DISTRIBUTED ESTIMATION USING REDUCED DIMENSIONALITY SENSOR OBSERVATIONS: A SEPARATION PERSPECTIVE

Chao Yu and Gaurav Sharma

Electrical and Computer Engineering Department,
University of Rochester, Rochester NY 14627, USA

ABSTRACT

We consider distributed estimation for a geographically dispersed sensor network, where sensors collect observations that are linearly pre-processed and transmitted over dimensionality-constrained channels. A central processor utilizes the received sensor data to obtain a linear estimate of the desired signal. In this scenario, we consider the optimal pre-processing at the sensors under a mean squared error (MSE) metric. In the single-sensor case, applying a modification of Sakrison's separation principle we show that the optimal pre-processing can be decomposed into two steps: a LMMSE estimate followed by a (linear) MSE optimal dimensionality reduction of the estimate. The latter is readily obtained as the well-known Karhunen-Loéve transform (KLT). Under the multi-sensor scenario, we extend this result to show that given the pre-processing at other nodes, each node's optimal linear pre-processing again reduces to a side-informed linear estimation followed by a side-informed version of the KLT. The separation perspective thus provides a simple and intuitive derivation of the optimal linear pre-processing under reduced dimensionality channels.

Index Terms— Distributed Estimation, Dimensionality Reduction, Separation, Distributed Source Coding,Sensor Networks

1. INTRODUCTION

Wireless sensor networks (WSN) enable a multitude of applications due to their low deployment and maintenance cost. We consider an application scenario of distributed estimation in a WSN where sensors collect and linearly pre-process observations about a desired signal, a central processor (CP) receives these observations and utilizes a linear estimator to estimate the desired signal. We consider specifically scenarios where the channel may be represented by a reduced-dimensionality constraint, a problem that has been previously formulated and addressed in [1, 2, 3]. An optimal linear mean squared error estimate (LMMSE) formulation is presented in [1, 3] where the receiver performs LMMSE estimate of the desired signal using pre-processed observations and the pre-processing is chosen to minimize the resulted MSE. A canonical correlation analysis (CCA) based approach is presented in [2]. The result in [2] is shown to be a two-step estimate-compress (EC) process. Non-ideal links with power constraints are also consider in [2]. Other related work include the Generalized KLT [4] and distributed KLT [5], the former can be mapped to a simplified version of our problem where only one sensor exists in the network, the latter considers a data acquisition model slightly different in which each sensor observes a partial, noiseless fraction of the source signal.

In this paper, we consider the problem of distributed estimation using reduced dimensionality sensor observations from a novel separation perspective inspired by Sakrison's result [6]. For a single-sensor scenario, under an MSE distortion criterion, Sakrison's separation principle [6] (with minor modification) implies that without any MSE penalty the linear pre-processing can be decomposed into two steps, a LMMSE estimation of the desired signal followed by optimal dimensionality reduction. The well-known KLT provides the latter step of optimal dimensionality reduction. Under a multiple sensor scenario, by considering the problem at a single node, we can once again utilize the separation perspective. In this case, the linearly preprocessed observations from other nodes may be treated as *side information* at the CP and an extension of Sakrison's principle in the side-informed scenario [7] applies. By utilizing this separation principle, we can readily show that the optimal linear pre-processing has a similar two-step decomposition of side-informed estimation followed by side-informed dimensionality reduction. Our formulation leads to a more direct and intuitive derivation of results developed in prior work in [1, 2, 3].

2. PROBLEM FORMULATION

The network we consider is depicted in Fig. 1 where N geographically dispersed sensors make observations about a desired signal $\mathbf{x}$, represented as a $k \times 1$ real-valued vector. The i^{th} sensor observation is denoted as a $p_i \times 1$ vector $\mathbf{y}_i$. Each sensing node locally pre-processses its observation data by a linear processing matrix that imposes the dimensionality con-

This work is supported in part by the National Science Foundation under grant number ECS-0428157.

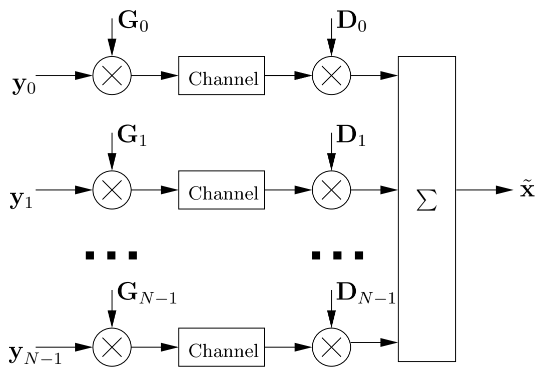

Fig. 1. Estimate $\mathbf{x}$ from observations $\mathbf{y}_0, \mathbf{y}_2, \cdots, \mathbf{y}_{N-1}$. The CP receives reduced dimensionality preprocessed versions of sensor observations and utilizes these in a linear estimator for the desired signal $\mathbf{x}$.

straint. Accordingly, we represent processing at the i^{th} sensor by a $r_i \times p_i$ matrix $\mathbf{G}_i$, where $r_i \leq p_i$. We assume that the channel is otherwise error free so that the pre-processed observations $\mathbf{G}_i\mathbf{y}_i$ are available at the CP. The CP utilizes these pre-processed observations in a linear estimator for $\mathbf{x}$, obtaining $\tilde{\mathbf{x}}$. This process can be represented as shown in Fig. 1, where the estimate $\tilde{\mathbf{x}}$ is obtained by linearly post-processing, by a $k \times r_i$ matrix $\mathbf{D}_i$, the received data from the i^{th} sensor and summing together the resulting values over all the sensors. For notational simplicity, we assume all signals are zero mean, non-zero mean values can be handled by a straightforward extension.

We consider the optimal choice of the pre and post-processing matrices $\mathbf{G}_i, \mathbf{D}_i$ under a MSE distortion metric, i.e.

$$\{\{\mathbf{G}_i^*\}_{i=0}^{N-1}, \{\mathbf{D}^*\}_{i=0}^{N-1}\} = \arg \min_{\{\mathbf{G}_i\}_{i=0}^{N-1}, \{\mathbf{D}\}_{i=0}^{N-1}} E\|\mathbf{x} - \sum_{i=0}^{N-1} \mathbf{D}_i\mathbf{G}_i\mathbf{y}_i\|^2 \quad (1)$$

Where E denotes the expectation operator. We begin by first considering in Section 3 a simplified scenario where only one sensor exists in the network. In Section 4, we consider the problem in the general multi-sensor case.

3. OPTIMAL LINEAR PROCESSING - SINGLE SENSOR CASE

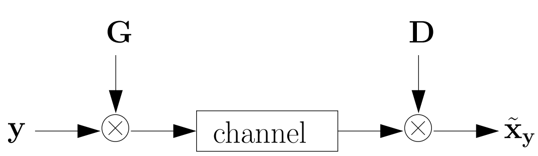

Fig. 2. Single sensor version of the network in Fig. 1

The single sensor realization of the network in Fig. 1 is shown in Fig. 2, where $\mathbf{y}$ is a $p \times 1$ observation vector, $\mathbf{G}$ is a $r \times p$ pre-processing matrix with $r \leq p$, and $\mathbf{D}$ is a $k \times r$ post-processing matrix, and other terms are as defined earlier. For this simplified case, the optimization (1) reduces to:

$$\{\mathbf{G}^*, \mathbf{D}^*\} = \arg \min_{\mathbf{G},\mathbf{D}} E\|\mathbf{x} - \tilde{\mathbf{x}}_{\mathbf{y}}\|^2 \quad (2)$$

where $\tilde{\mathbf{x}}_{\mathbf{y}} = \mathbf{DGy}$ denotes the estimate of $\mathbf{x}$ from received data. In order to obtain the optimal pre and post-processing matrices $\mathbf{G}^*, \mathbf{D}^*$, respectively in (2), we will utilize a minor variant of Sakrison's separation principle [6]. Since the demonstration of this principle is rather straightforward we replicate it here for clarity. We begin by introducing $\hat{\mathbf{x}}$, the LMMSE estimate of $\mathbf{x}$ from $\mathbf{y}$, i.e. [8]

$$\hat{\mathbf{x}}_{\mathbf{y}} = \mathbf{R}_{\mathbf{xy}}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{y} \quad (3)$$

where $\mathbf{R}_{ab} = E[\mathbf{ab}^T]$ denotes the cross-covariance of random vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{R}_a = E[\mathbf{aa}^T]$ denotes the auto-covariance of $\mathbf{a}$ and the presence of inevitable sensor noise assumes invertibility of $\mathbf{R}_{\mathbf{y}}^{-1}$. Now observe that objective function in (2) can be rewritten as:

$$E\|\mathbf{x} - \tilde{\mathbf{x}}_{\mathbf{y}}\|^2 = E\|\mathbf{x} - \hat{\mathbf{x}}_{\mathbf{y}} + \hat{\mathbf{x}}_{\mathbf{y}} - \tilde{\mathbf{x}}_{\mathbf{y}}\|^2 \quad (4)$$

$$= E\|\mathbf{x} - \hat{\mathbf{x}}_{\mathbf{y}}\|^2 + E\|\hat{\mathbf{x}}_{\mathbf{y}} - \tilde{\mathbf{x}}_{\mathbf{y}}\|^2 + 2E < \mathbf{x} - \hat{\mathbf{x}}_{\mathbf{y}}, \hat{\mathbf{x}}_{\mathbf{y}} - \tilde{\mathbf{x}}_{\mathbf{y}} > \quad (5)$$

where $< \mathbf{a}, \mathbf{b} >= \mathbf{a}^T\mathbf{b}$ denotes the inner product of two vectors $\mathbf{a}$ and $\mathbf{b}$. From the orthogonality property of the LMMSE estimators, the prediction error $(\mathbf{x} - \hat{\mathbf{x}}_{\mathbf{y}})$ is orthogonal to any linear function of the predictor inputs $\mathbf{y}$ [8]. Noting that both $\hat{\mathbf{x}}_{\mathbf{y}}$ and $\tilde{\mathbf{x}}_{\mathbf{y}}$ are linear functions of $\mathbf{y}$, we obtain:

$$E < \mathbf{x} - \hat{\mathbf{x}}_{\mathbf{y}}, \hat{\mathbf{x}}_{\mathbf{y}} - \tilde{\mathbf{x}}_{\mathbf{y}} >= 0 \quad (6)$$

Substituting (6) into (5), we obtain:

$$E\|\mathbf{x} - \tilde{\mathbf{x}}_{\mathbf{y}}\|^2 = E\|\mathbf{x} - \hat{\mathbf{x}}_{\mathbf{y}}\|^2 + E\|\hat{\mathbf{x}}_{\mathbf{y}} - \tilde{\mathbf{x}}_{\mathbf{y}}\|^2 \quad (7)$$

Equation (7) represents the adaption of Sakrison's separation principle [6] for the linear estimation and encoding scenario. While we do not require it here, we note that this separation did not explicitly utilize the dimensionality reduction characterization of our channel and applies in general provided the encoding is linear. Note that the first term on the right hand side $E\|\mathbf{x} - \hat{\mathbf{x}}_{\mathbf{y}}\|^2$ represents the mean squared estimation error of the LMMSE estimator for $\mathbf{x}$ from $\mathbf{y}$. This term is independent of the pre and post-processing matrices $\{\mathbf{G}, \mathbf{D}\}$.

Now consider the reduced dimensionality encoding problem shown in Fig. 3, where the estimate $\hat{\mathbf{x}}_{\mathbf{y}}$ is pre-processed by the $r \times k$ pre-processing matrix $\mathbf{G}'$ and a corresponding $k \times r$ post-processing matrix $\mathbf{D}'$ recovers an estimate of $\mathbf{x}$ from the received data $\mathbf{G}'\hat{\mathbf{x}}_{\mathbf{y}}$ at the receiver. The optimal pre and post-processing matrices for this problem are given by

$$\{\mathbf{G}'^*, \mathbf{D}'^*\} = \arg \min_{\mathbf{G}',\mathbf{D}'} E\|\hat{\mathbf{x}}_{\mathbf{y}} - \tilde{\mathbf{x}}_{\mathbf{y}}\|^2. \quad (8)$$

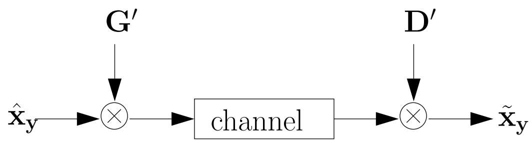

Fig. 3. Optimal reduced dimensionality encoding of the estimate $\hat{\mathbf{x}}_{\mathbf{y}}$

From the preceding observations, the optimization in (2) reduces to:

$$\begin{aligned}\{\mathbf{G}'^{*}, \mathbf{D}'^{*}\} &= \arg\min_{\mathbf{G}',\mathbf{D}'} E\|\hat{\mathbf{x}}_{\mathbf{y}} - \tilde{\mathbf{x}}_{\mathbf{y}}\|^2 \\ &= \arg\min_{\mathbf{G}',\mathbf{D}'} E\|(\mathbf{I} - \mathbf{D}'\mathbf{G}')\hat{\mathbf{x}}_{\mathbf{y}}\|^2 \quad (9)\end{aligned}$$

where $\tilde{\mathbf{x}}_{\mathbf{y}} = \mathbf{D}'\mathbf{G}'\hat{\mathbf{x}}_{\mathbf{y}}$. The objective function in (9) can be rewritten as:

$$\begin{aligned}E\|(\mathbf{I} - \mathbf{D}'\mathbf{G}')\hat{\mathbf{x}}_{\mathbf{y}}\|^2 &= tr((\mathbf{I} - \mathbf{D}'\mathbf{G}')\mathbf{R}^{\frac{1}{2}}_{\hat{\mathbf{x}}_{\mathbf{y}}}\mathbf{R}^{\frac{1}{2}}_{\hat{\mathbf{x}}_{\mathbf{y}}}(\mathbf{I} - \mathbf{D}'\mathbf{G}')^T) \\ &= \|\mathbf{R}^{\frac{1}{2}}_{\hat{\mathbf{x}}_{\mathbf{y}}} - \mathbf{A}\|_F^2 \quad (10)\end{aligned}$$

where $\mathbf{A} = \mathbf{D}'\mathbf{G}'\mathbf{R}^{\frac{1}{2}}_{\hat{\mathbf{x}}_{\mathbf{y}}}$, $\mathbf{R}_{\hat{\mathbf{x}}_y} = \mathbf{R}_{\mathbf{xy}}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{yx}}$ and $\|\cdot\|_F$ denotes the Frobenius norm. Note that rank$(\mathbf{G}') = r$ implies that rank$(\mathbf{A}) \leq r$. Now the problem

$$\mathbf{A}^* = \arg\min_{rank(\mathbf{A})\leq r} \|\mathbf{R}^{\frac{1}{2}}_{\hat{\mathbf{x}}_{\mathbf{y}}} - \mathbf{A}\|_F^2 \quad (11)$$

is the well-known low-rank matrix approximation problem for which the solution is obtained by a truncated SVD [9]. It is readily seen that this matrix approximation provides a corresponding solution to (9) as:

$$\begin{aligned}\mathbf{G}'^{*} &= (\mathbf{Q}^r_{\hat{\mathbf{x}}_{\mathbf{y}}})^T \quad (12)\\ \mathbf{D}'^{*} &= \mathbf{Q}^r_{\hat{\mathbf{x}}_{\mathbf{y}}} \quad (13)\end{aligned}$$

where $\mathbf{Q}^r_{\mathbf{a}}$ denotes the matrix whose r columns are the eigenvectors of $\mathbf{R}_{\mathbf{a}}$ corresponding to the r largest eigenvalues[1].

Next consider the network shown in Fig. 4, clearly this is a specific instance of the network of Fig. 2 and therefore the optimal pre and post-processing matrices for the system of Fig. 2 must offer MSE performance that is no worse than the network in Fig. 4. Conversely, from the separation implied by (7), the optimal pre and post-processing for the network of Fig. 3 achieve the same performance as the optimal pre and post-processing for the network of Fig. 2. It immediately follows that *without any MSE distortion penalty*, the optimal pre-processing for the network in Fig. 2 can be decomposed into two stages as shown in Fig. 4.

From Fig. 4, the optimal linear pre-processing matrix for $\mathbf{y}$ is a concatenation of the two-step estimate-compress process:

$$\mathbf{G}^* = (\mathbf{Q}^r_{\hat{\mathbf{x}}_{\mathbf{y}}})^T\mathbf{R}_{\mathbf{xy}}\mathbf{R}_{\mathbf{y}}^{-1}, \quad (14)$$

[1]These are the equivalently first r principal components for the random vector $\mathbf{a}$.

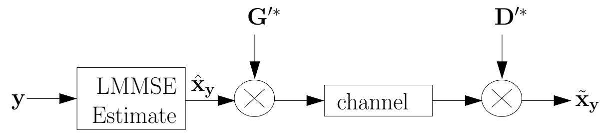

Fig. 4. The decomposed structure for optimal estimation of $\mathbf{x}$: An LMMSE estimate followed by transmission of $\hat{\mathbf{x}}_{\mathbf{y}}$

and $\mathbf{D}^* = \mathbf{D}'^{*}$.

Remark 1: The solution in (12) and (13) is not unique. Given any $r \times r$ invertible matrix $\mathbf{P}$, $\mathbf{P}\mathbf{G}'^{*}, \mathbf{D}'^{*}\mathbf{P}^{-1}$ is another solution for (9).

Remark 2: The result in (14) is equivalent to the result presented in [2]. According to our analysis, the solution presented by CCA can be obtained by a LMMSE estimate followed by a KLT compression. In [2], it was also noted that the optimal linear pre-processing is equivalent to this two-step process. Our use of the separation principle makes it inherent in the development.

Remark 3: Our separation result in (7) represents a slight modification of Sakrison's separation principle [6]. The latter utilizes an MMSE estimate as opposed to our LMMSE estimate, which assures orthogonality of the estimate to any function of the predictor inputs and therefore is applicable to general nonlinear encodings. For the Gaussian case, the MMSE and the LMMSE estimators coincide. Our results, however, apply in general for linear pre and post-processing and do not require the Gaussian assumption.

4. OPTIMAL LINEAR PRE-PROCESSING: MULTI-SENSOR CASE

Unlike the single-sensor scenario considered in the preceding section, for the general multi-sensor problem (1), a closed-form solution is not readily available. We therefore proceed as in prior literature on this problem [1, 2, 3] and consider the optimal choice of $\mathbf{G}_i, \{\mathbf{D}_i\}_{i=0}^{N-1}$ by assuming the values of $\{\mathbf{G}_j, j \in (0, \ldots, i-1, i+1, N-1)\}$ are known. Without loss of generality, we consider $i = 0$. Let $\bar{\mathbf{G}}_0 \stackrel{\text{def}}{=} \text{diag}(\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_{N-1})$ and $\bar{\mathbf{y}}_0 \stackrel{\text{def}}{=} [\ \mathbf{y}_1^T \ \ \mathbf{y}_2^T \ \ \ldots \ \ \mathbf{y}_n^T\ ]^T$, $\bar{\mathbf{D}}_0 \stackrel{\text{def}}{=} \text{diag}(\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_{N-1})$, then

$$\sum_{i=0}^{N-1} \mathbf{D}_i\mathbf{G}_i\mathbf{y}_i = \mathbf{D}_0\mathbf{G}_0\mathbf{y}_0 + \bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0,$$

With this notation, we can represent Fig. 1 in equivalent form as Fig. 5. The optimal $\mathbf{G}_0^*, \mathbf{D}_0^*, \bar{\mathbf{D}}_0^*$ can then be written as:

$$\{\mathbf{G}_0^*, \mathbf{D}_0^*, \bar{\mathbf{D}}_0^*\} = \arg\min_{\mathbf{G}_0,\mathbf{D}_0,\bar{\mathbf{D}}_0} E\|\mathbf{x} - \mathbf{D}_0\mathbf{G}_0\mathbf{y}_0 - \bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0\|^2 \quad (15)$$

Note that if the pre-processing matrices $\bar{\mathbf{G}}_0$ are in fact the optimal choices, the resulting solution $\mathbf{G}_0^*, \mathbf{D}_0^*$ combined with

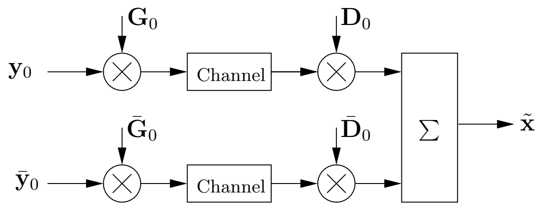

Fig. 5. An equivalent representation of the setup in Fig. 1, where $\bar{\mathbf{y}}_0$ represents all observations other than $\mathbf{y}_0$ and $\bar{\mathbf{G}}_0, \bar{\mathbf{D}}_0$ the corresponding pre and post processing.

$\bar{\mathbf{G}}_0$ in fact represents the solution to (1). The problem of seeking optimal choices of $\mathbf{G}_0, \bar{\mathbf{G}}_0$ simultaneously is computationally intractable [10], an iterative alternating optimization [1, 2, 3] can be utilized to obtain a locally optimal solution. The details are not given here for space consideration.

We demonstrate next that the optimization problem defined as (15) can be reduced into a form identical to (2). Our development in this case can be motivated by the side-informed generalization of Sakrison's separation principle presented in [7]. Once again, though, we utilize a slight modification of the side-informed separation principle applicable to optimal linear pre and post-processing. Let $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}_0$ represent the LMMSE estimate of $\mathbf{x}$ and $\mathbf{y}_0$, respectively, from $\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$. Now we can rewrite the objective function in (15) as:

$$\begin{aligned}
& E\|\mathbf{x} - \mathbf{D}_0\mathbf{G}_0\mathbf{y}_0 - \bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0\|^2 && (16)\\
= \; & E\|\mathbf{x} - \hat{\mathbf{x}} + \hat{\mathbf{x}} - \mathbf{D}_0\mathbf{G}_0\hat{\mathbf{y}}_0 + \mathbf{D}_0\mathbf{G}_0\hat{\mathbf{y}}_0 \\
& \quad -\mathbf{D}_0\mathbf{G}_0\mathbf{y}_0 - \bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0\|^2 && (17)\\
= \; & E\|\mathbf{x} - \hat{\mathbf{x}} + \mathbf{D}_0\mathbf{G}_0\hat{\mathbf{y}}_0 - \mathbf{D}_0\mathbf{G}_0\mathbf{y}_0\|^2 \\
& \quad +E\|\hat{\mathbf{x}} - \mathbf{D}_0\mathbf{G}_0\hat{\mathbf{y}}_0 - \bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0\|^2 && (18)\\
= \; & E\|\mathbf{t} - \mathbf{D}_0\mathbf{G}_0\mathbf{s}\|^2 \\
& \quad +E\|\hat{\mathbf{x}} - \mathbf{D}_0\mathbf{G}_0\hat{\mathbf{y}}_0 - \bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0\|^2 && (19)
\end{aligned}$$

where $\mathbf{s} = \mathbf{y}_0 - \hat{\mathbf{y}}_0$, $\mathbf{t} = \mathbf{x} - \hat{\mathbf{x}}$. The deduction from (17) to (18) relies on two observations resulting from the orthogonality principle which assures that the prediction error is orthogonal to any linear function of the predictor inputs [8]. Firstly we observe

$$(\mathbf{x} - \hat{\mathbf{x}} + \mathbf{D}_0\mathbf{G}_0\hat{\mathbf{y}}_0 - \mathbf{D}_0\mathbf{G}_0\mathbf{y}_0) \perp \bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0, \qquad (20)$$

where $\mathbf{a} \perp \mathbf{b}$ indicates $E(\mathbf{a}^T\mathbf{b}) = 0$, this can be seen from 1) $(\mathbf{x} - \hat{\mathbf{x}}) \perp \bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$, and 2) $(\mathbf{D}_0\mathbf{G}_0\mathbf{y}_0 - \mathbf{D}_0\mathbf{G}_0\hat{\mathbf{y}}_0) \perp \bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$ which is a variant form of $(\mathbf{y}_0 - \hat{\mathbf{y}}_0) \perp \mathbf{G}_0^T\mathbf{D}_0^T\bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$. The second observation enabling the deduction from (17) to (18) is

$$((\mathbf{x} - \mathbf{D}_0\mathbf{G}_0\mathbf{y}_0) - (\hat{\mathbf{x}} - \mathbf{D}_0\mathbf{G}_0\hat{\mathbf{y}}_0)) \perp (\hat{\mathbf{x}} - \mathbf{D}_0\mathbf{G}_0\hat{\mathbf{y}}_0), \quad (21)$$

which can be seen by defining $\mathbf{z} \stackrel{\text{def}}{=} (\mathbf{x} - \mathbf{D}_0\mathbf{G}_0\mathbf{y}_0)$ and noting that the LMMSE estimate of $\mathbf{z}$ from $\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$ is $\hat{\mathbf{z}} = (\hat{\mathbf{x}} - \mathbf{D}_0\mathbf{G}_0\hat{\mathbf{y}}_0)$.

Next we note that for any choice of $\mathbf{G}_0$, the optimal $\mathbf{D}_0$ and $\bar{\mathbf{D}}_0$ in (15) can be obtained as (the partitions of) the LMMSE estimation matrix for $\mathbf{x}$ from $[(\mathbf{G}_0\mathbf{y}_0)^T, (\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0)^T]^T$. At the optimal solution to (15) therefore

$$\mathbf{w} \stackrel{\text{def}}{=} \mathbf{x} - \mathbf{D}_0\mathbf{G}_0\mathbf{y}_0 - \bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0 \qquad (22)$$

represents the LMMSE prediction error about $\mathbf{x}$ from $[(\mathbf{G}_0\mathbf{y}_0)^T, (\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0)^T]^T$. From the orthogonality principle, $\mathbf{w} \perp \bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$. Now, the LMMSE estimate of $\mathbf{w}$ from $\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$ is

$$\hat{\mathbf{w}} = \hat{\mathbf{x}} - \mathbf{D}_0\mathbf{G}_0\hat{\mathbf{y}}_0 - \bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0, \qquad (23)$$

since $\mathbf{w} \perp \bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$, we have $\hat{\mathbf{w}} = \mathbf{0}$ and therefore $E\|\hat{\mathbf{w}}\|^2 = E\|\hat{\mathbf{x}} - \mathbf{D}_0\mathbf{G}_0\hat{\mathbf{y}}_0 - \bar{\mathbf{D}}_0\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0\|^2 = 0$ at the optimal solution of (15). It follows that

$$\{\mathbf{G}_0^*, \mathbf{D}_0^*\} = \arg\min_{\mathbf{G}_0, \mathbf{D}_0} E\|\mathbf{t} - \mathbf{D}_0\mathbf{G}_0\mathbf{s}\|^2 \qquad (24)$$

Eq. (24) is identical in form to (2). The separation principle is now applicable: in the presence of the side information $\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$ at the receiver, $\mathbf{s}$ plays the role of $\mathbf{y}$ from (2) and $\mathbf{t}$ the role of $\mathbf{x}$ from (2). Proceeding as we did for obtaining (13)-(14), we have the optimal choice of $\{\mathbf{G}_0, \mathbf{D}_0\}$:

$$\begin{aligned}
\mathbf{G}_0^* &= (\mathbf{Q}_{\hat{\mathbf{t}}_\mathbf{s}}^{r_0})^T \mathbf{R}_{\mathbf{ts}}\mathbf{R}_\mathbf{s}^{-1} && (25)\\
\mathbf{D}_0^* &= \mathbf{Q}_{\hat{\mathbf{t}}_\mathbf{s}}^{r_0} && (26)
\end{aligned}$$

where $\mathbf{Q}_{\hat{\mathbf{t}}_\mathbf{s}}^{r_0}$ is defined similarly as in (12), $\mathbf{t}, \mathbf{s}$ are defined in (19), and $\mathbf{R}_{\hat{\mathbf{t}}_\mathbf{s}} = \mathbf{R}_{\mathbf{ts}}\mathbf{R}_\mathbf{s}^{-1}\mathbf{R}_{\mathbf{st}}$. Note that $\mathbf{R}_{\mathbf{ts}}\mathbf{R}_\mathbf{s}^{-1}$ corresponds to a "side-informed" estimate for $\mathbf{x}$ given $\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$ and $\mathbf{Q}_{\hat{\mathbf{t}}_\mathbf{s}}^{r_0}$ corresponds to a "side-informed" KLT.

Remark 4: We observe that the estimation process at the CP in Fig. 5 can be viewed as a two-stage Kalman filter [8] with the static state vector $\mathbf{x}$ and two observations $\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0, \mathbf{G}_0\mathbf{y}_0$ in sequence. This Kalman filter achieves the LMMSE estimate $\hat{\mathbf{x}}$ from $\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$ in the first step, then updates its estimation based on the new observations $\mathbf{G}_0\mathbf{y}_0$:

$$\tilde{\mathbf{x}} = \hat{\mathbf{x}} + \mathbf{D}_0(\mathbf{G}_0\mathbf{y}_0 - \mathbf{G}_0\hat{\mathbf{y}}_0) \qquad (27)$$

$\mathbf{D}_0$ is the Kalman gain matrix. The MSE can be written as $E\|\mathbf{t} - \mathbf{D}_0\mathbf{G}_0\mathbf{s}\|^2$, where $\hat{\mathbf{x}}, \hat{\mathbf{y}}_0, \mathbf{t}, \mathbf{s}$ are defined earlier. We observe that $\mathbf{s}$ can be thought of as the *innovation* in $\mathbf{y}_0$ given $\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$. Similarly, $\mathbf{t}$ can be viewed as the *residue* of $\mathbf{x}$ given $\bar{\mathbf{G}}_0\bar{\mathbf{y}}_0$.

5. CONCLUSIONS

By viewing the problem of distributed estimation with reduced-dimensionality sensor observations from a novel separation perspective motivated by Sakrison's work [6], simple and intuitive derivations are obtained for known results on the optimal dimensionality reduction pre-processing. For the single-sensor case, the separation principle implies that, without incurring any MSE penalty, the optimal linear preprocessing can be decomposed into a LMMSE estimate followed by a MSE optimal dimensionality reduction, which is achieved by KLT. An extension to the multi-sensor case suggests a similar two-step decomposition: a side-informed LMMSE estimation followed by a side-informed version of KLT at each sensor node.

6. REFERENCES

[1] K. Zhang, X. Li, P. Zhang, and H. Li, "Optimal linear estimation fusion-part VI: sensor data compression," *Proc. of the Sixth Intl. Conf. of Information Fusion*, vol. 1, 2003.

[2] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed Estimation Using Reduced Dimensionality Sensor Observations," *IEEE Trans. Signal Proc.*, vol. 55, pp. 4284 – 4299, Aug 2007.

[3] E. Song, Y. Zhu, and J. Zhou, "Sensors' optimal dimensionality compression matrix in estimation fusion," *Automatica*, vol. 41, no. 12, pp. 2131–2139, 2005.

[4] Y. Hua and W. Liu, "Generalized Karhunen-Loeve transform," *IEEE Sig. Proc. Lett.*, vol. 5, no. 6, pp. 141–142, 1998.

[5] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The Distributed Karhunen-Loeve Transform," *IEEE Trans. Info. Theory*, vol. 52, no. 12, pp. 5177–5196, Dec 2006.

[6] D. Sakrison, "Source encoding in the presence of random disturbance," *IEEE Trans. Info. Theory*, vol. 14, pp. 165–167, 1968.

[7] C. Yu and G. Sharma, "On side-informed coding of noisy sensor observations," in *Proc.* 41^{st} *Asilomar Conf. on Signals, Systems & Computers*, Nov. 2007.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall, 1993.

[9] B. Noble and J. W. Daniel, *Applied Linear Algebra*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1977.

[10] Z. Luo, G. Giannakis, and S. Zhang, "Optimal linear decentralized estimation in a bandwidth constrained sensor network," *Proc. of the Intl. Symp. on Information Theory*, pp. 1441–1445, 2005.