Consensus-Based Distributed Estimation of Random Signals with Wireless Sensor Networks†

Ioannis D. Schizas and Georgios B. Giannakis
Dept. of ECE, University of Minnesota
200 Union Str. SE, Minneapolis, MN 55455, USA
Emails: {schizas,georgios}@ece.umn.edu

Abstract—We deal with distributed linear minimum mean-square error (LMMSE) estimation of a random signal vector based on observations collected across a wireless sensor network (WSN). We cast this decentralized estimation problem as the solution of multiple constrained convex optimization sub-problems. Using the method of multipliers in conjunction with a block coordinated descent approach we demonstrate how the resultant algorithm can be decomposed into a set of simpler tasks suitable for distributed implementation. Relative to existing alternatives, we establish that the novel decentralized algorithm guarantees convergence to the centralized LMMSE estimator for quite general (possibly nonlinear and/or non-Gaussian) data models. Through numerical experiments, we finally illustrate the convergence properties of the algorithm.

I. INTRODUCTION

Even though the gamut of WSN-driven applications is yet to be fully delineated, it is clear that the design of WSNs must be task-specific and adhering to stringent power and bandwidth constraints. The task we consider here is distributed LMMSE estimation of a stationary random signal vector using discrete-time samples collected across sensors. The constraints are respected by alleviating communication and computational costs. This becomes possible in an ad hoc WSN topology whereby each sensor communicates only with its neighbors and forms local estimates as linear functions of its own observations and those it receives from neighboring sensors. In this context, estimation of deterministic parameters using the notion of consensus averaging was considered in [5], [7] and [8]; see also [2] and [1]. Distributed estimation of Gaussian random parameters in a scalar linear model was also reported in [3] based on Jacobi’s iteration.

In the present paper, we consider decentralized estimation of random parameters in general (possibly nonlinear and non-Gaussian) data models. The novelty of this paper’s approach is twofold: i) formulation of the desired estimator as the solution of convex minimization sub-problems which are separable and thus amenable to distributed implementation; and ii) development of a decentralized algorithm that guarantees convergence of all local estimates to the desired centralized LMMSE estimator even when the data are non-Gaussian and the generating data model is nonlinear.

After stating the problem in Section II, we proceed to view LMMSE estimation as the optimal solution of a separable constrained convex minimization problem in Section III. The resultant formulation reveals that a properly selected subset of sensors suffices to impose the consensus requirement across the WSN and lead local estimates per sensor to converge to the centralized LMMSE estimate. We utilize the method of multipliers to find the LMMSE optimal solution as the minimum of the augmented Lagrangian. We then rely on a block coordinate descent algorithm to decompose the minimization of a suitably defined augmented Lagrangian into simple separable tasks. Convergence to the centralized LMMSE estimator across all sensors is guaranteed for general data models. Finally, we provide corroborating simulations in Section IV and conclude the paper in Section V.

II. PROBLEM STATEMENT

With reference to Fig. 1, consider an ad hoc WSN with \( J \) sensors represented by a graph whose vertices indicate sensor locations and edges denote the available communication links. The set of sensors having an active link with the \( j \)-th sensor comprise its neighborhood \( \mathcal{N}_j \), the cardinality of which we denote as \( |\mathcal{N}_j| \). Each sensor, say the \( j \)-th one, observes an \( L_j \times 1 \) vector \( x_j \) that is correlated with the \( p \times 1 \) random signal of interest \( s \). The data model relating \( s \) with \( \{x_j\}_{j=1}^J \) can be arbitrary, not necessarily linear and/or Gaussian. We are interested in estimating \( s \) across all \( J \) sensors using linear functions of the observation data \( x := [x_1^T \ldots x_J^T]^T \in \mathbb{R}^{L \times 1} \), where \( L := \sum_{j=1}^J L_j \) and \( T \) stands for transposition. To this end, we will assume without loss of generality that \( x \) and \( s \) are zero-mean with cross-covariance matrices \( C_{sx}, C_{xx} := E[ss^T] \), \( C_{xixj} := E[x_ix_j^T] \) for \( i,j \in [1,J] \). The classical (centralized) LMMSE estimator of \( s \) based on \( x \) is [4]:

\[
\hat{s} = C_{sx}C_{xx}^{-1}x,
\]

where \( C_{xx} := \{C_{xixj}\} \) and \( C_{sx} := \{C_{sx1}, \ldots, C_{sxJ}\} \).

With the samples (sub-vectors) in \( x \) scattered across the WSN, one approach to finding \( \hat{s} \) could be for each sensor to transmit \( x_j \) to a fusion center (FC) where the LMMSE estimator could be formed as in the centralized case [4], provided that the FC has also available the auto- and cross-covariance matrices \( C_{xx} \) and \( C_{sx} \). The advantages of the ad

† Prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory 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.
hoc topology considered here over the FC-based star topology is twofold: i) communications of remote sensors with the FC can be costly; and ii) a star topology is prone to FC failures.

In this paper’s ad hoc topology, we assume that the $j$-th sensor has available cross-correlation information between its observations and the other sensors’ observations and the parameter $s$, namely $C_{xx} := [C_{x_1x_1}, \ldots, C_{x_Jx_J}]$ and $C_{sx}$.

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 $C_{xx}$ and $C_{sx}$, but only a portion of them containing $1/J$ of the total covariance information. The challenge is that both the observations as well as the cross-correlation information ($C_{xx}$ and $C_{sx}$) are scattered across the WSN.

III. DISTRIBUTED LMMSE ESTIMATION

Our approach to developing a distributed (D-) LMMSE estimator is to write $\hat{s}$ as the solution of an appropriate convex optimization problem and then use the method of multipliers along with a block coordinate descent iteration [2], to split the original problem into simpler subtasks that can be implemented in parallel. Toward this objective, consider the $L \times 1$ vector $\hat{y} := C_{xx}^{-1}x$ partitioned as $\hat{y} = [\hat{y}_1^T \ldots \hat{y}_J^T]^T$, with $\hat{y}_j \in \mathbb{R}^{L \times 1}$. Writing the LMMSE estimator as $\hat{s} = \sum_{j=1}^J C_{sx_j}^{} \hat{y}_j^{}$, it can be easily shown that [6]:

**Proposition 1:** The LMMSE estimator $\hat{s}$ in (1) can be alternately obtained as

$$\hat{s}, \hat{y} = \arg \min_{s \in \mathbb{R}^{L 	imes 1}, y \in \mathbb{R}^L} \sum_{j=1}^J ||s - J C_{sx_j} y_j||_2^2.$$  (2)

Since summands in (2) are coupled through the vector $s$, separate minimization of each one of them (locally per sensor) will not ensure convergence to the global LMMSE optimal estimator in general. To guarantee global optimality of local estimates, we introduce the auxiliary vectors $s^j, y^j := [\hat{y}_1^T \ldots \hat{y}_j^T]^T$ for each sensor $j$ and impose the “consensus constraint” $s^i = \ldots = s^j$ to ensure that $\hat{s}^j = s^j \forall j \in [1, J]$. As a notational convention throughout the paper, the subscript stands for the corresponding sub-vector of a super-vector; and the superscript consistently denotes the sensor at which auxiliary variables are computed; e.g., $\hat{y}_k^j$ encapsulates the covariance information from sensor $i$ that is utilized by sensor $j$ to form its local estimate $\hat{s}^j$.

Guaranteeing consensus will turn out to require communicating information among neighboring sensors. These communications can be reduced if instead of all sensors we require consensus among a subset $B \subseteq \{1, \ldots, J\}$ of what we will henceforth term “bridge” sensors. These considerations prompt us to modify (2) as

$$\{\hat{s}^j, \hat{y}^j\}_{j=1}^J := \arg \min_{s^j \in \mathbb{R}^{L \times 1}, y^j \in \mathbb{R}^L} \sum_{j=1}^J ||s^j - J C_{sx_j} y_j||_2^2 \quad \text{s.t. } s^i = s^j, b \in B, j \in N_b, \quad y^j = y^b, b \in B, j \in N_b, \quad C_{sx_j} y_j = x_j, j \in [1, J].$$  (3)

To establish the equivalence between (2) and (3) at the optimum (when consensus is reached) conditions that are practically satisfied must be imposed on the set of bridge sensors, as specified next.

**Definition 1:** The set $B$ comprises a subset of bridge sensors if and only if

(a) $\forall j \in [1, J]$ there exists at least one $b \in B$ so that $b \in N_j$; and

(b) $\forall b \in B$ there exists a sensor $b_2 \in B$ so that the shortest path between $b_1$ and $b_2$ has at most two edges.

An example is depicted in Fig. 1, where the black sensors represent a possible selection of $B$ obeying (a) and (b). The set of bridge neighbors of the $j$-th sensor will be denoted as $M_j := N_j \cap B$. The equivalence between (2) and (3) is summarized in the following [6]:

**Proposition 2:** If $B$ is as in Definition 1, then the optimal solutions of (2) and (3) coincide; i.e.,

$$\hat{s} = \hat{s}^j, \quad j \in [1, J].$$  (4)

Proposition 2 asserts that consensus can be achieved across the WSN without utilizing all the $J$ sensors but only a subset of them.

A. The Algorithm of Alternating Multipliers

In this subsection, we will show how to solve (3) by combining the method of multipliers with a block coordinate descent approach [2]. Interestingly, this procedure will yield a distributed estimation algorithm whereby recursive updates per sensor guarantee convergence to the centralized LMMSE estimator $\hat{s}$ across all sensors.

Let $\{v^b_j, w^b_j\}_{b \in M_j}$ and $\{\mu^j\}_{j=1}^J$ denote the Lagrange multipliers to be updated at the $j$-th sensor and are associated with the constraints $s^j = s^b$, $y^j = y^b$ and $C_{sx_j} y^j = x_j$, respectively. Consider now the augmented Lagrangian function of the optimization problem in (3)

$$\mathcal{L}[u, \psi, \bar{u}, \bar{\psi}, v, w, \mu] = \sum_{j=1}^J ||s^j - J C_{sx_j} y_j||_2^2 \quad + \sum_{j=1}^J (\mu^j)^T (C_{sx_j} y_j - x_j) + \frac{c}{2} \sum_{j=1}^J ||C_{sx_j} y_j - x_j||_2^2 \quad + \sum_{b \in B} \sum_{j \in N_b} ||v^b_j^T (s^b - s^j) + (w^b_j)^T (y^j - y^b)||_2^2 \quad + \frac{c}{2} \sum_{b \in B} \sum_{j \in N_b} (||s^j - s^b||_2^2 + ||y^j - y^b||_2^2),$$  (5)
where $\mathbf{u} := \{s^j\}_{j=1}^J$, $\psi := \{y^j\}_{j=1}^J$, $\bar{\mathbf{u}} := \{s^b\}_{b \in \mathcal{B}}$, $\bar{\psi} := \{y^b\}_{b \in \mathcal{B}}$, $\mathbf{v} := \{v^j_k\}_{j \in [1,J]}$, $\mathbf{w} := \{w^j_k\}_{j \in [1,J]}$, $\mu := \{\mu^j_k\}_{j=1}^J$, and the constant $c > 0$ is arbitrary.

Applying the method of multipliers and block coordinate descent minimization, we can summarize our main result pertaining to the distributed solution of (5) as follows:

**Proposition 3:** For each sensor $j \in [1,J]$, consider per iteration the recursions

\[
\begin{align*}
\psi^i_j(k) &= \psi^i_b(k-1) + c(s^j(k) - s^b(k)), \quad b \in \mathcal{M}_j \\
\psi^i_b(k) &= \psi^i_b(k-1) + c(y^j(k) - y^b(k)), \quad b \in \mathcal{M}_j \\
\mu^i_j(k) &= \mu^i_j(k-1) + c(C_{x_j} y^j(k) - x_j), \\
y^j(k+1) &= F_j^{-1}(C_{x_j}(\mu^j(k) + 2J \psi^j)(k) + \sum_{b \in \mathcal{M}_j} [cy^b(k) - w^b(k)]), \\
s^j(k+1) &= \frac{2J}{(2 + c|M_j|)} C_{s_{x_j}} \psi^j(k+1) + \zeta^j(k), \\
s^b(k+1) &= \frac{1}{|\mathcal{B}|} \sum_{j \in \mathcal{B}} [c^{-1} \psi^b_j(k) + s^j(k+1)], \quad b \in \mathcal{B}, \\
y^b(k+1) &= \frac{1}{|\mathcal{B}|} \sum_{j \in \mathcal{B}} [c^{-1} \psi^b_j(k) + y^j(k+1)].
\end{align*}
\]

where

\[
\mathbf{F}_j := 2J^2c|M_j|((2 + c|M_j|)^{-1} C_{s_{x_j}} C_{s_{x_j}}^T \\
+ c^2 C_{x_j}^T C_{x_j} + c|M_j|I_{x \times L}), \\
\zeta^j(k) := (2 + c|M_j|)^{-1} \sum_{b \in \mathcal{M}_j} [cs^b(k) - \psi^b_j(k)],
\]

with $\mathbf{C}_{x_j} := [0_{p \times L_1}, \ldots, C_{s_{x_j}}, \ldots, 0_{p \times L_j}]^T$ and $c > 0$. Then, as $k \to \infty$ the WSN reaches consensus; i.e.,

\[
\lim_{k \to \infty} s^j(k) = \lim_{k \to \infty} s^b(k) = \bar{s}, \quad \forall j \in [1,J], b \in \mathcal{B}.
\]

**Proof:** We wish to show that (6)-(12) generate a series of local estimates converging to the optimal solution of the optimization problem in (3) when $\mathcal{B}$ is a bridge sensor subset, i.e., the LMMSE. We will establish this using the method of multipliers which consists of two steps [2]: i) minimization of the augmented Lagrangian in (5); and ii) recursive updates of the corresponding Lagrange multipliers. Let $\psi^i_j(k), \psi^i_b(k)$ and $\mu^i_j(k)$ denote the Lagrange multipliers at the $k$-th iteration. The aforementioned two steps during the $(k+1)$-st iteration are:

[S1] Set $\mathbf{v} := \mathbf{v}(k) = \{v^j_k(k)\}_{j \in [1,J]}$, and similarly fix $\mathbf{w} := \mathbf{w}(k)$ and $\mu := \mu(k)$ to obtain $\mathbf{u}(k+1), \psi(k+1), \bar{\mathbf{u}}(k+1), \bar{\psi}(k+1)$ by minimizing the augmented Lagrangian in (5) as:

\[
\text{arg min}_{\mathbf{u}, \psi, \bar{\mathbf{u}}, \bar{\psi}} \mathcal{L}_a[\mathbf{u}, \psi, \bar{\mathbf{u}}, \bar{\psi}, \mathbf{v}, \mathbf{w}, \mu].
\]

[S2] Update the Lagrange multipliers $\{v^j_k(k), w^j_k(k)\}_{j \in [1,J]}$ and $\{\mu^j_k\}_{j=1}^J$ by using the updates (6)-(8).

It is known that if $\mathcal{L}_a[\mathbf{u}, \psi, \bar{\mathbf{u}}, \bar{\psi}, \mathbf{v}, \mathbf{w}, \mu]$ is the augmented Lagrangian of a strictly convex optimization problem, then [S1]-[S2] converge to the unique global minimum for any constant $c > 0$ [2]. Notice that [S1] is not amenable to a distributed implementation since it requires joint minimization of (5) with respect to (wrt) $\mathbf{u}, \psi, \bar{\mathbf{u}}, \bar{\psi}$. We overcome this by applying a block coordinate descent method, where we minimize $\mathcal{L}_a[\mathbf{u}, \psi, \bar{\mathbf{u}}, \bar{\psi}, \mathbf{v}, \mathbf{w}, \mu]$ wrt one variable at a time, effectively replacing [S1] with the following steps:

[S1-a] For fixed $\bar{\mathbf{u}} = \bar{\mathbf{u}}(k)$ and $\bar{\psi} = \bar{\psi}(k)$, determine $\mathbf{u}(k+1)$ and $\psi(k+1)$ as

\[
\text{arg min}_{\mathbf{u}, \psi} \mathcal{L}_a[\mathbf{u}, \psi, \bar{\mathbf{u}}, \bar{\psi}, \mathbf{v}, \mathbf{w}(k), \mu(k)].
\]

Based on (5), we deduce that (14) is equivalent to $J$ separate optimization sub-problems:

\[
\text{arg min}_{s^{i_j} \psi^j} \mathcal{L}_a[\mathbf{u}(k+1), \psi(k+1), \bar{\mathbf{u}}(k+1), \bar{\psi}(k), \mathbf{v}(k), \mathbf{w}(k), \mu(k)].
\]

[S1-b] Setting $\{s^{i_j} = s^{i_j}(k+1)\}_{j=1}^J$ and $\{\psi^j = \psi^j(k+1)\}_{j=1}^J$ we can now minimize wrt $\bar{\mathbf{u}}$ to obtain

\[
\bar{\mathbf{u}}(k+1) = \text{arg min}_{\bar{\mathbf{u}}} \mathcal{L}_a[\mathbf{u}(k+1), \psi(k+1), \bar{\mathbf{u}}, \bar{\psi}(k), \mathbf{v}(k), \mathbf{w}(k), \mu(k)].
\]

Similar to [S1-a], this minimization can be decomposed as

\[
\text{arg min}_{s^{i_j}, \psi^j} \mathcal{L}_a[\mathbf{u}(k+1), \psi(k+1), \bar{\mathbf{u}}, \bar{\psi}(k), \mathbf{v}(k), \mathbf{w}(k), \mu(k)].
\]

From the last minimization problem, we finally obtain the desired equivalent decoupled formulation

\[
\psi^j(k+1) = \text{arg min}_{\psi^j} \mathcal{L}_a[\mathbf{u}(k+1), \psi(k+1), \bar{\mathbf{u}}(k+1), \bar{\psi}(k), \mathbf{v}(k), \mathbf{w}(k), \mu(k)].
\]

The cost functions in (15)-(17) are strictly convex; thus, their global optimal solution can be obtained by applying first-order optimality conditions; i.e., by simply equating the gradient to zero. It turns out that the optimal solutions of (15)-(17) are given by (9)-(12). The algorithm formed by [S1-a]-[S1-c] and [S2] belongs to the class of what are known as alternating multiplier schemes which similar to [S1]-[S2] also converge to the unique global minimum for any $c > 0$ [2, Chapter 3].

Recursions (6)-(12) constitute our distributed estimation algorithm (abbreviated as D-LMMSE) whereby all the sensors $j \in [1,J]$ keep track of the local estimate $s^j(k)$ along with $y^j(k)$ and the Lagrange multipliers $\{v^j_k(k), w^j_k(k)\}_{b \in \mathcal{M}_j}$ and $\mu^j_k(k)$. The sensors belonging to the subset $\mathcal{B}$ update also the consensus enforcing variables $s^b(k)$ and $y^b(k)$. During the $k$-th iteration, sensor $j$ receives the consensus variables $s^b(k)$ and $y^b(k)$ from all its neighbors in the subset $\mathcal{B}$, namely all $b \in \mathcal{M}_j$. Based on these consensus variables, it updates its Lagrange multipliers $\{v^j_k(k), w^j_k(k)\}_{b \in \mathcal{M}_j}$ and $\mu^j_k(k)$ using
positive definite. Taking advantage of these properties of $B$ in (12). This completes the ${k}$th iteration. Furthermore, the communication process involves transmission of $O(L)$ scalars per iteration which is expected since the covariance information contained in $C_{xx}$ is scattered across the network.

Remark 1: Relative to [3], the D-LMMSE algorithm in this paper is more general since it neither requires linearity nor Gaussianity in the data model. Furthermore, each sensor in [3] is assumed to have available information from the matrices $C_{ss}^{-1}$ and $C_{rn}^{-1}$. This requires an additional decentralized algorithm to disseminate these inverses across the WSN. A fair comparison between [3] and the present D-LMMSE algorithm does not appear possible since the former utilizes information from the inverse covariance matrices which facilitates estimation but leaves open the question of how and how costly acquiring this information is.

Remark 2: Similar to [3], [7], [8], [5], the communication graph is connected and the wireless links over which local information is exchanged among sensors are assumed ideal. This requires sufficiently strong error control codes to mitigate multiplicative fading and additive noise effects.

Algorithm 1 D-LMMSE

Initialize $\{s^b(0), y^b(0)\}_{b \in B}$ and $\{v^s(0), w^s(0)\}_{s \in S}$ and $\{\mu^l(0)\}_{l=1}^5$ to zero.

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

Bridge sensors $b \in B$: transmit $s^b(k), y^b(k)$ to its neighbors in $N_b$

All $j \in [1, J], j \neq b$: update $\{v^s_b(k), w^s_b(k)\}_{b \in M_j}$ using (6)-(8)

All $j \in [1, J], j \neq c$: transmit $\mu^s_c(k)$

end for

IV. NUMERICAL EXAMPLES

In this section, we test the convergence properties of our D-LMMSE estimation algorithm with the use of ad hoc WSNs. The WSN is generated by randomly placing nodes according to a uniform distribution in the unit square $[0, 1] \times [0, 1]$. We assume that two sensors are able to communicate – and thus have an edge in Fig. 3 (a) – if their Euclidean distance is less than 1/4. We further consider a linear data model $x_j = H_j s + n_j$, $j \in [1, J]$. Noise $n_j$ is spatially uncorrelated with $C_{n_j} = I_{s_j \times s_j}$. The covariance matrix of $s$ is $C_{xx} = 0.5 I_{p \times p}$, and the entries of $H_j$ are generated independently and uniformly distributed over the interval $[-0.5, 0.5]$. The metric used to demonstrate convergence is the total normalized error

$$E_{\text{norm}}(k) = \sum_{j=1}^J \frac{||w^j(k) - \bar{s}||^2_2}{||\bar{s}||^2_2},$$

where $s^j(k)$ is the local estimate at the $j$th sensor after the $k$th iteration. In Fig. 3 (b), we plot $E_{\text{norm}}(k)$ as a function of the iteration index $k$ for a WSN with 30 sensors with each sensor having 5 observations, i.e., $L_1 = \cdots = L_{30} = 5$, and $s$ incorporating $p = 5$ parameters; see also Fig. 3 (a). Fig. 4 (a) depicts a WSN with 50 sensors, $L_1 = \cdots = L_{50} = 20$ and $p = 5$, while Fig. 4 (b) displays $E_{\text{norm}}(k)$ versus iteration $k$. Clearly, the algorithm converges since $E_{\text{norm}}(k) \to 0$ as $k \to \infty$. Observe also that appropriate selection of the scalar $c$ leads to improved convergence rates.

V. CONCLUSIONS

We developed a distributed algorithm for estimating random signals using data collected by an ad-hoc WSNs based on successive refinement of local estimates. At each transmission cycle of the novel algorithm, information is exchanged among single-hop neighbors only; the information received from these neighbors is then used to improve their local estimate. The crux of our approach is to express the linear mean-square error (LMMSE) estimator as the solution of judiciously designed convex optimization problems. We then used the method of multipliers combined with block coordinate descent updates to enable parallel implementation. Our algorithm is guaranteed...
to converge to the centralized LMMSE optimal estimate for general non-Gaussian and nonlinear data models. Numerical tests corroborated our convergence results.

Future research topics include generalizing our approach to distributed estimation of dynamical (i.e., state-space) processes as well as a better understanding of the convergence rate as a function of the parameter $c$ and the effect of non-ideal local links.\(^1\)

REFERENCES


\(^1\)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.