\ \mathbf{0}\\ \vdots\\ \mathbf{0}\end{bmatrix}$$
$$:=\breve{\boldsymbol{\Phi}}(t-1)\breve{\mathbf{s}}(t-1)+\breve{\mathbf{w}}(t-1) \tag{47}$$

where $\breve{\mathbf{s}}(t):=[\mathbf{s}^T(t)\ldots\mathbf{s}^T(t-K)]^T$. The aggregate observations $\mathbf{x}(t):=[x_1(t)\ldots x_J(t)]^T$ obey

$$\mathbf{x}(t)=\breve{\mathbf{H}}(t)\breve{\mathbf{s}}(t)+\mathbf{n}(t) \tag{48}$$

where $\breve{\mathbf{h}}(t):=[\mathbf{h}(t),\mathbf{0}\ldots,\mathbf{0}]$ and $\mathbf{n}(t):=[n_1(t)\ldots n_J(t)]^T$ has covariance matrix $\mathbf{r}(t):=\mathrm{diag}(\sigma^2_{n_1}(t)\ldots,\sigma^2_{n_J}(t))$. Note that this state augmentation guarantees that the augmented noise $\breve{\mathbf{w}}(t-1)$ is uncorrelated across time. The latter readily implies that fixed-lag centralized KS can be implemented equivalently as a centralized KF on the augmented model in (47) and (48), as follows (cf. (41)–(44) and [3, p. 177]):

$$\breve{\mathbf{C}}(t|t-1)=\breve{\boldsymbol{\Phi}}(t-1)\breve{\mathbf{C}}(t-1|t-1)\breve{\boldsymbol{\Phi}}^T(t-1)$$
$$+\mathrm{diag}\,(\mathbf{Q}(t-1),\mathbf{0},\ldots,\mathbf{0}) \tag{49}$$
$$\hat{\breve{\mathbf{s}}}(t|t-1)=\breve{\boldsymbol{\Phi}}(t-1)\hat{\breve{\mathbf{s}}}(t-1|t-1) \tag{50}$$
$$\breve{\mathbf{C}}(t|t)=\left[\breve{\mathbf{H}}^T(t)\mathbf{R}^{-1}(t)\breve{\mathbf{H}}(t)+\breve{\mathbf{C}}^{-1}(t|t-1)\right]^{-1} \tag{51}$$
$$\hat{\breve{\mathbf{s}}}(t|t)=\hat{\breve{\mathbf{s}}}(t|t-1)+\breve{\mathbf{C}}(t|t)\breve{\mathbf{H}}^T(t)\mathbf{R}^{-1}(t)$$
$$\times\left[\mathbf{x}(t)-\breve{\mathbf{H}}(t)\hat{\breve{\mathbf{s}}}(t|t-1)\right] \tag{52}$$

where $\hat{\breve{\mathbf{s}}}(t|t-1)$ and $\hat{\breve{\mathbf{s}}}(t|t)$ are the predicted and filtered estimates of $\breve{\mathbf{s}}(t)$, while $\breve{\mathbf{C}}(t|t-1)$ and $\breve{\mathbf{C}}(t|t)$ denote the covariance matrices for the corresponding state estimation and prediction errors.

Since $\hat{\breve{\mathbf{s}}}(t+K|t+K):=[\hat{\mathbf{s}}^T(t+K|t+K)\ldots\hat{\mathbf{s}}^T(t|t+K)]^T$, the KF estimate for the augmented state $\breve{\mathbf{s}}(t+K)$ contains both a filtered estimate of the original state $\mathbf{s}(t+K)$ as well as smoothed estimates of $\mathbf{s}(t+K-k)$, for $k=1,\ldots,K$, using all the available data up to $t+K$.

B. The D-KS Algorithm

The second summand in (52) is the LMMSE estimator $E[\breve{\mathbf{s}}(t)|\tilde{\mathbf{x}}(t)]$ based on the innovations $\tilde{\mathbf{x}}(t):=\mathbf{x}(t)-\breve{\mathbf{H}}(t)\hat{\breve{\mathbf{s}}}(t|t-1)$. This estimator could be formed in a distributed fashion using the D-LMMSE estimator developed in Section IV-B. However, if D-(or RD-) LMMSE is run for a finite number of consensus steps K, to track a fast varying $\mathbf{s}(t)$, sensors can only use *local* estimates $E_j[\breve{\mathbf{s}}(t)|\tilde{\mathbf{x}}(t)]$ which are suboptimum because data becoming available during $(t,t+K)$ are not exploited. Next, we develop a distributed algorithm that guarantees any-time MSE optimality under ideal links, while being robust when noise is present.

To derive such an algorithm we will need local estimates of $\breve{\boldsymbol{\chi}}(t):=\breve{\mathbf{H}}^T(t)\mathbf{r}^{-1}(t)\mathbf{x}(t)$, $\breve{\boldsymbol{\mathcal{I}}}(t):=\breve{\mathbf{H}}^T(t)\mathbf{r}^{-1}(t)\breve{\mathbf{H}}(t)$ and the covariance $\mathbf{C}(t|t;t:t)$ [cf. (51) and (52)]. But since $\breve{\boldsymbol{\mathcal{I}}}(t):=\mathrm{diag}(\boldsymbol{\mathcal{I}}(t),\mathbf{0},\ldots,\mathbf{0})$ and $\breve{\boldsymbol{\chi}}(t):=[\boldsymbol{\chi}^T(t),\mathbf{0},\ldots,\mathbf{0}]^T$, it suffices to devise distributed estimators of $\boldsymbol{\mathcal{I}}(t)$ and $\boldsymbol{\chi}(t)$. Towards this objective, let us re-express vector $\boldsymbol{\chi}(t)$ in (46) as

$$\boldsymbol{\chi}(t):=\arg\min_{\boldsymbol{\chi}}\sum_{j=1}^{J}\left\|\boldsymbol{\chi}-J\mathbf{h}_j(t)\sigma^{-2}_{n_j}(t)x_j(t)\right\|_2^2 \tag{53}$$

where the term $J\mathbf{h}_j(t)\sigma^{-2}_{n_j}(t)x_j(t)$ is locally available at sensor j. Matrix $\boldsymbol{\mathcal{I}}(t)$ in (45) can be rewritten likewise. Then, we can readily utilize the alternating-direction method of multipliers to form as in Section V iterates whose successive differences yield estimates $\hat{\boldsymbol{\mathcal{I}}}_j(t;t:t+k)$ and $\hat{\boldsymbol{\chi}}_j(t;t:t+k)$ which converge to $\boldsymbol{\mathcal{I}}(t)$ and $\boldsymbol{\chi}(t)$ respectively as $k\to\infty$ under ideal channel links. Indeed, upon modifying (27) and (28) by setting $L=0$ and substituting $\mathbf{F}_j^{-1}(\mathbf{I}_p-\mathbf{G}_j)=(2+c_j|\mathcal{B}_j|)^{-1}\mathbf{I}_p$, $\mathbf{D}_j^1=c_j|\mathcal{B}_j|\mathbf{I}_p$, $\mathbf{D}_j^2=c_j\mathbf{I}_p$ and $\mathbf{D}_b=(\sum_{\beta\in\mathcal{N}_b}c_\beta)^{-1}\mathbf{I}_p$, the estimate $\hat{\boldsymbol{\chi}}_j(t;t:t+k)$ can be formed locally at the jth sensor after setting $\hat{\mathbf{x}}_j=2J(2+c_j|\mathcal{B}_j|)^{-1}\mathbf{h}_j(t)\sigma^{-2}_{n_j}(t)x_j(t)$ in (27), and computing the differences of the successive iterates produced after k steps by the modified pair of recursions (27), (28). Likewise, $\hat{\boldsymbol{\mathcal{I}}}_j(t;t:t+k)$ can be obtained after replacing the vector iterates in (27) and (28) with $p\times p$ matrices ($\boldsymbol{\phi}_j(k)$'s are in this case matrices), while setting $\hat{\mathbf{x}}_j\equiv 2J(2+c_j|\mathcal{B}_j|)^{-1}\mathbf{h}_j(t)\sigma^{-2}_{n_j}(t)\mathbf{h}_j^T(t)$.

Following the steps used to derive the RD-LMMSE global recursion in (38), we can write the local recursions for $\hat{\boldsymbol{\chi}}_j(t;t:t+k)$ and $\hat{\boldsymbol{\mathcal{I}}}_j(t;t:t+k)$ for $k=1,\ldots,K$, in compact form as

$$\hat{\boldsymbol{\chi}}_j(t;t:t+k)$$
$$=\mathbf{A}_j(k)\left[\hat{\boldsymbol{\chi}}^T(t;t:t+k-1),\hat{\boldsymbol{\chi}}^T(t;t:t+k-2)\right]^T \tag{54}$$
$$\hat{\boldsymbol{\mathcal{I}}}_j(t;t:t+k)$$
$$=\mathbf{A}_j(k)\left[\hat{\boldsymbol{\mathcal{I}}}^T(t;t:t+k-1),\hat{\boldsymbol{\mathcal{I}}}^T(t;t:t+k-2)\right]^T \tag{55}$$

where $\hat{\boldsymbol{\chi}}(t;t:t+k):=[\hat{\boldsymbol{\chi}}_1^T(t;t:t+k)\ldots\hat{\boldsymbol{\chi}}_J^T(t;t:t+k)]^T$ and $\hat{\boldsymbol{\mathcal{I}}}(t;t:t+k):=[\hat{\boldsymbol{\mathcal{I}}}_1^T(t;t:t+k)\ldots\hat{\boldsymbol{\mathcal{I}}}_J^T(t;t:t+k)]^T$,

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. The local estimates in (54) and (55) are initialized as follows:

$$\hat{\boldsymbol{\chi}}_j(t;t:t) = 2J\left(2 + c_j|\mathcal{B}_j|\right)^{-1}\mathbf{h}_j(t)\sigma_{n_j}^{-2}(t)x_j(t) \quad (56)$$

$$\hat{\boldsymbol{\mathcal{I}}}_j(t;t:t) = 2J\left(2 + c_j|\mathcal{B}_j|\right)^{-1}\mathbf{h}_j(t)\sigma_{n_j}^{-2}(t)\mathbf{h}_j^T(t) \quad (57)$$

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

Interestingly, it turns out that $\hat{\boldsymbol{\chi}}_j(t;t:t+k)$ and $\hat{\boldsymbol{\mathcal{I}}}_j(t;t:t+k)$ are linearly related. Specifically, we prove in Appendix E that:

Lemma 5: Under ideal channel links, $\hat{\boldsymbol{\chi}}_j(t;t:t+k)$ and $\hat{\boldsymbol{\mathcal{I}}}_j(t;t:t+k)$ are linearly related; i.e.,

$$\hat{\boldsymbol{\chi}}_j(t;t:t+k) = \hat{\boldsymbol{\mathcal{I}}}_j(t;t:t+k)\mathbf{s}(t) + \hat{\mathbf{n}}_j(t;t:t+k) \quad (58)$$

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

$$\boldsymbol{\mathcal{R}}_j(t;t:t+k) = \mathbf{A}_j(k)\boldsymbol{\mathcal{R}}(t;t:t+k-1)\mathbf{A}_j^T(k) \quad (59)$$

where $\boldsymbol{\mathcal{R}}(t;t:t+k)$ is the $2Jp \times 2Jp$ covariance matrix of the noise vector $[\hat{\mathbf{n}}_1^T(t;t:t+k),\ldots,\hat{\mathbf{n}}_J^T(t;t:t+k),\hat{\mathbf{n}}_1^T(t;t:t+k-1),\ldots,\hat{\mathbf{n}}_J^T(t;t:t+k-1)]^T$. The recursion in (59) is initialized by $\boldsymbol{\mathcal{R}}(t;t:t) = \mathrm{diag}(\boldsymbol{\mathcal{R}}_1(t;t:t),\ldots,\boldsymbol{\mathcal{R}}_J(t;t:t),\mathbf{0}_{Jp\times Jp})$, where $\boldsymbol{\mathcal{R}}_j(t;t:t) = (2J)^2(2 + c_j|\mathcal{B}_j|)^{-2}\sigma_{n_j}^{-2}(t)\mathbf{h}_j(t)\mathbf{h}_j^T(t)$. Note that $\boldsymbol{\mathcal{R}}_j(t;t:t+k)$ is the jth $p \times p$ diagonal block of $\boldsymbol{\mathcal{R}}(t;t:t+k)$.

In terms of the augmented state model, (58) can be expressed as

$$\hat{\boldsymbol{\chi}}_j(t;t:t+k) = \hat{\breve{\boldsymbol{\mathcal{I}}}}_j(t;t:t+k)\breve{\mathbf{s}}(t) + \hat{\mathbf{n}}_j(t;t:t+k) \quad (60)$$

where $\hat{\breve{\boldsymbol{\mathcal{I}}}}(t;t:t+k) := [\hat{\boldsymbol{\mathcal{I}}}(t;t:t+k),\mathbf{0}_{p\times p},\ldots,\mathbf{0}_{p\times p}]$. Our key idea is to view $\hat{\boldsymbol{\chi}}_j(t;t:t+k)$ in (60) 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 $\mathbf{x}_j(t)$, quantity $\hat{\boldsymbol{\chi}}_j(t;t:t+k)$ includes "consensus data" from neighboring sensors whose number increases as k increases. Since $\hat{\boldsymbol{\chi}}_j(t;t:t+k)$ contains more information than $\mathbf{x}_j(t)$, state estimates based on it will clearly exhibit improved performance.

What is left to derive is a distributed algorithm for computing $\boldsymbol{\mathcal{R}}_j(t;t:t+k)$, which we provide next (see Appendix F for the detailed derivation).

Lemma 6: Consider per sensor $j \in [1,J]$ the $p \times Jp$ matrix $\boldsymbol{\Psi}_j$ factor of $\boldsymbol{\mathcal{R}}_j$ obtained recursively as

$$\boldsymbol{\Psi}_j(t;t:t+k) = \mathbf{A}_j(k)\Big[\boldsymbol{\Psi}^T(t;t:t+k-1), \boldsymbol{\Psi}^T(t;t:t+k-2)\Big]^T \quad (61)$$

where $\boldsymbol{\Psi}^T(t;t:t+k) := [\boldsymbol{\Psi}_1^T(t;t:t+k),\ldots,\boldsymbol{\Psi}_J^T(t;t:t+k)]^T$ and (61) is initialized using

$$\boldsymbol{\Psi}_j(t;t:t) = \Big[\mathbf{0}_{p\times p}\ldots,\boldsymbol{\mathcal{R}}_j^{1/2}(t;t:t),\ldots,\mathbf{0}_p\Big] \quad (62)$$

$$\boldsymbol{\Psi}_j(t;t:t-1) = \mathbf{0}_{p\times Jp} \quad (63)$$

with $\boldsymbol{\mathcal{R}}_j^{1/2}(t;t:t)$ denoting the square root matrix of $\boldsymbol{\mathcal{R}}_j(t;t:t)$. Then, the modified noise covariance $\boldsymbol{\mathcal{R}}_j(t;t:t+k)$ can be obtained locally at sensor j as

$$\boldsymbol{\mathcal{R}}_j(t;t:t+k) = \boldsymbol{\Psi}_j(t;t:t+k)\boldsymbol{\Psi}_j^T(t;t:t+k). \quad (64)$$

Equations (61)–(64) express the local $\boldsymbol{\mathcal{R}}_j(t;t:t+k)$ recursively, thus allowing the D-KS algorithm to implement its iterations as summarized in the following proposition.

Proposition 5: Consider per sensor j the local augmented state estimates $\hat{\breve{\mathbf{s}}}_j(t+k|t+k;t+k:t+K)$, with $k = 0,1,\ldots,K$, obtained at time instant $t+K$ through the KS recursions

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

$$\breve{\mathbf{C}}_j(t+k-1|t+k-1;t+k-1:t+K) = \breve{\boldsymbol{\Phi}}^T(t+k-1)+\mathrm{diag}\left(\mathbf{Q}(t+k-1),\mathbf{0},\ldots,\mathbf{0}\right) \quad (65)$$

$$\hat{\breve{\mathbf{s}}}_j(t+k|t+k-1;t+k-1:t+K) = \breve{\boldsymbol{\Phi}}(t+k-1)\hat{\breve{\mathbf{s}}}_j(t+k-1|t+k-1;t+k-1:t+K) \quad (66)$$

$$\breve{\mathbf{C}}_j(t+k|t+k;t+k:t+K) = \Big[\hat{\breve{\boldsymbol{\mathcal{I}}}}_j^T(t+k;t+k:t+K)\boldsymbol{\mathcal{R}}_j^{-1}(t+k;t+k:t+K) \times \hat{\breve{\boldsymbol{\mathcal{I}}}}_j(t+k;t+k:t+K) + \breve{\mathbf{C}}_j^{-1}(t+k|t+k-1;t+k-1:t+K)\Big]^{-1} \quad (67)$$

$$\hat{\breve{\mathbf{s}}}_j(t+k|t+k;t+k:t+K) = \hat{\breve{\mathbf{s}}}_j(t+k|t+k-1;t+k-1:t+K) + \breve{\mathbf{C}}_j(t+k|t+k;t+k:t+K) \times \hat{\breve{\boldsymbol{\mathcal{I}}}}^T(t+k;t+k:t+K)\boldsymbol{\mathcal{R}}_j^{-1}(t+k;t+k:t+K) \times \Big[\hat{\boldsymbol{\chi}}_j(t+k;t+k:t+K) - \hat{\breve{\boldsymbol{\mathcal{I}}}}^T(t+k;t+k:t+K) \times \hat{\breve{\mathbf{s}}}_j(t+k|t+k-1;t+k-1:t+K)\Big] \quad (68)$$

corresponding to the state model in (47) and the observation model in (60). Then, the local estimates $\hat{\breve{\mathbf{s}}}_j(t+k|t+k;t+k:t+K)$ are MSE optimal in the sense that $\hat{\breve{\mathbf{s}}}_j(t+k|t+k;t+k:t+K) := E[\breve{\mathbf{s}}(t+k)|\{\hat{\boldsymbol{\chi}}_j(t';t':t'+K)\}_{t'=0}^{t},\{\hat{\boldsymbol{\chi}}_j(t';t':t+K)\}_{t'=t+1}^{t+k}]$. Further, as the number of local iterates $K \to \infty$ the local augmented state estimates converge to their centralized counterparts; i.e., $\forall\, j \in [1,J]$, $t = 0,1,\ldots$,

$$\lim_{K\to\infty}\hat{\breve{\mathbf{s}}}_j(t|t;t:t+K) = \hat{\breve{\mathbf{s}}}(t|t) \quad (69)$$

$$\lim_{K\to\infty}\breve{\mathbf{C}}_j(t|t;t:t+K) = \breve{\mathbf{C}}(t|t). \quad (70)$$

Proof: MSE optimality of $\hat{\breve{\mathbf{s}}}_j(t+k|t+k;t+k:t+K)$ holds because these estimates are derived by local KS recursions that adhere to the state model in (47) and (60). Thus, $\hat{\breve{\mathbf{s}}}_j(t+k|t+k;t+k:t+K)$ is the MMSE estimator of $\breve{\mathbf{s}}(t+k)$ at sensor j given the available data $\{\hat{\boldsymbol{\chi}}_j(t';t':t+K)\}_{t'=0}^{t+k}$ that contain "consensus enriched" information up to time $t+K$. Further,

recall that under ideal links $\lim_{K\to\infty}\hat{\breve{\boldsymbol{\mathcal{I}}}}_j(t;t:t+K)=\breve{\boldsymbol{\mathcal{I}}}(t)$ and $\lim_{K\to\infty}\hat{\boldsymbol{\chi}}_j(t;t:t+K)=\boldsymbol{\chi}(t)$ for $t=0,1,2,\ldots$. Thus, (65)–(68) coincide with (49)–(52) and convergence of $\hat{\breve{\mathbf{s}}}_j(t|t;t:t+K)$, $\breve{\mathbf{C}}_j(t|t;t:t+K)$ to $\hat{\breve{\mathbf{s}}}(t|t)$ and $\breve{\mathbf{C}}(t|t)$ follows readily. □

Over the interval $[t,t+K]$, D-KS produces a sequence of local MSE optimal state estimates $\hat{\breve{\mathbf{s}}}(t|t;t:t+k)$, for $k=0,1,\ldots,K$. Generally, the MSE associated with $\hat{\breve{\mathbf{s}}}(t|t;t:t+k)$, decreases as k increases since $\hat{\boldsymbol{\chi}}_j(t;t:t+k)$ improves with k. 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 a 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 state estimates $\hat{\breve{\mathbf{s}}}_j(t+k|t+k;t+k:t+K)$ with $k=0,1,\ldots,K$. Upon recalling that $\hat{\breve{\mathbf{s}}}_j(t+K|t+K;t+K:t+K):=[\hat{\mathbf{s}}_j^T(t+K|t+K;t+K:t+K),\ldots,\hat{\mathbf{s}}_j^T(t|t+K;t+K:t+K)]$, 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 $\{\hat{\mathbf{s}}_j^T(t+k|t+K;t+K:t+K)\}_{k=0}^{K-1}$, exploiting all the $\hat{\breve{\boldsymbol{\chi}}}_j$'s available up to time $t+K$.

Different from [2], D-KS consists only of local recursions through which each sensor obtains its local state estimate and corresponding covariance information by communicating only with its single-hop neighbors. The D-KS scheme is tabulated as Algorithm 2. The D-KS scheme is tabulated as Algorithm 2.

Remark 5: Besides being able to form filtered and smoothed state estimates, D-KS exhibits noise robustness and trades off delay for MSE reduction. Indeed, even though (58) holds only approximately in the presence of noise, $\hat{\boldsymbol{\chi}}_j$ and $\hat{\boldsymbol{\mathcal{I}}}_j$ are robust to noise because they are formed via differences as in (27) and (28).

Algorithm 2 D-KS Algorithm

For all $j\in[1,J]$ initialize $\breve{\mathbf{s}}_j(-1|-1)=\mathbf{0}$ and $\breve{\mathbf{C}}_j(-1|-1)=\mathbf{C}_{ss}(-1)$.

At time instant $t+K$

For all $j\in[1,J]$ and $k=0,\ldots,K$ form estimates $\hat{\boldsymbol{\chi}}_j(t+k;t+k:t+K)$ and $\hat{\boldsymbol{\mathcal{I}}}_j(t+k;t+k:t+K)$ using (54) and (55).

For all $j\in[1,J]$ and $k=0,\ldots,K$ form augmented state estimates $\hat{\breve{\mathbf{s}}}_j(t+k|t+k;t+k:t+K)$ using (65)–(68).

C. Numerical Examples

Here we test the D-KS algorithm in terms of estimation-error covariance, and compare it with [13], [19] and [2], abbreviated here SD-KF, OD-KF, and AD-KF, respectively. We consider $J=60$ sensors as in Section III-B and a scalar state process, with $\boldsymbol{\Phi}(k)=1$, $Q=4$, and initial conditions $E[s(-1)]=0$ and $\sigma_{ss}^2(-1)=1$. Sensor j acquires at time k the scalar observation $x_j(k)$ 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 $\breve{\boldsymbol{\mathcal{I}}}(t)$ and $\boldsymbol{\chi}(t)$ via (54) and (55) is set to $K=6$. The penalty coefficients are chosen as $c_j=4/|\mathcal{B}_j|$.

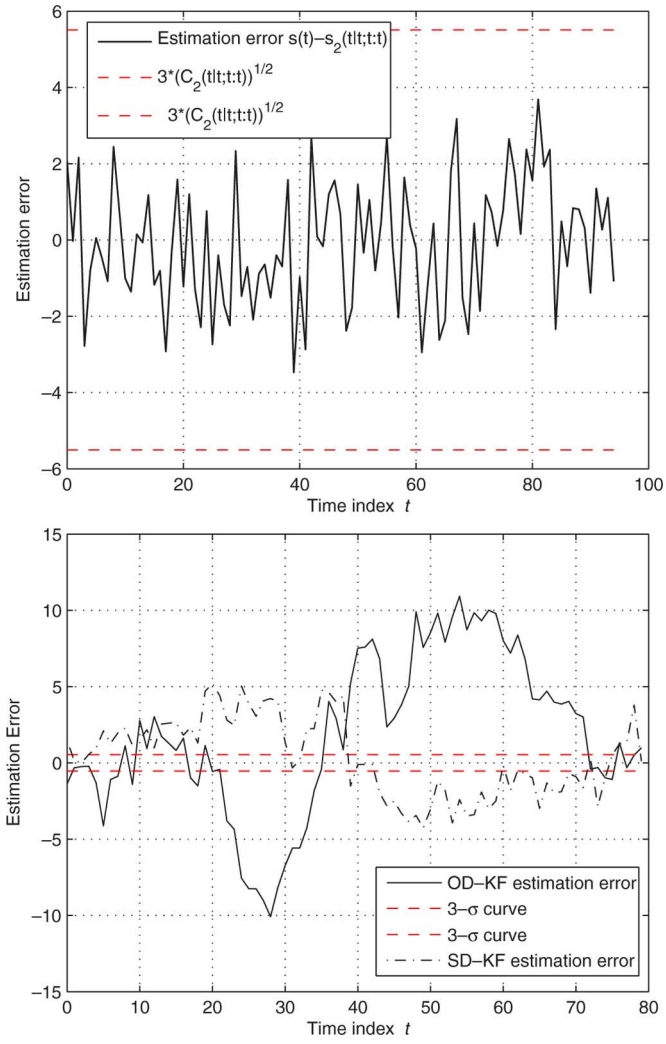

Fig. 4. The estimation error $s(t)-\hat{s}_2(t|t;t:t)$ is within the 3σ curves for D-KS (top); while the estimation error associated with SD-KF and OD-KF are not bounded by the corresponding 3σ curves (bottom).

Fig. 4 (top) depicts the estimation error $s(t)-\hat{s}_2(t|t;t:t)$ at sensor 2. Notice that this error falls within the $\pm3\sqrt{\mathbf{C}_j(t|t;t:t)}$ curves. This is reasonable since the local KS schemes are consistent with the data model of $\hat{\boldsymbol{\chi}}_j(t;t-k:t)$. However, this is not the case with SD-KF and OD-KF as can be seen in Fig. 4 (bottom) which being inconsistent with the underlying observation model, cannot provide MSE optimal estimates [4, p. 233].

Another 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 $k=0,\ldots,K$. This is to be contrasted with SD-KF that only provides an estimate for $s(t-K)$ after K iterations, without using the observations over the interval $[t-K,t]$. Fig. 5 (top) depicts the estimation error and corresponding 3σ 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. This can also be seen in Fig. 5 (bottom) where the MSE is plotted for various time delays. Observe that the MSE associated with the filtered (zero-delay) local state estimates $\hat{s}_j(t|t;t:t)$, is much smaller than the MSE attained when each sensor estimates $s(t)$ using only its own observations

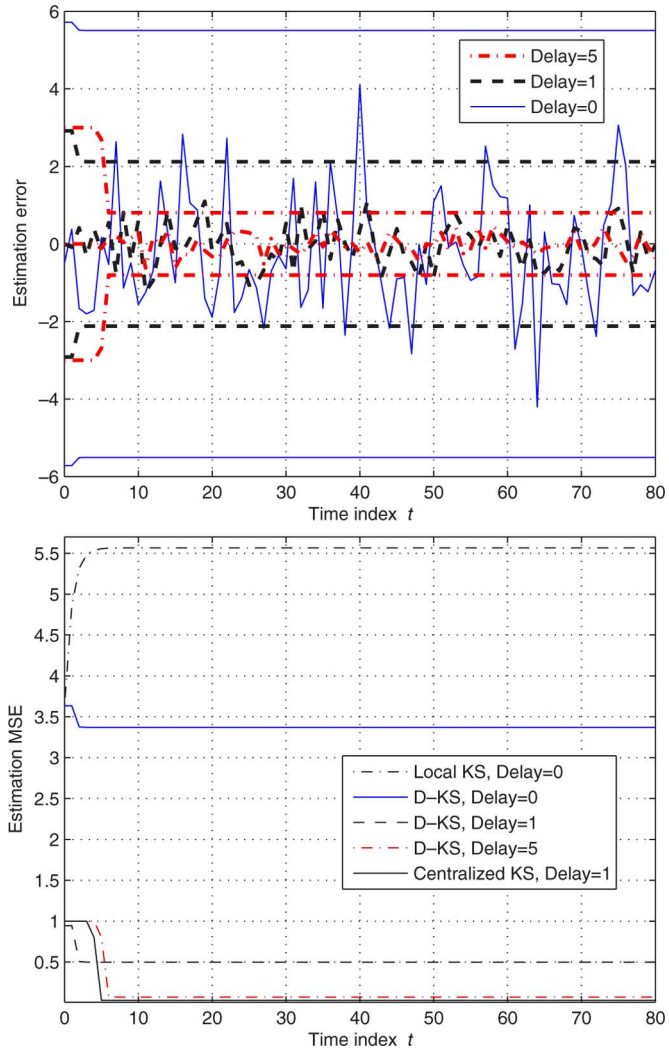

Fig. 5. Estimation error and 3σ curves versus t (top) and estimation MSE (bottom) for variable delays at sensor $j = 2$.

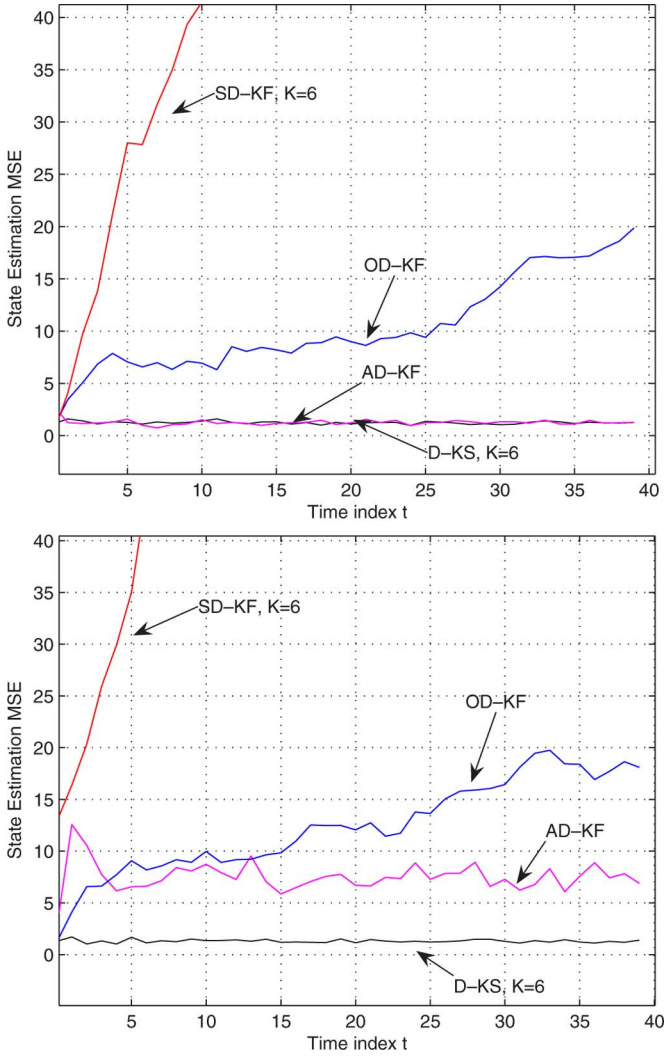

Fig. 6. Empirical estimation MSE versus time index t for D-KS, SD-KF, OD-KF and AD-KF under ideal channel links (top); and in presence of reception noise (bottom) at sensor $j = 2$

$x_j(t)$ (local filtering). Also, notice that the MSE is greatly reduced even when a one-step delay can be afforded. As the delay increases, the MSE associated with $\hat{s}_j(t-k|t;t:t)$ decreases and approaches that of the centralized KS.

We now examine the MSE achieved by the filtered estimates provided by SD-KF, OD-KF, AD-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. 6 depicts the MSE as a function of time under ideal communication links. 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. 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 $s(t)$. The AD-KF outperforms OD-KF and SD-KF, and its corresponding MSE converges close to the one achieved by D-KS. Though, recall that the corresponding covariance at every sensor in AD-KF requires global statistical information from all sensors in the network. In Fig. 6 (bottom) reception noise is present in the "refined" observations $\hat{\boldsymbol{\chi}}_j(t-k;t-k:t)$ for $k=0,1,\ldots,K$, and the weighted filtered state estimates in [2]. The noise-resilience properties of D-KS are apparent since the MSE converges to a finite value outperforming AD-KF, SD-KF and OD-KF, while the MSE for SD-KF and OD-KF grows unbounded.

VII. Concluding Remarks

We developed distributed algorithms for estimation of random signals using ad hoc WSNs based on successive refinement of local estimates. The essence of our approach is to express the desired estimator, either MAP or (L)MMSE, as the solution of pertinent convex optimization problems. We then relied on the alternating-direction method of multipliers to enable decentralized implementation. The decentralized estimators take into account all the available *a priori* information, while they allow for arbitrary correlations among sensor data. The novel framework does not require knowledge of the estimator in closed form, and leads to decentralized estimation algorithms that exhibit resilience to communication noise. Furthermore, the distributed LMMSE estimator is provably noise resilient. We also derived distributed estimators of non-stationary random signals. 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.

APPENDIX

A. Proof of Lemma 1

The cost in (11), call it $F_{\rm lmmse}(\mathbf{s},\mathbf{y})$, as well as the linear equality constraint $\mathbf{C}_{xx}\mathbf{y} = \mathbf{x}$ are convex. The optimal solution can thus be obtained by applying the first-order optimality conditions to the Lagrangian function $\mathcal{L}_{\rm lmmse}(\mathbf{s},\mathbf{y},\boldsymbol{\mu}) = F_{\rm lmmse}(\mathbf{s},\mathbf{y}) + \boldsymbol{\mu}^T(\mathbf{C}_{xx}\mathbf{y}-\mathbf{x})$, where $\boldsymbol{\mu}$ is the Lagrange multiplier corresponding to the equality constraint in (11). Specifically, the gradient of $\mathcal{L}_{\rm lmmse}(\mathbf{s},\mathbf{y},\boldsymbol{\mu})$ wrt $\mathbf{s}$ is $\nabla_{\mathbf{s}}\mathcal{L}_{\rm lmmse}(\mathbf{s},\mathbf{y},\boldsymbol{\mu}) = 2J\mathbf{s} - 2J\sum_{j=1}^{J}\mathbf{C}_{sx_j}(\mathbf{y})_j$. Setting $\nabla_{\mathbf{s}}\mathcal{L}_{\rm lmmse}(\mathbf{s},\mathbf{y},\boldsymbol{\mu}) = \mathbf{0}$, we obtain that the optimal $\mathbf{s}^*$ and $\mathbf{y}^*$ should satisfy $\mathbf{s}^* = \mathbf{C}_{sx}\mathbf{y}^*$. From the equality constraint in (11) and the invertibility of $\mathbf{C}_{xx}$ it follows that $\mathbf{y}^* = \mathbf{C}_{xx}^{-1}\mathbf{x}$, from which we deduce that $\mathbf{s}^* = \mathbf{C}_{sx}\mathbf{C}_{xx}^{-1}\mathbf{x} = \hat{\mathbf{s}}_{\rm lmmse}$. □

B. Proof of Proposition 2

We wish to show that (13)–(15) generate a series of local estimates converging to the optimal solution of (11), namely the LMMSE estimator, when $\mathcal{B}$ is a bridge sensor subset. We will establish this by showing that (13)–(15) correspond to the steps involved in the alternating-direction method of multipliers [6, pg. 255]. Toward this end, let $\mathbf{v}_j^b(k)$ and $\mathbf{w}_j^b(k)$ denote the Lagrange multipliers at the kth iteration. Moreover, define $\boldsymbol{\Gamma} := [\boldsymbol{\Gamma}_1^T \ldots \boldsymbol{\Gamma}_{|B|}^T]^T$ of size $(\sum_{b\in\mathcal{B}}|\mathcal{N}_b|)(L+p)\times J(L+p)$, where $\boldsymbol{\Gamma}_b := [\boldsymbol{\epsilon}_{b_1}^T \ldots \boldsymbol{\epsilon}_{b_{|\mathcal{N}_b|}}^T]^T \otimes \mathbf{I}_{L+p}$ and $\boldsymbol{\epsilon}_{b_1} \in \mathbb{R}^{J\times 1}$ denotes the vector with b_1th entry one and zero elsewhere, while $b_1 > b_2 > \ldots > b_{|\mathcal{N}_b|}$ are the indexes of the nonzero entries in the bth column of $\mathbf{E}$. Then, (12) can be equivalently written as

$$\{\hat{\mathbf{s}}_j,\hat{\mathbf{y}}_j\}_{j=1}^J = \arg\min G_1(\mathbf{s}_y) + G_2(\boldsymbol{\Gamma}\mathbf{s}_y),$$
$$\text{s.t. } \mathbf{C}_{xx}\mathbf{y}_j = \mathbf{x}_j,\ j=1,\ldots,J, \text{ and } \bar{\mathbf{s}}_y := \boldsymbol{\Gamma}\mathbf{s}_y \in \mathcal{C} \quad (71)$$

where $G_1(\mathbf{s}_y) := \sum_{j=1}^J \|\mathbf{s}_j - \bar{\mathbf{C}}_{sx_j}^T\mathbf{y}_j\|_2^2$, $G_2(\bar{\mathbf{s}}_y) = 0$, while $\mathbf{s}_y := [\mathbf{s}_{y,1}^T \ldots \mathbf{s}_{y,J}^T]^T$, and $\mathcal{C} \subset \mathbb{R}^{\sum_{b\in\mathcal{B}}|\mathcal{N}_b|(L+p)\times 1}$ is the polyhedral set defined so as for $\boldsymbol{\Gamma}_b\mathbf{s}_y := [(\bar{\mathbf{s}}_{y,b}^1)^T \ldots (\bar{\mathbf{s}}_{y,b}^{|\mathcal{N}_b|})^T]^T$ it holds that $\bar{\mathbf{s}}_{y,b}^1 = \ldots = (\bar{\mathbf{s}}_{y,b}^{|\mathcal{N}_b|}) = \bar{\mathbf{s}}_{y,b}$ for $b\in\mathcal{B}$. Inspection of (71) shows that it has the same form as the optimization problem in [6, p. 255, Eq. (4.76)]. Thus, the steps of the alternating-direction method of multipliers at $(k+1)$st iteration are as follows.

S1-a) Set $\mathbf{v} := \mathbf{v}(k) = \{\mathbf{v}_j^b(k)\}_{j\in[1,J]}^{b\in\mathcal{B}_j}$, and similarly fix $\mathbf{w} := \mathbf{w}(k)$, $\bar{\mathbf{s}} := \bar{\mathbf{s}}(k)$ and $\bar{\mathbf{y}} := \bar{\mathbf{y}}(k)$ to obtain $\boldsymbol{s}(k+1)$ and $\mathbf{y}(k+1)$ by solving the following minimization problem

$$[\boldsymbol{s}(k+1),\boldsymbol{y}(k+1)] = \arg\min_{\boldsymbol{s},\boldsymbol{y}} \mathcal{L}_a\left[\boldsymbol{s},\bar{\mathbf{s}}(k),\boldsymbol{y},\bar{\boldsymbol{\psi}}(k),\mathbf{v}(k),\mathbf{w}(k)\right]$$
$$\text{s.t. } \mathbf{C}_{x_j}\mathbf{y}_j = \mathbf{x}_j,\ j=1,\ldots,J \quad (72)$$

S1-b) For fixed $\mathbf{v} := \mathbf{v}(k) = \{\mathbf{v}_j^b(k)\}_{j\in[1,J]}^{b\in\mathcal{B}_j}$, and setting $\boldsymbol{s} = \boldsymbol{s}(k+1)$, $\boldsymbol{y} = \boldsymbol{y}(k+1)$ after completing step [S1-a], the consensus variables in $\bar{\mathbf{s}}(k+1)$ and $\bar{\mathbf{y}}(k+1)$ are obtained as

$$[\bar{\mathbf{s}}(k+1),\bar{\mathbf{y}}(k+1)] = \arg\min_{\bar{\mathbf{s}},\bar{\mathbf{y}}} \Bigg(-\sum_{b\in\mathcal{B}}\sum_{j\in\mathcal{N}_b}\left((\mathbf{v}_j^b(k))^T\bar{\mathbf{s}}_b + (\mathbf{w}_j^b(k))^T\bar{\mathbf{y}}_b\right) + \sum_{b\in\mathcal{B}}\sum_{j\in\mathcal{N}_b}\frac{c_j}{2}\|\mathbf{s}_j(k+1)-\bar{\mathbf{s}}_b\|_2^2 + \sum_{b\in\mathcal{B}}\sum_{j\in\mathcal{N}_b}\frac{d_j}{2}\|\mathbf{y}_j(k+1)-\bar{\mathbf{y}}_b\|_2^2 \Bigg). \quad (73)$$

S2) Update $\{\mathbf{v}_j^b(k),\mathbf{w}_j^b(k)\}_{j\in[1,J]}^{b\in\mathcal{B}_j}$ via (13).

Utilizing (13), we infer that (72) is equivalent to the following separate J subproblems:

$$[\mathbf{s}_j(k+1),\boldsymbol{y}_j(k+1)] = \arg\min_{\mathbf{s}_j,\boldsymbol{y}_j}\Bigg(\left\|\mathbf{s}_j - J\bar{\mathbf{C}}_{sx_j}^T\boldsymbol{y}_j\right\|_2^2 + \sum_{b\in\mathcal{B}_j}\left[(\mathbf{v}_j^b(k))^T\mathbf{s}_j + (\mathbf{w}_j^b(k))^T\boldsymbol{y}_j\right] \times \sum_{j\in\mathcal{B}_j}\left[\frac{c_j}{2}\|\mathbf{s}_j-\bar{\mathbf{s}}_b(k)\|_2^2 + \frac{d_j}{2}\|\mathbf{y}_j-\bar{\mathbf{y}}_b(k)\|_2^2\right]\Bigg),$$
$$\text{s.t. } \mathbf{C}_{xx}\boldsymbol{y}_j = \mathbf{x}_j,\ j=1,\ldots,J. \quad (74)$$

Similarly, $\bar{\mathbf{s}}_b(k+1)$ and $\bar{\mathbf{y}}_b(k+1)$ can be obtained by minimizing the cost function formed by keeping only the bth term of the sums in (73). Interestingly, each of the optimization problems in (74) can be solved locally at the corresponding sensor. After applying the KKT optimality conditions (see e.g., [7, p. 244]) to (74), we can readily obtain (14). Further, setting the gradient of the cost function formed by the bth sum terms in (73), wrt $\bar{\mathbf{s}}_{y,b}$, equal to zero we obtain (15). We have shown that the alternating direction method of multipliers applied to (12), boils down to (13)–(15). Since (12) is convex and $\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}$ is invertible, recursions (13)–(15) converge to the optimal solution $(\hat{\mathbf{s}}_{\rm lmmse},\hat{\mathbf{y}})$ [6, p. 257–260]. □

C. Proof of Lemma 3

With $\mathbf{v}_j^b(-1) = \mathbf{0}, \mathbf{w}_j^b(-1) = \mathbf{0}$, both (15) and (26) yield the same result for $k = -1$. Next, arguing by induction, suppose that $\mathbf{s}_{y,b}(n)$ is given by (26) $\forall\, n \le k$. After successive substitution of $\mathbf{v}_j^b(k)$ and $\mathbf{w}_j^b(k)$ into (13) and setting $\mathbf{v}_j^b(-1) = \mathbf{0}$ and $\mathbf{w}_j^b(-1) = \mathbf{0}$ we arrive at

$$\left[(\mathbf{v}_j^b(k+1))^T(\mathbf{w}_j^b(k+1))^T\right]^T = \mathbf{D}_j^2\sum_{n=1}^{k+1}(\mathbf{s}_{y,j}(n)-\bar{\mathbf{s}}_{y,b}(n)) \quad (75)$$

where $b\in\mathcal{B}_j$ and $j\in[1,J]$. Substituting (75) into (15), and using the induction step we can readily obtain (26). Further, substituting (75) into (14) and using (25), (26) follows after straightforward algebra. □

D. Proof of Equation (33)

Substituting (31) into (30), we obtain

$$\begin{aligned}\boldsymbol{\phi}_j(k+1)=&\hat{\mathbf{x}}_j+\left[\mathbf{I}_{L+p}-\mathbf{F}_j^{-1}(\mathbf{I}_{L+p}-\mathbf{G}_j)\mathbf{D}_j^1\right]\boldsymbol{\phi}_j(k)\\&+2\mathbf{F}_j^{-1}(\mathbf{I}_{L+p}-\mathbf{G}_j)\mathbf{D}_j^2\sum_{b\in\mathcal{B}_j}\sum_{j'\in\mathcal{N}_b}\mathbf{D}_{j'}^2\mathbf{D}_b\boldsymbol{\phi}_{j'}(k)\\&-\mathbf{F}_j^{-1}(\mathbf{I}_{L+p}-\mathbf{G}_j)\mathbf{D}_j^2\sum_{b\in\mathcal{B}_j}\sum_{j'\in\mathcal{N}_b}\mathbf{D}_{j'}^2\mathbf{D}_b\boldsymbol{\phi}_{j'}(k-1)\\&+\bar{\boldsymbol{\eta}}_j(k)+\bar{\boldsymbol{\eta}}_{b,j}(k).\end{aligned}\tag{76}$$

After stacking the first two terms in (76) we obtain the first two terms in (33) and the first term in $\mathbf{A}_1\boldsymbol{\phi}(k)$. Likewise, vectors $\bar{\boldsymbol{\eta}}(k)$ and $\bar{\boldsymbol{\eta}}_b(k)$ are obtained by stacking the noise terms in (76). Next, the second term in $\mathbf{A}_1\boldsymbol{\phi}(k)$ can be obtained after rewriting the sum of the third term in (76) as

$$\begin{aligned}\sum_{b\in\mathcal{B}_j}\sum_{j'\in\mathcal{N}_b}\mathbf{D}_{j'}^2\mathbf{D}_b\boldsymbol{\phi}_{j'}(k)=&\sum_{b\in\mathcal{B}}\mathbf{E}_{jb}\mathbf{I}_{L+p}(\mathbf{e}_b\otimes\mathbf{I}_{L+p})^T\\&\times\mathrm{diag}\left(\mathbf{D}_1^2\mathbf{D}_b,\ldots,\mathbf{D}_J^2\mathbf{D}_b\right)\boldsymbol{\phi}(k).\end{aligned}\tag{77}$$

Stacking the third term in (76) for $j=1,\ldots,J$ and using (77) we find the second term of $\mathbf{A}_1\boldsymbol{\phi}(k)$ in (33). Vector $\mathbf{A}_2\boldsymbol{\phi}(k)$ can be obtained similarly after stacking the fourth term in (76). □

E. Proof of Proposition 4

Matrix $\mathbf{A}$ satisfies the properties given in the following lemma:

Lemma 7: The eigenvalues $\{\lambda_{A,i}\}_{i=1}^{2J(L+p)}$ of $\mathbf{A}$ ordered so that $|\lambda_{A,1}|\geq\ldots\geq|\lambda_{A,2J(L+p)}|$ and the corresponding right and left eigenvectors $\mathbf{u}_{A,i}$ and $\mathbf{v}_{A,i}$ satisfy the following.

a) Eigenvalues $\lambda_{A,i}=1$ for $i=1,\ldots,L+p$; while $|\lambda_{A,i}|<1$ for $i\in[L+p+1,2J(p+L)]$.
b) The $L+p$ dominant left eigenvectors $\mathbf{v}_{A,i}:=[(\mathbf{v}_{A,i}^1)^T,(\mathbf{v}_{A,i}^2)^T]^T$ are given by

$$\left(\mathbf{v}_{A,i}^1\right)^T=\boldsymbol{\epsilon}_i^T[\mathbf{Y}_1\mathbf{F}_1\ldots\mathbf{Y}_J\mathbf{F}_J]\in\mathbb{R}^{J(L+p)\times1},\tag{78}$$

$$\left(\mathbf{v}_{A,i}^2\right)^T=-\left(\mathbf{v}_{A,i}^1\right)^T\mathbf{A}_2,\quad i=1,\ldots,L+p\tag{79}$$

where $\mathbf{Y}_j=\mathbf{P}+[\mathbf{C}_{sx}^T,\mathbf{I}_L]^T[\mathbf{C}_{xx}^{-1}]_j\mathbf{C}_{x_j}\times[[\mathbf{F}_j^{-1}]_{12},[\mathbf{F}_j^{-1}]_{22}]$ with $[\mathbf{C}_{xx}^{-1}]_j$ containing the columns of $\mathbf{C}_{xx}^{-1}$ with index from $\sum_{j'=1}^{j-1}L_{j'}+1$ to $\sum_{j'=1}^{j}L_{j'}$, $\mathbf{P}:=(2J)^{-1}\mathrm{diag}(\mathbf{I}_{p\times p},\mathbf{0}_{L\times L})$, and $[[\mathbf{F}_j^{-1}]_{12},[\mathbf{F}_j^{-1}]_{22}]$ denoting the $L\times(L+p)$ lower submatrix of $\mathbf{F}_j^{-1}$.
c) The $L+p$ dominant right eigenvectors $\{\mathbf{u}_{A,i}\}_{i=1}^{L+p}$ are $\mathbf{u}_{A,i}=[\boldsymbol{\epsilon}_i^T\ldots\boldsymbol{\epsilon}_i^T]^T\in\mathbb{R}^{2J(L+p)\times1}$, where $\boldsymbol{\epsilon}_i\in\mathbb{R}^{(L+p)\times1}$ denotes the vector with its ith entry one and zeros elsewhere.

Proof: Let $(\lambda,\mathbf{v})$ be an eigenvalue and the corresponding left eigenvector of $\mathbf{A}$. Upon defining $\mathbf{v}:=[\mathbf{v}_1^T\ \mathbf{v}_2^T]^T$ with $\mathbf{v}_1,\mathbf{v}_2\in\mathbb{C}^{J(L+p)\times1}$, and using that $\mathbf{v}^T\mathbf{A}=\lambda\mathbf{v}^T$ we find $\mathbf{v}_1^T(-\lambda^2\mathbf{I}+\lambda(\mathbf{I}-\mathbf{A}_1)-\mathbf{A}_2)=\mathbf{0}$, from which after multiplying both sides from the right with $\mathbf{F}_G\mathbf{v}_1^*$, (* denotes conjugation), we deduce that $\mathbf{v}_1^{\mathcal{H}}[-\lambda^2\mathbf{F}_G+\lambda(\mathbf{F}_G-\mathbf{A}_1\mathbf{F}_G)-\mathbf{A}_2\mathbf{F}_G]\mathbf{v}_1=0$, where $\mathcal{H}$ denotes conjugate transposition. Letting $\tilde{\mathbf{v}}_1^T:=\mathbf{v}_1^T\mathrm{diag}(\mathbf{F}_1^{-1},\ldots,\mathbf{F}_J^{-1})=[\tilde{\mathbf{v}}_{1,1}^T\ldots\tilde{\mathbf{v}}_{1,J}^T]$ and accounting for the structure of $\mathbf{A}_1$ and $\mathbf{A}_2$ we arrive at

$$\begin{aligned}&-\left(\sum_{j=1}^{J}\tilde{\mathbf{v}}_{1,j}^{\mathcal{H}}(\mathbf{I}-\mathbf{G}_j)\mathbf{F}_j\tilde{\mathbf{v}}_{1,j}\right)\lambda^2\\&+\left[\sum_{j=1}^{J}\tilde{\mathbf{v}}_{1,j}^{\mathcal{H}}\left(\mathbf{F}_j-\mathbf{G}_j\mathbf{F}_j-(\mathbf{I}-\mathbf{G}_j)\mathbf{D}_j^1\left(\mathbf{I}-\mathbf{G}_j^T\right)\right)\tilde{\mathbf{v}}_{1,j}\right.\\&\qquad\left.+2\sum_{b\in\mathcal{B}}\left\|\mathbf{D}_b^{1/2}\sum_{j'\in\mathcal{N}_b}\mathbf{D}_{j'}^2\left(\mathbf{I}-\mathbf{G}_{j'}^T\right)\tilde{\mathbf{v}}_{1,j'}\right\|_2^2\right]\lambda\\&\qquad-\sum_{b\in\mathcal{B}}\left\|\mathbf{D}_b^{1/2}\sum_{j'\in\mathcal{N}_b}\mathbf{D}_{j'}^2\left(\mathbf{I}-\mathbf{G}_{j'}^T\right)\tilde{\mathbf{v}}_{1,j'}\right\|_2^2=0.\end{aligned}\tag{80}$$

The roots of the second-order polynomial in (80) are given by $\lambda=\alpha_1\pm\sqrt{\alpha_1^2-\alpha_2}$, where

$$\begin{aligned}\alpha_1=&\frac{1}{2}\left(\sum_{j=1}^{J}\tilde{\mathbf{v}}_{1,j}^{\mathcal{H}}(\mathbf{I}-\mathbf{G}_j)\mathbf{F}_j\tilde{\mathbf{v}}_{1,j}\right)^{-1}\\&\times\left(\sum_{j=1}^{J}\tilde{\mathbf{v}}_{1,j}^{\mathcal{H}}\left(\mathbf{F}_j-\mathbf{G}_j\mathbf{F}_j-(\mathbf{I}-\mathbf{G}_j)\mathbf{D}_j^1\left(\mathbf{I}-\mathbf{G}_j^T\right)\right)\tilde{\mathbf{v}}_{1,j}\right.\\&\qquad\left.+2\sum_{b\in\mathcal{B}}\left\|\mathbf{D}_b^{1/2}\sum_{j'\in\mathcal{N}_b}\mathbf{D}_{j'}^2\left(\mathbf{I}-\mathbf{G}_{j'}^T\right)\tilde{\mathbf{v}}_{1,j'}\right\|_2^2\right)\end{aligned}\tag{81}$$

$$\begin{aligned}\alpha_2=&\left(\sum_{j=1}^{J}\tilde{\mathbf{v}}_{1,j}^{\mathcal{H}}(\mathbf{I}-\mathbf{G}_j)\mathbf{F}_j\tilde{\mathbf{v}}_{1,j}\right)^{-1}\\&\times\sum_{b\in\mathcal{B}}\left\|\mathbf{D}_b^{1/2}\sum_{j'\in\mathcal{N}_b}\mathbf{D}_{j'}^2\left(\mathbf{I}-\mathbf{G}_{j'}^T\right)\tilde{\mathbf{v}}_{1,j'}\right\|_2^2.\end{aligned}\tag{82}$$

Using (81) and (82) we will show that $|\lambda_{1,2}|\leq1$. If $\alpha_1^2<\alpha_2$ then $\lambda_{1,2}$ is complex. Applying the Cauchy–Schwartz inequality to (82) we obtain

$$\begin{aligned}\sum_{b\in\mathcal{B}}\left\|\mathbf{D}_b^{1/2}\sum_{j'\in\mathcal{N}_b}\mathbf{D}_{j'}^2\left(\mathbf{I}-\mathbf{G}_{j'}^T\right)\tilde{\mathbf{v}}_{1,j'}\right\|_2^2&\\\leq\sum_{j=1}^{J}\tilde{\mathbf{v}}_{1,j}^{\mathcal{H}}(\mathbf{I}-\mathbf{G}_j)\mathbf{D}_j^1\left(\mathbf{I}-\mathbf{G}_j^T\right)\tilde{\mathbf{v}}_{1,j}.&\end{aligned}\tag{83}$$

Further, utilizing the Schur complement (see e.g., [11, pg. 472]) it can be shown that $\mathbf{F}_j-\mathbf{G}_j\mathbf{F}_j-(\mathbf{I}-\mathbf{G}_j)\mathbf{D}_j^1(\mathbf{I}-\mathbf{G}_j^T)$ is positive semi-definite. Combining the latter with (83) we infer that $|\lambda_{1,2}|^2=\alpha_2\leq1$. If $\alpha_1^2>\alpha_2$, there are two cases of interest. In the first case $\lambda_{1,2}=\alpha_1-\sqrt{\alpha_1^2-\alpha_2}$, and since (83) implies that $\alpha_1\leq1$, we immediately have $\lambda_{1,2}<1$. Next, we consider $\lambda_{1,2}=\alpha_1+\sqrt{\alpha_1^2-\alpha_2}$ and define

$$\begin{aligned}\alpha_3:=&\left(\sum_{j=1}^{J}\tilde{\mathbf{v}}_{1,j}^{\mathcal{H}}(\mathbf{I}-\mathbf{G}_j)\mathbf{F}_j\tilde{\mathbf{v}}_{1,j}\right)^{-1}\\&\times\sum_{j=1}^{J}\tilde{\mathbf{v}}_{1,j}^{\mathcal{H}}\left(\mathbf{F}_j-\mathbf{G}_j\mathbf{F}_j-(\mathbf{I}-\mathbf{G}_j)\mathbf{D}_j^1\left(\mathbf{I}-\mathbf{G}_j^T\right)\right)\tilde{\mathbf{v}}_{1,j}.\end{aligned}\tag{84}$$

Since $\lambda_{1,2} = 1 - (\alpha_3/2) + ((\alpha_3^2/4) + \alpha_2(\alpha_2 + \alpha_3 - 1))^{1/2}$, we can easily infer from (83) that $\alpha_2 + \alpha_3 \leq 1$ and because $\alpha_2 \geq 0$ we have $\lambda_{1,2} \leq 1$. Thus, all the eigenvalues of $\mathbf{A}$ should have magnitude at most one. To prove that the geometric multiplicity of $\lambda = 1$ is $L + p$, we must show that for $\lambda = 1$ the matrix $\mathbf{T}(\lambda) := -\lambda^2\mathbf{I} + \lambda(\mathbf{A}_1 - \mathbf{A}_2) - \mathbf{A}_2$ has nullity $\nu(\mathbf{T}(\lambda)) = L + p$. To this end, notice that $\nu(\mathbf{T}(\lambda)) \leq \nu(\mathbf{T}(\lambda)\mathbf{F}_G)$. Specifically, using the structure of $\{\mathbf{G}_j\}_{j=1}^J$ we can readily show that $\nu(\mathbf{F}_G) = L$, while $\mathbf{F}_G\mathbf{v}_{A,i}^1 = \mathbf{0}$, for $\mathbf{v}_{A,i}^1$ as in (78) for $i = p+1, \ldots, p+L$. If $\mathbf{v}_1 \notin \text{null}(\mathbf{F}_G)$, then (80) gives the $\lambda_{1,2} = 1$ iff (83) is tight. This holds true iff $\mathbf{F}_G\mathbf{v}_1 = [\boldsymbol{\epsilon}_i^T \ldots \boldsymbol{\epsilon}_i^T]^T \in \mathbf{C}^{J(L+p)\times 1}$ for $i \in [1, L+p]$. It can be verified after some algebra that the latter is true iff $\mathbf{v}_1 = \mathbf{v}_{A,i}^1$ for $i = 1, \ldots, p$, with $\mathbf{v}_{A,i}^1$ as in (78). Further, holds that $\mathbf{v}_{A,i}^T\mathbf{A} = \mathbf{v}_{A,i}^T$ with $i = 1, \ldots, L+p$ which implies that $\mathbf{v}_{A,i}$ are left eigenvectors of $\mathbf{A}$ associated with the eigenvalue 1. Moreover, $\mathbf{v}_{A,i}$ are linearly independent. Thus, $\nu(\mathbf{T}(1)\mathbf{F}_G) = L + p$, and since the $L + p$ vectors in (78) are left eigenvectors of $\mathbf{A}$ associated with $\lambda = 1$, we find $\nu(\mathbf{T}(1)) = L + p$.

It remains to show that the algebraic multiplicity of $\lambda = 1$ is also $L+p$. This can be done using the Jordan canonical form (see e.g., [11]) $\mathbf{A} = \mathbf{T}\mathbf{J}_A\mathbf{T}^{-1}$, where $\mathbf{T} \in \mathbb{C}^{2Jp\times 2Jp}$ is invertible and $\mathbf{J}_A \in \mathbb{C}^{2Jp\times 2Jp}$ is a block diagonal matrix. Then, we argue by contradiction as in [17, App. H]

To prove (c) recall that the $L + p$ dominant eigenvectors, namely $\{\mathbf{u}_{A,i}\}_{i=1}^{L+p}$, should satisfy $\mathbf{A}\mathbf{u}_{A,i} = \mathbf{u}_{A,i}$ and $\mathbf{v}_{A,i}^T\mathbf{u}_{A,i} = 1$ with $\mathbf{v}_{A,i}$ as in (78). The vectors in (c) readily satisfy these conditions. □

Since $|\lambda_{A,i}| < 1$ for $i = L+p+1, \ldots, 2J(L+p)$, we have $\lim_{k\to\infty} \mathbf{A}^{k-1} = \sum_{i=1}^{L+p} \mathbf{u}_{A,i}\mathbf{v}_{A,i}^T$. Further, the $L + p$ dominant eigenvectors in (78) satisfy $(\mathbf{v}_{A,i}^1)^T(\mathbf{I} - \mathbf{A}_1) + (\mathbf{v}_{A,i}^2)^T = (\mathbf{v}_{A,i}^1)^T$ and $(\mathbf{v}_{A,i}^1)^T + (\mathbf{v}_{A,i}^2)^T\mathbf{A}_2 = \mathbf{0}$, from which it follows that

$$\begin{aligned}
&\lim_{k\to\infty} E\left[\delta\bar{\boldsymbol{\phi}}(k+1)\right] \\
&= \mathbf{A}_\infty\left[(\mathbf{I} - \mathbf{A}_1)^T\mathbf{I}\right]^T\hat{\mathbf{x}} = \sum_{i=1}^{L+p}\mathbf{u}_{A,i}\left(\mathbf{v}_{A,i}^1\right)^T\hat{\mathbf{x}} \\
&= \sum_{i=1}^{L+p}\mathbf{u}_{A,i}\boldsymbol{\epsilon}_i^T\sum_{j=1}^{J}\left[\left(\mathbf{C}_{sx}\left[\mathbf{C}_{xx}^{-1}\right]_j\mathbf{x}_j\right)^T, \left(\left[\mathbf{C}_{xx}^{-1}\right]_j\mathbf{x}_j\right)^T\right]^T \\
&= (\mathbf{1}_J \otimes \mathbf{I}_{L+p})\left[\hat{\mathbf{s}}_{\text{lmmse}}^T\hat{\mathbf{y}}^T\right]^T \qquad (85)
\end{aligned}$$

where the fourth equality (85) follow using $\mathbf{Y}_j$ in Lem. 7. The limit covariance matrix $\mathbf{C}_\eta(k)$ in (40) can be obtained after: i) rewriting $\mathbf{A}^n$ as $\mathbf{A}^n = \sum_{i=1}^{2J(L+p)}\lambda_{A,i}^n\mathbf{u}_{A,i}\mathbf{v}_{A,i}^T$, ii) using the property that $\mathbf{v}_{A,i}^T\boldsymbol{\Sigma}_A\mathbf{v}_{A,i'} = 0\ \forall i, i' \in [1, L+p]$, where $\boldsymbol{\Sigma}_A$ is the matrix between $\mathbf{v}_{A,i}^T$ and $\mathbf{v}_{A,i'}$ in (40), and iii) using the equality $\sum_{n=0}^{\infty}(\lambda_{A,i}\lambda_{A,i'})^n = (1 - \lambda_{A,i}\lambda_{A,i'})^{-1}$ since $|\lambda_{A,i}\lambda_{A,i'}| < 1, \forall\, i, i' \in [L+p+1, 2J(L+p)]$. □

F. Proof of Lemma 5

We first prove (58) for $k = 0$ by rewriting $\hat{\boldsymbol{\chi}}_j(t;t:t)$ as

$$\begin{aligned}
\hat{\boldsymbol{\chi}}_j(t;t:t) &= 2J\left(2 + c_j|\mathcal{B}_j|\right)^{-1}\sigma_{n_j}^{-2}(t) \\
&\quad\times\left[\mathbf{h}_j(t)\mathbf{h}_j^T(t)\mathbf{s}(t) + \mathbf{h}_j(t)n_j(t)\right]. \qquad (86)
\end{aligned}$$

From (86) we can readily verify that (58) holds true for $k = 0$, while the corresponding noise covariance is given by $\boldsymbol{\mathcal{R}}_j(t;t:t) = 4J^2(2 + c_j|\mathcal{B}_j|)^{-2}\sigma_{n_j}^{-2}(t)\mathbf{h}_j(t)\mathbf{h}_j^T(t)$. Arguing by induction, suppose now that (58) holds true for $\hat{\boldsymbol{\chi}}_j(t;t:t+k-k')$ for $k' = 1, \ldots, k$ and $\forall j$. We will show that (58) is also true for $\hat{\boldsymbol{\chi}}_j(t;t:t+k)$. From the induction hypothesis, $\hat{\boldsymbol{\chi}}(t;t:t+k-k') = \hat{\boldsymbol{\mathcal{I}}}(t;t:t+k-k')\mathbf{s}(t) + \hat{\mathbf{n}}(t;t:t+k-k')$, where $\hat{\mathbf{n}}(t;t:t+k-k') := [\hat{\mathbf{n}}_1^T(t;t:t+k-k')\ldots\hat{\mathbf{n}}_J^T(t;t:t+k-k')]^T$ for $k' = 1, \ldots, k$, implying that

$$\begin{aligned}
&\hat{\boldsymbol{\chi}}_j(t;t:t+k) \\
&= \mathbf{A}_j(k)\left[\hat{\boldsymbol{\chi}}^T(t;t:t+k-1), \hat{\boldsymbol{\chi}}^T(t;t:t+k-2)\right]^T \\
&= \mathbf{A}_j(k)\left[\hat{\boldsymbol{\mathcal{I}}}^T(t;t:t+k-1), \hat{\boldsymbol{\mathcal{I}}}^T(t;t:t+k-2)\right]^T\mathbf{s}(t) \\
&\quad + \mathbf{A}_j(k)\left[\hat{\mathbf{n}}^T(t;t:t+k-1), \hat{\mathbf{n}}^T(t;t:t+k-2)\right]^T \\
&= \hat{\boldsymbol{\mathcal{I}}}_j(t;t:t+k)\mathbf{s}(t) + \hat{\mathbf{n}}_j(t;t:t+k) \qquad (87)
\end{aligned}$$

where the third equality in (87) follows from (55) and $\hat{\mathbf{n}}_j(t;t:t+k)$ is equal to the noise term in the rhs of the second equality in (87). Notice that $\hat{\mathbf{n}}_j(t;t:t+k)$ is zero-mean Gaussian. From the second equality in (87) it follows that $\boldsymbol{\mathcal{R}}_j(t;t:t+k)$ is given by (59). Since $\hat{\mathbf{n}}_j(t;t:t)$ is uncorrelated across sensors, and $\hat{\mathbf{n}}_j(t;t:t-1) = \mathbf{0}$ we have $\boldsymbol{\mathcal{R}}(t;t:t) = \text{diag}(\boldsymbol{\mathcal{R}}_1(t;t:t), \ldots, \boldsymbol{\mathcal{R}}_J(t;t:t), \mathbf{0}_{Jp\times Jp})$. □

G. Proof of Lemma 6

It suffices to prove that $\bar{\boldsymbol{\Psi}}(t;t:t+k) := [\boldsymbol{\Psi}^T(t;t:t+k), \boldsymbol{\Psi}^T(t;t:t+k-1)]^T \times [\boldsymbol{\Psi}^T(t;t:t+k), \boldsymbol{\Psi}^T(t;t:t+k-1)] = \boldsymbol{\mathcal{R}}(t;t:t+k)$ for $k = 0, \ldots, K$. This follows readily for $k = 0$ after setting $\boldsymbol{\Psi}_j(t;t:t)$ as in Lemma 5. Using the second equality in (87) we can easily verify that $\boldsymbol{\mathcal{R}}(t;t:t+k) = [\mathbf{A}^T(k), \bar{\mathbf{I}}]^T\boldsymbol{\mathcal{R}}(t;t:t+k-1)[\mathbf{A}^T(k), \bar{\mathbf{I}}]$, where $\mathbf{A}(k) := [\mathbf{A}_1^T(k), \ldots, \mathbf{A}_J^T(k)]^T$ and $\bar{\mathbf{I}} := [\mathbf{I}_{Jp\times Jp}, \mathbf{0}_{Jp\times Jp}]$. Further, from (61) it follows that $\bar{\boldsymbol{\Psi}}(t;t:t+k) = [\mathbf{A}^T(k), \bar{\mathbf{I}}]^T\bar{\boldsymbol{\Psi}}(t;t:t+k-1)[\mathbf{A}^T(k), \bar{\mathbf{I}}]$. Since both $\boldsymbol{\mathcal{R}}(t;t:t+k)$ and $\bar{\boldsymbol{\Psi}}(t;t:t+k)$ are produced by the same recursion, while both are identically initialized $\bar{\boldsymbol{\Psi}}(t;t:t) = \boldsymbol{\mathcal{R}}(t;t:t)$, we find that (64) holds true. □

ACKNOWLEDGMENT

The authors would like to thank Prof. X. Ma for kindly pointing out [21].

REFERENCES

[1] M. Alanyali and V. Saligrama, "Distributed tracking in multi-hop networks with communication delays and packet losses," in *Proc. IEEE Workshop Stat. Signal Process.*, Jul. 2005, pp. 1190–1195.

[2] P. Alricksson and A. Rantzer, "Distributed Kalman filtering using weighted averaging," presented at the 17th Intl. Symp. Math Theory of Nets. Systems, Kyoto, Japan, Jul. 2006.

[3] B. D. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[4] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation*. New York: Wiley, 2001.

[5] S. Barbarossa and G. Scutari, "Decentralized maximum likelihood estimation for sensor networks composed of nonlinearly coupled dynamical systems," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3456–3470, Jul. 2007.

[6] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, 2nd ed. Belmont, MA: Athena Scientific, 1999.
[7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[8] V. Delouille, R. Neelamani, and R. Baraniuk, "Robust distributed estimation in sensor networks using the embedded polygons algorithm," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2998–3010, Aug. 2006.
[9] A. Dogandzic and B. Zhang, "Distributed estimation and detection for sensor networks using hidden Markov random field models," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 3200–3215, Aug. 2006.
[10] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, pp. 388–404, Mar. 2000.
[11] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
[12] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[13] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *Proc. 44th IEEE Conf. Dec., Eur. Contr. Conf.*, Seville, Spain, Dec. 2005, pp. 8179–8184.
[14] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.
[15] R. Olfati-Saber and J. S. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," in *Proc. 44th IEEE Conf. Dec., Eur. Control Conf.*, Seville, Spain, Dec. 2005, pp. 6698–6703.
[16] D. Scherber and H. C. Papadopoulos, "Distributed computation of averages over ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 776–787, Apr. 2005.
[17] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc WSNs with noisy links—Part I: Distributed estimation of deterministic signals," *IEEE Trans. Signal Process.* vol. 56, no. 1, pp. 350–364, Jan. 2008.
[18] S. F. Shah, A. Ribeiro, and G. B. Giannakis, "Bandwidth-constrained MAP estimation for wireless sensor networks," in *Proc. 39th Asilomar Conf. Signals, Systems, Computers*, Monterey, CA, Oct. 2005, pp. 215–219.
[19] D. P. Spanos, R. Olfati-Saber, and R. J. Murray, "Approximate distributed Kalman filtering in sensor networks with quantifiable performance," presented at the 4th Int. Symp. Inf. Processing in Sensor Networks, Los Angeles, CA, 2005.
[20] D. P. Spanos, R. Olfati-Saber, and R. J. Murray, "Distributed sensor fusion using dynamic consensus," presented at the IFAC, Prague, Czech Republic, 2005.
[21] J. Wu and H. Li, "A dominating-set-based routing scheme in ad hoc wireless networks," *Telecommun. Syst. J.*, vol. 3, pp. 63–84, Sep. 2001.
[22] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, pp. 65–78, Sep. 2004.
[23] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least mean-square deviation," *J. Parallel Distrib. Comput.*, vol. 67, pp. 33–46, Jan. 2007.
[24] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. 4th Int. Symp. Inf. Processing Sensor Networks*, Berkeley, CA, 2005, pp. 63–70.

Ioannis D. Schizas received the Diploma degree in computer engineering and informatic s(with Honors from the University of Patras, Patras, Greece, in 2004 and the M.Sc. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, in 2007. Since August 2004, he has been working towards the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis.

His research interests lie in the areas of communication theory, signal processing, and networking. His current research focuses on distributed signal processing with wireless sensor networks, and distributed compression and source coding.

Georgios B. Giannakis (F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, in 1981. From 1982 to 1986 he was with the Univ. of Southern California (USC), where he received the MSc. degree in electrical engineering in 1983, the M.Sc. degree in mathematics in 1986, and the Ph.D. degree in electrical engineering in 1986.

Since 1999 he has been a Professor with the Electrical and Computer Engineering Department at the University of Minnesota, Minneapolis, where he now holds an ADC Chair in Wireless Telecommunications. His general interests span the areas of communications, networking and statistical signal processing—subjects on which he has published more than 250 journal papers, 450 conference papers, two edited books, and two research monographs titled *Space-Time Coding for Broadband Wireless Communications* (New York: Wiley, 2006) and *Ultra-Wideband Wireless Communications* (Cambridge, U.K.: Cambridge Univ. Press, 2007). Current research focuses on diversity techniques, complex-field and space–time coding, multicarrier, cooperative wireless communications, cognitive radios, cross-layer designs, mobile ad hoc networks, and wireless sensor networks.

Dr. Giannakis is the (co-)recipient of six paper awards from the IEEE Signal Processing Society (SPS) and the IEEE Communications Society, including the G. Marconi Prize Paper Award in Wireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from EURASIP (2005), a Young Faculty Teaching Award, and the G. W. Taylor Award for Distinguished Research from the University of Minnesota. He has served the IEEE in a number of posts.

Stergios I. Roumeliotis (S'97–M'00) received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, in 1995, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, in 1997 and 2000 respectively.

From 2000 to 2002, he was a Postdoctoral Fellow at the California Institute of Technology, Pasadena. Since 2002, he has been an Assistant Professor in the Department of Computer Science and Engineering at the University of Minnesota, Minneapolis. His research interests include inertial navigation of aerial and ground autonomous vehicles, fault detection and identification, and sensor networks. Recently his research has focused on distributed estimation under communication and processing constraints and active sensing for reconfigurable networks of mobile sensors.

Dr. Roumeliotis is the recipient of the NSF CAREER award (2006) and the McKnight Land-Grant Professorship award (2006–2008), and he is the co-recipient of the One NASA Peer award (2006), and the One NASA Center Best award (2006). He has coauthored publications for which he received the Robotics Society of Japan Best Paper award (2007), the ICASSP Best Student Paper award (2006), and the NASA Tech Briefs Award (2004). He is currently serving as Associate Editor for the IEEE TRANSACTIONS ON ROBOTICS.

Alejandro Ribeiro (S'05) received the B.Sc. degree in electrical engineering from the Universidad de la Republica Oriental del Uruguay, Montevideo, Uruguay, in 1998 and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Minnesota, Minneapolis, in May 2005 and December 2006, respectively.

From 1998 to 2003, he was a member of the Technical Staff at Bellsouth Montevideo. He is currently a Research Associate in the Department of Electrical and Computer Engineering, University of Minnesota. His research interests lie in the areas of communication theory, signal processing, and networking. His current research focuses on wireless cooperative communications, random access, wireless ad hoc and sensor networks, and distributed signal processing.

Dr. Ribeiro is a Fulbright Scholar.