OPTIMAL DIMENSIONALITY REDUCTION FOR MULTI-SENSOR FUSION IN THE PRESENCE OF FADING AND NOISE

Ioannis D. Schizas, Georgios B. Giannakis, and Zhi-Quan Luo

Dept. of ECE, University of Minnesota
200 Union Street, Minneapolis, MN 55455, USA

ABSTRACT

We derive linear estimators of stationary random signals based on reduced-dimensionality observations collected at distributed sensors and communicated over wireless fading links to a fusion center, where additive noise is also present. Dimensionality reduction compresses sensor data to meet low-power and bandwidth constraints, while linearity in compression and estimation are well motivated by the limited computing capabilities wireless sensor networks are envisioned to operate with. For uncorrelated sensor data, we develop mean-square error (MSE) optimal estimators in closed-form; while for correlated sensor data, we derive sub-optimal iterative estimators which guarantee convergence at least to a stationary point. Performance analysis and corroborating simulations demonstrate the merits of the novel distributed estimators relative to existing alternatives.

1. INTRODUCTION

With the popularity of battery-powered wireless sensor networks (WSNs), distributed estimation relying on sensor data processed at a fusion center (FC) has attracted increasing interest recently. Constrained by limited power and bandwidth resources, existing approaches either take advantage of spatial correlations across sensor data to reduce transmission requirements [2, 6, 7], or, rely on severely quantized (possibly down to one bit) *digital* WSN data to form distributed estimators of *deterministic* parameters; see e.g., [4] and references therein. Distributed estimation of *random* signals has also been considered by [3, 6, 7], but results are restricted by one or more of the following assumptions: i) Gaussian signals and/or sensor data; ii) linear sensor observation models; and iii) ideal links; i.e., absence of fading and additive noise at the FC.

Overcoming limitations i)-iii), our goal in this paper is to form estimates at the FC of a *random* stationary vector based on *analog-amplitude* multi-sensor observations. To enable estimation under the stringent power and computing limitations of WSNs, we seek linear dimensionality reducing operators (data compressing matrices) per sensor along with linear operators at the FC, in order to minimize the mean-square error (MSE) in estimation. If sufficiently strong error-control codes are used, we can treat links as ideal and formulate this intertwined compression-estimation task as a canonical correlation analysis problem [5]. Here, we explicitly account for non-ideal links and develop distributed estimators generally applicable to nonlinear and non-Gaussian setups.

Specifically, with uncorrelated (decoupled) sensor data we derive in closed-form the MSE optimal compressing and estimation matrices and prove that the optimal solution amounts to optimally compressing the linear minimum mean-square error (LMMSE) signal estimate formed at each sensor (Section 3). With correlated (coupled) sensor observations, globally optimal distributed estimation has been shown to be NP-hard when reduced-dimensionality sensor data are concatenated at the FC [3]. For this case, we develop a block coordinate descent iterative estimator which always converges to a stationary point (Section 4). This distributed estimator subsumes a recent distributed reconstruction algorithm derived for Gaussian sources and ideal links in [2]. Our findings in Sections 3 and 4 are corroborated by numerical examples (Section 5). We conclude this paper in Section 6.

2. PROBLEM STATEMENT

Consider the WSN depicted in Fig. 1, comprising L sensors linked with an FC. Each sensor, say the ith one, observes an $N_i \times 1$ vector $\mathbf{x}_i$ that is correlated with a $p \times 1$ random signal of interest $\mathbf{s}$. Through a $k_i \times N_i$ fat matrix $\mathbf{C}_i$ each sensor transmits a compressed $k_i \times 1$ vector $\mathbf{C}_i\mathbf{x}_i$, using e.g., multicarrier modulation with one entry riding per subcarrier. Low-power and bandwidth constraints at the sensors encourage transmissions with $k_i \ll N_i$, while linearity in compression and estimation are well motivated by low-complexity requirements. Furthermore, we assume that:

(a1) *No information is exchanged among sensors, and each sensor-FC link comprises a $k_i \times k_i$ full rank fading multiplicative channel matrix $\mathbf{D}_i$ along with zero-mean additive FC noise $\mathbf{z}_i$, which is uncorrelated with $\mathbf{x}_i$, $\mathbf{D}_i$, and across channels; i.e., noise covariance matrices satisfy $\mathbf{\Sigma}_{z_i z_j} = \mathbf{0}$ for $i \neq j$. Matrices $\{\mathbf{D}_i, \mathbf{\Sigma}_{z_i z_i}\}_{i=1}^L$ are available at the FC.*

(a2) *Data $\mathbf{x}_i$ and the signal of interest $\mathbf{s}$ are zero-mean with full rank auto- and cross-covariance matrices $\mathbf{\Sigma}_{ss}$, $\mathbf{\Sigma}_{sx_i}$ and $\mathbf{\Sigma}_{x_i x_j}$ $\forall\, i, j \in [1, L]$, all of which are available at the FC.*

In multicarrier links, full rank of the channel matrices $\{\mathbf{D}_i\}_{i=1}^L$ is ensured if sensors do not transmit over subcarriers with zero channel gain. Matrices $\{\mathbf{D}_i\}_{i=1}^L$ can be acquired via training, and likewise the signal and noise covariances in (a1) and (a2) can be estimated via sample averaging as usual. With multicarrier (and generally any orthogonal) sensor access, the noise uncorrelatedness

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.

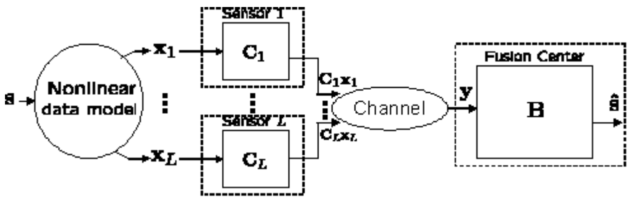


Fig. 1. Distributed setup for estimating a random signal $\mathbf{s}$

across channels is also well justified. Notice that unlike [2, 3, 6, 7], we neither confine ourselves to a linear signal-plus-noise model $\mathbf{x}_i = \mathbf{H}\mathbf{s} + \mathbf{n}_i$, nor we invoke any assumption on the distribution (e.g., Gaussianity) of $\{\mathbf{x}_i\}_{i=1}^L$ and $\mathbf{s}$. Equally important, we do not assume ideal channel links.

Sensors transmit over orthogonal channels so that the FC separates and concatenates the received vectors $\{\mathbf{y}_i(\mathbf{C}_i) = \mathbf{D}_i\mathbf{C}_i\mathbf{x}_i + \mathbf{z}_i\}_{i=1}^L$, to obtain the $(\sum_{i=1}^L k_i) \times 1$ vector:

$$\mathbf{y}(\mathbf{C}_1, \ldots, \mathbf{C}_L) = \text{diag}(\mathbf{D}_1\mathbf{C}_1, \ldots, \mathbf{D}_L\mathbf{C}_L)\mathbf{x} + \mathbf{z}, \tag{1}$$

Left multiplying $\mathbf{y}$ by a $p \times (\sum_{i=1}^L k_i)$ matrix $\mathbf{B}$, we form the linear estimate $\hat{\mathbf{s}}$ of $\mathbf{s}$. For a prescribed power P_i per sensor, our problem is to obtain under (a1)-(a2) MSE optimal matrices $\{\mathbf{C}_i^o\}_{i=1}^L$ and $\mathbf{B}^o$; i.e., we seek:

$$(\mathbf{B}^o, \{\mathbf{C}_i^o\}_{i=1}^L) = \arg \min_{\mathbf{B}, \{\mathbf{C}_i\}_{i=1}^L} E[\|\mathbf{s} - \mathbf{B}\mathbf{y}(\mathbf{C}_1, \ldots, \mathbf{C}_L)\|^2],$$
$$\text{s. to } \text{tr}(\mathbf{C}_i\mathbf{\Sigma}_{x_ix_i}\mathbf{C}_i^T) \leq P_i, \quad i \in \{1, \ldots, L\}. \tag{2}$$

3. DECOUPLED DISTRIBUTED ESTIMATION

Let us consider first the case $\mathbf{\Sigma}_{x_ix_j} \equiv \mathbf{0}$, $\forall i \neq j$, which shows up e.g., when matrices $\{\mathbf{H}_i\}_{i=1}^L$ in the linear model $\mathbf{x}_i = \mathbf{H}_i\mathbf{s} + \mathbf{n}_i$ are mutually uncorrelated and also uncorrelated with the noise vectors $\mathbf{n}_i$. Then, the multi-sensor optimization task in (2) reduces to a set of L decoupled problems. Specifically, it is easy to show that the cost function in (2) can be written as [5]:

$$J(\mathbf{B}, \{\mathbf{C}_i\}_{i=1}^L) = \textstyle\sum_{i=1}^L E[\|\mathbf{s} - \mathbf{B}_i(\mathbf{D}_i\mathbf{C}_i\mathbf{x}_i + \mathbf{z}_i)\|^2] - (L-1)\text{tr}(\mathbf{\Sigma}_{ss}), \tag{3}$$

where $\mathbf{B}_i$ is the $p \times k_i$ submatrix of $\mathbf{B} := [\mathbf{B}_1 \ldots \mathbf{B}_L]$. As the ith non-negative summand depends only on $\mathbf{B}_i, \mathbf{C}_i$ the MSE optimal matrices are given by

$$(\mathbf{B}_i^o, \mathbf{C}_i^o) = \arg \min_{\mathbf{B}_i, \mathbf{C}_i} E[\|\mathbf{s} - \mathbf{B}_i(\mathbf{D}_i\mathbf{C}_i\mathbf{x}_i + \mathbf{z}_i)\|^2],$$
$$\text{s. to } \text{tr}(\mathbf{C}_i\mathbf{\Sigma}_{x_ix_i}\mathbf{C}_i^T) \leq P_i, \; i \in \{1, \ldots, L\}. \tag{4}$$

Since the cost function in (4) corresponds to a single-sensor setup ($L = 1$), we will drop the subscript i for notational brevity and write $\mathbf{B}_i = \mathbf{B}, \mathbf{C}_i = \mathbf{C}, \mathbf{x}_i = \mathbf{x}, \mathbf{z}_i = \mathbf{z}, P = P_i$ and $k = k_i$. The Lagrangian for minimizing (3) can be easily written as:

$$J(\mathbf{B}, \mathbf{C}, \mu) = J_o + \text{tr}(\mathbf{B}\mathbf{\Sigma}_{zz}\mathbf{B}^T) + \mu[\text{tr}(\mathbf{C}\mathbf{\Sigma}_{xx}\mathbf{C}^T) - P] + \text{tr}[(\mathbf{\Sigma}_{sx} - \mathbf{B}\mathbf{D}\mathbf{C}\mathbf{\Sigma}_{xx})\mathbf{\Sigma}_{xx}^{-1}(\mathbf{\Sigma}_{xs} - \mathbf{\Sigma}_{xx}\mathbf{C}^T\mathbf{D}^T\mathbf{B}^T)], \tag{5}$$

where $J_o := \text{tr}(\mathbf{\Sigma}_{ss} - \mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xs})$ is the minimum attainable MMSE for linear estimation of $\mathbf{s}$ based on $\mathbf{x}$. Continuing, we derive a simplified form of (5) the minimization of which will provide closed-form solutions for the MSE optimal matrices $\mathbf{B}^o$ and $\mathbf{C}^o$.

Aiming at this simplification, consider the SVD $\mathbf{\Sigma}_{sx} = \mathbf{U}_{sx}\mathbf{S}_{sx}\mathbf{V}_{sx}^T$, and the eigen-decompositions $\mathbf{\Sigma}_{zz} = \mathbf{Q}_z\mathbf{\Lambda}_z\mathbf{Q}_z^T$ and $\mathbf{D}^T\mathbf{\Sigma}_{zz}^{-1}\mathbf{D} = \mathbf{Q}_{zd}\mathbf{\Lambda}_{zd}\mathbf{Q}_{zd}^T$, where $\mathbf{\Lambda}_{zd} := \text{diag}(\lambda_{zd,1} \cdots \lambda_{zd,k})$ and $\lambda_{zd,1} \geq \cdots \geq \lambda_{zd,k} > 0$. Notice, that $\lambda_{zd,i}$ captures the SNR of the ith entry in the received signal vector at the FC. Further, define $\mathbf{A} := \mathbf{Q}_x^T\mathbf{V}_{sx}\mathbf{S}_{sx}^T\,\mathbf{S}_{sx}\mathbf{V}_{sx}^T\mathbf{Q}_x$ with $\rho_a := \text{rank}(\mathbf{A}) = \text{rank}(\mathbf{\Sigma}_{sx})$, and $\mathbf{A}_x := \mathbf{\Lambda}_x^{-1/2}\mathbf{A}\mathbf{\Lambda}_x^{-1/2}$ with corresponding eigendecomposition $\mathbf{A}_x = \mathbf{Q}_{ax}\mathbf{\Lambda}_{ax}\mathbf{Q}_{ax}$, where $\mathbf{\Lambda}_{ax} = \text{diag}(\lambda_{ax,1}, \cdots, \lambda_{ax,\rho_a}, 0, \cdots, 0)$ and $\lambda_{ax,1} \geq \ldots \geq \lambda_{ax,\rho_a} > 0$. Moreover, let $\mathbf{V}_a := \mathbf{\Lambda}_x^{-1/2}\mathbf{Q}_{ax}$ denote the invertible matrix which simultaneously diagonalizes the matrices $\mathbf{A}$ and $\mathbf{\Lambda}_x$. Since matrices $(\mathbf{Q}_{zd}, \mathbf{Q}_x, \mathbf{V}_a, \mathbf{U}_{sx}, \mathbf{\Lambda}_{zd}, \mathbf{Q}_{zd}, \mathbf{D}, \mathbf{\Sigma}_{zz})$ are all invertible, for every matrix $\mathbf{C}$ (or $\mathbf{B}$) we can clearly find a unique matrix $\mathbf{\Phi}_C$ (correspondingly $\mathbf{\Phi}_B$) that satisfies:

$$\mathbf{C} = \mathbf{Q}_{zd}\mathbf{\Phi}_C\mathbf{V}_a^T\mathbf{Q}_x^T, \quad \mathbf{B} = \mathbf{U}_{sx}\mathbf{\Phi}_B\mathbf{\Lambda}_{zd}^{-1}\mathbf{Q}_{zd}^T\mathbf{D}^T\mathbf{\Sigma}_{zz}^{-1}, \tag{6}$$

where $\mathbf{\Phi}_C := [\phi_{c,ij}]$ and $\mathbf{\Phi}_B$ have sizes $k \times N$ and $p \times k$, respectively. Using (6), the Lagrangian in (5) becomes:

$$J(\mathbf{\Phi}_C, \mu) = J_o + \text{tr}(\mathbf{\Lambda}_{ax}) + \mu(\text{tr}(\mathbf{\Phi}_C\mathbf{\Phi}_C^T) - P) - \text{tr}\left((\mathbf{\Lambda}_{zd}^{-1} + \mathbf{\Phi}_C\mathbf{\Phi}_C^T)^{-1}\mathbf{\Phi}_C\mathbf{\Lambda}_{ax}\mathbf{\Phi}_C^T\right). \tag{7}$$

Applying the well known Karush-Kuhn-Tucker (KKT) conditions (e.g., [1, Ch. 5]) that must be satisfied at the minimum of (7), we prove in [5] that the matrix $\mathbf{\Phi}_C^o$ minimizing (7), is diagonal with diagonal entries:

$$\phi_{c,ii}^o = \begin{cases} \pm\sqrt{\left(\frac{\lambda_{ax,i}}{\mu^o\lambda_{zd,i}}\right)^{1/2} - \frac{1}{\lambda_{zd,i}}}, & 1 \leq i \leq \kappa \\ 0, & \kappa + 1 \leq i \leq k \end{cases}, \tag{8}$$

where κ is the maximum integer in $[1, k]$ for which $\{\phi_{c,ii}^o\}_{i=1}^\kappa$ are strictly positive, or, $\text{rank}(\mathbf{\Phi}_C^o) = \kappa$; and μ^o is chosen to satisfy the power constraint $\sum_{i=1}^\kappa(\phi_{c,ii}^o)^2 = P$ as:

$$\mu^o = \frac{(\sum_{i=1}^\kappa(\lambda_{ax,i}\lambda_{zd,i}^{-1})^{1/2})^2}{(P + \sum_{i=1}^\kappa \lambda_{zd,i}^{-1})^2}. \tag{9}$$

When $k > \rho_a$, the MMSE remains invariant [5]; thus, it suffices to consider $k \in [1, \rho_a]$. Summarizing, we have established that:

Theorem 1: *Under (a1), (a2), and for $k \leq \rho_a$, the matrices minimizing $J(\mathbf{B}_{p\times k}, \mathbf{C}_{k\times N}) = E[\|\mathbf{s} - \mathbf{B}_{p\times k}(\mathbf{D}\mathbf{C}_{k\times N}\mathbf{x} + \mathbf{z})\|^2]$, subject to $\text{tr}(\mathbf{C}_{k\times N}\mathbf{\Sigma}_{xx}\mathbf{C}_{k\times N}^T) \leq P$, are:*

$$\mathbf{C}^o = \mathbf{Q}_{zd}\mathbf{\Phi}_C^o\mathbf{V}_a^T\mathbf{Q}_x^T, \tag{10}$$
$$\mathbf{B}^o = \mathbf{\Sigma}_{sx}\mathbf{Q}_x\mathbf{V}_a\mathbf{\Phi}_C^{o\,T}\left(\mathbf{\Phi}_C^o\mathbf{\Phi}_C^{o\,T} + \mathbf{\Lambda}_{zd}^{-1}\right)^{-1}\mathbf{\Lambda}_{zd}^{-1}\mathbf{Q}_{zd}^T\mathbf{D}^T\mathbf{\Sigma}_{zz}^{-1},$$

where $\mathbf{\Phi}_C^o$ is given by (8), and the corresponding Lagrange multiplier μ^o is specified by (9). The MMSE is

$$J_{\min}(k) = J_o + \sum_{i=1}^{\rho_a}\lambda_{ax,i} - \sum_{i=1}^{k}\frac{\lambda_{ax,i}(\phi_{c,ii}^o)^2}{\lambda_{zd,i}^{-1} + (\phi_{c,ii}^o)^2}. \tag{11}$$

According to Theorem 1, the optimal weight matrix $\mathbf{\Phi}_C^o$ in $\mathbf{C}^o$ distributes the given power across the entries of the pre-whitened vector $\mathbf{V}_a^T\mathbf{Q}_x\mathbf{x}$ at the sensor in a waterfilling-like manner so as to balance channel strength and additive noise variance at the FC with the degree of dimensionality reduction that can be afforded. It is worth mentioning that (8) dictates a minimum power per sensor. Specifically, in order to ensure that $\text{rank}(\mathbf{\Phi}_C^o) = \kappa$ the power must satisfy:

$$P > \frac{\sum_{i=1}^{\kappa}(\lambda_{ax,i}\lambda_{zd,i}^{-1})^{1/2}}{\sqrt{\lambda_{ax,\kappa}\lambda_{zd,\kappa}}} - \sum_{i=1}^{\kappa}\lambda_{zd,i}^{-1}. \qquad (12)$$

The optimal matrices in Theorem 1 can be viewed as implementing a two-step scheme, where: i) we estimate $\mathbf{s}$ based on $\mathbf{x}$ at the sensor using the LMMSE estimate $\hat{\mathbf{s}}_{LM} = \mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{x}$; and ii) compress $\hat{\mathbf{s}}_{LM}$ at the sensor and reconstruct it at the FC using the optimal matrices $\mathbf{C}^o$ and $\mathbf{B}^o$ implied by Theorem 1 after replacing $\mathbf{x}$ with $\hat{\mathbf{s}}_{LM}$. For this estimate-first compress-afterwards (EC) interpretation, we prove in [5] that:

Corollary 1: *For $k \in [1, \rho_a]$, the $k \times N$ matrix in (10) can be written as $\mathbf{C}^o = \hat{\mathbf{C}}^o\mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}$, where $\hat{\mathbf{C}}^o$ is the $k \times p$ optimal matrix obtained by Theorem 1 when $\mathbf{x} = \hat{\mathbf{s}}_{LM}$. Thus, the EC scheme is MSE optimal in the sense of minimizing (3).*

Another interesting feature of the EC scheme implied by Theorem 1 is that the MMSE $J_{\min}(k)$ is non-increasing with respect to the reduced dimensionality k, given a limited power budget per sensor. Specifically, we establish in [5] that:

Corollary 2: *If $\mathbf{C}^o_{k_1\times N}$ and $\mathbf{C}^o_{k_2\times N}$ are the optimal matrices determined by Theorem 1 with $k_1 < k_2$, under the same channel parameters $\lambda_{zd,i}$ for $i = 1, \ldots, k_1$, and common power P, the MMSE in (11) is non-increasing; i.e., $J_{\min}(k_1) \geq J_{\min}(k_2)$ for $k_1 < k_2$.*

4. COUPLED DISTRIBUTED ESTIMATION

In this section, we allow the sensor observations to be correlated. Because $\mathbf{\Sigma}_{xx}$ is no longer block diagonal, decoupling of the multi-sensor optimization problem cannot be effected in this case. The pertinent MSE cost is [c.f. (2)]:

$$J(\{\mathbf{B}_i, \mathbf{C}_i\}_{i=1}^L) = E[\|\mathbf{s} - \textstyle\sum_{i=1}^L \mathbf{B}_i(\mathbf{D}_i\mathbf{C}_i\mathbf{x}_i + \mathbf{z}_i)\|^2]. \qquad (13)$$

Minimizing (13) does not lead to a closed-form solution and incurs complexity that grows exponentially with L [3]. For this reason, we resort to iterative alternatives which converge at least to a stationary point of the cost in (13). To this end, let us suppose temporarily that matrices $\{\mathbf{B}_l\}_{l=1,l\neq i}^L$ and $\{\mathbf{C}_l\}_{l=1,l\neq i}^L$ are fixed and satisfy the power constraints $\text{tr}(\mathbf{C}_l\mathbf{\Sigma}_{x_lx_l}\mathbf{C}_l^T) = P_l$, for $l = 1, \ldots, L$ and $l \neq i$. Upon defining the vector $\bar{\mathbf{s}}_i := \mathbf{s} - \sum_{l=1,l\neq i}^L(\mathbf{B}_l\mathbf{D}_l\mathbf{C}_l\mathbf{x}_l + \mathbf{B}_l\mathbf{z}_l)$ the cost in (13) becomes:

$$J(\mathbf{B}_i, \mathbf{C}_i) = E[\|\bar{\mathbf{s}}_i - \mathbf{B}_i\mathbf{D}_i\mathbf{C}_i\mathbf{x}_i - \mathbf{B}_i\mathbf{z}_i\|^2]\,, \qquad (14)$$

which being a function of $\mathbf{C}_i$ and $\mathbf{B}_i$ only, falls under the realm of Theorem 1. This means that when $\{\mathbf{B}_l\}_{l=1,l\neq i}^L$ and $\{\mathbf{C}_l\}_{l=1,l\neq i}^L$ are given, the matrices $\mathbf{B}_i$ and $\mathbf{C}_i$ minimizing (14) under the power constraint $\text{tr}(\mathbf{C}_i\mathbf{\Sigma}_{x_ix_i}\mathbf{C}_i^T) \leq P_i$ can be directly obtained from (10), after setting $\mathbf{s} = \bar{\mathbf{s}}_i$, $\mathbf{x} = \mathbf{x}_i$, $\mathbf{z} = \mathbf{z}_i$ and $\rho_a = \text{rank}(\mathbf{\Sigma}_{\bar{s}_ix_i})$ in Theorem 1. The corresponding auto- and cross- covariance matrices needed must also be modified appropriately, namely $\mathbf{\Sigma}_{ss} = \mathbf{\Sigma}_{\bar{s}_i\bar{s}_i}$ and $\mathbf{\Sigma}_{sx_i} = \mathbf{\Sigma}_{\bar{s}_ix_i}$. We have thus established the following result for coupled sensor observations:

Theorem 2: *If (a1) and (a2) are satisfied, and $k_i \leq \text{rank}(\mathbf{\Sigma}_{\bar{s}_ix_i})$, then for given matrices $\{\mathbf{B}_l\}_{l=1,l\neq i}^L$ and $\{\mathbf{C}_l\}_{l=1,l\neq i}^L$ satisfying $\text{tr}(\mathbf{C}_l\mathbf{\Sigma}_{x_lx_l}\mathbf{C}_l^T) = P_l$, the optimal $\mathbf{B}_i^o$ and $\mathbf{C}_i^o$ matrices minimizing $E[\|\mathbf{s} - \sum_{l=1}^L \mathbf{B}_l(\mathbf{D}_l\mathbf{C}_l\mathbf{x}_l + \mathbf{z}_l)\|^2]$ are provided by Theorem 1, after setting $\mathbf{x} = \mathbf{x}_i$, $\mathbf{s} = \bar{\mathbf{s}}_i$ and applying the corresponding covariance modifications.*

Theorem 2 suggests the following alternating algorithm for distributed estimation in the presence of fading and FC noise:

Algorithm 1:
Initialize randomly the matrices $\{\mathbf{C}_i^{(0)}\}_{i=1}^L$ and $\{\mathbf{B}_i^{(0)}\}_{i=1}^L$, such that $\text{tr}(\mathbf{C}_i^{(0)}\mathbf{\Sigma}_{x_ix_i}\mathbf{C}_i^{(0)^T}) = P_i$.
for $n = 1, \ldots$
  for $i = 1, \ldots, L$
    Given $\mathbf{C}_1^{(n)}, \mathbf{B}_1^{(n)}, \ldots, \mathbf{C}_{i-1}^{(n)}, \mathbf{B}_{i-1}^{(n)}, \mathbf{C}_{i+1}^{(n-1)}, \mathbf{B}_{i+1}^{(n-1)}, \ldots, \mathbf{C}_L^{(n-1)}, \mathbf{B}_L^{(n-1)}$, determine $\mathbf{C}_i^{(n)}, \mathbf{B}_i^{(n)}$ via Thm. 2.
  end
  If $|\text{MSE}^{(n)} - \text{MSE}^{(n-1)}| < \epsilon$ for a prescribed tolerance ϵ, then stop.
end

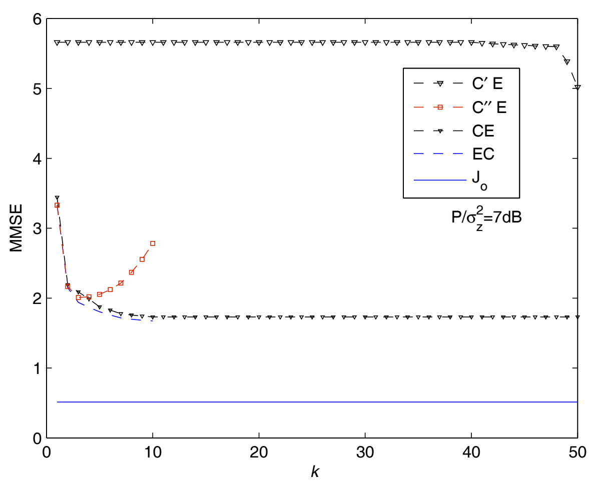
Notice that Algorithm 1 belongs to the class of block coordinate descent iterative schemes. At every step i during the nth iteration, it yields the optimal pair of matrices $\mathbf{C}_i^o, \mathbf{B}_i^o$, treating the rest as given. Thus, the $\text{MSE}^{(n)}$ cost per iteration is non-increasing and the algorithm always converges to a stationary point of (13). Beyond its applicability to possibly non-Gaussian and nonlinear model settings, it is the only available algorithm for handling fading and generally colored FC noise effects in distributed estimation.

5. NUMERICAL RESULTS

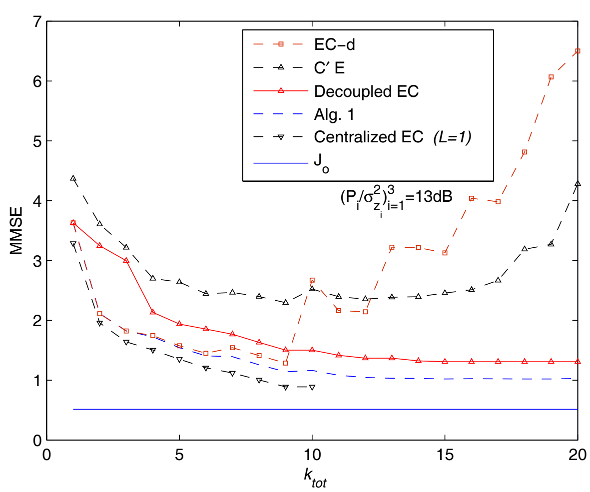
Here we test MMSE performance versus k for the EC scheme and the estimator returned by Algorithm 1. To assess the difference in handling noise effects, we also compare EC and Algorithm 1 with the schemes in [7] and [6], which we abbreviate as C′E and C″E because they perform compression (C) followed by estimation (E). Although C′E and C″E have been derived under ideal link conditions, we modify them here to account for $\mathbf{D}_i$. Our comparisons will further include an option we term CE, which compresses first the data and reconstructs them at the FC using $\mathbf{C}^o$ and $\mathbf{B}^o$ found by (10) after setting $\mathbf{s} = \mathbf{x}$, and then estimates $\mathbf{s}$ based on the reconstructed data vector $\hat{\mathbf{x}}$. For benchmarking purposes, we also plot J_o, achieved when estimating $\mathbf{s}$ based on uncompressed data transmitted over ideal links.

Test Case 1 (EC with uncorrelated sensor data): We consider first the decoupled case of Section 3, where MMSE performance is characterized by the single sensor ($L = 1$) setup. Fig. 2 (a) depicts the MMSE versus k for J_o, EC, CE, C′E and C″E for a linear model $\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}$, where $N = 50$ and $p = 10$. The matrices $\mathbf{H}$, $\mathbf{\Sigma}_{ss}$ and $\mathbf{\Sigma}_{nn}$, are selected randomly such that $\text{tr}(\mathbf{H}\mathbf{\Sigma}_{ss}\mathbf{H}^T)/\text{tr}(\mathbf{\Sigma}_{nn}) = 2$, while $\mathbf{s}$ and $\mathbf{n}$ are uncorrelated. We set $\mathbf{\Sigma}_{zz} = \sigma_z^2\mathbf{I}_k$, and select P such that $10\log_{10}(P/\sigma_z^2) = 7$dB. As expected J_o benchmarks all curves, while the worst performance is exhibited by C′E. Albeit suboptimal, CE comes close to the optimal EC. The monotonic decrease of MMSE with k for EC corroborates Corollary 2. Contrasting it with the increase C″E exhibits in MMSE beyond a certain k, we can appreciate the importance of coping with noise effects. This increase is justifiable since each entry of the compressed data in C″E is allocated a smaller portion of the given power as k grows. In EC however, the quality of channel links and the available power determine the number of the compressed components (which might lie in a vector space of dimensionality $\kappa \leq k$), and allocate power optimally among them.

Test Case 2 (Algorithm 1 with correlated sensor data): Here we consider a 3-sensor setup using the same linear model as in Test Case 1, while setting $N_1 = N_2 = 17$ and $N_3 = 16$. FC noise $\mathbf{z}_i$ is white with variance $\sigma_{z_i}^2$. The power P_i and variance $\sigma_{z_i}^2$ are chosen such that $10\log_{10}(P/\sigma_{z_i}^2) = 13$dB, for $i = 1,2,3$, and the tolerance quantity for the Algorithm 1 is set to $\epsilon = 10^{-3}$. Fig. 2 (b) depicts the MMSE as a function of the total number $k_{tot} = \sum_{i=1}^{3} k_i$ of compressed entries across sensors for: i) a centralized EC setup for which a single (virtual) sensor ($L = 1$) has available the data vectors of all three sensors; ii) the estimator returned by Algorithm 1; iii) the decoupled EC estimator which ignores sensor correlations; iv) the C$'$E and v) an iterative estimator developed in [5], denoted here as EC-d, which similar to C$'$E accounts for fading but ignores noise. Interestingly, our decentralized Algorithm 1 comes very close to the hypothetical single-sensor bound of the centralized EC estimator, while outperforming the decoupled EC one. Also worth noting is that EC-d performs close to Algorithm 1 for small values of k_{tot}, but as k_{tot} increases it behaves as bad as C$'$E.



(a)



(b)

Fig. 2. MMSE comparisons versus k for a centralized, $L = 1$ (a), and a distributed 3-sensor setup (b).

6. CONCLUSIONS

We derived algorithms for estimating stationary random signals based on reduced-dimensionality observations collected by power-limited wireless sensors linked with a fusion center. We dealt with non-ideal channel links that are characterized by multiplicative fading and additive noise. When data across sensors are uncorrelated, we established global mean-square error optimal schemes in closed-form and proved that they implement estimation followed by compression per sensor. Besides distributed estimation with reduced dimensionality decoupled observations, such closed-form solutions are valuable for all applications principal components and canonical correlation analysis are sought in the presence of multiplicative and additive noise.

For correlated sensor observations, we developed an algorithm that relies on block coordinate descent iterations which are guaranteed to converge at least to a local stationary point of the associate mean-square error cost. The optimal estimators allocate properly the prescribed power following a waterfilling-like principle to balance judiciously channel effects and additive noise at the fusion center with the degree of dimensionality reduction that can be afforded. Mean-square error performance of our novel estimators was evaluated both analytically and with numerical examples.[1]

7. REFERENCES

[1] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[2] M. Gastpar, P. L. Draggoti, and M. Vetterli, “The distributed Karhunen-Loève transform,” *IEEE Trans. on Info. Theory*, 2004 (submitted); http://www.eecs.berkeley.edu/∼gastpar/

[3] Z.-Q. Luo, G. B. Giannakis, and S. Zhang, “Optimal linear decentralized estimation in a bandwidth constrained sensor network,” *Proc. of the Intl. Symp. on Info. Theory*, Adelaide, Australia, Sept. 4-9, 2005; http://www.ece.umn.edu/users/luozq/pubs/

[4] A. Ribeiro and G. B. Giannakis, “Bandwidth-constrained distributed estimation for wireless sensor networks, Parts I & II,” *IEEE Trans. on Signal Processing*, 2006 (to appear); http://spincom.ece.umn.edu/

[5] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, “Distributed estimation using reduced dimensionality sensor observations,” *IEEE Trans. on Signal Proc.*, 2005 (submitted); http://spincom.ece.umn.edu/

[6] K. Zhang, X. R. Li, P. Zhang, and H. Li, “Optimal linear estimation fusion–Part VI: Sensor data compression,” *Proc. of the Intl. Conf. on Info. Fusion*, vol. 23, pp. 221–228, Queensland, Australia, 2003.

[7] Y. Zhu, E. Song, J. Zhou, and Z. You, “Optimal dimensionality reduction of sensor data in multisensor estimation fusion,” *IEEE Trans. on Signal Proc.*, vol. 53, pp. 1631–1639, May 2005.

[1]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.