RATE-DISTORTION ANALYSIS OF BACKWARD ADAPTIVE TRANSFORM CODING SCHEMES

David Mary, Dirk T.M. Slock

Institut Eurécom
2229 route des Crêtes, B.P. 193, 06904 Sophia Antipolis Cedex, FRANCE
Tel: +33 4 9300 2916/2606; Fax: +33 4 9300 2627
e-mail: {David.Mary,Dirk.Slock}@eurecom.fr

ABSTRACT

The main advantage of backward over forward adaptive coding schemes is to update the coding parameters with the data available at the decoder, avoiding thereby any excess bit rate. In this work, the performances of two practical backward adaptive transform coding schemes are analyzed in terms of rate and distortion for two transforms: the KLT (Karhunen-Loève transform) and the LDU (based on a Lower-Diagonal-Upper factorization of the covariance matrix R of the data) transform. For both algorithms, we model the expected distortion w.r.t. the number of vectors available at the decoder. Our analysis shows that for an algorithm using Sheppard's correction on the second order moment estimates, the distortion should converge to the target distortion. Without this correction, the effects of backward adaptation are shown to move the actual r(D) point of the system from the target point by the same term for both transforms. Simulations results confirming the theoretic analysis are then presented.

1. INTRODUCTION

For non- or locally- stationary data, the efficiency of transform coding relies on the updating of the coding parameters according to the source statistics changes. These updates aim to keep the performance of the structure close to a predetermined rate-distortion trade-off. Classically, they are sent as side information to the decoder, though this excess bit rate could be saved by using closed-loop, or backward adaptive algorithms.

We propose to model the effects of the backward adaptation for two simple algorithms and for two different transforms, the unitary KLT and the causal LDU transform [5, 6], which are known to be optimal for Gaussian signals in the ideal case (where R is constant and known from the coder and the decoder).A transform coding scheme is in the ideal case designed to reach a target point of the rate-distortion function $D(r)$. Some value of the distortion D will for example be chosen to be acceptable for the purpose of some application, resulting in an average bit rate r dedicated to represent the quantized signals. Assuming now that we use some backward adapted algorithm, an interesting question is to know if the corresponding distortion will converge or not to the target distortion, and if yes, how fast. Also, one may desire to control the rate which will then actually be required to represent the resulting quantized signals.

Although classical transform coding theory may appear as rather old and routine, the problem of backward adaptation in transform coding has, to our knowledge, received few attention until recently. Indeed, the interdependence of quantization and estimation noise on the several estimates makes the analysis of the recursive backward adaptation somehow delicate. Some convergence results have however been proven in the unitary case [1]. A theoretic comparison between causal and unitary approaches was lead in [3, 4], which did however not describe how practical backward adaptive transform algorithms would perform. This is the aim of this work, where we lead an analysis based on small perturbations.

Section 2 reviews and formalizes some results from the ideal coding schemes. Section 3 states how both the transformations and the quantization stepsize are perturbed, and the two adaptive algorithms are presented. Section 4 derives the distortion analysis for the two proposed algorithms and section 5 considers the problem of the rates. The last section presents some simulations results.

2. TRANSFORM CODING

2.1. Framework

Consider a stationary Gaussian vectorial source $\{X\}$. This source may be composed of any scalar sources $\{x_i\}$. In the classical transform coding framework, a linear transformation T is applied to each N-vector X_k to produce an N-vector $Y_k = TX_k$ whose components are independently quantized using scalar quantizers Q_i. A number of bits r_i is attributed to each Q_i under the constraint $\sum_i r_i = Nr$. For an entropy constrained scalar quantizer of a Gaussian source y_i, the high resolution distortion is $E(y_{i,k}^q - y_{i,k})^2 = \sigma_{q_i}^2 = c2^{-2r_i}\sigma_{y_i}^2$, where $c = \frac{\pi e}{6}$. At the decoder, the quantized vectors are computed by recovering the quantized value Y^q from the received codeword, and applying $X^q = T^{-1}Y^q$.

An important property of commonly used transformations is that $E\|\tilde{X}\|^2_{(T)} = E\|\tilde{Y}\|^2_{(T)} = D$, where $\|\tilde{X}\|^2_{(T)}$ denotes the variance of the quantization error on X, obtained for a transformation T.

2.2. Quantization Stepsize and Optimal Bit Assignment

The optimal bit assignment yields the well known distortion for the vectorial signal $\{Y\}$: $E\|\tilde{Y}\|^2_T = \frac{1}{N}\sum_{i=1}^N \sigma_{q_i}^2 = Nc2^{-2r} \times (\prod_{i=1}^N \sigma_{y_i}^2)^{1/N} = N\sigma_q^2$. Thus, the noise $\sigma_{q_i}^2$ should be independent of i. The number of bits assigned to the ith component is then $r + \frac{1}{2}\log_2 \sigma_{y_i}^2/(\prod_{i=1}^N \sigma_{y_i}^2)^{\frac{1}{N}}$. Under high resolution assumption, the quantization noise resulting from quantization with stepsize Δ_i is a uniformly distributed (over $[-\frac{\Delta_i}{2}, \frac{\Delta_i}{2}[$) random variable (r.v.), with variance $\sigma_{q_i}^2 = \frac{\Delta_i^2}{12}$. A simple way of realizing the optimal bit assignment is therefore to quantize all the components with an equal stepsize Δ. If now the y_i^q are entropy coded, the bitrate

(correponding to the average lengths of the codewords representing the transformed signals) is given by the zeroth order discrete entropy H of these signals. For Gaussian signals $r_i = H(y_i^q) \approx \frac{1}{2}\log_2 2\pi e \sigma_{y_i}^2 - \log_2 \Delta$. It can then easily be checked that choosing $\Delta = \sqrt{2\pi e}2^{-r}(\prod_{i=1}^{N}\sigma_{y_i}^2)^{\frac{1}{2N}} = \sqrt{2\pi e}2^{-r}\det(TRT^T)^{\frac{1}{2N}}$ ensures $r_i = H(y_i^q) = r \quad \forall i$. Both approaches are equivalent as long as the rate dedicated to each component is at least 1 bit per sample [1]. The corresponding distortion is then

$$D = \frac{\Delta^2}{12} = \frac{\pi e}{6}2^{-2r}\det(TRT^T)^{\frac{1}{N}}. \qquad (1)$$

2.3. Optimal Transforms

In the causal case, $Y = LX = X - \overline{L}X^q$, where $\overline{L}X^q$ is the reference vector. The output X^q is $Y^q + \overline{L}X^q$. Note that the reconstruction error $\tilde{X}$ equals the quantization error $\tilde{Y}$:

$$\tilde{X} = X - X^q = X - \overline{L}X^q - Y^q = Y - Y^q = \tilde{Y}, \qquad (2)$$

as in ADPCM. If we neglect the fact that (2) uses quantized data, one shows [6, 5] that the optimal (unimodular) L in terms of coding gain is such that $LRL^T = diag\{\sigma_{y_1}^2,...\sigma_{y_N}^2\}$, where $diag\{...\}$ represents a diagonal matrix whose elements are $\sigma_{y_i}^2$. In other words, the components y_i are the prediction errors of x_i with respect to the past values of X, the $X_{1:i-1}$, and the optimal coefficients $-L_{i,1:i-1}$ are the optimal prediction coefficients. The distortion is then

$$\begin{aligned} D_0 &= \frac{\pi e}{6}2^{-2r}\det(LRL^T)^{\frac{1}{N}} = \frac{\pi e}{6}2^{-2r}\det(VRV^T)^{\frac{1}{N}} \\ &= \frac{\pi e}{6}2^{-2r}\det\Lambda^{\frac{1}{N}} = \frac{\pi e}{6}2^{-2r}\det(R)^{\frac{1}{N}}, \end{aligned}$$

$$(3)$$

where V denotes a KLT of R, and Λ its eigenvalue matrix.

If we take now into account the fact that (2) uses quantized data, the actual prediction error variances $\sigma_{y_i}'^2$ are greater than the optimal ones σ_i^2 [6] due to a quantization noise feedback similar as that occuring in ADPCM, and are given by (σ_q^2 is the variance of the quantization noise)

$$\prod_{i=1}^{N}\sigma_{y_i}'^2 \approx \det(R)\left(1 + \sigma_q^2\sum_{i=1}^{N}(\frac{1}{\lambda_i} - \frac{1}{\sigma_i^2})\right), \qquad (4)$$

3. BACKWARD ADAPTIVE ALGORITHMS

As can be seen from (1) and (3), the design of a transform coding scheme indeed results, as every source coding problem, from a rate distortion trade-off. The higher the bitrate dedicated to the transform signals, the less the resulting distortion. Since T depends on R, as well as Δ (for a given target rate r, Δ is related to R by (1)), changes in the statistics of the source require to update T and Δ if one want the system to perform close to a chosen target point of the $r(D)$ function. We now propose two algorithms updating T and Δ with the data available at the decoder only.

Suppose that the coder deals with locally stationary data. We assume zero mean independent identically distributed (i.i.d.) Gaussian vectors (which is for example the case if the sampling period is high in comparison with their typical correlation time), and that the first N vectors are very accurately quantized and sent (without transformation) to the decoder.

Algorithm [1]:
- Step 1: The decoder disposes then of a first estimate of the covariance matrix $\widehat{R}_N = \frac{1}{N}\sum_{i=1}^{N}X_iX_i^T$.
- Step 2: A transform $\widehat{T}_N$ is computed such that $\widehat{T}_N\widehat{R}_N\widehat{T}_N^T$ is diagonal, where $\widehat{T}_N$ is either a KLT, either an LDU factorization of $\widehat{R}_N$, and a stepsize $\widehat{\Delta}_N^{[1]}$ is computed by $\sqrt{2\pi e}2^{-r}\det(\widehat{T}_N\widehat{R}_N\widehat{T}_N^T)^{\frac{1}{2N}}$.
- Step 3: These parameters are used to transform and quantize the $(N+1)$th vector by $Y^q = [\widehat{V}_N X_{N+1}]_{\widehat{\Delta}_N^{[1]}}$ in the unitary case, or $Y^q = [X_{N+1} - \widehat{\overline{L}}X^q]_{\widehat{\Delta}_N^{[1]}}$ in the causal case, where $[.]_\Delta$ denotes uniform quantization with stepsize Δ. The expected distortion for the $(N+1)$th vector is then $D^{[1]}(N+1) = E\widehat{\Delta}_N^{[1]^2}/12$.
- Step 4: Back to Step 1: the decoder disposes then of an estimate of the covariance matrix $\widehat{R}_{N+1} = \frac{1}{N+1}(\sum_{i=1}^{N}X_iX_i^T + X_{N+1}^q X_{N+1}^{qT})$, from which $\widehat{T}_{N+1}$ and $\widehat{\Delta}_{N+1}$ can be computed, used to code the $(N+2)$th vector, and so on.

Algorithm [2]:
A simple improvement to the previous algorithm can be made by using the following result. Suppose that the $\{X_K\}$ are quantized without transformation using the same (constant) stepsize Δ. Then it can be shown [1] that

$$EX_K^q X_K^{qT} = R^q = R + \frac{\Delta^2}{12}I + C, \qquad (5)$$

where I denotes the Identity matrix and $C \to 0$ elementwise as $\Delta \to 0$. In the previous algorithm now, if the stepsize converges to some small stepsize $\Delta_\infty(T)$, one may expect that the estimate of the covariance matrix converges to some $R + \frac{\Delta_\infty^2(T)}{12}I$. Thus, a better estimate of R can be computed after a certain amount of vectors, say N_1, by substracting $\frac{\widehat{\Delta}_K^2}{12}I$ to the current estimate of R. This correction on the estimate of the second order moment of the data by their quantized version is usually referred to as Sheppard's correction [2]. Except from this difference concerning $\widehat{R}$, the steps of Algorithm [2] are the same as in Algorithm [1].

The estimate of the covariance matrix for the second algorithm can now be expressed as

$$\begin{aligned} \widehat{R}_K^{[2]} &= \frac{1}{K}(\sum_{i=1}^{N}X_iX_i^T + \sum_{i=N+1}^{N_1}X_i^{[1]q}X_i^{[1]qT} + \sum_{i=N_1+1}^{K}X_i^{[2]q}X_i^{[2]qT}) - \frac{\widehat{\Delta}_{K-1}^{[2]^2}}{12}I \\ &= \frac{1}{K}\left(NR + \sum_{i=N+1}^{N_1}(R + D^{[1]}(i)I) + \sum_{i=N_1+1}^{K}(R + D^{[2]}(i)I)\right) \\ &\quad + \frac{1}{K}\left(\sum_{i=1}^{N}\Delta R_i^{(1)} + \sum_{i=N+1}^{N_1}\Delta R_i^{q[1](1)} + \sum_{i=N_1+1}^{K}\Delta R_i^{q[2](1)}\right) - \frac{\widehat{\Delta}_{K-1}^{[2]^2}}{12}I \end{aligned}$$

$$(6)$$

where $D(i)$ denotes the distortion obtained for the ith vector, and where we used the following notation:
- superscript $^{[j]}$ refers to algorithm [j],
- superscript ^{q} refers to quantization,
- superscript $^{(1)}$ refers to estimation noise occuring by estimating a covariance matrix R by the estimate $XX^T = \widehat{R}^{(1)} = R + \Delta R^{(1)}$,
- subscript $_K$ refers to the total number of vectors available at the decoder (except indeed from X_i, which denotes the ith vector).

The corresponding estimate for the first algorithm $\widehat{R}_K^{[1]}$ can also be computed from (6), where in this case the underlined terms vanish.

By writing $\frac{\widehat{\Delta}_{K-1}^{[2]^2}}{12} = D^{[2]}(K) + \delta D^{[2]}(K)$, the estimate (6) can

also be written as $\widehat{R}_K^{[2]} = R + \Delta R_K^{[2]}$, with

$$
\Delta R_K^{[2]} = \underbrace{\left[\frac{1}{K} \left(\sum_{i=N+1}^{N_1} D^{[1]}(i) + \sum_{i=N_1+1}^{K} D^{[2]}(i) \right) - D^{[2]}(K) \right] I}_{\Delta R_{K,det}^{[2]}}
$$
$$
+ \underbrace{\frac{1}{K} \left(\sum_{i=1}^{N} \Delta R_i^{(1)} + \sum_{i=N+1}^{N_1} \Delta R_i^{q[1](1)} + \sum_{i=N_1+1}^{K} \Delta R_i^{q[2](1)} \right) - \frac{\widehat{\Delta}_{K-1}^{[2]^2}}{12} I,}_{\Delta R_{K,sto}}
$$

(7)

where $\Delta R_{K,det}^{[2]}$ is a deterministic diagonal matrix, and $\Delta R_{K,sto}^{[2]}$ is a stochastic matrix. The update of the transform (to simplify the notations, the subscript $[2]$ will be omitted for $\widehat{T}_K^{[2]}$) is then computed so that $\widehat{T}_K \widehat{R}_K^{[2]} \widehat{T}_K^T$ is diagonal, and the updated stepsize $\widehat{\Delta}_K^{[2]} = \sqrt{2\pi e}2^{-r} \det\left(\widehat{T}_K \widehat{R}_K^{[2]} \widehat{T}_K^T \right)^{\frac{1}{2N}}$ is used to quantize the $(K+1)$th transform vector. The expected distortion is then

$$
D_{K+1}^{[2]} = E\widehat{\Delta}_K^{[2]^2}/12 = \frac{\pi e}{6} 2^{-2r} \det(\widehat{T}_K \widehat{R}_K^{[2]} \widehat{T}_K^T)^{\frac{1}{N}}.
$$

(8)

Using the unimodularity property of the transforms and considering $\Delta R_K^{[2]}$ in (7) as a perturbation term on R, one should compute in both unitary and causal cases (tr denotes the trace operator)

$$
D_{K+1}^{[2]} = \sqrt{2\pi e}2^{-r} \det\left(\widehat{R}_K^{[2]} \right)^{\frac{1}{2N}} \approx D_0[1 + \frac{1}{N}Etr\{\Delta R_K^{[2]} R^{-1}\}
$$
$$
+ \frac{1}{2N^2}E(tr\{\Delta R_K^{[2]} R^{-1}\})^2 - \frac{1}{2N}Etr\{\Delta R_K^{[2]} R^{-1} \Delta R_K^{[2]} R^{-1}\}],
$$

(9)

4. DISTORTION ANALYSIS

In order to compute the three expectations in (9), we can describe the r.v.s involved in (7) as follows. The elementary term $\{\Delta R_i^{(1)}\}$ corresponds to "one-shot" estimates of R based on a single observation. Since the $\{X_K\}$ are i.i.d., so is $\Delta R_i^{(1)}$. The elementary terms $\{\Delta R_i^{q[1,2](1)}\}$ correspond to "one-shot" estimates of $R + E(\widehat{\Delta}_{i-1}^{[1,2]^2}/12)I$ which, from (5), can be approximated as $R + D^{[1,2]}(i)I$. These terms are indeed not identically distributed. They are neither independent since $\Delta R_i^{q[1,2](1)}$ depends on $\widehat{\Delta}_{i-1}^{[1,2]}$, which depends on $\widehat{R}_{i-1}^{[1,2]}$, which in turns depends on $\Delta R_i^{q[1,2](1)}$. However, we assume that this is the case, since this dependence concerns only the noise part of the quantized vectors. Because of the quantization noise, the X_K^q are not Gaussian; again, for high resolution, we assume that this is however the case. The following result is now necessary to compute (9). Let $\Delta R_l^{(1)} = R_l = X_k X_k^T$ be the (symmetric) estimate of some $R_l = [\underline{r}_{l_1} \ldots \underline{r}_{l_N}]$ by means of one real zero mean Gaussian vector X_k, with $EX_k X_k^T = R_l$. Then it can be shown that $\Delta R_l^{(1)}$ is a zero mean r.v., and that among the N^2 blocks of $Evec\Delta R_l^{(1)} vec^T \Delta R_l^{(1)}$, the (i,j)th block $(Evec\Delta R_l^{(1)} vec^T \Delta R_l^{(1)})_{block(i,j)}$ equals $(R_l \otimes R_l)_{block(i,j)}$ $+\underline{r}_{l_j}\underline{r}_{l_i}^T$, where $\otimes$ denotes the Kronecker product. If now $R_l = R + D_l I$, the previous expression may, for highly correlated sources, be approximated as
$$
Evec\Delta R_l^{(1)} vec^T \Delta R_l^{(1)} \approx 2R_l \otimes R_l \approx 2[R \otimes R + D_l(R \otimes I + I \otimes R)].
$$

(10)

The first term of (9) may be written as

$$
\frac{1}{N}Etr\{\Delta R_K^{[2]} R^{-1}\} = \frac{1}{N}\left(tr\{\Delta R_{K,det}^{[2]} R^{-1}\} + \underbrace{Etr\{\Delta R_{K,sto}^{[2]} R^{-1}\}}_{0} \right)
$$
$$
\approx \frac{1}{N}\left[\frac{1}{K}\Sigma_{tot} - D^{[2]}(K) \right] tr\{R^{-1}\},
$$

(11)

with $\Sigma_{tot} = \sum_{i=N+1}^{N_1} D^{[1]}(i) + \sum_{i=N_1+1}^{K} D^{[2]}(i)$. The second term leads to

$$
\frac{1}{2N^2}E(tr\{\Delta R_K^{[2]} R^{-1}\})^2 = \frac{1}{2N^2}E(tr\{R^{-\frac{1}{2}}\Delta R_K^{[2]} R^{-\frac{1}{2}}\})^2
$$
$$
\approx \frac{1}{2N^2}(vec^T R^{-\frac{1}{2}})(R^{-\frac{1}{2}} \otimes I)
$$
$$
\times \underbrace{Evec\Delta R_K^{[2]} vec^T \Delta R_K^{[2]}}_{vec\Delta R_{K,det}^{[2]} vec^T \Delta R_{K,det}^{[2]} + Evec\Delta R_{K,sto}^{[2]} vec^T \Delta R_{K,sto}^{[2]}}(R^{-\frac{1}{2}} \otimes I)vecR^{-\frac{1}{2}}
$$

(12)

where the term corresponding to the deterministic part can be computed using the fact that $\Delta R_{K,det}^{[2]}$ is diagonal. The stochastic term in (12) generates, according to (7), fours terms, which can be computed using (10). The second term in (9) leads finally to

$$
\frac{1}{2N^2}E(tr\{\Delta R_K^{[2]} R^{-1}\})^2 \approx \frac{1}{KN} + tr\{R^{-1}\}(\frac{2\Sigma_{tot}}{K^2 N^2})
$$
$$
+ \frac{(tr\{R^{-1}\})^2}{2N^2}\left[\left(\frac{\Sigma_{tot}}{K} - D^{[2]}(K) \right)^2 + E\delta D_K^{[2]^2} \right],
$$

(13)

where for the purpose of this first order analysis, only the dominating terms needs to be retained. For example, the term $E\delta D^{[2]^2}(K) = E\widehat{\Delta}_K^{[2]^4}/12^2 - D_K^{[2]^2}$ will be neglected. Concerning the third term of (9), let G be $R^{-\frac{1}{2}}\Delta R_K^{[2]} R^{-\frac{1}{2}}$. Then we have $vecG = (R^{-\frac{1}{2}} \otimes R^{-\frac{1}{2}})vec\Delta R_K^{[2]}$, and we get

$$
-\frac{1}{2N}Etr\{\Delta R_K^{[2]} R^{-1} \Delta R_K^{[2]} R^{-1} = E - \frac{1}{2N}tr\{GG\}
$$
$$
= -\frac{1}{2N}Etr\{vecGvec^T G\}
$$
$$
= -\frac{1}{2N}Etr\{(R^{-\frac{1}{2}} \otimes R^{-\frac{1}{2}})Evec\Delta R_K^{[2]} vec^T \Delta R_K^{[2]}(R^{-\frac{1}{2}} \otimes R^{-\frac{1}{2}})\}
$$

(14)

where again, the arising terms can be computed using (10). Finally, the distortion occuring with the second algorithm can be approximated by the recursive expression

$$
D_{K+1}^{[2]} \approx D_0 \times
$$
$$
\left[1 + \frac{1}{K}(\frac{1}{N} - N) + tr\{R^{-1}\}\left(\frac{1}{N}\left[\frac{1}{K}\left(\sum_{i=N+1}^{N_1} D^{[1]}(i) + \sum_{i=N_1+1}^{K} D^{[2]}(i) \right) - D^{[2]}(K) \right] \right) \right].
$$

(15)

Inspecting the vanishing terms in (6), we obtain then the following recursive expression for the algorithm without correction

$$
D_{K+1}^{[2]} \approx D_0 \left[1 + \frac{1}{K}(\frac{1}{N} - N) + \frac{tr\{R^{-1}\}}{KN}\left(\sum_{l=N+1}^{N_1} D^{[1]}(i) \right) \right].
$$

(16)

On the one hand, the recursive expression (15) shows that the algorithm based on the Sheppard's correction should, as $K \to \infty$, converge to the target distortion D_0. On the other hand, the model provided by (16) does not converge to D_0 but to some $D_\infty > D_0$. It can be shown that

$$
D_\infty \approx \frac{D_0}{1 - D_0 \frac{tr\{R^{-1}\}}{N}}.
$$

(17)

5. RATE ANALYSIS

This section analyzes the bitrate required to entropy code the transform signals as $K \to \infty$. For the algorithm using the correction on the second order moment estimate, one should compute

$$r^{[2]}_{(T)} \approx \frac{1}{2N} \sum_{i=1}^{N} \log_2 2\pi e \sigma^2_{y_i,\infty} - \log_2 \Delta_0 \approx r + \frac{1}{2N} \log_2 \frac{\prod_{i=1}^{N} \sigma^2_{y_i,\infty}}{\det R}, \tag{18}$$

where $\sigma^2_{y_i,\infty}$ are the variances of the transform signals obtained by using the transform based on the asymptotic estimate $\widehat{R}^{[2]}_{\infty}$, which in this case is R. Thus, the estimated KLT and LDU should converge to the optimal transforms. The variances of the transform signals in the unitary case are then λ_i and

$$r^{[2]}_{(V)} = r. \tag{19}$$

In the causal case, a quantization noise feedback occurs which increases the variances of the transform signals, because the reference signal is composed with quantized data. Using (4) with $\sigma^2_q = \Delta^2_0/12$ to evaluate (18) gives

$$r^{[2]}_{(L)} \approx r + \frac{D_0}{2\ln 2} \sum_{i=1}^{N} \left(\frac{1}{\lambda_i} - \frac{1}{\sigma^2_i} \right). \tag{20}$$

As a conclusion, though the target distortion is reached in both cases, the unitary approach yields to lowest asymptotic rate. For the algorithm [1] now, one should compute

$$r^{[1]}_{(T)} = \frac{1}{N} \sum_{i=1}^{N} H(y^q_i) \approx \frac{1}{2N} \sum_{i=1}^{N} \log_2 2\pi e \sigma^2_{y_i,\infty} - \log_2 \Delta_\infty \tag{21}$$

where, this time, the $\sigma^2_{y_i,\infty}$ are the variances of the transform signals obtained by using the transform based on the asymptotic estimate $\widehat{R}^{[1]}_{\infty} \approx R + \frac{\Delta^2_\infty}{12} I$. In the unitary case, since a KLT of R is also a KLT of $R + \frac{\Delta^2_\infty}{12} I$, the $\sigma^2_{y_i,\infty}$ should again be equal to the λ_i. Using (17), we obtain

$$r^{[1]}_{(V)} \approx r - \frac{D_0}{2N\ln 2} tr\{R^{-1}\}. \tag{22}$$

In the causal case, the noise feedback in (4) involves this time $\Delta_\infty = 12 D^{1/2}_\infty$, and computing (21) yields

$$r^{[1]}_{(L)} \approx r^{[2]}_{(L)} - \frac{D_0}{2N\ln 2} tr\{R^{-1}\}. \tag{23}$$

Thus, the effect of not using the Sheppard correction in the backward adaptive algorithms is, for both transforms, to deplace the actual rate-distortion point from the targeted point by a rate $\delta r \approx \frac{D_0}{2N\ln 2} tr\{R^{-1}\}$ (or equivalently, as given by (17), by a distortion $\delta D_0 \approx D^2_0 tr\{R^{-1}\}/N$).

6. SIMULATIONS

For the simulations, we generated real Gaussian i.i.d. vectors with covariance matrix $R_{XX} = H R_{AR1} H^T$. R_{AR1} is the covariance matrix of an AR(1) process with $\rho = 0.9$. H is a diagonal matrix whose ith entry is $(N-i+1)^{1/3}$, $N = 3$. The target rate is 3 b/s. Figure (1) plots the averaged observed distorsions for the KLT and the LDU versus K, the theoretic model as given by (16), and the theoretic asymptotic distortion from (17). The optimal distortion is given by (3). Similar results are shown in Figure (2) for the algorithm [2], where the Sheppard's correction is applied after $N_1 = 60$ vectors, and where the theoretic model is given by (15).

7. REFERENCES

[1] V. Goyal, J. Zhuang, M. Vetterli, "Transform Coding with Backward Adaptive Updates," *IEEE trans. on Inf. Theory*, vol. 46, no. 4, July 2000.

[2] L. Cheded, P.A. Payne "The Exact Impact of Amplitude Quantization on Multi-Dimensional, High-Order Moments Estimation" *Signal Processing, Elsevier* no. 39, 1994.

[3] D. Mary, Dirk T.M. Slock, "Comparison between Unitary and Causal Approaches to Backward Adaptive Transform Coding of Vectorial Signals," *ICASSP02*, Orlando, Florida, May 2002.

[4] D. Mary, Dirk T.M. Slock, "Backward Adaptive Transform Coding of Vectorial Signals: a Comparison between Unitary and Causal Approaches," *EUSIPCO 2002*, Toulouse, France, September 2002.

[5] S.-M. Phoong, Y.-P. Lin, "Prediction-Based Lower Triangular Transform," *IEEE trans. on Sig. Proc.*, vol. 48, July 2000.

[6] D. Mary, Dirk T.M. Slock, "Vectorial DPCM Coding and Application to Wideband Coding of Speech," *ICASSP 2001*, Salt Lake City, USA, May 2001.
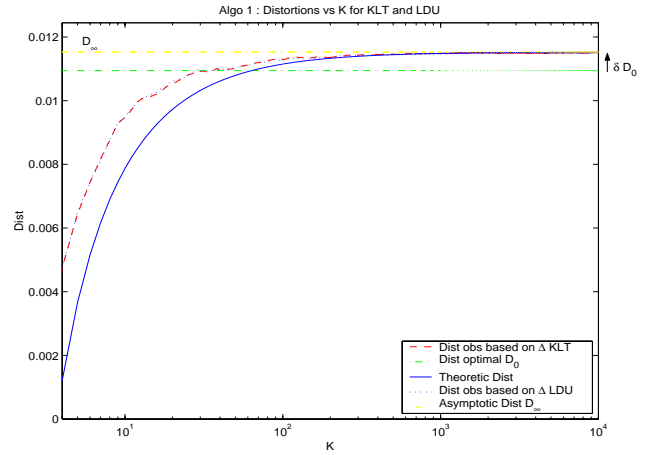
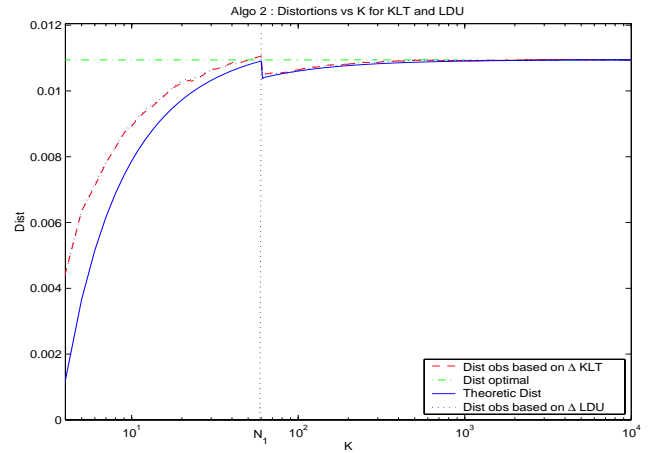**Fig. 1**. Distortions for Algorithm [1] vs K.

Fig. 2. Distortions for Algorithm [2] vs K.