The Distributed Karhunen-Loève Transform*

Michael Gastpar⋆, Pier-Luigi Dragotti° and Martin Vetterli⋆†

March 21, 2003

⋆Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA 94720, USA
Email: gastpar@eecs.berkeley.edu

°Department of Electrical and Electronic Engineering
Imperial College, London, SW7 2BT, UK
Email: p.dragotti@ic.ac.uk

†Audiovisual Communications Lab
Ecole Polytechnique Fédérale (EPFL), Lausanne, Switzerland
Email: martin.vetterli@epfl.ch

Abstract

The Karhunen-Loève transform (KLT) is a key element of many signal processing tasks, including approximation, compression, and classification. Many recent applications involve *distributed* signal processing where it is not generally possible to apply the KLT to the signal; rather, the KLT must be approximated in a distributed fashion. This paper investigates such distributed approximations to the KLT. First, we present explicit solutions to special cases, including a partial KLT (where only a subset of the sources is observed), a conditional KLT (where some sources act as side information), and the combination of these two special cases. These results are used to derive an algorithm that finds the best distributed approximation to the KLT.

This distributed transform has potential applications in sensor networks, distributed databases, surveillance systems and hyper-spectral imagery.

1 Introduction

The approximation or compression of an observed signal is a central and widely studied problem in signal processing and communication. The Karhunen-Loève transform (KLT) has always played a pivotal role in this context. Assume, for instance, that the observed signal is a random vector X with covariance matrix Σ and that the statistics of the source are known. Then to solve the approximation problem, one can apply the KLT to X to obtain uncorrelated components and the best K-order approximation of the source is given by the K components corresponding to the K largest eigenvalues of Σ. In the case of compression, the uncorrelated components can be compressed independently and more rate can be allocated to the components related to the largest eigenvalues of Σ. This compression process is widely known as transform coding and, if the input source is a Gaussian source,

*The material in this paper was presented in part at MMSP 2002 [1] and at DCC 2003 [2], and was supported in part by the Swiss National Science Foundation in the framework of the National Competence Center in Research for Mobile Information and Communication Systems.

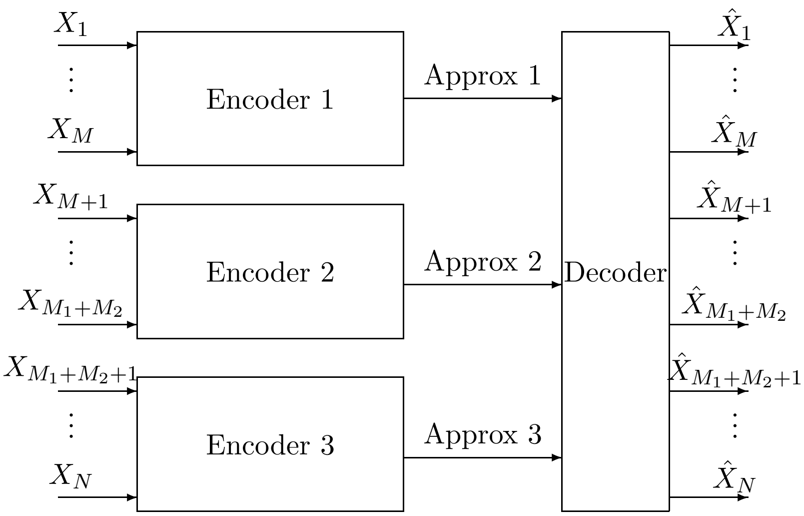

Figure 1: The distributed KLT problem: Distributed compression of multiple correlated vector sources.

it is possible to show that it is optimal [3]. For on excellent review on transform coding and on the optimality of the KLT in this context, we refer to the exposition in [4].

In the present work, we investigate a related scenario where there are multiple sensors each observing only a part of the vector X (see Figure 1). These sensors transmit an approximation of the observed subvector to a fusion center and cannot communicate with each others. Thus, signal processing must be done in a distributed fashion and the full KLT cannot be applied to the data. Therefore, the original approximation and compression problems change in these circumstances significantly. In this paper, we show how the concept of the KLT extends to such a scenario.

Notice that, in one particular case, the distributed compression problem has been already solved by Wyner and Ziv [5] and some constructive approaches to implement the Wyner and Ziv coder have been proposed recenty [6, 7, 8, 9, 10]. We extend the result of Wyner and Ziv to the case of correlated vectors and this leads to the introduction of a new transform called the conditional KLT.

The paper is organized as follows: Section 2 formally states the problem leading to the distributed KLT. Section 3 and 4 studies special cases of the general problem. Those special cases are called the partial and the conditional KLT. Explicit solutions are found for both cases. For the conditional KLT, it is shown in Section 4 that the problem splits into separate Wyner-Ziv problems. The two solutions are combined in Section 5. In Section 6, we show that the results of this paper directly lead to an algorithm to solve the distributed KLT problem. Finally, we conclude in Section 7.

2 The Distributed KLT Problem

The problem leading to the distributed KLT is shown in Figure 1: There are L terminals, each of which samples a part of the random vector X of length N,

$$X \stackrel{def}{=} (X_1, X_2, \ldots, X_N), \tag{1}$$

with zero mean and covariance matrix Σ. The terminals cannot communicate with each other. Each terminal furnishes a certain approximation of its samples to a central decoder. The goal of the central decoder is to produce an estimate $\hat{X}$ in such a way as to minimize the mean-squared error $E||X - \hat{X}||^2$.

For the approximation furnished by the terminals, two different scenarios are of interest to us:

1. *Approximation.* Terminal i furnishes a k_i-dimensional approximation of its sampled vector. What are the best approximation spaces for the L terminals?

2. *Compression.* Terminal i furnishes a compressed description using R_i bits per sample. For a required maximum average distortion D, what are the rate tuples $(R_1, R_2, \ldots, R_L)$ permitting to satisfy the constraints? In other words, the goal of this consideration is to determine the achievable rate region for a fixed distortion D.

To illustrate the point, suppose that X is a vector of jointly Gaussian random variables. If there is only one terminal that senses all the components of the vector X, the solution for both the approximation and the compression scenario is given by first applying the KLT to the vector X, yielding a transformed vector Y with independent components. For the approximation framework, the best k-dimensional approximation space is given by the eigenvectors corresponding to the k largest eigenvalues. For the compression framework, the bit allocation between the components of Y is determined by the eigenvalues.

If there are multiple terminals, each sensing only a part of the vector X, as illustrated in Figure 1, then it is not possible to apply the KLT in general; rather, it has to be approximated in a distributed fashion. What is the best such approximation to the full KLT of the vector X? Recall that in our considerations, the "best" approximation is the one minimizing the mean-squared error $E||X - \hat{X}||^2$.

3 The Partial KLT

In this section, we study the problem of partial observation or subsampling, as shown in Figure 2. For simplicity, we assume that the random vector X has mean zero. Without loss of generality, we suppose that the first M components of X are sampled, and collect them in the vector

$$X_S \stackrel{def}{=} (X_1, X_2, \ldots, X_M), \tag{2}$$

with zero mean and covariance matrix Σ_S. The non-sampled components of X are denoted by

$$X_{S^c} \stackrel{def}{=} (X_{M+1}, X_{M+2}, \ldots, X_N), \tag{3}$$

with zero mean and covariance matrix Σ_{S^c}. Moreover, we denote the covariance matrix between X_S and X_{S^c} as $\Sigma_{SS^c} \stackrel{def}{=} Cov(X_S, X_{S^c})$. The presence of the hidden part X_{S^c} — not observed, but to be reconstructed — alters the problem significantly.

As outlined above, two points of view are of particular interest to us:

1. *Approximation.* The M-dimensional vector X_S of correlated random variables is approximated in a k-dimensional space. What is the best such space? If there is no

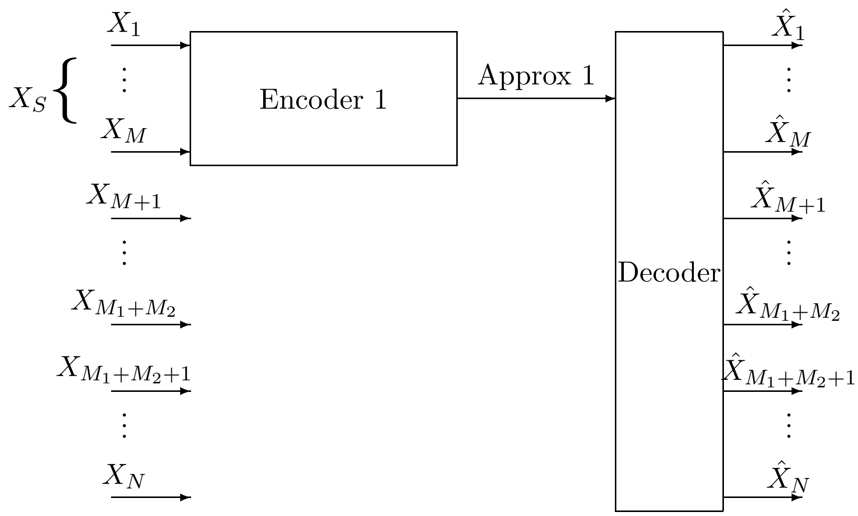

Figure 2: Compression of a subsampled set of correlated random variables.

hidden part ($M = N$), the best choice is well known to be the eigenvectors corresponding to the k largest eigenvalues of Σ_S. But if there is a hidden part ($M < N$), it is not optimal simply to take the k largest eigenvalues of Σ_S since the non-sampled part may depend crucially on some of the smaller eigenvalues. In this section, we determine the optimal k-dimensional space.

2. *Compression.* The M-dimensional vector X_S of correlated random variables is compressed using a total of R bits. What is the optimal compression for a decoder that wants to minimize the distortion $E||X - \hat{X}||^2 = \sum_{i=1}^{N} E|X_i - \hat{X}_i|^2$? For the compression problem, much of our consideration is limited to the case where X is a vector of jointly *Gaussian* random variables. For that case, if there is no hidden part ($M = N$), the best compression is well known: apply the Karhunen-Loève transform (KLT). This gives M independent random variables that can be compressed separately from one another. The bits are divided up according to “inverse water-filling”: the stronger components receive more, the weaker less. But if there is a hidden part ($M < N$), this is no longer optimal: some otherwise unimportant part of X_S may be vital for X_{S^c}. In this section, we show how the transform has to be altered to account for the hidden part.

The discussion of this section is limited to the case where X_{S^c} is related to X_S by

$$X_{S^c} = AX_S + V, \tag{4}$$

where A is a constant matrix, and V is a random vector independent of X_S.

Denoting the M-dimensional identity matrix by I_M, the *partial KLT* is defined as follows:

Definition 1. Denote the eigendecomposition as $\binom{I_M}{A}\Sigma_S(I_M A^H) = Q\Lambda Q^H$, where Λ is a diagonal matrix and Q is unitary. The *partial KLT* of X_S with respect to X_{S^c} is the linear transform characterized by the matrix $P = Q^H\binom{I_M}{A}$.

The transformed version of X_S is denoted by $Y_S = PX_S$, and the variances of the components of Y by $\sigma_i^2 = Var(Y_i^2)$. Note that only M components of Y_S are non-zero.

Lemma 1 (properties of the partial KLT). *The partial KLT has the following key properties:*

1. $\text{rank}(P) = M$*, but* P *is not generally a unitary matrix.*
2. *The components of* Y_S *are uncorrelated. If* X_S *is a vector of jointly Gaussian random variables, then they are independent.*

Proof. The first property follows directly from the fact that Q is a unitary matrix, and that $\binom{I_M}{A}$ has rank M. The second property follows by evaluating

$$Cov(Y_S, Y_S) \quad = \quad PCov(X_S, X_S)P^H = Q^H \binom{I_M}{A} Cov(X_S, X_S)(I_M A^H)Q. \tag{5}$$

By construction of Q, the last expression is a diagonal matrix, implying that the components of Y_S are uncorrelated. □

3.1 Approximation Problem

The goal is to minimize the estimation error $E||X - \hat{X}||^2$. Simply by the definition of the symbol $|| \cdot ||^2$,

$$\begin{aligned} E||X - \hat{X}||^2 &= E||X_S - \hat{X}_S||^2 + E||X_{S^c} - \hat{X}_{S^c}||^2 \\ &= E||X_S - \hat{X}_S||^2 + E||AX_S + V - \hat{X}_{S^c}||^2 \end{aligned} \tag{6}$$

The key step is to relate $\hat{X}_S$ and $\hat{X}_{S^c}$. This is enabled by the following standard lemma (see e.g. [11, Thm.34.8]):

Lemma 2. *Irrespective of the statistics of* W *and* X_S*, it is true that the minimum mean-square error estimator of* X_S *given* W *can be expressed as*

$$\hat{X}_S \quad = \quad E[X_S|W]. \tag{7}$$

By the linearity of expectation, $E[AX_S|W] = AE[X_S|W]$*.*

From this insight, it follows that

$$\hat{X}_{S^c} \quad = \quad A\hat{X}_S. \tag{8}$$

This permits to replace $\hat{X}_{S^c}$ to obtain

$$\begin{aligned} E||X - \hat{X}||^2 &= E||X_S - \hat{X}_S||^2 + E||AX_S + V - A\hat{X}_S||^2 \\ &= E||X_S - \hat{X}_S||^2 + E||AX_S - A\hat{X}_S||^2 + E||V||^2 \\ &= E||\binom{I_M}{A}X_S - \binom{I_M}{A}\hat{X}_S||^2 + E||V||^2, \end{aligned} \tag{9}$$

where the second equality follows because V and X_S are independent.

The key step is to rewrite this in terms of the partial KLT of X_S as

$$E||X - \hat{X}||^2 \quad = \quad E||Y_S - \hat{Y}_S||^2 + E||V||^2 = \sum_{i=1}^{N} E|Y_i - \hat{Y}_i|^2 + E||V||^2. \tag{10}$$

Suppose that the approximation consists of k components of Y_S and denote the set of their indices by T. The resulting distortion is

$$E||X - \hat{X}||^2 \quad = \quad \sum_{i \in T^c} \sigma_i^2 + E||V||^2, \tag{11}$$

where T^c is the complement of T in $\{1, \ldots, N\}$. This permits to characterize the optimum choice of a set T as follows:

Theorem 3. *The best k-dimensional approximation space for the subsampling problem of Figure 2 is spanned by the columns of P^H corresponding to the k largest variances σ_i^2, i.e., the k largest eigenvalues of $\binom{I_M}{A}\Sigma_S(I_M A^H)$.*

Example 1. A toy example illustrating the basic issue is the following: Suppose that a Gaussian random vector X has mean zero and the following covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0.1 & 0.1 \\ 0 & 0.1 & 0.25 & 0 \\ 0.1 & 0.25 & 1 & 0.25 \\ 0.1 & 0 & 0.25 & 1 \end{pmatrix}. \tag{12}$$

Suppose that the first two components are sampled by the terminal, i.e., $M = 2$. Since X is jointly Gaussian, the matrix A in Equation (4) is found to be $A = \Sigma_{S^cS}\Sigma_S^{-1}$. The terminal is asked to provide a 1-dimensional approximation. For $\sigma_1^2 = 0.11$, applying the usual KLT to the first two components is simple in this example: the first two components are uncorrelated, hence the KLT is the identity. Picking the eigenvector corresponding to the larger eigenvalue of Σ_S incurs a distortion of $D_{klt} = 1.9182$. Using the partial KLT discussed in this section, and hence making the optimal choice, results in a distortion of $D_{pklt} = 1.3795$, and the transform is

$$P = \begin{pmatrix} 1.1119 & 2.6353 \\ 1.1902 & -0.5524 \end{pmatrix}. \tag{13}$$

It is clear that this matrix is substantially different from the usual KLT (applied to the first two components).

3.2 Compression Problem

The rate-distortion function for the scenario of Figure 2 follows directly from standard arguments. More precisely, the following holds:

Theorem 4 (rate-distortion function of subsampling). *For the rate-distortion problem of Figure 2,*

$$R_S(D) = \min I(X_S; \hat{X}) \tag{14}$$

where the minimum is over all conditional densities $p(\hat{x}|x_S)$ that satisfy $Ed(X, \hat{X}) \leq D$.

Proof. The proof is given in Appendix A.

If the distortion measure is the mean-squared error, this can be simplified considerably:

Corollary 5 (rate-distortion function of subsampling under MSE). *For the rate distortion problem of Figure 2,*

$$R_S(D) = \min I(Y_S; \hat{Y}_S) \tag{15}$$

where $Y_S = PX_S$, where the minimum is over all conditional densities $p(\hat{y}_S|y_S)$ that satisfy

$$\sum_{i=1}^{N} E|Y_{S,i} - \hat{Y}_{S,i}|^2 + E||V||^2 \leq D. \tag{16}$$

Proof. This is a corollary to Theorem 4. The fact that $\hat{X}$ is the minimum mean-squared error estimate implies (through Lemma 2) that we can rewrite $I(X_S; \hat{X}) = I(X_S; \hat{X}_S)$. But since the partial KLT has (full) rank M, $I(X_S; \hat{X}_S) = I(Y_S; \hat{Y}_S)$. The simplified distortion expression follows directly from Equation (10). □

The rest of the discussion of the compression problem is limited to the case where X is a vector of jointly *Gaussian* random variables. This clearly satisfies (4); the vector V turns out to be Gaussian, too.

In the Gaussian case, the components of Y_S are *independent* random variables. Therefore, the minimum of $I(Y_S, \hat{Y}_S)$ is achieved by a $\hat{Y}_S$ whose ith component $\hat{Y}_i$ depends only on Y_i. This leads to the following theorem.

Theorem 6 (rate-distortion function of the subsampled Gaussian). *The rate-distortion function for the subsampled Gaussian, illustrated in Figure 2, is given by*

$$R_S(D) = \min_{D_i} \sum_{i=1}^{M} \max \left\{ \frac{1}{2} \log_2 \frac{\sigma_i^2}{D_i}, 0 \right\}, \tag{17}$$

where σ_i^2 are the M eigenvalues of the matrix $\binom{I_M}{A} \Sigma_S (I_M A^H)$, and the minimum is over all sets $\{D_i\}_{i=1}^M$ satisfying $\sum_{i=1}^M D_i \leq D - E||V||^2$.

Proof. Corollary 5 proves that the subsampling problem is equivalent to the standard mean-square rate-distortion problem for vector sources. In the Gaussian case, this is particularly simple because the components of the vector source Y_S are independent (which follows from the properties of the partial KLT). The solution to this problem is well known [12], see also [13] and [14, p. 347]. □

This theorem says that an optimal compression system is the one given in Figure 3: Apply the partial KLT, and compress the components separately, using the appropriate bit allocation. The key difference to the usual Gaussian rate-distortion problem is that the partial KLT is *not* an orthonormal transform.

Example 2. Consider again the covariance matrix given in Example 1, with $\sigma_1^2 = 0.11$. The first two components are sampled and can be encoded using a total rate R. The systematic error for this example is $E||V||^2 = 1.1932$. The rate-distortion trade-off is shown in Figure 4. The solid line is the rate-distortion function $R_S(D)$ (i.e., incorporating the partial KLT). The dotted line is the performance for a compression scheme that ignores the hidden part when encoding. At decoding time, the hidden part is estimated optimally from the available information. The figure witnesses a clear advantage for the partial KLT, illustrating the fact that the hidden part does alter the compression problem significantly. In the limit of low rates, as $R \to 0$, it is clear that both schemes have the same performance: No information is transmitted. In the limit of high rates, both schemes end up encoding the observations perfectly, and again, the same distortion results.

Remark (best sensor placement). For given statistics Σ and desired distortion D, what is the best "placement" of M sensors? In other words, what choice of M components of X minimizes the rate $R_S(D)$ at the desired distortion D? The solution to this problem is given by Theorem 6: Compute $R_S(D)$ for all sets S with cardinality M.

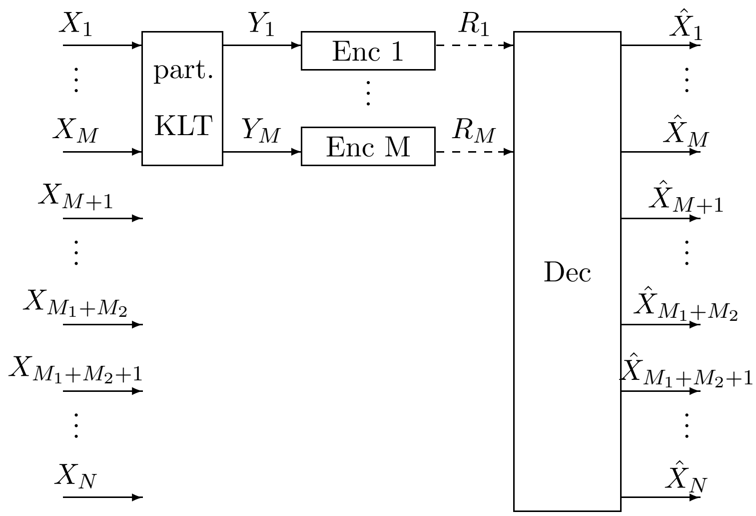

Figure 3: Compression of a subsampled set of correlated Gaussian sources using the partial KLT. This system is shown to perform optimally.

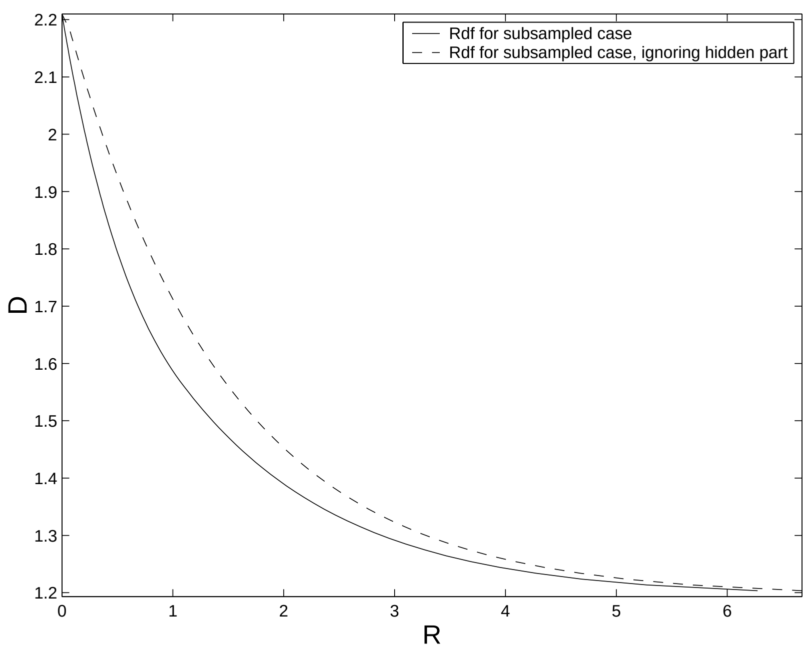

Figure 4: The rate-distortion function for the subsampling scenario of Example 2.

4 The Conditional KLT

In this section, we study the scenario of Figure 5. This is (in some sense) the complement of Figure 2: Here, the non-encoded random variables are known perfectly to the decoder, whereas there, they were not known at all. Intermediate cases will be studied in the next section. By analogy to Section 3, two points of view are of interest to us:

1. *Approximation.* The M-dimensional random vector X_S is approximated in a k-dimensional space. What is the best such space if at reconstruction time, we know a

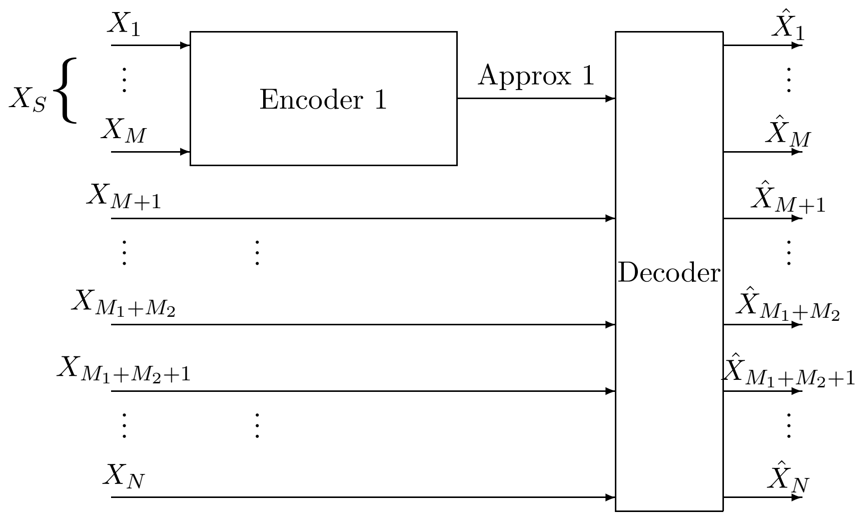

Figure 5: Compression of a subsampled set of correlated random variables for a decoder that has side information.

random vector X_{S^c} which is correlated with X_S? This best k-dimensional space can be determined easily using the conditional KLT.

2. *Compression.* The M-dimensional random vector X_S is compressed using a total of R bits for a decoder that has access to X_{S^c}. What is the optimal compression scheme? For $M = 1$ and $N = 2$, the problem of Figure 5 has been solved by Wyner and Ziv [5]. Here, we restrict attention to the case where X is a jointly *Gaussian* random vector, and we extend the result of [5] to arbitrary M and N. We show that the solution can be found using the *conditional KLT*: It transforms X_S into a vector Y_S whose components are conditionally independent given X_{S^c}. Just like in the standard KLT, each such component is then compressed *separately* by applying the Wyner-Ziv solution; we determine the bit allocation between these M Wyner-Ziv problems.

The main tool of this section is the *conditional KLT*:

Definition 2. The *conditional KLT* of X_S with respect to X_{S^c} exists if

$$\{Cov(X_i, X_j | X_{S^c} = x_{S^c})\}_{i,j} \quad = \quad \Sigma_{S|S^c}, \tag{18}$$

i.e., $Cov(X_i, X_j | X_{S^c} = x_{S^c})$ does not depend on the value of x_{S^c}. In that case, it is the unitary matrix C such that

$$C \Sigma_{S|S^c} C^H \quad = \quad \mathrm{diag}(\lambda_1, \ldots, \lambda_M), \tag{19}$$

where diag(x) denotes a diagonal matrix with elements x on the diagonal, and ^{H} the Hermitian transpose.

Remark. If $Cov(X_i, X_j | X_{S^c} = x_{S^c})$ depends on the value of x_{S^c}, one can define a conditional KLT of X_S with respect to the event $X_{S^c} = x_{S^c}$. For the scope of this paper, however, we restrict to cases according to the above definition.

The transformed version of X_S is denoted by

$$Y_S = CX_S, \tag{20}$$

and $\lambda_i^2 = Var(Y_i|X_{S^c} = x_{S^c})$, which by assumption does not depend on x_{S^c}.

Lemma 7 (properties of the conditional KLT). *The conditional KLT has the following key properties:*

1. *C is an orthonormal transform*
2. *The components of the vector Y_S are conditionally uncorrelated given X_{S^c}. If X is a vector of jointly Gaussian random variables, then they are conditionally independent.*

Proof. The first property is a direct consequence of the fact that C is the matrix of eigenvectors of Σ_S. The second property follows by evaluating

$$Cov(Y_S, Y_S|X_{S^c}) = CCov(X_S, X_S|X_{S^c})C^H = C\Sigma_{S|S^c}C^H. \tag{21}$$

By construction of C, the last expression is a diagonal matrix. □

The discussion of this section is limited to X_S for which the conditional KLT with respect to X_{S^c} exists. This is true for the interesting case where X_S and X_{S^c} are related by

$$X_S = BX_{S^c} + U, \tag{22}$$

where B is a constant matrix, and U is a random vector independent of X_{S^c}, with covariance matrix Σ_U. In that case, $\Sigma_{S|S^c} = \Sigma_U$, and the conditional KLT is given by the eigendecomposition of Σ_U,

$$C\Sigma_U C^H = \text{diag}(\lambda_1, \ldots, \lambda_M). \tag{23}$$

4.1 Approximation Problem

For any fixed value of the side information $X_{S^c} = x_{S^c}$, the goal is to minimize the mean-squared error,

$$E\left[||X_S - \hat{X}_S||^2 \middle| X_{S^c} = x_{S^c} \right]. \tag{24}$$

It is established that since the conditional KLT of X_S with respect to X_{S^c} exists (in the sense of Definition 2), the optimal approximation X_{S^c} does *not* depend on the value x_{S^c}. The key step is to rewrite this in the conditional KLT domain, to obtain

$$\begin{aligned} E\left[||X_S - \hat{X}_S||^2 \middle| X_{S^c} = x_{S^c} \right] &= E\left[||CX_S - C\hat{X}_S||^2 \middle| X_{S^c} = x_{S^c} \right] \qquad (25) \\ &= E\left[||Y_S - \hat{Y}_S||^2 \middle| X_{S^c} = x_{S^c} \right] \\ &= \sum_{i=1}^{M} E\left[|Y_i - \hat{Y}_i|^2 \middle| X_{S^c} = x_{S^c} \right], \qquad (26) \end{aligned}$$

where (25) follows from Property 1) of the conditional KLT, (26) from Property 2).

Denote the set of k dimensions for the approximation by T, and the set of $M-k$ discarded dimensions by T^c. The resulting distortion is

$$E\left[||X_S - \hat{X}_S||^2 \Big| X_{S^c} = x_{S^c}\right] \;=\; \sum_{i\in T^c} E\left[|Y_i - \hat{Y}_i|^2 \Big| X_{S^c} = x_{S^c}\right], \tag{27}$$

which clearly assumes its minimum by choosing $\hat{Y}_i = E[Y_i|X_{S^c} = x_{S^c}]$. This estimate *does* depend on the value of the side information, x_{S^c}; it is formed only at the destination. Hence,

$$E\left[||X_S - \hat{X}_S||^2 \Big| X_{S^c} = x_{S^c}\right] \;=\; \sum_{i\in T^c} Var\left(Y_i | \, X_{S^c} = x_{S^c}\right)$$

By assumption, the conditional KLT exists (in the sense of Definition 2), which implies that $Var\left(Y_i| \, X_{S^c} = x_{S^c}\right)$ does not depend on x_{S^c}. Therefore, we can write

$$E\left[||X_S - \hat{X}_S||^2 \Big| X_{S^c} = x_{S^c}\right] \;=\; \sum_{i\in T^c} \lambda_i^2. \tag{28}$$

This expression is minimized if T contains the indices corresponding to the largest λ_i^2. We have proved the following theorem:

Theorem 8. *The best k-dimensional approximation space for the side information problem of Figure 5 is spanned by the k columns of C^H corresponding to the k largest conditional variances λ_i^2, i.e., the k largest eigenvalues of $\Sigma_{S|S^c}$.*

Remark. This theorem also says that knowing the actual value of the side information x_{S^c} at the encoder does not change the solution to our problem: The optimal choice of k dimensions to approximate X_S remains the same. (Recall that this is true under (22).) Note however that the *coefficients* of the approximation do change when x_{S^c} is known at the encoder, since in that case, Bx_{S^c} can be subtracted at the encoder.

Example 3. A toy example illustrating the basic issue is the following: Suppose that a Gaussian random vector X has mean zero and the following covariance matrix:

$$\Sigma \;=\; \begin{pmatrix} \sigma_1^2 & 0 & 0.1 & 0.1 \\ 0 & 0.1 & 0.25 & 0 \\ 0.1 & 0.25 & 1 & 0.25 \\ 0.1 & 0 & 0.25 & 1 \end{pmatrix}. \tag{29}$$

Suppose that the first two components are sampled by the terminal, i.e., $M = 2$. Since X is jointly Gaussian, the matrix B in Equation (22) is easily found to be $B = \Sigma_{SS^c}\Sigma_S^{-1}$, and the covariance matrix of U can be written as $\Sigma_U = \Sigma_S - \Sigma_{SS^c}\Sigma_{S^c}^{-1}\Sigma_{S^cS}$. The terminal is asked to provide a 1-dimensional approximation. For $\sigma_1^2 = 0.1$, applying the usual KLT to the first two components is simple in this example: the first two components are uncorrelated, hence the KLT is the identity. Picking the eigenvector corresponding to the larger eigenvalue of Σ_S incurs a distortion of $D_{klt} = 0.0720$. Using the conditional KLT discussed in this section, and hence making the optimal choice, results in a distortion of $D_{cklt} = 0.0264$, and the transform is

$$C \;=\; \begin{pmatrix} -0.9447 & 0.3280 \\ -0.3280 & -0.9447 \end{pmatrix}. \tag{30}$$

4.2 Compression Problem

The discussion is limited to the case where X is a vector of jointly Gaussian random variables. Relationship (22) holds, i.e.,

$$X_S = BX_{S^c} + U, \tag{31}$$

where U is a Gaussian random vector independent of X_{S^c}. Hence, in the Gaussian case, the conditional KLT is simply the standard KLT of the random vector U.

From the results of [5, 15], the smallest R (in Figure 5) permitting a distortion of D is

$$R(D) = \min_{p(w|x_S)} I(X_S; W|X_{S^c}), \tag{32}$$

where the minimization is over all auxiliary random variables W for which there exists a function $\hat{X}_S(W, X_{S^c})$ such that $E||X_S - \hat{X}_S(W, X_{S^c})||^2 \leq D$.

This can be rewritten in the conditional KLT domain:

$$R(D) = \min_{p(w|y_S)} I(Y_S; W|X_{S^c}), \tag{33}$$

where the minimization is over all auxiliary random variables W for which there exists a function $\hat{Y}_S(W, X_{S^c})$ such that $E||Y_S - \hat{Y}_S(W, X_{S^c})||^2 \leq D$.

Due to Property 1) of the conditional KLT, the distortion constraint is unchanged. Property 2) permits to simplify the mutual information expression. One can artificially introduce auxiliary random variables $W_1, W_2, \ldots, W_M$, where W_i is allowed to depend arbitrarily on Y_S. With this, we can write out

$$\begin{aligned} R &= \min_{p(w|y_S)} I(Y_S; W|X_{S^c}) && (34)\\ &\geq \min_{p(w_1,\ldots,w_M|y_S)} I(Y_S; W_1, \ldots, W_M|X_{S^c}) && (35)\\ &\stackrel{(a)}{\geq} \min_{p(w_1,\ldots,w_M|y_S)} \sum_{i=1}^{M} I(Y_i; W_i|X_{S^c}) && (36) \end{aligned}$$

where (a) holds because $Y_1, \ldots, Y_M$ are conditionally independent given X_{S^c}. Equality is attained in (a) when the auxiliary W_i depends only on Y_i, rather than on all of Y_S.

This permits to rewrite (36) as

$$R \geq \sum_{i=1}^{M} \min_{p(w_i|y_S)} I(Y_i; W_i|X_{S^c}). \tag{37}$$

The solution to the minimization problem inside the sum has been found by Wyner and Ziv [5]. Suppose that the distortion for the i-th component is D_i. Using their result, we can give the following theorem:

Theorem 9. *The rate-distortion function for the problem with side information, illustrated in Figure 5, is given by*

$$R(D) = \min_{D_i} \sum_{i=1}^{M} \max\left\{\frac{1}{2}\log_2 \frac{\lambda_i^2}{D_i}, 0\right\} \tag{38}$$

where λ_i^2 are the M eigenvalues of the matrix $\Sigma_{S|S^c}$ and the minimum is over all sets $\{D_i\}_{i=1}^M$ satisfying $\sum_{i=1}^M D_i \leq D$.

This result is identical to the result found for Gaussian sources with memory in [16]. The theorem says that the compression problem of Figure 5 can be optimally solved by the system shown in Figure 6: A conditional KLT, followed by *separate* compression of each component (using the techniques described in [5]).

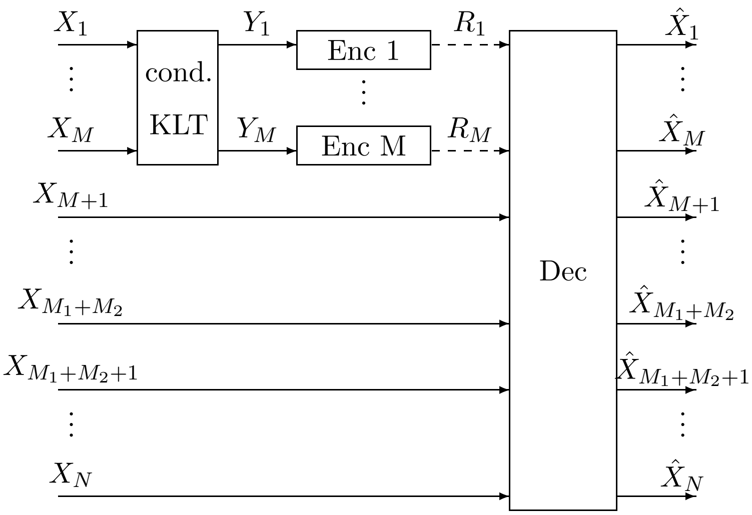

Figure 6: Compression of a set of correlated Gaussian sources for a decoder that has side information using the conditional KLT. This system is shown to perform optimally.

5 Combining the Partial and the Conditional KLT

The solutions found in Section 3 and 4 can be used to determine the solution for the problem illustrated in Figure 7: there is side information *and* a hidden part. The vector X_{S^c} is transformed by the linear transform C_2 into a vector Y_{S^c},

$$Y_{S^c} \quad = \quad C_2 X_{S^c}. \tag{39}$$

The first k_2 components of Y_{S^c} are side information, denoted by Y'_{S^c}; the remaining components are the hidden part, denoted by Y''_{S^c}. For notational purposes, we write

$$Y'_{S^c} \quad = \quad C_{2a} X_{S^c}, \tag{40}$$

where C_{2a} are simply the k_2 first rows of the matrix C_2. The discussion of this section is limited to the case where X is a vector of jointly Gaussian random variables.

5.1 Approximation Problem

The goal is to minimize the distortion $E\left[||X-\hat{X}||^2|Y'_{S^c}=y'_{S^c}\right]$ in the scenario depicted in Figure 7. Recall that in the present section, we assume that X is a vector of jointly Gaussian random variables. Since $Y'_{S^c}=C_{2a}X_{S^c}$, X is also jointly Gaussian with Y'_{S^c}. Therefore, we can write

$$X_{S^c} \quad = \quad AX_S + A_2 Y'_{S^c} + V, \tag{41}$$

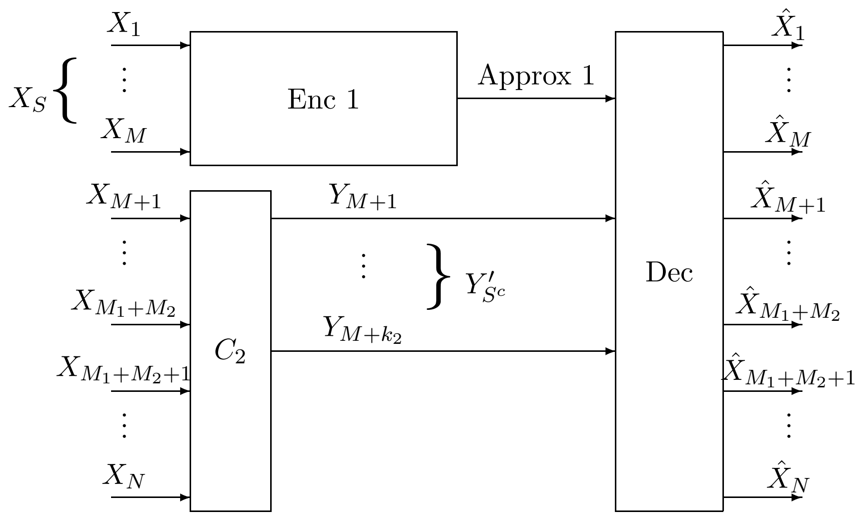

Figure 7: Compression involving subsampling and side information.

where (X_S, Y'_{S^c}) and V are independent Gaussian random vectors. By the same token, we can write

$$\binom{I_M}{A} X_S \quad = \quad BY'_{S^c} + U, \tag{42}$$

where Y'_{S^c} and U are independent Gaussian random vectors. We denote the covariance matrix of the vector U by Σ_U. The parameters in Equations (41) and (42) follow form standard results on multivariate Gaussian distributions; the corresponding formulae are given in Appendix B. By analogy to the derivation of the partial KLT in Section 3, we can write the distortion as

$$\begin{aligned}
&E\left[||X - \hat{X}||^2 | Y'_{S^c} = y'_{S^c}\right] \\
&\quad = E\left[||X_S - \hat{X}_S||^2 | Y'_{S^c} = y'_{S^c}\right] + E\left[||X_{S^c} - \hat{X}_{S^c}||^2 | Y'_{S^c} = y'_{S^c}\right] \\
&\quad = E\left[||X_S - \hat{X}_S||^2 | Y'_{S^c} = y'_{S^c}\right] + E\left[||AX_S + A_2 Y'_{S^c} + V - \hat{X}_{S^c}||^2 | Y'_{S^c} = y'_{S^c}\right] \\
&\quad = E\left[||X_S - \hat{X}_S||^2 | Y'_{S^c} = y'_{S^c}\right] + E\left[||AX_S + V - A\hat{X}_S||^2 | Y'_{S^c} = y'_{S^c}\right] \qquad (43) \\
&\quad = E\left[||X_S - \hat{X}_S||^2 | Y'_{S^c} = y'_{S^c}\right] + E\left[||AX_S - A\hat{X}_S||^2 + ||V||^2 | Y'_{S^c} = y'_{S^c}\right] \\
&\quad = E\left[||\binom{I_M}{A} X_S - \binom{I_M}{A} \hat{X}_S||^2 | Y'_{S^c} = y'_{S^c}\right] + E||V||^2, \qquad (44)
\end{aligned}$$

where (43) follows again from the fact that $\hat{X}_{S^c}$ is the mininum mean-squared error estimator of X_{S^c}, using Lemma 2. Equation (42) implies that the conditional KLT of $\binom{I_M}{A} X_S$ with respect to Y'_{S^c} exists, and we denote it by C_P. Define

$$Y_S \quad = \quad C_P \binom{I_M}{A} X_S. \tag{45}$$

Note that Y_S is a vector of length $N - k_2$, but only M of the components Y_i of Y_S have non-zero conditional variance $Var(Y_i|Y'_{S^c}) = \nu_i^2$. Since C_P is a unitary transform,

$$\begin{aligned} & E\left[||\binom{I_M}{A}X_S - \binom{I_M}{A}\hat{X}_S||^2|Y'_{S^c} = y'_{S^c}\right] + E||V||^2 \\ &= E\left[||C_P\binom{I_M}{A}X_S - C_P\binom{I_M}{A}\hat{X}_S||^2|Y'_{S^c} = y'_{S^c}\right] + E||V||^2. \\ &= E\left[||Y_S - \hat{Y}_S||^2|Y'_{S^c} = y'_{S^c}\right] + E||V||^2. \end{aligned} \tag{46}$$

Then, since the components of Y_S are conditionally independent given Y'_{S^c}, this can be further simplified to yield

$$E\left[||X - \hat{X}||^2|Y'_{S^c} = y'_{S^c}\right] = \sum_{i=1}^{N-k_2} E\left[|Y_i - \hat{Y}_i|^2\middle| Y'_{S^c} = y'_{S^c}\right] + E||V||^2 \tag{47}$$

Suppose that the approximation consists of k components of Y_S and denote the set of their indices by T. The resulting distortion is

$$E\left[||X - \hat{X}||^2|Y'_{S^c} = y'_{S^c}\right] = \sum_{i \in T^c} \nu_i^2 + E||V||^2, \tag{48}$$

where T^c is the complement of T in $\{1, \ldots, N - k_2\}$. This permits to characterize the optimum choice of a set T as follows:

Theorem 10 (partial-conditional KLT). *The best k-dimensional subspace is spanned by the k columns of $(I_M A^H)C_P^H$ corresponding to the k largest conditional variances ν_i^2, i.e., the k largest eigenvalues of the conditional covariance matrix $Cov(\binom{I_M}{A}X_S, \binom{I_M}{A}X_S|Y'_{S^c}) = \Sigma_U$.*

5.2 Compression Problem

For the scenario of Figure 7, suppose now that Encoder 1 furnishes an R bit approximation to its observation. The rate-distortion function for this scenario follows almost immediately from the analysis of Section 4.2. It can be described by the following theorem:

Theorem 11. *The rate-distortion function for the problem with side information, illustrated in Figure 5, where X is a vector of jointly Gaussian random variables, is given by*

$$R(D) = \min_{D_i} \sum_{i=1}^{M} \max\left\{\frac{1}{2}\log_2 \frac{\nu_i^2}{D_i}, 0\right\} \tag{49}$$

where ν_i^2 are the eigenvalues of the matrix Σ_U, and where the minimum is over all D_i satisfying $\sum_{i=1}^{M} D_i + E||V||^2 \leq D$.

The proof of this theorem is similar to the arguments given in Section 4.2. It is given in detail in Appendix C.

6 The Distributed KLT Algorithm

Let us now return to the problem of Figure 1. We restrict attention to the case where X is a vector of jointly Gaussian random variables, and present again our solution to the approximation problem. The analysis of the corresponding compression problem is currently under investigation and will be presented in [17], extending results of [18] to the case of Gaussian vector sources, and the results of [19] to the scenario of Figure 1.

6.1 Approximation Problem

In Figure 1, consider Terminal i. Suppose that all other terminals have furnished a k_j-dimensional approximation to their sensed part $X_{(j)}$ of the vector X. In particular, suppose that terminal j applies a transform C_j to the sensed part $X_{(j)}$, and that the first k_j components are the approximation furnished by Terminal j. What is the optimum for Terminal i?

This is precisely the partial-conditional situation discussed in Section 5: X_S is now the vector $X_{(i)}$ sensed by Terminal i. The conditional part Y'_{S^c} is given by the approximations furnished by all other terminals, and the hidden part Y''_{S^c} by what the other terminals do not include into their approximation.

Following Theorem 10, the optimal solution for terminal i can be characterized as follows:

Theorem 12. *For fixed C_j, $\forall j \neq i$, the optimum C_i^* is given by the partial-conditional KLT (Theorem 10), with the side information Y'_{S^c} given by the union of the k_j first components of $C_jX_{(j)}$, $\forall j \neq i$, and the hidden part Y''_{S^c} given by the union of the $M_j - k_j$ last components of $C_jX_{(j)}$, $\forall j \neq i$.*

While we have not found a closed-form expression for the optimal spaces, the theorem suggests a simple algorithm:

Algorithm 1 (distributed KLT). *Input:* 1. *Covariance matrix Σ.* 2. *$j_1, j_2, \ldots, j_L$: the first j_1 components of X are $X_{(1)}$, the next j_2 components of X are $X_{(2)}$, and so on.* 3. *$k_1, k_2, \ldots, k_L$: Terminal i furnishes a k_i-dimensional approximation to $X_{(i)}$.*
Initialize by picking arbitrary unitary matrices $C_1, C_2, \ldots, C_L$.
Then, iterate the following, in turn for each i, $i = 1, 2, \ldots, L$:

> *Compute the best partial-conditional transform C_i^* as described in Theorem 12, and arrange the rows of C_i^* in such a way that the first k rows yield the best k-dimensional approximation.*

This algorithm is illustrated in Figure 8. The figure shows one iteration of the algorithm: The transform matrices C_2 and C_3 are kept fixed while Encoder 1 is picked optimally. By Theorem 12, the optimal choice of Encoder 1 is indeed composed of a transform matrix C_1^*, followed by an appropriate choice of k_1 components in the transform domain.

It is easy to show that the incurred mean-squared error cannot increase from one iteration to the next. This is a direct consequence of Theorem 12; we formulate it in the following lemma:

Lemma 13 (property of the distributed KLT algorithm). *Denote the estimate of X after the nth iteration by $\hat{X}^{(n)}$. This is the minimum mean-square estimate of X based on the first k_i rows of $C_iX_{(i)}$, for $i = 1, 2, \ldots, L$. Then,*

$$E||X - \hat{X}^{(n)}||^2 \quad \geq \quad E||X - \hat{X}^{(n+1)}||^2, \tag{50}$$

i.e., the distortion is a non-increasing function of the iteration number.

Early numerical studies suggest a rapid convergence for well-behaved covariance matrices Σ. The convergence behavior of this algorithm is currently under investigation and will be documented in [17]. An convergence curve for the setup of Example 4 is given in Figure 9.

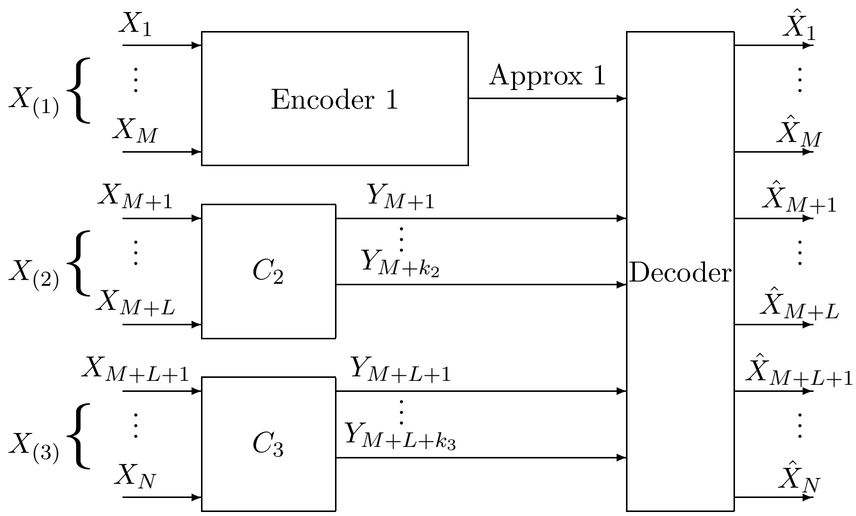

Figure 8: An iteration of the distributed KLT algorithm: The transform matrices C_2 and C_3 are kept fixed while Encoder 1 is picked optimally.

Example 4. A toy example illustrating the basic issue is the following:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0.1 & 0.1 \\ 0 & 0.1 & 0.25 & 0 \\ 0.1 & 0.25 & 1 & 0.25 \\ 0.1 & 0 & 0.25 & 1 \end{pmatrix}. \tag{51}$$

Suppose that the first two components are sampled by the terminal 1, i.e., $X_{(1)} = (X_1, X_2)$, and the last two components by terminal 2, i.e., $X_{(2)} = (X_3, X_4)$. Both terminals are asked to provide a 1-dimensional approximation. For $\sigma_1^2 = 0.11$, if each terminal applies the marginal KLT to its observation. a distortion of $D_{klt} = 0.8207$ is incurred. Note that the KLT's are simple: Terminal 1 applies the identity transform, and Terminal 2 applies

$$C_{2,marginal} = \begin{pmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{pmatrix}. \tag{52}$$

Using the distributed KLT algorithm discussed in this section, and hence making the optimal choice, results in a distortion of $D_{dklt} = 0.3457$, and the transforms are

$$C_1 = \begin{pmatrix} 0.6968 & 2.6205 \\ -0.9820 & 0.1996 \end{pmatrix}. \tag{53}$$

$$C_2 = \begin{pmatrix} 0.2385 & 0.9758 \\ 0.9717 & -0.2366 \end{pmatrix}. \tag{54}$$

The convergence of the distributed KLT algorithm, when C_2 is initially the identity matrix, is shown in Figure 9. The figure shows the error in the middle of the nth iteration, and at the end of the n iteration. Finally, if the entire vector could be handled jointly and the goal is to find the best two-dimensional approximation, a distortion of $D_{jointklt} = 0.1243$ is feasible.

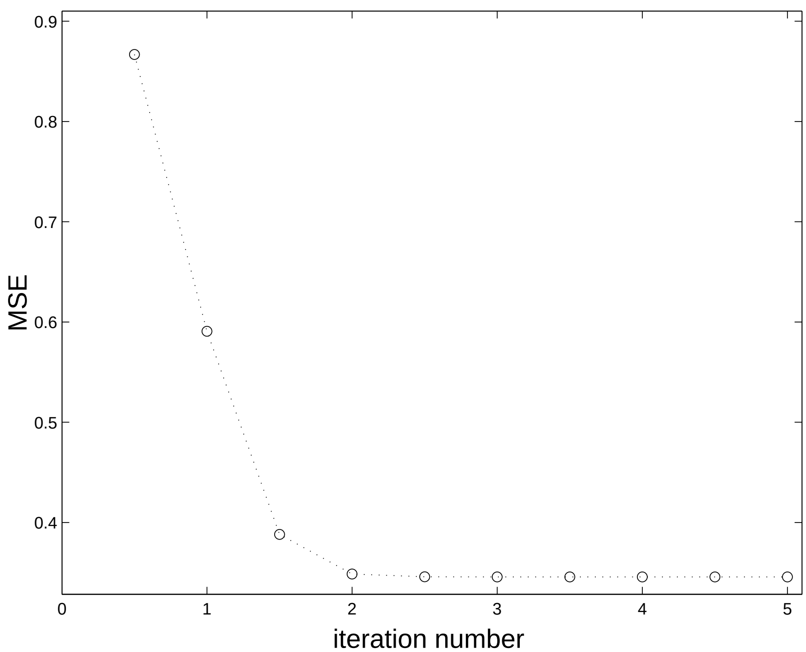

Figure 9: Convergence for Example 4.

Example 5. For the scenario of Example 4, we could also look for the best approximation with $k_1 + k_2 = 2$. It turns out that in the distributed case, it is best to describe $X_{(2)}$ entirely, and to drop $X_{(1)}$, leading to a distortion $D_{mklt} = D_{dklt} = 0.1273$. Note that since the best strategy is simply to retain $X_{(2)}$ completely, it is quite clear that the distributed KLT cannot gain over the marginal KLT here.

Example 6. Suppose Σ is a Toeplitz matrix with first row $(1, \rho, \rho^2, \ldots)$, X_S contains the odd-indexed components of X, and X_{S^c} the even-indexed. For $N = 40$, $M = 20$, $k_1 = k_2 = 10$, and $\rho = 0.7$, the marginal KLT, i.e., the standard KLT applied to each part separately, leads to a distortion $D_{mklt} = 8.3275$, while the distributed KLT gives $D_{dklt} = 6.8464$. Hence, even in this seemingly symmetric scenario, the distributed KLT is *substantially* different from the standard KLT. For comparison, the full standard KLT, applied to the entire vector X, would give $D = 4.5195$.

By analogy to Example 5, we can again determine the minimum distortion under the constraint $k_1 + k_2 \leq 20$. For the marginal KLT, it turns out that $D_{min} = 7.0134$, which is achieved when $k_1 = 20$ or $k_2 = 20$. It can be verified that $D_{dklt} = 6.8464$ is the smallest distortion for the distributed KLT.

7 Conclusions

This paper derives a distributed approximation to the Karhunen-Loève transform. The KLT naturally arises in approximation and compression problems involving vectors of correlated random variables. If the entire vector is available, then the KLT can be applied, furnishing an equivalent vector of uncorrelated random variables. This simplifies the task of approximation and, in the Gaussian case, of optimal compression (in the information-theoretic sense). However, if multiple terminals observe each only a part of the vector X, then it is not generally possible to apply the KLT; rather, the KLT must be approximated in a distributed fashion. Our derivation of such a distributed approximation to the KLT is presented in four steps. First, a *partial* KLT is introduced: Only a part of the vector X of correlated random variables is observed, but the entire vector is to reconstructed. Second, a *conditional* KLT is defined: Only a part of the vector X is observed, and the rest is furnished at reconstruction time. Third, the partial and conditional KLT are combined: Only a part of the vector X is observed. Of the remaining components of X, one part is furnished at reconstruction time (like in the conditional KLT scenario), while the rest remains hidden (like in the partial KLT scenario). Fourth, a *distributed* KLT is presented in the shape of an iterative algorithm. Some evidence of the convergence of this algorithm is presented. A more detailed analysis will be presented in [17]. This distributed transform has potential applications in sensor networks, distributed databases, surveillance systems and hyper-spectral imagery.

A Proof of Theorem 4

Proof of Theorem 4. For the converse part, it has to be established that there does not exist a source code with fewer bits than what Theorem 4 claims. This proof is by analogy to [14, pp. 349–351]. The first step is to establish that the considered rate-distortion function is convex and non-increasing. This follows straightforwardly from [14, Lemma 13.4.1]. Then, the chain of inequalities [14, (13.58)-(13.70)] applies to X_S instead of X:

$$
\begin{aligned}
nR &\geq H(\hat{X}^n) && (55)\\
&\geq I(X_S^n; \hat{X}^n) && (56)\\
&= H(X_S^n) - H(X_S^n|\hat{X}^n) && (57)\\
&= \sum_{i=1}^n H(X_S^{(i)}) - \sum_{i=1}^n H(X_S^{(i)}|\hat{X}^n, X_S^{(i-1)}, \ldots, X_S^{(1)}) && (58)\\
&\geq \sum_{i=1}^n H(X_S^{(i)}) - \sum_{i=1}^n H(X_S^{(i)}|\hat{X}^{(i)}) && (59)\\
&= \sum_{i=1}^n I(X_S^{(i)}; \hat{X}^{(i)}) && (60)\\
&\geq \sum_{i=1}^n R_S(Ed(X^{(i)}, \hat{X}^{(i)})) && (61)\\
&= n\sum_{i=1}^n \frac{1}{n} R_S(Ed(X^{(i)}, \hat{X}^{(i)})) && (62)\\
&\geq nR_S(\frac{1}{n}\sum_{i=1}^n Ed(X^{(i)}, \hat{X}^{(i)}))) && (63)\\
&= nR_S(Ed(X^n, \hat{X}^n)). && (64)
\end{aligned}
$$

The achievability is a direct consequence of the arguments [14, pp.353-356], the only modification being to replace X by X_S. □

B The parameters of Equations (41) and (42)

The matrices A and A_2 in Equation (41) can be calculated from standard results about multivariate Gaussian distributions. It is found to be

$$\begin{pmatrix} A \\ A_2 \end{pmatrix} = Cov(X_{S^c}, X) \begin{pmatrix} I_M & 0 \\ 0 & C_{2a} \end{pmatrix}^H \left(\begin{pmatrix} I_M & 0 \\ 0 & C_{2a} \end{pmatrix} \Sigma \begin{pmatrix} I_M & 0 \\ 0 & C_{2a} \end{pmatrix}^H \right)^{-1} . \quad (65)$$

Similarly, the covariance matrix Σ_U of the random vector U in Equation (42) also follows from standard results about multivariate Gaussian distributions. It can be written as

$$\Sigma_U = \binom{I_M}{A} \Sigma_S (I_M A^H) - \binom{I_M}{A} \Sigma_{SS^c} C_{2a}^H (C_{2a} \Sigma_{S^c} C_{2a}^H)^{-1} C_{2a} \Sigma_{SS^c}^H (I_M A^H). \quad (66)$$

C Proof of Theorem 11

Proof of Theorem 11. From the results of [5, 15], the smallest R (in Figure 7) permitting a distortion of D is

$$R(D) = \min_{p(w|x_S)} I(X_S; W | Y'_{S^c}), \quad (67)$$

where the minimization is over all auxiliary random variables W for which there exists a function $\hat{X}(W, Y'_{S^c})$ such that $E||X - \hat{X}(W, Y'_{S^c})||^2 \leq D$.

It was established above that the distortion term can be rewritten as

$$E\left[||X - \hat{X}||^2 | Y'_{S^c} = y'_{S^c}\right] = E\left[||\binom{I_M}{A} X_S - \binom{I_M}{A} \hat{X}_S||^2 | Y'_{S^c} = y'_{S^c}\right] + E\left[||V||^2\right]. \quad (68)$$

The next step is to rewrite this using the conditional KLT C_P of $\binom{I_M}{A} X_S$ with respect to Y'_{S^c}. Since the conditional KLT is a unitary transform, the overall transform $C_P \binom{I_M}{A}$ has full rank, i.e., $\text{rank}(C_P \binom{I_M}{A}) = M$. Therefore, the mutual information expression can be rewritten as

$$R(D) = \min_{p(w|y_S)} I(Y_S; W | Y'_{S^c}), \quad (69)$$

where the minimization is over all auxiliary random variables W for which there exists a function $\hat{Y_S}(W, Y_{S^c})$ such that

$$\sum_{i=1}^{M} E\left[|Y_i - \hat{Y}_i|^2 \middle| Y'_{S^c} = y'_{S^c}\right] + E||V||^2 \leq D. \quad (70)$$

Due to the properties of the conditional KLT, the components $(Y_1, Y_2, \ldots, Y_M)$ are conditionally independent given Y'_{S^c}. To continue, we can again artificially introduce auxiliary random variables $W_1, W_2, \ldots, W_M$, where W_i is allowed to depend arbitrarily on Y_S. With this, we can write out

$$R = \min_{p(w|y_S)} I(Y_S; W | Y'_{S^c}) \quad (71)$$

$$\geq \min_{p(w_1,\ldots,w_M|y_S)} I(Y_S; W_1, \ldots, W_M | Y'_{S^c}) \quad (72)$$

$$\stackrel{(a)}{\geq} \min_{p(w_1,\ldots,w_M|y_S)} \sum_{i=1}^{M} I(Y_i; W_i | Y'_{S^c}), \quad (73)$$

where (a) holds because $Y_1, \ldots, Y_M$ are conditionally independent given Y'_{S^c}. Equality is attained in (a) when the auxiliary W_i depends only on Y_i, rather than on all of Y_S.

This permits to rewrite (36) as

$$R \geq \sum_{i=1}^{M} \min_{p(w_i|y_S)} I(Y_i; W_i | Y'_{S^c}). \tag{74}$$

The solution to the minimization problem inside the sum has been found by Wyner and Ziv [5]. Let the distortion for the i-th component be D_i. Then,

$$\min_{p(w_i|y_i)} I(Y_i; W_i | X_{S^c}) \quad = \quad \max \left\{ \frac{1}{2} \log_2 \frac{\nu_i^2}{D_i}, 0 \right\}, \tag{75}$$

where $\nu_i^2 = Var(Y_i | Y'_{S^c})$. Therefore, the ν_i^2 are simply the eigenvalues of the matrix Σ_U, and the theorem follows. □

References

[1] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," in *Proc 2002 International Workshop on Multimedia Signal Processing*, (St. Thomas, US Virgin Islands), December 2002.

[2] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed, partial, and conditional Karhunen-Loève transforms," in *Proc 2003 IEEE Data Compression Conference (DCC)*, (Snowbird, UT), March 2003.

[3] J.-Y. Huang and P. Schultheiss, "Block quantization of correlated gaussian random variables," *IEEE Trans. Commun. Systems*, vol. 11, pp. 289–296, September 1963.

[4] V. K. Goyal, "Theoretical foundations of transform coding," *IEEE Signal Processing Magazine*, vol. 18, pp. 9–21, September 2001.

[5] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the receiver," *IEEE Transactions on Information Theory*, vol. IT–22, pp. 1–11, January 1976.

[6] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," in *Data Compression Conference*, (Snowbird, Utah), pp. 158–167, March 1999.

[7] S. Pradhan and K. Ramchandran, "Distributed source coding: Symmetric rates and applications to sensor network," in *Data Compression Conference*, (Snowbird, Utah), pp. 363–372, March 2000.

[8] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, pp. 51–60, March 2002.

[9] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Communication Letters*, vol. 5, pp. 417–419, October 2001.

[10] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Data Compression Conference*, (Snowbird, Utah), pp. 252–261, March 2002.

[11] P. Billingsley, *Probability and Measure.* New York: Wiley, 3 ed., 1995.

[12] R. A. McDonald and P. M. Schulthess, "Information rates of Gaussian signals under criteria constraining the error spectrum," *Proceedings of the IEEE*, vol. 52, pp. 415–416, 1964.

[13] T. Berger, *Rate Distortion Theory: A Mathematical Basis For Data Compression.* Englewood Cliffs, NJ: Prentice-Hall, 1971.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[15] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Information and Control*, vol. 38, pp. 60–80, 1978.

[16] S. S. Pradhan, "On rate-distortion function of Gaussian sources with memory in the presence of side information at the decoder," *Project Report, ECE 480. University of Illinois at Urbana-Champaign*, 1998.

[17] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *to be submitted to IEEE Transactions on Information Theory*, 2003.

[18] T. Berger, "Multiterminal source coding," *Lectures presented at CISM Summer School on the Information Theory Approach to Communications*, July 1977.

[19] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Transactions on Information Theory*, vol. IT–43, pp. 1912–1923, November 1997.