

The Distributed, Partial, And Conditional Karhunen-Loève Transforms

Michael Gastpar^{*} Pier-Luigi Dragotti[°] Martin Vetterli^{†*}

^{*}Department of EECS
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

[†]Institute of Communication Systems
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.

Applications of our results to sensor networks and to distributed databases are discussed.

1 Introduction

Distributed signal processing and communication is an integral part of many of the crucial contemporary applications. Consider for example a scene filmed by multiple cameras, or environmental data recorded by multiple sensors: the signals at the sensors are correlated. In the non-distributed case, one would apply the KLT, thus obtaining uncorrelated data streams which can now be compressed separately from each other. Suppose however that communication between the sensors is expensive, or that they cannot communicate at all. Then, signal processing must be done in a distributed fashion, and the full KLT cannot be applied to the data. 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 in the non-distributed case, we refer to

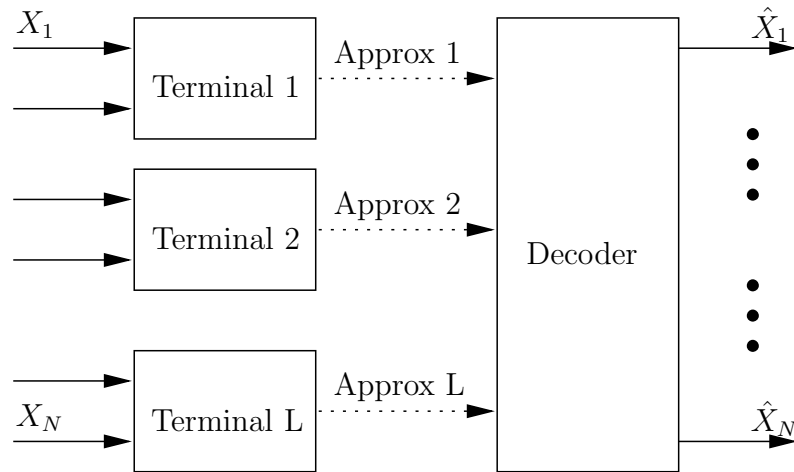


Figure 1: The distributed KLT problem: Distributed compression of multiple correlated vector sources.

the excellent exposition in [1]. The importance of distributed source coding is further elaborated e.g. in [2].

Section 2 formally states the problem leading to the distributed KLT. Section 3 then 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 partial KLT, we show in Section 3.1 that it leads to an original rate allocation problem. For the conditional KLT, it is shown in Section 3.2 that the problem splits into separate Wyner-Ziv problems. Finally, the two solutions are combined in Section 3.3. In Section 4, we show that the results of this paper directly lead to an algorithm to solve the distributed KLT problem. Finally, Section 5 discusses key applications of the distributed KLT introduced in this paper.

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 , with zero mean and covariance matrix Σ .¹ 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

¹Throughout the present paper, we assume for simplicity that Σ has full rank.

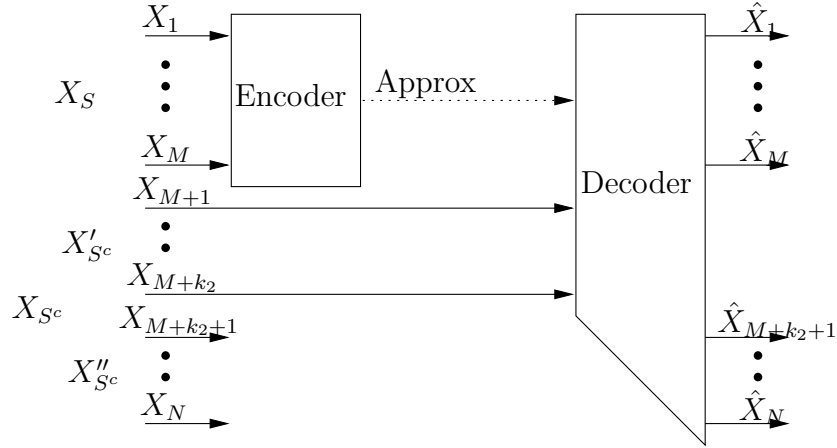


Figure 2: The problem of the partial and the conditional KLT.

sample. For a required maximum average distortion D , what are the rate tuples (R_1, R_2, \dots, 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 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 ? In our considerations, the “best” approximation is the one minimizing the mean-squared error $E\|X - \hat{X}\|^2$.

3 Partial and Conditional KLTs

As a first part of the solution, consider the system of Figure 1 with only two terminals: Terminal 1 senses $X_S = (X_1, X_2, \dots, X_M)$, with covariance matrix Σ_S , and Terminal 2 senses the rest of the vector X , namely $X_{S^c} = (X_{M+1}, X_{M+2}, \dots, X_N)$. The first k_2 components of X_{S^c} are available at the decoder, the remaining components are discarded. This is illustrated in Figure 2. In this section, we present explicit solutions for this problem. First, we discuss the *partial* KLT, i.e., the case $k_2 = 0$. Then, we examine the *conditional* KLT, i.e., the case $k_2 = M$. Finally, we combine our results to obtain the general case shown in Figure 2.

3.1 The Partial KLT

In this section, we study the problem of partial observation or subsampling, as shown in Figure 2 with $k_2 = 0$, i.e., without side information. The presence of the hidden part X_{S^c} — not observed, but to be reconstructed — alters the problem significantly.

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 the variances of the components of Y by $\sigma_i^2 = \text{Var}(Y_i^2)$.

Properties of the partial KLT.

1. P is an orthonormal transform
2. The components of Y_S are uncorrelated.

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

$$X_{S^c} = AX_S + V, \quad (1)$$

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

Approximation Problem

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, as we show next.

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

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

The minimum mean-square error estimate satisfies $\hat{X}_{S^c} = A\hat{X}_S$. The key step is to apply the (partial) KLT to X_S . Denote $Y_S = PX_S$ and note that the best estimate of Y_S is simply $\hat{Y}_S = P\hat{X}_S$. Hence,

$$\begin{aligned} E\|X - \hat{X}\|^2 &= E\|Y_S - \hat{Y}_S\|^2 + E\|AP^{-1}Y_S - AP^{-1}\hat{Y}_S\|^2 + E\|V\|^2 \\ &= \sum_{i=1}^M (1 + a_i) E|Y_{S,i} - \hat{Y}_{S,i}|^2 + E\|V\|^2, \end{aligned} \quad (3)$$

where $a_i = \sum_{j=1}^M |(AP^{-1})_{ji}|^2$, 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 2 (with $k_2 = 0$) is composed of the k eigenvectors of Σ_S corresponding to the k largest modified eigenvalues $(1 + a_i)\sigma_i^2$.*

Example 1. A toy example illustrating the basic issue is the following:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & 0.1 & 0.25 \\ 0 & 0.25 & 1 \end{pmatrix}. \quad (4)$$

Suppose that the first two components are sampled by the terminal, i.e., $M = 2$. The terminal is asked to provide a 1-dimensional approximation. If $\sigma_1^2 \leq 0.1$, both the partial KLT and the usual KLT suggest the same choice of a 1-dimensional space, hence they achieve the same distortion. However, if $\sigma_1^2 > 0.1$, choosing the approximation space according to the usual KLT is sometimes suboptimal. For example, for $\sigma_1^2 = 0.11$, the usual KLT results in a distortion of $D_{klt} = 1.1$, while the partial KLT achieves a distortion of $D_{pkt} = 0.485$.

Compression Problem

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 $E\|X - \hat{X}\|^2 = \sum_{k=1}^N E|X_k - \hat{X}_k|^2$? We consider only 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), and compress each component separately, the rate allocation being given by the eigenvalues of the covariance matrix. 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} . The following theorem is proved in [3, 4]:

Theorem 2 (rate-distortion function of the subsampled Gaussian). *The rate-distortion function for the subsampled Gaussian, illustrated in Figure 2 (with $k_2 = 0$), is given by*

$$R_S(D) = \min_{D_i} \sum_{i=1}^M \max \left\{ \frac{1}{2} \log_2 \frac{\sigma_i^2}{D_i}, 0 \right\}, \quad (5)$$

where the minimum is over all D_i satisfying $\sum_{i=1}^M (1 + a_i)D_i \leq D - E\|V\|^2$.

Remark 1. The theorem says that in spite of the hidden part, the first stage of the source coding scheme remains the same: simply take the KLT of the sampled part. The rate allocation, however, depends on the hidden part: through the coefficients a_i .

Remark 2 (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 2: Compute $R_S(D)$ for all sets S with cardinality M .

3.2 The Conditional KLT

In this section, we study the scenario of Figure 2 with $k_2 = N - M$, i.e., without a hidden part. 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

$$\{Cov(X_i, X_j | X_{S^c} = x_{S^c})\}_{i,j} = \Sigma_{S|S^c}, \quad (6)$$

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, the conditional KLT C of X_S with respect to X_{S^c} is the eigendecomposition of $\Sigma_{S|S^c}$, i.e., $C\Sigma_{S|S^c}C^T = \text{diag}(\lambda_1, \dots, \lambda_M)$.

Remark 3. 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 $Y_S = CX_S$, and $\lambda_i^2 = Var(Y_i | X_{S^c} = x_{S^c})$, which by assumption does not depend on x_{S^c} .

Properties of the conditional KLT:

1. C is an orthonormal transform
2. The components of the vector Y_S are conditionally uncorrelated given X_{S^c} .

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

$$X_S = BX_{S^c} + U, \quad (7)$$

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, $\Sigma_{S|S^c} = \Sigma_U$.

Approximation Problem

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.

The goal is to minimize the conditional distortion $E \left[\|X_S - \hat{X}_S\|^2 \middle| X_{S^c} = x_{S^c} \right]$. The key step is to rewrite this in the conditional KLT domain, which eventually permits the following shape:

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

where T denotes the set of the k dimensions that are used for the approximation, and T^c its complement (within S). The best choice of the set T does not depend on the actual value of x_{S^c} :

Theorem 3. *The best k -dimensional approximation space for the side information problem of Figure 2 (with $k_2 = N - M$) is composed of the k conditional eigenvectors (rows of C) corresponding to the k largest conditional variances λ_i^2 .*

Example 2. Take Σ the $N \times N$ Toeplitz matrix of $(1, \rho, \rho^2, \dots)$. Consider $N = 100$, $\rho = 0.8$, $S = (1, 3, 5, 7, \dots, 99)$, and $k = 25$. If we simply apply the (“marginal”) KLT to the sampled part and retain the eigenvectors corresponding to the k largest eigenvalues, the final distortion is $D_{klt} = 5.5141$. If we apply the conditional KLT with respect to the non-sampled part, the distortion is $D_{cklt} = 2.6195$, i.e., considerably smaller.

Compression Problem

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 2 (with $k_2 = N - M$) has been solved by Wyner and Ziv [5]. Here, we restrict our attention to the case where X is a vector of jointly Gaussian random variables, 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; the bit allocation between these M Wyner-Ziv problems is determined by the following theorem:

Theorem 4. *The rate-distortion function for the problem with side information, illustrated in Figure 2 (with $k_2 = N - M$), is given by*

$$R(D) = \min_{D_i} \sum_{i=1}^M \max \left\{ \frac{1}{2} \log_2 \frac{\lambda_i^2}{D_i}, 0 \right\} \quad (8)$$

where the minimum is over all D_i satisfying $\sum_{i=1}^M D_i \leq D$.

For a complete proof, see [4]. This result has also been found in the context of Gaussian sources with memory [6].

3.3 Combining the Partial and the Conditional KLT

The solutions found in Section 3.1 and 3.2 can be used to determine the solution for the problem illustrated in Figure 2: there is side information *and* a hidden part. Our analysis is limited to the case where X is a vector of jointly Gaussian random variables. Again, we can consider the problems of *approximation* and *compression*; here, we present our solution to the approximation problem. The intuitively pleasing solution is as follows: The conditional KLT with respect to X'_{S^c} is applied (as if X''_{S^c} was not there). The best choice of k components for the approximation problem, or the best rate allocation for the compression problem, depends both on X'_{S^c} and X''_{S^c} . It is determined by the modified conditional eigenvalues $(1 + c_i)\lambda_i$ defined below.

Approximation Problem

Consider the distortion $E \left[\|X - \hat{X}\|^2 | X'_{S^c} = x'_{S^c} \right]$. Due to the joint Gaussianity of X , we can write

$$X_S = BX'_{S^c} + U, \quad (9)$$

where X_S and U are independent Gaussian random vectors. Hence, the conditional KLT of X_S with respect to X'_{S^c} exists, and we obtain

$$\begin{aligned} E \left[\|X - \hat{X}\|^2 | X'_{S^c} = x'_{S^c} \right] \\ = \sum_{k=1}^M E \left[|Y_k - \hat{Y}_k|^2 | X'_{S^c} = x'_{S^c} \right] + E \left[\|X''_{S^c} - \hat{X}''_{S^c}\|^2 | X'_{S^c} = x'_{S^c} \right] \end{aligned} \quad (10)$$

Again due to the joint Gaussianity of X , we can write

$$X''_{S^c} = A \begin{pmatrix} X_S \\ X'_{S^c} \end{pmatrix} + V = A_1 X_S + A_2 X'_{S^c} + V, \quad (11)$$

where A_1 and A_2 are defined accordingly, and V is independent of X_S and X'_{S^c} . From the arguments leading to the partial KLT, we find that the distortion can be written as

$$E \left[\|X - \hat{X}\|^2 | X'_{S^c} = x'_{S^c} \right] = \sum_{k=1}^M (1 + c_k) E \left[|Y_k - \hat{Y}_k|^2 | X'_{S^c} = x'_{S^c} \right] + E \|V\|^2,$$

where $c_i = \sum_{j=1}^{N-M-k_2} |(A_1 C^{-1})_{ji}|^2$. The best k -dimensional approximation to X_S is therefore easily determined:

Theorem 5 (partial-conditional KLT). *The best k -dimensional subspace is a subspace of the conditional KLT of X_S with respect to X'_{S^c} . Out of this M -dimensional conditional KLT space, the best k vectors are the ones with largest modified conditional eigenvalues $(1 + c_i)\lambda_i^2$, where the modification c_i depends on the hidden part X''_{S^c} .*

4 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.

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 ?

Clearly, this is the partial-conditional situation outlined in Figure 2: X_S is now the vector $X_{(i)}$ sensed by Terminal i . The conditional part is given by the approximations furnished by all other terminals, and the hidden part by what the other terminals do not include into their approximation.

The optimal solution for terminal i can be characterized as follows:

Theorem 6. *For fixed $C_j, \forall j \neq i$, the optimum C_i^* is given by the partial-conditional KLT (Theorem 5), conditioned on the k_j first components of $C_j X_{(j)}, \forall j \neq i$.*

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, \dots, 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, \dots, k_L : Terminal i furnishes a k_i -dimensional approximation to $X_{(i)}$. Initialize by picking arbitrary unitary matrices C_1, C_2, \dots, C_L . Then, iterate the following, in turn for each $i, i = 1, 2, \dots, L$:*

Compute the conditional KLT C_i of $X_{(i)}$ with respect to the union of the k_j first components of $C_j X_{(j)}, \forall j \neq i$, and order the rows of C_i such that the k_i first are the best k_i -dimensional approximation in the sense of Theorem 5.

Example 3. Suppose Σ is a Toeplitz matrix with first row $(1, \rho, \rho^2, \dots)$, X_S contains the odd-indexed components of X , and X_{S^c} the even-indexed. For $N = 40$ and $\rho = 0.7$, the standard KLT applied to each part separately leads to a distortion $D_{klt} = 8.3275$, 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 $D = 4.5195$.

Early numerical studies suggest a rapid convergence for well-behaved covariance matrices Σ . The convergence behavior of this algorithm is currently under investigation [4].

5 Applications

5.1 Distributed Compression in Sensor Networks

Figure 1 is a certain sensor network situation: Suppose there are L stations, each of which has a certain number of sensors. All L stations report to a central unit, whose goal is to find an estimate \hat{X} . The quality of the estimate is assessed by $E\|X - \hat{X}\|^2$. What is the best operation of the L stations? Under the hypotheses of this paper, the optimal operation is for the stations to apply first the distributed KLT to their observed vector X_S . In the approximation framework, each station then selects the appropriate components according to the rules of the partial-conditional KLT (Theorem 5). The compression framework is currently being studied [4].

5.2 Distributed Database Retrieval

Distributed databases have found widespread interest, in particular for music and image data. Figure 1 can be seen as a simple model of a distributed database; in that case, each terminal is a server. X_i are the documents added to the database. For storage limitations, each database only stores an approximation to the documents. We assume that the documents of the various databases are generally correlated.

A user specifies the number i if she wants to retrieve document X_i . The databases furnish an approximation \hat{X}_i . Suppose that the quality of the database retrieval system is measured by the overall average distortion,

$$D = \sum_{i=1}^N E|X_i - \hat{X}_i|^2. \quad (12)$$

One of the fundamental question is the following: what is the optimum way for database l to compress its documents? Here, the “optimum” way is the one minimizing D .

Our theory applies directly to this setup: in the approximation framework, database l applies the distributed KLT to the data, and retains the appropriate k_l -dimensional approximation; the compression framework is currently under investigation.

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