Signal Processing 7 (1984) 231-249
North-Holland

ON THE APPROXIMATION OF THE DISCRETE KARHUNEN-LOEVE TRANSFORM FOR STATIONARY PROCESSES

Michael UNSER, member EURASIP

Signal Processing Laboratory, Swiss Federal Institute of Technology, 16 Ch. de Bellerive, 1007 Lausanne, Switzerland

Received 11 January 1984
Revised 3 April 1984 and 19 July 1984

Abstract. Suboptimal fast transforms are useful substitutes to the optimal Karhunen-Loève transform (KLT). The selection of an efficient approximation for the KLT must be done with respect to some performance criterion that might differ from one application to another. A general class of criterion functions including most of the commonly used performance measures is introduced. They are shown to be optimized by the KLT. Various properties of the eigenvectors of the symmetric Toeplitz covariance matrix of a wide sense stationary process are reviewed. Several transforms such as the complex or real, odd and even Fourier transforms (DFT, DOFT, DREFT, DROFT), the cosine and even sine transforms (DCT, DEST) are obtained from the decomposition of a symmetric Toeplitz matrix in the sum of a circulant and a skew circulant matrix. These transforms are compared on the basis of a general performance criterion and appear to be good substitutes for the optimal KLT. Finally, it is shown that these transforms are asymptotically equivalent in performances to the KLT of an arbitrary wide sense stationary process.

Zusammenfassung. Gewisse orthogonale Transformationen erlauben eine effiziente Approximation der optimalen Karhunen-Loève Transformation (KLT). Die Wahl einer sub-optimalen Transformation wird mit Hilfe eines Performanz-Kriteriums gemacht, dieses kann je nach Applikation verschieden sein. Deshalb wird eine allgemeine Klasse von Kriterien eingeführt, die die meisten gebräuchlichen Kriterien umfasst. Es wird gezeigt, dass diese Familie durch die KLT optimiziert wird. Gewisse Eigenschaften von Eigenvektoren einer symmetrischen Toeplitz Matrix werden gezeigt. Eine zirkulare Dekomposition der Kovarianz Matrix eines im weiten Sinne stationären Prozesses wird eingeführt. Diese führt zur Summe zweier block-zirkulanten Matrizen, wovon eine symmetrisch und die andere anti-symmetrisch ist. Eine Anzahl reelle und komplexe orthogonale Transformationen (DFT, DOFT, DREFT, DROFT, DCT, DEST) die den einen oder anderen Term (oder beide teilweise) dieser Dekomposition, diagonalisieren werden vorgeschlagen. Diese Transformationen werden anhand des generellen Performanz-Kriteriums verglichen. Dank ihrer fast optimalen Performanzen erlauben sie eine effiziente Approximation der KLT. Letztlich wird gezeigt, dass jede Transformation asymptotisch der KLT aequivalent ist, und dies für alle im weiteren Sinne stationären Prozesse.

Résumé. Certaines transformations orthogonales rapides permettent une approximation efficace de la transformation optimale de Karhunen-Loève (KLT). Le choix d'une transformation sous-optimale appropriée s'effectue généralement sur la base d'un critère de performance; celui-ci peut être différent d'une application à l'autre. A cet effet, nous introduisons une classe générale de critères de performances qui englobe la majorité des mesures de performances couramment utilisées. Il est montré que cette famille de fonctions est optimisée par la KLT. Certaines propriétés des vecteurs propres d'une matrice symétrique de Toeplitz sont présentées. Une décomposition circulaire de la matrice de covariance d'un processus stationnaire au sens large est introduite. Celle-ci fait apparaitre la somme de deux matrices à blocs circulants symétriques et anti-symétriques. Un certain nombre de transformations orthogonales complexes ou réelles (DFT, DOFT, DREFT, DROFT, DCT, DEST), permettant une diagonalisation de l'un des termes (ou éventuellement une diagonalisation partielle des deux termes) apparaissant dans cette décomposition, sont alors proposées. Ces transformations sont comparées sur la base du critère général de performance. De par leur comportement très proche de l'optimum, elles permettent une approximation efficace de la KLT. Il est finalement établi que chacune de ces transformations est asymptotiquement équivalente en performances à la KLT de n'importe quel processus stationnaire au sens large.

Keywords. Karhunen-Loève transform, Karhunen-Loève transform approximation, stationary process, Toeplitz matrix, eigenvalue decomposition.

0165-1684/84/$3.00

1. Introduction

Orthogonal transforms are widely used in the area of digital signal processing [1]. Research efforts and applications include domains such as image processing, speech processing, pattern recognition, communication systems and generalized filtering. Given the second order statistical properties of a N-dimensional random vector, the optimal transform for data representation or compression [1], data analysis [2] or data processing [3] is the well known Karhunen-Loève transform (KLT) which is defined as follows.

Consider an N component random vector $\boldsymbol{x}$ with associated covariance matrix:

$$\boldsymbol{C}_N = E\{(\boldsymbol{x}-E\{\boldsymbol{x}\})(\boldsymbol{x}-E\{\boldsymbol{x}\})^{\mathrm{T}}\}. \tag{1}$$

The rows of the corresponding orthogonal Karhunen-Loève $N \times N$ transform matrix $\boldsymbol{\Phi}_N$ are the eigenvectors of the symmetric positive definite covariance matrix $\boldsymbol{C}_N$. The essential property of this transform is that it produces uncorrelated transform coefficients $\boldsymbol{y} = \boldsymbol{\Phi}_N \boldsymbol{x}$:

$$E\{(\boldsymbol{y}-E\{\boldsymbol{y}\})(\boldsymbol{y}-E\{\boldsymbol{y}\})^{\mathrm{T}}\} = \boldsymbol{\Phi}_N \boldsymbol{C}_N \boldsymbol{\Phi}_N^{\mathrm{T}} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ 0 & \cdots & \lambda_N \end{bmatrix}, \tag{2}$$

where the λ_i's are the eigenvalues or characteristic roots of $\boldsymbol{C}_N$. Thereby, this representation admits processing schemes in which each datum is manipulated independently from the others. The KLT also minimizes the approximation error when the number of coefficients that are used to represent the input data block is less than N.

In most of the signal processing applications, the random vector $\boldsymbol{x}$ is considered as a realisation of a wide sense stationary process which leads to a symmetric Toeplitz covariance matrix $\boldsymbol{C}_N$. Unfortunately, despite the simple structure of this type of matrices, no simple way to compute the eigenvectors seems to be known until now. The KLT is usually computed using standard—computationally costly—numerical iterative eigenvector extraction methods [4]. There is no unique KLT for all random processes, and it is, in general, not possible to find a fast (FFT-type) algorithm to compute the transform coefficients. It is possible, however, to approximate closely the KLT of a wide sense stationary process by suboptimal transforms such as the DFT [5], the DCT [6] or other sinusoidal transforms [7]. The use of this type of transforms, was justified until now by the fact that they belong to a family of transforms asymptotically equivalent to the KLT of a first order Markov process [7].

In this paper, a new approach to the approximation of KLT, that does not restrict itself to a specific class of stationary processes (such as the Markov-1 family), is presented. It is based on a circular decomposition of a symmetric Toeplitz matrix which emphasizes two extreme cases for which the exact KLT can be derived analytically. Interesting solutions for the KLT in one of these extreme cases are the complex discrete Fourier (DFT), the discrete real even Fourier (DREFT), the complex discrete odd Fourier (DOFT) and the discrete real odd Fourier (DROFT) transforms. The well known discrete cosine (DCT) and even sine (DEST) are shown to be closely related to this decomposition. All these transforms can be used as efficient substitutes of the optimal KLT. An important result of this work is that all these transforms are asymptotically equivalent in performances to the KLT of an arbitrary wide sense stationary process. This means that the performance degradation, resulting from the substitution of the KLT by one of these suboptimal transforms, vanishes to zero as the block size increases.

The present article is organized as follows. Section 2 is mainly concerned with the notion of optimality of the KLT. Important properties of the eigenvector decomposition of persymmetric matrices (including the family of symmetric Toeplitz matrices) are reviewed in Section 3. The central decomposition of a symmetric Toeplitz matrix in the sum of a circulant and a skew-circulant matrix is introduced in Section

4. These matrices are respectively diagonalized by the complex even and odd Fourier transforms (DFT, DOFT). In Section 5, it is shown how to form equivalent or closely related real orthogonal transforms by grouping all pairs of complex conjugate basis vectors. These real transforms are generally better substitutes to the KLT than the associated complex transforms. The asymptotic performance equivalence with the KLT is finally established in Section 6.

2. Criterion functions for transform selection

The Karhunen–Loève expansion is optimal for a large variety of problems. The optimality of this transform has to be understood in the sense that it maximizes or minimizes a certain criterion (or cost) function that may depend on the particular type of application. The usefulness of the KLT is mainly due to the fact that it provides a canonical decorrelated representation of the data. It is the transform which diagonalizes the covariance matrix.

Let us consider an $N\times N$ real or complex transform matrix $\boldsymbol{U}_N=[\boldsymbol{u}_1\boldsymbol{u}_2\cdots\boldsymbol{u}_N]^{\mathrm{T}}$ which is applied to the input data vector $\boldsymbol{x}$. The performances in data representation associated with this particular transform is usually measured on the covariance matrix of the transformed coefficients given by $\boldsymbol{U}_N\boldsymbol{C}_N\boldsymbol{U}_N^{\mathrm{H}}$, where $\boldsymbol{U}_N^{\mathrm{H}}$ is the Hermitian transpose of $\boldsymbol{U}_N$. Most commonly used performance measures (distortion rate, basis restriction errors) are computed from the diagonal elements only. A relatively general class of criterion or performance functions is introduced as

$$\zeta(\boldsymbol{U}_N\boldsymbol{C}_N\boldsymbol{U}_N^{\mathrm{H}})=\sum_{i=1}^{N}G(\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{C}_N\boldsymbol{u}_i^{*}),\tag{3}$$

where $G(\cdot)$ is a continuous, monotonously increasing (or decreasing), convex (or concave) function. It is shown in appendix A that under the constraint of an energy preserving transform with normalized row vector $\boldsymbol{u}_i$, ζ is bounded by

$$NG(\mathrm{tr}(\boldsymbol{C}_N)/N)\leq\zeta(\boldsymbol{U}_N\boldsymbol{C}_N\boldsymbol{U}^{\mathrm{H}})\leq\zeta^*=\sum_{i=1}^{N}G(\lambda_i),\tag{4}$$

if $G(\cdot)$ is monotonously increasing convex or decreasing concave, or

$$NG(\mathrm{tr}(\boldsymbol{C}_N)/N)\geq\zeta(\boldsymbol{U}_N\boldsymbol{C}_N\boldsymbol{U}_N^{\mathrm{H}})\geq\zeta^*=\sum_{i=1}^{N}G(\lambda_i),\tag{5}$$

if $G(\cdot)$ is monotonously decreasing convex or increasing concave.

The λ_i's are the eigenvalues of the covariance matrix $\boldsymbol{C}_N$. The optimum value ζ^* corresponds to the Karhunen–Loève transform. It is important to note that if the diagonal elements of the covariance matrix $\boldsymbol{C}_N$ are all equal (which is the case for a wide sense stationary process), the criterion function associated to the identity-transform (or any permutation matrix) is minimum. Therefore, any non trivial transformation will improve the situation. From this point of view, the original coordinate system is the worst for data representation. Any linear transform will decrease the cross correlation between samples and result in a covariance matrix closer to diagonal.

The relative improvement of performance obtained for a given transform $\boldsymbol{U}_N$ can be computed from the normalized coefficient

$$\xi(\boldsymbol{U}_N\boldsymbol{C}_N\boldsymbol{U}_N^{\mathrm{H}})=(\zeta(\boldsymbol{U}_N\boldsymbol{C}_N\boldsymbol{U}_N^{\mathrm{H}})-\zeta(\boldsymbol{C}_N))/(\zeta(\boldsymbol{\Phi}_N\boldsymbol{C}_N\boldsymbol{\Phi}_N^{\mathrm{T}})-\zeta(\boldsymbol{C}_N)),\tag{6}$$

which will be one when the transform is equivalent to the Karhunen–Loève transform and zero when no improvement is obtained, independent of the fact that the KLT corresponds to a maximum or a minimum of the criterion function.

The class of criterion function, introduced in this section, includes most of the functions that are commonly used in the literature and that may be different from one problem to another. Some examples are given below.

a) Energy criterion

Let $\Lambda_N^{(u)} = \mathrm{diag}(U_N C_N U_N^H)$ be the diagonal matrix obtained after transformation, in setting all non-diagonal elements to zero. The energy criterion associated to transform U_N is defined as

$$\zeta_1(U_N C_N U_N^H) = |\Lambda_N^{(u)}|^2 = \sum_{i=1}^{N} (u_i^T C_N u_i^*)^2 = |U_N^H \Lambda_N^{(u)} U_N|^2, \tag{7}$$

where $|\cdot|^2$ is the Hilbert-Schmidt norm of a matrix and is invariant by orthogonal similarity transform. In this particular case $G(\cdot)$ is a monotonously increasing convex function. The advantage of this particular criterion is that its optimal value will simply be $\zeta_1^* = \zeta(\Phi_N C_N \Phi_N^T) = |C_N|^2$ and can be evaluated without requiring the knowledge of the eigenvalues of C_N. The normalized coefficient ξ_1 which is often used as stop criterion in iterative eigenvalue extraction methods [4], is given by

$$\xi_1(U_N C_N U_N^H) = |\Lambda_N^{(u)}|^2 / |C_N|^2. \tag{8}$$

b) Bit rate

The optimal bit rate achievable in coding the transformed coefficient of a gaussian process as N independent sources is [5]

$$\zeta_2(U_N C_N U_N^H) = \sum_{i=1}^{N} \log(u_i^T C_N u_i^* / D); \tag{9}$$

where $D < \min\{u_i^T C_N u_i^*\}$ is the mean square error. Here $G(\cdot)$ is a monotonously increasing concave function.

c) Entropy

The trace of a matrix is invariant under orthogonal similarity transform and corresponds, in the case of a covariance matrix, to the average squared norm of the input random vector. The relative energy contribution of the transformed coefficients can be combined in a entropy measure [8].

$$\zeta_3(U_N C_N U_N^H) = -\sum_{i=1}^{N} \gamma_i \log(\gamma_i) \quad \text{with} \quad 0 \leq \gamma_i = u_i^T C_N u_i^* / \mathrm{tr}(C_N) \leq 1. \tag{10}$$

The minimization of this criterion ($G(\cdot)$ is a monotonously decreasing convex function) by the KLT leads to an interesting interpretation. Among all unitary transforms, the KLT is the one which produces coefficients with the most spread out variance distribution.

Another method of comparing different transforms is to evaluate their ability in data compression. If the transformed coefficients are ranked in a decreasing order of their variances $\sigma_1^2 > \sigma_2^2 > \sigma_3^2 > \cdots > \sigma_N^2$ with $\sigma_i^2 = u_i^T C_N u_i^*$, then

$$J_m(U_N C_N U_N^H) = \sum_{i=m+1}^{N} \sigma_i^2 \Big/ \sum_{k=1}^{N} \sigma_k^2 \tag{11}$$

is called the 'basis restriction error' [7]. It represents the normalized mean square error between the original random vector and an approximation, computed from m of its most significant transformed coefficient. This quantity is minimum for the KLT for any $m = 1, \ldots, N$.

3. Symmetric/skew-symmetric transform family

A symmetric Toeplitz matrix is a special member of a more general family of matrices: doubly symmetric or persymmetric matrices. Interesting properties of such matrices, which are symmetric along both main diagonals, have been investigated in [9, 10]. A fundamental property is that orthogonal transforms diagonalizing such matrices (in particular symmetric Toeplitz matrices) must be members of the Symmetric/Skew Symmetric (S/SS) orthogonal transform family introduced below.

Let $\boldsymbol{J}_N$ be a $N \times N$ permutation operator that reverses the rows (resp. columns) when applied to the left (resp. right)

$$\boldsymbol{J}_N = \begin{bmatrix} 0 & \cdots & 1 \\ & 1 & \\ 1 & \cdots & 0 \end{bmatrix}. \tag{12}$$

An $N \times N$ square transform matrix $\boldsymbol{U}_N$ (or any transform with identical but permuted row vectors)

$$\boldsymbol{U}_N = [\boldsymbol{u}_1 \boldsymbol{u}_2 \cdots \boldsymbol{u}_p \mid \boldsymbol{v}_1 \boldsymbol{v}_2 \cdots \boldsymbol{v}_Q]^{\mathrm{T}}, \tag{13}$$

where $P = [(N+1)/2]$ and $Q = [N/2]$ ($[x]$ denotes the smallest integer greater or equal to x), is a member of the S/SS orthogonal transform family when it has P symmetric and orthogonal row vectors

$$\boldsymbol{u}_i = \boldsymbol{J}_N \boldsymbol{u}_i \quad \text{and} \quad \boldsymbol{u}_i^{\mathrm{T}} \cdot \boldsymbol{u}_j = \delta_{i,j}, \quad (i, j = 1, \ldots, P) \tag{14}$$

as well as Q skew-symmetric and orthogonal row vectors

$$\boldsymbol{v}_i = -\boldsymbol{J}_N \boldsymbol{v}_i \quad \text{and} \quad \boldsymbol{v}_i^{\mathrm{T}} \cdot \boldsymbol{v}_j = \delta_{i,j}, \quad (i, j = 1, \ldots, Q). \tag{15}$$

The transforms belonging to this family have the following interesting properties.

Property 1: the transform matrix $\boldsymbol{U}_N$ is orthogonal: $\boldsymbol{U}_N \cdot \boldsymbol{U}_N^{\mathrm{T}} = \boldsymbol{I}_N$. This property is simply due to the fact that any symmetric vector will always be orthogonal to any skew symmetric vector.

Property 2: Let $\boldsymbol{C}_N$ be a persymmetric covariance matrix such as $\boldsymbol{C}_N = \boldsymbol{J}_N \boldsymbol{C}_N \boldsymbol{J}_N$. The application of the transform $\boldsymbol{U}_N$ to a random vector with covariance matrix $\boldsymbol{C}_N$ will result in mutually decorrelated symmetric and skew-symmetric components:

$$\boldsymbol{U}_N \cdot \boldsymbol{C}_N \cdot \boldsymbol{U}_N^{\mathrm{H}} = \left[\begin{array}{c|c} \boldsymbol{C}_P^+ & \boldsymbol{O}_{P,Q} \\ \hline \boldsymbol{O}_{Q,P} & \boldsymbol{C}_Q^- \end{array}\right], \tag{16}$$

where $\boldsymbol{O}_{P,Q}$ denotes a $P \times Q$ null matrix.

Property 3 [10]: Let $\boldsymbol{C}_N$ be a persymmetric matrix with distinct eigenvalues. Then $\boldsymbol{C}_N$ has $P = [(N+1)/2]$ symmetric and $Q = [N/2]$ skew symmetric eigenvectors. The eigenvector transform matrix is therefore a member of the S/SS transform family. In the case where the eigenvectors are not distinct, it is always possible to find a S/SS transform that diagonalizes $\boldsymbol{C}_N$.

Other results concerning the eigenvectors of symmetric Toeplitz matrices may be found in [11]. Properties 2 and 3 are very important for our purpose, which is the approximation of the Karhunen-Loève transform of a stationary process. The choice of a S/SS transform will mostly result in a performance improvement when compared with other closely related transforms not sharing this property. Another important point is that a transform not belonging to this family can generally not (unless the eigenvalues are multiple) diagonalize a symmetric Toeplitz matrix.

4. A circular decomposition for symmetric Toeplitz matrices

In this section, a decomposition of a symmetric Toeplitz matrix is derived. This decomposition leads to a sum of two matrices. Two kinds of transforms, which will be introduced later on, diagonalize either one of these matrices fully or both partially.

Consider a symmetric $N \times N$ Toeplitz matrix $\boldsymbol{C}_N$ completely characterized by its first row elements $c_0, \ldots, c_{N-1}$:

$$\boldsymbol{C}_N = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & & c_{N-1} \\ c_1 & c_0 & c & \cdots & & c_{N-2} \\ & \cdot & \cdot & \cdot & & \\ c_{N-2} & \cdot & & c_1 & c_0 & c_1 \\ c_{N-1} & \cdot & & & c_1 & c_0 \end{bmatrix} = \text{Toeplitz}(c_0, c_1, \ldots, c_{N-1}) \tag{17}$$

and define the corresponding symmetric Toeplitz matrix obtained in reversing the non diagonal elements of the first row:

$$\boldsymbol{D}_N = \begin{bmatrix} c_0 & c_{N-1} & c_{N-2} & \cdots & & c_1 \\ c_{N-1} & c_0 & c_{N-1} & \cdots & & c_2 \\ & \cdot & \cdot & \cdot & & \\ c_2 & \cdot & & c_{N-1} & c_0 & c_{N-1} \\ c_1 & \cdot & & & c_{N-1} & c_0 \end{bmatrix} = \text{Toeplitz}(c_0, c_{N-1}, \ldots, c_1). \tag{18}$$

We now define the matrices

$$\boldsymbol{A}_N = \tfrac{1}{2}(\boldsymbol{C}_N + \boldsymbol{D}_N) = \text{Toeplitz}(a_0, a_1, \ldots, a_{N-1}), \tag{19}$$

$$\boldsymbol{B}_N = \tfrac{1}{2}(\boldsymbol{C}_N - \boldsymbol{D}_N) = \text{Toeplitz}(b_0, b_1, \ldots, b_{N-1}). \tag{20}$$

It is easily verified that

$$a_i = \tfrac{1}{2}(c_i + c_{N-i}), \quad (i = 1, \ldots, N-1); \quad a_0 = c_0, \tag{21}$$

$$b_i = \tfrac{1}{2}(c_i - c_{N-i}), \quad (i = 1, \ldots, N-1); \quad b_0 = 0. \tag{22}$$

A circular decomposition of the matrix $\boldsymbol{C}_N$ is obtained by

$$\boldsymbol{C}_N = \boldsymbol{A}_N + \boldsymbol{B}_N, \tag{23}$$

where $\boldsymbol{A}_N$ is a circulant matrix [9] and $\boldsymbol{B}_N$ in a form that we define as skew circulant. These matrices are both symmetric Toeplitz and have the following symmetry and skew symmetry properties:

$$a_i = a_{N-i}, \qquad (i = 1, \ldots, N-1), \tag{24}$$

$$b_i = -b_{N-i}, \quad (i = 1, \ldots, N-1). \tag{25}$$

The interest of the decomposition given by (23) is that it is possible to give an expression for the eigenvalues and eigenvectors of the matrices A_N and B_N.

Let us define the permutation matrices

$$P_N = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & 0 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & & 0 & 1 \\ 1 & 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix} \quad \text{and } Q_N = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & 0 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 \\ -1 & 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix}. \tag{26}$$

Circulant and skew circulant matrices can be expressed as polynomial functions of P_N and Q_N:

$$A_N = \sum_{k=0}^{N-1} a_k P_N^k \quad \text{and} \quad B_N = \sum_{k=0}^{N-1} b_k Q_N^k. \tag{27}$$

Therefore, the eigenvectors of P_N (resp. Q_N) are the eigenvectors of A_N (resp. B_N). By solving the characteristic equation for P_N and Q_N, it is found that the characteristic roots are:

$$P_N: \quad \lambda_m = \exp(\pm j2\pi(m-1)/N), \quad (m = 1, \ldots, N), \tag{28}$$

$$Q_N: \quad \mu_n = \exp(\pm j\pi(2n-1)/N), \quad (n = 1, \ldots, N), \tag{29}$$

with corresponding eigenvectors given by the complex discrete Fourier transform (DFT) for P_N (and A_N):

$$u_m(k) = \frac{1}{\sqrt{N}} \exp\{\pm j2\pi(m-1)(k-1)/N\}; \quad (k, m = 1, \ldots, N) \tag{30}$$

and the complex discrete odd Fourier transform (DOFT) for Q_N (and B_N)

$$v_m(k) = \frac{1}{\sqrt{N}} \exp\{\pm j\pi(2m-1)(k-1)/N\}; \quad (k, m = 1, \ldots, N). \tag{31}$$

From equation (27), it follows that the eigenvalues of A_N and B_N are given by

$$A_N: \quad \lambda_i^{(a)} = \sum_{k=0}^{N-1} a_k \lambda_i^k = \sum_{k=0}^{N-1} a_k \exp(\pm j2\pi(i-1)k/N), \tag{32}$$

$$B_N: \quad \mu_i^{(b)} = \sum_{k=0}^{N-1} b_k \mu_i^k = \sum_{k=0}^{N-1} b_k \exp(\pm j\pi(2i-1)k/N). \tag{33}$$

Using the additional properties (24) and (25), it can be shown that

$$\lambda_i^{(a)} = \lambda_{N-i}^{(a)} = \sum_{k=0}^{N-1} a_k \cos\{2\pi(i-1)k/N\}, \tag{34}$$

$$\mu_i^{(b)} = \mu_{N-i-1}^{(b)} = \sum_{k=0}^{N-1} b_k \cos\{\pi(2i-1)k/N\}, \tag{35}$$

Hence, the complex conjugate eigenvectors of A_N and B_N will correspond to identical eigenvalues.

The circular decomposition of a Toeplitz matrix into two matrices with known eigenvectors does unfortunately not solve the problem of the eigenvector extraction of a symmetric Toeplitz matrix which, in most cases, can only be solved by applying numerical iterative methods. Nevertheless, the DFT or DOFT will diagonalize the covariance matrix C_N in the extreme cases where it is circulant or skew-circulant. It is therefore suggested to use either the DFT or the DOFT as a substitute of the exact eigen-decomposition (or Karhunen–Loève transform). The approximation of the KLT by the DFT (resp. DOFT) is particularly effective when the matrix B_N (resp. A_N) given by the circular decomposition is close to diagonal.

Considering an aribtrary positive definite Toeplitz matrix C_N, it will be shown next that it is possible to obain real orthogonal transforms with improved performances, by simple combination of the basis vectors of the DFT and DOFT.

5. Real transforms for the approximation of the KLT

The complex even and odd Fourier transforms (DFT and DOFT) exhibit a conjugate symmetry:

$$\boldsymbol{u}_i = \boldsymbol{u}^*_{N-i+2}, \quad (m=2,\ldots,N), \tag{36}$$

$$\boldsymbol{v}_i = \boldsymbol{v}^*_{N-i}, \quad (m=1,\ldots,N-1). \tag{37}$$

Considering an arbitrary complex vector $\boldsymbol{w}$ and its complex conjugate $\boldsymbol{w}^*$, we have that:

$$\boldsymbol{w}^{\mathrm{T}} C_N \boldsymbol{w}^* = (\boldsymbol{w}^*)^{\mathrm{T}} C_N \boldsymbol{w}. \tag{38}$$

Thus, for any covariance matrix C_N, the variances associated with the projection of the input data on two complex conjugate vectors of the DFT or DOFT are the same. This property will result in a relatively important residual correlation between the corresponding couples of transform coefficients. It is shown next, that closely related real transforms can be used to obtain better performances than the previously introduced complex transforms (DFT and DOFT).

5.1. Real even and odd Fourier transforms

The DFT (resp. DOFT) are not the only matrices to diagonalize A_N (resp. B_N) since the eigenvalues are not unique and can be grouped as $N/2$ equal pairs of eigenvalues (with complex conjugate eigenvectors). It is therefore possible to construct real transforms that also diagonalize A_N (resp. B_N) in choosing suitable linear combinations between all pairs of complex conjugate basis vectors. Consider a complex basis vector $\boldsymbol{u}_i$ and its complex conjugate $\boldsymbol{u}_i^*$. The real vectors $\boldsymbol{r}_i$ and $\boldsymbol{s}_i$ as

$$\boldsymbol{r}_i = \frac{1}{\sqrt{2}} \cdot \{\exp(\mathrm{j}\theta_i) \cdot \boldsymbol{u}_i + \exp(-\mathrm{j}\theta_i) \cdot \boldsymbol{u}_i^*\}, \tag{39}$$

$$\boldsymbol{s}_i = \frac{-\mathrm{j}}{\sqrt{2}} \cdot \{\exp(\mathrm{j}\theta_i) \cdot \boldsymbol{u}_i - \exp(-\mathrm{j}\theta_i) \cdot \boldsymbol{u}_i^*\}, \tag{40}$$

where θ_i is an arbitrary phase, can be shown to be orthogonal and to form a basis of the plane spanned by $\boldsymbol{u}_i$ and $\boldsymbol{u}_i^*$. Using this property, it is possible to generate a family of sinusoidal/cosinusoidal orthogonal transforms that diagonalize A_N or B_N. If, for example, the θ_i's are chosen to be zero, the associated transform coefficients will correspond to the real and imaginary parts of the DFT (or DOFT). Of particular interest is the case where the θ_i's are chosen so that the vectors $\boldsymbol{r}_i$ are symmetric and the vectors $\boldsymbol{s}_i$ are skew symmetric. The corresponding transforms—referred to as the discrete real even Fourier transform (DREFT) and the discrete real odd Fourier transform (DROFT)—are member of the S/SS transform family. The associated basis vectors may be found in Table 1.

For an arbitrary covariance matrix C_N, the real sinusoidal/cosinusoidal transforms will generally lead to an improvement in performances when compared to the corresponding complex transforms. This can

Table 1

Basis vectors of the principal complex and real transforms obtained from the circular decomposition of a symmetric Toeplitz matrix

Transform	Basis vectors	
	$u_m(k)$; $m=1,\dots,N$; $k=1,\dots,N$ or $u_p(k)$ and $u_q(k)$; $p=1,\dots,[(N+1)/2]$, $q=1,\dots,[N/2]$	
Discrete Fourier Transform (DFT)	$\frac{1}{\sqrt{N}}\exp\left\{\pm j\frac{2\pi(m-1)(k-1)}{N}\right\}$ with $j=\sqrt{-1}$	
Discrete Odd Fourier Transform (DOFT)	$\frac{1}{\sqrt{N}}\exp\left\{\pm j\frac{\pi(2m-1)(k-1)}{N}\right\}$	
Discrete Real Even Fourier Transform (DREFT) (symmetric/skew symmetric)	$\frac{1}{\sqrt{N}}$; $p=1$, $k=1,\dots,N$	$\sqrt{\frac{2}{N}}\sin\left\{\frac{(2k-1)q\pi}{N}\right\}$; $q=1,\dots,[(N-1)/2]$
	$\sqrt{\frac{2}{N}}\cos\left\{\frac{(2k-1)(p-1)\pi}{N}\right\}$; $p=2,\dots,[(N+1)/2]$	$\frac{1}{\sqrt{N}}\sin\left\{(2k-1)\frac{\pi}{2}\right\}$; $q=N/2$, (if N even)
Discrete Real Odd Fourier Transform (DROFT) (symmetric/skew symmetric)	$\sqrt{\frac{2}{N}}\sin\left\{\frac{(2k-1)(2p-1)\pi}{2N}\right\}$; $p=1,\dots,[N/2]$	$\sqrt{\frac{2}{N}}\cos\left\{\frac{(2k-1)(2q-1)\pi}{2N}\right\}$; $q=1,\dots,[N/2]$
	$\frac{1}{\sqrt{N}}\sin\left\{(2k-1)\frac{\pi}{2}\right\}$; $p=[(N+1)/2]$, (if N odd)	
Discrete Cosine Transform (DCT)	$\frac{1}{\sqrt{N}}$; $m=1$, $k=1,\dots,N$ $\sqrt{\frac{2}{N}}\cos\left\{\frac{(2k-1)(m-1)\pi}{2N}\right\}$; $m=1,\dots,N$	
Discrete Even Sine Transform (DEST)	$\sqrt{\frac{2}{N}}\sin\left\{\frac{(2k-1)m\pi}{2N}\right\}$; $m=1,\dots,N-1$ $\frac{1}{\sqrt{N}}\sin\left\{(2k-1)\frac{\pi}{2}\right\}$; $m=N$	

be shown in the following way. The general criterion function introduced in Section 2 can be decomposed in a sum of sub-criterions associated with each plane spanned by a complex vector and its conjugate. Following the same reasoning as in appendix A, it is easily shown that the optimal value of the sub-criterion is obtained when the projections on the two associated basis vectors are decorrelated. On the other hand, the worst value of the sub-criterion is achieved when the two variances are the same—which is precisely the case when the complex transform is used (DFT or DOFT). These results can be summarized as:

Property 4: All real sinusoidal/cosinusoidal orthogonal transforms obtained in grouping the conjugate vectors of an orthogonal complex transform will perform as well, or better (in the sense of a family of performance criterions) as the original complex transform. The best possible performance is achieved by the associated symmetric/skew-symmetric real transform that decorrelates all pairs of grouped components.

This is illustrated in Fig. 1 which shows the improvement of performances obtained by using the DREFT as opposed to the real Fourier transform (real and imaginary parts) and the complex DFT for different correlation values of a first order Markov process. The performance criterion used is the entropy coefficient—the same type of graph can, of course, be obtained with any criterion function belonging to the family introduced in Section 2. The basis restriction errors obtained with these transforms for a first order Markov process with correlation 0.9, using 16 coefficients, are shown in Fig. 2. As expected, the best performances are obtained with the DREFT.

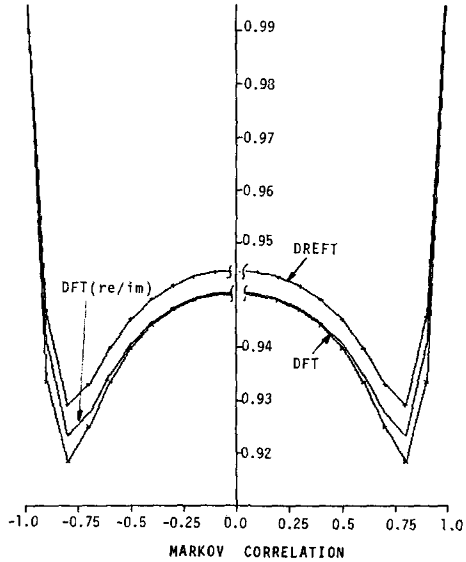

Fig. 1. Normalized entropy coefficient of the real and complex even Fourier transforms as a function of the correlation for a first order Markov process of duration $N = 16$.

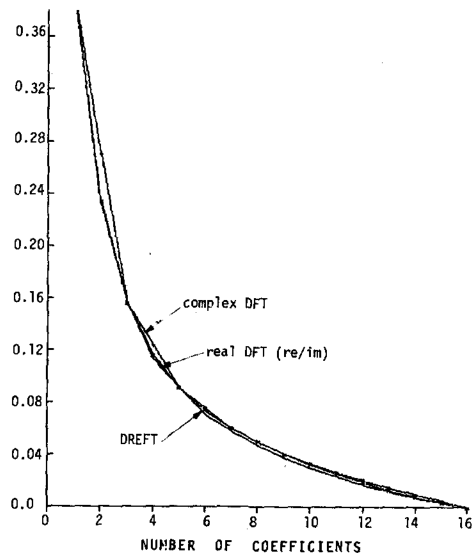

Fig. 2. Basis restriction error of the real and complex even Fourier transforms for a first order Markov process of duration $N = 16$ and correlation $\rho = 0.9$.

5.2. Related cosine and sine transforms

The previously introduced transforms share the property of diagonalizing either $\boldsymbol{A}_N$ or $\boldsymbol{B}_N$ obtained from the circular decomposition of a symmetric Toeplitz matrix. Among these transforms, the DREFT and DROFT are shown to be the most interesting in the sense that they lead to the best performances in approximating the optimal Karhunen-Loève expansion. These transforms belong to the S/SS transform family, and it is therefore possible to use property 1 to form two closely related transforms obtained in combining the symmetric basis vectors of the DREFT with the skew-symmetric basis vectors of the DROFT, and vice versa. These new transforms are also members of the S/SS transform family and have the interesting property of partially diagonalizing $\boldsymbol{A}_N$ and $\boldsymbol{B}_N$ and can be predicted to be interesting candidates for the approximation of the KLT. The corresponding basis vectors are given in Table 1. It is very interesting to note that the first transform is the well known discrete cosine transform (DCT) that has initially been proposed by Ahmed *et al.* [6]. The second transform has been introduced by Jain [7] as an interesting member of a very general sinusoidal transform family and has been referred to as the discrete even sine transform (DEST). The DCT is the most popular transform used to approximate the KLT transform of a wide sense stationary process and is known to provide excellent results when applied to the analysis of a process with an essentially lowpass spectral power density function. The DCT corresponds almost to the KLT of a first order markov process with correlation coefficient 0.9, see for example [6] and [12].

5.3. Performance comparison

In order to compare these various transforms, we have considered the representation of a first order Markov process with correlation coefficient ρ. The associated $N \times N$ covariance matrix is given by eq. (17) with

$$c_k = \sigma^2 \rho^k, \tag{41}$$

where σ^2 is the variance of each component. The KLT of such a process is a member of the sinusoidal transform family introduced by [7]. The frequencies associated with the eigenvectors are generally non-harmonic and can be obtained from the solution of a transcendental equation [13]. The effectiveness of the DREFT, DROFT, DCT and DEST for the representation of a N-component vector with first order Markov statistics are compared in Figs. 3 and 4 on the basis of the normalized entropy coefficient (eq. (6) and (9)) and the energy criterion (eq. (8)) computed for different values of ρ. It is interesting to note that the performances of the DCT and DEST depend on the sign of the correlation coefficient. The DCT provides a very close approximation of the KLT for positive correlation values close to one but behaves very poorly when ρ is negative—this situation is reversed for the DEST. The DREFT and DROFT have symmetric characteristics and can be seen to be less performing than the DCT or DEST (depending on the sign of ρ) but present the advantage of having a more constant behavior (less dependent on ρ). The DREFT should generally be prefered to the DROFT unless ρ is known to be small. For comparison, the performances obtained with the discrete sine transform (DST), which is known to provide an excellent approximation of the KLT for small values of ρ, have been included [14]. This particular transform performs better than the others for $|\rho| < 0.5$. The basis restriction error associated with these different transforms for $N = 16$ and $\rho = 0.9$ is shown in Fig. 5. In this particular case, the DCT can be seen to be an almost perfect approximation of the KLT. When selecting a transform for a particular application, it should be kept in mind that it is generally more interesting to choose a transform that provides a close approximation of the KLT for strongly correlated data (e.g. DCT or DREFT). In the case of weakly correlated data, there is not much gain to be expected from one representation to another.

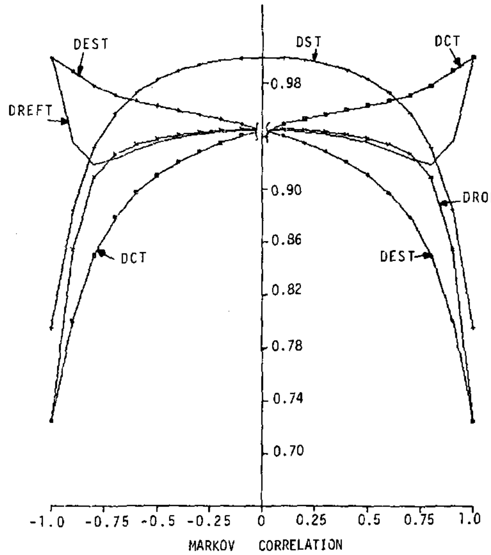

Fig. 3. Normalized entropy coefficient of various sinusoidal transforms as a function of the correlation for a first order Markov process of duration $N = 16$.

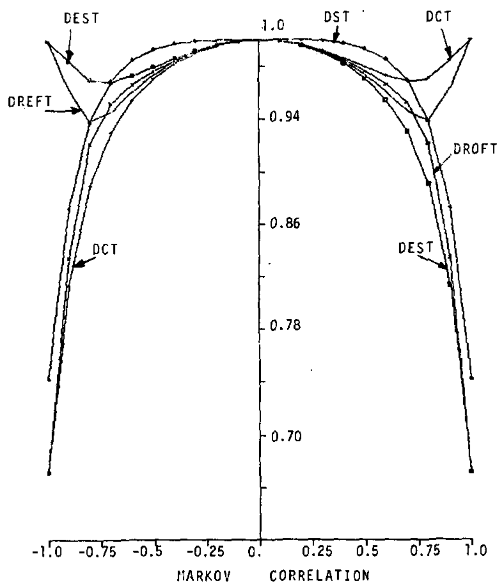

Fig. 4. Normalized energy criterion of various sinusoidal transforms as a function of the correlation for a first order Markov process of duration $N = 16$.

5.4. Computational considerations

An important advantage of selecting a suboptimal sinusoidal transform obtained from the circular decomposition of a symmetric Toeplitz matrix as a substitute of the KLT is, that it is possible to compute the transformed coefficients using a fast algorithm with a number of operations of the order of $N \log(N)$. This type of algorithm can be designed from the well known FFT algorithm using some auxiliar computa-

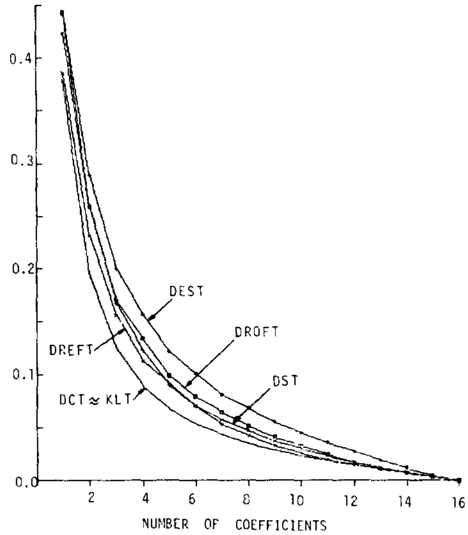

Fig. 5. Basis restriction error of various sinusoidal transforms for a first order Markov process of duration $N = 16$ and correlation $\rho = 0.9$.

tions [7]. It is also possible to take full advantage of the real harmonic structure of these transforms and to obtain specialized algorithms with a significant improvement in the number of operations over the classical schemes using the FFT [15].

6. Asymptotic equivalence

In terms of some performance criterion, all the previously introduced sinusoidal/cosinusoidal transforms can be shown to be asymptotically equivalent to the optimal KLT transform in the case of an arbitrary wide sense stationary process. To establish this equivalence, let us first consider some definitions.

6.1. Asymptotic equivalence—definition

Two matrices $\boldsymbol{A}_N$ and $\boldsymbol{B}_N$ are said to be asymptotically equivalent if the weak (or Hilbert-Schmidt) norm of the difference vanishes as N increases

$$\lim_{N\to\infty} |\boldsymbol{A}_N - \boldsymbol{B}_N|^2 = 0. \tag{42}$$

Theorem 1 [16]. *Two asymptotically equivalent symmetric real matrices* $\boldsymbol{A}_N$ *and* $\boldsymbol{B}_N$ *with associated bounded eigenvalues* $\{\alpha_{i,N}\}$ *and* $\{\beta_{i,N}\}$ *are asymptotically equally distributed, i.e.*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} F(\alpha_{i,N}) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} F(\beta_{i,N}), \tag{43}$$

where $F(\cdot)$ *is an arbitrary continuous function.*

In most applications, the choice of the KLT is justified by the fact that it optimizes a criterion function or performance measure as defined in Section 2. Considering the very large class of functions defined by eq. (3), the performance criterion associated to an orthogonal transform U_N can be expressed as

$$\zeta(U_N C_N U_N^{\mathrm{H}}) = \frac{1}{N} \sum_{i=1}^{N} G(\mu_{i,N}) \cdot N, \tag{44}$$

where $\mu_{i,N}$ are the eigenvalues of the matrix $C_N^{(u)}$ defined as

$$C_N^{(u)} = U_N^{\mathrm{H}} \operatorname{diag}(U_N C_N U_N^{\mathrm{H}}) U_N. \tag{45}$$

The diagonal matrix $\operatorname{diag}(U_N C_N U_N^{\mathrm{H}})$ is formed by setting the non diagonal elements of $U_N C_N U_N^{\mathrm{H}}$ to zero. The matrix C_N is diagonalized by the transform U_N. From eq. (43) which holds for any continuous function $F(\cdot)$, it follows that the asymptotic equivalence between $C_N^{(u)}$ and C_N implies:

$$\lim_{N\to\infty} \zeta(U_N C_N U_N^{\mathrm{H}}) = \lim_{N\to\infty} \zeta^*(\Phi_N C_N \Phi_N^{\mathrm{T}}) = \lim_{N\to\infty} \sum_{i=1}^{N} G(\lambda_{i,N}), \tag{46}$$

where λ_i are the eigenvalues of the matrix C_N.

In such a situation, the transformation U_N will asymptotically lead to performances equivalent to the optimal KLT. The asymptotic equivalence between $C_N^{(u)}$ and C_N is a sufficient condition for (46) and therefore provides a relatively simple method to demonstrate the asymptotic performance equivalence between a transform U_N and the Karhunen–Loève expansion.

6.2. Discrete complex odd and even Fourier transforms

Let the F_N and H_N be the $N \times N$ even and odd DFT rotation matrices. The asymptotic equivalence between C_N and $C_N^{(f)} = F_N^{\mathrm{H}} \operatorname{diag}(F_N C_N F_N^{\mathrm{H}}) F_N$ has been established in [5] under the constraint of a summable covariance function. A similar procedure can be used to show the asymptotic equivalence between C_N and $C_N^{(h)} = H_N^{\mathrm{H}} \operatorname{diag}(H_N C_N H_N^{\mathrm{H}}) H_N$. This demonstrates that the performance degradation resulting from the use of the even or odd DFT (as opposed to the Karhunen–Loève transform) in coding or filtering or data representation vanishes as $N \to \infty$.

The approximation matrices $C_N^{(f)}$ and $C_N^{(h)}$ can be shown to be related to the symmetric Toeplitz matrix C_N by

$$C_{i,j}^{(f)} = \frac{N - |i-j|}{N} C_{i,j} + \frac{|i-j|}{N} C_{i,N-j}, \tag{47}$$

$$C_{i,j}^{(h)} = \frac{N - |i-j|}{N} C_{i,j} - \frac{|i-j|}{N} C_{i,N-j}. \tag{48}$$

It is easily verified that $C_N^{(f)}$ is a circulant matrix and $C_N^{(h)}$ a skew-circulant matrix. It is worthwhile to note that both matrices are symmetric Toeplitz. As it has been shown previously, $C_N^{(f)}$ is diagonalized by F_N and by any real DFT obtained by choosing a suitable linear combination between all pairs of complex conjugate vectors of F_N with associated identical eigenvalues. The same remark is valid for $C_N^{(h)}$ which is diagonalized by the complex odd DFT or any real odd DFT.

6.3. Discrete real odd and even Fourier transform

It has previously been shown that the use of any real odd or even orthogonal Fourier transform will always result in an improvement of the performance criterion when compared to the corresponding complex transform. For this reason, the asymptotic performance equivalence of the complex even or odd

Fourier transform with the Karhunen-Loève expansion is also valid for the corresponding real transforms. The improvement of the performance criterion (e.g. entropy coefficient) as N increases is demonstrated in Fig. 6 for the particular case where the DREFT is applied to a first order Markov process.

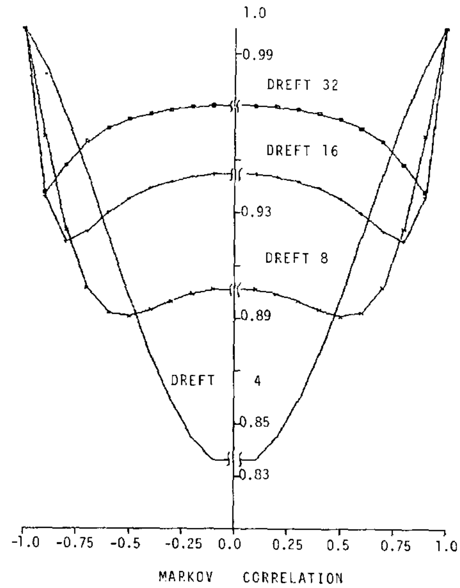

Fig. 6. Asymptotic convergence of the normalized entropy coefficient of the DREFT as a function of the correlation for a first order Markov process.

Let G_N be the $N \times N$ rotation matrix associated to a real even orthogonal Fourier transform. The asymptotic equivalence of C_N and $C_N^{(g)}$ is a stronger condition than the asymptotic performance equivalence of G_N and the KLT and will be proved next. The Hilbert-Schmidt norm between C_N and $C_N^{(g)}$ which is invariant under any similarity transform can be rewritten as

$$|C_N - C_N^{(g)}|^2 = |C_N|^2 - |\mathrm{diag}(G_N C_N G_N^{\mathrm{H}})|^2. \tag{49}$$

Let us recall that $|\mathrm{diag}(G_N C_N G_N^{\mathrm{H}})|^2$ is the energy performance criterion associated to the transform G_N and is a member of the criterion family introduced in Section 2. A particular implication of property 4 is that

$$|\mathrm{diag}(G_N C_N G_N^{\mathrm{H}})|^2 \geq |\mathrm{diag}(F_N C_N F_N^{\mathrm{H}})|^2. \tag{50}$$

This equation implies that

$$|C_N - C_N^{(g)}|^2 \leq |C_N - C_N^{(f)}|^2. \tag{51}$$

Taking the limit as $N \to \infty$

$$\lim_{N\to\infty} |C_N - C_N^{(g)}|^2 \leq \lim_{N\to\infty} |C_N - C_N^{(f)}|^2 = 0, \tag{52}$$

which proves the asymptotic equivalence of any symmetric Toeplitz matrix C_N and $C_N^{(g)}$. The same proof is also valid when G_N is a real odd Fourier transform and when F_N is substituted by H_N.

6.4. Discrete even cosine and sine transforms

In order to provide a simple proof for the asymptotic performance equivalence of the DCT, the DEST and the KLT, it is necessary to introduce the following theorem (Appendix B).

Theorem 2. *The eigenvalues of two asymptotically equivalent persymmetric real matrices $\boldsymbol{A}_N$ and $\boldsymbol{B}_N$ which are strongly bounded (bounded eigenvalues) can be divided into two sets of asymptotically equally distributed eigenvalues*

$$\lim_{N\to\infty} \frac{2}{N} \sum_{i=1}^{N/2} F(\alpha_{i,N}^{+}) = \lim_{N\to\infty} \frac{2}{N} \sum_{i=1}^{N/2} F(\beta_{i,N}^{+}), \tag{53}$$

$$\lim_{N\to\infty} \frac{2}{N} \sum_{i=1}^{N/2} F(\alpha_{i,N}^{-}) = \lim_{N\to\infty} \frac{2}{N} \sum_{i=1}^{N/2} F(\beta_{i,N}^{-}), \tag{54}$$

where $\alpha_{i,N}^{+}$ and $\beta_{i,N}^{+}$ are the eigenvalues corresponding to the symmetric eigenvectors of $\boldsymbol{A}_N$ and $\boldsymbol{B}_N$ and $\alpha_{i,N}^{-}$ and $\beta_{i,N}^{-}$ are the eigenvalues corresponding to the skew symmetric eigenvectors of $\boldsymbol{A}_N$ and $\boldsymbol{B}_N$. $F(\cdot)$ is an arbitrary continuous function.

Let us consider the symmetric/skew symmetric real odd and even Fourier transforms. Applying Theorem 2, and considering an arbitrary function $G(\cdot)$ satisfying the requirements introduced in Section 3, we have:

$$\lim_{N\to\infty} \sum_{i=1}^{N/2} G(\alpha_{i,N}^{+}) = \lim_{N\to\infty} \sum_{i=1}^{N/2} G(\beta_{i,N}^{+}) = \lim_{N\to\infty} \sum_{i=1}^{N/2} G(\mu_{i,N}^{+}), \tag{55}$$

$$\lim_{N\to\infty} \sum_{i=1}^{N/2} G(\alpha_{i,N}^{-}) = \lim_{N\to\infty} \sum_{i=1}^{N/2} G(\beta_{i,N}^{-}) = \lim_{N\to\infty} \sum_{i=1}^{N/2} G(\mu_{i,N}^{-}), \tag{56}$$

with

$$\lim_{N\to\infty} \sum_{i=1}^{N/2} G(\mu_{i,N}^{+}) + \lim_{N\to\infty} \sum_{i=1}^{N/2} G(\mu_{i,N}^{-}) = \lim_{N\to\infty} \zeta_{\mathrm{KLT}}^{*}(\boldsymbol{C}_N), \tag{57}$$

where $\alpha_{i,N}^{+}$, $\beta_{i,N}^{+}$, $\mu_{i,N}^{+}$ (resp. $\alpha_{i,N}^{-}$, $\beta_{i,N}^{-}$, $\mu_{i,N}^{-}$) are the eigenvalues of $\boldsymbol{C}_N^{(f)}$, $\boldsymbol{C}_N^{(h)}$ and $\boldsymbol{C}_N$ corresponding to the symmetric (resp. skew-symmetric) eigenvectors of these matrices. Recalling that the DCT (resp. DEST) can be obtained by taking the symmetric vectors of the DREFT (resp. DROFT) and the skew-symmetric basis vectors of the DROFT (resp. DREFT), it follows immediately that

$$\lim_{N\to\infty} \zeta_{\mathrm{DCT}}(\boldsymbol{C}_N) = \lim_{N\to\infty} \sum_{i=1}^{N/2} G(\alpha_{i,N}^{+}) + \lim_{N\to\infty} \sum_{i=1}^{N/2} G(\beta_{i,N}^{-}) = \lim_{N\to\infty} \zeta_{\mathrm{KLT}}^{*}(\boldsymbol{C}_N), \tag{58}$$

$$\lim_{N\to\infty} \zeta_{\mathrm{DEST}}(\boldsymbol{C}_N) = \lim_{N\to\infty} \sum_{i=1}^{N/2} G(\alpha_{i,N}^{-}) + \lim_{N\to\infty} \sum_{i=1}^{N/2} G(\beta_{i,N}^{+}) = \lim_{N\to\infty} \zeta_{\mathrm{KLT}}^{*}(\boldsymbol{C}_N). \tag{59}$$

For an arbitrary wide sense stationary process, this proves that the DCT and DEST are asymptotically equivalent in performances to the optimal Karhunen-Loève transform. This equivalence was already known to hold in the case of some restricted families of stationary processes such as first order Markov processes ([12] for the DCT) and finite p-order Markov processes ([7] for the general sinusoidal transform family including the DCT and DEST). The asymptotic convergence of the normalized entropy coefficient of the DCT, when applied to the representation of a first order Markov process, is illustrated in Fig. 7.

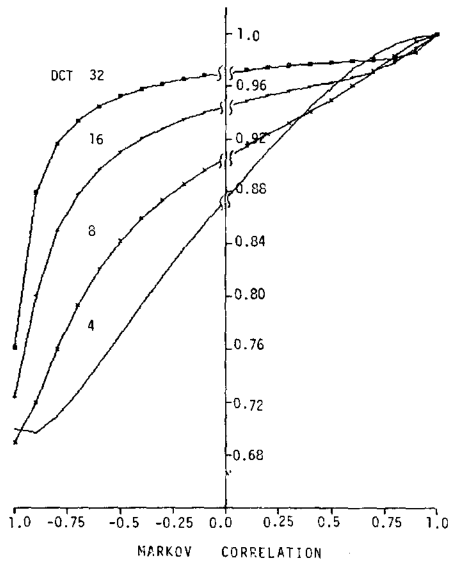

Fig. 7. Asymptotic convergence of the normalized entropy coefficient of the DCT as a function of the correlation for a first order Markov process.

The results established in this section are more general than those previously reported in the literature. It should also be noted that the mathematics involved are significantly simpler than what can be found in [12] and [7].

7. Conclusion

A new approach to the approximation of the Karhunen–Loève transform of a wide sense stationary process has been presented. It is based on a decomposition of a symmetric Toeplitz matrix in the sum of a circulant and a skew circulant matrix. Various transforms (DFT, DOFT, DREFT, DROFT, DCT, DEST) have been shown to be closely related to this decomposition and can therefore provide efficient substitutes of the KLT. These transforms have been compared on the basis of a general family of performance criterions that are optimized by the KLT. All these transforms can be computed using fast algorithm and are therefore useful in many signal processing applications such as data compression and pattern recognition. No particular restriction has been made on the class of the stationary processes whose KLT can be approximated by one of these transforms. Their assymptotic performance equivalence with the optimal KLT has been shown to be valid for an arbitrary stationary wide sense process.

Appendix A: Optimality of the Karhunen–Loève transform

Let $G(\cdot)$ be a monotonously concave or convex function. The problem is to find an extremum of the performance criterion

$$\zeta(\boldsymbol{U}_N\boldsymbol{C}_N\boldsymbol{U}_N^{\mathrm{T}}) = \sum_{i=1}^{N} G(\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{C}_N\boldsymbol{u}_i) \tag{a1}$$

under the constraints that U_N is an $N \times N$ energy preserving transform

$$\sum_{i=1}^{N} u_i^T C_N u_i = \mathrm{Tr}(C_N) \tag{a2}$$

and that its row vectors are normalized:

$$u_i^T \cdot u_i = 1, \quad (i = 1, \ldots, N). \tag{a3}$$

Applying the Lagrange multiplier method, we define the function

$$L(U_N C_N U_N^T) = \sum_{i=1}^{N} G(u_i^T C_N u_i) - \alpha \left(\sum_{i=1}^{N} u_i^T C_N u_i - \mathrm{tr}(C_N) \right) + \sum_{i=1}^{N} \beta_i (u_i^T u_i - 1). \tag{a4}$$

The optimum transform is obtained by solving:

$$\nabla_{u_i} \{L(U_N C_N U_N^T)\} = 2 C_N u_i \frac{\partial G}{\partial \sigma_i^2} - 2\alpha C_N u_i - 2\beta_i u_i = 0, \quad \text{where } \sigma_i^2 = u_i^T C_N u_i. \tag{a5}$$

Two distinct cases have to be considered:

1) $\partial G / \partial \sigma_i^2 - \alpha = 0$,

which in conjunction with constraint (a2) gives that

$$u_i^T C_N u_i = u_j^T C_N u_j = \mathrm{Tr}(C_N)/N, \quad \forall i, j. \tag{a6}$$

It can be verified that ζ is minimum (resp. maximum) if $G(\cdot)$ is a monotonously increasing convex or decreasing concave (resp. monotonously increasing concave or decreasing convex) function:

$$\min\{\zeta(U_N C_N U_N^T)\} = N \cdot G(\mathrm{tr}(C_N)/N). \tag{a7}$$

2) $\partial G / \partial \nabla_i^2 - \alpha \neq 0$.

In this case, taking into account eq. (a3), (a5) can be rewritten as

$$C_N u_i = (u_i^T C_N u_i) \cdot u_i = \beta_i u_i. \tag{a8}$$

The u_i's solution of this equation are the eigenvectors of the matrix C_N with corresponding eigenvalues β_i $(i = 1, \ldots, N)$. This case clearly corresponds to a maximum (resp. minimum) when $G(\cdot)$ is monotonously increasing convex or decreasing concave (resp. monotonously increasing concave or decreasing convex) giving:

$$\max(\zeta(U_N C_N U_N^T) = \sum_{i=1}^{N} G(\beta_i). \tag{a9}$$

Appendix B: Asymptotic equivalence for persymmetric real matrices

Let us introduce the unitary transform

$$V_N = \frac{1}{\sqrt{2}} \left[\begin{array}{c|c} I_{N/2} & -J_{N/2} \\ \hline J_{N/2} & I_{N/2} \end{array} \right],$$

which, when applied to a even persymmetric matrix $\boldsymbol{A}_N$, produces two half dimension matrices

$$\boldsymbol{V}_N\boldsymbol{A}_N\boldsymbol{V}_N^{\mathrm{T}}=\left[\begin{array}{c|c}\boldsymbol{A}_{N/2}^{+} & \boldsymbol{O}_{N/2}\\ \hline \boldsymbol{O}_{N/2} & \boldsymbol{A}_{N/2}^{-}\end{array}\right],$$

where $\boldsymbol{O}_{N/2}$ is the $N/2\times N/2$ null matrix. In the case where N is odd, a slightly modified transform matrix $\boldsymbol{V}_N$ has to be chosen.

Let $\{\alpha_{i,N}^{+}\}$, $(i=1,\ldots,N/2)$ (resp. $\{\alpha_{i,N}^{-}\}$) denote the eigenvalues of the matrix $\boldsymbol{A}_{N/2}^{+}$ (resp. $\boldsymbol{A}_{N/2}^{-}$). It can easily be shown that the $\alpha_{i,N}^{+}$'s (resp. $\alpha_{i,N}^{-}$) correspond to the eigenvalues of $\boldsymbol{A}_N$ associated to the symmetric (resp. skew symmetric) eigenvectors. The asymptotic equivalence between $\boldsymbol{A}_N$ and $\boldsymbol{B}_N$ can be rewritten as

$$\lim_{N\to\infty}|\boldsymbol{A}_N-\boldsymbol{B}_N|^2=\lim_{N\to\infty}|\boldsymbol{V}_N(\boldsymbol{A}_N-\boldsymbol{B}_N)\boldsymbol{V}_N^{\mathrm{T}}|^2=0$$

and implies that

$$\lim_{N\to\infty}|\boldsymbol{A}_{N/2}^{+}-\boldsymbol{B}_{N/2}^{+}|^2=0 \quad\text{and}\quad \lim_{N\to\infty}|\boldsymbol{A}_{N/2}^{-}-\boldsymbol{B}_{N/2}^{-}|^2=0.$$

The application of Theorem 1 to the matrices $\boldsymbol{A}_{N/2}^{+}$ and $\boldsymbol{B}_{N/2}^{+}$ (resp. $\boldsymbol{A}_{N/2}^{-}$ and $\boldsymbol{B}_{N/2}^{-}$) results in Theorem 2.

References

[1] N. Ahmed and K.R. Rao, *Orthogonal Transforms for Digital Signal Processing*, Springer, New York, 1975.

[2] K. Fukunaga and W.L. Koontz, "Application of the Karhunen-Loève expansion to feature selection and ordering", *IEEE Trans. Computers*, Vol. C-19, pp. 311-318, April 1970.

[3] W.K. Pratt, "Generalized Wiener filtering computation techniques", *IEEE Trans. on Computers*, Vol. C-21, No. 7, pp. 636-641, July 1972.

[4] *Applied Numerical Methods*, John Wiley & Sons, New York, pp. 210-268.

[5] J. Pearl, "On coding and filtering stationary signals by discrete fourier transforms", *IEEE Trans. Inform. Theory*, Vol. IT-19, pp. 229-232, March 1973.

[6] N. Ahmed, T. Natarajan and K.R. Rao, "Discrete cosine transform", *IEEE Trans. Comput.*, Vol. C-23, pp. 90-93, Jan. 1974.

[7] A.K. Jain, "A sinusoidal family of unitary transforms", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-1, No. 4, pp. 356-365, Oct. 1979.

[8] S. Watanabe, "Karhunen-Loève expansion and factor analysis", *Trans. Fourth Prague Conf. on Information Theory*, 1965, Czechoslovak Academy of Sciences, Prague, 1967.

[9] P.A. Roebuck and S. Barnett, "A survey of Toeplitz matrices and related matrices", *Int. Journal of systems sciences*", Vol. 9, No. 8, pp. 921-934, 1978.

[10] A. Cantoni and P. Butler, "Properties of The Eigenvectors of Persymmetric Matrices with applications to communication theory", *IEEE Trans. Communication* Vol. COM-24, pp. 804-809, Aug. 1976.

[11] J. Makhoul, "On the Eigenvectors of symmetric Toeplitz matrices", *IEEE Trans. on Acoustic*, Speech and Signal Processing, Vol. ASSP-29, No. 4, pp. 368-872, Aug. 1981.

[12] M. Hamidi and J. Pearl, "Comparison of the cosine and Fourier transforms of Markov-1 signals", *IEEE Trans. on Acoustic, Speech and Signal Processing*, Vol. ASSP-24, pp. 428-429, Oct. 1976.

[13] W.D. Ray and R.M. Driver, "Further decomposition of the Karhunen-Loève series representation of a stationary random process", *IEEE Trans. Information Theory*, Vol. IT-16, pp. 663-668, Nov. 1970.

[14] A.K. Jain, "A fast Karhunen-Loève transform for a class of stochastic processes", *IEEE Trans. Commun.*, Vol. COM-24, pp. 1023-1029, Sept. 1976.

[15] M. Vetterli and H. Nussbaumer, "Simple FFT and DCT algorithms with reduced number of operations", *Signal Processing*, Vol. 6, No. 4, July 1984.

[16] R.M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices", *IEEE Trans. Inform. Theory*, Vol. IT-18, pp. 725-730, Nov. 1972.