

PLT VERSUS KLT

See-May Phoong

Dept. of EE & Inst. of Comm Engr.
National Taiwan Univ.
Taipei, Taiwan, R.O.C.

Yuan-Pei Lin

Dept. Electrical and Control Engr.
National Chiao Tung Univ.
Hsinchu, Taiwan, R.O.C.

ABSTRACT

In this paper a new class of transforms named Prediction-based Lower triangular Transform (PLT) is introduced for signal compression. The PLT is a nonunitary transform that yields the same coding gain as the Kahunen-Loeve transform (KLT). Unlike KLT, the derivation of PLT does not involve any eigen problem. An M -dimensional PLT can be solved by Levinson-Durbin recursion formula in $\mathcal{O}(M^2)$. Moreover the complexity of PLT is less than one half of KLT. For AR(1) inputs, the PLT has a closed form expression and needs only $(M - 1)$ multiplications.

1. INTRODUCTION

Transform coding has found many applications in various areas of signal processing and communication [1]. Fig. 1 shows a transform coder implemented using multirate building blocks. It is well known that the optimal unitary transform that yields the maximum coding gain is the KLT. The KLT is a unitary matrix that decorrelates the input autocorrelation matrix. It consists of eigenvectors of the input autocorrelation matrix. Due to its signal dependence and computational cost, the KLT is often only used as a benchmark for performance comparison. In many applications, sub-optimal signal independent transforms like discrete cosine transform (DCT) are often used.

Recently there are interests in applying the KLT to universal transform coding. By estimating the statistics from quantized data, the authors in [2] derive a class of universal transform coders using KLT. No side information is needed because both encoder and decoder can access the quantized data. Promising experimental results are demonstrated. In [3], the authors introduce a classification based method using KLT. The signal space is divided into a number of classes and a fixed transform is designed for each class. In the proposed two-stage algorithm, the encoder uses a collection of transform/bit allocation pairs. In [4], it was shown that under the assumption that the quantization noise

is white, KLT is no loss of generality among all transforms (unitary or nonunitary). That means, for a given statistics the coding gain of the best non unitary transform cannot be better than that of KLT. In this paper, we introduce a class of optimal nonunitary transform that has the same coding performance as KLT. In addition to its excellent coding performance, the new transform has many other features that make it an attractive choice for signal compression. Many results in this paper will be stated without proof. Readers are referred to [5] for details.

2. PREDICTION-BASED LOWER TRIANGULAR TRANSFORMS

Consider Fig. 1. Let $\mathbf{x}(n)$ and $\mathbf{y}(n)$ be respectively the input and output vectors of the transform \mathbf{T} . Assuming that the input is real and WSS, then the $M \times M$ autocorrelation matrices are related as

$$\mathbf{R}_y = \mathbf{T} \mathbf{R}_x \mathbf{T}^T. \quad (1)$$

Since \mathbf{R}_x is symmetric, its eigenvectors are orthonormal. By choosing these eigenvectors as the column vectors of \mathbf{T} , \mathbf{R}_y is diagonal. In other words, the transform coefficients $y_i(n)$ are uncorrelated. Such a unitary transform \mathbf{T} is the well-known KLT. One can show [1] that KLT maximizes the coding gain of the transform coder. Let σ_x^2 be the input variance and $\sigma_{KLT}^2(k)$ be the variances of $x_k(n)$ for KLT. Then under optimal bit allocation, the coding gain of KLT is given by [1]

$$\mathcal{CG}_{KLT} = \frac{\sigma_x^2}{\prod_{k=0}^{M-1} [\sigma_{KLT}^2(k)]^{1/M}} = \frac{\sigma_x^2}{[\det \mathbf{R}_x]^{1/M}}. \quad (2)$$

However the KLT is not the only transform that has this decorrelation property. In fact, there exists a lower triangular matrix \mathbf{T} such that the transform coefficients $y_k(n)$ are decorrelated. To see this, we need to use the following lemma from matrix theory [6]:

Lemma 1 *The LU decomposition of matrices [6]: Let \mathbf{A} be an M by M nonsingular matrix. Suppose that all of its*

THIS WORK WAS SUPPORTED BY NSC 87-2218-E-002-053 AND NSC 87-2213-E-009-052, TAIWAN, R.O.C.

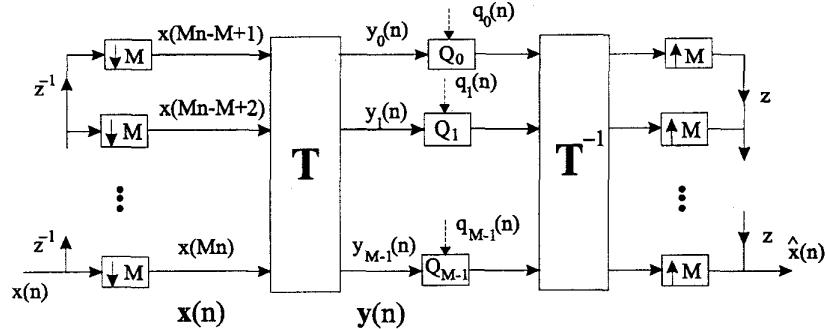


Figure 1: An M -dimensional transform coder.

principle submatrices \mathbf{A}_K are nonsingular. Then \mathbf{A} can be written as

$$\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{U}, \quad (3)$$

in which \mathbf{L} (respectively \mathbf{U}) is a lower (respectively upper) triangular matrix with all diagonal entries equal to 1, and \mathbf{D} is a diagonal matrix. Moreover the matrices \mathbf{L} , \mathbf{U} and \mathbf{D} are unique. In particular, \mathbf{D} is determined by $\det[\mathbf{D}_K] = \det[\mathbf{A}_K]$, where $K = 1, \dots, N$.

It is not difficult to show that if the matrix \mathbf{A} is symmetric, then the unique matrices \mathbf{U} and \mathbf{L} in Lemma 1 are related as $\mathbf{U} = \mathbf{L}^T$. Applying this fact to (1), we immediately see that there is a unique lower triangular matrix \mathbf{P} such that $\mathbf{P}\mathbf{R}_x(M)\mathbf{P}^T$ is a diagonal matrix. Such a transform \mathbf{P} will be called Prediction-based Lower triangular Transform (PLT). The reason for this name will become clear later. The geometric mean (GM) of the subband variances σ_{PLT}^2 becomes

$$\prod_{k=0}^{M-1} [\sigma_{PLT}^2(k)]^{1/M} = \det[\mathbf{P}\mathbf{R}_x\mathbf{P}^T]^{1/M} = \det[\mathbf{R}_x]^{1/M}. \quad (4)$$

The GM of $\sigma_{PLT}^2(k)$ is the same as that of $\sigma_{KLT}^2(k)$. However the PLT is not unitary and one can show [5] that the quantization noise $q_i(n)$ in Fig. 1 will be amplified by the inverse PLT at the decoder. Therefore the coding gain of the PLT is less than that of KLT in (2) if the traditional transform coding structure in Fig. 1 is used. To solve this problem, we will introduce a novel minimum noise structure for the PLT in the next section.

Given any input with autocorrelation matrix \mathbf{R}_x , PLT can be obtained by using the Gaussian elimination in $\mathcal{O}(M^3)$. However since \mathbf{R}_x is Toeplitz, the computation of PLT can be done in $\mathcal{O}(M^2)$. To see this, let $p_{k,i}$ be the coefficients of \mathbf{P} in the k th row and let

$$P_k(z) = 1 + p_{k,k-1}z^{-1} + \dots + p_{k,0}z^{-k}, \quad (5)$$

for $k = 1, \dots, M - 1$. If we take $P_k(z)$ as the k th order prediction filter of $x(n)$, then the transform coefficients $y_k(n)$ are the k th order prediction error $e_k(n - M + k + 1)$. Using the orthogonality principle from linear prediction theory [1], one can show that $E\{y_k(n)y_j(n)\} = 0$ for $k \neq j$. From Lemma 1, we know that the lower triangular matrix with such a decorrelation property is unique. Therefore the matrix \mathbf{P} formed by the prediction filter coefficients is the PLT. Hence \mathbf{P} is called the *prediction-based* lower triangular matrix. Using the Levinson-Durbin fast algorithm, all the k th order prediction filters (for $k = 1, \dots, M - 1$) can be obtained in $\mathcal{O}(M^2)$. The k th subband variance for PLT, $\sigma_{PLT}^2(k)$, is equal to the k th order prediction error variance, $\mathcal{E}(k)$. Summarizing the results, we have

Theorem 1 The PLT: Consider the transform coder in Fig. 1. Given any wide sense stationary input $x(n)$, there exists a unique lower triangular matrix \mathbf{P} with unity diagonal elements such that the transform coefficients $y_k(n)$ are uncorrelated. The unique optimal lower triangular transform can be obtained by choosing $P_k(z)$ in (5) as the k th-order optimal prediction error filter. Moreover the autocorrelation matrix of the subband vector $\mathbf{y}(n)$ is

$$\mathbf{R}_y = \text{diag}(\mathcal{E}(0), \mathcal{E}(1), \dots, \mathcal{E}(M - 1)), \quad (6)$$

where $\mathcal{E}(k)$ is the prediction error variance of $P_k(z)$.

3. LADDER-BASED AND MINLAB STRUCTURES FOR PLT

Using two different factorization forms of lower triangular matrices, we are able to find two structurally PR implementations using ladder structure for PLT [5]. In this paper, we will discuss only one of the structures and readers are referred to [5] for the other. Note that the lower triangular matrix \mathbf{P} can be decomposed as

$$\mathbf{P} = \mathbf{P}_1 \mathbf{P}_2 \dots \mathbf{P}_{M-1}, \quad (7)$$

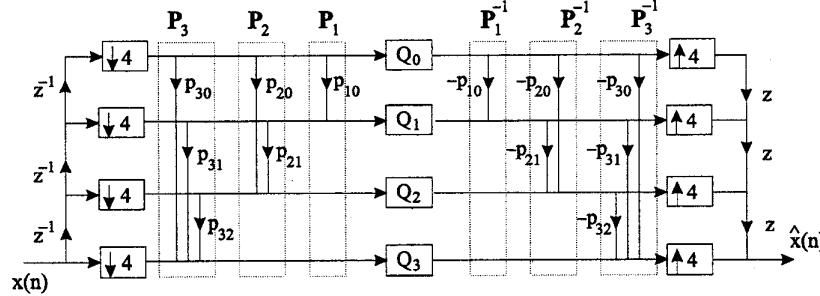


Figure 2: Ladder-based implementation of a four-channel PLT coder.

where the elementary matrix \mathbf{P}_k is

$$\mathbf{P}_k = \begin{pmatrix} \mathbf{I}_k & \mathbf{0}_{M-k-1 \times k} \\ p_{k,0} \dots p_{k,k-1} & 1 \dots 0 \\ \mathbf{0}_{k \times M-k-1} & \mathbf{I}_{M-k-1} \end{pmatrix}.$$

Using (7), we have $\mathbf{P}^{-1} = \mathbf{P}_{M-1}^{-1} \dots \mathbf{P}_1^{-1}$. The inverses of these elementary matrices, \mathbf{P}_k^{-1} , can be obtained by replacing the nontrivial elements $p_{k,i}$ in \mathbf{P} with $-p_{k,i}$. Fig. 2 shows the ladder-based implementation obtained from this factorization for the case of $M = 4$. The advantages of the ladder structure are two-fold: (1) In the absence of quantizers, the PLT coder continues to have PR even when all the multipliers $p_{k,i}$ in the structure are quantized to a finite precision. (2) The inverse transform also has $p_{k,i}$ as its multiplier and it can be obtained by inspection.

However, the PLT is a nonunitary matrix, so is its inverse. Hence the PLT coder does not have the energy preservation property. In general, the quantization noise generated in the subbands will be amplified at the decoder. To see this, consider Fig. 2. The inputs to the multipliers $p_{k,i}$ at the encoder are the unquantized data while the inputs to the multipliers $p_{k,i}$ at the decoder are quantized data. That means, the predictors at the encoder use unquantized data as their observations while the predictors at the decoder use the quantized data. It is this mismatch that causes the noise amplification. To avoid the mismatch of observations, one can modify the structure so that the inputs to the multipliers $p_{k,i}$ at the encoder are the quantized data instead of the original unquantized values. The encoder of the modified structure for $M = 4$ case is shown in Fig. 3 and the decoder is the same as Fig. 2. From the figure, one can verify that the structure has the unity noise gain property. This property holds even for correlated and colored quantization noise. The implementation in Fig. 3 will be referred to as MINIMUM NOISE LAdder-based Biorthogonal (MINLAB) structure for PLT.

Using the unity noise gain property and the fact that the GM of the subband variances is equal to $\det[\mathbf{R}_x]$, one can prove that the coding gain for the MINLAB PLT coder is the same as KLT.

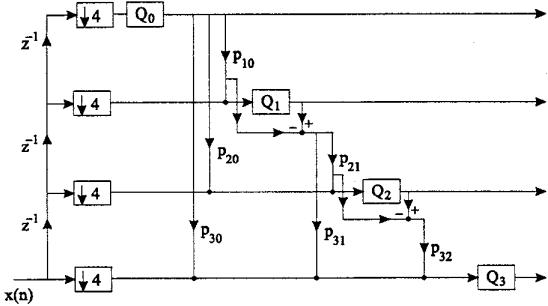


Figure 3: The encoder of a PLT MINLAB structure, the decoder is the same as Fig. 2.

Remark: As MINLAB coder has the unity noise gain property, its average output noise variance is minimized if all quantizers have the same noise variance. Therefore the **equal stepsize rule** is optimal and entropy coding can be used to encode the outputs of Q_k .

4. OTHER ATTRACTIVE FEATURES OF PLT

Optimal Universal Transform Coders: Since the MINLAB structures are structurally PR, even when the multipliers $p_{k,i}$ in Fig. 3 are time-dependent, the PR property continues to hold. The statistics of the input can be adaptively estimated from the quantized data and these informations can be used to update the prediction filters. Since both the encoder and decoder have access to the quantized data, there is no need to send or store any side information. Given any input signal, we can initialize the PLT as $\mathbf{P}^{(0)} = \mathbf{I}$. After each input vector $\mathbf{x}(n)$ is encoded with $\mathbf{P}^{(n)}$, the statistics can be up-

dated and the transform $\mathbf{P}^{(n+1)}$ can be computed in $\mathcal{O}(M^2)$ using the Levinson-Durbin fast algorithm. After a few iterations, if the statistics of the input does not vary too fast, the rate of adaptation can be reduced. And the transform can be updated only after a number of input vectors are encoded.

Lossless/Lossy Transform Coders: In many applications, it is desired that a lossy coding system becomes lossless when a sufficient bit rate is available. Since the multipliers of KLT are real numbers, in practice they have to be quantized. In general KLT with quantized multipliers will not have the PR property. Therefore KLT in general cannot be used for lossless coding. On the other hand, the MINLAB structures for PLT can implement both lossy and lossless coders after some minor modifications. To see this, assume that the input values $x(n)$ are integers. If the outputs of all the multipliers (in both encoder and decoder) in Fig. 3 are quantized to integer, then the PR property continues to hold. This quantization can be roundoff, truncation or ceiling quantization. Recall that PLT coder is optimal when equal stepsize rule is applied to Q_k in Fig. 3. Therefore we can set the stepsize Δ_k of quantizer Q_k to the same value and entropy coding can be used to encode the quantized subband signals. If all $\Delta_k = n > 1$, then we have a lossy PLT coder. If all $\Delta_k = 1$, then the PLT coder becomes lossless. Therefore we can implement both lossy and lossless coding with the same PLT coder by simply adjusting the stepsizes Δ_k .

5. AR(1) INPUTS

If the input is an AR(1) process with correlation ρ , then all the prediction error polynomials $P_k(z)$ in (5) will have the same form $(1 - \rho z^{-1})$. The PLT in this case has the following closed form

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}. \quad (8)$$

Once ρ is known, we can find \mathbf{P} by inspection, no computation is needed. Therefore the optimal universal PLT coder introduced in Section 4 becomes very simple, we need to estimate only one parameter ρ . The coding gain in this case becomes

$$CG_{PLT,MIN}(M) = \left[\frac{1}{1 - \rho^2} \right]^{\frac{M-1}{M}}. \quad (9)$$

When M is large, the above gain approaches the coding gain of a DPCM coder.

Also note that the transform in (8) is almost independent of the input signal. An $M \times M$ PLT for AR(1) process needs only $(M-1)$ multiplications and additions. Thus its complexity is lower than the DCT which has a complexity of $\mathcal{O}(M \log M)$. Moreover the PLT in (8) is optimal for all AR(1) processes, unlike DCT which is optimal only when ρ approaches 1.

6. COMPARISON WITH KLT

We will conclude the paper by providing a comparison of PLT and KLT:

1. PLT has the same coding performance as the KLT.
2. The design cost of PLT is much lower than that of KLT. The implementational of PLT is less than one half of KLT.
3. Unlike KLT, PLT has a structurally PR implementation using simple building blocks.
4. PLT coders can implement both lossy and lossless compression while KLT in general cannot be used for lossless coding.
5. Unlike KLT, PLT has the simple form in (8) for AR(1) inputs. In this case, the M -dimensional PLT takes only $M - 1$ multiplications and additions for each input block. Moreover it is almost signal independent.

7. REFERENCES

- [1] N. S. Jayant, and P. Noll, *Digital coding of waveforms*, Prentice Hall, 1984.
- [2] V. Goyal, J. Zhuang, M. Vetterli, and C. Chan, "Transform coding using adaptive bases and quantization," in *Proc. IEEE Int. Conf. Image Proc.*, pp. 365-368, 1996.
- [3] M. Effros, and P. A. Chou, "Weighted universal transform coding: Universal image compression with the Karhunen-Loeve transform," in *Proc. IEEE Int. Conf. Image Proc.*, pp. 61-64, 1995.
- [4] P. P. Vaidyanathan, "Theory of optimal orthonormal coder," *IEEE Trans. Signal Proc.*, June 1998.
- [5] S. M. Phoong, and Y. P. Lin, "Prediction-based lower triangular transform," *Preprint*, National Taiwan University, Oct. 1998 (submitted to IEEE Trans. SP).
- [6] R. A. Horn, and C. R. Johnson, *Matrix analysis*, Cambridge University Press, 1985