The Distributed Karhunen-Loève Transform

Michael Gastpar, Pier Luigi Dragotti and Martin Vetterli
Audiovisual Communications Laboratory (LCAV)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
Email: Michael.Gastpar@epfl.ch, Pierluigi.Dragotti@epfl.ch, Martin.Vetterli@epfl.ch

***Abstract*— The Karhunen-Loève transform is a key element of many signal processing tasks, including classification and compression. In this paper, we consider distributed signal processing scenarios with limited communication between correlated sources, and we investigate a distributed Karhunen-Loève transform (KLT). In particular, a partial KLT (where only a subset of sources are observed) and a conditional KLT (where some sources act as side information) are posed and solved in a rate-distortion sense. The partial KLT leads to an original bit allocation problem, while the conditional KLT leads to a Wyner Ziv solution which is separable at the sources.**

These two cases can be seen as extreme cases of a distributed KLT.

I. Introduction

MANY of the crucial contemporary applications involve distributed signal processing and communication. Consider for example a scene filmed by multiple cameras. Clearly, the signals are correlated. If they are processed together, standard means such as the KLT can be employed. Suppose however that communication is expensive, and hence, that the signal processing must be done in a distributed fashion separately at every camera. In this paper, we show how the concept of the KLT extends to such a distributed scenario. For a state of the art of the key results on the KLT, we refer to the excellent exposition in [1]. For the importance of distributed source coding and results on distributed compression, we refer to [2].

It is thus natural to pose the following questions:

Given a correlated source vector (e.g. a jointly Gaussian vector) where individual entries are observed in distinct physical locations and communication is a bottleneck, what are appropriate distributed source compression schemes.

Given that in the non-distributed case, the KLT is often the answer, it is natural to see to what extent there exists a distributed equivalent.

A first version of the problem appears if only some of the sources are observed, but all of them need to be reconstructed. We term this the partial KLT problem. In Section II, we give a solution to this problem, showing that it leads to an original rate allocation problem. Note that this solution can also be used to decide how many and which sources need to be observed, given a certain rate.

A second version of the problem appears when certain sources are available at the receiver, a problem we call the conditional KLT, for which the solution is presented in Section III. It is shown that the problem splits into separate Wyner-Ziv problems after taking a KLT that makes the observed entries independent, conditionally on the side information. This is shown to be optimal.

In Section IV, we outline the implication of the above results on a fully distributed KLT (of which the two previous cases are extreme points). This problem is similar to multiterminal lossy source coding, which is an open problem, and we conjecture that using partial and conditional KLT's achieves the best known performance.

II. The Partial KLT

In this section, we study the problem of partial observation or subsampling, depicted in Figure 1: There are N correlated

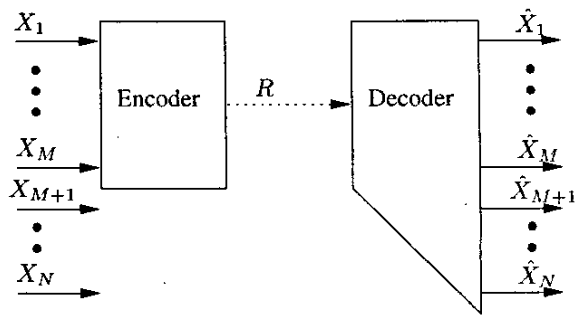

Fig. 1. Compression of a subsampled set of correlated random variables.

sources of which only M are sampled. We consider the vector X of N random variables,

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

with zero mean and covariance matrix Σ.[1] We get to sample only M of these. Without loss of generality, we assume that the *first* M random variables are sampled, and we denote their vector by

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

[1]Throughout the present paper, we assume that Σ has full rank.

0-7803-7713-3/02/$17.00 © 2002 IEEE

with covariance matrix Σ_S, and the vector of non-sampled random variables by

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

The presence of the hidden part X_{S^c} — not observed, but to be reconstructed — alters the problem significantly. 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 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 $E||X - \hat{X}||^2 = \sum_{k=1}^{N} E|X_k - \hat{X}_k|^2$? For the compression problem, our considerations are limited to the case where X is a vector of jointly *Gaussian* random variables. 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 "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 that the solution is still given by the KLT, but that the bit allocation has to be modified.

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

Definition 1: The *partial KLT* of X is the KLT of its sampled part X_S. It is denoted by P.

The transformed version of X_S is denoted by $Y_S = PX_S$, and $\sigma_i^2 = Var(Y_i^2)$.

Properties of the partial KLT.

1) Orthogonal transform
2) The components of Y_S are uncorrelated. If X_S is a vector of jointly Gaussian random variables, then they are independent.

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

1) Approximation Problem: The goal is to minimize the estimation error $E||X - \hat{X}||^2$. The key step is to rewrite this using the partial KLT:

$$\begin{aligned} &E||X - \hat{X}||^2 \\ &\stackrel{(a)}{=} E||X_S - \hat{X}_S||^2 + E||AX_S - A\hat{X}_S||^2 + E||V||^2 \\ &\stackrel{(b)}{=} E||Y_S - \hat{Y}_S||^2 + E||AY_S - A\hat{Y}_S||^2 + E||V||^2 \\ &\stackrel{(c)}{=} \sum_{i=1}^{M}(1 + a_i)E|Y_S - \hat{Y}_S|^2 + E||V||^2, \end{aligned} \tag{5}$$

where (a) follows from standard arguments about the minimum mean-squared error (details see [3]), (b) follows from Property 1) of the partial KLT, and (c) from Property 2), and where

$$a_i = \sum_j |(AP^{-1})_{ji}|^2, \tag{6}$$

i.e. a_i is the sum of the squares of column i of the matrix AP^{-1}.

Theorem 1: The best k-dimensional approximation space for the subsampling problem of Figure 1 is composed of the k eigenvectors corresponding to the k largest *modified* eigenvalues $(1 + a_i)\sigma_i^2$.

2) Compression Problem: 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.

Standard rate-distortion theory determines the minimum rate R (in Figure 1) needed to achieve a distortion D to be

$$R_S(D) = \min I(X_S; \hat{X}_S) \tag{7}$$

where the minimum is over all conditional densities $p(\hat{x}_S|x_S)$ that satisfy $E||X - \hat{X}||^2 \leq D$. Details can be found in [3].

The key step is to transform this problem into the partial KLT domain. Since the partial KLT is an orthogonal transform,

$$R_S(D) = \min I(Y_S; \hat{Y}_S) \tag{8}$$

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

$$\sum_{i=1}^{M}(1 + a_i)E|Y_S - \hat{Y}_S|^2 + E||V||^2 \leq D. \tag{9}$$

For the last equation, we have used (5). Since moreover, the components of Y_S are *independent* random variables, the minimum of $I(Y_S, \hat{Y}_S)$ is achieved by a $\hat{Y}_S$ whose kth component $\hat{Y}_k$ depends only on Y_k. This leads to the following theorem.

Theorem 2: The rate-distortion function for the subsampled Gaussian, illustrated in Figure 1, is given by

$$R_S(D) = \min_{D_i} \sum_{i=1}^{M} \max\left\{\frac{1}{2}\log_2 \frac{\sigma_i^2}{D_i}, 0\right\}, \tag{10}$$

where the minimum is over all D_i satisfying

$$\sum_{i=1}^{M}(1 + a_i)E|Y_S - \hat{Y}_S|^2 \leq D - E||V||^2. \tag{11}$$

For the complete proof, we refer to [3].

This theorem says that an optimal compression system is the one given in Figure 2: Apply the partial KLT, and compress the components separately, using the appropriate bit allocation.

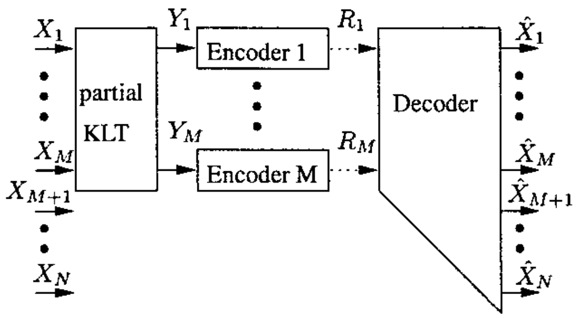

Fig. 2. Compression of a subsampled set of correlated Gaussian sources using the partial KLT. This system is shown to perform optimally.

The appropriate bit allocation is found by solving the optimization problem of Theorem 2. For the standard case ($M = N$), the problem can be solved, e.g., using Lagrange multipliers and the Kuhn-Tucker conditions (see e.g. [4, p. 348]). The Kuhn-Tucker conditions for the more general case $M < N$ are

$$-\frac{1}{2\ln 2}\frac{1}{D_i} + \lambda(1+a_i)\begin{cases} = 0 & \text{if } D_i < \sigma_i^2, \\ \leq 0, & \text{if } D_i \geq \sigma_i^2, \end{cases} \tag{12}$$

which means that the rate allocation is different from the non-subsampled case. Further details will be given in [3].

III. The Conditional KLT

In this section, we study the scenario of Figure 3. This is (in some sense) the complement of Figure 1: Here, the non-encoded random variables are known perfectly to the decoder, whereas there, they were not known at all. Intermediate cases will be outlined in the next section and studied in [3]. By analogy to Section II, two points of view are of interest to us:

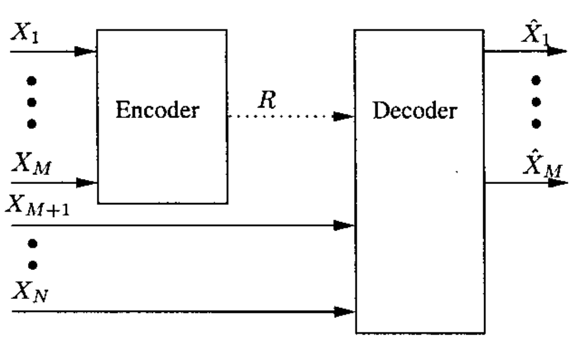

Fig. 3. Compression of a subsampled set of correlated random variables for a decoder that has side information.

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 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 3 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} is the matrix C satisfying

$$C\Sigma_{S|S^c}C^T = I_M, \tag{13}$$

where $\Sigma_{S|S^c}$ is the conditional covariance matrix with entries $\{\Sigma_{S|S^c}\}_{i,j} = Cov(X_i, X_j|X_{S^c})$, I_M denotes the M-dimensional identity matrix, and ^{T} the matrix transpose.

The transformed version of X_S is denoted by $Y_S = CX_S$, and $\lambda_i^2 = Var(Y_i|X_{S^c})$.

Properties of the conditional KLT:

1) Orthogonal 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.

1) Approximation Problem: The goal is to minimize the conditional distortion $E[||X_S - \hat{X}_S||^2|X_{S^c}]$. The key step is to rewrite this in the conditional KLT domain as

$$E\left[||X_S - \hat{X}_S||^2\Big|X_{S^c}\right] \stackrel{(a)}{=} E\left[||Y_S - \hat{Y}_S||^2\Big|X_{S^c}\right] \stackrel{(b)}{=} \sum_{i=1}^{M} E\left[|Y_i - \hat{Y}_i|^2\Big|X_{S^c}\right],$$

where (a) follows from Property 1) of the conditional KLT, (b) from Property 2). The last expression can be rewritten in terms of the conditional variances λ_i^2, yielding:

Theorem 3: The best k-dimensional approximation space for the side information problem of Figure 3 is composed of the k conditional eigenvectors (rows of C) corresponding to the k largest conditional variances λ_i^2.

For a detailed proof, see [3].

2) Compression Problem. The discussion is limited to the case where X is a vector of jointly Gaussian random variables.

The following relationship holds:

$$X_S = BX_{S^c} + U, \qquad (14)$$

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], [6], the smallest R (in Figure 3) permitting a distortion of D is

$$R(D) = \min_{p(w|x_S)} I(X_S; W|X_{S^c}), \qquad (15)$$

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:

$$R(D) = \min_{p(w|y_S)} I(Y_S; W|X_{S^c}), \qquad (16)$$

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

$$\begin{aligned} R &= \min_{p(w|y_S)} I(Y_S; W|X_{S^c}) & (17)\\ &\geq \min_{p(w_1,\ldots,w_M|y_S)} I(Y_S; W_1, \ldots, W_M|X_{S^c}) & (18)\\ &\overset{(a)}{\geq} \min_{p(w_1,\ldots,w_M|y_S)} \sum_{k=1}^{M} I(Y_k; W_k|X_{S^c}) & (19) \end{aligned}$$

where (a) holds because $Y_1, \ldots, Y_M$ are conditionally independent given X_{S^c}. It can be shown that equality is attained in (a) when the auxiliary W_k depends only on Y_k, rather than on all of Y_S. For details, see [3]. This permits to rewrite (19) as $R \geq \sum_{k=1}^{M} \min_{p(w_k|y_S)} I(Y_k; W_k|X_{S^c})$. The solution to the minimization problem inside the sum has been found by Wyner and Ziv [5]. Suppose that the distortion for the k-th component is D_k. Using their result, we can give the following theorem:

Theorem 4: The rate-distortion function for the problem with side information, illustrated in Figure 3, is given by

$$R(D) = \min_{D_i} \sum_{i=1}^{M} \max\left\{\frac{1}{2}\log_2 \frac{\lambda_i^2}{D_i}, 0\right\} \qquad (20)$$

where the minimum is over all D_i satisfying $\sum_{i=1}^{M} D_i \leq D$. For a complete proof, see [3]. This result has also been found in the context of Gaussian sources with memory [7].

The theorem says that the compression problem of Figure 3 can be optimally solved by the system shown in Figure 4: A conditional KLT, followed by *separate* compression of each component (using the techniques described in [5]).

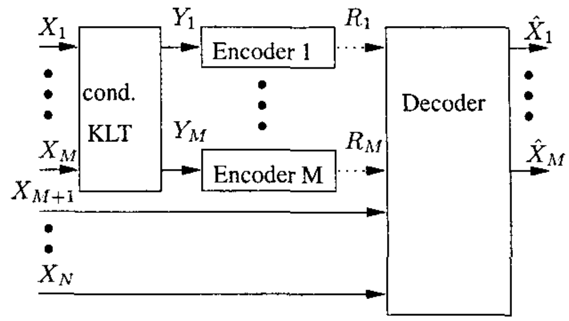

Fig. 4. Compression of a set of correlated Gaussian sources for a decoder that has side information using the conditional KLT. This system is shown to perform optimally.

IV. The Distributed KLT

The settings of Figures 1 and 3 are not truly distributed settings: In both cases, there is only one encoder and only one decoder.

A more interesting scenario is with *two* encoders and one decoder: In Figure 3, add a second encoder, namely for X_{S^c}. The decoder receives both codewords, and has to reconstruct X. What are the best achievable rate pairs for the two encoders? For the case $M = 1$ and $N = 2$, the best known region was given in [8]. Our conjecture is that the extension of this region can be achieved by a system that uses combinations of partial and conditional KLT's on the two sets of random variables. We call these combinations the *distributed KLT*. This will be further studied in [3].

Another interesting setting is with two encoders and *two* decoders. Simple instances of this can be solved immediately using the results of this paper. Take for example Figure 1 and add a second encoder, namely for X_{S^c}. The output of this second is observed *only* by a second decoder. Then, each encoder applies a partial KLT to its observed random variables, followed by the appropriate bit allocation, given by Theorem 2.

References

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

[2] 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.

[3] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *to be submitted*, 2002.

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

[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] 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.

[7] 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.

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