Application of biorthogonal filter functions to pattern recognition and feature extraction

Eugene A. Margerum

MARCH 1980

U.S. ARMY CORPS OF ENGINEERS
ENGINEER TOPOGRAPHIC LABORATORIES
FORT BELVOIR, VIRGINIA 22060
APPLICATION OF BIORTHOGONAL FILTER FUNCTIONS TO PATTERN RECOGNITION AND FEATURE EXTRACTION

Dr. Eugene A. Margerum

U. S. Army Engineer
Topographic Laboratories
Fort Belvoir, Virginia 22060

March 1980

Approved for Public Release; Distribution Unlimited

A mathematical method is developed for generating a set of filter functions from a given set of signatures. The filter function set of functions is bi-orthogonal to the set of signature functions; therefore, any one filter function gives a perfect response to one signature, and a response to all other given signatures is completely suppressed. The method can be used to decompose superpositions of signatures as well as for improving separation of measured parameters for pattern recognition. It can also be used to suppress interferences from the background when it is included in the given set of signatures.
20. Continued

A method of adding new filter functions to an existing set without complete recomputation (adaptive learning) is discussed.
PREFACE

The work reported on was done under DA Project 4A161101A91D, Task 01, Work Unit 1891D010064, "Feature Extraction from Signal Structure Utilizing Biorthogonal Filter Functions."

The work was performed during 1978, Dr. F. Rohde was the Team Leader, Center for Theoretical and Applied Physical Sciences; and Mr. M. Crowell, Jr., was the Director, Research Institute.

COL Philip R. Hoge, CE, LTC William T. Stockhausen, CE, and COL Daniel L. Lycan, CE were Commanders and Directors and Mr. Robert P. Macchia was Technical director of the Engineer Topographic Laboratories during the study and report preparation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREFACE</td>
<td>1</td>
</tr>
<tr>
<td>ILLUSTRATIONS</td>
<td>2</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>3</td>
</tr>
<tr>
<td>A QUALITATIVE DESCRIPTION OF THE METHOD</td>
<td>4</td>
</tr>
<tr>
<td>MATHEMATICAL DERIVATION</td>
<td>7</td>
</tr>
<tr>
<td>FILTERING OF SIGNATURES</td>
<td>10</td>
</tr>
<tr>
<td>APPROXIMATION IN THE MEAN</td>
<td>12</td>
</tr>
<tr>
<td>ADAPTIVE LEARNING</td>
<td>15</td>
</tr>
<tr>
<td>GENERAL REMARKS</td>
<td>20</td>
</tr>
<tr>
<td>CONCLUSION</td>
<td>20</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>21</td>
</tr>
</tbody>
</table>

## ILLUSTRATIONS

<table>
<thead>
<tr>
<th>NUMBER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Characteristic Patterns for Three Objects</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Linear Superposition of the Reference Patterns</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Construction of Biorthogonal Filter Functions</td>
<td>6</td>
</tr>
</tbody>
</table>
APPLICATION OF BIORTHONGONAL FILTER FUNCTIONS TO PATTERN RECOGNITION AND FEATURE EXTRACTION

INTRODUCTION

It is typical of many problems of pattern recognition that the number of possible environmental states to be considered may be \( \textit{infinite} \), but that the number of solutions or classifications that can be assigned remains \( \textit{finite} \). A set of measurements is performed on an environmental system, and by using the measurements, a \( \textit{single} \) category is assigned to this pattern of measurements. Thus, the system is recognized as belonging to only one class.

In cases where terrain is examined, the environmental system must sometimes be viewed as a collection of objects belonging to several different classes. Again, in cases where remotely sensed data is used, the pattern of measurements obtained can consist of a \( \textit{superposition} \) of patterns from objects belonging to different classes. In these two cases, one must determine the relative proportion of the different components. In other cases, the object to be identified, may be viewed against any one of a number of different backgrounds that produce strong interferences in measurements made on the system and it becomes desirable to separate or filter the object parameters from the background.

In the first two cases, methods more general than the normally used simple pattern recognition scheme are required. In the third, at least some additional processing is indicated. It will be shown that mathematical expansion in biorthogonal sets of functions can be used to advantage in solving these problems.

The methods to be developed are general and the approach is phenomenological since they can be applied irrespective of the type of measurements performed on the system. For example, the spatial frequency spectrum may be used, or some functions that depend on it; or the electromagnetic spectrum may be used, or some integrated responses for wide spectral bands. As an alternative, other types of measurements or combinations of mixed types of measurements may be used.

In most pattern classification schemes, the “classing” of an object depends on its correlation with patterns of an idealized reference pattern set or alphabet. With the method to be introduced here, a biorthogonal filter function is constructed corresponding to each member of the reference set. The filter function has the (biorthogonal) property
that it has zero correlation with all members of the reference set except the one with which it corresponds. This property has the effect of suppressing cross correlation with unwanted patterns for standard pattern recognition problems as well as for the ones to be emphasized here. The method is also adaptive in the sense that when a new reference pattern is added to the alphabet, the new filter function can be constructed and the old ones revised to yield a new enlarged set of filter functions without starting the computation from the beginning.

A QUALITATIVE DESCRIPTION OF THE METHOD

A set of three functions or reference patterns that may characterize three different objects is shown in figure 1. A superposition of these same functions is given schematically in figure 2. The problem is to determine which components are present in the superposition and in what amounts. In a real case, the number of reference functions might be expected to be much greater than three, and the superposition may contain a small number of them, e.g. two or three. Some of the reference patterns could consist of typical background noise patterns that need to be suppressed in the process of system identification.

The method for constructing the filter functions is demonstrated schematically by means of a different example shown in figure 3. Three reference patterns are shown in the upper part of the diagram as A, B, and C. Their corresponding biorthogonal filter functions are given below them by A', B', and C', respectively. They have the property that if any unprimed function is multiplied by a primed function and integrated, the result will be zero except for the case where the filter function corresponds with the reference pattern. It is significant that each filter function can be represented as a linear combination of the given reference functions as shown by the equations at the bottom of figure 3.
Figure 1. Characteristic Patterns for Three Objects

Figure 2. Linear Superposition of the Reference Patterns
Figure 3. Construction of Biorthogonal Filter Functions

\[ A' = 3A - 2B + C \quad B' = 2A + 2B - C \quad C' = A' - B + C \]
MATHEMATICAL DERIVATION

The set reference patterns is assumed to be given by a set of \( N \) linearly independent functions \( f_j(x) \), where \( j = 1, 2, \ldots, N \). It is possible to construct a second set of functions \( \phi_k(x) \), where \( k = 1, 2, \ldots, N \), such that they have the property

\[
\int_a^b f_j(x) \phi_k(x) \, dx = \delta_{jk}
\]

(1)

where \( \delta_{jk} \) is the Kronecker delta. Sets of functions \( \phi_k(x) \) obeying equation (1) are said to be biorthogonal (normalization will be understood as well) to the first set \( f_j(x) \) over the interval \([a, b]\).

For a finite number of reference functions, it will usually be possible to construct many different sets of biorthogonal filter functions. However, if the form of the filter functions is limited to linear superpositions of the reference functions

\[
\phi_k(x) = \sum_{j=1}^{N} \alpha_{jk} f_j(x)
\]

(2)

The symmetric matrix \( \beta \) is defined in terms of its components by

\[
\beta_{lm} = \int_a^b f_l(x) f_m(x) \, dx
\]

(3)

and it will possess an inverse if the functions \( f_j(x) \) form a linearly independent set and the determinant of \( \beta \) is consequently non-zero.

If equation 2 is multiplied by \( f_m(x) \) and integrated from \( a \) to \( b \) and if equations 1 and 3 are used, then

\[
\int_a^b f_m(x) \phi_j(x) \, dx = \sum_{l=1}^{N} \alpha_{jl} \int_a^b f_m(x) f_l(x) \, dx
\]

(4)
\[
\delta_{jm} = \sum_{l=1}^{N} \alpha_{jl} \beta_{lm} \quad (5)
\]

It is found that the matrix \( \alpha \) is the inverse of \( \beta \). The computation of the components of \( \alpha \) is all that is required to find the biorthogonal filter functions of the required form as expressed in equation 2.

The fact that a reciprocal relationship exists between the reference functions can be readily demonstrated by forming the symmetrical autocorrelation coefficient matrix for the functions \( \phi_j(x) \),

\[
\int_{a}^{b} \phi_j(x) \phi_k(x) \, dx = \sum_{l=1}^{N} \alpha_{jl} \int_{a}^{b} f_l(x) \phi_k(x) \, dx = \sum_{l=1}^{N} \alpha_{jl} \delta_{lk} \quad (6)
\]

\[
= \sum_{l=1}^{N} \alpha_{jl} \delta_{lk} = \alpha_{jk} \quad (7)
\]

where use has been made of equations 2 and 1, and by forming the sum

\[
\sum_{k=1}^{N} \beta_{jk} \phi_k(x) = \sum_{k=1}^{N} \sum_{l=1}^{N} \beta_{jk} \alpha_{kl} f_l(x) = \sum_{l=1}^{N} \delta_{jl} f_l(x) = f_j(x) \quad (9)
\]
where equations 2 and 5 have been used. To emphasize their character, the reciprocal relationships of equations 2 and 11 are given in more explicit terms below:

\[
\phi_k(x) = \sum_{j=1}^{N} f_j(x) \int_{a}^{b} \phi_k(\xi) \phi_j(\xi) \, d\xi
\]

\[
f_k(x) = \sum_{j=1}^{N} \phi_j(x) \int_{a}^{b} f_k(\xi) f_j(\xi) \, d\xi
\]

Since every member of either set of functions is expressible as a linear combination of the members of the other set, either set may be used as a basis to span the same function space.
FILTERING OF SIGNATURES

Suppose the function \( \mu(x) \) consists of a linear superposition of patterns \( f_j(x) \) for a known set of features, then

\[
\mu(x) = \sum_{j=1}^{N} a_j f_j(x) \quad (14)
\]

If the biorthogonal set of functions \( \phi_j(x) \) is known, the presence and relative amounts of each component pattern is readily found.

\[
\int_{a}^{b} \mu(x) \phi_j(x) \, dx = \sum_{j=1}^{N} a_j \int_{a}^{b} f_j(x) \phi_k(x) \, dx \quad (15)
\]

\[
= \sum_{j=1}^{N} a_j \delta_{jk} \quad (16)
\]

\[
= a_k \quad (17)
\]

It should be emphasized that once the filter function \( \phi_k(x) \) is known, the presence or absence and the amount of the k-th component can be found without determining the presence or amounts of the other possible components. This is of value when searching for a particular feature in a background consisting of other reference patterns where it is not desirable to perform computations for all of the components present.

The equations to be presented next are relevant if the correlation coefficients of the superposition with the set of reference patterns is known. Their derivation is straightforward and follows from the reciprocal properties of biorthogonal sets of functions.

\[
\mu(x) = \sum_{k=1}^{N} b_k \phi_k(x) \quad (18)
\]
The relationship between the \( a \)'s and \( b \)'s is determined by

\[
\int_a^b \mu(x) f_m(x) \, dx = \sum_{k=1}^N b_k \int_a^b \phi_k(x) f_m(x) \, dx
\]  
(19)

\[
= \sum_{k=1}^N b_k \delta_{km}
\]  
(20)

\[
= b_m
\]  
(21)

\[
\int_a^b \phi_l(x) \mu(x) \, dx = \sum_{k=1}^N b_k \int_a^b \phi_l(x) \phi_k(x) \, dx
\]  
(22)

\[
a_l = \sum_{k=1}^N b_k \alpha_{kl}
\]  
(23)

or

\[
b_m = \sum_{l=1}^N a_l \beta_{lm}
\]  
(24)

since the matrix \( \beta \) is the inverse of \( \alpha \).
The best fit of the function $\mu(x)$ in a least-square sense results in the same coefficients as those given by equation 21, and the proof of this follows in an analogous fashion to that used for orthogonal functions. The function

$$1 = \int_a^b \left[ \mu(x) - \sum_{l=1}^N \gamma_l f_l(x) \right]^2 \, dx$$

is to be minimized by adjusting the coefficients $\gamma_l$.

$$\frac{\partial I}{\partial \gamma_m} = 0 = 2 \int_a^b \left[ \mu(x) - \sum_{l=1}^N \gamma_l f_l(x) \right] f_m(x) \, dx$$

$$\int_a^b \mu(x) f_m(x) \, dx = \sum_{l=1}^N \gamma_l \int_a^b f_l(x) f_m(x) \, dx$$

$$= \sum_{l=1}^N \gamma_l \beta_{lm}$$

By multiplying both sides by $\alpha_{mn}$ and summing over $m$,

$$\int_a^b \mu(x) \sum_{m=1}^N \alpha_{mn} f_m(x) \, dx = \sum_{l=1}^N \sum_{m=1}^N \gamma_l \beta_{lm} \alpha_{mn}$$

$$\int_a^b \mu(x) \phi_n(x) \, dx = \sum_{l=1}^N \gamma_l \delta_{ln}$$

$$a_n = \gamma_n$$
it is found that the coefficients $\gamma$ and $\alpha$ are the same.

The quantity

$$S = \int_a^b [\nu(x) - \sum_{n=1}^N a_n f_n(x)]^2 \, dx$$

may be used as a measure of the fit of $\nu(x)$ by the set of functions $f_n(x)$ in cases where $\nu(x)$ is a pattern from the real world environment. A representation in terms of reference functions is not usually exact, and it is important to know the quality of the fit and the improvement accomplished by changing or introducing additional reference (basis) functions.

$$S = \int_a^b [\nu(x)]^2 \, dx - 2 \sum_{n=1}^N a_n \int_a^b \nu(x) f_n(x) \, dx$$

$$+ \sum_{n=1}^N \sum_{m=1}^N a_n a_m \int_a^b f_n(x) f_m(x) \, dx$$

$$= \int_a^b [\nu(x)]^2 \, dx - 2 \sum_{n=1}^N a_n b_n + \sum_{n=1}^N \sum_{m=1}^N a_n a_m \beta_{mn}$$

$$= \int_a^b [\nu(x)]^2 \, dx - 2 \sum_{n=1}^N a_n b_n + \sum_{n=1}^N a_n b_n$$

$$= \int_a^b [\nu(x)]^2 \, dx - \sigma$$

$$\sigma = \sum_{n=1}^N a_n b_n$$
or

\[ \sigma = \sum_{n=1}^{N} \sum_{m=1}^{N} a_n a_m \beta_{nm} \]  \hspace{1cm} (38)

or

\[ \sigma = \sum_{n=1}^{N} \sum_{m=1}^{N} b_n b_m \alpha_{nm} \]  \hspace{1cm} (39)

may be of use depending on the computational context.
ADAPTIVE LEARNING

The ability of the system to recognize patterns may be increased by the addition of new reference patterns. This requires the computation of a new set of filter functions. In this section it will be shown that this may be accomplished by using the old set without recomputing and starting from the beginning.

Suppose the biorthonormal set \[ \{ \phi_k(x): k = 1, N - 1 \} \] has been found corresponding to the set \[ \{ f_k(x): k = 1, N - 1 \} \] and then the latter set is enlarged to \[ \{ f_k(x): k = 1, N \} \] by the addition of \( f_N(x) \). A new set of biorthogonal filter functions \[ \{ \tilde{\phi}_k(x): k = 1, N \} \] will result. The old coefficients \( \beta_{lk} \) where \( l, k \leq N - 1 \) will remain unchanged, but the new coefficients \( \beta_{Nk}, \beta_{IN}, \) and \( \beta_{NN} \) must be computed and appended. Also, a new set of coefficients \( \tilde{\alpha}_{kl} \); \( k, l \leq N \) must be found to replace the old ones \( \alpha_{kl} \); \( k, l \leq N - 1 \). However, use can be made of the old coefficients \( \alpha_{kl} \) to reduce the labor involved in finding \( \alpha_{kl} \).

The new and old filter functions are given by

\[
\tilde{\phi}_k(x) = \sum_{l=1}^{N} \tilde{\alpha}_{kl} f_l(x) \tag{40}
\]

and

\[
\phi_m(x) = \sum_{n=1}^{N-1} \alpha_{mn} f_n(x) \tag{41}
\]

respectively.

\[
\int_a^b \tilde{\phi}_k(x) \phi_m(x) \, dx = \sum_{l=1}^{N} \tilde{\alpha}_{kl} \int_a^b f_l(x) \phi_m(x) \, dx \tag{42}
\]
Combining equations 44 and 47 yields

\[ \tilde{\alpha}_{k m} = \alpha_{m k} - \tilde{\alpha}_{k N} \int_a^b f_N(x) \phi_m(x) \, dx \]  

Next, formulae for the coefficients of the type \( \tilde{\alpha}_{k N} \) and \( \tilde{\alpha}_{N m} \) will be found.

\[ \int_a^b \phi_N(x) \phi_m(x) \, dx = \sum_{n=1}^{N-1} \alpha_{m n} \int_a^b \phi_N(x) f_n(x) \, dx \]

\[ = \sum_{n=1}^{N-1} \alpha_{m n} \delta_{N n} \]

\[ = 0 \]
Also

\[\int_{a}^{b} \tilde{\phi}_{N}(x) \phi_{m}(x) \, dx = \sum_{l=1}^{N} \tilde{\alpha}_{Nl} \int_{a}^{b} f_{l}(x) \phi_{m}(x) \, dx \]

(52)

\[= \sum_{l=1}^{N-1} \tilde{\alpha}_{Nl} \delta_{lm} + \tilde{\alpha}_{NN} \int_{a}^{b} f_{N}(x) \phi_{m}(x) \, dx \]

(53)

By combining equations 51 and 53

\[0 = \sum_{l=1}^{N-1} \tilde{\alpha}_{Nl} \delta_{lm} + \tilde{\alpha}_{NN} \int_{a}^{b} f_{N}(x) \phi_{m}(x) \, dx \]

(54)

\[\tilde{\alpha}_{Nm} = - \tilde{\alpha}_{NN} \int_{a}^{b} f_{N}(x) \phi_{m}(x) \, dx \]

(55)

and by inserting equation 55 into 48, the coefficients of type \(\tilde{\alpha}_{kN}\) and \(\tilde{\alpha}_{Nm}\) are eliminated.

\[\tilde{\alpha}_{kM} = \alpha_{mk} + \tilde{\alpha}_{NN} \left( \int_{a}^{b} f_{N}(x) \phi_{m}(x) \, dx \right) \left( \int_{a}^{b} f_{N}(x) \phi_{k}(x) \, dx \right) \]

(56)

For convenience of notation, let

\[\lambda_{m} = \int_{a}^{b} f_{N}(x) \phi_{m}(x) \, dx = \sum_{n=1}^{N-1} \alpha_{mn} \int_{a}^{b} f_{N}(x) f_{n}(x) \, dx = \sum_{n=1}^{N-1} \alpha_{mn} \beta_{Nn} \]

(57)

Then, equation 56 becomes

\[\tilde{\alpha}_{km} = \alpha_{mk} + \tilde{\alpha}_{NN} \lambda_{m} \lambda_{k} \]

(58)
For the new set of filter functions, the normalization condition for \( \phi_N \) associated with the biorthogonality is

\[
1 = \int_a^b \phi_N(x) f_N(x) \, dx
\]

\[
= \sum_{l=1}^N \beta_{NI} \int_a^b \tilde{\phi}_N(x) \tilde{\phi}_l(x) \, dx
\]

\[
= \sum_{l=1}^{N-1} \beta_{NL} \tilde{\alpha}_{NI} + \beta_{NN} \tilde{\alpha}_{NN}
\]

\[
= -\sum_{l=1}^{N-1} \beta_{NI} \tilde{\alpha}_{NN} \lambda_l + \beta_{NN} \tilde{\alpha}_{NN}
\]

where the substitution

\[
\tilde{\alpha}_{NI} = -\tilde{\alpha}_{NN} \lambda_l
\]

has been taken from equations 55 and 57 above.

Equation 62 can now be solved for \( \tilde{\alpha}_{NN} \).

\[
\tilde{\alpha}_{NN} = \frac{1}{\beta_{NN} - \sum_{l=1}^{N-1} \beta_{NI} \lambda_l}
\]

\[
= \frac{1}{\beta_{NN} - \sum_{l=1}^{N-1} \sum_{m=1}^{N-1} \beta_{NI} \alpha_{lm} \beta_{Nm}}
\]
The formulae needed for the revision of the filter functions are all derived. The steps in the computational procedure are given below:

1. \( \beta_{jk} \) remain unchanged for \( j, k = 1, N - 1 \)

2. Compute \( \beta_{Nj} = \beta_{jN} \) for \( j = 1, N \) from
   \[
   \beta_{Nj} = \int_a^b f_N(x) f_j(x) \, dx
   \]

3. Compute \( \lambda_l = \sum_{m=1}^{N-1} \alpha_{lm} \beta_{Nm} \) for \( l = 1, N - 1 \)

4. Calculate \( \tilde{\alpha}_{NN} = \frac{1}{\beta_{NN} - \sum_{l=1}^{N-1} \beta_{Nl} \lambda_l} \)

5. Compute \( \tilde{\alpha}_{Nl} = -\tilde{\alpha}_{NN} \lambda_l \) for \( l = 1, N - 1 \)

6. Compute \( \tilde{\alpha}_{km} = \alpha_{km} + \tilde{\alpha}_{NN} \lambda_m \lambda_k \)

or

\[
\tilde{\alpha}_{km} = \alpha_{km} - \tilde{\alpha}_{Nk} \lambda_m \text{ for } k, m = 1, N - 1.
\]

Where possible, use the symmetry of \( \tilde{\alpha}_{km} \) to fill out its values.

The procedure given here amounts to a learning process where the old filter functions are modified (retrained) to ignore the new pattern being added to the set of reference functions.
GENERAL REMARKS

In practical implementation, the integrals will often be replaced by summation over some finite number of points. The function space (with an infinite number of components) is thereby reduced to a finite dimensional vector space. To retain linear independence of the reference patterns, the number of components must be at least as great as the number of reference patterns. By way of contrast, no such condition is necessary (although it may hold accidentally) in most normal pattern recognition procedures. The gain for fulfilling this condition is the ability to decompose superpositions of patterns into their component parts.

Even if the filter functions were used with the usual pattern recognition procedures, the separability of many patterns should be increased. In addition, the weight of any single component can be obtained by itself by using its appropriate filter function. It should also be mentioned that typical backgrounds and interferences can be included in the reference pattern set, and their usual tendency to interfere can be somewhat suppressed in this manner.

CONCLUSION

The concept of biorthogonality has been developed into a method of pattern recognition that can be applied to a wide variety of data types. This treatment leads to mathematical methods for constructing filter functions that give perfect response to one member of the set of given functions while giving zero response to all of the others. Furthermore, this response can be found by probing the environmental state with only the one filter function.

The mathematical system is capable of learning new patterns and revising the old filter functions so that they have no response to the new patterns. In contrast to standard pattern recognition methods where one classification is made, this method permits the decomposition of a superposition of many patterns into its component parts.
BIBLIOGRAPHY


