OPTIMAL DELAYED DECODING OF PREDICTIVELY ENCODED SOURCES

Vinay Melkote and Kenneth Rose

Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106
{melkote, rose}@ece.ucsb.edu

ABSTRACT

Predictive coding eliminates redundancy due to correlations between the current and past signal samples, so that only the innovation, or prediction residual, needs to be encoded. However, the decoder may, in principle, also exploit correlations with future samples. Prior decoder enhancement work mainly applied a non-causal filter to smooth the regular decoder reconstruction. In this work we broaden the scope to pose the problem: Given an allowed decoding delay, what is the optimal decoding algorithm for predictively encoded sources? To exploit all information available to the decoder, the proposed algorithm recursively estimates conditional probability densities, given both past and available future information, and computes the optimal reconstruction via conditional expectation. We further derive a near-optimal low complexity approximation to the optimal decoder, which employs a time-invariant lookup table or codebook approach. Simulations indicate that the latter method closely approximates the optimal delayed decoder, and that both considerably outperform the competition.

Index Terms— Predictive coding, smoothing, delayed decoding, DPCM, recursive estimate

1. INTRODUCTION

Predictive coding is widely used in signal compression, for instance, in differential pulse code modulation (DPCM), motion compensated video coding in H.264, sub-band adaptive DPCM in G.722 speech coding, etc. The development of predictive coding schemes generally assumes an autoregressive model of the source [1]-[4]. In this paper we will largely focus on a first-order autoregressive (or Markov) model of the signal (generalization of the proposed schemes to higher order processes will be briefly discussed). The source consists of a zero-mean stationary sequence $\{x_n\}$ of real-valued random variables with,

$$x_n = \rho x_{n-1} + z_n \, . \qquad (1)$$

The random variables $\{z_n\}$ are independent and identically distributed (i.i.d), with probability density function (pdf) $p_Z(z)$, and are referred to as the innovations of the process. The correlation coefficient of adjacent samples is ρ. The objective of predictive coding is to de-correlate the current sample from the past so that only the new information (or innovation) in the current sample needs to be coded. One of the simplest forms of predictive coding is DPCM [1]. The DPCM encoder generates a prediction $\tilde{x}_n$, based on prior reconstructions, subtracts it from the current sample x_n to generate the prediction error e_n, which is quantized using a scalar quantizer $\mathcal{Q}$. $\mathcal{Q}$ is specified by the mappings $f_{\mathcal{Q}}(x) : \mathbf{R} \rightarrow \mathbf{I}$, and $g_{\mathcal{Q}}(x) : \mathbf{I} \rightarrow \mathbf{R}$, where $\mathbf{R}$ is the real line, and $\mathbf{I}$, a countable index set. The quantization index $i_n = f_{\mathcal{Q}}(e_n)$ is entropy coded and sent to the decoder, which generates $\hat{x}_n = \tilde{x}_n + g_{\mathcal{Q}}(i_n)$. At high rate, $\hat{x}_{n-1} \approx x_{n-1}$, implying the optimal prediction

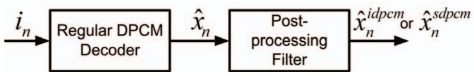
$$\tilde{x}_n = \rho \hat{x}_{n-1}. \qquad (2)$$

Even at low bit-rates this form of the predictor is commonly employed.

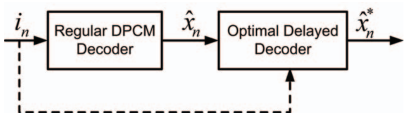
In an autoregressive source model (such as (1)), x_n is correlated not just with samples from the past, but also with the future, i.e., with $\{x_l\}_{n+1}$ (henceforth, with respect to any sequence $\{a_n\}$, the notation $\{a_l\}_m^k$, $\{a_l\}_m$, and $\{a_l\}^k$ denote, respectively, the truncated sequences $\{a_l : m \leq l \leq k\}$, $\{a_l : l \geq m\}$, and $\{a_l : l \leq k\}$. Thus, $\{x_l\}_{n+1}$ denotes $\{x_l : l \geq n+1\}$). At high rate, the prediction error $e_n \approx z_n \; \forall n$ and hence $\{i_n\}$ are approximately i.i.d. In this case future indices $\{i_l\}_{n+1}$ provide no additional information on x_n. In practical, limited bit-rate scenarios, such approximations do not hold, and $\{i_l\}_{n+1}$ contains information on x_n, which can be used to improve its reconstruction at the decoder. Naturally, this entails decoding delay. This fact has been previously exploited by the interpolative DPCM (IDPCM)[3], and smoothed DPCM (SDPCM) [4] approaches, both of which *smooth* (i.e., filter) the regular DPCM outputs $\{\hat{x}_n\}$ with a suitable *non-causal* post-processor to generate refined estimates, as depicted in Fig. 1a. While more details are provided in Sec. 2, suffice it to say that the design of the post-processor in either scheme is heuristic and depends on assumptions, for instance about the quantizer resolution, which preclude performance guarantees relative to regular DPCM. A more significant shortcoming is that by merely filtering $\{\hat{x}_n\}$ (see Fig. 1a), while disregarding the indices $\{i_n\}$, as well as knowledge about the exact operation of modules such as the quantizer and predictor, these methods under-utilize the information available at the decoder. The following simple argument illustrates this suboptimality: The decoder knows $(\tilde{x}_n, i_n, \mathcal{Q})$ that determine the effective quantization interval $\mathcal{I}_n = \{x \in \mathbf{R} : f_{\mathcal{Q}}(x - \tilde{x}_n) = i_n\}$ in which x_n *must* lie, but simple smoothing of $\{\hat{x}_n\}$ may produce a reconstruction that lies outside $\mathcal{I}_n$.

In contrast, we propose an estimation-theoretic (ET) delayed decoder that *optimally* combines all the information (i.e., indices $\{i_l\}^{n+L}$ or equivalently the intervals $\{\mathcal{I}_l\}^{n+L}$) available at the decoder, for a given delay (or 'look-ahead') L, in a recursively obtained conditional pdf, to guarantee the best reconstruction of the current sample x_n. Fig. 1b contrasts the proposed approach with the prior work. This optimal delayed decoder in turn motivates an approximate decoder whose reconstruction is of the form $\tilde{x}_n + c(\{i_l\}_n^{n+L})$ (implementable as a time-invariant codebook over indices $\{i_l\}_n^{n+L}$), which is observed in experiments to provide near-optimal performance, and with the obvious benefit of low-complexity decoding (even compared to time-invariant linear smoothing). Simulation re-

The work was supported in part by the NSF under grant CCF-0917230, the University of California MICRO Program, Applied Signal Technology Inc., Cisco Systems Inc., Qualcomm Inc., and Sony Ericsson, Inc..



(a) IDPCM [3] or SDPCM [4]



(b) Proposed optimal delayed decoder

Fig. 1. Prior work merely smooth the regular DPCM reconstructions; the Optimal Delayed Decoder exploits all available information

sults demonstrate that both proposed methods substantially outperform IDPCM and SDPCM. We note that the ET approach in [5] to (zero delay) scalable predictive coding, which efficiently utilizes base-layer quantization interval information for enhancement layer prediction, was an early inspiration for this work.

2. PRELIMINARIES

We briefly describe prior work, [3] and [4], as applied to (1).

2.1. IDPCM

IDPCM [3] consists of determining the coefficients $b_l, \; -L' \leq l \leq L, \; l \neq 0$, that minimize $E[(x_n - \sum_{l=-L',l\neq 0}^{L} b_l x_{n+l})^2]$, and obtaining the smoothed estimate of x_n as,

$$\hat{x}_n^{idpcm} = \sum_{l=-L'}^{-1} b_l \hat{x}_{n+l}^{idpcm} + \sum_{l=1}^{L} b_l \hat{x}_{n+l} \qquad (3)$$

where $\hat{x}_n$ are the regular DPCM reconstructions. It is shown in [3] that b_l values depend only on the auto-correlation matrix (i.e., only on ρ), irrespective of the bit-rate, or process distributions. It is easily seen that for the first order process (1), $b_{-1} = b_1 = \frac{\rho}{1+\rho^2}$, and for $l \notin \{-1, 1\}$, $b_l = 0$, i.e., the look-ahead L is automatically restricted to 1. We note that for higher order processes, IDPCM entails altering the predictor as well [3].

2.2. SDPCM

In SDPCM [4] a Kalman filtering-based fixed-lag smoother is used. The autoregressive process (1) provides the following 'plant' model for a fixed-lag (i.e., look-ahead) L:

$$\underbrace{\begin{bmatrix} x_n \\ x_{n-1} \\ \vdots \\ x_{n-L} \end{bmatrix}}_{\mathbf{x}_n} = \underbrace{\begin{bmatrix} \rho & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\Phi} \mathbf{x}_{n-1} + \begin{bmatrix} z_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} . \qquad (4)$$

The quantization process provides the 'observation' model:

$$\hat{x}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \mathbf{x}_n + w_n, \qquad (5)$$

where w_n is the DPCM quantization noise, assumed to be white. Kalman filtering [6] is now used to obtain the best estimate $\hat{\mathbf{x}}(n|n)$ of $\mathbf{x}_n$, given observations up to time n, and the smoothed estimate

$$\hat{x}_n^{sdpcm} = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \hat{\mathbf{x}}(n+L|n+L) \qquad (6)$$

Although the Kalman filter is generally adaptive, for the time-invariant model given by (4) and (5), it can be shown that it has a 'steady state' [6] when $|\rho| < 1$, implying a time-invariant smoother. Since $\{x_n\}$ are correlated, the assumption that $\{w_n\}$ are white is invalid, especially at low bit-rates. SDPCM, like IDPCM, ignores the process distribution.

3. THE OPTIMAL DELAYED DECODER

We now describe the proposed optimal delayed decoder. The DPCM encoder is not altered in any way. Therefore, $\{i_l\}^n$ known at the decoder, determine $\{\mathcal{I}_l\}^n$ exactly, $\{i_l\}^n \Leftrightarrow \{\mathcal{I}_l\}^n$. Let mean squared error (MSE) be the distortion criterion. Then the *optimal estimate* $\hat{x}_n^*$, for a fixed look-ahead L, is given by

$$\hat{x}_n^* = E[x_n|\{i_l\}^{n+L}] = E[x_n|\{\mathcal{I}_l\}^{n+L}] \,, \qquad (7)$$

the expectation over the conditional pdf $p(x_n|\{\mathcal{I}_l\}^{n+L})$. Thus, $\hat{x}_n^*$ can be obtained if this density is known. We use the streamlined notation $p(\cdot)$ to denote any pdf or probability, and add a subscript whenever the interpretation is not evident from the arguments. Note that the above conditional pdf automatically limits the optimal estimate to the interval $\mathcal{I}_n$. We now write

$$p(x_n|\{\mathcal{I}_l\}^{n+L}) = \frac{p(x_n|\{\mathcal{I}_l\}^n)p(\{\mathcal{I}_l\}_{n+1}^{n+L}|x_n)}{\int p(x_n|\{\mathcal{I}_l\}^n)p(\{\mathcal{I}_l\}_{n+1}^{n+L}|x_n)dx_n} \,. \qquad (8)$$

Unless otherwise indicated, integrals are over $\mathbf{R}$. The above equality is obtained using Bayes' rule, and the Markov property of the process (1): given x_n, the probability of events $\{\mathcal{I}_l\}_{n+1}^{n+L}$, is independent of any other information from the past (i.e., $\{\mathcal{I}_l\}^n$). Note that $p(x_n|\{\mathcal{I}_l\}^n)$ is the pdf of x_n conditioned on all information up to the current time n. An optimal *zero delay* decoder would simply use the mean of this pdf as the estimate of x_n. The optimal *delayed* decoder though, weighs this pdf with $p(\{\mathcal{I}_l\}_{n+1}^{n+L}|x_n)$ representing the probability, given x_n, of the known future outcomes, to obtain the composite pdf $p(x_n|\{\mathcal{I}_l\}^{n+L})$ in (8) that incorporates all known information up to the fixed delay L. The estimate of x_n is then $\hat{x}_n^*$ of (7). We next provide recursion formulas to calculate $p(x_n|\{\mathcal{I}_l\}^n)$ and $p(\{\mathcal{I}_l\}_{n+1}^{n+L}|x_n)$.

Suppose we have $p(x_{n-1}|\{\mathcal{I}_l\}^{n-1})$, the pdf of x_{n-1} conditioned on all information up to time $n-1$. The corresponding, 'one sample advanced' pdf, $p(x_n|\{\mathcal{I}_l\}^n)$ is obtained as follows. The first step is to find the pdf of x_n conditioned on information up to $n-1$:

$$\begin{aligned} p(x_n|\{\mathcal{I}_l\}^{n-1}) &= \int p(x_{n-1}, x_n|\{\mathcal{I}_l\}^{n-1})dx_{n-1} \\ &= \int p(x_{n-1}|\{\mathcal{I}_l\}^{n-1})p(x_n|x_{n-1})dx_{n-1} \\ &= \int p(x_{n-1}|\{\mathcal{I}_l\}^{n-1})p_Z(x_n - \rho x_{n-1})dx_{n-1} \qquad (9) \end{aligned}$$

The second equality is due to the Markov property of the process (1). Employing the fact that z_n is independent of x_{n-1} and using (1) yields (9). Next update this pdf to include the additional information on interval $\mathcal{I}_n$, available at time n, by appropriate conditioning:

$$p(x_n|\{\mathcal{I}_l\}^n) = \begin{cases} \frac{p(x_n|\{\mathcal{I}_l\}^{n-1})}{\int_{\mathcal{I}_n} p(x_n|\{\mathcal{I}_l\}^{n-1})dx_n} & x_n \in \mathcal{I}_n \\ 0 & else \end{cases} \qquad (10)$$

Substituting the scaled auxiliary variable $y = \rho x_{n-1}$ in (9) shows that it is simply a convolution. In practice, discretized versions of the densities are used, with this convolution efficiently implemented using an FFT, combined with an interpolation/re-sampling operation between updates, for the required axial scaling.

Now consider the following. Say, at time $n+1$ we have the probability $p(\{\mathcal{I}_l\}_{n+2}|x_{n+1})$ of *all* future outcomes given x_{n+1}. The following 'one sample retreat' procedure provides the corresponding probability, $p(\{\mathcal{I}_l\}_{n+1}|x_n)$, at time n.

$$p(\{\mathcal{I}_l\}_{n+1}|x_n) = \int_{\mathcal{I}_{n+1}} p(\{\mathcal{I}_l\}_{n+2}, x_{n+1}|x_n)dx_{n+1}$$

$$= \int_{\mathcal{I}_{n+1}} p(\{\mathcal{I}_l\}_{n+2}|x_{n+1}, x_n)p(x_{n+1}|x_n)dx_{n+1}$$

$$= \int_{\mathcal{I}_{n+1}} p(\{\mathcal{I}_l\}_{n+2}|x_{n+1})p_Z(x_{n+1} - \rho x_n)dx_{n+1}. \quad (11)$$

But for a given look-ahead L, the future information available to reconstruct x_n is just $\{\mathcal{I}_l\}_{n+1}^{n+L}$, i.e., the only knowledge about samples x_l, $l > n+L$, is that they are real. Hence, effectively $\mathcal{I}_l = \mathbf{R}$, $l > n+L$ (obviously with probability one). Thus we initialize $p(\{\mathcal{I}_l\}_{n+L+1}|x_{n+L}) = 1$ and employ (11) L times to 'retreat' from time $n+L$ to n, and obtain the probability $p(\{\mathcal{I}_l\}_{n+1}|x_n) = p(\{\mathcal{I}_l\}_{n+1}^{n+L}|x_n)$ of the known future outcomes given x_n.

The optimal delayed decoder, for a look-ahead L, is thus:

Optimal Delayed Decoder

At time $n+L$

1. Decode (as in regular DPCM) index i_{n+L}, and use $\tilde{x}_{n+L}$ to obtain $\hat{x}_{n+L}$, and the interval $\mathcal{I}_{n+L}$.
2. Update pdf $p(x_{n-1}|\{\mathcal{I}_l\}^{n-1})$ to $p(x_n|\{\mathcal{I}_l\}^n)$, that combines all information available up to time n.
3. Obtain probability $p(\{\mathcal{I}_l\}_{n+1}^{n+L}|x_n)$ as a function of x_n, that combines information about all available future outcomes (relative to time n).
4. Use (7) and (8) to obtain the optimal estimate $\hat{x}_n^*$.

Note that $p(x_n|\{\mathcal{I}_l\}^n)$ needs to be suitably initialized, say at $n=0$. If $|\rho| < 1$, the effect of this initialization wears out with time.

4. CODEBOOK-BASED DELAYED DECODER

With look-ahead $L=1$, the substitution $x_{n+1} = e + \tilde{x}_{n+1}$ in (11) (where $\tilde{x}_{n+1}$ is given by (2)) yields

$$p(\mathcal{I}_{n+1}|x_n) = \int_{\mathcal{I}_\mathcal{Q}(i_{n+1})} 1 \cdot p_Z(e - \rho(x_n - \hat{x}_n))de \,. \quad (12)$$

Here $\mathcal{I}_\mathcal{Q}(i) = \{x \in \mathbf{R} : f_\mathcal{Q}(x) = i\}$ are *time-invariant* intervals characteristic of the quantizer. Therefore $p(\mathcal{I}_{n+1}|x_n)$ is really a function of the form $\Gamma(x_n - \hat{x}_n, i_{n+1})$, i.e., for a given value of the index i_{n+1}, the shape of this function is independent of n. It merely translates, by $\hat{x}_n$, along the real line. By induction it can be shown that $p(\{\mathcal{I}_l\}_{n+1}^{n+L}|x_n)$ is, in fact, of the form $\Gamma(x_n - \hat{x}_n, \{i_l\}_{n+1}^{n+L})$ (the proof is omitted for conciseness). Thus the L-step retreat procedure in Sec. 3 can *equivalently* be replaced by a procedure that collects indices $\{i_l\}_{n+1}^{n+L}$, and translates, by $\hat{x}_n$, the corresponding function read from a codebook (of functions), to obtain $p(\{\mathcal{I}_l\}_{n+1}^{n+L}|x_n)$.

Consider now the issue of predictive quantizer design [1]. If the quantizer thresholds (i.e., $\mathcal{I}_\mathcal{Q}(\cdot)$, or equivalently $f_\mathcal{Q}(\cdot)$) are fixed, the reconstructions $g_\mathcal{Q}(i_n)$ are obtained as

$$g_\mathcal{Q}(i_n) = E[e_n|i_n] = \int e_n p(e_n|i_n)de_n = \frac{\int_{\mathcal{I}_\mathcal{Q}(i_n)} e_n p(e_n)de_n}{\int_{\mathcal{I}_\mathcal{Q}(i_n)} p(e_n)de_n} \,,$$

where $p(e_n)$ is the marginal prediction error density *at time* n. But e_n (and thus $p(e_n)$) is itself dependent on $\{g_\mathcal{Q}(i_l)\}^{n-1}$, through the prediction $\tilde{x}_n$. Thus, to obtain a time invariant $g_\mathcal{Q}(\cdot)$ a recursive optimization needs to be performed, which at convergence yields a corresponding stationary marginal prediction error density $p_E(e)$ [1, 2]. We now make the following approximation

$$p(x_n|\{\mathcal{I}_l\}^n) \approx p(x_n|\tilde{x}_n, \mathcal{I}_n) = \begin{cases} \frac{p_E(x_n - \tilde{x}_n)}{\int_{\mathcal{I}_n} p_E(x_n - \tilde{x}_n)dx_n} & x_n \in \mathcal{I}_n \\ 0 & else \end{cases} \quad (13)$$

where conditioning on $\tilde{x}_n$ is in lieu of the optimal combination (in $p(x_n|\{\mathcal{I}_l\}^n)$) of all past information. Therefore, in case of a predictor of the form (2), we have the approximate estimate,

$$\hat{x}_n^{approx} = \frac{\int_{\mathcal{I}_n} x_n p_E(x_n - \rho\hat{x}_{n-1})p(\{\mathcal{I}_l\}_{n+1}^{n+L}|x_n)dx_n}{\int_{\mathcal{I}_n} p_E(x_n - \rho\hat{x}_{n-1})p(\{\mathcal{I}_l\}_{n+1}^{n+L}|x_n)dx_n} \quad (14)$$

$$= \rho\hat{x}_{n-1} + \frac{\int_{\mathcal{I}_\mathcal{Q}(i_n)} e p_E(e)\Gamma(e - g_\mathcal{Q}(i_n), \{i_l\}_{n+1}^{n+L})de}{\int_{\mathcal{I}_\mathcal{Q}(i_n)} p_E(e)\Gamma(e - g_\mathcal{Q}(i_n), \{i_l\}_{n+1}^{n+L})de} \quad (15)$$

$$\triangleq \rho\hat{x}_{n-1} + c(\{i_l\}_n^{n+L}) \quad (16)$$

Using (13) in evaluating (7) yields (14). Replacing $p(\{\mathcal{I}_l\}_{n+1}^{n+L}|x_n)$ with its canonical form $\Gamma(x_n - \hat{x}_n, \{i_l\}_{n+1}^{n+L})$, a change of variable to $e = x_n - \rho\hat{x}_{n-1}$, and recognizing that $\hat{x}_n - \rho\hat{x}_{n-1} = g_\mathcal{Q}(i_n)$, together yield (15). Thus we have an approximate delayed decoder implementable as a codebook-based correction $c(\{i_l\}_n^{n+L})$, to $\rho\hat{x}_{n-1}$.

5. RESULTS

We compare the proposed optimal and codebook-based decoders, at $L=1$ and 3, with IDPCM ($L=1$ necessarily), and SDPCM at $L=3$ (which outperforms SDPCM at $L=1$ or 2). Experiments include cases where $\{z_n\}$ are drawn from a zero-mean Gaussian, or from a zero-mean Laplace distribution. Note that the transform coefficients of motion compensated, prediction residual blocks in video coding, for instance, are modeled well by a Laplace distribution. All the competing schemes use the same DPCM encoder, with the one-step predictor (2). The quantizer $\mathcal{Q}$ is a symmetric uniform threshold quantizer (UTQ), suitably scaled to vary the bit-rate, estimated as the first order entropy of the output indices. At every rate, and for every choice of innovation density $p_Z(z)$, while the thresholds of the quantizer are fixed, the reconstructions $g_\mathcal{Q}(\cdot)$ are optimized recursively as in [1], with $p_E(e)$, used in the codebook approach, obtained as a by-product. We emphasize that the proposed decoding schemes are independent of the choice of $\mathcal{Q}$ and that the UTQ is only a representative case. Each point on the graphs (Figs. 2 - 5) has been obtained as an average of 20 trials, with a random sequence of 2000 samples each. Results are presented in terms of SNR gain at the same rate, over the regular DPCM decoder.

The results demonstrate that both proposed decoders, substantially outperform SDPCM, and IDPCM, with the the latter schemes not always guaranteed to perform better than regular DPCM. In the case of the processes considered, the performance of the codebook decoder is very close to that of the optimal approach. Fig. 3 magnifies the boxed region in Fig. 2 to show the performance gap between the two methods. Given the observed quality of this approximation, the curves for the optimal approach have been omitted in Figs. 4 and 5, to avoid clutter. Note that even with just a 1 sample delay the proposed approaches provide better performance than SDPCM at a higher delay ($L=3$). At low bit-rates, as L increases, the gain over regular DPCM due to both proposed schemes increases, i.e., lower the bit-rate, higher the information in $\{i_l\}_{n+1}^{n+L}$, about x_n. At high bit-rates, there is diminishing return from increasing L. The poor performance of IDPCM and SDPCM at high rate is attributed to the observation in Sec. 1. As the rate increases, the interval $\mathcal{I}_n$ shrinks, resulting in higher likelihood of the smoothed estimates, $\hat{x}_n^{sdpcm}$ and $\hat{x}_n^{idpcm}$, falling outside it. The optimal, and codebook-based delayed decoders, by design, account for this interval information. The results for the Gaussian innovations at different values of ρ indicate, as expected, that higher correlation offers more to be gained by increasing L.

Even at $L = 0$ (i.e., *no delay*), one would expect the optimal estimate in (7) to improve over regular DPCM. For the process and quantizers considered, such gains were observed to be insignificant.

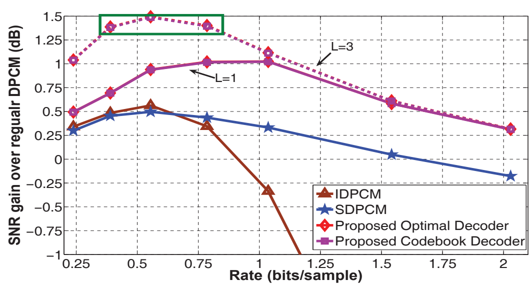


Fig. 2. Performance comparison of different delayed decoders for a Gaussian autoregressive process with $\rho = 0.95$.

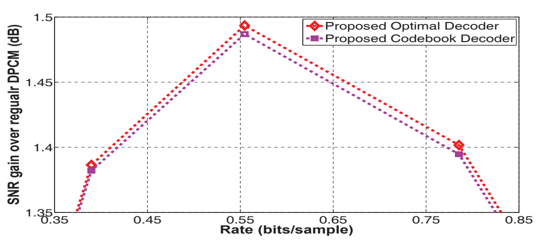


Fig. 3. Magnification of the boxed region in Fig. 2, showing the performance gap between the proposed optimal delayed decoder and its codebook-based approximation.

6. GENERALIZATION TO HIGHER ORDER PROCESSES

A K^{th} order autoregressive process of random scalars in $\mathbf{R}$, can be equivalently viewed as a first order process of random vectors in $\mathbf{R}^K$, by a formulation similar to (4), albeit with a different Φ. The current structures of the recursions in Sec. 3, the optimal, and codebook-based delayed decoders still hold, but the intervals $\mathcal{I}_n$ are replaced by corresponding K-dimensional segments $\underline{\mathcal{I}}_n$, and integrals, auxiliary functions in the algorithms, and densities are all defined in the vector space $\mathbf{R}^K$. Difficulties arise in the implementation of the K-dimensional convolutions in (9) and (11). But note that the codebook-based delayed decoder still retains its low complexity structure (albeit with increased storage), with performance likely close to that of the optimal decoder.

7. CONCLUSIONS

An optimal delayed decoding algorithm for predictively encoded autoregressive sources, based on an ET approach that recursively calculates the conditional density of the current sample given all information known to the decoder for a certain look-ahead, is proposed. Irrespective of the bit-rate, or innovation probability density, the algorithm ensures optimal reconstruction. The optimal delayed decoder in turn motivates a codebook-based approach, which has performance almost indistinguishable from that of the former, with the obvious advantage of low implementation complexity. Simulations demonstrate the considerable performance gains of either approach, over existing smoothing post-filters. Future directions include encoder modifications in light of the proposed algorithms, and implementations within video and speech coding schemes.

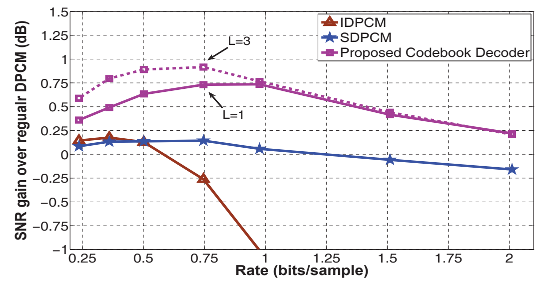


Fig. 4. Performance comparison of different delayed decoders for a Gaussian autoregressive process with $\rho = 0.8$.

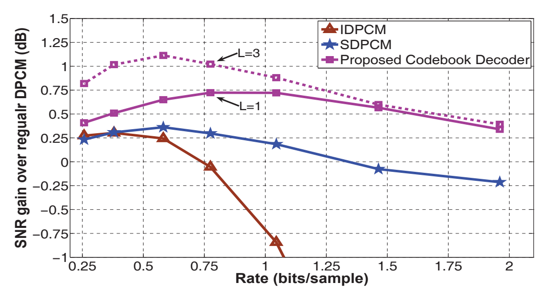


Fig. 5. Performance comparison of different delayed decoders for an autoregressive process with Laplacian innovations, and $\rho = 0.95$.

8. REFERENCES

[1] N. Farvardin and J.W. Modestino, "Rate-distortion performance of DPCM schemes for autoregressive sources," *IEEE. Trans. Info. Theory*, vol. 31, pp. 402–418, May 1985.

[2] A. Hayashi, "Differential pulse code modulation of stationary gaussian inputs," *IEEE. Trans. Comm.*, vol. 26, pp. 1137–1147, Aug 1978.

[3] M.L. Sethia and J.B. Anderson, "Interpolative DPCM," *IEEE. Trans. Comm.*, vol. 32, pp. 729–736, June 1984.

[4] W-W. Chang and J.D. Gibson, "Smoothed DPCM codes," *IEEE. Trans. Comm.*, vol. 39, pp. 1351–1359, September 1991.

[5] K. Rose and S.L. Regunathan, "Toward optimality in scalable predictive coding," *IEEE. Trans. Image Proc.*, vol. 10, pp. 965–976, July 2001.

[6] J.M. Mendel, *Lessons in estimation theory for signal processing, communications, and control*, Prentice Hall PTR, Upper Saddle River, NJ, 1995.