DISCRETE ORTHOGONAL FUNCTIONS APPLIED TO FILTERING AND DATA COMPRESSION

By

L. KÓNYA

Department of Automation, Technical University, Budapest

Received December 22, 1976

Presented by Prof. Dr. F. CSÁKI

Discrete orthogonal transformations provide a useful tool of computer aided signal processing. In signal processing an important problem is the separation of the signal from the noise. Another problem is the efficient storage of the signal in the computer memory. The paper gives a short review of this field.

Mathematical bases

A function $g(t)$ may be expanded into a series of orthogonal functions [4]:

$$g(t)=\sum_{i=1}^{\infty} a_i f_i(t) \tag{1}$$

For the orthogonal system of functions $\{f_i(t)\}$:

$$\frac{1}{l}\int_0^l f_i(t)\cdot f_j(t)\,\mathrm{d}t=\begin{cases}0 & \text{if} \quad i\neq j\\ 1 & \text{if} \quad i=j\end{cases} \tag{2}$$

The coefficients $\{a_i\}$ are determined by the relationship [4]:

$$a_i=\frac{1}{l}\int_0^l f_i(t)g(t)\,\mathrm{d}t \qquad i=1, 2, \ldots \tag{3}$$

where l is the range of orthogonality.

Such an expansion is e.g. the Fourier series, where the orthogonal functions are a set of trigonometric functions, and the orthogonality range is $(0, 2\pi)$ [5]. The above equations may be written in discrete matrix form by dividing $(0, l)$ into $N=2^n (n>0$, integer) subintervals, and within each interval $f_i(t)$ and $g(t)$ are supposed to be constant.

Introducing the designation $k=t/N$, we obtain the relationships

$$g(k)=\sum_{i=1}^{N} a_i f_i(k) \qquad k=1,2,\ldots,N$$

and

$$a_i=\frac{1}{N}\sum_{k=1}^{N} g(k)\cdot f_i(k) \qquad i=1,2,\ldots,N \tag{5}$$

which may be written, introducing the symbols

$$\begin{gathered}\mathbf{a}^{\mathrm{T}}=[a_1, a_2, \ldots, a_N]\\ \mathbf{g}^{\mathrm{T}}=[g(1), g(2), \ldots, g(N)]\\ \mathbf{f}_i^{\mathrm{T}}=[f_i(1), f_i(2), \ldots, f_i(N)] \qquad i=1,2,\ldots,N\end{gathered} \tag{6}$$

$$\mathbf{F}^{\mathrm{T}}=[\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N] \tag{6a}$$

in the form of

$$\mathbf{a}=\frac{1}{N}\mathbf{Fg}; \qquad \mathbf{g}=\mathbf{F}^{-1}\mathbf{a}. \tag{7}$$

This shows that the connection between the sampled functions and the expansion coefficients is given by an orthogonal transformation, which may be represented by a matrix multiplication. For the matrix $\mathbf{F}$ it is true that

$$\begin{gathered}\mathbf{FF}^{\mathrm{T}}=N\mathbf{I} \qquad \text{where } \mathbf{I} \text{ is the unity matrix}\\ \mathbf{F}^{\mathrm{T}}=\mathbf{F}^{-1} \qquad \text{where T is the symbol of transposition.}\end{gathered} \tag{8}$$

The rows and columns of $\mathbf{F}$ are orthogonal to each other which is the presentation of (2) in discrete form. What is the apparent meaning of the transformation? The original vector $\mathbf{g}$ formed from the sampled function values is a vector of an N-dimensional space. Upon orthogonal transformation it becomes a vector of another space characterized by the transformation matrix [4].

Performing the matrix multiplication required for the transformation may be accelerated by applying the fast transformation technique developed for several discrete systems of orthogonal functions, e.g. Fast Fourier Transformation (FFT), Walsh–Hadamard Transformation (WHT), Haar Transformation (HT), Discrete Cosine Transformation (DCT) [1].

Fast transformations significantly reduce the time and space requirement.

The generalized Wiener filtering

The mathematical model is shown in Fig. 1.

The input is the vector **z**, sum of signal **x** and noise **w**; **F** designates the orthogonal transformation, **A** is the filter matrix, and $\hat{\mathbf{x}}$ is the filtered vector. The problem is to find **A** so that $\hat{\mathbf{x}}$ approximates **x** best. Let us minimize the expression:

$$E\{\|\hat{\mathbf{x}}-\mathbf{x}\|^2\}=tr(E\{(\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{x}}-\mathbf{x})^T\})=\varepsilon. \tag{9}$$

Here $E\{\ldots\}$ means the expected value and tr is the trace of the matrix. From the figure if follows that:

$$\hat{\mathbf{x}}=\mathbf{F}^T\mathbf{A}\mathbf{F}\mathbf{z}. \tag{10}$$

By substituting this into (9) we have:

$$E\{\mathbf{z}^T\mathbf{F}^T\mathbf{A}^T\mathbf{F}\mathbf{F}^T\mathbf{A}\mathbf{F}\mathbf{z}\}-2E\{\mathbf{z}^T\mathbf{F}^T\mathbf{A}^T\mathbf{F}^T\mathbf{x}\}+E\{\|\mathbf{x}\|^2\}=\varepsilon\to\min. \tag{11}$$

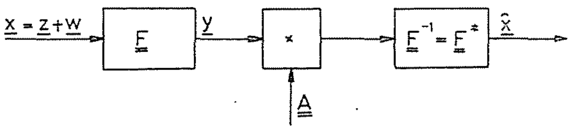

Fig. 1.

The condition of the minimum is that the derivative of (11) with respect to **A** equals zero:

$$E\{\mathbf{A}\mathbf{F}\mathbf{z}\mathbf{z}^T\mathbf{F}^T\}=E\{\mathbf{F}\mathbf{x}\mathbf{z}^T\mathbf{F}^T\}. \tag{12}$$

Here we utilized the identities of the differentiation according to a matrix and the exchangeability between the differentiation and the expected value formation. As $E\{\ldots\}$ needs only be formed for the variable values, we obtained for the optimal **A**:

$$\begin{aligned}\mathbf{A}_{opt}&=\mathbf{F}E\{\mathbf{x}\mathbf{z}^T\}\mathbf{F}^T[\mathbf{F}E\{\mathbf{z}\mathbf{z}^T\}\mathbf{F}^T]^{-1}=\\&=\mathbf{F}E\{\mathbf{x}\mathbf{z}^T\}[E\{\mathbf{z}\mathbf{z}^T\}]^{-1}\mathbf{F}^T\end{aligned} \tag{13}$$

Here the relationships (8) and the identity $(\mathbf{AB})^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1}$ were utilized. The relationship may also be written by the covariance matrices [1, 2].
By definition:

$$\begin{aligned}\mathbf{K}_x&=E\{(\mathbf{x}-E\{\mathbf{x}\})(\mathbf{x}-E\{\mathbf{x}\})^T\}=\\&=E\{\mathbf{x}\mathbf{x}^T\}-E\{\mathbf{x}\}E\{\mathbf{x}\}^T\end{aligned} \tag{14}$$

The quantities appearing in (13) are:

$$E\{\mathbf{x}\mathbf{z}^T\} = E\{\mathbf{x}(\mathbf{x}+\mathbf{w})^T\} = E\{\mathbf{x}\mathbf{x}^T\} + E\{\mathbf{x}\mathbf{w}^T\} \tag{15}$$

$$E\{\mathbf{z}\mathbf{z}^T\} = E\{\mathbf{x}\mathbf{x}^T\} + E\{\mathbf{w}\mathbf{w}^T\} + E\{\mathbf{x}\mathbf{w}^T\} + E\{\mathbf{w}\mathbf{x}^T\}. \tag{16}$$

Assuming

$$E\{\mathbf{x}\} = \frac{1}{M}\sum_{i=1}^{M} \mathbf{x}_i = 0 \tag{17}$$

and $\mathbf{x}$, $\mathbf{w}$ are uncorrelated

$$E\{\mathbf{w}\mathbf{x}^T\} = E\{\mathbf{x}\mathbf{w}^T\} = 0 \tag{18}$$

then

$$E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{K}_x + \mathbf{K}_w. \tag{19}$$

With the above taken into account, (13) may be written in the form:

$$\mathbf{A}_{opt} = \mathbf{F}\mathbf{A}_r\mathbf{F}^T; \qquad \mathbf{A}_r = \mathbf{K}_x(\mathbf{K}_x + \mathbf{K}_w)^{-1}. \tag{20}$$

ε_{min} is obtained by substituting $\mathbf{A}_{opt}$ into (9) and by performing the appointed operations [1]:

$$\varepsilon_{min} = tr(\mathbf{K}_x - \mathbf{K}_x(\mathbf{K}_x + \mathbf{K}_w)^{-1}\mathbf{K}_x). \tag{21}$$

In the transformed domain the covariance matrix may also be determined.
Since

$$\mathbf{y} = \mathbf{F}\mathbf{z} \tag{22}$$

$$\mathbf{K}_y = E\{\mathbf{y}\mathbf{y}^T\} - E\{\mathbf{y}\}E\{\mathbf{y}\}^T = \mathbf{F}E\{\mathbf{z}\mathbf{z}^T\}\mathbf{F}^T - \mathbf{F}E\{\mathbf{z}\}E\{\mathbf{z}\}^T\mathbf{F}^T$$

i.e.

$$\mathbf{K}_y = \tilde{\mathbf{K}}_z = \mathbf{F}\mathbf{K}_z\mathbf{F}^T. \tag{23}$$

Here $\tilde{\mathbf{K}}_z$ is the transformed covariance matrix.
As $\mathbf{K}$ is a real symmetrical matrix,

$$tr(\mathbf{K}) = tr(\mathbf{F}\mathbf{K}\mathbf{F}^T) = tr(\mathbf{K}). \tag{24}$$

By substituting (22) and (23) into (21):

$$\varepsilon_{min} = tr(\mathbf{K}_x - \mathbf{K}_x[\mathbf{K}_x + \mathbf{K}_w]^{-1}\mathbf{K}_x). \tag{25}$$

On the basis of (21), (24), (25) the value of ε_{min} is seen to be independent of the orthogonal transformation $\mathbf{F}$. Accordingly it is advised to select $\mathbf{F}$ in a way that the transformation requires little computational work. From (10) it follows to be useful to select $\mathbf{F}$ in a way, that $\mathbf{A}_{opt}$ contains as few non-zero elements as possible, besides the data of the given random characteristics.

Karhunen–Loeve transformation

First let us examine the transformation $\mathbf{F}$, which results in an optimal diagonal filter $\mathbf{A}_{od}$. Let us utilize the following theorem: If λ_i and t_i $(i=1, 2, \ldots, N)$ are the eigenvalues and the eigenvectors, respectively, of a real symmetrical matrix $\mathbf{Q}$, then transform

$$\mathbf{TQT}^{\mathrm{T}}=\mathbf{D}=\mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_{\mathrm{N}})$$

where (26)

$$\mathbf{T}=[t_1, t_2, \ldots, t_{\mathrm{N}}]$$

results in a diagonal matrix [4]. From (26) it is evident, that if $\mathbf{F}$ is chosen to be the eigenvector matrix of matrix $\mathbf{A}_r$, then

$$\mathrm{A}_{\mathrm{od}}=\mathbf{FA}_r\mathbf{F}^{\mathrm{T}} \tag{27}$$

is the desired optimum diagonal matrix. The above orthogonal transformation whose basis vectors are eigenvectors of the specified $\mathbf{A}_r$ covariance-matrix is called the Karhunen–Loeve Transformation (KLT).

Supoptimal diagonal filters

Filters, $\mathrm{A}_{\mathrm{opt}}=\mathbf{FA}_r\mathbf{F}^{\mathrm{T}}$ will be discussed which may be computed by the fast transformations. After transformation, the main diagonal elements of $\mathrm{A}_{\mathrm{opt}}$ will be kept:

$$\mathrm{A}_d=\mathrm{diag}(a_{11}, \ldots, a_{\mathrm{NN}}). \tag{28}$$

From (20), using (23)

$$\mathrm{A}[\tilde{\mathbf{K}}_x+\tilde{\mathbf{K}}_w]=\tilde{\mathbf{K}}_x \tag{29}$$

we get, keeping the diagonal elements of $\mathbf{A}$:

$$a_{ii}=\frac{\tilde{\mathbf{K}}_x(i, i)}{\tilde{\mathbf{K}}_x(i, i)+\tilde{\mathbf{K}}_w(i, i)} \qquad i=1, 2, \ldots, N \tag{30}$$

where $\tilde{\mathbf{K}}_x(i, i)=\mathbf{FK}_x\mathbf{F}^{\mathrm{T}}$ denotes the i-th diagonal element of $\tilde{\mathbf{K}}_x$.
Using (9), (21), (23) the mean square error is [1]:

$$\varepsilon_d=tr(\tilde{\mathbf{K}}_x)-tr(\mathbf{A}_d\tilde{\mathbf{K}}_x) \tag{31}$$

but as $\mathbf{A}_d$ is diagonal

$$\varepsilon_d = tr(\tilde{\mathbf{K}}_x) - \sum_{i=1}^{N} \frac{\tilde{\mathbf{K}}_x^2(i, i)}{\tilde{\mathbf{K}}_x(i, i) + \tilde{\mathbf{K}}_w(i, i)} \tag{32}$$

here we utilized (30). In the case of KLT:

$$\varepsilon_{\mathrm{KLT}} = \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \frac{\alpha_i^2}{\beta_i} \tag{33}$$

where α_i and β_i $(i=1, 2, \ldots, N)$ are the eigenvalues of $\mathbf{K}_x$ and $\mathbf{K}_x+\mathbf{K}_w$ respectively. As a conclusion: the ideal transformation method is KLT, which results in an optimal diagonal filter, and the error (33) equals the minimum value (25). The disadvantage of this method is that the eigenvectors of $\mathbf{A}_r$ must be determined, rather high in computation demand with increasing dimension, and the performance of the operation (10) requires a full matrix multiplication. In the case of suboptimal diagonal filters, if fast transformation methods are used, the computational work decreases considerably, and the error will be as described by (32). The transformations can be compared on the basis of (32). The transformation is good if the value of (32) is low.

Data Compression

The basic problem is the following: vector $\mathbf{x}$ of N elements is given. By $\mathbf{y}=\mathbf{F}\mathbf{x}$ transformation a new vector is obtained. Some of its elements are omitted before the inverse transformation, or replaced by selected elements. The objective is to select a subset of M components of $\mathbf{y}$, where M is substantially less than N. The remaining $N-M$ components can then be discarded without introducing an objectionable error, when the signal is reconstructed using the retained M components of $\mathbf{y}$. With (6a):

$$\mathbf{x} = \mathbf{F}^{\mathrm{T}}\mathbf{y} = \sum_{i=1}^{N} y_i \mathbf{f}_i \tag{34}$$

$$\mathbf{x}^{\mathrm{T}} = [x_1, \ldots, x_N] \tag{35}$$

$$\mathbf{y}^{\mathrm{T}} = [y_1, \ldots, y_N]. \tag{36}$$

We retain the elements $\{y_i,\ i=1, 2, \ldots, M\}$ $(M<N)$, and substitute the $N-M$ components of $\mathbf{y}$ by preselected constants b_i to obtain

$$\hat{\mathbf{x}}(M) = \sum_{i=1}^{M} y_i \mathbf{f}_i + \sum_{i=M+1}^{N} b_i \mathbf{f}_i \tag{37}$$

where $\hat{\mathbf{x}}(M)$ denotes the estimate of $\mathbf{x}$.

The error vector:

$$\varepsilon = \mathbf{x} - \hat{\mathbf{x}}(M) = \sum_{i=M+1}^{N} (y_i - b_i)\mathbf{f}_i. \tag{38}$$

Using (38) the mean square error is

$$\varepsilon(M) = E\{\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon}\} = E\left\{\sum_{i=M+1}^{N}\sum_{j=M+1}^{N}(y_i - b_i)(y_j - b_j)\mathbf{f}_i^{\mathrm{T}}\mathbf{f}_j\right\} \tag{39}$$

$\varepsilon(M)$ is seen to depend on the transformation $\mathbf{F}$ and on $\{b_i\}$. The process of choosing the optima b_i and $\mathbf{f}_i$ is carried out in two steps. The optimum b_i are obtained from (39) as follows:

$$b_i = E\{\mathbf{y}_i\} \qquad i = M+1, \ldots, N. \tag{40}$$

But as

$$y_i = \mathbf{f}_i^{\mathrm{T}}\mathbf{x} \tag{41}$$

(40) may be written as:

$$b_i = \mathbf{f}_i^{\mathrm{T}}E\{\mathbf{x}\} \qquad i = M+1, \ldots, N. \tag{42}$$

Substituting (42) into (39) and using (41):

$$\varepsilon(M) = \sum_{i=M+1}^{N} \mathbf{f}_i^{\mathrm{T}}E\{(\mathbf{x}-E\{\mathbf{x}\})(\mathbf{x}-E\{\mathbf{x}\})^{\mathrm{T}}\}\mathbf{f}_j = \sum_{i=M+1}^{N}\mathbf{f}_i^{\mathrm{T}}\mathbf{K}_x\mathbf{f}_i. \tag{43}$$

Relationship (14) was utilized here.

To obtain the optimum $\mathbf{f}_i$ not only $\varepsilon(M)$ has to be minimized with respect to $\mathbf{f}_i$, but also the constraint $\mathbf{f}_i^{\mathrm{T}}\mathbf{f}_i = 1$ has to be satisfied. This may be performed with the help of the Lagrange multiplicator, and the following final result is obtained:

$$\mathbf{K}_x\mathbf{f}_i = \gamma_i\mathbf{f}_i \quad i = M+1, \ldots, N \tag{44}$$

where γ_i denotes the Lagrange multiplicator. By definition (44) implies that $\mathbf{f}_i$ is the eigenvector of $\mathbf{K}_x$ and γ_i is the i-th corresponding eigenvalue. Here also the KLT transformation proves ideal. The error:

$$\varepsilon_{\min}(M) = \sum_{i=M+1}^{N} \gamma_i. \tag{45}$$

The matrix $\mathbf{K}_y = \mathbf{F}\mathbf{K}_x\mathbf{F}^{\mathrm{T}}$ will be diagonal if $\mathbf{F}$ is the matrix of KLT. From (45) it follows that, omitting a component y_k the mean square error increases by γ_k, the corresponding eigenvalue. Thus the set of y_i with the greatest M eigenvalues should be selected, and the remaining y_i discarded since they can be substituted by the constants b_i. From (42) it follows that $b_i = 0$ if $E\{\mathbf{x}\} = 0$. For all other transformations, however, $\mathbf{K}_y$ has non-zero off diagonal terms.

Thus a logical criterion for selecting transformed components is to retain the set of M components with the greatest variances—the remaining $N-M$ components can be discarded. This selection is called the variance criterion. As it has been mentioned in the case of the Wiener filtering, the efficiency of the various transformations may be determined by error examination, here by the variance criterion. The presence of few elements of relatively great amplitudes in the main diagonal of $\mathbf{K}_y$ is favourable.

For example, we have vectors of 128 elements. Storing these as words in a computer memory, every vector would occupy 128 places. With the help of this procedure, selecting the first 43 elements with the greatest variances, it is enough to store these, representing a 3-fold data compression without the change of the character of $\mathbf{x}$ at the resetting, i.e. "$\hat{\mathbf{x}}$ is very similar to $\mathbf{x}$". In pattern recognition problems, data compression corresponds to feature selection [1, 3].

Applications

An important field of application of the method may be in computer aided medical diagnostics, where it may help classification and storage of various physiological curves (e.g. electrocardiograms). Another possibility is pattern recognition [1]. It may be used also in processing seizmograms. It must be added, that the processing supposes availability of certain hardware, crucial for the utilization of the methods in practice.

Summary

The paper deals with the discrete orthogonal transformations, as a modern and useful tool of computer aided signal processing.

In signal processing, a frequent problem is to separate the signal from the noise (filtering). Another problem may be the storage of the signal in the computer memory. The size of the memory may be reduced, if only the most important characteristics of the signal are stored. These tasks may be performed effectively by the application of orthogonal transformation.

The purpose of this paper is a short review of this field and calling attention to the theme.

References

1. AHMED, N.–RAO, K. R.: Orthogonal Transforms for Digital Signal Processing. Springer-Verlag, Berlin, 1975.
2. PRATT, W. K.: Generalized Wiener Filtering Computation Techniques. IEEE Trans. on Comp. vol C21 No. 7. July 1972.
3. NILSSON, N. J.: Learning Machines, McGraw–M.U New York 1965.
4. KORN, A. G.–KORN, T. M.: Mathematical Handbook for Scientists and Engineers. Műszaki Könyvkiadó, Budapest 1975. (in Hung)
5. SZŐKEFALVY-NAGY, B.: Real functions and Series of Functions. Tankönyvkiadó, Budapest, 1965.
6. Proc. 1972. Symp. Applications of Walsh Functions. Naval Research Laboratory, Washington.

László Kónya, H—1521 Budapest