In the Proceedings of Data Compression Conference 2000

Rate Distortion Behavior of Threshold-Based Nonlinear Approximations

Claudio Weidmann and Martin Vetterli*
Dept. of Electrical Engineering
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
{weidmann,vetterli}@lcav.de.epfl.ch

Abstract

Many transform coders use a type of nonlinear approximation that selects all coefficients with magnitudes above a threshold, encodes their positions, and then quantizes their values. This kind of two-stage scheme is very effective at low rates. We derive an upper bound on the operational rate distortion function of such threshold-based nonlinear approximations. The bound is applied to the spike process, which is a generic model for sparse transform coefficients, and to a Gaussian mixture process that is a better model for wavelet coefficients in image transform codes. The results exhibit the same change in distortion decay between low and high rates which is typical for such image coders.

1 Introduction

The term "nonlinear approximation" obviously has many different meanings. We give a definition which is significant in the context of signal approximation in orthonormal bases and, on the practical side, in transform coding. Threshold-based, quantized nonlinear approximation is a two-stage process: in the first pass, the positions of the coefficients above threshold are encoded. After that, the values of these so-called *significant* coefficients are quantized and encoded. The main contribution of this paper is an upper bound on the operational distortion rate that is achievable with such two part schemes. In conjunction with computed $D(R)$ functions this bound can be used to study the performance of these algorithms for a given coefficient model. It is trivially also an upper bound on the true distortion rate function.

After the necessary definitions and an example, we derive the bound in Section 3. In the Sections thereafter, we use the bound to study two models of transform coefficients. A short discussion ends the paper.

*M. Vetterli is also with the Department of EECS, UC Berkeley, Berkeley CA 94720

2 Definitions

Definition 1 (Threshold-based Nonlinear Approximation (NLA)) *Given a threshold T, a vector of i.i.d. real random variables $X_1^N = [X_1, \ldots, X_N]$ is approximated by the binary position vector U_1^N with*

$$U_i = 1_{|X_i| \geq T} \tag{1}$$

and a value vector V_1^N with

$$V_i = U_i X_i \tag{2}$$

This process is sometimes called hard thresholding in the denoising literature. As alternative to (2), we could define a variable length value vector of length $K = \sum_{i=1}^{N} U_i$ that contains only the nonzero components of V.

Definition 2 (Quantized Nonlinear Approximation (QNLA)) *Same definition as above, but with quantized values $\hat{V}_i = U_i \hat{X}_i$.*

In this paper we will restrict our attention to i.i.d. real random variables with symmetric probability density functions $f(x) = f(-x)$, though most results can be extended to non-symmetric pdf's. The density of the thresholded random variable is:

$$\bar{f}^{(T)}(v) := \begin{cases} 0, & |v| < T \\ \frac{1}{\mu(T)} f(v), & |v| \geq T \end{cases} \tag{3}$$

where μ is the probability of a non-zero value:

$$\mu(T) := \mathrm{E}[U] = \Pr\{|X| \geq T\} = 2 \int_T^\infty f(x)\, dx \tag{4}$$

If V (respectively $\hat{V}$) is defined as in (2), the position vector U is actually redundant. But in many practical QNLA schemes, a two part code is used: first, the position information U_i is sent. Only if it equals one, a code for $\hat{V}_i \neq 0$ is appended (scalar quantization). This can be extended to vectors. It can be easily shown that the two part code with scalar quantization has the same asymptotic rate as directly coding $\hat{V}$ [9]. That is, NLA followed by scalar quantization (the two part code) is equivalent to simple scalar quantization with a prescribed zero bin $(-T, T)$ (aka dead zone quantizer), i.e. there is only an implementation difference.

Example We consider two and four level Lloyd-Max quantization of thresholded Gaussian random variables. Figure 3 (at end) shows the parametric distortion rate curves achieved with this method, together with the Gaussian $D(R)$ as a lower bound and the Gish-Pierce high rate asymptote for uniform quantization: $D(R - 0.255\,\text{bits})$. The two curves start at the respective (R, D) points for Lloyd-Max quantization $(T = 0)$ and end in $(R, D) = (0, 1)$ for $T \to \infty$. They show that by varying the threshold one can achieve a whole set of (R, D) points which are actually below simple time sharing of the original Lloyd-Max solutions. ~~It is remarkable that NLA with four level quantization at a rate of 2.1 bits per sample is only 0.8 dB above the lower bound.~~ (Simulation error; at $R = 2$ bits curve is on Gish-Pierce asymptote.)

3 Upper Bound on NLA Distortion Rate

Nonlinear approximation as considered in this paper is a two-stage process, consisting of thresholding followed by quantization of the samples that are above the threshold. Ideally one would use a variable-dimension vector quantizer, since the number of samples varies from block to block (assuming data is processed in blocks). In practical schemes, scalar quantization is often preferred thanks to its simplicity and speed. And at low rates it performs quite well, as illustrated by the example in Section 2.

In general we can never achieve a rate distortion optimal encoding, if we use two-stage quantized nonlinear approximation to lossily encode a block of i.i.d. random variables. Except in "constructed" cases, hard thresholding yields a suboptimal encoder even if it is followed by an optimal quantizer. In this Section we present an upper bound on distortion rate as a tool to study the low rate performance of threshold-based quantized nonlinear approximations.

An optimal two-stage encoder is composed of an ideal lossless encoder for the position variable U, indicating the samples above threshold, and a R-D optimal encoder for the values V of these samples. Therefore we first characterize the R-D behavior of thresholded random variables.

$D(R)$ of a single thresholded random variable V can be upper bounded with the distortion rate function of a Gaussian with the same variance. The unnormalized variance is

$$A(T) := 2\int_T^\infty f(x)x^2\,dx \tag{5}$$

where $f(x)$ is the pdf of the non-thresholded rv and $A(0) = \sigma^2$ its variance. After normalization with (4) the upper bound becomes

$$D_V(R_V) \le \mathrm{E}_{\bar{f}}V^2 2^{-2R_V} = \frac{A(T)}{\mu(T)}2^{-2R_V} \tag{6}$$

For each sample we need $H(U) = h(\mu(T))$ bits to indicate whether it is *significant*, i.e. above threshold (h is the binary entropy function). Given the total rate per sample R, the rate R_V available to code each significant sample is

$$R_V(T,R) = \frac{R - h(\mu(T))}{\mu(T)}. \tag{7}$$

Taking into account the distortion resulting from the uncoded sub-threshold samples, we can write the following bound on the distortion rate function of the original rv X:

$$D(R) \le \mu(T)D_V(R_V(T,R)) + (1-\mu(T))\mathrm{E}[X^2 \big| |X| < T]$$

or

$$D(R) \le B(T,R) := A(T)\exp\left(-2\ln(2)\frac{R - h(\mu(T))}{\mu(T)}\right) + \sigma^2 - A(T). \tag{8}$$

Direct minimization of $B(t,R)$ with respect to t is very difficult. Instead we sweep the threshold t from 0 to ∞ and compute candidate points of the convex hull of the resulting bounds $D(R) \leq B(t,R)$ in the (R,D) plane. To further simplify the task, we assume that two bounds $B(t,r)$ and $B(t+\Delta t, r)$ have a common tangent (equivalently: they intersect each other). This tangent can then be used to find candidate convex hull points. We get the parametric upper bound outlined below.

Proposition 1 *The distortion rate function of threshold-based nonlinear approximation of a memoryless continuous random variable X with symmetric pdf $f(x)$ and variance σ^2 is upper bounded by*

$$D(\bar{R}(t)) \leq A(t)\left[\exp\left(-2\ln(2)\frac{\bar{R}(t)-h(\mu(t))}{\mu(t)}\right)-1\right]+\sigma^2, \quad \forall\, t \geq 0 : \exists \bar{R}(t) \tag{9}$$

where the rate $\bar{R}(t)$ is given by (primes denote derivatives)

$$\bar{R}(t) = \frac{\mu(t)}{2\ln(2)}\left[\frac{2h(\mu(t))}{\mu(t)} - 2h'(\mu(t)) - \rho(t) - \mathrm{W}_{-1}\left(-\rho(t)e^{-2h'(\mu(t))-\rho(t)}\right)\right] \tag{10}$$

with

$$\rho(t) := \frac{\mu(t)}{A(t)}t^2 \tag{11}$$

The expression for $\bar{R}(t)$ involves Lambert's W *function, which solves $\mathrm{W}(x)e^{\mathrm{W}(x)} = x$. The subscript -1 indicates the second real branch of* W*, taking values on $[-1,-\infty[$. A real-valued solution for $\bar{R}$ exists only if the argument of the* W *function is larger than $-1/e$ (see remarks).*

Sketch of proof: We take two curves of the family (8), say $B(t,r_0)$ and $B(t+\Delta t, r_1)$, and determine their common tangent by solving the following system of equations:

$$\frac{\partial}{\partial r}(B)(t,r_0) = \frac{\partial}{\partial r}(B)(t+\Delta t, r_1) := s \tag{12}$$

$$\frac{B(t,r_0)-B(t+\Delta t,r_1)}{r_0-r_1} = s \tag{13}$$

We solve (12) for r_1, plug this solution into (13) and solve for r_0. Taking the limit for $\Delta t \to 0$ leads to (10). This solution is purely symbolic in the sense that for a specific pdf $f(x)$ and threshold t no common tangent might exist. That case is explained next.

Remarks As pointed out above, for some values of the threshold t there might be no solution to the common tangent problem. This is the case when increasing t produces a bound $B(t+\Delta t, r)$ that for all r is larger than $B(t,r)$. It can be shown that there is no common tangent for all t satisfying

$$\min_r \frac{\partial}{\partial t}(B)(t,r) \geq 0. \tag{14}$$

Computing the above minimum yields a reassuring result: (14) is true exactly iff the argument of the W function in (10) is less than or equal to $-1/e$, i.e. when there is no real-valued solution. But this condition alone does not guarantee that we find only convex hull points, since one of the constituent bounds $B(t,r)$ might be below all others for a threshold t satisfying (14). In all examples we have studied, it happened to be the trivial bound $B(0,r)$, if at all. For Gaussian data, obviously no upper bound can be better than $B(0,r)$, while for Laplacian all thresholds above a critical value yield slight improvements. For pdf's that are even more peaked around zero, such as those studied in the next two Sections, the critical threshold is (almost) zero. Informally, these "forbidden" threshold values mean that the reduction in number of coded significant samples is not sufficient to offset the increased position rate $h(\mu(t))$. From this reasoning it becomes evident that such t can only lie between 0 and $t_{0.5}$, with $\mu(t_{0.5}) = 0.5$. On the other hand, this means that the bound will be useful at low rates (i.e. for high thresholds), a rate region for which few bounds are known.

4 Spike Processes

Nonlinear approximation plays a key role in the design and understanding of transform coding methods [4]. While it is difficult to come up with a realistic statistical model for e.g. wavelet coefficients in an image coder, transforms that are popular in compression generally concentrate a large part of the signal energy in relatively few coefficients. This is the case for wavelet transforms of piecewise regular signals, since there will be only a small number of non-zero coefficients, mostly around the signal singularities [4]. But even transforms of the Fourier family show a similar behavior if they operate on small blocks, like the JPEG image coding standard. Another example are the coefficients in LPC speech coding. A strongly simplified model would be a transform vector that is zero everywhere except in a few positions, where large "spikes" stick out. In [8] we presented a deterministic spike model: one assumes to know a priori that in a vector of length N there will be exactly K nonzero entries. In the present paper we introduce a probabilistic definition that generalizes this concept.

Definition 3 (Bernoulli-Gaussian (BG) spike process) *An i.i.d. Bernoulli-Gaussian spike source (process) emits a memoryless random variable X that is the product of a Bernoulli rv with* $\Pr\{1\} = p$, $\Pr\{0\} = 1 - p$ *with a zero mean Gaussian rv. Using the $\delta(\cdot)$ distribution, its pdf can be written as*

$$f(x) = (1-p)\delta(x) + p\frac{1}{\sqrt{2\pi}\sigma}e^{-x^2/2\sigma^2}. \tag{15}$$

This pdf can also be seen as a mixture of two zero mean Gaussian random variables, with one of them having zero variance (a special case of the model studied in Section 5). The bound (9) can be easily evaluated if one replaces T by $T+\epsilon$ in the lower integration boundary of (4), with an arbitrarily small number $\epsilon > 0$. By doing this we exclude the Dirac $(1-p)\delta(x)$ from the integral, and hence we have $\mu(T) \leq p \quad \forall T \geq 0$. This is obviously correct, since we never have to code a spike of zero amplitude.

Figure 1(a) shows the bound and the empirical $D(R)$ for $p = 0.11$. The asymptote shown is actually the trivial upper bound $B(0,R)$, i.e. when all spikes are coded (thus

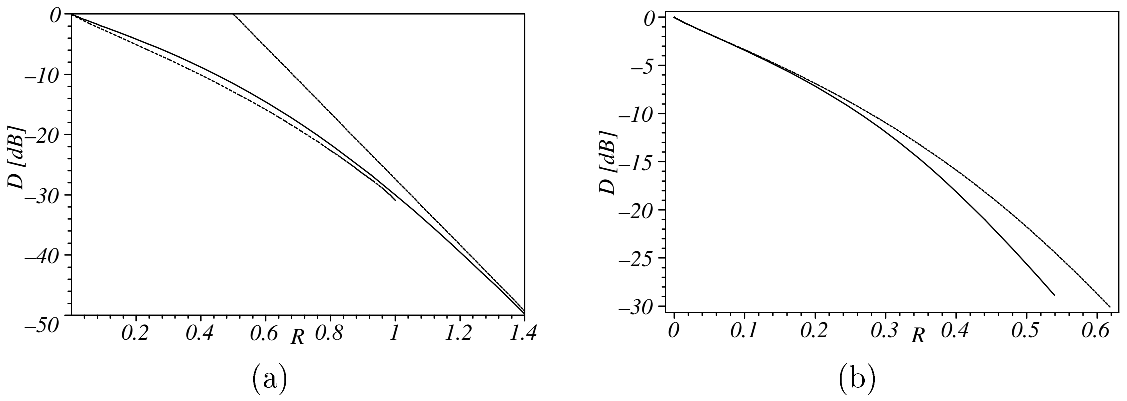


Figure 1: (a) Bernoulli-Gaussian spike with $p = 0.11$: empirical $D(R)$, upper bound (9) and trivial upper bound $B(0, R)$ (bottom to top curve). Normalized to unit variance. (b) Upper bounds for memoryless BG spike with $p = 1/16$ (top curve) and first order Markov Gaussian spike with $\alpha = 1/30$ and $\beta = 1/2$, yielding same marginal probability $p = 1/16$ (bottom curve).

at least $R = h(0.11) = 0.5$ bits are required before distortion decreases). The figure illustrates the change in $D(R)$ behavior between low and high rates that is typical of spike processes, regardless whether the continuous part of the pdf is a Gaussian or some other density.

4.1 First Order Markov Spikes

A natural extension of the memoryless model is to consider bursts of spikes. To model such a bursty behavior we can simply replace the Bernoulli process by a first order binary Markov process, with states S_0 (no spike) and S_1 (spike). Thus we get a spike process with correlated positions, but uncorrelated values. The Markov process is specified by two transition probabilities, $\alpha = \Pr\{S_0 \to S_1\}$ and $\beta = \Pr\{S_1 \to S_0\}$.

To compute a bound like (9) we first threshold the Markov-Gaussian spike process. It is easy to show that the resulting process is the product of a new first order Markov process and a *thresholded* i.i.d. Gaussian process. The parameters of the new ("thresholded") Markov process can be computed from α, β and the threshold t. Then we compute the entropy rate of the thresholded process, i.e. the spike position indicator (1), and use it to replace $h(\mu(T))$ in equation (7).

Figure 1(b) compares upper bounds for memoryless and Markov spikes with the same marginal spike probability, $p = \alpha/(\alpha+\beta)$. At low rates the two bounds are quite close, while at higher rates their horizontal spacing approaches the difference between zero-th and first order entropy (entropy rate). The low rate behavior can be explained as follows: low rates correspond to high thresholds, which means that the thresholded spikes are very sparse and there is thus less dependency among them. Therefore the difference between zero-th and first order entropy gets smaller. This finding indicates that for QNLA it is harder to exploit the correlation between neighboring coefficients at low rates .

5 Gaussian Mixture Models

In the previous Section we have briefly studied the spike process, which we proposed as a model for sparse transform coefficients. But its $R(D)$ behavior is very different from the one observed in actual image coders. Therefore we tried to find a better model, that allows e.g. a comparison with Mallat and Falzon's analysis of the low rate behavior of image transform coders [6]. Using high rate quantization results, they were able to derive the $D(R)$ curves that are typical for these image compression systems, i.e. steeper distortion decay at low rates that slows down to the usual exponential decay ($D \propto 2^{-2R}$) at rates above 1 bit per pixel. On the other hand, their result cannot be directly applied to assess the performance of a threshold-based compression scheme for a specific coefficient model, since they rely on approximations to measured statistics.

Here we take a different approach: we compute our bound for a simple i.i.d. model of wavelet coefficients and compare it with the empirical $D(R)$ curve obtained with the Blahut-Arimoto algorithm. If the curve and the bound stay close together, then we may use the quantities used in the bound as estimates for practical encoder design. The most interesting one is certainly the rate R_V allocated to each significant coefficient, or more generally the rate tradeoff between coding the position U (significance map) and the values V.

As an example of a coefficient model, we have chosen a simple i.i.d. Gaussian mixture model, where a hidden Bernoulli-p source picks one of two zero mean memoryless Gaussian sources. It is also apparent that this is a generalization of the spike model, where one source has zero variance. The model pdf is:

$$f(x) = pf(x|S=1) + (1-p)f(x|S=2) \tag{16}$$

where S is the hidden state selecting a distribution, and

$$f(x|S=i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-x^2/2\sigma_i^2}. \tag{17}$$

Such models have been used quite successfully in various applications, see [3] and references therein. To get realistic estimates for the parameters, we used a version of the EM algorithm [2] on the wavelet coefficients of the Lena image (transformed with SPIHT [7]). The scaling coefficients were discarded beforehand, since the model (16) does not fit them (due to nonzero mean); and therefore they are not accounted for in the rate distortion bound.

Figure 2(a) shows the obtained bound and the empirical $D(R)$ computed with Blahut's algorithm. Up to the knee, which is typical for image coding distortion rate curves, the distortion decays more rapidly. This means that mainly the sparse coefficients from the high variance source are retained by the thresholding operation. At higher rates, also the coefficients from the low variance source start being significant. The empirical $D(R)$ curve crosses the bound for numerical precision reasons: the pdf input to Blahut's algorithm should be more finely quantized around zero, where the low variance source is determinant. Another apparent fact is that at higher rates the bound becomes loose, because it converges to the trivial Gaussian upper bound $B(0, R)$. At high rate (very low threshold), $D(R)$ of a thresholded rv is close to $D(R)$ of the non-thresholded rv, which is far from the Gaussian upper bound for the mixture density in our example. A tighter upper bound is presented in the next Section.

The rate tradeoff between position and value codes is shown in Figure 4. Quite interestingly, the rate per significant sample never falls below 2.2 bits and remains below 5.5 bits in a wide range of low rates. Only at abysmally low rates it starts to grow again, e.g. $R_V = 2.8$ at $R = 10^{-8}$. The "plateau" at 1 bit per sample is again a sign that low variance samples are getting significant (at that point, the threshold t equals about $2\sigma_1$).

5.1 Other $D(R)$ Bounds for Gaussian Mixtures

Since Gaussian mixtures are a popular tool to approximate unknown pdf's, it is interesting to have rate distortion bounds that are tighter than (9). We derive an upper bound (without assuming NLA) and a lower bound.

Upper bound In Section 3 we bounded the rate distortion function of the samples above threshold with eqn. (6), since we were looking at threshold-based NLA. But a better $D(R)$ bound may result if also the samples below threshold are considered, i.e. upper bounded as Gaussians. Optimizing the bound then becomes a weighted rate allocation problem between two random variables with variance $\sigma^2_{T,0} := (\sigma^2 - A_T)/(1 - \mu_T)$, weight $1 - \mu_T$, and variance $\sigma^2_{T,1} := A_T/\mu_T$, weight μ_T, respectively (for readability the argument T is shown as index). The average rate per sample is $R = h(\mu_T) + (1-\mu_T)R_{V0} + \mu_T R_{V1}$, and the average distortion $\bar{D} = (1-\mu_T)\sigma^2_{T,0}2^{-2R_{V0}} + \mu_T\sigma^2_{T,1}2^{-2R_{V1}}$. We assume that $\bar{D} \leq \sigma^2_{T,0}$, which after carrying out the optimization corresponds to

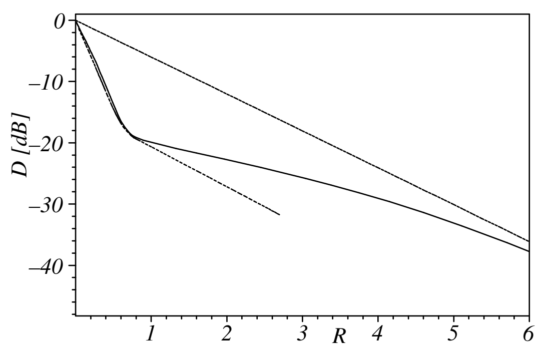
$$R \geq R_{min}(T) := \frac{1}{2}\mu_T \log_2 \frac{A_T(1-\mu_T)}{\mu_T(\sigma^2 - A_T)} + h(\mu_T).$$

An upper bound for all $R \geq R_{min}(T)$ is

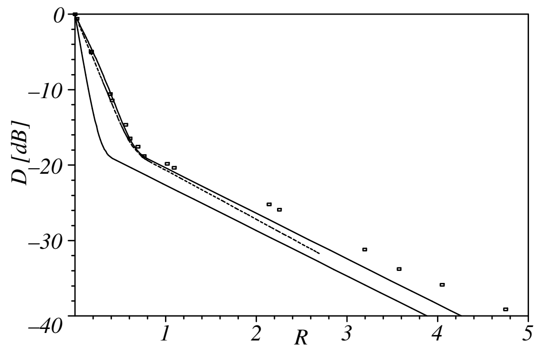
$$\begin{aligned} D(R) &\leq D_{ub}(T, R) := C(T)2^{-2R}, \\ \text{where } C(T) &= 2^{3h(\mu_T) + (1-\mu_T)\log_2(\sigma^2 - A_T) + \mu_T \log_2 A_T}. \end{aligned} \tag{18}$$

To plot a bound for all $R \geq 0$, we numerically search $t_{min} \in [0, \infty)$ that minimizes $C(t)$, yielding the best asymptotic ($\forall R \geq R_{min}$) upper bound of the family (18). Then we sweep t from 0 to ∞ to get candidate points $(R_{min}(t), D_{ub}(t, R_{min}(t)))$. If $\bar{R}(t)$ given by (10) is larger than $R_{min}(t)$, we keep the point. In the opposite case, the NLA upper bound (9) is tighter.

Lower bound The Gaussian mixture source can be seen as a finite alphabet discrete memoryless source S that switches between $|\mathcal{S}|$ Gaussian sources $\mathcal{N}(m_s, \sigma_s^2)$ with selection probabilities $w_s = Pr\{S = s\}$. A lower bound on $D(R)$ is found by assuming



(a)



(b)

Figure 2: (a) Gaussian mixture model for wavelet (detail) coefficients: upper bound on NLA distortion rate function and empirical $D(R)$. The top curve is the trivial Gaussian upper bound $B(0, R)$. Model parameters, normalized to unit variance: $p = 0.9141$, $\sigma_1^2 = 0.01207$ and $\sigma_1^2 = 11.51$. (b) Upper and lower distortion rate bounds for Gaussian mixture model. The middle curve is the empirical $D(R)$, the boxes denote (R, D) points achieved with a practical QNLA encoder.

that an oracle provides the hidden state variable S to the source encoder. Since $S \to X \to \widehat{X}$ form a Markov chain, we have

$$I(X; \widehat{X}|S) \leq I(X; \widehat{X}). \tag{19}$$

But $R_{lb}(D) := \min_{p(\hat{x}|x,s) \in Q_D} I(X; \widehat{X}|S)$ (with $Q_D = \{p(\hat{x}|x, s) : \mathrm{E}(X - \widehat{X})^2 \leq D\}$) can be computed exactly by solving the following standard rate allocation problem:

$$\min D_{lb}(R_{lb}) = \min_{\{R_s\}} \sum w_s \sigma_s^2 2^{-2R_s} \qquad \text{subject to } \sum w_s R_s = R_{lb} \text{ and } R_s \geq 0.$$

This yields the lower bound $D(R) \geq D_{lb}(R)$.

Figure 2(b) shows these bounds, together with the (R, D) points achieved by a simple QNLA scheme with octavely decreasing thresholds (on $3 \cdot 10^5$ pseudo-random samples from mixture source; significance maps are entropy coded, sign and refinements bits left uncoded). At low rates, thresholding with simple scalar quantization performs very close to the R-D optimum.

6 Conclusion

The example in Figure 3 shows that NLA followed by Lloyd-Max quantization can achieve remarkable low bit rate performance, and at a very reasonable computational cost (linear in the block size). Similar observations were already made in [5]. This is one reason why QNLA is popular in image coding applications that require low complexity and embedding. These two features are harder to achieve simultaneously with methods such as vector quantization or trellis-coded quantization.

In all our examples the bound is close to the actual $D(R)$ at low rates. This means that two-stage threshold-based nonlinear approximation is a good choice at low rates, provided that the quantizer stage is close to optimal. In fact simple scalar quantization is more than sufficient, as Figure 2(b) demonstrates. Section 4.1 illustrates that one should not expect much tighter low rate bounds by including correlation in the coefficient model. The bound (9) provides also an explicit rate tradeoff between the encoding of the significance map and the encoding of the sample values. This could be used in practical encoder design, at least at low target rates. For higher rates the bound becomes too loose due to the Gaussian (over)estimate in (8), and therefore the rate tradeoff will be less reliable.

Finally Gaussian mixture models are shown to be a useful tool for the rate distortion analysis of wavelet image coding. By allowing more than two mixture components, the actual R-D behavior can be approximated even better than in Figure 2.

References

[1] T. Berger, "Rate Distortion Theory", Prentice-Hall, 1971.

[2] O. Cappé, "A set of MATLAB/OCTAVE functions for the EM estimation of hidden Markov models with Gaussian state-conditional distributions ", http://tsi.enst.fr/ cappe/mfiles/h2m.tar.gz.

[3] M. S. Crouse, R. D. Nowak and R. G. Baraniuk, "Wavelet-Based Statistical Signal Processing Using Hidden Markov Models", IEEE Trans. Signal Processing, vol. 46, pp. 886-902, April 1998.

[4] D. L. Donoho, M. Vetterli, R. A. DeVore and I. Daubechies, "Data Compression and Harmonic Analysis", IEEE Trans. Inform. Theory, vol. 44, pp. 2435-2476, October 1998.

[5] N. Farvardin, J. W. Modestino, "Optimum quantizer performance for a class of non-Gaussian memoryless sources", IEEE Trans. Info. Theory, vol. 30, pp. 485 - 497, May 1984.

[6] S. Mallat and F. Falzon, "Analysis of Low Bit Rate Image Transform Coding", IEEE Trans. Signal Processing, vol. 46, pp. 1027-1042, April 1998.

[7] A. Said and W.A. Pearlman, "A new fast and efficient image codec on set partitioning in hierarchical trees", IEEE Trans. on CSVT, Vol. 6, pp. 243-250, June 1996.

[8] C. Weidmann and M. Vetterli, "Rate-Distortion Analysis of Spike Processes", in Proceedings of DCC'99, Snowbird, Utah, March 1999.

[9] C. Weidmann and M. Vetterli, "Quantized Nonlinear Approximation of Gaussian Random Variables", submitted to ISIT'2000.

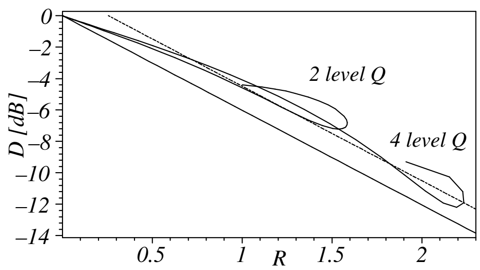


Figure 3: QNLA of Gaussian rv: thresholding followed by two-, resp. four-level Lloyd-Max quantization. Lower bound is Gaussian $D(R)$, dotted is Gish-Pierce asymptote.

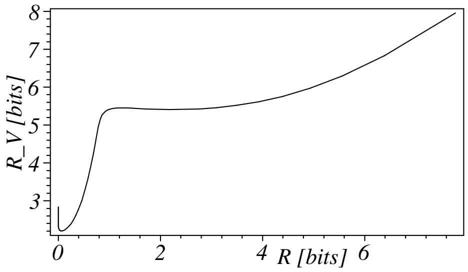


Figure 4: Gaussian mixture model: Rate per significant coefficient R_V (10) as a function of the average rate per coefficient. (Same parameters as in Fig. 2.)