CONSENSUS-BASED DISTRIBUTED PARAMETER ESTIMATION IN AD HOC WIRELESS SENSOR NETWORKS WITH NOISY LINKS

Ioannis D. Schizas, Alejandro Ribeiro, and Georgios B. Giannakis

Dept. of ECE, University of Minnesota, USA

ABSTRACT

We deal with distributed estimation of deterministic vector parameters using ad hoc wireless sensor networks (WSNs). We cast the decentralized estimation problem as the solution of multiple convex optimization subproblems. Using the method of alternating multipliers we derive algorithms which are decomposable into a set of simpler tasks suitable for distributed implementation. Different from existing alternatives, our approach does not require knowing the desired estimator in closed-form thus allowing for distributed nonlinear estimation. Our algorithms have guaranteed convergence under ideal channel links, while they exhibit noise resilience provably established for the distributed best linear unbiased estimator (BLUE).

Index Terms— Distributed Estimation, Distributed Algorithms

1. INTRODUCTION

A popular application of WSNs is decentralized estimation of unknown deterministic signal vectors using samples collected across sensors. The estimation task can be performed iteratively in a distributed fashion based on successive refinements of local estimates maintained at individual sensors. Each iteration of the estimation algorithm comprises a communication step where the sensors interchange information with their neighbors, and an update step where each sensor uses this information to refine its local estimate. In this context, estimation of deterministic parameters in linear Gaussian models was considered in [6] using the notion of consensus averaging. However, consensus averaging schemes are challenged by the presence of noise (non-ideal sensor links), exhibiting a statistical behavior similar to that of a random walk, and eventually diverging [5].

An alternative to consensus averaging uses the notion of mutually coupled oscillators [1]. Experimental results in [1] suggest that coupled oscillators exhibit noise robustness, but convergence has not been established analytically. In addition, [6] and [1] require the desired estimator to be known in closed-form.

Here we focus on decentralized estimation of deterministic parameter vectors in general (possibly nonlinear and/or non-Gaussian) data models. Novelties of our approach include: i) formulation of the desired estimator as the solution of convex minimization subproblems that exhibit a separable structure and are thus amenable to distributed implementation; ii) unlike [1] and [6], it leads to decentralized algorithms even when the desired estimator is not available in closed-form, as is frequently the case with the maximum likelihood estimator (MLE); and iii) provably noise-resilient algorithms.

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

Specifically, we view MLE as the optimal solution of a separable constrained convex minimization problem in Section 3, and utilize the method of alternating multipliers to construct the corresponding decentralized estimation algorithm. In Section 4 we consider distributed linear estimation using the BLUE, for which we develop a noise-robust distributed algorithm. Numerical results in Section 5 demonstrate the merits of our algorithms with respect to (wrt) [1] and [6].

2. PRELIMINARIES AND PROBLEM STATEMENT

Consider an ad hoc WSN with J sensors. We allow single-hop communications only, so that the j-th sensor communicates solely with nodes i in its neighborhood $\mathcal{N}_j \subseteq [1, J]$. Assuming that links are symmetric, the WSN is modelled as an undirected graph whose vertices are the sensors and its edges represent the available communication links; see Fig. 1. The graph connectivity is summarized in the so called adjacency matrix $\mathbf{E} \in \mathbb{R}^{J \times J}$ for which $\mathbf{E}_{ji} = \mathbf{E}_{ij} = 1$ if $i \in \mathcal{N}_j$, while $\mathbf{E}_{ji} = 0$ if $i \notin \mathcal{N}_j$.

The WSN is deployed to estimate a $p \times 1$ deterministic unknown parameter vector $\mathbf{s}$ based on distributed random observations $\{\mathbf{x}_j \in \mathbb{R}^{L_j \times 1}\}_{j=1}^{J}$. The $\mathbf{x}_j$ observation is taken at the j-th sensor and has probability density function (pdf) $p_j(\mathbf{x}_j; \mathbf{s})$. We further assume that observations are independent across sensors. If $p_j(\mathbf{x}_j; \mathbf{s})$ is known, the MLE is then given by

$$\hat{\mathbf{s}}_{ML} := \arg \min_{\mathbf{s} \in \mathbb{R}^{p \times 1}} -\sum_{j=1}^{J} \ln[p_j(\mathbf{x}_j; \mathbf{s})]. \tag{1}$$

Another estimation scenario arises when the observations adhere to a model for which $E[\mathbf{x}_j] = \mathbf{H}_j \mathbf{s}$ but different from (1), only the covariance matrix $\mathbf{\Sigma}_{x_j x_j} := E[(\mathbf{x}_j - E[\mathbf{x}_j])(\mathbf{x}_j - E[\mathbf{x}_j])^T]$, and the matrix $\mathbf{H}_j$ are known per sensor. This setup arises frequently, and includes as a special case the popular linear model $\mathbf{x}_j = \mathbf{H}_j \mathbf{s} + \mathbf{n}_j$. A pertinent approach in this scenario is to apply the BLUE, which for zero-mean uncorrelated sensor observations is given by

$$\hat{\mathbf{s}}_{BL} := (\sum_{j=1}^{J} \mathbf{H}_j^T \mathbf{\Sigma}_{x_j x_j}^{-1} \mathbf{H}_j)^{-1} \sum_{j=1}^{J} \mathbf{H}_j^T \mathbf{\Sigma}_{x_j x_j}^{-1} \mathbf{x}_j, \tag{2}$$

where ^{T} stands for transposition. Both (1) and (2) will be considered. In particular, we will develop iterative algorithms based on single-hop communications that generate time iterates $\mathbf{s}_j(k)$ so that:

(s1) If $p_j(\mathbf{x}_j; \mathbf{s})$ is known only at the j-th sensor, the local iterates converge as $k \to \infty$ to the MLE, i.e., $\lim_{k \to \infty} \mathbf{s}_j(k) = \hat{\mathbf{s}}_{ML}$.

(s2) If $\mathbf{\Sigma}_{x_j x_j}, \mathbf{H}_j$ are known at the j-th sensor and the block matrix $\mathbf{H} := [\mathbf{H}_1^T \ldots \mathbf{H}_J^T]^T$ has full column rank, then $\lim_{k \to \infty} \mathbf{s}_j(k) = \hat{\mathbf{s}}_{BL}$.

The decentralized algorithm developed under scenario (s1) is attractive for ML estimation in nonlinear data models. The linear estimator considered in (s2) is encountered in many cases of practical

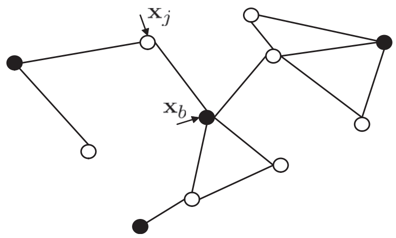

Fig. 1. An ad-hoc wireless sensor network.

interest. The BLUE is outperformed by the MLE but has lower complexity and remains applicable even when only $\{\mathbf{H}_j\}_{j=1}^J$ and $\{\mathbf{\Sigma}_{x_j x_j}\}_{j=1}^J$ are known. Clearly, when $\mathbf{x}_j$ adheres to a linear model $\mathbf{x}_j = \mathbf{H}_j \mathbf{s} + \mathbf{n}_j$ and $\mathbf{n}_j$ is Gaussian, $\hat{\mathbf{s}}_{ML} = \hat{\mathbf{s}}_{BL}$ and consequently (s1) coincides with (s2).

The iterates $\mathbf{s}_j(k)$ will turn out to exhibit resilience to communication noise. Specifically, if $\mathbf{t}_j^i(k) \in \mathbb{R}^{p\times 1}$ represents[1] a vector transmitted from the j-th to the i-th sensor at time slot k, the corresponding vector $\mathbf{r}_i^j(k) \in \mathbb{R}^{p\times 1}$ received by the i-th sensor is

$$\mathbf{r}_i^j(k) = \mathbf{t}_j^i(k) + \mathbf{z}_i^j(k), \tag{3}$$

where $\mathbf{z}_i^j(k) \in \mathbb{R}^{p\times 1}$ represents zero-mean additive white Gaussian noise (AWGN), at sensor i, assumed uncorrelated across sensors and time with covariance $\mathbf{\Sigma}_{zz} = \sigma^2 \mathbf{I}_p$, where $\mathbf{I}_p$ denotes the $p \times p$ identity matrix. Furthermore, we assume that:

(a1) The communication graph is connected.

(a2) The pdf $p_j(\mathbf{x}_j; \mathbf{s})$ is log-concave wrt $\mathbf{s}$.

Similar to [1,6], (a1) ensures utilization of all observation vectors by the decentralized scheme, while (a2) is satisfied by a number of pdfs encountered in practice; see e.g., [3].

3. DISTRIBUTED MLE

In this section we consider decentralized estimation of $\hat{\mathbf{s}}_{ML}$ in (s1), under (a1) and (a2). Since summands in (1) are coupled through $\mathbf{s}$, it is not straightforward to decompose the unconstrained optimization problem in (1). This prompts us to define the auxiliary variable $\mathbf{s}_j$ to represent the local estimate of $\mathbf{s}$ at sensor j and consider the constrained optimization problem

$$\{\hat{\mathbf{s}}_j\}_{j=1}^J := \arg\min -\sum_{j=1}^J \ln[p_j(\mathbf{x}_j; \mathbf{s}_j)], \tag{4}$$
$$\text{s. to } \mathbf{s}_j = \bar{\mathbf{s}}_b,\ b \in \mathcal{B}, j \in \mathcal{N}_b,$$

where $\mathcal{B} \subseteq [1, J]$ is a subset of "bridge" sensors maintaining local vectors $\bar{\mathbf{s}}_b$ that are utilized to impose consensus among local estimates across all sensors. If e.g., $\mathcal{B} = [1, J]$, then (a1) and the constraint $\mathbf{s}_j = \bar{\mathbf{s}}_b,\ b \in \mathcal{B}, j \in \mathcal{N}_b$ will render $\mathbf{s}_j = \mathbf{s}_i\ \forall\ i, j$. In such a case (1) and (4) are equivalent in the sense that $\hat{\mathbf{s}}_j = \hat{\mathbf{s}}_{ML}\ \forall\ j \in [1, J]$. In fact, a milder requirement on $\mathcal{B}$ is sufficient to ensure equivalence of (1) and (4), as described in the following definition.

Definition 1 *Set $\mathcal{B}$ is a subset of bridge sensors 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) $\forall\ b_1 \in \mathcal{B}$ *there exists a sensor $b_2 \in \mathcal{B}$ such that the shortest path between b_1 and b_2 has at most two edges.*

[1] Throughout the paper, subscripts denote the sensor at which variables are "controlled" (e.g., computed at) , while superscripts specify the sensor at which the variable is communicated to.

For the WSN in Fig. 1 a possible selection of sensors forming a bridge sensor subset $\mathcal{B}$, i.e., obeying (a) and (b), is represented by the black nodes. For future reference, the set of bridge neighbors of the j-th sensor will be denoted as $\mathcal{B}_j := \mathcal{N}_j \cap \mathcal{B}$, and its cardinality by $|\mathcal{B}_j|$ for $j = 1, \ldots, J$. Roughly speaking, condition (a) in Definition 1 ensures that every node has a bridge sensor neighbor; while condition (b) ensures that all the bridge variables $\{\bar{\mathbf{s}}_b\}_{b\in\mathcal{B}}$ are equal. Together, they provide a necessary and sufficient condition for the equivalence between (1) and (4) as asserted by the following result [4].

Proposition 1 *The optimal solutions of (1) and (4) coincide; i.e.,*

$$\hat{\mathbf{s}}_{ML} = \hat{\mathbf{s}}_j,\ \forall\ j \in [1, J], \tag{5}$$

if and only if $\mathcal{B}$ is a subset of bridge sensors as per Definition 1.

3.1. The Algorithm of Alternating Multipliers

Here we solve (1) using the method of multipliers to obtain a distributed algorithm for computing the MLE $\hat{\mathbf{s}}_{ML}$. The method of multipliers exploits the decomposable structure of the augmented Lagrangian [2, Chpt. 3]. Let $\mathbf{v}_j^b$ denote the Lagrange multiplier associated with the constraint $\mathbf{s}_j = \bar{\mathbf{s}}_b$, where the multipliers $\{\mathbf{v}_j^b\}_{b\in\mathcal{B}_j}$ are kept at the j-th sensor. The augmented Lagrangian for (4) is

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

where $\boldsymbol{s} := \{\mathbf{s}_j\}_{j=1}^J$, $\bar{\mathbf{s}} := \{\bar{\mathbf{s}}_b\}_{b\in\mathcal{B}}$ and $\mathbf{v} := \{\mathbf{v}_j^b\}_{j\in[1,J]}^{b\in\mathcal{B}_j}$. The constants $\{c_j > 0\}_{j=1}^J$ are penalty coefficients corresponding to the constraints $\mathbf{s}_j = \bar{\mathbf{s}}_b,\ \forall\ b \in \mathcal{B}_j$. Combining the method of multipliers with a block coordinate descent iteration [2, Chpt.3], we have established the following result [4].

Proposition 2 *For a time index k consider iterates $\mathbf{v}_j^b(k)$, $\mathbf{s}_j(k)$ and $\bar{\mathbf{s}}_b(k)$ defined by the recursions*

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

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

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

for all sensors $j \in [1, J]$; and let the initial values of the Lagrange multipliers $\{\mathbf{v}_j^b(-1)\}_{b\in\mathcal{B}_j}$, the local estimates $\{\mathbf{s}_j(0)\}_{j=1}^J$ and the consensus variables $\{\bar{\mathbf{s}}_b(0)\}_{b\in\mathcal{B}}$ be arbitrary. Assuming ideal communication links and the validity of (a1) and (a2), the iterates $\mathbf{s}_j(k)$ converge to the MLE $\hat{\mathbf{s}}_{ML}$ as $k \to \infty$; i.e.,

$$\lim_{k\to\infty} \mathbf{s}_j(k) = \lim_{k\to\infty} \bar{\mathbf{s}}_b(k) = \hat{\mathbf{s}}_{ML},\ \ \forall j \in [1, J],\ b \in \mathcal{B}. \tag{10}$$

We then say that as $k \to \infty$ the WSN reaches consensus.

The recursions in (7)-(9) constitute our distributed (D-) MLE algorithm. All sensors $j \in [1, J]$ keep track of the local estimate $\mathbf{s}_j(k)$ along with the Lagrange multipliers $\{\mathbf{v}_j^b(k)\}_{b\in\mathcal{B}_j}$. The sensors belonging to $\mathcal{B}$ also update the consensus-enforcing variables

$\bar{\mathbf{s}}_b(k)$. During the k-th iteration, sensor j receives the consensus variables $\bar{\mathbf{s}}_b(k)$ from all its neighbors in the subset $\mathcal{B}$, namely all $b \in \mathcal{B}_j$. Based on these consensus variables, it updates the Lagrange multipliers $\{\mathbf{v}_j^b(k)\}_{b\in\mathcal{B}_j}$ using (7), which are then used to compute $\mathbf{s}_j(k+1)$ via (8). After determining $\mathbf{s}_j(k+1)$, sensor j transmits to each of its neighbors $b \in \mathcal{B}_j$ the vector $\mathbf{v}_j^b(k) + c_j\mathbf{s}_j(k+1)$. Each sensor $b \in \mathcal{B}$ receives $\mathbf{v}_j^b(k) + c_j\mathbf{s}_j(k+1)$ from all its neighbors $j \in \mathcal{N}_b$, and proceeds to compute $\bar{\mathbf{s}}_b(k+1)$ using (9). This completes the k-th 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 then initialize the $(k+1)$-st iteration. Note that the cost function in (8) is strictly convex; thus, the optimal solution $\mathbf{s}_j(k+1)$ is unique and can be obtained using e.g., Newton's method.

In the presence of noise, $\bar{\mathbf{s}}_b(k)$ and $\mathbf{s}_j(k)$ in (7)-(9) are corrupted as described in (3). Then, these recursions can be thought of as a stochastic gradient algorithm; see e.g., [2, Sec. 7.8]. This implies that noise causes $\mathbf{s}_j(k)$ to fluctuate around $\hat{\mathbf{s}}_{ML}$ with the magnitude of fluctuations being proportional to the noise variance. However, $\mathbf{s}_j(k)$ is guaranteed to remain within a ball around $\hat{\mathbf{s}}_{ML}$ with high probability [4]. This should be contrasted with [5] which suffers from catastrophic noise propagation.

4. NOISE-ROBUST DISTRIBUTED BLUE

In this section, we consider decentralized estimation of $\hat{\mathbf{s}}_{BL}$ in (s2), under (a1). Similar to Section 3, we write $\hat{\mathbf{s}}_{BL}$ as the optimal solution of a constrained convex minimization problem whose constraints are exactly the same as in (4), and its cost function has the form $\|\mathbf{\Sigma}_{x_jx_j}^{-1/2}\mathbf{H}_j\mathbf{s}_j - \mathbf{\Sigma}_{x_jx_j}^{-1/2}\mathbf{x}_j\|_2^2$ (see [4]). Then, we utilize the method of alternating multipliers to obtain a decentralized algorithm with iterates $\mathbf{s}_j(k)$ converging to $\hat{\mathbf{s}}_{BL}$, and we obtain a recursive scheme as suggested in Proposition 2 with Eq. (8) replaced by

$$\mathbf{s}_j(k+1) = \hat{\mathbf{x}}_j - \mathbf{B}_j^{-1}(\textstyle\sum_{b\in\mathcal{B}_j}\mathbf{v}_j^b(k) + c_j \displaystyle\sum_{b\in\mathcal{B}_j}\bar{\mathbf{s}}_b(k)), \quad (11)$$

where $\mathbf{B}_j := 2\mathbf{H}_j^T\mathbf{\Sigma}_{x_jnx_j}^{-1}\mathbf{H}_j + c_j|\mathcal{B}_j|\mathbf{I}_p$, while the vector $\hat{\mathbf{x}}_j$ is $\hat{\mathbf{x}}_j := \mathbf{B}_j^{-1}2\mathbf{H}_j^T\mathbf{\Sigma}_{x_jx_j}^{-1}\mathbf{x}_j$. Vector $\hat{\mathbf{x}}_j$ can be interpreted as a regularized version of the local BLUE $(\mathbf{H}^T\mathbf{\Sigma}_{x_jx_j}\mathbf{H}_j)^{-1}\mathbf{H}_j^T\mathbf{\Sigma}_{x_jx_j}^{-1}\mathbf{x}_j$. The decentralized algorithm described by the recursions (7), (9) and (11), abbreviated as D-BLUE, guarantees convergence of all local estimates to the centralized BLUE i.e., $\lim_{k\to\infty}\mathbf{s}_j(k) = \hat{\mathbf{s}}_{BL}$ under ideal channel links.

Building on D-BLUE, we will derive another distributed scheme that exhibits improved noise resilience and is amenable to convergence analysis. To this end, let us initialize the recursions in (7), (9) and (11) with $\{\mathbf{v}_b^j(-1) = \mathbf{0}\}_{b\in\mathcal{B}_j}^{j\in[1,J]}$, $\{\bar{\mathbf{s}}_b(-1) = \mathbf{0}\}_{b\in\mathcal{B}}$ and $\{\mathbf{s}_j(0) = \hat{\mathbf{x}}_j\}_{j=1}^J$. Upon substituting (7), in (9) and (11), and stacking local estimates in the vector $\boldsymbol{s}(k) := [\mathbf{s}_1{}^T(k)\ldots\mathbf{s}_J{}^T(k)]^T$, and defining $\hat{\mathbf{x}} := [\hat{\mathbf{x}}_1^T,\ldots,\hat{\mathbf{x}}_J^T]^T$, we find that the D-BLUE iteration is equivalent to [4]:

$$\boldsymbol{s}(k+1) = \boldsymbol{s}(k) - \mathbf{A}_1\boldsymbol{s}(k) - \mathbf{A}_2\boldsymbol{s}(k-1), \quad (12)$$

where $\boldsymbol{s}(0) = \hat{\mathbf{x}}$ and $\boldsymbol{s}(-1) = \mathbf{0}$, and matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are found in [4] to be $\mathbf{A}_1 = (\text{diag}(c_1|\mathcal{B}_1|\ldots c_J|\mathcal{B}_J|)\otimes\mathbf{I}_p)\mathbf{B}^{-1} - 2\mathbf{B}^{-1}\mathbf{W}_E$ and $\mathbf{A}_2 = \mathbf{B}^{-1}\mathbf{W}_E$, with $\mathbf{B} := \text{diag}(\mathbf{B}_1,\ldots,\mathbf{B}_J)$, $\mathbf{e}_b$ denoting the b-th column of the adjacency matrix $\mathbf{E}$, and

$$\mathbf{W}_E = \mathbf{D}_c\sum_{b\in\mathcal{B}}\frac{1}{\sum_{\beta\in\mathcal{N}_b}c_\beta}(\mathbf{e}_b\otimes\mathbf{I}_p)(\mathbf{e}_b\otimes\mathbf{I}_p)^T\mathbf{D}_c, \quad (13)$$

with $\mathbf{D}_c := \text{diag}(c_1,\ldots,c_J)\otimes\mathbf{I}_p$ and $\otimes$ denoting the Kronecker product. It follows from (12) that $\boldsymbol{s}(k)$ is a second-order vector AR process. Careful examination of (9) and (11) reveals that each sensor, say the j-th, updates its local estimate $\mathbf{s}_j(k+1)$ using information from neighboring sensors within a radius of two hops. Indeed, $\mathbf{s}_j(k+1)$ is updated using the consensus variables $\bar{\mathbf{s}}_b(k)$ for $b \in \mathcal{B}_j$, formed using the local estimates of all sensors within the set $\{\mathcal{N}_b\}_{b\in\mathcal{B}_j}$, which contains all the sensors within a distance of either a single hop or two hops from sensor j. As a result, D-BLUE has the potential of achieving higher convergence rates wrt [1, 6] because it utilizes more information across time and space.

Starting from (12), we next develop a decentralized noise-robust algorithm for BLUE. The idea is to introduce an auxiliary vector $\boldsymbol{\phi}(k) := [\boldsymbol{\phi}_1^T(k)\ldots\boldsymbol{\phi}_J^T(k)]^T$, with $\boldsymbol{\phi}_j(k)$ kept at the j-th sensor. We will show that successive differences of $\boldsymbol{\phi}_j(k)$ converge to the BLUE; i.e., $\lim_{k\to\infty}[\boldsymbol{\phi}_j(k+1) - \boldsymbol{\phi}_j(k)] = \hat{\mathbf{s}}_{BL}$. 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 robustness to noise. This is akin to the principle for noise suppression utilized in the approach of coupled oscillators in [1], where a *continuous-time* differential (state) equation is involved per sensor, and the information is encoded in the derivative of the state. The desired *discrete-time* recursion for $\boldsymbol{\phi}_j(k)$ is introduced next [4].

Proposition 3 *If $\boldsymbol{\phi}(0) = \hat{\mathbf{x}}$ and $\boldsymbol{\phi}(-1) = \boldsymbol{\phi}(-2) = 0$, the second-order recursion for $k \geq 0$*

$$\boldsymbol{\phi}(k+1) = \hat{\mathbf{x}} + \boldsymbol{\phi}(k) - \mathbf{A}_1\boldsymbol{\phi}(k) - \mathbf{A}_2\boldsymbol{\phi}(k-1), \quad (14)$$

yields iterates $\boldsymbol{\phi}(k)$ whose difference $\delta\boldsymbol{\phi}(k) := \boldsymbol{\phi}(k+1) - \boldsymbol{\phi}(k)$ equals the iterates $\mathbf{s}(k)$ of (12), i.e., $\boldsymbol{s}(k) = \boldsymbol{\phi}(k) - \boldsymbol{\phi}(k-1)$.

Proposition 3 links (14) with (12); and since $\lim_{k\to\infty}\mathbf{s}_j(k) = \hat{\mathbf{s}}_{BL}$, we have that $\lim_{k\to\infty}\boldsymbol{\phi}_j(k) - \boldsymbol{\phi}_j(k-1) = \lim_{k\to\infty}\delta\boldsymbol{\phi}_j(k-1) = \hat{\mathbf{s}}_{BL}$ [cf. Proposition 3]. Thus, for the recursion in (14) the BLUE is obtained at each sensor from the difference between subsequent states. Interestingly, (14) can be implemented in a distributed fashion and is equivalent to the following recursion per sensor [4]

$$\boldsymbol{\phi}_j(k+1) = \hat{\mathbf{x}}_j + (\mathbf{I} - c_j|\mathcal{B}_j|\mathbf{B}_j^{-1})\boldsymbol{\phi}_j(k) + c_j\mathbf{B}_j^{-1}\sum_{b\in\mathcal{B}_j}\bar{\boldsymbol{\psi}}_b(k)$$

$$\bar{\boldsymbol{\psi}}_b(k) = \sum_{j\in\mathcal{N}_b}\frac{c_j}{\sum_{\beta\in\mathcal{N}_b}c_\beta}\left[2\boldsymbol{\phi}_j(k) - \boldsymbol{\phi}_j(k-1)\right],\ b\in\mathcal{B}. \quad (15)$$

In the presence of communication noise, $\bar{\boldsymbol{\psi}}_b(k)$ and $2\boldsymbol{\phi}_j(k)-\boldsymbol{\phi}_j(k-1)$ in the two recursions of (15) are replaced by $\bar{\boldsymbol{\psi}}_b(k) + \mathbf{z}_j^b(k)$ and $2\boldsymbol{\phi}_j(k) - \boldsymbol{\phi}_j(k-1) + \mathbf{z}_b^j(k)$, respectively. These recursions implement the following steps: (i) all sensors $j \in [1, J]$ receive the vectors $\bar{\boldsymbol{\psi}}_b(k)+\mathbf{z}_j^b(k)$ from $b \in \mathcal{B}_j$ to form the (noisy iterate) $\boldsymbol{\phi}_j(k+1)$; and (ii) bridge sensors receive $2\boldsymbol{\phi}_j(k+1) - \boldsymbol{\phi}_j(k) + \mathbf{z}_b^j(k)$ from $j \in \mathcal{N}_b$ to form $\bar{\boldsymbol{\psi}}_b(k+1)$. Adding the noise terms in the two equations of (15) as described, and stacking the first equation for $j \in [1, J]$ we obtain the noisy version of (14) as

$$\boldsymbol{\phi}(k+1) = \hat{\mathbf{x}} + \boldsymbol{\phi}(k) - \mathbf{A}_1\boldsymbol{\phi}(k) - \mathbf{A}_2\boldsymbol{\phi}(k-1) + \bar{\mathbf{z}}(k) + \bar{\mathbf{z}}_b(k), \quad (16)$$

where $\bar{\mathbf{z}}(k) := [\bar{\mathbf{z}}_1^T(k)\ldots\bar{\mathbf{z}}_J^T(k)]^T$ and $\bar{\mathbf{z}}_b(k) := [\bar{\mathbf{z}}_{b,1}^T(k)\ldots\bar{\mathbf{z}}_{b,J}^T(k)]^T$ have entries

$$\bar{\mathbf{z}}_j(k) = c_j\mathbf{B}_j^{-1}\sum_{b\in\mathcal{B}_j\setminus\{j\}}\mathbf{z}_j^b(k),\quad \bar{\mathbf{z}}_{b,j}(k) = c_j\mathbf{B}_j^{-1}\sum_{\substack{b\in\mathcal{B}_j,\\ j'\in\mathcal{N}_b, j'\neq b}}\frac{c_{j'}\mathbf{z}_b^{j'}(k)}{\sum_{\beta\in\mathcal{N}_b}c_\beta}.$$

The decentralized algorithm resulting from the recursions in (15) and its noisy counterpart summarized in (16) is abbreviated as robust distributed (RD-) BLUE. Next, let $\mathbf{A}$ be a $2Jp\times 2Jp$ matrix formed by the $Jp\times Jp$ submatrices $[\mathbf{A}]_{11} = \mathbf{I}_{Jp} - \mathbf{A}_1$, $[\mathbf{A}]_{12} = -\mathbf{A}_2$,

$[\mathbf{A}]_{21} = \mathbf{I}_{Jp}$ and $[\mathbf{A}]_{22} = \mathbf{0}_{Jp}$. Furthermore, let $\lambda_{A,i}$, $\mathbf{u}_{A,i}$ and $\mathbf{v}_{A,i}$ denote the ith largest in magnitude eigenvalue of $\mathbf{A}$ and the corresponding right and left eigenvectors, respectively. Define also, $\boldsymbol{\Sigma}_{\bar{z}\bar{z}_b} = \text{diag}(\boldsymbol{\Sigma}_{\bar{z}\bar{z}}, \boldsymbol{\Sigma}_{\bar{z}_b\bar{z}_b})$ and $\bar{\boldsymbol{\Sigma}}_{\bar{z}\bar{z}_b} = \text{diag}(\boldsymbol{\Sigma}_{\bar{z}\bar{z}} + \boldsymbol{\Sigma}_{\bar{z}\bar{z}_b}, \mathbf{0}_{Jp})$, whose entries have finite magnitude (since σ^2 is finite). Moreover, let $\delta\bar{\boldsymbol{\phi}}(k) := [\delta\boldsymbol{\phi}(k)^T \quad \delta\boldsymbol{\phi}(k-1)^T]^T$. Interestingly, it turns out that [4]:

Proposition 4 *The RD-BLUE algorithm summarized in* (16) *reaches consensus in the mean i.e.,*

$$\lim_{k\to\infty} E[\delta\boldsymbol{\phi}_j(k)] := \lim_{k\to\infty} E[\boldsymbol{\phi}_j(k+1) - \boldsymbol{\phi}_j(k)] = \hat{\mathbf{s}}_{BL}, \quad j \in [1, J].$$

Matrix $\boldsymbol{\Sigma}_n(k) := [(\delta\bar{\boldsymbol{\phi}}(k) - E[\delta\bar{\boldsymbol{\phi}}(k)])(\delta\bar{\boldsymbol{\phi}}(k) - E[\delta\bar{\boldsymbol{\phi}}(k)])^T]$ *converges to*

$$\lim_{k\to\infty} \boldsymbol{\Sigma}_n(k) = \bar{\boldsymbol{\Sigma}}_{\bar{z}\bar{z}_b} + \sum_{i=p+1}^{2Jp} \sum_{i'=p+1}^{2Jp} \frac{\mathbf{u}_{A,i}\mathbf{u}_{A,i'}^T}{1 - \lambda_{A,i}\lambda_{A,i'}} \mathbf{v}_{A,i}^T \cdot \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_1 \\ -\mathbf{I} & -\mathbf{I} \end{bmatrix} \boldsymbol{\Sigma}_{\bar{z}\bar{z}_b} \begin{bmatrix} \mathbf{A}_1^T & -\mathbf{I} \\ \mathbf{A}_1^T & -\mathbf{I} \end{bmatrix} \mathbf{v}_{A,i'}. \quad (17)$$

Furthermore, the entries of $\boldsymbol{\Sigma}_n(k)$ *are bounded.*

Proposition 4 establishes convergence of RD-BLUE in the mean. It also shows that even though noise causes the local estimates to fluctuate around BLUE, their variance remains bounded as $k \to \infty$. Iterates in the consensus average approach of [5] are obtained via a first-order vector AR process. In order to effect consensus, the largest eigenvalue of the matrix defining the AR recursion has to be 1. This entails, alas, an unstable AR process and leads to catastrophic noise propagation. For the coupled oscillators in [1] the consensus is achieved in the derivative of a continuous-time state. Noise resilience is thus expected, and indeed observed in simulations, but not formally established. As per Proposition 4, RD-BLUE is shown to achieve consensus in the mean with local iterates remaining within a ball of consensus with high probability.

5. NUMERICAL EXAMPLES

Here we test the convergence of D-BLUE and RD-BLUE, and compare them with the coupled oscillators (CO) based BLUE in [1] and the consensus average (CA) BLUE in [6]. Furthermore, we examine the noise resilience properties of the aforementioned schemes in the presence of communication noise. We consider a WSN with $J = 60$ sensors. Nodes in the WSN are randomly placed in the unit square $[0, 1] \times [0, 1]$ with uniform distribution. Each sensor collects 5 observations, i.e., $\{L_j\}_{j=1}^{50}$, while $\mathbf{s}$ incorporates $p = 2$ parameters. We consider a linear model, where the entries of $\mathbf{H}_j$ are random uniformly distributed over $[-0.5, 0.5]$ and $\{\mathbf{n}_j\}_{j=1}^{J}$ are zero-mean AWGN with $\boldsymbol{\Sigma}_{n_j n_j} = 0.5\mathbf{I}_{p\times p}$. Noise variance σ^2 is adjusted so that $\text{SNR} := 10\log_{10}\frac{\hat{\mathbf{s}}_{BL}/p}{\sigma^2}$ assumes specific values.

Fig. 2 depicts the average noise variance per sensor, namely the $\text{trace}(\boldsymbol{\Sigma}_n(k))/J$, versus iteration index k, after incorporating noise in the sensor links so that SNR = 20dB. Specifically, the noise variance per sensor is computed via ensemble averaging across sensors and across 50 different realizations of the RD-BLUE,D-BLUE, CO-BLUE and CA-BLUE. For a fair comparison between RD-BLUE and CO-BLUE we set the penalty coefficients in RD-BLUE and D-BLUE to $c_j = 1/|\mathcal{B}_j|$ for $j = 1, \ldots, 60$, while in CO-BLUE the corresponding parameter c is set equal to the value that attains the highest convergence rate. The selected parameters $\{c_j\}_{j=1}^{60}$ guarantee that the steady-state noise variance for both RD-BLUE and CO-BLUE is the same with $\text{trace}(\lim_{k\to} \boldsymbol{\Sigma}_n(k)) = 1.4 \cdot 10^{-3}$. For CA-BLUE we adopt the max-degree and Metropolis weights [6]. As expected, CA-BLUE eventually diverges in the presence of noise. Notice that the D-BLUE exhibits noise resilience, at the expense of higher steady-state variance than the RD-BLUE. But RD-BLUE achieves a higher convergence rate relative to CO-BLUE while the steady state-noise variance is the same for both schemes. Thus, RD-BLUE is flexible to tradeoff convergence rate for steady-state error variance[2].

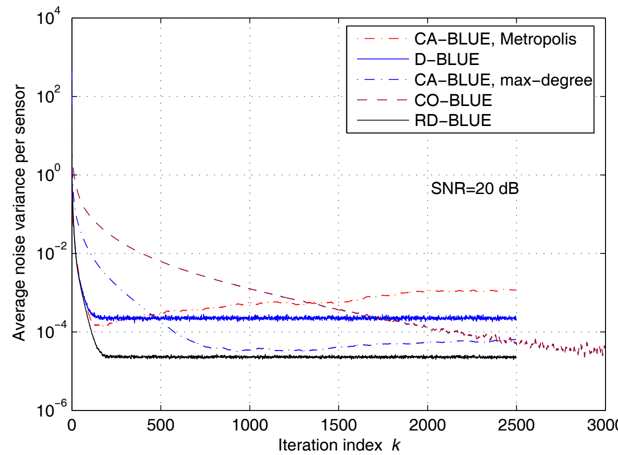

Fig. 2. Average noise variance per sensor vs. k for D-BLUE, RD-BLUE, CA-BLUE and CO-BLUE.

6. REFERENCES

[1] S. Barbarossa and G. Scutari, "Decentralized Maximum Likelihood Estimation for Sensor Networks Composed of Nonlinearly Coupled Dynamical Systems," *IEEE Trans. on SP.*, 2007; see also *Intl .Conf. on ASSP*, vol. IV, pp. 385–388, Toulouse, France, May 2006.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific, Second Edition, 1999.

[3] A. Ribeiro and G. B. Giannakis, "Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks," *IEEE Trans. On Sig. Proc.*, vol. 54, pp. 2784–2796, 2006.

[4] 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. on Signal Proc.*, September 2006 (submitted); http:// spincom.ece.umn.edu/.

[5] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed Average Consensus With Least-Mean-Square Deviation," *Journal of Par. and Distr. Comput.*, submitted May 2005; http://www.stanford.edu/∼boyd/.

[6] L. Xiao, S. Boyd, and S. Lall, "A Scheme for Robust Distributed Sensor Fusion Based on Average Consensus," in *Proc. of the 4th Intl. Symp. on Info. Proc. in Sensor Net.*, Berkeley, pp. 63–70.

[2] The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U. S. Government.