POWER-EFFICIENT DIMENSIONALITY REDUCTION FOR DISTRIBUTED CHANNEL-AWARE KALMAN TRACKING USING WIRELESS SENSOR NETWORKS

Hao Zhu, Ioannis D. Schizas and Georgios B. Giannakis

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

ABSTRACT

Estimation and tracking of nonstationary dynamical processes is of paramount importance in various applications including localization and navigation. The goal of this paper is to perform such tasks in a distributed fashion using data collected at power-limited sensors communicating with a fusion center (FC) over noisy links. For a prescribed power budget, linear dimensionality reducing operators are derived per sensor to account for the sensor-FC channel and minimize the mean-square error (MSE) of Kalman filtered state estimates formed at the FC. Using these operators and state predictions fed back from the FC online, sensors compress their local innovation sequences and communicate them to the FC where tracking estimates are corrected. Analysis and corroborating simulations confirm that the novel channel-aware distributed tracker outperforms competing alternatives.

Index Terms— Distributed tracking, Kalman Filtering

1. INTRODUCTION

With the advent of power-limited wireless sensor networks (WSNs), distributed estimation and tracking of dynamical processes using sensor data processed at a fusion center (FC) has drawn a lot of interest recently. To comply with stringent bandwidth limitations, existing approaches entail sensors communicating their analog or quantized compressed data to the FC where a Kalman filter (KF) is implemented to track the dynamical process of interest [2, 3]. However, sensor-power as well as the sensor-FC channels are not accounted for in these approaches.

In this paper, power-efficient and channel-aware KF-based tracking is derived based on analog-amplitude reduced dimensionality multi-sensor data. Specifically, linear dimensionality reducing operators (matrices) are derived to compress sensor data and minimize the MSE of state estimates at the FC, while adhering to power constraints prescribed at each sensor. An attractive feature of the present scheme is the utilization of feedback from the FC to the sensors which allows them to remove redundant information from their observations and gain in power efficiency; see also [4] where feedback was also advocated (without channel-aware dimensionality reduction) to track an auto-regressive process. The novel approach subsumes as a special case the results reported in [1] for distributed estimation of stationary random signals using reduced-dimensionality sensor observations.

After formulating the problems in Section 2, we develop our reduced-dimensionality KF scheme (Section 3). Further, in Section 3.1 the MSE optimal linear dimensionality reducing operators are specified for a single-sensor setup; whenever the latter is impossible, we develop a block coordinate descent algorithm (Section 3.2). Our theoretical results are corroborated by numerical examples in Section 4.

2. PROBLEM STATEMENT AND PRELIMINARIES

Consider the WSN in Fig. 1 which depicts L sensors \{S_\ell\}_{\ell=1}^L linked with an FC. Each sensor observes an \(N_\ell \times 1\) vector \(x_\ell(n)\) that is correlated with the \(p \times 1\) state process of interest \(s(n)\). The latter obeys the discrete-time vector recursion

\[
s(n) = A(n)s(n - 1) + w(n)
\]

where \(w(n)\) denotes zero-mean additive white Gaussian noise (AWGN) with covariance matrix \(\Sigma_{ww}(n)\). Sensor \(S_\ell\) observes

\[
x_\ell(n) = H_\ell(n)s(n) + v_\ell(n), \quad \ell = 1, \ldots, L
\]

where \(v_\ell(n)\) is zero-mean, uncorrelated (in time), Gaussian observation noise with (cross-)covariance matrix \(\Sigma_{v_\ell v_\ell}(n)\).

Using a \(k_\ell \times N_\ell\) matrix \(C_\ell(n)\) with \(k_\ell \leq N_\ell\), sensor \(S_\ell\) forms the reduced-dimensionality vector \(C_\ell(n)[x_\ell(n) - \hat{x}_\ell(n|n - 1)]\), where \(\hat{x}_\ell(n|n - 1)\) is a vector subtracted from \(x_\ell(n)\) to save transmission power. (It will turn out that \(\hat{x}_\ell(n|n - 1)\) is a predictor of \(x_\ell(n)\) based on data to be specified soon.) Furthermore, we assume that:

(a1) The \(S_\ell\)-to-FC link comprises a \(k_\ell \times k_\ell\) full rank channel matrix \(D_\ell\) along with zero-mean AWGN \(z_\ell(n)\) at the FC with covariance \(\Sigma_{zz_\ell}\), which is uncorrelated with \(s(n)\), \(\{x_\ell(n)\}_{\ell=1}^L\) and across channels. Sensors transmit over orthogonal channels so that the FC can receive separately the vectors

\[
y_\ell(n) = D_\ell C_\ell(n)[x_\ell(n) - \hat{x}_\ell(n|n - 1)] + z_\ell(n), \quad \forall \ell.
\]

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.
(a2) The covariance matrices $\Sigma_{ww}(n)$, $\{\Sigma_{v,v}(n)\}_{\ell=1}^{L}$ as well as matrices $A(n)$ and $\{H_{\ell}(n)\}_{\ell=1}^{L}$ are known at the FC.

Assumption (a1) is satisfied for e.g., multicarrier transmissions where each entry of the compressed vector rides on a subcarrier with nonzero constant channel gain. Matrices $\{D_{\ell}\}_{\ell=1}^{L}$ can be acquired during a training phase, while the matrices in (a2) can be obtained from the physics of the problem or using sample averaging of training samples.

The FC concatenates $y_\ell(n)$, for $\ell = 1, \ldots, L$, to form
\[
y(n) = \text{diag}(D_1C_1(n), \ldots, D_LC_L(n))[x(n) - \hat{x}(n|n-1)] + z(n)
\]
where $x(n) := [x_1^T(n) \ldots x_L^T(n)]^T$ and likewise for the aggregate predictor $\hat{x}(n|n-1)$ and FC noise vector $z(n)$.

We are interested in tracking $s(n)$ using the received data $\{y(k)\}_{k=0}^{n}$. Specifically, we wish to find at the FC the MMSE optimal state estimate $\hat{s}(n|n) = E[s(n)|y(0), \ldots, y(n)]$, recursively (via a KF) based on the observations given by (3). To this end, notice that the filtered state estimate $\hat{s}(n|n)$ can be obtained from the predictor $\hat{s}(n|n-1) = A(n)\hat{s}(n-1|n-1)$ after correcting it using the innovations $\hat{y}(n|n-1) = y(n) - \hat{y}(n|n-1)$, where $\hat{y}(n|n-1) = E[y(n)|y(0), \ldots, y(n-1)];$ i.e., see e.g., [5, Chapter 13]
\[
\hat{s}(n|n) = \hat{s}(n|n-1) + E[\hat{s}(n)|\hat{y}(n|n-1)].
\] (4)

Equation (4) guides the selection of the local predictor at sensor $\hat{s}(n|n-1)$ which is summarized in the following lemma.

**Lemma 1:** If $x_\ell(n|n-1) = E[x_\ell(n)|y(0), \ldots, y(n-1)]$ then $\hat{y}_\ell(n|n-1) = 0$

**Proof:** The predictor of $y_\ell(n)$ based on the past observations can be written as [c.f. (a1)]
\[
E[\hat{D}_\ell C_\ell(n)[x_\ell(n) - \hat{x}_\ell(n|n-1)]|y(0), \ldots, y(n-1)] = \hat{D}_\ell C_\ell(n)[x_\ell(n) - \hat{x}_\ell(n|n-1)] = 0
\]
Q.E.D.

Lemma 1 implies that $\hat{y}_\ell(n|n-1) = y_\ell(n)$. This choice of $x_\ell(n|n-1)$ enables each sensor $S_\ell$ to save considerable power because it transmits the non-redundant local innovation $x_\ell(n) - \hat{x}_\ell(n|n-1)$, which has smaller variance than its raw data $x_\ell(n)$. In addition, this innovation that is received at the FC as $y_\ell(n)$ coincides with $\hat{y}_\ell(n|n-1)$, which is all that the FC needs to implement the correction step (4) of the KF. Since from (2) it holds that $x_\ell(n|n-1) = H_\ell(n)\hat{s}(n|n-1)$, for sensor $S_\ell$ to form $\hat{x}_\ell(n|n-1)$ the FC must feed back the predictor $\hat{s}(n|n-1)$.

To proceed with the dimensionality reduction task, we note that Gaussianity and uncorrelatedness of $z_\ell(n)$ and $v_\ell(n)$ imply that $E[s(n)|\hat{y}(n|n-1)]$ in (4) is a linear function of $\hat{y}(n|n-1) = \hat{y}(n)$, i.e., $E[s(n)|\hat{y}(n|n-1)] = B(n)\hat{y}(n|n-1)$. Furthermore, it follows from (3) and (4) that $\hat{s}(n|n)$ depends on both $C_\ell(n)$ and $B(n)$. Thus, for a prescribed transmit power $P_\ell$ per sensor $S_\ell$ the optimization problem is:

\[
\begin{align*}
\{B(n), \{C_\ell^T(n)\}_{\ell=1}^{L}\} &= \arg \min_{\{B(n), \{C_\ell(n)\}_{\ell=1}^{L}\}} E[|s(n) - \hat{s}(n|n)|^2] \\
\text{s.t. } & \{\text{tr}(C_\ell^T(n)\Sigma_{\hat{x},\hat{x}}(n)C_\ell^T(n)) \leq P_\ell\}_{\ell=1}^{L}
\end{align*}
\] (5)

where $\Sigma_{\hat{x},\hat{x}}(n)$ is the covariance matrix of $\hat{x}_\ell(n|n-1) := x_\ell(n) - \hat{x}_\ell(n|n-1)$ that is compressed at sensor $\ell$.

**3. REDUCED-DIMENSIONALITY KF**

In this section we first outline the steps of the reduced dimensionality distributed tracker at the sensors and FC, and then explain how matrices $\{C_\ell(n)\}_{\ell=1}^{L}$ and $B(n)$ are selected.

Supposing that all quantities at step $n-1$ are available, the FC relies on $\hat{s}(n|n-1)$ to obtain the state predictor
\[
\hat{s}(n|n-1) = A(n)\hat{s}(n|n-1) - 1
\] (6)
and on $M(n|n-1) := E[(s(n|n-1) - \hat{s}(n|n-1))(s(n|n-1) - \hat{s}(n|n-1))]$ to update $M(n|n-1)$, which is the covariance matrix of state innovation $\hat{s}(n|n-1) := s(n|n-1) - \hat{s}(n|n-1)$
\[
M(n|n-1) = A(n)M(n|n-1)A^T(n) + \Sigma_{ww}(n)
\] (7)
With optimal $\{C_\ell(n)\}_{\ell=1}^{L}$ and $B(n)$ assumed determined as described in the ensuing section, the FC feeds back to sensor $S_\ell$ matrix $C_\ell(n)$ and vector $\hat{s}(n|n-1)$. Note that the feedback channel can be assumed reliable since the FC does not have stringent power and communication limitations.

Using the feedback and its local observation, sensor $S_\ell$ forms the innovation $x_\ell(n|n-1)$ and transmits to the FC the reduced-dimensionality vector $C_\ell(n)\hat{x}_\ell(n|n-1)$. The FC receives each $y_\ell(n)$ separately and relies on $y(n)$ to obtain the filtered estimate (corrector)
\[
\hat{s}(n|n) = \hat{s}(n|n-1) + \sum_{\ell=1}^{L} B_\ell(n)\hat{y}_\ell(n|n-1)
\] (8)
where $B_\ell(n)$ represents the columns of matrix $B(n)$ with indices $\{\sum_{m=1}^{\ell-1} k_m + 1\}$ through $\sum_{m=1}^{\ell} k_m$ and $B(n) := \{B_\ell(n) \ldots B_L(n)\}$. (Recall also that $\hat{y}_\ell(n|n-1) = y_\ell(n)$.) Note that $B(n)$ is the gain of the KF based on the reduced-dimensionality data in $y(n)$. Further, as $B(n)$ is chosen so that $\hat{s}(n|n)$ is the LMMSE estimator, based on the orthogonality principle and the linearity of the expected value operator the filtered error covariance matrix (ECM) can be updated...
as
\[
\mathbf{M}(n|n) = E[(\mathbf{s}(n|n - 1) - \sum_{\ell=1}^{L} \mathbf{B}_\ell(n)\mathbf{y}_\ell(n|n - 1))\mathbf{s}(n|n - 1)^T] = [\mathbf{I} - \sum_{\ell=1}^{L} \mathbf{B}_\ell(n)\mathbf{D}_\ell\mathbf{C}_\ell(n)\mathbf{H}_\ell(n)]\mathbf{M}(n|n - 1). \tag{9}
\]

3.1. MSE Optimal Dimensionality-Reducing Matrices
Consider the (cross-) covariance matrices of the state and observation innovations given by
\[
\Sigma_{\tilde{x}_s|x_s}(n) := E[(\mathbf{s}(n|n - 1) - \mathbf{x}_s^T(n|n - 1)]^T \mathbf{M}(n|n - 1) \mathbf{H}_s^T(n)\] (10)
\[
\Sigma_{\tilde{x}_s|x_s}(n) := E[(\mathbf{x}_s(n|n - 1) - \mathbf{H}_s(n)\mathbf{M}(n|n - 1)\mathbf{H}_s^T(n) + \Sigma_{e_{s_s}|e_s}(n). \tag{11}
\]
Focusing first on the solution of (5) for a single sensor, and the Lagrangian function for \( L = 1 \) is
\[
J(B(n), C_1(n), \mu) = J_o(n) + \text{tr}[\Sigma_{\tilde{x}_s|x_s}(n) - \mathbf{B}(n)\mathbf{D}_1 C_1(n)\Sigma_{\tilde{x}_s|x_s}(n)] - \text{tr}[\Sigma_{\tilde{x}_s|x_s}(n)\Sigma_{\tilde{x}_s|x_s}(n) - \mathbf{C}_1(n)] + \mu[\text{tr}[\mathbf{C}_1(n)\Sigma_{\tilde{x}_s|x_s}(n)] - P_1] \tag{12}
\]
where \( \mu \) is the corresponding Lagrange multiplier and \( J_o(n) := \text{tr}[\mathbf{M}(n|n - 1) - \Sigma_{\tilde{x}_s|x_s}(n)\Sigma_{\tilde{x}_s|x_s}(n)] \) is the minimum MSE achieved when estimating \( \mathbf{s}(n|n - 1) \) based on \( \mathbf{x}_s(n|n - 1) \) without channel distortion and additive noise at the FC. Interestingly, the Lagrangian function for determining \( C_1(n) \) and \( B_o(n) \) resembles the formulation in [1], which deals with distributed estimation via reduced-dimensionality observations of stationary signals. The important difference here is that we consider estimation of nonstationary processes.

To simplify (12), consider the SVD \( \Sigma_{\tilde{x}_s|x_s}(n) = \mathbf{U}_{\tilde{x}_s}(n)\mathbf{S}_{\tilde{x}_s}(n)\mathbf{V}_{\tilde{x}_s}^T(n) \), the eigen-decompositions \( \Sigma_{\tilde{x}_s|x_s}(n) = \mathbf{Q}_s(n)\mathbf{A}_s(n)\mathbf{Q}_s^T(n) \), where \( \mathbf{Q}_s(n) := \text{diag} (\lambda_{s,1}(n), \ldots, \lambda_{s,N_s}(n)) \) and \( \mathbf{A}_s(n) := \lambda_{s,1}(n) \mathbf{I} = \ldots \leq \lambda_{s,N_s}(n) \), and \( \mathbf{A}_s(n) = \lambda_{s,1}(n) \mathbf{I} \), where \( \lambda_{s,1}(n) \leq \lambda_{s,2}(n) \leq \ldots \leq \lambda_{s,N_s}(n) \). Notice that \( \lambda_{s,1}(n) \) captures the SNR of the \( n \)-th entry in the received signal vector at the FC. Further, define \( \mathbf{G}(n) := \mathbf{Q}_s^T(n)\mathbf{V}_{\tilde{x}_s}(n)\mathbf{S}_{\tilde{x}_s}(n)\mathbf{Q}_s^T(n)\mathbf{V}_{\tilde{x}_s}^T(n)\mathbf{Q}_s^T(n) \) with \( \rho_g := \text{rank}(\mathbf{G}(n)) = \text{rank}(\Sigma_{\tilde{x}_s|x_s}(n)) \), and \( \mathbf{G}_x(n) := \mathbf{A}_s^{-1/2}(n)\mathbf{G}(n)\mathbf{A}_s^{-1/2}(n) \) with eigen-decomposition \( \mathbf{G}_x(n) = \mathbf{Q}_x(n)\mathbf{A}_x(n)\mathbf{Q}_x^T(n) \), where \( \mathbf{Q}_x(n) := \text{diag}(\lambda_{x,1}(n), \ldots, \lambda_{x,N_x}(n), 0, \ldots, 0) \) and \( \lambda_{x,1}(n) \geq \ldots \geq \lambda_{x,N_x}(n) \). Moreover, let \( \mathbf{V}_y(n) := \mathbf{A}_x^{-1/2}(n)\mathbf{Q}_y(n) \) denote the invertible matrix which simultaneously diagonalizes the matrices \( \mathbf{G}(n) \) and \( \mathbf{A}_x(n) \). Since \( \mathbf{Q}_x(n), \mathbf{V}_y(n), \mathbf{U}_{\tilde{x}_s}(n), \mathbf{A}_x(n), \mathbf{D}_1, \Sigma_{\tilde{x}_s|x_s} \) are all invertible, for every matrix \( C_1(n) \) (or \( B(n) \)) we can clearly find a unique matrix \( \Phi_C \) (corresponding to \( \Phi_B \)) that satisfies:
\[
\mathbf{C}_1(n) = \mathbf{Q}_{2d} \Phi_C \mathbf{V}_y^T(n)\mathbf{Q}_{2d}^T(n) \tag{13}
\]
\[
\mathbf{B}(n) = \mathbf{U}_{\tilde{x}_s}(n) \Phi_B \mathbf{A}_x^{-1} \mathbf{D}_1^T \Sigma_{\tilde{x}_s|x_s}^{-1}
\]
where \( \Phi_C := [\phi_{C,i,j}] \) and \( \Phi_B \) have sizes \( k_1 \times N_1 \) and \( p \times k_1 \), respectively. Using (13), the Lagrangian in (12) becomes
\[
J(\Phi_C, \mu) = J_o(n) + \text{tr}(\mathbf{A}_x\mathbf{A}_x^T) + \mu(\text{tr}(\mathbf{C}_1\Phi_C^T) - P_1)
\]

Applying the well known Karush-Kuhn-Tucker (KKT) conditions (see e.g., [6, Ch. 5]) that must be satisfied at the minimum of (14), we obtain that \( \Phi_C \) minimizing (14), is diagonal with diagonal entries:
\[
\phi_{C,i,i}^o = \begin{cases}
\pm \left( \frac{\lambda_{x,i}(n)\lambda_{s,1}(n)}{\mu^o} \right)^{1/2}, & i = 1, \ldots, \kappa \\
0, & i = \kappa + 1, \ldots, \kappa + k_1 \end{cases}
\]
where \( \kappa \) is the maximum integer in \([1, k_1]\) for which \( \{\phi_{C,i,i}^o\}_{i=1}^\kappa \) are strictly positive; and \( \mu^o \) is the Lagrange multiplier given by
\[
\mu^o = \frac{\left(\sum_{i=1}^\kappa (\lambda_{x,i}(n)\lambda_{s,1}(n))^{1/2}\right)^2}{\text{(15)}}, \tag{16}
\]
Summarizing, we can establish that:
**Proposition 1:** Under (a1), (a2) and for \( L = 1 \) the optimal matrices \( \Phi_C(n) \) and \( B^o(n) \) that minimize (5) are:
\[
\Phi_C(n) = \mathbf{Q}_{2d}\Phi_C^o \mathbf{V}_y^T(n)\mathbf{Q}_{2d}^T(n) \tag{17}
\]
\[
B^o(n) = \mathbf{U}_{\tilde{x}_s}(n)\Phi_B^o \mathbf{A}_x^{-1}\mathbf{D}_1^T\Sigma_{\tilde{x}_s|x_s}^{-1} \tag{18}
\]
where \( \Phi_C^o \) is given by (15), and the corresponding Lagrange multiplier \( \mu^o \) is specified by (16).

The optimal matrices \( \Phi_C(n) \) and \( B^o(n) \) obtained in Proposition 1 are given in closed-form as a function of the state and observation model parameters \( \mathbf{A}(n), \Sigma_{w}(n), \mathbf{H}_s(n) \) and \( \Sigma_{e_{s_s}|e_s} \), as well as the channel matrix \( \mathbf{D}_1 \), the received noise covariance \( \Sigma_{\tilde{x}_s|x_s} \) and the transmit-power \( P_1 \).

Intuitively, the optimal matrix \( \Phi_C(n) \) in Proposition 1 selects the entries of \( \mathbf{x}_s(n|n - 1) \) in which \( s(n) \) is strongest and the channel imperfections weakest, and allocates power among them in a water-filling like manner. Matrix \( B^o(n) \) ensures that \( B^o(n)\mathbf{y}_s(n|n - 1) \) is the MMSE estimate of \( \mathbf{s}(n|n - 1) \) based on the non-redundant innovation \( \mathbf{y}(n|n - 1) \), and is a function of \( \Phi_C(n) \). It is also worth mentioning that (15) dictates a minimum power per sensor. Indeed, in order to ensure that \( \text{rank}(\Phi_C) = \kappa \), we must have \( \mu^o < \lambda_{y,\kappa}(n) \), which implies that the power must satisfy
\[
P_1 > \frac{\sum_{i=1}^{\kappa} (\lambda_{x,i}(n)\lambda_{s,1}(n))^{1/2}}{\lambda_{y,\kappa}(n)} - \sum_{i=1}^{\kappa} \lambda_{s,1}^{-1}. \tag{18}
\]
3.2. Dimensionality-Reducing Matrices for Multi-Sensor

In the multi-sensor scenario it turns out that the minimization problem in (5) does not lead to a closed-form solution, and in general incurs complexity that grows exponentially with \( L \), see e.g., [1]. For this reason, we resort to a block coordinate descent algorithm which converges at least to a stationary point of the cost in (5). Specifically, we suppose temporarily that matrices \( \{B_l(n)\}_{l=1}^L \) and \( \{C_l(n)\}_{l=1}^L \) are fixed and satisfy the power constraints \( \text{tr}(C_l(n)\Sigma_{x_l,\hat{x}_l}(n)C_l^T(n)) = P_l \), for \( l = 1, \ldots, L \) and \( l \neq \ell \). Upon defining

\[
\tilde{s}_l(n) := s(n) - y(n) - 1 - \sum_{l=1, l \neq \ell}^L B_l(n)\tilde{y}_l(n)|n-1|
\]

the cost in (5) can be written as

\[
J(B_l(n), C_l(n)) = E[\|\tilde{s}_l(n) - B_l(n)\tilde{y}_l(n)|n-1]\|^2 \quad (19)
\]

which is a function of \( C_l(n) \) and \( B_l(n) \). Interestingly, (19) falls under the realm of Proposition 1. This means that when \( \{B_l(n)\}_{l=1}^L \) and \( \{C_l(n)\}_{l=1}^L \) are given, the matrices \( B_l(n) \) and \( C_l(n) \) minimizing (19) under the power constraint

\[
\text{tr}(C_l(n)\Sigma_{x_l,\hat{x}_l}(n)C_l^T(n)) \leq P_l
\]

can be directly obtained from (17), after setting \( \tilde{s}_l(n) \) and \( \tilde{y}_l(n) \) in Proposition 1.

In order to apply the corresponding matrix decomposition, we need to update the following covariance matrices

\[
\Sigma_{x_{\ell,\hat{x}_\ell}}(n) := E[\tilde{s}_\ell(n)\tilde{s}_\ell^T(n)]
\]

(20)

\[
\Sigma_{\hat{x}_\ell,\hat{x}_\ell}(n) = E[\tilde{s}_\ell(n)\tilde{y}_\ell^T(n)]
\]

(21)

Based on this setup, we have the following proposition:

**Proposition 2:** Under (a1) and (a2) and for given matrices \( \{B_l(n)\}_{l=1, l \neq \ell}^L \) and \( \{C_l(n)\}_{l=1, l \neq \ell}^L \) satisfying \( \text{tr}(C_l(n)\Sigma_{x_\ell,\hat{x}_\ell}(n)C_l^T(n)) = P_l \), the optimal \( B_l^*(n) \) and \( C_l^*(n) \) matrices minimizing (19) subject to \( \text{tr}(C_l(n)\Sigma_{x_\ell,\hat{x}_\ell}(n)C_l^T(n)) \leq P_l \) are provided by Proposition 1, after setting \( \tilde{s}_\ell(n) = s(n)|n-1 \) \( \hat{y}_\ell(n) = \hat{y}_\ell(n)|n-1 \), and applying the corresponding (cross-)covariance modifications in (20) and (21).

Based on Proposition 2, a block coordinate descent algorithm follows where the FC determines in an alternating fashion (successively for \( \ell = 1, \ldots, L \)) the matrices \( C_l(n) \) and \( B_l(n) \), which are guaranteed to attain at least a stationary point (local optimum) of (5).

**Algorithm 1** Fusion Center: Solving for Optimal Matrices

Initialize randomly the matrices \( \{C_l^{(0)}(n)\}_{l=1}^L \) and \( \{B_l^{(0)}(n)\}_{l=1}^L \) so that \( \text{tr}(C_l^{(0)}(n)\Sigma_{x_l,\hat{x}_l}(n)C_l^{(0)}(n)) = P_l \)

for \( k = 1, \ldots, L \) do

for \( \ell = 1, \ldots, L \) do

Given \( C_l^{(k)}(n), B_l^{(k)}(n), \ldots, C_{l-1}^{(k-1)}(n), B_{l-1}^{(k-1)}(n), C_{l+1}^{(k)}(n), B_{l+1}^{(k)}(n), \) determine \( C_l^{(k+1)}(n), B_l^{(k+1)}(n) \) using Proposition 2.

end for

if \( J((B_l^{(k)}(n))_{l=1}^L, \{C_l^{(k)}(n)\}_{l=1}^L) < J((B_l^{(k-1)}(n))_{l=1}^L, \{C_l^{(k-1)}(n)\}_{l=1}^L) \) then

break

end if

end for

**Remark:** The novel distributed KF using reduced dimensionality sensor data is channel- and power-aware. The MMSE optimal dimensionality-reducing matrices in the single-sensor setup as well as those in the multi-sensor case select the most `informative` entries of \( x_l(n)|n-1 \) and sent them through the most reliable subchannels of \( D_l \). On the other hand, the channel-unaware approach in [3] is challenged by catastrophic error accumulation when AWGN is present at the FC. Further, the feedback link from the FC to sensors enables forming and reducing the dimensionality of the innovation \( \tilde{x}_l(n)|n-1 \) that has smaller dynamic range than \( x_l(n) \), and thus effects power savings. Further, the novel reduced dimensionality KF offers a neat generalization of the stationary results in [1] to nonstationary Markov processes.

The reduced dimensionality KF scheme is summarized in Algorithm 2, which is run by the FC to track the dynamic process \( s(n) \):

**Algorithm 2** Fusion Center: Reception and Estimation

Require: prior estimate \( \hat{s}(0)|0 \) and ECM \( M(0)|0 \)

for \( n = 1, 2, \ldots, \) do

Compute \( \hat{s}(n)|n-1 \) and \( M(n)|n-1 \) using (6) and (7)

Compute optimal matrices \( \{B^*_l(n)\}_{l=1}^L \) and \( \{C^*_l(n)\}_{l=1}^L \) using Algorithm 1

FC transmits to sensor \( S_l \) \( C_l^*(n) \) and \( \hat{s}(n)|n-1 \) for \( \ell = 1, \ldots, L \).

FC receives \( y(n) \) from sensor transmissions

FC computes \( \hat{s}(n)|n \) and \( M(n)|n \) using (8) and (9)

end for

4. SIMULATIONS

Here we test the performance of the reduced-dimensionality KF and compare it with [3]. Consider a WSN deployed for
instance measuring e.g., room temperature. A common state zero-acceleration propagation model is adopted from [2]

\[ s(n) = \left( \begin{array}{c} T(n) \\ \dot{T}(n) \end{array} \right) = \left( \begin{array}{cc} 1 & \alpha \\ 0 & 1 \end{array} \right) s(n-1) + w(n) \]

with \( \alpha = 0.05 \) and covariance matrix

\[ \Sigma_{ww}(n) = \left( \begin{array}{cc} \alpha^3/3 & \alpha^2/2 \\ \alpha^2/2 & \alpha \end{array} \right). \]

The WSN comprises with \( L = 3 \) sensors, and sensor \( S_i \) measures the temperature

\[ x_i(n) = T(n) - \beta \dot{T}(n) + v(n) \]

with \( \beta_1 = 0.1, \beta_2 = 0.2, \beta_3 = 0.3 \) and \( \Sigma_{v_v} = 2 \) for \( \ell = 1, 2, 3 \). Each sensor acquires data vectors of size \( N_{\ell} = 10 \) and reduces them to \( k_\ell = 1 \). Further, the reception noise covariance is set to \( \Sigma_{z_{v_z}} = 2^2 \) for \( \ell = 1, 2, 3 \) so that the SNR := 10 \( \log_{10} P_z/\sigma_z^2 \) is equal to 20 dB.

Figure 2 depicts the estimation MSE along with that of [3] obtained using the trace of \( M(n|n) \) (theoretical), as well as Monte Carlo simulations (empirical). Note that in the novel channel-aware approach the MSE reaches steady-state, while in the channel-unaware approach [3] the MSE eventually diverges due to the accumulation of errors.

Fig. 2. Estimation MSE at FC versus time \( n \).

6. REFERENCES