PREDICTIVE VECTOR QUANTIZER DESIGN FOR DISTRIBUTED SOURCE CODING

Pradeepa Yahampath

Department of Electrical and Computer Engineering
University of Manitoba, Winnipeg, MB R3T 5V6, Canada
pradeepa@ee.umanitoba.ca

ABSTRACT

This paper investigates the design of a system of predictive vector quantizers for distributed sources with memory, in which linear prediction is used to exploit the source memory, while distributed quantization is used to exploit the correlation between sources. A training-based algorithm is proposed for jointly designing the predictors, binning functions, and reconstruction codebooks of the given system to match the intra- and inter-source correlations. In order to demonstrate the effectiveness of the algorithm, experimental results obtained by designing both scalar and vector quantizers for a set of distributed Gauss-Markov sources are presented. While the optimality of these designs is unknown, it is shown that they convincingly outperform several other alternatives.

Index Terms- Distributed source coding, vector quantization, predictive coding

1. INTRODUCTION

Recently, there has been a tremendous interest in distributed source coding, due to its potential applications in emerging sensor networks. A typical problem of interest is the separate encoding and joint decoding of observations from a spatially distributed set of correlated sources (sensors). It is known that the loss of coding efficiency due to lack of communication between sensors can be compensated for by a joint decoder which exploits the inter-source redundancy [1].

The particular problem considered in this paper was first studied in [2] for the case of two memoryless sources, where the sources were modeled as noisy observations of some random variable to be estimated at a central decoder. An information theoretic achievable rate region of an optimal system was obtained and the problem of designing a good system was also considered. In particular, each distributed encoder was implemented as a two stage encoder, in which the first stage is an ordinary quantizer optimized for its input (i.e. a "pre-quantizer") and the second stage is a quantizer (operating on the discrete output of the first stage) which exploits the inter-source correlation to reduce the transmission bit-rate. The latter operation is generally referred to as *binning* [3]. In subsequent research (e.g. [3]), efficient binning strategies based on channel codes and their application to distributed quantizer design have been considered.

This work was supported by a grant from NSERC of Canada.

The above mentioned work focuses on exploiting inter-source correlation, but not the memory in individual sources. In practice, *predictive coding* (e.g. DPCM) is widely used to quantize sources with memory (e.g. speech). Therefore, designing predictive quantizers for distributed coding is of great interest. The predictive quantizer design in a distributed setting poses an interesting optimization problem in which the predictor coefficients must be chosen differently, compared to an ordinary predictive coder for the given source. A method for designing distributed scalar predictive quantizers is presented in [4]. However, that method requires the use of a restricted class of distributed uniform quantizers, as well as an exhaustive search to find the best predictor coefficients.

In contrast, this paper presents an algorithm for designing predictive vector quantizers (PVQ) for distributed sources, by generalizing the PVQ design algorithms of the form [5], [6]. The proposed algorithm iteratively improves an appropriately chosen initial system based on a training set of source vectors. Different to [4], the proposed method can be used to design vector quantizers and the optimal predictor coefficients are obtained as a part of the design algorithm, as opposed to using an exhaustive search over all admissible filters. In order to simplify our preliminary study, we identify binning as a problem of index assignment [7], and conveniently use combinatorial optimization to obtain good binning functions. This approach can be useful in optimizing distributed quantizers in general as well. The possibility exists for incorporating channel coding-based binning methods [3] within the proposed algorithm as well.

2. SYSTEM DESCRIPTION

The system considered in this paper is shown in Fig.1. It encodes two statistically dependent sources represented by discrete-time stationary ergodic random processes $\{\mathbf{X}_n\}$ and $\{\mathbf{Y}_n\}$ where the d-dimensional random vectors $\mathbf{X}_n \in \mathbb{R}^d$ and $\mathbf{Y}_n \in \mathbb{R}^d$ have a joint probability density function (pdf) $f_{\mathbf{XY}}(\mathbf{x}_n, \mathbf{y}_n)$ (we use upper case to denote random variables and lower case to denote realizations). Each source is quantized by its own predictive encoder, where $\mathbf{U}_n$ and $\mathbf{V}_n$ are the respective predictions errors. It is assumed that the encoders cannot communicate with each other. However, on the receiver side, the prediction errors $\hat{\mathbf{U}}'_n$ and $\hat{\mathbf{V}}'_n$ are jointly decoded.

In general, there will be a statistical dependence between $\mathbf{U}_n$ and $\mathbf{V}_n$, characterized by the joint pdf $f_{\mathbf{UV}}(\mathbf{u}_n, \mathbf{v}_n)$. Let us define more precisely the encoder and decoder for $\mathbf{X}_n$ (definition of the system for $\mathbf{Y}_n$ follows from symmetry). The prediction error $\mathbf{U}_n = \mathbf{X}_n - \tilde{\mathbf{X}}_n$, where $\tilde{\mathbf{X}}_n$ is the prediction generated by a vector linear predictor β_x, is quantized by an N_u level vector quantizer ϵ_u whose output is an integer $I_n \in \{1, \ldots, N_u\}$. As usual in predictive coding, the prediction is based on a locally reconstructed signal $\hat{\mathbf{X}}_n = \tilde{\mathbf{X}}_n + \hat{\mathbf{U}}_n$, where $\hat{\mathbf{U}}_n$ is produced by the local decoder δ_u based on I_n. The transmission of I_n requires a rate of $\log_2 N_u$ bits/vector. In order to achieve a rate reduction by exploiting the correlation between $\mathbf{U}_n$ and $\mathbf{V}_n$, I_n is mapped to $I'_n = \phi_u(I_n)$ prior to transmission, where $I'_n \in \{1, \ldots, M_u\}$ and $M_u \leq N_u$. The "binning" function ϕ_u is chosen based on $f_{\mathbf{UV}}(\mathbf{u}_n, \mathbf{v}_n)$. As a result of binning, the required transmission rate reduces to $\log_2 M_u$ bits/vector.

We measure the performance of the system by it's mean square error (MSE)

$$D = E\|\mathbf{X}_n - \hat{\mathbf{X}}'_n\|^2 + E\|\mathbf{Y}_n - \hat{\mathbf{Y}}'_n\|^2, \tag{1}$$

and define the optimal system as a one that achieves the minimum MSE (MMSE) for the given joint pdf $f_{\mathbf{XY}}(\mathbf{x}_n, \mathbf{y}_n)$. Our goal is to design a system optimal for the empirical distribution of a realization (i.e. a training set) drawn from this joint density. In this case, we replace (1) by the corresponding sample average based on the training set.

3. DESIGN ALGORITHM

In order to motivate a design strategy, it is worth emphasizing that there is a compromise between distributed coding and predictive coding effected by the choice of the linear predictor. On the one hand, for distributed coding to be effective, there must be sufficient correlation between prediction errors $\mathbf{U}_n$ and $\mathbf{V}_n$. On the other hand, for predictive coding of individual sequences $\{\mathbf{X}_n\}$ and $\{\mathbf{Y}_n\}$ to be effective, the temporal correlation in the prediction error sequences $\{\mathbf{U}_n\}$ and $\{\mathbf{V}_n\}$ must be low (in single source PVQ, one strives to render prediction error a white process). The point is that the predictors affect both intra-sequence and inter-sequence correlations. Therefore, for a given $f_{\mathbf{XY}}(\mathbf{x}_n, \mathbf{y}_n)$, the optimal predictors are different to those in a single source PVQ (i.e. prediction error whitening is no longer the optimal strategy).

The most successful approach to practical PVQ design is the iterative system improvement [8], [5], [6]. The basic philosophy behind this approach is as follows. Given a source training set, one finds the input sequence to each component in some initial system. Then, based on an appropriate distortion measure, each component is optimized for its input training set with the other components fixed, ignoring the effect of feedback. The system is then replaced by a new system with updated components, the input sequence to each component is recomputed in closed-loop, and the above component updates are repeated iteratively, until the average distortion of the system is sufficiently low. Our goal is to develop a similar approach to design a distributed PVQ. As such, we divide the system into there parts and attempt to optimize each for a fixed input as follows.

1) Distributed vector quantizers for prediction errors and the joint decoder: Assume fixed inputs $\{\mathbf{U}_n\}$, $\{\mathbf{V}_n\}$. We first design the vector quantizers ϵ_u and ϵ_v for their respective inputs using the generalized Lloyd algorithm [8]. Then, keeping these pre-quantizers fixed, we optimize the binning functions ϕ_u and ϕ_v and the joint decoder, as described in Sec. 4.

2) Local decoders δ_u and δ_v: Assume fixed inputs $\{I_n\}$, $\{J_n\}$. If the joint decoding is very effective, then (I_n, J_n) can be correctly determined from (I'_n, J'_n) with probability almost one. In that case, the quantized prediction errors reconstructed by the local decoders and those reconstructed by the joint decoder will be nearly identical. However, if the correlation between $\{\mathbf{U}_n\}$ and $\{\mathbf{V}_n\}$ is not high enough, there will be a mismatch between the prediction errors in the encoders and the joint decoder. In order to effect a graceful degradation of this mismatch error with decreasing correlation between the two sources, we choose each local decoder so as to minimize the mean square mismatch errors $E\|\hat{\mathbf{U}}_n - \hat{\mathbf{U}}'_n\|^2$ and $E\|\hat{\mathbf{V}}_n - \hat{\mathbf{V}}'_n\|^2$ respectively. For given $\{I_n\}$ and $\{J_n\}$, local decoders which achieve this condition are given by

$$\begin{aligned} \delta^*_u(i_n) &= E\{\hat{\mathbf{U}}'_n | i_n\}, \\ \delta^*_v(j_n) &= E\{\hat{\mathbf{V}}'_n | j_n\}, \end{aligned} \tag{2}$$

Given a joint decoder, these codebooks can be estimated.

3) Vector prediction filters β_x and β_y: Assuming L-th order linear prediction, the prediction for $\mathbf{X}_n$ at the receiver can be written as

$$\tilde{\mathbf{X}}'_n = \sum_{k=1}^{L} \mathbf{A}_k \hat{\mathbf{X}}'_{n-k}, \tag{3}$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_L$ are the $d \times d$ coefficient matrices. Let the corresponding coefficient matrices for $\tilde{\mathbf{Y}}'_n$ be $\mathbf{B}_1, \ldots, \mathbf{B}_L$. Also let $\tilde{\mathbf{X}}''_n = \mathbf{X}_n - \hat{\mathbf{U}}'_n$ and $\tilde{\mathbf{Y}}''_n = \mathbf{Y}_n - \hat{\mathbf{V}}'_n$, which are the ideal predictions desired at the receiver. Then, the distortion in (1) can be expressed as

$$E\|\tilde{\mathbf{X}}''_n - \sum_{k=1}^{L} \mathbf{A}_k \hat{\mathbf{X}}'_{n-k}\|^2 + E\|\tilde{\mathbf{Y}}''_n - \sum_{k=1}^{L} \mathbf{B}_k \hat{\mathbf{Y}}'_{n-k}\|^2 \tag{4}$$

Given a training set of input-output pairs $(\hat{\mathbf{X}}'_n, \tilde{\mathbf{X}}''_n)$ of the filter β_x, the problem at hand is to find the filter coefficients

$\mathbf{A}_1, \ldots, \mathbf{A}_L$ which minimizes the first expectation in (4). Also a similar problem can be formulated for β_y with respect to the second expectation in (4). The solution to these two linear estimation problems can be obtained by solving a set of vector-form Yule-Walker equations. We omit the details for brevity, but refer to Sec. 13.3 of [8].

The design algorithm repeatedly applies the above three steps, starting from some initial system. After each iteration the input sequence to each component is recomputed based on the new design. The iterations are stopped when the average distortion of the resulting the design does not change significantly between successive iterations.

4. OPTIMIZATION OF BINNING FUNCTIONS

In this section, we consider the problem of choosing the binning functions ϕ_u and ϕ_v (integer-to-integer quantizers) so that $E\|\mathbf{U}-\hat{\mathbf{U}}'\|^2 + E\|\mathbf{V}-\hat{\mathbf{V}}'\|^2$, is minimized (without loss of generality time subscript is omitted). It is straight forward to show that (details omitted)

$$E\|\mathbf{U}-\hat{\mathbf{U}}'\|^2 = D_{I,J} + \sum_{i=1}^{N_u}\sum_{j=1}^{N_v} \|\mathbf{g}_{i,j}^{(u)} - \hat{\mathbf{u}}'_{i',j'}\|^2 P_{i,j} \quad (5)$$

where $\hat{\mathbf{u}}'_{j',j'}$ is the codebook used by the joint decoder (at receiver) with $i' = \phi_u(i)$ and $j' = \phi_v(j)$, $D_{I,J}$ is the mean square quantization error if the joint decoding of $\mathbf{U}$ were based on i, j (as opposed to i', j'), using the MMSE optimal codebook $\mathbf{g}_{i,j}^{(u)} = E\{\mathbf{U}|i,j\}$, and $P_{i,j}$ is the joint probability of the index pair (i, j). It is easy to see that the MMSE optimal joint decoder codebook based on i', j' is simply $\hat{\mathbf{u}}'_{i',j'} = E\{\mathbf{U}|i',j'\}$. Now, notice that the second term in (5) represents the additional MSE incurred (in coding $\mathbf{U}$) by the rate reduction due to binning. Considering both $\mathbf{U}$ and $\mathbf{V}$, the total additional MSE due to binning is

$$\sum_{i=1}^{N_u}\sum_{j=1}^{N_v} \left[\|\mathbf{g}_{i,j}^{(u)} - \hat{\mathbf{u}}'_{i',j'}\|^2 + \|\mathbf{g}_{i,j}^{(v)} - \hat{\mathbf{v}}'_{i',j'}\|^2\right] P_{i,j}. \quad (6)$$

Thus, we have a combinatorial optimization problem of minimizing (6), with respect to the two integers-to-integer mappings ϕ_u and ϕ_v. To solve this problem we can adapt the simulated annealing-based index assignment optimization algorithm given in [7]. In addition to using the cost function (6), the main difference here is the way in which the solution vector is represented and the random perturbations are effected. We describe this.

Binning functions ϕ_u and ϕ_v can be viewed as some assignment of N_u integers to a set of M_u bins and N_v integers to M_v bins respectively. The perturbation of a solution in simulated annealing can be effected by moving an element from a randomly chosen bin to another randomly chosen bin. Note that the solution space can be significantly reduced by considering that, in a valid solution, every bin must contain at least one element. With this formulation, the index assignment algorithm in [7] can be easily adapted to optimizing binning functions in distributed quantization.

5. SIMULATION RESULTS AND DISCUSSION

In order to demonstrate the effectiveness of the algorithm proposed above, we use this algorithm to design distributed quantizers for two jointly Gaussian random variables X_n and Y_n with correlation coefficient $\gamma = E\{X_nY_n\}/\sigma_x\sigma_y$, where $EX_n = EY_n = 0$ and $\sigma_x^2 = EX_n^2$, $\sigma_y^2 = EY_n^2$. Furthermore, the sequences $\{X_n\}$ and $\{Y_n\}$ are two first-order Gauss-Markov processes defined by

$$\begin{aligned} X_n &= \rho_1 X_{n-1} + R_n \\ Y_n &= \rho_2 Y_{n-1} + S_n \end{aligned} \quad (7)$$

where R_n and S_n are independent zero-mean Gaussian iid sequences, and $|\rho_1|, |\rho_2| < 1$. In our experiments, we set $\sigma_x^2 = \sigma_y^2 = 1.0$ (i.e. average signal power is 1.0). In the following, the performance of a quantizer system is measured by the average distortion (1), expressed as a signal-to-noise ratio (SNR).

First, we consider the design of predictive scalar quantizers (also known as DPCM) with first order predictors. Each source is quantized at the rate of 3 bits/sample [the rate of each pre-quantizer (ϵ_u and ϵ_v) was chosen to be 6 bits/sample]. In order to initialize the design algorithm, we chose purely source-optimized quantizers designed as in [5]. In our first set of experiments, we considered two sources with inter-source correlation $\gamma = 0.98$. The proposed algorithm was used to design distributed predictive quantizers for different values of intra-source correlation, ρ_1 and ρ_2 (for simplicity, let $\rho_1 = \rho_2$). The performance of these designs are compared with several other alternatives in Fig. 2. Here, *distributed non-predictive* refers to memoryless distributed quantizers designed by simulated annealing as described in Sec. 4, *Non-distributed predictive* refers to ordinary predictive quantizers designed for each source as in [5], and *Non-distributed non-predictive* refers to memoryless, Lloyd-max quantizers designed for each source [8]. The values shown along the curves in parenthesis are the predictor coefficients found by the design algorithm (due to symmetry, the predictor coefficients for two sources are nearly identical). These results reveal that, the proposed distributed predictive quantizers provide the "best of both worlds". On the one hand, the performance of the distribute predictive quantizer approaches that of the independent predictive quantizer when the intra-source correlation is high. At this extreme the gain due to predictive coding dominates the overall performance. On the other hand, the performance of the distribute predictive quantizer approaches that of the memoryless distributed quantizer when the intra-source correlation becomes low (i.e. the gain due to predictive coding becomes negligible). Between these extremes, the distributed predictive quantizers yield a substantial performance

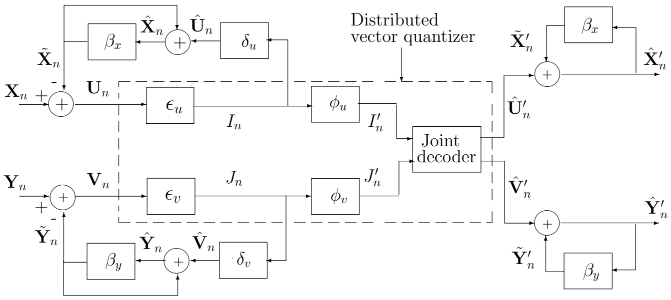

Fig. 1. Proposed distributed PVQ system.

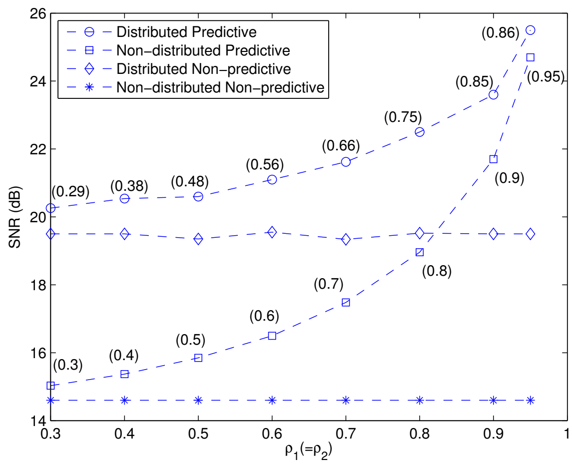

Fig. 2. Performance comparison of distributed predictive scalar quantizers with other alternatives at the rate of 3 bits/sample per source, 1st-order prediction, and $\gamma = 0.98$. The number in parenthesis is the predictor coefficient.

gain over predictive-only and distributed-only quantizers, by choosing the predictor coefficients, the binning functions, and the joint codebooks to match the intra- and inter-source correlations. We also observed similar results in a second set of experiments in which the quantizers were designed for different γ and fixed ρ_1, ρ_2.

We have also investigated the design of predictive vector quantizers. In this case, we considered 2-dimensional vector quantization of the sames sources in (7) at the rate of 3 bits/vector (1.5 bits/sample), using 1st-order vector predictors. Again, we performed a number of experiments which confirmed the observations made with scalar quantization above. To save space, we summarize in Table 1 a comparison of SNR performance for the case $\rho_1 = \rho_2 = 0.9$ and $\gamma = 0.98$. This example clearly shows the difference between predictor matrices of the distributed and non-distributed quantizers.

While it appears difficult to establish the optimality of the designs obtained by the given algorithm (a problem common to any predictive quantizer design in general, see [8] pp. 493),

System	SNR (dB)	Predictor Matrix
Distributed Predictive	16.35	0.47 0.13 0.60 0.12
Non-distributed Predictive	14.02	0.81 0.00 0.90 0.00
Distributed Non-predictive	14.43	
Non-distributed Non-predictive	10.78	

Table 1. Comparison of 2-dimensional distributed PVQ with other alternatives for the case $\gamma = 0.98$ and $\rho_1 = \rho_2 = 0.9$. The rate is 3 bits/vector. The ordering of predictor matrix $\mathbf{A}$ is $[\tilde{x}_{n+1}\ \tilde{x}_n]^T = \mathbf{A}[\hat{x}_{n-1}\ \hat{x}_{n-2}]^T$.

numerous experiments (also involving other source models) indicated that this algorithm is a very effective way of designing predictive VQ systems that exploit both intra- and inter-source correlation of distributed sources. It is straightforward to extend the algorithm to a set-up with many sources, such as a sensor network. An interesting avenue of further work is to incorporate well known channel coding based binning methods [3] into the given optimization algorithm.

6. REFERENCES

[1] A. D. Wyner and H. Ziv, "The rate-distortion function for source coding with side information," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[2] T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Trans. Inform. Theory*, vol. 33, no. 6, pp. 773–786, Nov. 1987.

[3] S. S. Pradhan and K. Ramachandran, "Distributed source coding: Symmetric rates and applications to sensor network," in *Data Comp. Conf.*, March 2000, pp. 363–372.

[4] E. Tuncel, "Predictive coding of correlated sources," in *IEEE Inform. Theory Workshop*, Oct 2004, pp. 111–116.

[5] V. Cuperman and A. Gersho, "Vector predictive coding of speech at 16 kbits/s," *IEEE Trans. Commun.*, vol. COM-33, no. 7, pp. 83–89, 1985.

[6] P. Yahampath and P. Rondeau, "Multiple desctiption predictive vector quantization with applications to low-bit rate speech coding over networks," *IEEE Trans. Speech Audio Processing*, To appear.

[7] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Inform. Theory*, vol. 36, no. 4, pp. 799–809, July 1990.

[8] A. Gersho and R. M. Gray, *Vector Quantization*, Kluwer Academic Publishers, 1992.