ITW2004, San Antonio, Texas, October 24 – 29, 2004

Predictive Coding of Correlated Sources

Ertem Tuncel
Dept. of Electrical Engineering
University of California, Riverside
E-mail: ertem@ee.ucr.edu

Abstract — **A lossy coding scheme is proposed for separate encoding and joint decoding of two correlated sequences. The algorithm simultaneously exploits the correlation between the sequences (using a binning-based quantization scheme) and that between the samples of each sequence (using linear prediction). Under the proposed coding regime, optimal prediction filter design fundamentally deviates from the traditional approach. More specifically, it is, in general, not optimal to employ first-order prediction for noisy observations of a first-order Markov source, even when the noise is negligibly small. Moreover, even if the prediction filter is constrained to be of degree 1, the optimal filter coefficient is different from the correlation coefficient of the Markov source. In the particular example treated in this paper, it is shown that optimal first- and second-order prediction respectively enjoy up to 0.9dB and 1.15dB improvement over the traditional approach.**

Index terms **- Binning, coding of correlated sources, linear prediction, Markov sources.**

I. Introduction

The theory of distributed source coding was initiated by the seminal work of Slepian and Wolf [6]. They analyzed the fundamental limits of lossless compression when a pair of data sequences are encoded, simultaneously and *separately*, to be jointly decoded by a central decoder. It is implied by the separate encoding that the correlation between the two sequences can only be exploited through the underlying joint probability distribution, but not through actual realizations of data. The need for this type of compression arises in sensor networks where bandwidth and power are both severely constrained, and therefore communication is only permitted between each encoder and the central decoder, but not between the encoders themselves.

Wyner and Ziv [7] later on introduced distributed *lossy* compression by extending the work of Slepian and Wolf to a special distributed scenario where a prescribed level of distortion is allowed in the reconstruction of *one* of the sequences when the other sequence is transmitted losslessly in a traditional (point-to-point) fashion. That scenario was then extended to truly distributed cases [1], [2], where (possibly different levels of) distortion is allowed in the reconstruction of both sequences.

A practical distributed lossy compression algorithm can most naturally be considered in the general (and optimal) framework of vector quantization (VQ) [5], applied to the specific distributed setting [3]. The design problem can be posed as the minimization of a cost function, and an iterative method based on the generalization of the well-known Lloyd algorithm [5] can be devised, as done in [3]. On the other hand, VQ-based algorithms suffer from an even more severe initialization problem than the ordinary VQ design [5], because of the increased degree of freedom brought by binning. Also, design and storage complexities still increase exponentially in the block-length N, which must be large in order to increase compression efficiency when there is high temporal correlation.

In this work, we propose predictive coding of correlated sources as a methodology to circumvent the mentioned difficulties of VQ-based algorithms. The quantization followed by the prediction step is to be performed over *scalar* prediction error samples. Coding of correlated scalars is a relatively easier task in terms of complexity and initialization, and a preliminary work in this direction [4] already yielded promising results. For that purpose, we propose a family of codes based on a deterministic binning strategy. Our codes effectively *cover* the joint support set, i.e., they reliably encode the pairs that are in the support, but neglect the distortion that is caused by occasional pair of samples that fall out of the support. We also show that in the high-resolution uniform quantization regime, the average distortion achieved by our scheme is the same as that which would be achieved if both of the scalars were available at a single sensor node [5, Section 8.3]. Therefore, the proposed coding scheme exhibits the same phenomenon as the Slepian-Wolf theorem reveals, this time in *lossy* coding. That is, the potential disadvantage brought by the distributed nature of data vanishes in the high-resolution regime.

The task of finding optimal prediction filters is not as straightforward as in the point-to-point communication scenario. In the latter, one simply finds the filter coefficients that minimize the variance of the prediction error by solving the well-known *normal* equations [5, Section 4.3]. This, in turn, approximately minimizes the end-to-end distortion in a closed-loop scheme [5, Section 7.3]. The fundamental difference here is that, in our framework, the distortion achieved at any time instant n after the prediction step becomes approximately proportional to the square-root of the determinant of the covariance matrix for the two prediction error samples at time n. Minimization of this distortion results in prediction coefficients that are different from those which would be obtained by solving the normal equations. We also show that, in general, it also results in a different filter *order*.

For demonstration purposes, we work on a correlated sequence pair generated as the noisy observations of a first order Gauss-Markov sequence, where the SNR is 30dB, i.e., the noise is negligibly small. We show for this sequence pair that the SNR gain of (i) second-order optimal prediction over no prediction is up to 4dB, (ii) first-order optimal prediction over traditional prediction is up to 0.9dB, and finally (iii) second-order optimal prediction over traditional prediction is up to 1.15dB.

II. A Family of Lossy Codes for Correlated Sources

We propose a systematic family of codes, parameterized by three integers, W, N_X, and N_Y, for separate encoding and joint decoding of two correlated random variables X and Y (Although the proposed codes will be exclusively used on prediction error sequences, the discussion in this section will be for a generic pair $\{X, Y\}$). The dynamic ranges of both X and Y are divided into WN_XN_Y uniform intervals, thereby defining a uniform grid on the two dimensional plane. We denote the resultant cells by $C_{i,j}$, and define

$$\mathcal{R}_W = \bigcup_{2|i-j|\leq W} C_{i,j} \ .$$

The parameters N_X, N_Y, and W are chosen so as to satisfy

$$\Pr[\mathcal{R}_W] \approx 1 \ . \tag{1}$$

For example, in Figure 1, $N_X = N_Y = 2$ and $W = 3$, and therefore both dynamic ranges are divided into 12 intervals, creating 144 cells. The shaded area indicates $\mathcal{S}_{XY}$, the *support* of the joint distribution of $\{X, Y\}$, and as seen from the figure, the choice $W = 3$ satisfies (1).

The key observation (which we are able to prove rigorously for any odd W and any $N_X, N_Y > 1$) is that, the codeword assignment

$$\begin{aligned}
\mathcal{C}_X(i) &= \left(\mathcal{C}_Y(i) + \left\lfloor \frac{i}{WN_Y} \right\rfloor W\right) \bmod WN_X \\
\mathcal{C}_Y(j) &= j \bmod WN_Y
\end{aligned}$$

for the intervals $0 \leq i, j \leq WN_XN_Y - 1$ satisfying $2|i-j| \leq W$ is *uniquely decodable* within the approximation (1). That is, any cell $C_{i,j} \in \mathcal{R}_W$ can be uniquely identified given $\{\mathcal{C}_X(i), \mathcal{C}_Y(j)\}$. To see that, denote by $i_1, i_2, \ldots, i_{N_X}$ the solutions of $\mathcal{C}_Y(i) = k$ where $0 \leq k < WN_Y$. Observe that $\mathcal{C}_X(i_n)$ forms a linear sequence in modulo WN_X with "slope" W. That, in turn, implies that the subsets $\{\mathcal{C}_X(i_n - (W-1)/2), \ldots, \mathcal{C}_X(i_n + (W-1)/2)\}$ do not overlap for any n.

The resultant rates of transmission achieved by the code $\{\mathcal{C}_X, \mathcal{C}_Y\}$ is given by

$$R_X = \lceil \log_2(WN_X) \rceil \tag{2}$$
$$R_Y = \lceil \log_2(WN_Y) \rceil \tag{3}$$

Also, the corresponding lengths of the uniform intervals are

$$\Delta_X = \frac{2d_X}{WN_XN_Y} \tag{4}$$
$$\Delta_Y = \frac{2d_Y}{WN_XN_Y} \tag{5}$$

where d_X and d_Y denote the dynamic range of X and Y, respectively.

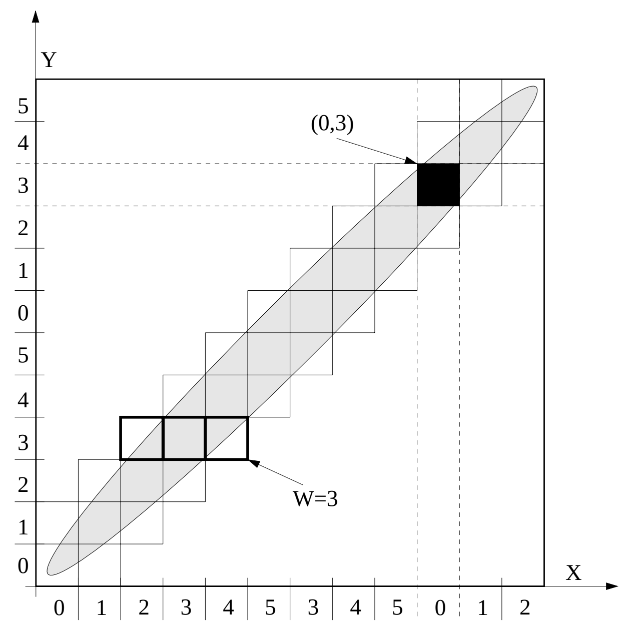

Figure 1: Proposed coding scheme with $W = 3$, $N_X = 2$, and $N_Y = 2$. The reader can verify that the decoder can uniquely identify the cells in $\mathcal{R}_W$. For example, although there are four cells which are encoded as $\{0, 3\}$, only the indicated cell has significant probability.

Assuming further that $\{X, Y\}$ is a Gaussian distribution with zero mean and covariance

$$\mathbf{C} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix} ,$$

we can set the support $\mathcal{S}_{XY}$ to be the set of all (x, y) that satisfy

$$[x \quad y]\, \mathbf{C}^{-1} \begin{bmatrix} x \\ y \end{bmatrix} \leq k^2 \ ,$$

where k is appropriately chosen so that $\Pr[\mathcal{S}_{XY}] \approx 1$. The dynamic ranges then become

$$d_X = \max_x \{(x,y) \in \mathcal{S}_{XY}\} = k\sigma_X$$

and similarly $d_Y = k\sigma_Y$. We define the *width* and the *height* of $\mathcal{S}_{XY}$ as

$$\begin{aligned} w_X &= 2\max_x \{(x,0) \in \mathcal{S}_{XY}\} = 2k\sigma_X\sqrt{1-\rho^2} \\ w_Y &= 2\max_y \{(0,y) \in \mathcal{S}_{XY}\} = 2k\sigma_Y\sqrt{1-\rho^2} \end{aligned}$$

and for any W, choose N_X and N_Y so that

$$\frac{W}{WN_XN_Y} \approx \frac{w_X}{2d_X} = \frac{w_Y}{2d_Y}$$

which implies

$$\frac{1}{N_XN_Y} \approx \sqrt{1-\rho^2} \ . \tag{6}$$

As can be verified from Figure 1, with this choice of N_X and N_Y, the code satisfies $\mathcal{S}_{XY} \subset \mathcal{R}_W$ assuming a large number of intervals, i.e., $WN_XN_Y \gg 1$. Once N_X and N_Y are fixed to satisfy (6), W determines the achieved rate-distortion point. More specifically, from (2)-(5), and using the fact that the quantization noise is approximately uniform in the high-resolution regime, we obtain

$$\begin{aligned} D_X &\approx \frac{\Delta_X^2}{12} \\ &= \frac{d_X^2}{3W^2N_X^2N_Y^2} \\ &= \frac{k^2\sigma_X^2}{3W^2N_XN_Y}\sqrt{1-\rho^2} \\ &= \frac{k^2}{3}\sigma_X^2\sqrt{1-\rho^2}\,2^{-(R_X+R_Y)} \end{aligned} \tag{7}$$

and similarly

$$D_Y \approx \frac{k^2}{3}\sigma_Y^2\sqrt{1-\rho^2}\,2^{-(R_X+R_Y)} \ . \tag{8}$$

If we pursue the minimization of the average distortion $\frac{D_X+D_Y}{2}$ for a fixed total rate R_X+R_Y, it is implied by (7) and (8) that we must target $D_X \approx D_Y$ [5, Section 8.3]. This requirement is automatically satisfied if $\sigma_X^2 = \sigma_Y^2$. On the other hand, if $\sigma_X^2 \neq \sigma_Y^2$, then one can always construct a *second layer* of cells (with no binning structure) inside each original cell $C_{i,j}$, thereby equalizing D_X and D_Y. This results in

$$D_X \approx D_Y \approx \frac{k^2}{3}\sqrt{\det(\mathbf{C})}\,2^{-(R_X+R_Y)} \ , \tag{9}$$

where R_X and R_Y refer to the total rates after the equalization of D_X and D_Y. In other words, for a fixed total rate, the quality of the coder is determined by $\sqrt{\det(\mathbf{C})}$.

We close this section with the demonstration of a phenomenon which is somewhat reminiscent of the Slepian-Wolf result [6]. Within the above approximations, the following lemma proves that our coding scheme achieves the same distortion as a hypothetical *joint* coding scenario where *rotated* variables X' and Y' are encoded using uniform scalar quantizers.

Lemma 1 *Given the total rate $\bar{R} = R_X + R_Y$, application of the Karhunen-Loeve Transform (KLT) on a jointly Gaussian pair $\{X, Y\}$ followed by optimal bit allocation results in an average distortion given by* (9)*.*

Proof: The eigenvalues of $\mathbf{C}$ can be calculated as

$$\begin{aligned} \lambda_1 &= \frac{\sigma_X^2+\sigma_Y^2+\sqrt{(\sigma_X^2+\sigma_Y^2)^2-4\sigma_X^2\sigma_Y^2(1-\rho^2)}}{2} \\ \lambda_2 &= \frac{\sigma_X^2+\sigma_Y^2-\sqrt{(\sigma_X^2+\sigma_Y^2)^2-4\sigma_X^2\sigma_Y^2(1-\rho^2)}}{2} \end{aligned}$$

Employing uniform quantizers on the KLT coefficients X' and Y' with bit rates R'_X and R'_Y yields

$$D'_X \approx \frac{k^2\lambda_1 2^{-2R'_X}}{3} \ , \ D'_Y \approx \frac{k^2\lambda_2 2^{-2R'_Y}}{3} \ .$$

As mentioned above, to minimize $\frac{D'_X+D'_Y}{2}$ subject to fixed $R'_X + R'_Y = \bar{R}$, one needs to set $D'_X = D'_Y$ [5], which yields

$$\frac{1}{2}\log\frac{\lambda_1}{\lambda_2} = R'_X - R'_Y \ ,$$

and hence

$$\begin{aligned} R'_X &= \frac{\bar{R}}{2} + \frac{1}{4}\log\frac{\lambda_1}{\lambda_2} \\ R'_Y &= \frac{\bar{R}}{2} - \frac{1}{4}\log\frac{\lambda_1}{\lambda_2} \ . \end{aligned}$$

Therefore,

$$\begin{aligned} D'_X = D'_Y &= \frac{k^2\lambda_1}{3}2^{-2\left(\frac{\bar{R}}{2}+\frac{1}{4}\log\frac{\lambda_1}{\lambda_2}\right)} \\ &= \frac{k^2}{3}\sqrt{\lambda_1\lambda_2}\,2^{-\bar{R}} \\ &= \frac{k^2}{3}\sigma_X\sigma_Y\sqrt{1-\rho^2}\,2^{-\bar{R}} \\ &= \frac{k^2}{3}\sqrt{\det(\mathbf{C})}2^{-(R_X+R_Y)} \ , \end{aligned}$$

which completes the proof.

■

III. Prediction Filter Design

In this section, we turn our attention back to distributed sequence pairs $\{X[n], Y[n]\}$, and discuss how prediction filters must be designed. We will discuss how optimal linear prediction filters are different from those which would be obtained by solving normal equations

$$\mathbf{R}\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{bmatrix} = \begin{bmatrix} r[1] \\ r[2] \\ \vdots \\ r[K] \end{bmatrix} ,$$

of order K, where $R_{i,j} = r[i-j]$ for $1 \leq i, j \leq K$, and $r[\cdot]$ denotes the autocorrelation function. The reason is that, as (9) suggests, minimizing the variance of the prediction error is no longer the optimal strategy. Rather, as we will show in the next section, it may result in significant SNR loss.

One may initially argue that closed-loop prediction (prediction of incoming samples from quantized past samples) is not possible in the context of separate coding of correlated sources, because it is not obvious how the encoders can replicate the data reconstructed at the decoder without knowing the output of each other. On the other hand, closed-loop prediction is preferable as the distortion at the quantizer output truly reflects the end-to-end distortion. Now observe that by putting the reconstruction vectors on a grid, as opposed to optimally designing them for each individual cell $C_{i,j}$, one can allow the encoders to replicate the data at the decoder end. More specifically, when the grid codebook structure is adopted, $\hat{e}_X[n]$, reconstruction of the prediction error $e_X[n]$, depends only on the interval which $e_X[n]$ falls into (and likewise for $\hat{e}_Y[n]$), provided $\{e_X[n], e_Y[n]\} \in \mathcal{R}_W$. Also, if the prediction filter is constrained to be minimum-phase, occasional errors caused by $\{e_X[n], e_Y[n]\} \notin \mathcal{R}_W$ will be quickly forgotten thanks to the stability of the inverse filter at the decoder. Even though the grid constraint on the codebook slightly decreases the compression efficiency at the quantizer output, it results in dramatically better reconstruction at the decoder end compared to an open-loop scheme[1].

The prediction error sequences in a closed-loop scheme are given by

$$\begin{aligned} e_X[n] &= h_X[n] * \hat{X}[n] \\ e_Y[n] &= h_Y[n] * \hat{Y}[n] \end{aligned}$$

where $*$ denotes convolution, and $h_X[n]$ and $h_Y[n]$ are real causal minimum-phase FIR sequences with $h_X[0] = h_Y[0] = 1$. Under the high-resolution regime, however, quantized output sequences $\hat{X}[n]$ and $\hat{Y}[n]$ can be approximated with their true values, and therefore

$$\begin{aligned} e_X[n] &\approx h_X[n] * X[n] \\ e_Y[n] &\approx h_Y[n] * Y[n] \,. \end{aligned}$$

Assuming that $X[n]$ and $Y[n]$ are zero-mean and jointly wide-sense stationary (WSS) sequences, we have

$$\begin{aligned} r_{e_X e_X}[k] &= h_X[k] * h_X[-k] * r_{XX}[k] \\ r_{e_Y e_Y}[k] &= h_Y[k] * h_Y[-k] * r_{YY}[k] \\ r_{e_X e_Y}[k] &= h_X[k] * h_Y[-k] * r_{XY}[k] \end{aligned}$$

where $r_{fg}[k] = E\{f[m]g[m-k]\}$ is the correlation sequence between the generic sequences f and g. We focus on the special case where $r_{XX}[k] = r_{YY}[k]$ and $h_X[n] = h_Y[n]$. It then follows from (9) that the achieved quantization distortion for $\{e_X[n], e_Y[n]\}$ is proportional to

$$\alpha(h_X) = \sqrt{r_{e_X e_X}[0]^2 - r_{e_X e_Y}[0]^2} \,.$$

Even though it is still tedious to minimize $\alpha(h_X)$ over all real causal minimum-phase FIR filters, one can further constrain the prediction order K and perform a search over the K-dimensional coefficient space.

In this paper, we adopt a slightly better approach, and use (6) only as a guideline to choose $N_X N_Y$, but not as an exact formula. More specifically, we choose $N_X N_Y$ as the largest integer that satisfies

$$\frac{W-2}{W N_X N_Y} \geq \sqrt{1-\rho^2} \,, \qquad (10)$$

where the extra factor $\frac{W-2}{W}$ is used as a "cushion" to ensure that (1) is satisfied, and use

$$D_X \approx D_Y \approx \frac{k^2 \sigma_X^2}{3W^2 N_X^2 N_Y^2} \,. \qquad (11)$$

For a fixed W, we exhaustively search over all K-dimensional minimum-phase filter coefficients, and for each set of coefficients, obtain rate-distortion pairs using (10) and (11). We then compute the convex-hull of the rate-distortion curves corresponding to each W.

IV. Experimental Results

A. The Setup

Let the process $Z[n]$ be defined as a first-order Gauss-Markov process, i.e.,

$$Z[n] = \eta Z[n-1] + \epsilon[n]$$

where $0 < \eta < 1$, and $\epsilon[n]$ is white Gaussian noise with zero mean and variance σ_ϵ^2 and is independent of $Z[n]$. It can be easily shown that

$$r_{ZZ}[k] = \sigma_Z^2 \eta^{|k|}$$

[1]The disadvantage of open-loop prediction is that inverse filters amplify the distortion. For example, if the prediction is first order with coefficient a, the inverse filter amplifies the quantizer distortion by $\frac{1}{1-a^2}$.

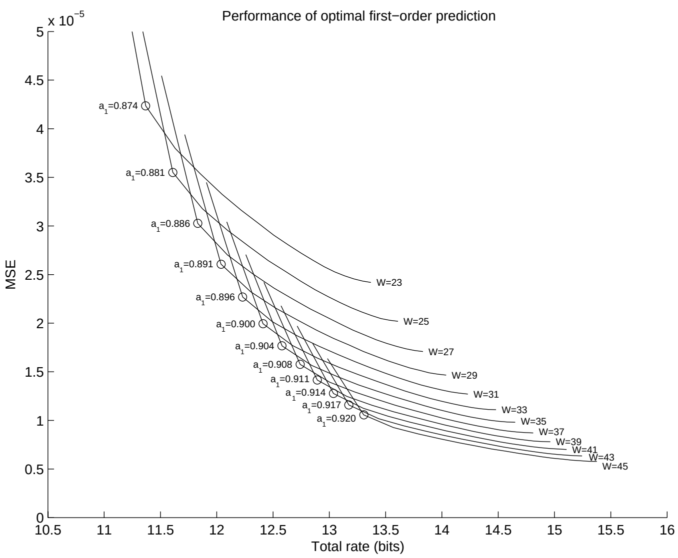

Figure 2: Rate-distortion curves for first-order prediction, obtained for various W. The convex-hull is indicated with circles, accompanied by the optimal values of a_1.

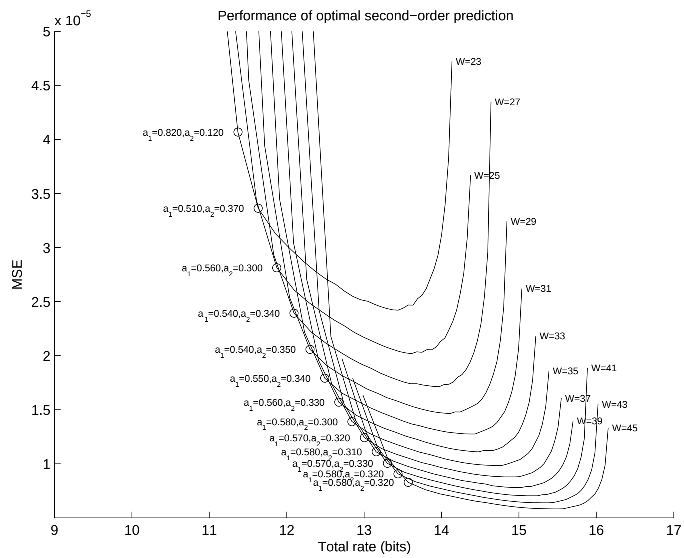

Figure 3: Rate-distortion curves for second-order prediction. The convex-hull is indicated with circles, accompanied by the optimal values of a_1 and a_2.

where $\sigma_Z^2 = \frac{\sigma_\epsilon^2}{1-\eta^2}$. Now, let two distributed sensors observe noisy versions of this process, i.e.,

$$\begin{aligned} X[n] &= Z[n] + \tau[n] \\ Y[n] &= Z[n] + \gamma[n] \end{aligned}$$

where $\tau[n]$ and $\gamma[n]$ are both white Gaussian noise signals with zero mean and variance σ_τ^2, and both are independent of $X[n]$ and $Y[n]$. Therefore,

$$r_{XX}[k] = r_{YY}[k] = r_{ZZ}[k] + \sigma_\tau^2\delta[k] = \sigma_Z^2\eta^{|k|} + \sigma_\tau^2\delta[k]$$

where $\delta[k]$ denotes the Kronecker delta function. Similarly

$$r_{XY}[k] = r_{ZZ}[k] = \sigma_Z^2\eta^{|k|} .$$

We will consider filters of order up to 2. For $H_X(z) = 1 - a_1z^{-1} - a_2z^{-2}$, it is easy to obtain

$$\begin{aligned} &r_{e_Xe_X}[0] \\ &\quad= (1 + a_1^2 + a_2^2)(\sigma_Z^2 + \sigma_\tau^2) - 2[a_1(1-a_2)\eta + a_2\eta^2]\sigma_Z^2 \\ &r_{e_Xe_Y}[0] \\ &\quad= (1 + a_1^2 + a_2^2 - 2\eta a_1(1-a_2) - 2a_2\eta^2)\sigma_Z^2 . \end{aligned}$$

Also, by definition,

$$\rho = \frac{r_{e_Xe_Y}[0]}{r_{e_Xe_X}[0]}.$$

B. Performance Comparisons

We compare the performance of our coding scheme (with optimized filter coefficients) to a purely binning-based scheme, where there is no prediction filter (i.e., $a_1 = a_2 = 0$ in our own scheme), and to the traditional approach, where filter coefficients are calculated using normal equations.

We have set $\eta = 0.95$, $\sigma_Z^2 = 1$, and $\sigma_\tau^2 = 0.001$, thereby setting the SNR of the observed sequences $X[n]$ and $Y[n]$ to

$$10\log_{10}\left(\frac{\sigma_Z^2}{\sigma_\tau}\right) = 30\text{dB} .$$

In this high-SNR setting, $X[n]$ and $Y[n]$ are themselves *approximately* first-order Markov processes, and solving the normal equations for a prediction filter of order 1 yields $a_1 = 0.949$. However, as we show next, we observe up to 1.15dB gain over the traditional filter $H_X(z) = 1 - 0.949z^{-1}$ when the prediction filter is designed as described in Section III.

Figure 2 depicts the (operational) rate-distortion curves obtained by exhaustively searching over all first-order filters, and for each filter, using (10) and (11) with $k = 4$. As described before, we obtain a curve for each W, and compute the convex-hull for the characterization of the rate-distortion performance. We observe that the optimal prediction filter of order 1 changes as we traverse the overall rate-distortion curve. In fact, we have also observed that the optimal filter coefficient converges to $a_1 = 0.949$ when the rate is sufficiently large (not shown on the figure). On the other hand, minimization of (9) yields $a_1 = 0.899$. This discrepancy is due to the inaccuracy of the approximation (6).

Similarly, the rate-distortion curves for second-order prediction is shown in Figure 3, together with the optimal filter

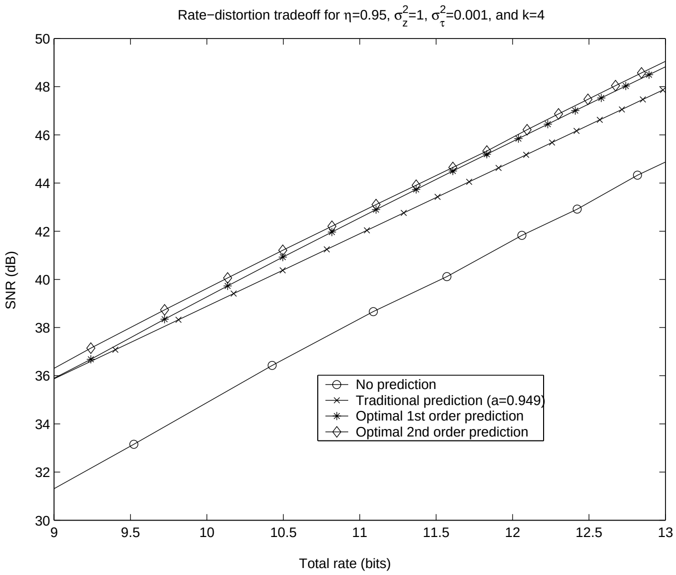

Figure 4: Comparison of SNR results for $k = 4$, $\eta = 0.95$, $\sigma_Z^2 = 1$, and $\sigma_\tau^2 = 0.001$.

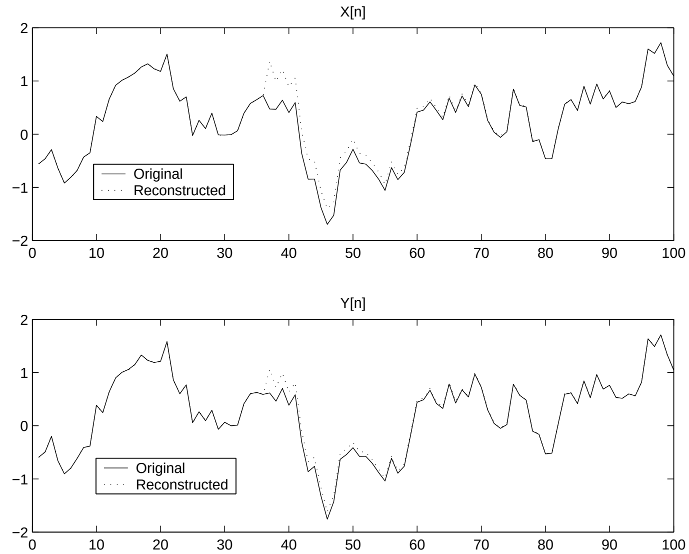

Figure 5: The original and the reconstructed sequences for $W = 35$, $N_X = 2$, $N_Y = 3$, $a_1 = 0.58$, and $a_2 = 0.3$.

coefficients as we traverse the overall rate-distortion curve. This time we observed that the optimal filter coefficients converge to $a_1 = 0.62$, $a_2 = 0.3$, whereas minimization of (9) yields $a_1 = 0.63$, $a_2 = 0.28$.

Figure 4 is a comparison of the above first- and second-order prediction filters to no prediction and traditional prediction. We observe up to 4dB improvement over the no-prediction method which exploits only the correlation *between* the samples. On the other hand, optimal first- and second-order prediction filter design in our regime respectively show up to 0.9dB and 1.15dB improvement over traditional prediction, which attempts to strip off all the dependency *within* each sequence. The success of our scheme is that it strikes an optimal balance between these two extremes.

We also exhibit in Figure 5 the output of our prediction scheme for a pair of randomly generated sequences of length 100,000. In the figure, only a portion of length 100 is depicted, as that interval contains the only time instant where $\{e_X[n], e_Y[n] \notin \mathcal{R}_W\}$. The error propagates because the encoders and the decoder is out of synch, but the effect of this propagation quickly diminishes as time proceeds.

References

[1] T. Berger, "Multiterminal source encoding," in *The Information Theory Approach to Communications*, G. Longo, Ed., CISM Courses and Lectures 229. Springer, New York, 1978.

[2] T. Berger and R. W. Yeung, "Multiterminal source encoding with one distortion criterion," *IEEE Transactions on Information Theory*, 35(2):228–236, March 1989.

[3] M. Fleming, Q. Zhao, and M. Effros, "Network vector quantization," *IEEE Transactions on Information Theory*, 50(8):1584-1604, August 2004.

[4] T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Transactions on Information Theory*, 33(6):773–787, November 1987.

[5] A. Gersho, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston, MA, 1992.

[6] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, 19(4):471–480, July 1973.

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