

TRANSFORM CODING USING ADAPTIVE BASES AND QUANTIZATION

Vivek K Goyal, Jun Zhuang, Martin Vetterli, and Christopher Chan

Dept. of Electrical Engineering and Computer Science
University of California, Berkeley

{vkgoyal,zhuang,martin}@eecs.berkeley.edu, wychan@engr.sgi.com

ABSTRACT

This paper considers the problem of universal transform coding based on estimating the Karhunen-Loève transform from quantized data. The use of quantized data in the estimation allows the encoder and decoder to maintain the same state without any side information. A theorem is presented that proves, under certain conditions, that consistent estimation of all the required moments is possible from uniformly scalar quantized data regardless of the quantization coarseness. This consistent estimation requires the solution of nonlinear equations. Very simple approximations that avoid these nonlinear equations are used to develop a practical adaptive coding technique. Promising experimental results obtained with this method are presented.

1. INTRODUCTION

Although there are many popular and effective adaptive lossless compression methods, such as Lempel-Ziv coding [1], dynamic Huffman coding [2], and arithmetic coding [3], there are no adaptive lossy compression methods with such universal acceptance. A primary reason for this disparity is computational complexity: While asymptotically optimal universal lossy compression schemes have been presented in the literature, these schemes suffer from high complexity because they depend on high dimensional vector quantization. One way to limit complexity, as is pursued here, is to limit all quantization to scalars or vectors of fixed, relatively small, dimension.

This paper approaches the problem of universal lossy coding by considering various adaptive schemes that require no offline training. A basic requirement for an encoder to achieve universality is for it to “learn” the statistics of the unknown or time-varying source. The standard way to make a coding system adaptive is to, given a block of data, use the data (and perhaps prior blocks) to develop a model of the source and design an encoder optimal for the source prior to coding the data. In order for the receiver to correctly decode the data, it must be informed

V. K. Goyal and M. Vetterli are also with the Laboratoire de Communications Audiovisuelles, Département d'Électricité, École Polytechnique Fédérale de Lausanne, Switzerland.

C. Chan is now with Silicon Graphics, Inc., Mountain View, California.

This work was supported by the US Dept. of Defense through an NDSEG fellowship and the US and Swiss National Science Foundations.

of the adaptation of the encoder; hence the parameters of the encoder are sent along with the coded data. The main sources of performance degradation compared to an optimal “omniscient” source coder are modelling error and the cost of sending encoder parameters to the receiver. This work has a fundamentally different spirit. It asks, “How well can one do *without* sending any encoder parameters to the receiver?” For the receiver to be able to correctly decode the data without being explicitly informed of the encoder state, it is necessary that all encoder adaptation depend only on information already available at the decoder at no additional cost, *i.e.* the *coded* data stream. In the parlance of communication, we are avoiding *side information*, or using *backward adaptation*. With this strategy, one avoids the price of side information but the efficiency of estimating source statistics is reduced. Also, there is some performance penalty from the requirement that adaptation be *strictly* causal.

A method for adaptive scalar quantization without side information ([4],[5]) is reviewed and generalization to vector quantization (VQ) is discussed. For adaptive VQ, a scheme using some side information is also presented [6]. The bulk of the paper addresses the issue of adaptive transform coding. It is shown that the transform (and, as above, the scalar quantizers) can be adapted based purely on past quantized data to give an optimal universal transform coding system. Simpler suboptimal systems (which still require no side information) and experimental results are also presented.

2. ADAPTIVE SCALAR QUANTIZATION WITH NO SIDE INFORMATION

A straightforward way of improving a transform coding system is to improve the performance of its scalar quantization. One systematic way to achieve adaptivity is to assume a parametric signal model, estimate the model parameters, and design the quantizer accordingly. In light of the desire for backward adaptivity, we would like to know how coarsely the data can be quantized and still allow model parameters to be determined. The answer was described first in [5] and is summarized below.

Let X have a parametric p.d.f. $f_{\theta_1, \theta_2, \dots, \theta_k}(x)$. Let $-\infty = b_0 < b_1 < \dots < b_{n+1} = \infty$ and denote the probability of X assuming a value in the i th bin $[b_i, b_{i+1})$ by $F_i = \int_{b_i}^{b_{i+1}} f_{\theta_1, \theta_2, \dots, \theta_k}(x) dx$, $i = 0, \dots, n$. Neglecting degeneracy conditions, if $n \geq k$, then the parameters $\theta_1, \theta_2, \dots, \theta_k$ can

be determined from the bin probabilities F_1, F_2, \dots, F_n .¹ If the bin probabilities are only approximately known or if the parametric p.d.f. model is not exact, the parameters can be estimated using an appropriate fitting criterion.

In general, this approach will lead to a nonlinear system of equations relating the parameters and bin probabilities. Ortega and Vetterli [4] proposed using a continuous, piecewise linear model for the signal p.d.f. This method is particularly practical because it yields a tridiagonal linear system relating the parameters and bin probabilities. In addition, the performance is quite good even when the source p.d.f. can not be closely approximated by a piecewise linear function [5].

3. ADAPTIVE VECTOR QUANTIZATION

3.1. Adaptive VQ with no side information

One approach to backward adaptive VQ is to simply extend the ideas of the previous section. It is interesting to note that even very coarse scalar quantization can yield enough information to fit a reasonable parametric model. For example, quantizing with only three bins will yield $3^k - 1$ independent probability estimates, where k is the vector dimension. For any $k \in \mathbb{Z}^+$, $3^k - 1 \geq \frac{1}{2}k^2 + \frac{3}{2}k$, so this quantization is fine enough to fit a multivariate Gaussian signal model.

This approach has not been pursued because of computational complexity. The complexity of parameter estimation depends on the parametric model itself, which is not necessarily prohibitively complicated in the vector case. The difficulty stems from the fact that even given a parametric source model, the design of good vector quantizers is very complicated. One lower complexity alternative to adaptive VQ is presented in the following subsection.

3.2. Dictionary-based Adaptive VQ

Given the difficulty in designing vector quantizers based on parametric source models, iterative approaches have normally been taken for this purpose. As an alternative, [6] presented a one-pass technique for VQ codebook construction and adaptation for individual source signals. The technique, called rate-distortion Lempel-Ziv (RDLZ), is an effort to extend the Lempel-Ziv algorithm to a practical universal lossy compression algorithm. A source signal is quantized using a VQ codebook, which in turn is modified during the quantization process. Adaptation is achieved without iteration on any particular segment of the source signal, but by accumulating source statistics as encoding proceeds. Side information must be transmitted since codebook adaptation, such as adding or moving code vectors, cannot be inferred otherwise. Determination of the amount of side information transmitted is based on a rate-distortion criterion. Fig. 1 shows a portion of a composite image coded with RDLZ (from top to bottom). It is apparent that after the transition from Barbara to text, the codebook adapts to become more appropriate for text.

¹Since $\sum_{i=0}^n F_i = 1$, only n of the $n+1$ F_i 's are independent.

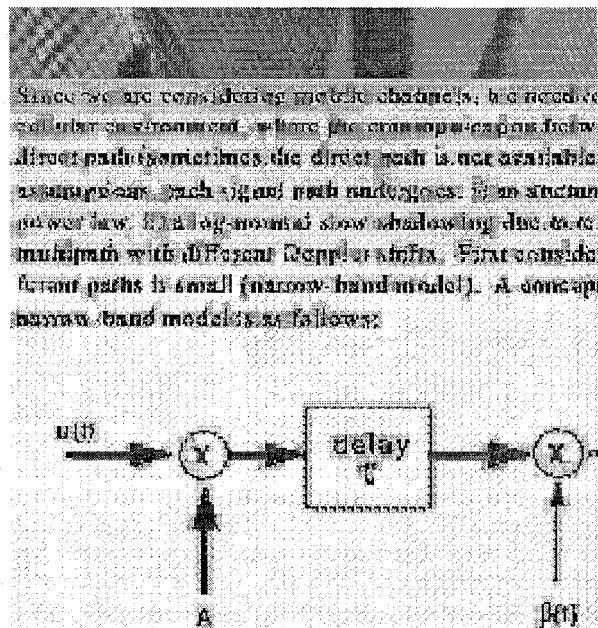


Figure 1: RDLZ coding of a composite image.

4. ADAPTIVE TRANSFORM CODING WITH NO SIDE INFORMATION

4.1. Exposition

The study of adaptive transform coding has been a rich research area for several years. Recently, very good transform coding results have been reported using classification based methods, *i.e.* schemes in which the signal space is divided into a finite set of classes and a fixed transform is designed for each class [7]. Classification methods generally rely on training for defining classes and/or designing a transform code for each class. In this paper we consider only methods that do not rely on *a priori* training.

The optimality of the Karhunen-Loève Transform (KLT) for transform coding of a stationary source under high resolution and arbitrary real bit assignment assumptions is well known [8, §8.6]. Adaptive versions of KLT coding are not widely used because in traditional adaptive systems (with side information) transmission of the KLT coefficients can be prohibitively expensive.² By advocating a backward adaptive KLT, we completely avoid performance penalties associated with side information. Instead, there is a potential for performance degradation due to noise in estimating signal statistics.

The following theorem establishes the feasibility of backward adaptive KLT coding for a Gaussian source.

Theorem 1: Let $X = [X_1, \dots, X_k]^T$, $X \sim N(0, \Sigma)$, where Σ is an unknown matrix with positive diagonal. Let \hat{X} be a scalar quantized version of X such that for $n \in \mathbb{Z}$ either

- (i) $X_i \in [n\Delta_i, (n+1)\Delta_i) \Rightarrow \hat{X}_i = (n + \frac{1}{2})\Delta_i$; or
- (ii) $X_i \in [(n - \frac{1}{2})\Delta_i, (n + \frac{1}{2})\Delta_i) \Rightarrow \hat{X}_i = n\Delta_i$.

²In practical adaptive transform coding system, 20 to 40 percent of the available bit rate is assigned to side information [9, §2.3].

Then for any set of positive, finite quantization step sizes $\Delta_1, \dots, \Delta_k$, all moments of X can be recovered exactly from the first and second order moments of \hat{X} .

Proof: Since the quantization of X is in an infinite number of bins, we could proceed as suggested by §2. Instead, we follow a more explicit strategy which yields relationships of independent interest. For simplicity, the proof below considers only case (i); case (ii) is similar. First note that since X is a Gaussian random variable, all its moments can be expressed in terms of its first and second order moments. Having already specified that X has mean zero, it is completely characterized by $\sigma_{ij} = E[X_i X_j]$, $1 \leq i, j \leq k$. (Similarly denote $E[\hat{X}_i \hat{X}_j]$ by $\hat{\sigma}_{ij}$ and σ_{ii} by σ_i^2 .)

Simplifying expressions from [10] relating the moments of a signal and its uniformly quantized version gives

$$\hat{\sigma}_i^2 = \sigma_i^2 + \Delta_i^2 \left[\frac{1}{12} + \sum_{m=1}^{\infty} e^{-2m^2 \pi^2 \sigma_i^2 / \Delta_i^2} \left(\frac{1}{m^2 \pi^2} + \frac{4\sigma_i^2}{\Delta_i^2} \right) \right] \quad (1)$$

and, for $i \neq j$,

$$\hat{\sigma}_{ij} = (1 + \delta_{ij})\sigma_{ij} + \mu_{ij}, \quad (2)$$

where

$$\delta_{ij} = 2 \left(\sum_{m_1=1}^{\infty} e^{-2m_1^2 \pi^2 \sigma_i^2 / \Delta_i^2} + \sum_{m_2=1}^{\infty} e^{-2m_2^2 \pi^2 \sigma_j^2 / \Delta_j^2} \right)$$

$$\mu_{ij} = \sum_{m_1, m_2=1}^{\infty} \frac{\Delta_i \Delta_j}{m_1 m_2 \pi^2} e^{-2\pi^2 \left(\frac{m_1^2 \sigma_i^2}{\Delta_i^2} + \frac{m_2^2 \sigma_j^2}{\Delta_j^2} \right)} \sinh \left(\frac{4\pi^2 \sigma_{ij} m_1 m_2}{\Delta_i \Delta_j} \right)$$

For any positive Δ_i , (1) describes a monotonic relationship between σ_i^2 and $\hat{\sigma}_i^2$, so each σ_i^2 can be determined from $\hat{\sigma}_i^2$. With the σ_i 's thusly determined, (2) describes a monotonic relationship between σ_{ij} and $\hat{\sigma}_{ij}$, so the σ_{ij} 's can similarly be determined. \square

Before discussing the application of Theorem 1 to KLT coding, it is worthwhile to comment on (1). It shows, reassuringly, that as $\Delta_i \rightarrow 0$, $\hat{\sigma}_i^2 \rightarrow \sigma_i^2$. Also, as $\Delta_i/\sigma_i \rightarrow \infty$, $\partial[\hat{\sigma}_i^2]/\partial[\sigma_i^2] \rightarrow 0$. This reflects the intuitive idea that recovering σ_i^2 in a numerically stable manner is difficult when Δ_i/σ_i is very large, *i.e.* the quantization is extremely coarse. Similar conclusions can be drawn from (2).

Applying Theorem 1 to recover the moments of the unquantized signal involves finding the roots of (1) and (2). This is clearly very complicated, but may be simplified by the fact that, as long as the Δ/σ 's are not too large, the sums can be approximated by a few terms. Furthermore, for a reasonable range of quantizer step sizes, the relative error incurred by replacing (1) and (2) by very simple approximations is small. Consider replacing the bracketed term in (1) by $\frac{1}{12}$ and replacing (2) by $\hat{\sigma}_{ij} \approx \sigma_{ij}$. The relative errors in making these approximations are shown in Fig. 2. It is assumed that $\frac{\Delta_i}{\sigma_i} = \frac{\Delta_j}{\sigma_j}$ and $\sigma_{ij} = \sigma_i \sigma_j$.³

The experimental results presented in the final subsection are for a system that does not use the corrections suggested in the proof of Theorem 1. This is partially justified

³This value of σ_{ij} maximizes $\hat{\sigma}_{ij}$. Naturally $\sigma_{ij} = 0$ would maximize the relative error.

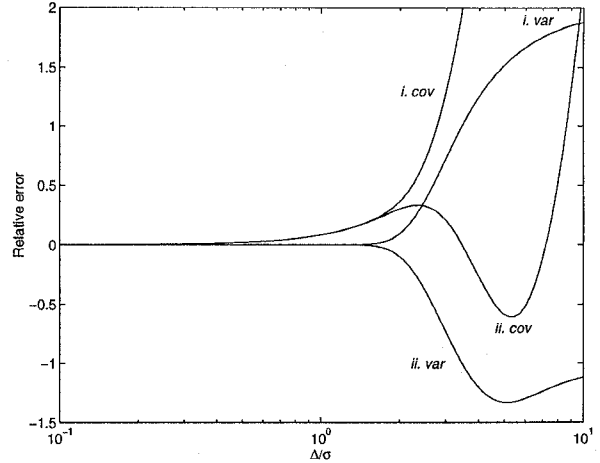


Figure 2: Relative errors in variance and covariance estimates induced by replacing (1) and (2) by simple approximations. (*i.* and *ii.* refer to the quantizer types in Theorem 1.)

by Fig. 2, but furthermore by the fact that finding a good estimate of the KLT for a source does not necessarily require a good estimate of its covariance matrix.⁴ The following theorem provides a situation where the KLT estimated directly from quantized data will be a consistent estimate of the KLT of the unquantized source. It presumes that the KLT of the source is known and can be viewed as a weak local stability result for the backward adaptive KLT coding described in the final subsection.

Theorem 2: Let $X = [X_1, \dots, X_k]^T$ be Gaussian and have KLT T , *i.e.* T is an orthonormal matrix such that $TR_X T^T = \Lambda$, where $R_X = E[XX^T]$ and Λ is a diagonal matrix with non-increasing diagonal elements. Let $\hat{X} = T^T q(TX)$, where q is a scalar quantization function that induces zero-mean quantization noise. Then, regardless of the quantizer resolution, $TE[\hat{X}\hat{X}^T]T^T$ is a diagonal matrix, so the KLT of \hat{X} differs from T by at most a permutation. **Proof:** Let $Y = TX$ and $e = q(Y) - Y$. Note that since Y is Gaussian and has uncorrelated components, the components are independent. Since the quantization is scalar, we have furthermore that e_i is independent of e_j and Y_j for $i \neq j$. Then

$$TE[\hat{X}\hat{X}^T]T^T = TE[T^T q(Y)q(Y)^T T]T^T$$

$$= E[q(Y)q(Y)^T] = E[YY^T] + E[eY^T + Ye^T + ee^T]$$

In the final expression, the first expectation gives Λ , while the three remaining expectations yield diagonal matrices because of the zero-mean and independence conditions. \square

4.2. Experimental results

Consider the following method for transform coding of a scalar source. The source samples are formed into vectors of size $k = 8$. These vectors are transformed by left multiplication by a unitary matrix T and then uniformly scalar quantized with stepsize Δ . Assume the quantized values will

⁴The KLT can be determined from the covariance matrix, but the converse is not true.

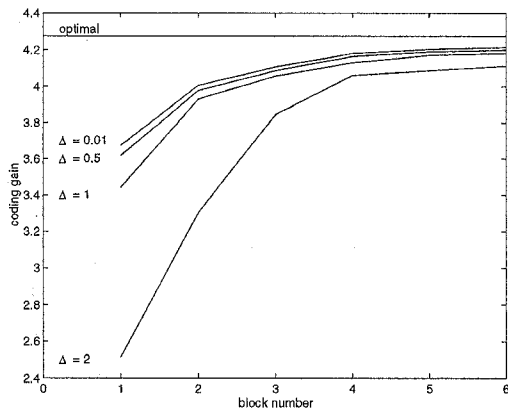


Figure 3: Coding gain of adaptive system compared to optimal coding gain for various coarseness levels.

be coded with an adaptive entropy coder.⁵ At the decoder, and at the encoder for adaptation purposes, reconstruction is performed by left multiplication by T^T . After every block of N vectors, the transform T is updated to be the empirical KLT of the last m blocks of *quantized* data, where m is sometimes called the *memory factor* and $m = \infty$ is used to indicate that all past data is used in the autocorrelation estimation. An unbiased autocorrelation estimate is used in the KLT calculation.

This backward adaptive KLT coding scheme has been tested on zero-mean, unit variance AR(1) sources. The results of the first experiment, given in Fig. 3, show that after a suitable amount of data has become available, close to optimal coding gain is achieved. The correlation coefficient of the source is $a = 0.9$, thus the optimal coding gain is ≈ 4.28 . An initial transform of $T = I$, a block length of $N = 2$, a memory factor of $m = \infty$, and quantization stepsizes $\Delta = 0.01, 0.5, 1$, and 2 were used. This experiment shows the relative importance of the number of samples available and the coarseness of the data. A system that uses unquantized data in adaptation (thus requiring side information) would have coding gain approximately as shown by the $\Delta = 0.01$ curve; the relatively small difference between this curve and the remaining curves shows that the dependence on the coarseness is quite small.

A principle disadvantage of static coding algorithms is performance loss due to a mismatch between a source and the source assumed in the design. The second experiment demonstrates the advantage of using this adaptive method in the case of mismatch. The source has correlation coefficient $a = 0.99$ and is coded using an initial transform of $T = I$, a block length of $N = 50$, and a memory factor of $m = 2$. In Fig. 4 performance is compared to the performances of KLT coders designed for correlation coefficients of $a = -0.9, 0, 0.5, 0.9$, and the true value of 0.99 . For the adaptive coder, the rates and distortions given are ave-

⁵ Assuming an equally weighted distortion measure and identical normalized component p.d.f.'s, using equal component quantizer stepsizes and no overload is essentially the same as optimal bit allocation. Also note that this is constant distortion operation; constant rate operation could be achieved by varying Δ according to a rate measure.

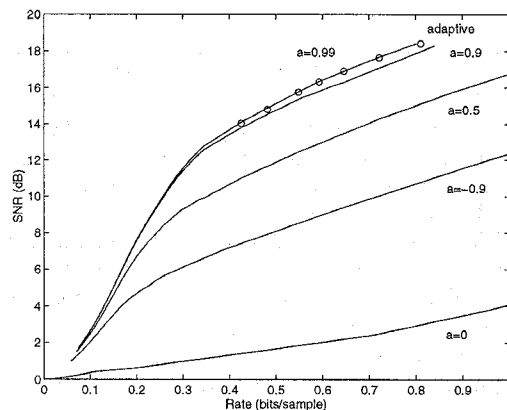


Figure 4: Performance comparison between adaptive system and fixed coders. Source has correlation $a = 0.99$ and fixed coders are designed for various values of a , including the true value. \circ 's mark performance of the adaptive system.

rages over blocks 2 through 26. As the mismatch becomes larger, the performance gain from the adaptive system also becomes larger. The loss in performance due to adaptation compared to a properly matched static design is minimal.

5. REFERENCES

- [1] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Trans. Info. Theory*, 23:337–343, May 1977.
- [2] R. S. Gallager. Variations on a theme by Huffman. *IEEE Trans. Info. Theory*, 24:668–674, Nov. 1978.
- [3] G. G. Langdon and J. Rissanen. Compression of black-white images with arithmetic coding. *IEEE Trans. Comm.*, 29:858–867, June 1981.
- [4] A. Ortega and M. Vetterli. Adaptive quantization without side information. In *Proc. IEEE Int. Conf. Image Proc.*, vol. 3, pp. 856–860, Austin, Nov. 1994.
- [5] B. Yu. A statistical analysis of adaptive scalar quantization based on quantized past. Submitted to *IEEE Trans. Info. Theory*, 1995.
- [6] C. Chan and M. Vetterli. Lossy compression of individual signals based on string matching and one pass codebook design. In *Proc. ICASSP*, vol. 4, pp. 2491–2494, Detroit, May 1995.
- [7] M. Effros and P. A. Chou. Weighted universal transform coding: Universal image compression with the Karhunen-Loève transform. In *Proc. IEEE Int. Conf. Image Proc.*, vol. II, pp. 61–64, 1995.
- [8] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Acad. Pub., Boston, 1992.
- [9] H. S. Malvar. *Signal Processing with Lapped Transforms*. Artech House, Norwood, MA, 1992.
- [10] L. Cheded and P. A. Payne. The exact impact of amplitude quantization on multi-dimensional, high-order moments estimation. *Signal Proc.*, 39(3):293–315, Sep. 1994.