Dimensionality Reduction for Distributed Estimation in the Infinite Dimensional Regime

Olivier Roy, *Student Member, IEEE*, and Martin Vetterli, *Fellow, IEEE*

***Abstract*—Distributed estimation of an unknown signal is a common task in sensor networks. The scenario usually envisioned consists of several nodes, each making an observation correlated with the signal of interest. The acquired data is then wirelessly transmitted to a fusion center that aims at estimating the desired signal within a prescribed accuracy. Motivated by the obvious processing limitations inherent to such distributed infrastructures, we seek to find efficient compression schemes that account for limited available power and communication bandwidth. In this paper, we propose a transform-based approach to this problem where each sensor provides the fusion center with a low-dimensional approximation of its local observation by means of a suitable linear transform. Under the mean squared error criterion, we derive the optimal solution to apply at one sensor assuming all else being fixed. This naturally leads to an iterative algorithm whose optimality properties are exemplified using a simple though illustrative correlation model. The stationarity issue is also investigated. Under restrictive assumptions, we then provide an asymptotic distortion analysis, as the size of the observed vectors becomes large. Our derivation relies on a variation of the Toeplitz distribution theorem, which allows us to provide a reverse "water-filling" perspective to the problem of optimal dimensionality reduction. We illustrate, with a first-order Gauss–Markov model, how our findings allow for the computation of analytical closed-form distortion formulas that provide an accurate estimation of the reconstruction error obtained in the finite-dimensional regime.**

***Index Terms*—Asymptotic eigenvalue distribution, distributed approximation and estimation, Karhunen–Loève transform (KLT), large Toeplitz matrices, principal component analysis, Toeplitz distribution theorem.**

Manuscript received August 16, 2006; revised November 23, 2007. This work was supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under Grant 5005-67322. The material in this paper was presented in part at the Allerton Conference on Communication, Control and Computing, Monticello, IL, September 2005, and at the IEEE International Symposium on Information Theory, Seattle, WA, July 2006. This paper is reproducible, meaning that all the results presented here can be reproduced using the code and data available at http://rr.epfl.ch.

O. Roy is with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: olivier.roy@epfl.ch).

M. Vetterli is with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland and also with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: martin.vetterli@epfl.ch).

Communicated by A. Høst-Madsen, Associate Editor for Detection and Estimation.

Color versions of Figs. 2, 3, and 5–8 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2008.917635

I. Introduction

Wireless sensor networks comprising a multitude of battery-operated sensing devices have recently emerged as a promising technology for a wide range of applications [2]. Typical examples are the reconstruction of a tridimensional scene by means of multiview images acquired with a network of cameras [3], acoustic beamforming using an array of microphones [4], or environmental monitoring by way of nodes measuring temperature, light or pressure variations [5]. All these scenarios involve a set of acquiring devices distributed over a large area and making local observations correlated with a signal (or field) of interest [6]. Each sensor then transmits information about its measurements to a fusion center whose primary goal is to merge the received data in order to recover the original signal within a prescribed accuracy. In this context, efficient transmission strategies should reduce the amount of information conveyed to the fusion center to account for the stringent power consumption and communication bandwidth limitations imposed by such wireless architectures.

In this paper, we adopt a transform-based approach to the aforementioned problem. More precisely, each sensor (or terminal) applies a suitable linear transform to its random observation vector so as to reduce its dimensionality. The fusion center then reconstructs the random source vector of interest by appropriately combining the low-dimensional approximations provided by the terminals. The case with two terminals is depicted in Fig. 1. Under the mean squared error fidelity criterion considered in this paper, we seek to find the optimal transform matrix to be applied at each terminal, accounting for correlation across the observations and the fact that intersensor communication is precluded. Under these assumptions, we derive the optimal strategy at one terminal assuming all else being fixed. This local perspective suggests an iterative algorithm, referred to as Algorithm 1, that is proved to converge to a stationary point of the underlying cost function. The possible suboptimality of this distributed algorithm is illustrated with a simple yet insightful example. We then prove that, if the involved signals exhibit some stationarity properties, it is possible to provide an asymptotic analysis of the distortion incurred by our distributed transform, as the size of the vectors becomes large. This aims at taking into account the correlation structure of the sources across space or time. To this end, we define the infinite block-length distortion and demonstrate how a variation of the Toeplitz distribution theorem [7]–[9] allows us to compute the distortion in closed form. The optimality results provided in the finite-dimensional regime are then extended to the infinite case and an asymptotic counterpart to Algorithm 1, referred to as Algorithm 2, is given. We

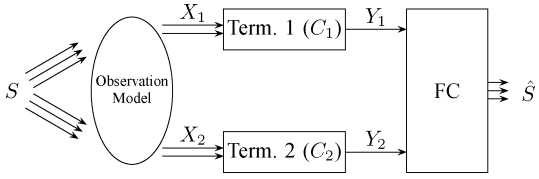

Fig. 1. Block diagram of the distributed approximation and estimation problem with two terminals. Terminal l observes a vector X_l correlated with the signal of interest S and provides the FC with a k_l-dimensional approximation $Y_l = C_l X_l$ of the observed vector ($l = 1, 2$). Based on the information received, the FC computes the reconstruction $\hat{S}$.

also show, with a first-order Gauss–Markov correlation model, how the infinite-dimensional distortion analysis considered in this paper allows us to derive closed-form analytical formulas of the infinite block-length distortion for various scenarios of interest. A crucial observation is that, in this example, our asymptotic analysis provides an excellent approximation of the distortion obtained in the finite dimensional regime already for small block lengths, highlighting the practicality of our approach.

In the literature, the problem of distributed estimation with communication constraints has been studied by various researchers over the past decade (see, for example, [10]–[14]). While our asymptotic analysis appears to be new, the transform-based approach considered here in the finite-dimensional regime is equivalent to the linear estimation fusion problem investigated by Zhang *et al.* in [15] and [16]. Our solution is, however, different from theirs and allows us to conveniently describe the optimal architecture as a two-stage process. Recently, and in parallel to our work, Schizas *et al.* [17] extended the results of [15] and [16] to the important case of nonideal communications links (fading and noise) and non-Gaussian signals. While of great interest, this generality will not be considered here since we chiefly aim at conveying the basic ideas behind our asymptotic study. Our construction builds upon the distributed Karhunen–Loève transform (KLT) devised by Gastpar *et al.* [13]. It is, however, slightly more general as the vectors of interest do not need to be directly observed at the sensors. These transforms extend the concept of principal component analysis [18], [19] to a distributed setup. As shown in [13], they also enjoy some optimality properties in the compression (rate-distortion) framework [20]. They thus provide a transform coding perspective [21] to the (remote) multiterminal source coding problem whose general solution remains unknown to date. Moreover, various optimality results can be proved in the high-rate regime, as demonstrated by Rebollo-Monedero *et al.* in [22] and [23]. From a more practical viewpoint, they have been successfully applied, for example, to distributed audio and video coding [24]–[26].

The outline of this paper is as follows. In Section II, we present the distributed approximation and estimation problem. Section III offers a local perspective to the general optimization task by means of an iterative algorithm. The optimality of the proposed scheme is also investigated and some stationarity issues are discussed. A distortion analysis in the infinite-dimensional regime is then provided in Section IV. Using these findings, we derive closed-form analytical distortion formulas for a simple correlation model in Section V. Section VI concludes this paper.

II. Distributed Approximation and Estimation Problem

The considered setup consists of L terminals and a fusion center (FC). Terminal l ($l = 1, 2, \ldots, L$) observes a vector $X_l \in \mathbb{C}^{n_l}$ correlated with a random vector $S \in \mathbb{C}^m$ that needs to be estimated at the FC. For simplicity, we will assume that $n_l \leq m$ for all l. Limited by communication constraints, terminal l provides the FC with a k_l-dimensional approximation Y_l of its observed vector X_l ($k_l \leq n_l$). This is achieved by means of the linear transformation

$$Y_l = C_l X_l \tag{1}$$

where C_l is a transform matrix of size $k_l \times n_l$. The FC then computes the reconstruction $\hat{S}$ based on the vectors provided by the terminals. For ease of notation, we will write $X = (X_1^T, X_2^T, \ldots, X_L^T)^T \in \mathbb{C}^n$ and $Y = (Y_1^T, Y_2^T, \ldots, Y_L^T)^T \in \mathbb{C}^k$ with $n = \sum_{l=1}^{L} n_l$ and $k = \sum_{l=1}^{L} k_l$. The involved signals are assumed to be zero-mean jointly Gaussian random vectors with (cross-)covariance matrices known at both the sensors and the FC. In the sequel, R_U denotes the covariance matrix of a vector U, R_{UV} denotes the cross-covariance matrix between U and V, and $R_{U|V}$ denotes the covariance matrix of the prediction error $U - \mathrm{E}[U|V]$, where $\mathrm{E}[\cdot|\cdot]$ denotes the conditional expectation. Under the mean squared error (MSE) criterion considered in this paper, the optimal reconstruction $\hat{S}$ is given by the conditional expectation of S given Y. In the jointly Gaussian case, this can be simply expressed as [27, Sec. IV.B]

$$\hat{S} = \mathrm{E}[S|Y] = \mathrm{E}[S|CX] = R_{SX} C^* (C R_X C^*)^{-1} C X \tag{2}$$

where * denotes the conjugate (Hermitian) transpose and C is the block diagonal matrix with diagonal blocks $C_1, C_2, \ldots, C_L$. Note that, in the sequel, the involved covariance matrices are assumed to be invertible. The singular case can be handled similarly using the modifications explained in Appendix I. For given k_l's, the goal is thus to find the block diagonal transform C that minimizes the mean-squared estimation error [27, Sec. IV.B]

$$\mathrm{E}[\|S - \hat{S}\|^2] = \mathrm{tr}\left(R_S - R_{SX} C^* (C R_X C^*)^{-1} C R_{SX}^*\right) \tag{3}$$

where $\|\cdot\|$ stands for the Euclidean norm and tr denotes the trace operator. A solution to this optimization problem seems difficult to derive because the optimal matrix C is constrained to have a particular block-diagonal form that reflects the distributed nature of the setup. In the next section, we investigate a local perspective to the above minimization task where the transforms C_l are optimized in turn at each terminal. Moreover, the benefits, tradeoffs, and complexity issues that are inherent to this distributed estimation problem are fully captured by a two-terminal scenario ($L = 2$) as illustrated in Fig. 1. For simplicity of exposure, we will thus concentrate on this setup for the rest of the discussion. The case where $L > 2$ can be similarly treated by fixing a group of $L - 1$ terminals while optimizing the remaining one.

III. Local Perspective in the Finite-Dimensional Regime

We provide a local perspective to the problem illustrated in Fig. 1. In the first part, we derive the optimal transform to be applied at one terminal assuming the other is fixed. We then describe an iterative algorithm (Algorithm 1) based on this result and discuss its optimality. In the second part, we address some stationarity issues which are of particular interest in this discussion.

A. Local Optimality

Let us assume that one of the terminals (say terminal 2) provides a low-dimensional approximation $Y_2 = C_2 X_2$ of its local observation and that we wish to find an optimal transform C_1 that needs to be applied at terminal 1. The solution to this problem is characterized by the following theorem.[1]

Theorem 3.1: The optimal transform $C_1 \in \mathbb{C}^{k_1 \times n_1}$ is given by

$$C_1 = \bar{C}_1^* R_{S\bar{X}_1} R_{\bar{X}_1}^{-1} \tag{4}$$

where $\bar{C}_1 \in \mathbb{C}^{m \times k_1}$ is the matrix whose columns are the k_1 eigenvectors of the matrix $R_S - R_{S|\bar{X}_1}$ corresponding to the k_1 largest eigenvalues and where $\bar{X}_1 = X_1 - \mathrm{E}[X_1|Y_2]$ with $Y_2 = C_2 X_2$. The resulting minimum mean squared error (MMSE) is computed as

$$\mathrm{E}[\|S - \hat{S}\|^2] = \mathrm{tr}(R_{S|X_1,Y_2}) + \sum_{i=1}^{n_1 - k_1} \lambda_i \tag{5}$$

where λ_i denote the n_1 largest eigenvalues of $R_{S|Y_2} - R_{S|X_1,Y_2}$ arranged in increasing order.

Proof: See Appendix II. □

Theorem 3.1 shows that the optimal transform amounts to first compute the best estimate of S (Wiener filtering) as if the error vector $\bar{X}_1$ (or, equivalently, the side information Y_2) were available at the encoder and then simply apply a KLT on this estimate. In other words, it says that we should send the part of S that can be predicted by X_1 but not by Y_2. Furthermore, it is seen in the proof of Theorem 3.1 that the availability of the side information Y_2 at both the encoder and the decoder results in the same MMSE. In particular, the minimum distortion $\mathrm{tr}(R_{S|X_1,Y_2})$ in (5) corresponds to the part of S that can be estimated neither by X_1 nor by Y_2, while the sum corresponds to the subspace approximation error. Note that here, k_1 and k_2 are assumed to be fixed. One may investigate how a total budget k can be optimally shared among the two terminals by considering the $k+1$ possible pairs (k_1, k_2) satisfying $k_1 + k_2 = k$.

Similarly to the iterative algorithms developed in [15], [17], and [13], the local perspective offered by Theorem 3.1 suggests an iterative approach to the quest of an optimal distributed transform coding architecture. Namely, terminals 1 and 2 both select an arbitrary initial transform matrix of size $k_1 \times n_1$ and $k_2 \times n_2$, respectively.[2] In turn, each terminal then updates its transform as described by Theorem 3.1 . Note that Theorem 3.1 is stated from the perspective of terminal 1. It should be clear that an optimal transform for terminal 2 is obtained by simply exchanging the role of the two terminals. This allows us to replace the original bidimensional block-component optimization problem by an iterative one-dimensional block-component minimization task for which conclusive results can be found. The algorithm can be described as follows.

Algorithm 1

Select a fixed tolerance ε and initial matrices $C_1^{(0)}$ and $C_2^{(0)}$ of size $k_1 \times n_1$ and $k_2 \times n_2$, respectively.

while (true)

  for $l = 1, 2$

    Compute $C_l^{(s)}$ as given by Theorem 3.1.

  end

  If $|\mathrm{MSE}^{(s)} - \mathrm{MSE}^{(s-1)}| < \varepsilon$ then stop.

  $s = s + 1$;

end

The outcome of this iterative algorithm is, however, not guaranteed to be a global minimum of the cost function but just a stationary point that is either a local minimum or a saddle point. This property follows straightforwardly from [13, Th. 9]. To further illustrate this fact, we will now consider a simple numerical example.

Example 3.2: Suppose that the source S is a Gaussian random vector with mean zero and covariance matrix

$$R_S = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

for some correlation parameter $\rho \in (0, 1)$. The observed vectors X_1 and X_2 are noisy version of the source, i.e., $X_1 = S + N_1$ and $X_2 = S + N_2$, where N_i is a Gaussian random vector with mean zero and covariance matrix $R_{N_i} = \sigma^2 I_2$ $(i = 1, 2)$. Here, I_2 denotes the identity matrix of size 2×2. The vectors S, N_1, and N_2 are assumed to be independent. Both terminals are required to provide a one-dimensional approximation of their observation. Because the performance is invariant under scaling, the transforms applied at the terminals may be parameterized as $C_1 = (\cos(\theta)\ \sin(\theta))$ and $C_2 = (\cos(\vartheta)\ \sin(\vartheta))$ for some scalars θ and ϑ. We plot in Fig. 2(a) the distortion surface obtained for $\rho = 0.5$ and $\sigma^2 = 1$. A top view of a portion of the optimization surface is depicted in Fig. 2(b). We observe that the point (θ_0, ϑ_0) corresponds to a local minimum along both (optimization) directions θ and ϑ, but is, however, a saddle point of the overall cost function. Algorithm 1 will thus stop if it reaches

[1] By construction, an optimal transform for this distributed approximation and estimation problem inherits the nonuniqueness of the KLT in the centralized scenario [18], [19]. The optimal solution is thus not unique in general.

[2] Strictly speaking, only one terminal needs to be initialized in a two-terminal setup. The initial transform of the other terminal is computed at the first iteration.

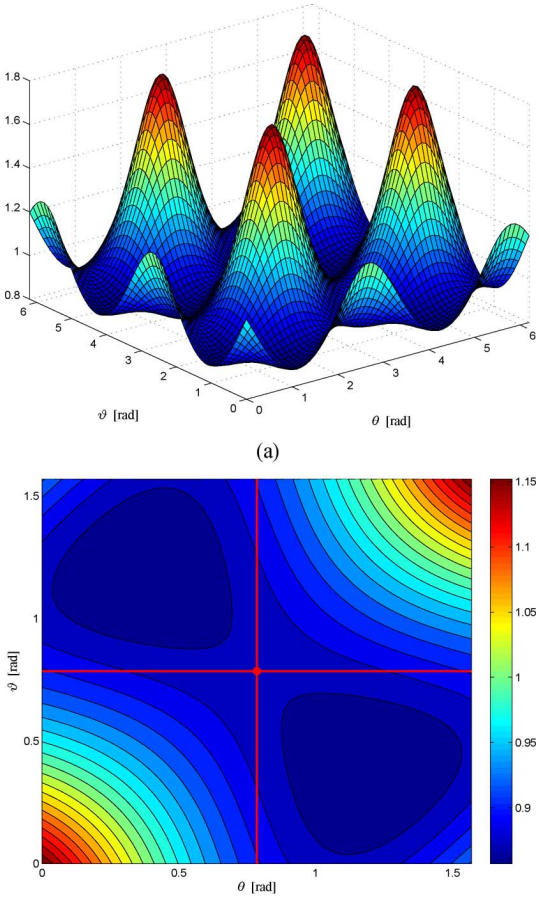

Fig. 2. Distortion surface of Example 3.2 with $\rho = 0.5$ and $\sigma^2 = 1$. (a) Cost function to be minimized. (b) Top view of a portion of the optimization surface. The point (θ_0, ϑ_0) is a local minimum along both the optimization directions θ and ϑ (indicated by the horizontal and vertical lines) but only a saddle point of the overall cost function.

this point, yielding a suboptimal solution in this case. Probabilistic methods, such as simulated annealing [28], may be used to increase the probability of reaching the global optimum. A thorough exposure of such techniques is, however, beyond the scope of this work.

B. Stationarity

Another optimality issue that arises with the iterative algorithm described previously, and which is of central interest in our discussion, is the question of stationarity. Note that here, the term stationarity refers either to space or time stationarity since both can be handled equivalently. Most signals encountered in practice do not exhibit such stationarity properties. However, in many practical applications (e.g., multichannel audio coding [29]), the signals are processed on a frame-by-frame basis so as to reduce the computational complexity. Every frame is then assumed to be the samples of a real, periodic, and time stationary random process such that the involved covariance matrices are (Hermitian) circulant. In this case, the decorrelating transform is the discrete Fourier transform (DFT) [8, Sec. 3.1]. Moreover, the (real) eigenvalues are simply obtained by taking the DFT of the first row of such matrices. The aforementioned practical motivations suggest that we analyze the proposed distributed transform coding algorithm with processes that exhibit additional stationary properties. Because the next section specifically addresses the distortion behavior in the limit of large vector size, the finite-dimensional analysis provided here will further assume that the source and observed vectors are of the same length ($m = n_1 = n_2$) and that the (cross-)covariance matrices involving S, X_1, or X_2 are circulant. This aims at analyzing the important remote sensing scenario where the observations are filtered (with a linear and time-invariant filter) and noisy versions of the vector of interest. Returning to Algorithm 1, we observe that the MSE incurred at each step (Theorem 3.1) can be computed using the eigenvalues of some carefully chosen covariance matrices, namely, $R_{S|Y_2}$ and $R_{S|X_1,Y_2}$. In the next proposition, we show that, with proper initialization, the iterations of Algorithm 1 will only involve circulant matrices, hence making the distortion at each step analytically tractable.

Proposition 3.3: Assume that the (cross-)covariance matrices involving S, X_1, and X_2 are circulant and that the initial transforms of Algorithm 1 can be written under the form

$$C_l^{(0)} = \bar{F}_l^* G_l, \qquad \text{for } l = 1, 2 \tag{6}$$

where G_l is the $m \times m$ circulant matrix and $\bar{F}_l$ is the $m \times k_l$ matrix whose k_l columns are chosen among the columns of the $m \times m$ DFT matrix F_m defined by

$$(F_m)_{i,k} = \frac{1}{\sqrt{m}} e^{-j\frac{2\pi}{m}ik}, \qquad \text{for } i, k = 0, 1, \ldots, m-1.$$

Then, at each step s of Algorithm 1, the transforms $C_l^{(s)}$ can be expressed under the form (6) and the covariance matrices involved in the computation of the MSE given by (5) are (Hermitian) circulant.

Proof: See Appendix III. □

Initial transforms that satisfy the assumptions of Proposition 3.3 include the scenario where the two terminals disregard each other's presence and apply an optimal strategy in this context, i.e., by selecting

$$C_1^{(0)} = \bar{F}_1^* R_{SX_1} R_{X_1}^{-1} \quad \text{and} \quad C_2^{(0)} = \bar{F}_2^* R_{SX_2} R_{X_2}^{-1} \tag{7}$$

where the columns of $\bar{F}_l$ are the eigenvectors corresponding to the k_l largest eigenvalues of $R_{SX_l} R_{X_l}^{-1} R_{SX_l}^*$. A careful initialization thus ensures that, at each step of Algorithm 1, the covariance matrices involved in the computation of the MSE are of (Hermitian) circulant form. Their (real) eigenvalues can hence be easily characterized analytically. While not being crucial when the observed vectors are of finite length, an analogous property will allow us to derive closed-form analytical expressions of the distortion incurred by the distributed transform in the infinite-dimensional regime. This will be investigated in greater details in Section IV. Again, it is important to emphasize that the particular form imposed by Proposition 3.3 on the initial transforms may not conduct the proposed iterative algorithm to the global minimum. In other words, the ability to analytically track the MSE obtained at each iteration may come at the cost

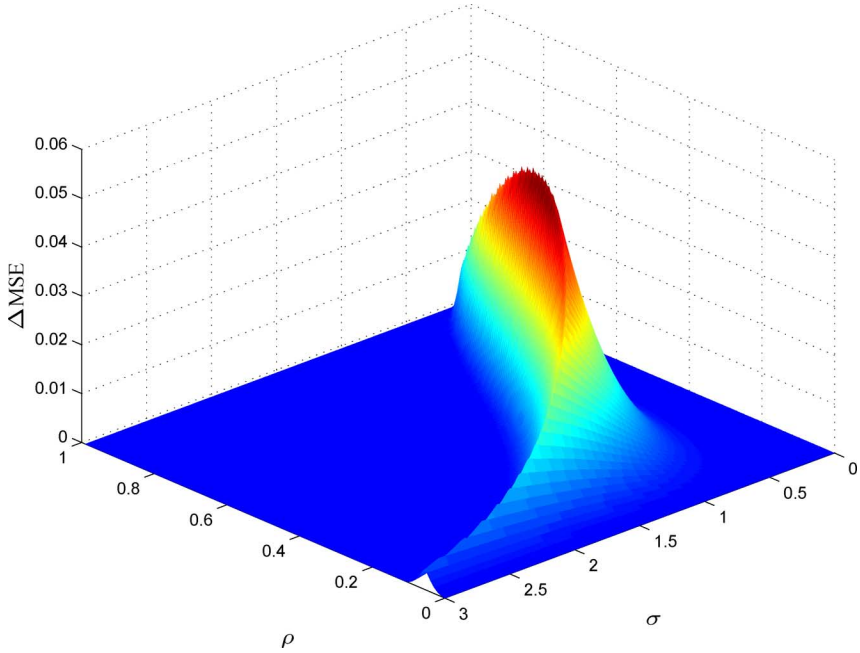

Fig. 3. Difference, in MSE, between the outcome of Algorithm 1 and the globally optimal solution for different values of ρ and σ. We observe the possible suboptimality of the proposed iterative scheme.

of a suboptimal solution. This fact is illustrated by the following numerical example.

Example 3.4: Consider the vectors S, X_1, and X_2 defined in Example 3.2. It is easily seen that the (cross-)covariance matrices are given by

$$R_S = R_{SX_1} = R_{SX_2} = R_{X_1X_2} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

and

$$R_{X_1} = R_{X_2} = \begin{pmatrix} 1+\sigma^2 & \rho \\ \rho & 1+\sigma^2 \end{pmatrix}$$

i.e., are Hermitian circulant. Here again, both terminals are required to provide a one-dimensional approximation of their observation. Algorithm 1 was run for different (ρ, σ)-pairs using the initial transforms (7). In Fig. 3, we plot the loss (ΔMSE) incurred by Algorithm 1 with respect to the optimal solution obtained by inspection. While global optimality is achieved for a wide range of values, the considered initialization leads the iterative algorithm to a suboptimal solution in some cases.

IV. Local Perspective in the Infinite-Dimensional Regime

In the previous section, we analyzed the distortion incurred by Algorithm 1 in the finite-dimensional regime under the additional assumption that the involved (cross-)covariance matrices are circulant. The main goal of this section is to extend this analysis to the infinite-dimensional regime, i.e., as we let the length m of the source and observed vectors go to infinity. More precisely, we show in the sequel how a variation of the Toeplitz distribution theorem [8, Th. 4.2] (see, also, [9, Ch. 7]) can provide closed-form characterizations of the reconstruction error incurred by the distributed transform derived in Theorem 3.1, hence for each step of Algorithm 1. In the context of data compression, the Toeplitz distribution theorem can be used to derive rate-distortion functions of stationary random processes in both centralized [30] and distributed [31] scenarios. The infinite block-length distortion analysis provided here is similar in spirit but is carried from a pure approximation (dimensionality reduction) standpoint, keeping up with the perspective adopted so far.

Let us study the setup considered in the finite case but letting m tend to infinity. In this context, terminal l ($l = 1, 2$) provides the FC with a description of size $k_l = \lfloor \alpha_l m \rfloor$ with $\alpha_l \in [0, 1]$, i.e., where only a fraction α_l of transformed coefficients is kept. Here, $\lfloor \cdot \rfloor$ denotes the floor operation. It is worth noting that both sizes of the observed and transformed vectors go to infinity whereas the ratio remains constant and is given by[3] $\alpha_l \sim k_l/m$. The study of the MSE under the above asymptotic considerations is achieved by means of the following definition.

Definition 4.1 (Infinite Block-Length Distortion): Let $\alpha_l \sim k_l/m$ for $l = 1, 2$. The infinite block-length distortion (IBLD) $D(\alpha_1, \alpha_2)$ is defined as

$$D(\alpha_1, \alpha_2) = \lim_{m \to \infty} \frac{D_m(k_1, k_2)}{m}$$

if the limit exists.

In Definition 4.1, $D_m(k_1, k_2)$ denotes the distortion in the finite-dimensional regime when terminal l provides a k_l-dimensional representation of its m-dimensional observation by means of a transform matrix C_l. Note that the normalization factor can be somewhat chosen arbitrarily. Here, we set it as the size m of the source vector S. From the local perspective

[3]The notation $\alpha_l \sim k_l/m$ means that $\lim_{m \to \infty} k_l/m = \alpha_l$.

developed in Section III-A, the minimum IBLD obtained when terminal 2 is fixed can be obtained from Theorem 3.1 by letting m tend to infinity. The ability to describe the optimal processing to be done at terminal 1 and to compute the resulting IBLD at the FC thus heavily depends on the knowledge we have about the eigenvalues of the involved covariance matrices in the limit of large vector sizes. In the sequel, we show that under additional stationarity assumptions, conclusive results can be found.

Before we proceed, let us establish some notation. We will denote by $\{T_m(\Phi)\}$ a sequence of $m \times m$ Toeplitz matrices, with eigenvalues $\lambda_{m,i}$, obtained from an absolutely summable sequence $\{t_k\}$. Its discrete-time Fourier transform (DTFT) $\Phi(\omega)$ exists, is continuous, and is bounded.[4] It is referred to as power spectral density (PSD). In the case the function $\Phi(\omega)$ is real, we denote its essential infimum and essential supremum by b_Φ and B_Φ, respectively. In the sequel, we will require the set over which a real PSD $\Phi(\omega)$ is constant to be of (Lesbegue) measure zero, i.e.,

$$\int_{\omega:\Phi(\omega)=x} d\omega = 0 \tag{8}$$

for all $x \in [b_\Phi, B_\Phi]$. In particular, we will assume that $\Phi(\omega)$ is nonzero almost everywhere (a.e.). The above technical condition ensures the continuity of the corresponding limiting eigenvalue distribution $F_\Phi(x)$ given in this case by [8, Corollary 4.1]

$$\begin{aligned} F_\Phi(x) &= \lim_{m\to\infty} F_{\Phi,m}(x) \\ &= \lim_{m\to\infty} \frac{|\{\lambda_{m,i} : \lambda_{m,i} \le x\}|}{m} \\ &= \frac{1}{2\pi}\int_{\omega\in\Omega} d\omega \end{aligned}$$

where $\Omega = \{\omega \in [0, 2\pi] : \Phi(\omega) \le x\}$. Here, $|\Omega|$ corresponds to the cardinality of the set Ω.

Because we aim at studying the distortion behavior of processes that exhibit stationarity properties, we will assume that the (cross)-covariance matrices involved in Theorem 3.1 are asymptotically Toeplitz (see [8, Sec. 2.3]). We thus follow a similar approach to that of Section III-B, because, in the limit, asymptotically Toeplitz matrices enjoy similar properties to those of circulant matrices [8, Th. 2.1 and Th. 5.3]. Note that a random vector process whose covariance matrix is asymptotically equivalent to the Toeplitz matrix $T_m(\Phi)$ will be said to have a PSD $\Phi(\omega)$ even if, strictly speaking, it is only stationary in the limit of large m.

As a means to provide an asymptotic counterpart to Theorem 3.1, let us define the m-dimensional vector $Z_l = \bar{C}_l Y_l$, where $\bar{C}_l$ is given by Theorem 3.1. Because $\bar{C}_l^* \bar{C}_l = I_{k_l}$, Y_l can be obtained from Z_l and vice versa. The distortion formula of Theorem 3.1 can hence be equivalently expressed as a function of Z_l by simply replacing Y_l by Z_l. Considering Z_l instead of Y_l allows us to investigate, as $m \to \infty$, the optimal way to process the sequence $X_l[m]$ observed at terminal l assuming that the strategy adopted by the other terminal is fixed. Let us consider again the problem from the perspective of terminal 1. Assume that the process $Z_2[m]$ provided by terminal 2 is expressed as $Z_2[m] = C_2[m] * X_2[m]$ for some fixed filter $C_2[m]$, with transfer function $C_2(\omega)$, chosen as to retain only a fraction α_2 of transformed coefficients. Here, $*$ denotes the convolution operator. Let us define the set

$$\Omega_1 = \{\omega \in [0, 2\pi] : \Phi_1(\omega) > x_{\alpha_1}\} \tag{9}$$

with $\Phi_1(\omega) = \Phi_{S\bar{X}_1}(\omega)\Phi_{\bar{X}_1}^{-1}(\omega)\Phi_{S\bar{X}_1}^*(\omega)$ and where x_{α_1} satisfies $F_{\Phi_1}(x_{\alpha_1}) = 1-\alpha_1$. The complementary set of Ω_1 in $[0, 2\pi]$ is denoted Ω_1^c. Furthermore, we define the process $\bar{X}_1[m]$ as

$$\bar{X}_1[m] = X_1[m] - H_2[m] * Z_2[m]$$

where the transfer function of $H_2[m]$ is given by $H_2(\omega) = \Phi_{X_1Z_2}(\omega)\Phi_{Z_2}^{-1}(\omega)$. Using these definitions, the choice of an optimal strategy for terminal 1 can be described as follows.

Theorem 4.2: The optimal process $Z_1[m]$ provided by terminal 1 when only a fraction α_1 of transformed coefficients are retained is given by

$$Z_1[m] = C_1[m] * X_1[m]$$

where the transfer function $C_1(\omega)$ can be expressed as

$$C_1(\omega) = \Phi_{S\bar{X}_1}(\omega)\Phi_{\bar{X}_1}^{-1}(\omega)\, 1_{\Omega_1}(\omega). \tag{10}$$

The resulting infinite block-length distortion $D(\alpha_1, \alpha_2)$ is given by

$$\begin{aligned} D(\alpha_1, \alpha_2) = &\frac{1}{2\pi}\int_{\omega\in[0,2\pi]} \Phi_{S|X_1,Z_2}(\omega)\, d\omega \\ &+ \frac{1}{2\pi}\int_{\omega\in\Omega_1^c} \Phi_{S|Z_2}(\omega) - \Phi_{S|X_1,Z_2}(\omega)\, d\omega. \end{aligned} \tag{11}$$

Proof: See Appendix IV. □

The optimal strategy amounts thus to bandlimit the process $X_1[m]$ such that the set of retained frequencies Ω_1 is of measure $2\pi\alpha_1$. The choice of these frequencies is given by $\Phi_1(\omega)$ which depends on the process $Z_2[m]$ provided by terminal 2. It is seen in the proof of Theorem 4.2 that the term $\Phi_{S\bar{X}_1}(\omega)\Phi_{\bar{X}_1}^{-1}(\omega)$ is not needed for optimality because it is assumed to be nonzero a.e. We provide it here by analogy to the finite-dimensional case. Note also that, in a strictly analogous manner to (5), the distortion given by (11) is the sum of a first term that is not influenced by terminal 1, and a second term that amounts to integrate $\Phi_1(\omega)$ in a reverse "water-filling" fashion where x_{α_1} plays the role of the "water level." This fact is illustrated in Fig. 4(a). As α_1 ranges from 0 to 1, x_{α_1} scopes from B_{Φ_1} to b_{Φ_1} as dictated by the relation

$$F_{\Phi_1}(x_{\alpha_1}) = 1 - \alpha_1. \tag{12}$$

The result is intuitively clear: the second term of the distortion corresponds to the sum of the eigenvalues below an admissible threshold x_{α_1}. The relationship between α_1 and x_{α_1} is given by the limiting eigenvalue distribution $F_{\Phi_1}(x)$. A closed-form

[4]While 2π-periodic, the DTFTs considered in this paper will have ω as an argument instead of $e^{j\omega}$. This considerably lightens the notation.

analytical expression of the IBLD thus depends on the ability to find x_{α_1} satisfying (12). For an arbitrary PSD $\Phi_1(\omega)$, this might turn out to be difficult. However, if $\Phi_1(\omega)$ is symmetric (i.e., the corresponding covariance matrix is real-valued) and strictly monotonic in $[0,\pi]$, then a simple geometrical argument reveals that the set Ω_1 in (9) can be expressed as

$$\Omega_1 = [0, \pi\alpha_1) \cup (2\pi - \pi\alpha_1, 2\pi] \tag{13}$$

if $\Phi_1(\omega)$ is strictly decreasing in $[0,\pi]$ [see Fig. 4(b)], and

$$\Omega_1 = (\pi(1-\alpha_1), \pi(1+\alpha_1))$$

if $\Phi_1(\omega)$ is strictly increasing in $[0,\pi]$. Interestingly, under these restrictive assumptions, the knowledge of the limiting eigenvalue distribution is not required to compute the IBLD. This will greatly simplify the derivation presented in Section V.

Similarly to the finite-dimensional regime, Theorem 4.2 can be used to describe the optimal step of an iterative algorithm, hereafter referred to as Algorithm 2. If the assumptions of the previous theorem are satisfied at initialization, the consecutive steps simply amount to the choice of the optimal filters $C_1^{(s)}(\omega)$ and $C_2^{(s)}(\omega)$ as described by (10). In fact, by strict analogy to Proposition 3.3, the (cross-)covariance matrices involved in the computations of Theorem 4.2 remain asymptotically Toeplitz over the iterations. Algorithm 1 can thus be restated in its infinite-dimensional version as follows.

Algorithm 2

Select a fixed tolerance ε and arbitrary bandlimiting filters $C_1^{(0)}(\omega)$ and $C_2^{(0)}(\omega)$ such that the sets Ω_1 and Ω_2 are of measure $2\pi\alpha_1$ and $2\pi\alpha_2$, respectively.

while (true)

for $l = 1, 2$

Compute $C_l^{(s)}(\omega)$ as given by Theorem 4.2.

end

If $|\text{MSE}^{(s)} - \text{MSE}^{(s-1)}| < \varepsilon$ then stop.

$s = s + 1$;

end

In this case, the quest for optimality amounts to bandlimit, in turn at each terminal, the observed spectrum with respect to what is provided by the other terminal. The algorithm thus reduces to a "spectrum shaping game" whose outcome provides the necessary filters. Similarly to Section III-A, it is important to point out that the particular initialization imposed by Theorem 4.2 may lead Algorithm 2 to a suboptimal solution. The design of an optimal distributed dimensionality reduction architecture thus remains a challenging task.

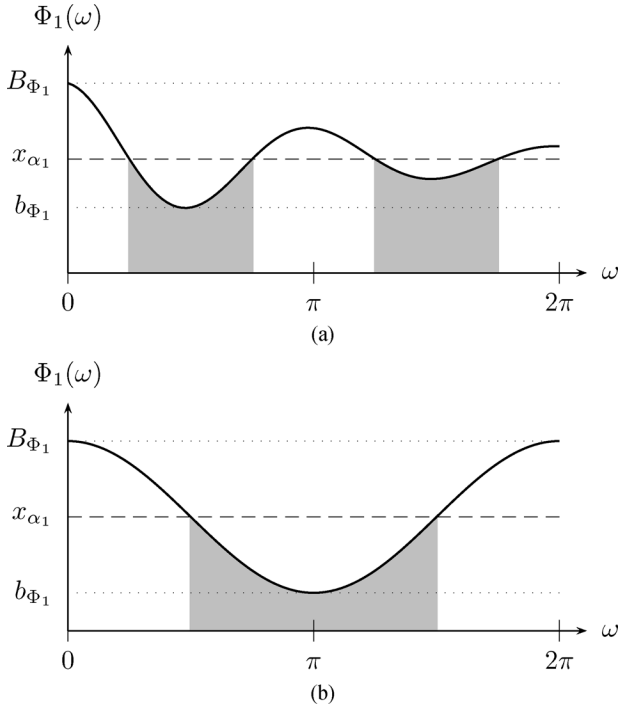

Fig. 4. Computation of the second term in the distortion of Theorem 4.2 in a reverse "water-filling" fashion (up to a scaling factor 2π). The value x_{α_1} plays the role of the "water level." When $\alpha_1 = 0$, $x_{\alpha_1} = B_{\Phi_1}$ and $\Phi_1(\omega)$ is integrated over $[0, 2\pi]$, i.e., the distortion is maximal. When $\alpha_1 = 1$, $x_{\alpha_1} = b_{\Phi_1}$ and the distortion is minimal. (a) $\Phi_1(\omega)$ is arbitrary. (b) $\Phi_1(\omega)$ is symmetric around π and strictly decreasing in $[0,\pi]$.

V. First-Order Gauss–Markov Process

We apply the results obtained previously to a first-order Gauss–Markov process. Owing to its simplicity, it is particularly suited for the analytical development that will be presented in the sequel. We derive analytical closed-form formulas for the IBLD that allow us to assess the performance of the distributed estimation scheme in the infinite-dimensional regime and to precisely compare different scenarios of interest. More importantly, we show that, in this example, our asymptotic analysis provides an accurate estimation of the distortion incurred in the finite case with vectors of small dimension. We also relate these analytical results to the general two-terminal distortion surface obtained numerically using Algorithm 1.

Let us consider a first-order Gauss–Markov process $X[m]$ with correlation parameter ρ, i.e., a random process that satisfies

$$X[m] = \rho\, X[m-1] + W[m], \qquad \text{for } m \in \mathbb{Z}$$

where $\rho \in (0,1)$[5] and where $W[m]$ is a white Gaussian noise with PSD $\Phi_W(\omega) = 1 - \rho^2$. We will consider the case where terminal 1 samples the odd coefficients of $X[m]$ and terminal 2 observes the even ones. In vector notation, we can define the $2m$-dimensional source vector $S = (X_1^T, X_2^T)^T$ with m-dimensional observed vectors $X_1 = (X[1], X[3], \ldots, X[2m-1])^T$ and $X_2 = (X[2], X[4], \ldots, X[2m])^T$. The sequences $S[m]$, $X_1[m]$,

[5]The case $\rho \in (-1, 0)$ follows immediately by considering $|\rho|$.

and $X_2[m]$ are zero-mean stationary random processes whose PSDs can be computed as

$$\Phi_S(\omega) = \frac{1-\rho^2}{1+\rho^2-2\rho\cos\omega}$$
$$\Phi_{X_l}(\omega) = \frac{1-\rho^4}{1+\rho^4-2\rho^2\cos\omega}$$
$$\Phi_{X_1X_2}(\omega) = \frac{\rho(1-\rho^2)(1+e^{-j\omega})}{1+\rho^4-2\rho^2\cos\omega}$$

for $l = 1, 2$. This setup is motivated, for example, by super-resolution imaging problems, where two subsampled versions of the same signal are acquired by cheap sensing devices in order to build a higher resolution image. In this case, correlation (hence stationarity) is considered across space. The analysis of the distortion in this scenario allows us to precisely quantify the gain achievable by the use of a low-resolution image as perfect side information, i.e., when one terminal entirely conveys its observed signal to the FC. It also gives a useful characterization of the loss incurred due to the need of interpolating missing samples, i.e., in the case one terminal does not transmit anything.

A. Centralized Scenario

We first consider the centralized scenario where the two terminals are allowed to process their observations jointly. In this case, the observed process is $S[m]$ and a fraction α of transformed coefficients is kept. The next proposition provides an analytical closed-form formula for the IBLD under these assumptions.

Proposition 5.1: The infinite block-length distortion in the centralized scenario where a fraction α of transformed coefficients is kept is given by

$$D(\alpha) = 1 - \frac{2}{\pi}\arctan\left(\frac{1+\rho}{1-\rho}\tan\left(\frac{\pi\alpha}{2}\right)\right).$$

Proof: See Appendix V-A. □

We provide in Fig. 5(a) the IBLD obtained in Proposition 5.1. We show in Fig. 5(b) how the IBLD is approximated for $\rho = 0.6$ and $m = 12$. We observe that even for small values of m, the asymptotic analysis presented here offers a very good approximation of the distortion in the finite-dimensional regime. We also compute in Fig. 5(c) the approximation error $e(m) = \frac{1}{m}\sum_{k=0}^{m-1} |D_m(k) - D(k/m)|^2$ to quantify the quality of the estimate with respect to the size of the source vector. The observed decay suggests, once again, that the results obtained by our asymptotical analysis approximate accurately the distortion we would compute with a finite number of measurements. This can be explained by the rapid decay of the correlation function $R_S[k]$ in this particular example.

B. Perfect Side Information Scenario

Let us now consider the case where terminal 1 needs to transmit $X_1[m]$ using a fraction α_1 of transformed coefficients and that the process $X_2[m]$ acts as side information, i.e., is perfectly conveyed to the FC ($\alpha_2 = 1$). The next proposition provides an analytical closed-form formula for the IBLD under these assumptions.

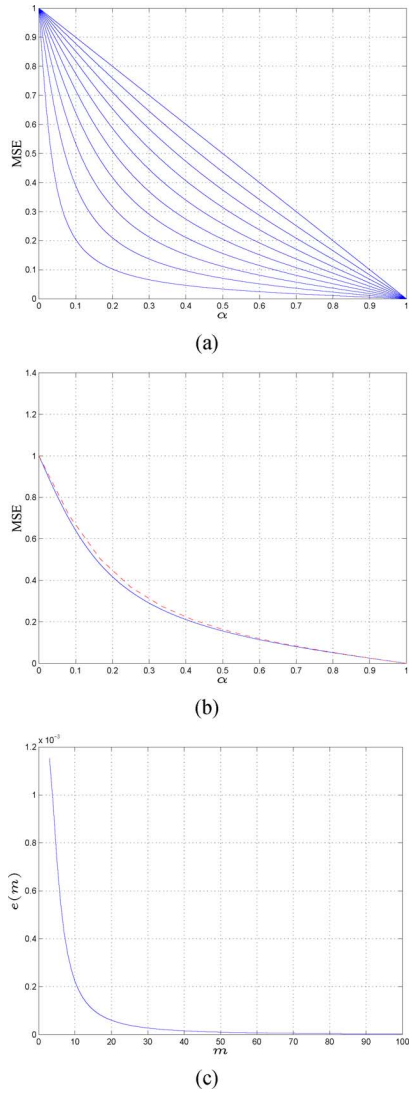

Fig. 5. IBLD in the centralized scenario. (a) IBLD for $\rho = 0, 0.1, \ldots, 0.9$ (from top to bottom). (b) IBLD (solid) and its approximation (dashed) for $\rho = 0.6$ and $m = 12$. (c) Approximation error $e(m)$ as a function of the size of the source vector for $\rho = 0.6$.

Proposition 5.2: The infinite block-length distortion in the perfect side information scenario where a fraction α_1 of transformed coefficients is kept is given by

$$D(\alpha_1, 1) = \frac{1-\rho^2}{2(1+\rho^2)}(1-\alpha_1).$$

Proof: See Appendix V-B. □

We show the IBLD obtained in Proposition 5.2 in Fig. 6(a). It is seen in the proof of the proposition that the prediction error of the odd coefficients by the even ones has uncorrelated components, i.e., the error process is white. This is due to the first-order property of the Gauss–Markov process and it yields a linear decrease in distortion. When $X_2[m]$ is completely available at the FC, the IBLD of $S[m]$ is equivalent to that of $X_1[m]$ up to a factor $1/2$. The above IBLD is thus half the reconstruction error of the process $X_1[m]$ with $X_2[m]$ as side information. In contrast, replacing ρ by ρ^2 and setting $\alpha = \alpha_1$ in Proposition 5.1

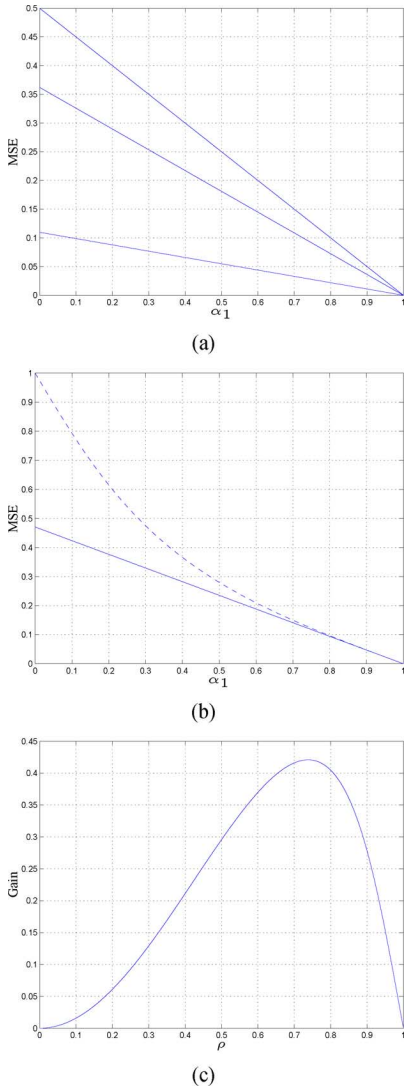

Fig. 6. IBLD in the perfect side information scenario (renormalized by a factor $1/2$). (a) IBLD for $\rho = 0, 0.4, 0.8$ (from top to bottom). (b) IBDL with (solid) and without (dashed) side information for $\rho = 0.6$. (c) Gain due to side information as a function of ρ for $\alpha = 0.1$.

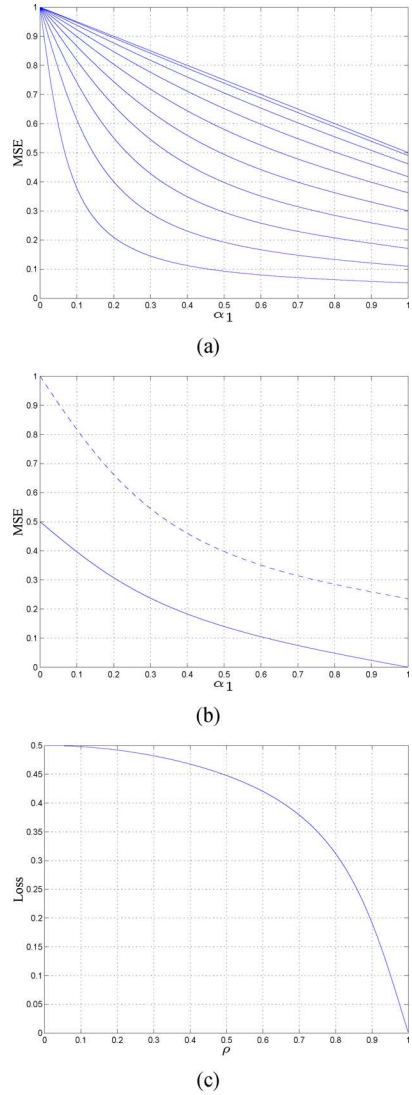

Fig. 7. IBLD in the partial-observation scenario. (a) IBLD for $\rho = 0, 0.1, \ldots, 0.9$ (from top to bottom). (b) IBDL with (dashed) and without (solid) hidden part for $\rho = 0.6$. (c) Loss due to the hidden part as a function of ρ for $\alpha = 0.1$.

amounts to computing the IBLD in the absence of side information at the FC. We compare these two scenarios in Fig. 6(b). We clearly see the gain achieved by providing the FC with some correlated side information. The exact gain can be expressed using Propositions 5.1 and 5.2 with the aforementioned modifications. As $\rho \to 0$, the processes $X_1[m]$ and $X_2[m]$ become uncorrelated, i.e., the presence of the side information does not provide any gain. When $\rho \to 1$, the correlation among the components of $X_1[m]$ allows us to perfectly recover the discarded coefficients without the need for $X_2[m]$. However, between these two extreme cases, a substantial gain is achieved by the use of noisy side information. This is depicted in Fig. 6(c).

C. Partial Observation Scenario

We treat now the case where terminal 1 needs to transmit $X_1[m]$ using a fraction α_1 of transformed coefficients and where the process $X_2[m]$ is completely discarded by terminal 2, i.e., acts as a hidden part ($\alpha_2 = 0$). The next proposition provides an analytical closed-form formula for the IBLD under these assumptions.

Proposition 5.3: The infinite block-length distortion in the partial observation scenario where a fraction α_1 of transformed coefficients is kept is given by

$$D(\alpha_1, 0) = 1 + \frac{\alpha_1(1-\rho^2)}{2(1+\rho^2)} - \frac{2}{\pi} \arctan\left(\frac{1+\rho^2}{1-\rho^2} \tan\left(\frac{\pi\alpha_1}{2}\right)\right).$$

Proof: See Appendix V-C. □

The IBLD obtained in Proposition 5.3 is depicted in Fig. 7(a). We also compare in Fig. 7(b) the IBLD obtained with and without hidden part. We clearly see the loss incurred by having to reconstruct the missing information at the FC. Furthermore, increasing ρ allows us to estimate more and more accurately the missing data, hence reducing the gap between the two distortions as shown in Fig. 7(c). The exact loss can be expressed from Proposition 5.3 and by replacing ρ by ρ^2,

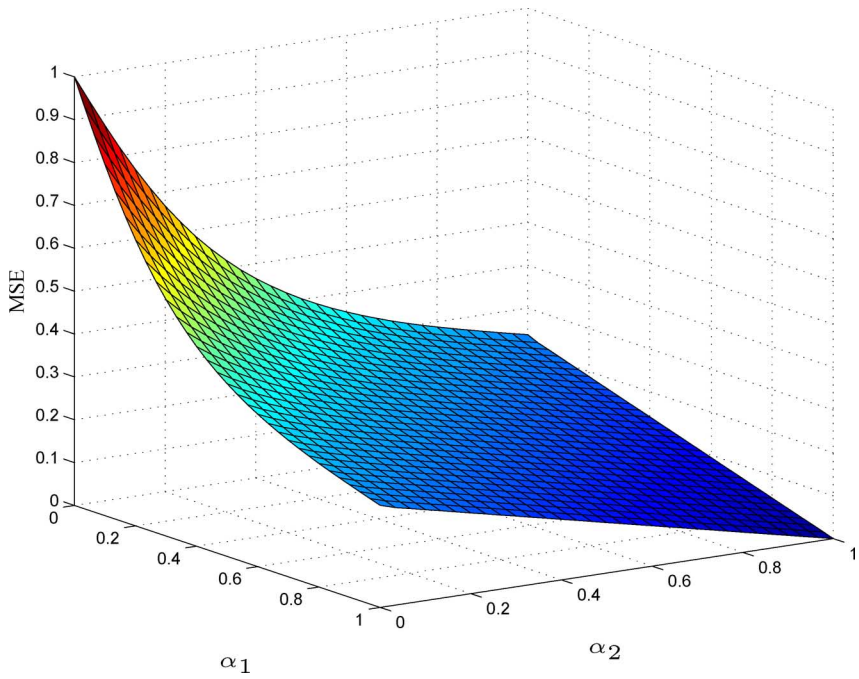

Fig. 8. IBLD in the general scenario for $\rho = 0.6$. The inside of the distortion surface is obtained using Algorithm 1 and a block length $m = 30$.

setting $\alpha = \alpha_1$, and normalizing by a factor 2 the distortion obtained in Proposition 5.1.

D. General Scenario

Let us now consider again the general two-terminal setup introduced in Section II. Assume that terminal l only keeps a fraction α_l of transformed coefficients. We can then conveniently represent the entire IBLD surface as a function of α_1 and α_2, both taking values in $[0, 1]$. This is shown in Fig. 8 for $\rho = 0.6$. The inside of the distortion surface is obtained numerically using Algorithm 1 with a block length of size $m = 30$, whereas the analytical expression of the border is perfectly known. In fact, it corresponds to the IBLD obtained for the partial observation scenario ($\alpha_1 = 0$ or $\alpha_2 = 0$) and the perfect side information scenario ($\alpha_1 = 1$ or $\alpha_2 = 1$). The inherent symmetry is due to the fact that $\Phi_{X_1}(\omega)$ and $\Phi_{X_2}(\omega)$ are equal. This perspective shows that even if the overall distortion surface is not known, it is constrained by its borders. In fact, the perfect side information and partial observation scenarios, respectively, provide a lower bound and an upper bound of the IBLD.

VI. Conclusion

We have addressed the problem of estimating an unknown random vector source at a fusion center using a set of distributed sensors (terminals). We have proposed a transform-based approach where each terminal applies a suitable linear transform in order to provide a low-dimensional approximation of its observation. Under the MSE criterion, the optimal transform to apply at one terminal assuming all else being fixed has been derived. Based on this result, we have developed an iterative algorithm whose optimality has been discussed. The equivalent problem has then been studied in the infinite-dimensional regime, i.e., as the size of the involved vectors goes to infinity. In this regime, similar optimality results have been proved under some restrictive stationarity assumptions. Finally, we have illustrated with a simple correlation model how the considered asymptotic analysis may be used to derive closed-form distortion formulas that provide an accurate estimate of the MSE in the finite-dimensional regime.

Appendix I
Minimum Mean Square Estimate With Singular Covariance Matrices

Let $S \in \mathbb{C}^m$ and $X \in \mathbb{C}^n$ be zero-mean jointly Gaussian random vectors. In this case, the conditional expectation is a linear operation, that is $\mathrm{E}[S|X] = A^*X$, where the $m \times n$ matrix A is obtained from the Wiener–Hopf equations

$$R_X A = R_{XS}. \tag{14}$$

When R_X is invertible, the solution directly follows as $A = R_X^{-1} R_{XS}$. When R_X is singular (say of rank $k < n$), we consider its eigenvalue decomposition

$$R_X = UDU^* = [U_1 \quad U_2] \begin{bmatrix} D_1 & O_{k\times n-k} \\ O_{n-k\times k} & O_{n-k\times n-k} \end{bmatrix} \begin{bmatrix} U_1^* \\ U_2^* \end{bmatrix}$$

and rewrite (14) as

$$DU^*A = U^*R_{XS}$$

or, equivalently,

$$\begin{bmatrix} D_1 U_1^* \\ O_{n-k\times n} \end{bmatrix} A = \begin{bmatrix} U_1^* R_{XS} \\ U_2^* R_{XS} \end{bmatrix}. \tag{15}$$

The diagonal matrix D_1 contains the k nonzero (real) eigenvalues of R_X. The unitary matrix of eigenvectors U is partitioned accordingly into U_1 and U_2. It thus follows that the el-

ements of the vector $Y = U_2^* X$ have variance zero. Using Cauchy–Schwarz's inequality [32, Th. C.1.1], we have that

$$0 \leq |\mathrm{E}[Y_i S_j^*]| \leq \sqrt{\mathrm{E}[|Y_i|^2]\mathrm{E}[|S_j|^2]} = 0$$

for $i = 1, 2, \ldots, n-k$ and $j = 1, 2, \ldots, m$, such that $U_2^* R_{XS} = U_2^* \mathrm{E}[XS^*] = \mathrm{E}[YS^*] = O_{n-k,m}$. The equality (15) thus reduces to

$$D_1 U_1^* A = U_1^* R_{XS}.$$

The above system has km equations and nm unknowns, hence an infinite number of solutions ($k < n$). One possible solution is obtained using the Moore–Penrose pseudoinverse of $D_1 U_1^*$ given by

$$(D_1 U_1^*)^\dagger = U_1 D_1^* (D_1 U_1^* U_1 D_1^*)^{-1} = U_1 D_1^{-1}$$

which yields the solution

$$A = (D_1 U_1^*)^\dagger U_1^* R_{XS} = U_1 D_1^{-1} U_1^* R_{XS}.$$

Appendix II
Proof of Theorem 3.1

The proof makes use of the following two lemmas.

Lemma II.1: Let X, Y, and Z be zero-mean jointly Gaussian random vectors. Define $\bar{Z} = Z - \mathrm{E}[Z|Y]$. It holds that $\mathrm{E}[X|Y,Z] = \mathrm{E}[X|Y] + \mathrm{E}[X|\bar{Z}]$.

Proof: We can write

$$\begin{aligned}
&\mathrm{E}[X|Y,Z] \\
&\overset{(a)}{=} \begin{pmatrix} R_{XY} & R_{XZ} \end{pmatrix} \begin{pmatrix} R_Y & R_{YZ} \\ R_{YZ}^* & R_Z \end{pmatrix}^{-1} \begin{pmatrix} Y \\ Z \end{pmatrix} \\
&\overset{(b)}{=} \begin{pmatrix} R_{XY} & R_{XZ} \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} \left(R_Y - R_{YZ} R_{\bar{Z}}^{-1} R_{YZ}^*\right)^{-1} & -R_Y^{-1} R_{YZ} R_{\bar{Z}}^{-1} \\ -R_{\bar{Z}}^{-1} R_{YZ}^* R_Y^{-1} & R_{\bar{Z}}^{-1} \end{pmatrix} \begin{pmatrix} Y \\ Z \end{pmatrix} \\
&\overset{(c)}{=} \begin{pmatrix} R_{XY} & R_{XZ} \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} R_Y^{-1} + R_Y^{-1} R_{YZ} R_{\bar{Z}}^{-1} R_{YZ}^* R_Y^{-1} & -R_Y^{-1} R_{YZ} R_{\bar{Z}}^{-1} \\ -R_{\bar{Z}}^{-1} R_{YZ}^* R_Y^{-1} & R_{\bar{Z}}^{-1} \end{pmatrix} \begin{pmatrix} Y \\ Z \end{pmatrix} \\
&= R_{XY} R_Y^{-1} Y \\
&\quad + \left(R_{XZ} - R_{XY} R_Y^{-1} R_{YZ}\right) R_{\bar{Z}}^{-1} \left(Z - R_{YZ}^* R_Y^{-1} Y\right) \\
&= R_{XY} R_Y^{-1} Y + R_{X\bar{Z}} R_{\bar{Z}}^{-1} \bar{Z} \\
&= \mathrm{E}[X|Y] + \mathrm{E}[X|\bar{Z}]
\end{aligned}$$

where (a) follows from the expression of the conditional expectation in the jointly Gaussian case [27, Sec. IV.B], (b) from the inversion formula of a partitioned matrix [33, Sec. 0.7.3], and (c) from that of a small rank adjustment [33, Sec. 0.7.4]. □

Lemma II.2: Let X, Y, and Z be zero-mean jointly Gaussian random vectors. Define $\bar{Z} = Z - \mathrm{E}[Z|Y]$. Then, $R_X - R_{X|\bar{Z}} = R_{X|Y} - R_{X|Y,Z}$.

Proof: Using Lemma II.1, we can write

$$\begin{aligned}
\mathrm{E}[X|\bar{Z}] &= \mathrm{E}[X|Y,Z] - \mathrm{E}[X|Y] \\
&= (X - \mathrm{E}[X|Y]) - (X - \mathrm{E}[X|Y,Z]).
\end{aligned}$$

By the orthogonality principle, $\mathrm{E}[X|\bar{Z}]$ is orthogonal to both Y and $X - \mathrm{E}[X|\bar{Z}]$, hence to $X - \mathrm{E}[X|Y,Z]$. We thus have

$$\begin{aligned}
&\mathrm{E}\left[(X - \mathrm{E}[X|Y])(X - \mathrm{E}[X|Y,Z])^*\right] \\
&\quad = \mathrm{E}\left[\left(X - \mathrm{E}[X|Y] - \mathrm{E}[X|\bar{Z}]\right)(X - \mathrm{E}[X|Y,Z])^*\right] \\
&\quad = \mathrm{E}\left[(X - \mathrm{E}[X|Y,Z])(X - \mathrm{E}[X|Y,Z])^*\right].
\end{aligned}$$

This allows us to write

$$\begin{aligned}
R_X - R_{X|\bar{Z}} &= \mathrm{E}\left[\mathrm{E}[X|\bar{Z}]\mathrm{E}[X|\bar{Z}]^*\right] \\
&= \mathrm{E}\left[((X - \mathrm{E}[X|Y]) - (X - \mathrm{E}[X|Y,Z])) \right. \\
&\qquad \left. \cdot ((X - \mathrm{E}[X|Y]) - (X - \mathrm{E}[X|Y,Z]))^*\right] \\
&= R_{X|Y} - R_{X|Y,Z}
\end{aligned}$$

yielding the claimed equality. □

Proof of Theorem 3.1: It was shown in [15, Th. 2] that an optimal solution must be such that the MMSE is given by

$$\mathrm{E}[\|S - \hat{S}\|^2] = \mathrm{tr}(R_{S|Y_2}) - \sum_{i=n_1-k_1+1}^{n_1} \tilde{\lambda}_i$$

where $\tilde{\lambda}_i$ denote the eigenvalues of the matrix $R_{S\bar{X}_1}^* R_{S\bar{X}_1} R_{\bar{X}_1}^{-1}$ arranged in increasing order. Using Lemma II.2, we have that

$$R_{S|Y_2} - R_{S|X_1,Y_2} = R_S - R_{S|\bar{X}_1} = R_{S\bar{X}_1} R_{\bar{X}_1}^{-1} R_{S\bar{X}_1}^*.$$

Using the determinant formula $\det(AB+I) = \det(BA+I)$, it follows that the n_1 largest eigenvalues of $R_{S|Y_2} - R_{S|X_1,Y_2}$ are given by $\lambda_i = \tilde{\lambda}_i$, the $m - n_1$ remaining ones being zero. The MSE incurred by an optimal transform can thus be written as

$$\mathrm{E}[\|S - \hat{S}\|^2] = \mathrm{tr}(R_{S|X_1,Y_2}) + \sum_{i=1}^{n_1-k_1} \lambda_i.$$

It remains to show that the transform given by (4) provides this MSE. We can write

$$\begin{aligned}
&\mathrm{E}[\|S - \hat{S}\|^2] \\
&\quad = \mathrm{E}[\|S - \mathrm{E}[S|C_1 X_1, Y_2]\|^2] \\
&\quad \overset{(a)}{=} \mathrm{E}[\|S - \mathrm{E}[S|Y_2] - \mathrm{E}[S|C_1 \bar{X}_1]\|^2] \\
&\quad \overset{(b)}{=} \mathrm{E}[\|S - R_{SY_2} R_{Y_2}^{-1} Y_2 \\
&\qquad - R_{S\bar{X}_1} C_1^* \left(C_1 R_{\bar{X}_1} C_1^*\right)^{-1} C_1 \bar{X}_1\|^2] \\
&\quad \overset{(c)}{=} \mathrm{tr}\left(R_{S|Y_2}\right) - \mathrm{tr}\left(R_{S\bar{X}_1} C_1^* \left(C_1 R_{\bar{X}_1} C_1^*\right)^{-1} C_1 R_{S\bar{X}_1}^*\right) \\
&\quad = \mathrm{tr}\left(R_{S|Y_2}\right) \\
&\qquad - \mathrm{tr}\Big(R_{S\bar{X}_1} R_{\bar{X}_1}^{-1} R_{S\bar{X}_1}^* \bar{C}_1 \left(\bar{C}_1^* R_{S\bar{X}_1} R_{\bar{X}_1}^{-1} R_{S\bar{X}_1}^* \bar{C}_1\right)^{-1} \\
&\qquad\qquad \cdot \bar{C}_1^* R_{S\bar{X}_1} R_{\bar{X}_1}^{-1} R_{S\bar{X}_1}^*\Big) \\
&\quad \overset{(d)}{=} \mathrm{tr}\left(R_{S|Y_2}\right) - \sum_{i=n_1-k_1+1}^{n_1} \lambda_i \\
&\quad = \mathrm{tr}\left(R_{S|X_1,Y_2}\right) + \sum_{i=1}^{n_1-k_1} \lambda_i
\end{aligned}$$

where (a) follows from Lemma II.1 and (b) from the expression of the conditional expectation in the jointly Gaussian case [27, Sec. IV.B]. The equality (c) follows from the definition of the Euclidean norm, the fact that expectation and trace commute, and from the orthogonality of $\bar{X}_1$ and Y_2 (orthogonality principle). Finally, (d) follows from the properties of the trace and the definition of $\bar{C}_1$. Because $\mathrm{E}[S|C_1X_1, Y_2] = \mathrm{E}[S|C_1\bar{X}_1, Y_2]$, the availability of Y_2 at the encoder does not change the MMSE. In this case, the optimal transform is still given by (4), but the transmitted coefficients are different because the optimal transform can be applied on $\bar{X}_1$ instead of X_1. □

Appendix III
Proof of Proposition 3.3

Assume that the transform applied by one terminal (say terminal 2) at step s is of the form (6), i.e., can be written as

$$C_2 = \bar{F}_2^* G_2.$$

To prove the first part of the claim, we simply need to show that the transform selected by the other terminal (terminal 1) is of the same form. The result then follows by induction on s and the fact that, at $s = 0$, the claim is trivially verified by initialization. The transform matrix to be applied at terminal 1 is given by Theorem 3.1. In particular, we have

$$\begin{aligned} R_{\bar{X}_1} &= R_{X_1} - R_{X_1X_2}C_2^*\left(C_2R_{X_2}C_2^*\right)^{-1}C_2R_{X_1X_2}^* \\ &= R_{X_1} - R_{X_1X_2}G_2^*\bar{F}_2\left(\bar{F}_2^*G_2R_{X_2}G_2^*\bar{F}_2\right)^{-1}\bar{F}_2^*G_2R_{X_1X_2}^*. \end{aligned}$$

Because G_2 and the involved (cross-)covariance matrices are circulant, the matrix $R_{\bar{X}_1}$ is easily proved to be circulant using the properties of circulant matrices [8, Th. 3.1] and the fact that the product $\bar{F}_2^*G_2R_{X_2}G_2^*\bar{F}_2$ is diagonal. A similar argument shows that $R_{S\bar{X}_1}$ is also circulant. The matrix $\bar{C}_1$ of Theorem 3.1 diagonalizes the circulant matrix $R_{S\bar{X}_1}R_{\bar{X}_1}^{-1}R_{S\bar{X}_1}^*$ and hence satisfies the condition on $\bar{F}_1$. The transform matrix C_1 is thus of the desired form (6).

The second part of the claim follows by noticing that the matrix $R_{S|X_1,Y_2}$ can be written using the inversion formula of a partitioned matrix [33, Sec. 0.7.3] as the sum, product, and inverse of circulant matrices. It is thus circulant. A similar argument holds for $R_{S|Y_2}$. □

Appendix IV
Proof of Theorem 4.2

The proof consists of two parts. In part A, we prove that the minimum achievable distortion is given by (11). In part B, we show that the filtering strategy (10) achieves this distortion.

A. Minimum Achievable Distortion

Let us first define the function

$$1_x(\lambda) = \begin{cases} 1, & \text{if } \lambda \le x \\ 0, & \text{if } \lambda > x \end{cases}$$

and prove the following two lemmas.

Lemma IV.1: Let $T_m(\Phi)$ be a sequence of $m \times m$ Hermitian Toeplitz matrices such that $\Phi(\omega)$ satisfies the condition given in (8). Then, for any nonnegative integer s

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}(\lambda_{m,i})^s\,1_x(\lambda_{m,i}) = \frac{1}{2\pi}\int_{\omega\in\Omega}\Phi^s(\omega)\,d\omega$$

where $\lambda_{m,i}$ denote the eigenvalues of $T_m(\Phi)$ and $\Omega = \{\omega \in [0, 2\pi] : \Phi(\omega) \le x\}$.

Proof: Let us proceed by induction. For $s = 0$, we know that the assertion holds true by [8, Th. 4.1], i.e.,

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}1_x(\lambda_{m,i}) = \frac{1}{2\pi}\int_{\omega\in\Omega}d\omega.$$

Assume it has been proved for $s-1$; we now prove it for s. We first note that the left-hand side of the assertion can be expressed as

$$\begin{aligned} \frac{1}{m}\sum_{i=1}^{m}\lambda_{m,i}^s 1_x(\lambda_{m,i}) &= \frac{1}{m}\sum_{i=1}^{m}\lambda_{m,i}^{s-1}\min\{x,\lambda_{m,i}\} \\ &\quad - \frac{x}{m}\sum_{i=1}^{m}\lambda_{m,i}^{s-1}\left(1 - 1_x(\lambda_{m,i})\right) \\ &= \frac{1}{m}\sum_{i=1}^{m}\lambda_{m,i}^{s-1}\min\{x,\lambda_{m,i}\} \\ &\quad - \frac{x}{m}\sum_{i=1}^{m}\lambda_{m,i}^{s-1} + \frac{x}{m}\sum_{i=1}^{m}\lambda_{m,i}^{s-1}1_x(\lambda_{m,i}). \end{aligned}$$

Because $\min$ is a continuous function, we can apply the eigenvalue distribution theorem [8, Th. 4.2] to the first and second summations and our induction assumption to the third one to obtain

$$\begin{aligned} &\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}\lambda_{m,i}^s 1_x(\lambda_{m,i}) \\ &\quad= \frac{1}{2\pi}\int_{\omega\in[0,2\pi]}\Phi^{s-1}(\omega)\min\{x,\Phi(\omega)\}\,d\omega \\ &\qquad - \frac{x}{2\pi}\int_{\omega\in[0,2\pi]}\Phi^{s-1}(\omega)\,d\omega + \frac{x}{2\pi}\int_{\omega\in\Omega}\Phi^{s-1}(\omega)\,d\omega \\ &\quad= \frac{1}{2\pi}\int_{\omega\in\Omega}\Phi^s(\omega)\,d\omega \end{aligned}$$

which yields the desired result. □

Because any polynomial in λ can be expressed as a linear combination of monomials λ^s $(s \ge 0)$, Lemma IV.1 can be straightforwardly extended to polynomials. Furthermore, invoking the Stone–Weierstrass theorem [8, Th. 2.3] immediately yields the following generalization.

Lemma IV.2: Let $T_m(\Phi)$ be a sequence of Hermitian Toeplitz matrices such that $\Phi(\omega)$ satisfies the condition given in (8). Then, for any function $G(x)$ continuous on $[b_\Phi, B_\Phi]$

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m}G(\lambda_{m,i})\,1_x(\lambda_{m,i}) = \frac{1}{2\pi}\int_{\omega\in\Omega}G(\Phi(\omega))\,d\omega$$

where $\lambda_{m,i}$ denote the eigenvalues of $T_m(\Phi)$ and $\Omega = \{\omega \in [0, 2\pi] : \Phi(\omega) \leq x\}$.

Note that Lemma IV.2 allows us to carry similar computations on functions with a finite number of discontinuities by simply isolating the continuous pieces. Extension to an infinite but countable number of discontinuities may be envisioned but special care has to be taken with the function $G(x)$ so as to satisfy Lebesgue's theorem assumptions [32, Th. 12].

We can now turn to the proof of part A. The optimal finite-dimensional distortion $D_m(k_1, k_2)$ is computed from Theorem 3.1 by replacing Y_2 by Z_2 to obtain

$$D_m(k_1,k_2) = \mathrm{tr}\left(T_m\left(\Phi_{S|X_1,Z_2}\right)\right) + \sum_{i=1}^{m-\lfloor \alpha_1 m \rfloor} \lambda_{m,i}$$

where $\lambda_{m,i}$ denote the eigenvalues of $T_m\left(\Phi_{S|Z_2} - \Phi_{S|X_1,Z_2}\right)$ arranged in increasing order. The first term in the IBLD directly follows from the limiting eigenvalue distribution theorem [8, Th. 4.2], i.e.,

$$\lim_{m\to\infty} \frac{1}{m}\mathrm{tr}\left(T_m\left(\Phi_{S|X_1,Z_2}\right)\right) = \frac{1}{2\pi}\int_{\omega\in[0,2\pi]} \Phi_{S|X_1,Z_2}(\omega)\, d\omega.$$

Now using Lemma II.2, we have that

$$\begin{aligned}\Phi_1(\omega) &\triangleq \Phi_{S\bar{X}_1}(\omega)\Phi^{-1}_{\bar{X}_1}(\omega)\Phi^{*}_{S\bar{X}_1}(\omega)\\ &= \Phi_S(\omega) - \Phi_{S|\bar{X}_1}(\omega)\\ &= \Phi_{S|Z_2}(\omega) - \Phi_{S|X_1,Z_2}(\omega).\end{aligned}$$

The second term thus follows from the fact that

$$\begin{aligned}&\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m-\lfloor \alpha_1 m \rfloor} \lambda_{m,i}\\ &= \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m} \lambda_{m,i} 1_{x_{\alpha_1}}(\lambda_{m,i})\\ &= \frac{1}{2\pi}\int_{\omega\in\Omega_1^c} \Phi_{S|Z_2}(\omega) - \Phi_{X_1,Z_2}(\omega)\, d\omega\end{aligned}$$

where x_{α_1} is chosen such that the fraction of eigenvalues smaller than x_{α_1} is equal to $1-\alpha_1$, i.e., such that $F_{\Phi_1}(x_{\alpha_1}) = 1-\alpha$. The last equality follows from Lemma IV.2 with $G(x) = x$ and $\Omega_1 = \{\omega \in [0, 2\pi] : \Phi_1(\omega) > x_{\alpha_1}\}$.

B. Optimal Filtering Strategy

The distortion can be expressed as

$$\begin{aligned}\mathrm{E}[\|S-\hat{S}\|^2] &= \mathrm{E}[\|S - \mathrm{E}[S|Z_1,Z_2]\|^2]\\ &= \frac{1}{2\pi}\int_{\omega\in[0,2\pi]} \Phi_{S|Z_1,Z_2}(\omega)\, d\omega\\ &= \frac{1}{2\pi}\int_{\omega\in\Omega_1} \Phi_{S|Z_1,Z_2}(\omega)\, d\omega\\ &\quad + \frac{1}{2\pi}\int_{\omega\in\Omega_1^c} \Phi_{S|Z_1,Z_2}(\omega)\, d\omega.\end{aligned}$$

We now distinguish two cases.

1) When $\omega \in \Omega_1$, $C_1(\omega) \neq 0$ a.e. since we assumed that the involved PSDs are nonzero a.e. Thus, $\Phi_{S|\bar{Z}_1}(\omega) = \Phi_{S|\bar{X}_1}(\omega)$ a.e. where $\bar{X}_1[m] = X_1[m] - H_2[m] * Z_2[m]$ with $H_2(\omega) = \Phi_{X_1Z_2}(\omega)\Phi^{-1}_{Z_2}(\omega)$. From Lemma II.2, $\Phi_{S|Z_1,Z_2}(\omega) = \Phi_{S|Z_2}(\omega) + \Phi_{S|\bar{Z}_1}(\omega) - \Phi_S(\omega)$ a.e. It thus follows that $\Phi_{S|Z_1,Z_2}(\omega) = \Phi_{S|X_1,Z_2}(\omega)$ a.e.
2) When $\omega \in \Omega_1^c$, $C_1(\omega) = 0$ and $\Phi_{S|Z_1,Z_2}(\omega) = \Phi_{S|Z_2}(\omega)$.

We can thus write

$$\begin{aligned}\mathrm{E}[\|S-\hat{S}\|^2] &= \frac{1}{2\pi}\int_{\omega\in\Omega_1} \Phi_{S|X_1,Z_2}(\omega)\, d\omega\\ &\quad + \frac{1}{2\pi}\int_{\omega\in\Omega_1^c} \Phi_{S|Z_2}(\omega)\, d\omega\\ &= \frac{1}{2\pi}\int_{\omega\in[0,2\pi]} \Phi_{S|X_1,Z_2}(\omega)\, d\omega\\ &\quad + \frac{1}{2\pi}\int_{\omega\in\Omega_1^c} \Phi_{S|Z_2}(\omega) - \Phi_{S|X_1,Z_2}(\omega)\, d\omega\end{aligned}$$

which proves part B and hence the theorem. Note that the term $\Phi_{S\bar{X}_1}(\omega)\Phi^{-1}_{\bar{X}_1}(\omega)$ in $C_1(\omega)$ is not needed for optimality, because it is assumed to be nonzero a.e. This solution is, however, provided by analogy to the finite-dimensional case. □

Appendix V
Proofs of Propositions 5.1, 5.2, and 5.3

Proof of Proposition 5.1: The IBLD in the centralized scenario can be computed using Theorem 4.2 assuming that $S[m] = X_1[m]$, $\alpha_1 = \alpha$, and $Z_2[m] = 0$ (i.e., $\alpha_2 = 0$). We can readily check that $\Phi_S(\omega)$ is strictly positive, symmetric, and strictly decreasing in $[0, \pi]$. The IBLD can thus be expressed using (13) and known integration formulas [34, p. 181] as

$$\begin{aligned}D(\alpha) &= \frac{1}{2\pi}\int_{\omega\in\Omega_1^c} \Phi_S(\omega)\, d\omega\\ &= \frac{1-\rho^2}{2\pi}\int_{\pi\alpha}^{2\pi-\pi\alpha} \frac{1}{1+\rho^2-2\rho\cos\omega}\, d\omega\\ &= 1 - \frac{2}{\pi}\arctan\left(\frac{1+\rho}{1-\rho}\tan\left(\frac{\pi\alpha}{2}\right)\right)\end{aligned}$$

and the proof follows. □

Proof of Proposition 5.2: The IBLD in the perfect side information scenario can be obtained by setting $Y_2 = X_2$ in Theorem 3.1. In this case, using the definition of the vector S, $R_{S|X_1,Y_2} = R_{S|X_1,X_2} = 0_{2m}$ and

$$R_{S|X_2} = \begin{pmatrix} R_{X_1|X_2} & 0_m \\ 0_m & 0_m \end{pmatrix}$$

where 0_m denotes the all-zero matrix of size $m \times m$. Thus, the m largest eigenvalues of $R_{S|X_2}$ are those of $R_{X_1|X_2}$, the

m remaining ones being zero. Using the inverse formula of a Kac–Murdoch–Szegö matrix [9], we have that

$$\begin{aligned} R_{X_1|X_2} &= R_{X_1} - R_{X_1X_2}R_{X_2}^{-1}R_{X_1X_2}^{*} \\ &= \operatorname{diag}\left(1-\rho^2, \frac{1-\rho^2}{1+\rho^2}, \ldots, \frac{1-\rho^2}{1+\rho^2}\right) \end{aligned}$$

i.e., the nonzero eigenvalues $\lambda_{m,i}$ of $R_{S|X_2}$ are all given by $(1-\rho^2)/(1+\rho^2)$ except the maximum one which is equal to $1-\rho^2$. Because the IBLD is not affected by the change of a finite number of eigenvalues, it can be computed from Theorem 3.1 as

$$\begin{aligned} D(\alpha_1,1) &= \lim_{m\to\infty} \frac{\sum_{i=1}^{m-k_1}\lambda_{m,i}}{2m} \\ &= \lim_{m\to\infty} \frac{1-\rho^2}{2(1+\rho^2)}\left(1-\frac{k_1}{m}\right) \\ &= \frac{1-\rho^2}{2(1+\rho^2)}(1-\alpha_1) \end{aligned}$$

where $\alpha_1 \sim k_1/m$. Note that the normalization factor $2m$ accounts for the size of the source vector S as compared to the subsampled observation vectors X_1 and X_2. □

Proof of Proposition 5.3: The IBLD in the partial observation scenario is obtained by setting $Y_2 = 0$ in Theorem 3.1. In this case, using the definition of the vector S, we have

$$R_{S|X_1,Y_2} = R_{S|X_1} = \begin{pmatrix} R_{X_2|X_1} & 0_m \\ 0_m & 0_m \end{pmatrix}.$$

Because $R_{X_2|X_1} = R_{X_1|X_2}$, the IBLD corresponding to the first term of the distortion in (5) is equal to the IBLD of the perfect side information scenario with $\alpha_1 = 0$, i.e.,

$$D_1(\alpha_1,0) = D(0,1) = \frac{1-\rho^2}{2(1+\rho^2)}.$$

The IBLD corresponding to the second term can be obtained by noticing that

$$\begin{aligned} R_{S|Y_2} - R_{S|X_1,Y_2} &= R_S - R_{S|X_1} \\ &= R_{X_1,X_2} - R_{X_1,X_2|X_1} \\ &= \begin{pmatrix} R_{X_1} \\ R_{X_1X_2}^{*} \end{pmatrix} R_{X_1}^{-1}\begin{pmatrix} R_{X_1} & R_{X_1X_2} \end{pmatrix}. \end{aligned}$$

For conforming matrices, we have that $\det(I+AB) = \det(I+BA)$. Thus, for $\lambda \neq 0$, we can write

$$\begin{aligned} &\det(R_{S|Y_2} - R_{S|X_1,Y_2} - \lambda I_{2m}) \\ &\quad = \det\left(\begin{pmatrix} R_{X_1} \\ R_{X_1X_2}^{*} \end{pmatrix} R_{X_1}^{-1}\begin{pmatrix} R_{X_1} & R_{X_1X_2} \end{pmatrix} - \lambda I_{2m}\right) \\ &\quad = \det\left(R_{X_1}^{-1}\begin{pmatrix} R_{X_1} & R_{X_1X_2} \end{pmatrix}\begin{pmatrix} R_{X_1} \\ R_{X_1X_2}^{*} \end{pmatrix} - \lambda I_m\right) \\ &\quad = \det(R_{X_1} + R_{X_1}^{-1}R_{X_1X_2}R_{X_1X_2}^{*} - \lambda I_m). \end{aligned}$$

Hence, the m largest eigenvalues $\lambda_{m,i}$ of $R_{S|Y_2} - R_{S|X_1,Y_2}$ are those of $R_1 = R_{X_1} + R_{X_1}^{-1}R_{X_1X_2}R_{X_1X_2}^{*}$, the m remaining ones being zero. The matrix R_1 is easily seen to be asymptotically equivalent to the Toeplitz matrix $T_m(\Phi_1)$ with $\Phi_1(\omega)$ defined as

$$\begin{aligned} \Phi_1(\omega) &= \Phi_{X_1}(\omega) + \Phi_{X_1}^{-1}(\omega)\Phi_{X_1X_2}(\omega)\Phi_{X_1X_2}^{*}(\omega) \\ &= \frac{(1-\rho^2)(1+4\rho^2+\rho^4+2\rho^2\cos\omega)}{(1+\rho^2)(1+\rho^4-2\rho^2\cos\omega)}. \end{aligned}$$

We can readily check that $\Phi_1(\omega)$ is strictly positive, symmetric, and strictly decreasing in $[0,\pi]$. The IBLD corresponding to the second term of the distortion in (5) can thus be evaluated using Lemma IV.2, (13), and integration formulas from [34, p. 181] as

$$\begin{aligned} &D_2(\alpha_1,0) \\ &= \lim_{m\to\infty} \frac{\sum_{i=1}^{m-k_1}\lambda_{m,i}}{2m} \\ &= \frac{1}{4\pi}\int_{\omega\in\Omega_1^c} \Phi_1(\omega)\,d\omega \\ &= \frac{1}{4\pi}\int_{\pi\alpha_1}^{2\pi-\pi\alpha_1} \frac{(1-\rho^2)(1+4\rho^2+\rho^4+2\rho^2\cos\omega)}{(1+\rho^2)(1+\rho^4-2\rho^2\cos\omega)}\,d\omega \\ &= 1 - \frac{(1-\alpha_1)(1-\rho^2)}{2(1+\rho^2)} \\ &\quad - \frac{2}{\pi}\arctan\left(\frac{1+\rho^2}{1-\rho^2}\tan\left(\frac{\pi\alpha_1}{2}\right)\right) \end{aligned}$$

where $\alpha_1 \sim k_1/m$. Note that the normalization factor $2m$ accounts for the size of the source vector S as compared to the subsampled observation vectors X_1 and X_2. The result follows by adding $D_1(\alpha_1,0)$ and $D_2(\alpha_1,0)$. □

ACKNOWLEDGMENT

The authors would like to thank M. Gastpar for the fruitful discussions and the three anonymous reviewers for their insightful comments.

REFERENCES

[1] M. Schwab, M. Karrenbach, and J. Claerbout, "Making scientific computations reproducible," *Comput. Sci. Eng.*, vol. 2, no. 6, pp. 61–67, 2000.

[2] T.-H. Lin, W. J. Kaiser, and G. J. Pottie, "Integrated low-power communication system design for wireless sensor networks," *IEEE Commun. Mag.*, vol. 42, no. 12, pp. 142–150, Dec. 2004.

[3] N. Gehrig and P. L. Dragotti, "Distributed compression in camera sensor networks," in *Proc. IEEE Workshop Multimedia Signal Process.*, Sep.–Oct. 2004, pp. 311–314.

[4] J. C. Chen, K. Yao, and R. E. Hudson, "Source localization and beamforming," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 30–39, Mar. 2002.

[5] T. Schmid, H. Dubois-Ferrière, and M. Vetterli, "SensorScope: Experiences with a wireless building monitoring sensor network," presented at the Workshop Real-World Wireless Sens. Netw., Stockholm, Sweden, Jun. 20–21, 2005.

[6] M. Gastpar, M. Vetterli, and P. Dragotti, "Sensing reality and communicating bits: A dangerous liaison," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 70–83, Jul. 2006.

[7] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 6, pp. 725–730, Nov. 1972.

[8] R. M. Gray, "Toeplitz and circulant matrices: A review," *Found. Trends Commun. Inf. Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[9] U. Grenander and G. Szegö, *Toeplitz Forms and their Applications*, 2nd ed. New York: Chelsea, 1984.
[10] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 27–41, Jul. 2006.
[11] P. Ishwar, R. Puri, K. Ramchandran, and S. S. Pradhan, "On rate-constrained distributed estimation in unreliable sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 765–775, Apr. 2005.
[12] A. Ribeiro and G. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks – Part I: Gaussian PDF," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.
[13] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5177–5196, Dec. 2006.
[14] H. Nurdin, R. R. Mazumdar, and A. Bagchi, "On estimation and compression of distributed correlated signals with incomplete observations," *Math. Theory Netw. Syst.*, Jul. 2004.
[15] K. Zhang, "Best linear unbiased estimation fusion with constraints," Ph.D. dissertation, Dept. Electr. Eng., Univ. New Orleans, New Orleans, LA, 2003.
[16] K. S. Zhang, X. R. Li, P. Zhang, and H. F. Li, "Optimal linear estimation fusion – Part VI: Sensor data compression," in *Proc. Int. Conf. Inf. Fusion*, Jul. 2003, vol. , pp. 221–228.
[17] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4282–4299, Aug. 2007.
[18] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educat. Psychol.*, vol. 41, pp. 417–441, 1933.
[19] K. Karhunen, "Über lineare Methoden in der Wahrscheinlichkeitsrechnung," *Annales Academiae Scientiarum Fennicae, Series A1: Mathematica-Physica*, vol. 37, pp. 3–79, 1947.
[20] T. Berger, "Multiterminal source coding," presented at the CISM Summer School on the Information Theory Approach to Communications, Udine, Italy, Jul. 1977.
[21] V. Goyal, "Theoretical foundations of transform coding," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 9–21, Sep. 2001.
[22] D. Rebollo-Monedero, S. Rane, and B. Girod, "Wyner-Ziv quantization and transform coding of noisy sources at high rates," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Nov. 2004, vol. 2, pp. 2084–2088.
[23] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod, "High-rate quantization and transform coding with side information at the decoder," *EURASIP J. Signal Process.*, vol. 86, Special Issue on Distributed Source Coding, no. 11, pp. 3160–3179, 2006.
[24] A. Majumdar, K. Ramchandran, and I. Kozintsev, "Distributed coding for wireless audio sensors," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct. 2003, pp. 209–212.
[25] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," presented at the 40th Annu. Allerton Conf. Commun. Control Comput., Monticello, IL, Oct. 2–4, 2002.
[26] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, Special Issue on Advances in Video Coding and Delivery, no. 1, pp. 71–83, Jan. 2005.
[27] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 2001.
[28] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
[29] F. Baumgarte and C. Faller, "Binaural cue coding—Part II: Schemes and applications," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 520–531, Nov. 2003.
[30] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
[31] O. Roy and M. Vetterli, "Rate-constrained beamforming for collaborating hearing aids," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2006, pp. 2809–2813.
[32] P. Brémaud, *Mathematical Principles of Signal Processing: Fourier and Wavelet Analysis*. New York: Springer-Verlag, 2002.
[33] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[34] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. London, U.K.: Academic, 1994.