BEN-TCHIKOU, Abdel-Hamid, 1936-
ON THE KARHUNEN-LOEVE EXPANSION.

The Ohio State University, Ph.D., 1969
Engineering, electrical

University Microfilms, Inc., Ann Arbor, Michigan
ON THE KARHUNEN-LOEVE EXPANSION

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Electrical Engineering in the
Graduate School of The Ohio State University

By

Abdel Hamid Ben-Tchikou, B.S.E.E., M.Sc.

The Ohio State University

1969

Approved by

[Signature]

Adviser
Department of Electrical Engineering
PLEASE NOTE:

Not original copy. Blurred and faint type on several pages. Filmed as received.

UNIVERSITY MICROFILMS.
ACKNOWLEDGEMENT

The author expresses his deep gratitude to his adviser, Professor C. E. Warren for introducing him to the subject of statistical communication theory for suggesting the topic of their research, and for his patience, encouragements and constant guidance throughout it.

He also would like to express his thanks to Professors H. Hemami and H. Ko for their criticisms and suggestions.

He thanks U.N.E.S.C.O. and the Polytechnic Institute of the University of Algiers for their support.

Finally he thanks Miss Janet Patterson and Mr. Vance Bown for typing this dissertation in critical circumstances.
January 29, 1936. . . . Born, Medea, Algeria


1963-1964 . . . . . . . Studies at the University of Southern California (U.S.C.); B.S.E.E.


1966 . . . . . . . . . . . Teaching Associate of Math at U.S.C.

1966-1969 . . . . . . . Studies at the Ohio State University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Electrical Engineering

Studies in Communication theory: Professor C. E. Warren

Studies in Information Science and coding: Professor J. Tou and Professor R. Lackey

Studies in Electromagnetic theory: Professor W. Rusch
TABLE OF CONTENTS

ACKNOWLEDGMENTS .............................................. ii
VITA ........................................................... iii
LIST OF TABLES .................................................... v
LIST OF FIGURES .................................................. vi
INTRODUCTION ..................................................... 1

Chapter

I. TOWARDS THE UNIFICATION OF THE THEORY OF THE KARHUNEN-LOEVE EXPANSION ........................................... 5

The usual K.L. expansion
Generalized forms of the K.L. expansion
General properties of the K.L.
The Gaussian case
The K.L. and canonical representations
The K.L. and Reproducing Kernel Hilbert spaces (R.K.H.S.)

II. COVARIANCE KERNELS AND INTEGRAL EQUATIONS ............. 73

Properties of covariance kernels
Solution techniques for integral equations
associated with the K.L.

III. SOME APPLICATIONS OF THE K.L. ............................. 92

Suggested applications
Classical applications

IV. SUMMARY AND AREAS FOR FURTHER RESEARCH .................. 117

APPENDIX

A-1 ............................................................. 119
A-2 ............................................................. 122
A-3 ............................................................. 125

iv
<table>
<thead>
<tr>
<th>APPENDIX</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A-4</td>
<td>127</td>
</tr>
<tr>
<td>A-5</td>
<td>130</td>
</tr>
<tr>
<td>A-6</td>
<td>133</td>
</tr>
<tr>
<td>A-7</td>
<td>142</td>
</tr>
<tr>
<td>A-8</td>
<td>145</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>149</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>30 - 32</td>
</tr>
<tr>
<td>Table 2</td>
<td>100</td>
</tr>
<tr>
<td>Table 3</td>
<td>109</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 5-1</td>
<td>62</td>
</tr>
<tr>
<td>Figure 9-1</td>
<td>92</td>
</tr>
<tr>
<td>Figure 9-2</td>
<td>97</td>
</tr>
<tr>
<td>Figure 10-1</td>
<td>112</td>
</tr>
</tbody>
</table>
INTRODUCTION

Over two decades ago, the Karhunen-Loeve Expansion (for abbreviation the K.L.) was introduced by Kac and Siegert in the engineering literature [1]. It was later on studied and used more systematically by Grenander [2] as a basic tool for statistical inference. Today its use is extensive in communication and information problems, especially in detection, estimation, filtering and recognition problems.

It is felt that many extensions of the forms of the K.L. and its properties have been made ad hoc. A major concern of the present research is to meet the need for the unification of the K.L. expansion theory. There are many reasons to account for the attractiveness of the K.L. as a basic tool in engineering today. One of them is that, in an age of digitalization and computers, the K.L. brings out a similarity between analog and digital processes. It therefore allows a continuity of reasoning and often a rigorous justification for passages to the limit.

Moreover, since more and more precision is sought for in modern systems, and since this is becoming technically more possible and economically more feasible, the optimality properties of the K.L. (which we discuss in sections III and IV) underline its value. Then with the widening introduction of generalized frequency concepts into the practical area (e.g. the Walsh functions), since new devices and filters are produced, we may hope that the generalized frequency which the K.L. introduces will be of use in the near future.
There has been existing a scholarly controversy about the usefulness of the K.L. expansion, or more precisely, about canonical representations (for their definition, cf. section V) among which the K.L. is the most prominent for its optimal properties. On the one side are those who, with V. S. Pugachev believe that "the method of canonical representations of random functions is the foundation of modern statistical theory of optimum systems"; and on the other, are those who, with E. Parzen, think that "reproducing kernel Hilbert spaces provide a more powerful and more elegant means of achieving in a unified manner the results which Pugachev has sought to unify by the method of canonical expansions"[3]. With respect to this problem, we will try to clarify, in the engineering language, the fact that the K.L. is but one concretization of the Hilbert space induced by a second-order random process.

Finally, it is felt that the K.L. may help in new areas, in particular in linear random channel modeling, where it would bring not only more efficiency of representation, but also more rigor in some assumptions; and that its generalized forms are of help in extending its application to classical areas, like detection, estimation and pattern recognition.

The present thesis is divided into three chapters.

a) Chapter I concerns itself with the unification of the K.L. theory, and deals mostly with the properties that depend on the random space $\Omega$. The unity is found through the use of general finite measure spaces and Schmidt's expansion and
approximation theorems. This allows not only to make all possible extensions of the K.L., but also to generalize the properties of the K.L. to those extensions in quite a natural and straightforward manner. The underlying structural unity of the different forms of the K.L. not only clarifies the basic concepts, but may show the way to new properties. In particular, it is seen that the limitation of the K.L. to the finite interval is due only to the fact that we need a finite measure space so as to be able to apply the Cauchy-Schwartz inequality; and that therefore there is a natural extension of the K.L. for the infinite interval as well. The relationship of the K.L. to the R.K.H.S. (Reproducing kernel Hilbert space) is reviewed; and some detailed study of the discrete case of the K.L. is made because of its potential application on the computer. Also, an extension of some recent results about the relationship between canonical representations and integral canonical ones is made to non-stationary processes.

We have tried briefly to bring together the properties of the K.L. that depend on the Gaussian character of the process. In particular, the relationship of the independence of the coefficients of the K.L. to Gaussian distribution is studied in some detail in view of the results of P. Pierre [4].

Last, possible invariance properties of the K.L. are investigated. It is clearly found that the K.L. is not very "generous" from this point. More generally, the effect of transformations on the K.L. is investigated; and a result of T. Kadota [5] about
differentiation is slightly improved.

b) Chapter II deals mostly with the difficult problems relating to the solution of Fredholm integral equations with covariance kernels. The main results about their eigenvalues and eigenfunctions are reviewed. The main concern here was to bring together solution techniques, analytical and numerical; and in particular, among the latter ones, a state-variable approach due to Baggeroer and a very unsophisticated yet very useful matrix approach based on the discrete K.L. Among analytical techniques, attention is drawn to the potential power of an old technique due to Fredholm and applicable to vector integral equations.

c) In Chapter III, areas of application of the K.L. are examined. In particular a model for the random time-varying linear channel is presented, based on the K.L. technique; and compared with the classical tapped delay line model. This is a very active area where practical, efficient models, are neither numerous nor entirely satisfactory.

An extension is made in the application of the K.L. to the problem of detection, in the vector as well as the discrete case; another in the application of the K.L. to the estimation problem.
CHAPTER I

TOWARDS THE UNIFICATION OF THE THEORY OF THE KARHUNEN-LOEVE EXPANSION

SECTION I

THE USUAL K. L. EXPANSION

In what follows we will denote by \( X(\omega, t) \) or \( X(t) \) a stochastic process, with

\[ \omega \in \Omega, \quad t \in T \]

where \( \Omega \) is the probability space and where \( T \) is a closed bounded interval. We will denote by \( \overline{X}(\omega, t) \) or \( \overline{X}(t) \) the complex conjugate of \( X(\omega, t) \); by \( \overrightarrow{X}(t) \) a vector random process, i.e. a sequence of random processes \( \{ X^1(t), X^2(t), \ldots, X^n(t) \} \)

By \( E\{ X(t) \} \) we denote the average of the random variable \( X(t) \), i.e.

\[ E\{ X(t) \} = \int_{\Omega} X(\omega, t) \, d\mu(\omega) \]

where \( \mu(\omega) \) is the probability measure over \( \Omega \).

We will consider mainly second-order, zero-mean random processes, i.e. processes such that

\[
E\{ X(t) \} = 0, \quad E\{ |X(t)|^2 \} = \int_{\Omega} |X(\omega, t)|^2 \, d\mu(\omega) < \infty
\]

(1-1)
We define as usual a random process to be \textit{continuous in the quadratic mean (q.m.)} as a process for which

\[ \mathcal{L} \text{. i. m. } \quad E \left\{ \left| X(s) - X(t) \right|^2 \right\} = 0 \]  

It is known that \( X(t) \) is continuous in q.m. if and only if

\[ R(t, s) = E \left\{ X(t) X(s) \right\} \] \( \text{is continuous in } t \text{ and } s, \) or equivalently, if it is continuous at \( t = s. \)

\section*{A. Existence theorem.}

The K.L. expansion is named after the two mathematicians, K. Karhunen [6] and M. Loeve [7], who discovered it separately around the year 1946 while they were investigating the properties of the Hilbert space induced by second-order random functions. It is defined by the following existence theorem.

\textbf{Theorem 1.1.} A zero-average random function \( X(t) \) continuous in quadratic mean on a closed interval \( T \) has on \( T \) an orthogonal decomposition

\[ X(t) = \sum_i \lambda_i \phi_i(t) \quad \phi_i(t) \]  

with

\[ E \left\{ \phi_i(t) \phi_j(t) \right\} = \delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \]  

\[ \int_T \phi_i(t) \phi_j(t) \, dt = \delta_{i,j} \]
if, and only if, the $\lambda_i$ are the eigenvalues, and the $\phi_i(t)$ are the orthonormalized eigenfunctions of its covariance function $R(t,s)$. Then the series in (1-3) converge in q.m. uniformly on $T$; and

$$\sum_{i=1}^{\infty} \lambda_i \phi_i(t) \phi_i(s)$$

(1-7)

where $T$ is a closed and bounded interval.

The proof of Mercer's theorem, in a more general context, is given in appendix A-2.

B. Remarks and examples. Let us note that:

1. The K.L. allows the expansion of a random function into a
denumerable sum of elementary uncorrelated functions $\lambda_i \phi_i(t)$,
Moreover, it separates the time dependence from the dependence on the random space, since the \( \phi_{i}^{*}(t) \) are completely deterministic, and the \( \mathcal{X}_{i} \) do not depend on \( t \). This makes the K.L. a very convenient tool when operating on a random function: the execution of different linear operations is reduced to operations on ordinary non-random functions.

2. Given the autocorrelation function of the zero-mean process \( X(t) \), the K.L. says that there is one and only one way to expand it in such a fashion that there is simultaneous orthogonality on the two spaces, \( \Omega \) and \( T \). Recalling the general importance of orthogonality for the ease in carrying out calculations, and especially for random variables where it is the necessary requirement for independence, we can already expect an unusual role for the K.L. In a sense, the K.L. for a second order process is like the principal axis reduction for a second order surface, or the diagonalization of a square matrix.

3. We note however, that the coordinate functions \( \phi_{i}^{*}(t) \) depend on \( T \), and are not easy to calculate in general. The K.L. is therefore a very specific though a most natural representation of a random process. From this fact there arise some of its limitations. The K.L. is unlike the "jack-of-all-trades", the familiar representation.

4. We note that

\[
\mathbb{E} \left\{ |X(t)|^2 \right\} = \sum_{i=1}^{\infty} \lambda_i^2 \phi_{i}^{*2}(t)
\]

and

\[
\int \mathbb{E} \left\{ |X(t)|^2 \right\} dt = \sum_{i=1}^{\infty} \lambda_i^2
\]
This shows the K.L. to act as a filter which unveils the "natural modes" of the random process, i.e. the $\phi_i(t)$ are "generalized frequencies" along which the energy is partitioned. The K.L. truly generalizes the Fourier series expansion.

5. Let us remark also that the $\xi_t$ are not samples in the ordinary sense; that the uniformity of the convergence of series (1-3) in $t$ is important because it legitimizes integration term by term, and that when an eigenvalue is repeated $n$ times ($n$ necessarily finite), we take any $n$ corresponding independent eigenfunctions and orthonormalize them by the Gram-Schmidt process.

We will now show a few examples:

1. **Band-limited white noise.** For this case

$$R(t,s) = \frac{\sin \omega_c (t-s)}{(t-s)}$$

The coordinate functions, i.e. the solutions of

$$\lambda \phi(t) = \int_0^T \frac{\sin \omega_s (t-s)}{(t-s)} \phi(s) ds$$

are the prolate spheroidal coordinates and have been studied extensively and tabulated [17]

2. **The Wiener process.** Here $R(t,s) = \min(t,s)$.

If we take $T = [0,1]$, then the K.L. expansion is

$$\chi(t) = \sum_{i=1}^{\infty} \frac{2^{3/2}}{(2i-1)i!} \xi_i \sin \left( \frac{i-1/2}{2} \right) \eta_i^t$$
The eigenfunctions are sinusoids, but they are not harmonically related. In fact, in section VII, we will see that a necessary and sufficient condition for the eigenfunctions to be harmonically related sinusoids is that the process be periodic in q.m.

3. Periodic processes. These are the ones for which there exists a number $T_0$ such that $R(t + T) = R(t)$, and are of course wide-sense stationary. Then the K.L. is the Fourier series:

$$X(t) = \sum_{\xi = -\infty}^{+\infty} \xi_i \exp \left[ -j i \omega \xi t \right], \quad \omega = \frac{2 \pi}{T_0}$$

SECTION II

GENERALIZED FORMS OF THE KARHUNEN-LOEVE EXPANSION

In the following, we will state Schmidt's expansion theorem for an $L^2$ kernel (defined below) in the general context of measure theory. We will show how its systematic use allows us to present in a unified manner all the results discovered so far about the K.L. in both its usual and generalized forms. As a result this will give us a better understanding of the underlying structure of the K.L., and will allow a natural and easy extension of the properties of the 'usual' K.L. to the generalized forms. In particular the proofs of theorems 2 - 3 and 2 - 4 are new; we bring in the mathematical rigor that is lacking in [9] and [56], and we examine in detail the cases where the random process is discrete or a function of several variables.
A. Schmidt's expansion theorem.

First, let us give some definitions and explain our notations. We let $\Omega_1$ and $\Omega_2$ represent two finite-measure spaces with respective measures $\mu_1$ and $\mu_2$. $L_2(\mu_1)$ and $L_2(\mu_2)$ are the corresponding spaces of square-integrable functions. In our work $\Omega_1$ will be usually a probability space. The most important measures for engineering applications are the ordinary Lebesgue measure, the Lebesgue-Stieltjes measure (in particular the probability measure), the counting measure and simple combinations of these. Then let us define an $L^2$-kernel.

Definition 2.1. We say that $K(s,t)$ with $s \in \Omega_1$ and $t \in \Omega_2$ is an $L^2$-kernel if it satisfies the following conditions:

a) $K(s,t)$ is measurable on the product space $\Omega_1 \times \Omega_2$ and

$$\int_{\Omega_1 \times \Omega_2} |K(s,t)|^2 \, d\mu_1(x) \, d\mu_2(y) < \infty \quad (2-1)$$

b) For each value of $s$, $K(s,t)$ is a measurable function of $t$ and

$$\int_{\Omega_2} |K(s,t)|^2 \, d\mu_2(t) < \infty \quad (2-2)$$

c) For each value of $t$, $K(s,t)$ is a measurable function of $s$ and

$$\int_{\Omega_1} |K(s,t)|^2 \, d\mu_1(s) < \infty \quad (2-3)$$

d) $K(s,t)$ is not equivalent to the null function.

Let us remark that the requirement of measurability is not very stringent in physical problems, since a measurable function can be roughly defined as one which is continuous except on a set of measure
zero. Condition (2-1) is the strongest of all. It in fact implies that (2-2) and (2-3) hold, except possibly on sets of measure zero with respect to \( \mu_1 \) and \( \mu_2 \) respectively. Conditions b), c), and d) are therefore needed to obtain strong results in the following.

Some background is in order at this point about non-Hermitian \( L^2 \) kernels. Their properties are studied by considering the adjoint kernel \( K^* \) of \( K \), i.e.

\[
K^*(s, t) = \overline{K(t, s)}
\]

where the upper bar denotes the complex conjugate. Then the kernels \( K^*K \) and \( KK^* \) are Hermitian and non-null, and they possess sets of eigenvalues and eigenfunctions designated by respectively, i.e.

\[
\lambda_n e^m_n = KK^* \mu_n
\]

\[
\lambda_n e^m_n = K^*K \nu_n
\]

The above compact operator notation means the following:

\[
Ku = \int_{\Omega_2} K(s, t) u(t) d\mu_2(t) = (Ku)(s)
\]

\[
KK^*u = \int_{\Omega_2} K(s, t) d\mu_2(t) \int_{\Omega_1} K^*(t, \tau) \mu_n(\tau) d\mu(\tau)
\]

\[
= \int_{\Omega_2} K(s, t) d\mu_2(t) \int_{\Omega_1} \overline{K(\tau, t)} \mu_n(\tau) d\mu_1(\tau)
\]

\[
KK^*u_n = \int_{\Omega_2 \times \Omega_1} K(s, t) \overline{K(\tau, t)} \mu_n(\tau) d\mu(\tau) d\mu_2(t)
\]

\[
= (KK^*u_n)(s)
\]
Therefore, by (2-4) we have

$$\lambda_{n}^{e} \mu_{n}(s) = \int_{\Omega_{2}} \int_{\Omega_{1}} K(s,t) \overline{K(t,r)} \mu_{n}(r) d\mu_{1}(r) d\mu_{2}(t)$$

(2-4)

Similarly, by (2-5) is meant that

$$\lambda_{n}^{e} \nu_{n}(t) = \int_{\Omega_{1}} \int_{\Omega_{2}} K(s,t) \overline{K(s,r)} \nu_{n}(r) d\mu_{1}(s) d\mu_{2}(r)$$

or

$$\lambda_{n}^{e} \nu_{n}(t) = \int_{\Omega_{2}} \int_{\Omega_{1}} K(s,t) K(s,r) \nu_{n}(r) d\mu_{1}(s) d\mu_{2}(r)$$

(2-5)

Let us remark that we have been using Fubini's theorem to go from

$$\int_{\Omega_{1}} \int_{\Omega_{2}} f \quad \rightarrow \quad \int_{\Omega_{1}} \int_{\Omega_{2}} f$$

since we know that \( \mu_{n}(s) \) and \( \nu_{n}(t) \) do exist.

Then, taking \( \lambda_{n} \) to be the positive square root of \( \lambda_{n}^{e} \), it can be shown [8, p. 143-144] that:

**Theorem 2.1.** For every \( L^{2} \)-kernel \( K(s,t) \) there exists a system \( (\mu_{n}, \nu_{n} ; \lambda_{n}) \) called singular for \( K \) such that:
\[ \lambda_n \mathcal{M}_n(s) = \int_{\Omega_2} K(s,t) \nu_n(t) \, d\mu_2(t). \]  
(2-6)

\[ \lambda_n \nu_n(t) = \int_{\Omega_1} \overline{K}(t,s) \mathcal{M}_n(s) \, d\mu_1(s). \]  
(2-7)

From the theory of Hermitian kernels [8], the systems \((\mathcal{M}_n)\) and \((\nu_n)\) can always be chosen to be orthonormal individually, i.e.

\[ \int_{\Omega_1} \mathcal{M}_n(s) \mathcal{M}_m(s) \, d\mu_1(s) = \delta_{nm} \]  
(2-8)

\[ \int_{\Omega_2} \nu_n(t) \nu_m(t) \, d\mu_2(t) = \delta_{nm} \]  
(2-9)

Under these conditions we have the important expansion theorem due to E. Schmidt.

**Theorem 2.2.** With respect to the singular system \((\mathcal{M}_n, \nu_n; \lambda_n)\) associated with \(K(s,t)\), and with \(\lambda_n \geq \lambda_{n+1}, \lambda_n > 0\), we have

\[ K(s,t) = \sum_{n=1}^{\infty} \lambda_n \mathcal{M}_n(s) \nu_n(t) \quad (L^2) \]  
(2-10)
where the sign $\equiv$ indicates that the equality is almost everywhere.

The proof is given in Appendix A-1, and is a simple generalization of the one given in [8].

B. Some concrete cases.

The connection between theorem 2.2 and the K.L. expansion was apparently first noticed by H. P. Kramer [9] but has remained largely unexploited. In what follows it will become clear that the K.L., in its usual and generalized forms, is but one concrete form of theorem 2.2. Let therefore

$$K(s,t) = X(\omega, t) = \tilde{X}(t)$$

where $X(t)$ is a second-order process

$$E \{ |X(t)|^2 \} = \int_{\Omega} |X(\omega, t)|^2 d\mu(\omega) < \infty$$

(2-3)'

where we have taken the probability space $\Omega$ to be $\Omega_1$, with measure $\mu_1$ equal to the probability distribution.

We let also

$$\Omega_2 = [0, T] = \text{closed interval}$$
\[ \mu_s(t) = t = \text{ordinary Lebesgue-measure} \]

Since \( \Omega_2 \) is finite and \((2-3)'\) holds, \( X(\omega,t) \) is an \( \mathcal{L}^2 \) kernel, and we can write

\[
X(t) = X(\omega,t) = \sum_n \lambda_n \mathcal{X}_n \Phi_n(t)
\]

(2-10)

with \( \mathcal{X}_n \) defined by \((2-5)\)' and \((2-6)\) respectively; i.e.

\[
\lambda_n \mathcal{X}_n = \int_T \int \Omega X(\omega,t) X(\omega,\tau) \Phi_n(\tau) d\mu(\omega) d\mu(\tau)
\]

(2-5)

or

\[
\lambda_n \mathcal{X}_n = \int_T \mathcal{R}(t,\tau) \Phi_n(\tau) d\mu(\tau)
\]

(2-6)

In all the following, we will assume the autocorrelation function \( \mathcal{R} \) to be real. Let us remark however, that \((2-10)\) is not quite the K.L. expansion, as we have not established the uniform convergence of the series in \((2-10)\) with respect to \( t \). And we are interested in this property because it allows important time operations (e.g. integration) to be performed equally well on both members of \((2-10)\). We cannot expect uniform convergence unless we require stronger conditions on \( \mathcal{X}(t) \), for instance, continuity in the quadratic mean (q.m.). We then assume that
is continuous in both variables $t$ and $\tau$ in the domain of interest, 
$\Omega_2 \times \Omega_2$. Moreover, we will assume that $\Omega_2$ is a compact space, 
as this condition is needed for the proof of the property that follows:

Lemma 2.1. If $X(t)$ is continuous in q.m. and if $t \in \Omega_2$, where 
$\Omega_2$ is compact, then the convergence in (2-10) is uniform in $t$.

The proof is given in Appendix A-3.

From now on, we will assume the continuity in the mean of the 
random function, and, consider in some detail concrete cases.

a) The K.L. expansion extends trivially to the case

$\Omega_2 = \prod_{i=1}^{D} \tau_i = \prod_{i=1}^{D} [\alpha_i, \beta_i]$ 

where $[\alpha_i, \beta_i]$ is a closed bounded interval with the usual Lebes-
gue measure; i.e., we can write

$X(t, t_2, \ldots, t_D) = \sum_{n=1}^{\infty} \lambda_n \omega_n \phi_n(t_1, t_2, \ldots, t_D)$

with

$\lambda_n \phi_n(t_1, \ldots, t_D) = \int_{\prod \tau_i} R_X(t_1, \ldots, t_D, \tau_1, \ldots, \tau_D) \phi_n(\tau_1, \ldots, \tau_D) d\tau_1 \ldots d\tau_D$

and

$\int_{\prod \tau_i} \phi_n(t_1, \ldots, t_D) \phi_m(t_1, \ldots, t_D) dt_1 \ldots dt_D = \delta_{nm}$

$E\{\omega_n \omega_m\} = \delta_{nm}$

It is remarkable that the random coefficients $\omega_n$ are not functions 
of any of the time-variables $A_i$. 
Let us note that the $t_i$'s are not necessarily coordinates in the same physical space. The representation could be, therefore, useful for time-space varying fields (like turbulence and electromagnetic fields) in a bounded domain.

All that is needed from $\Omega$ is that it be closed and bounded. For instance we could imagine sending a random signal over a finite succession of closed intervals and that we would like the corresponding K.L. expansion. Obviously the matter of determining the K.L. would become very complicated, if not totally unmanageable. But for theoretical purposes, such a K.L. expansion could be very helpful.

Let us note that when $R(t_1, ..., t_D; \tau_1, ..., \tau_D) = R(t, \tau)$ is stationary and isotropic (or "spherically" symmetrical), i.e. when

$$R(t, \tau) = R(|t-\tau|) = R(\rho)$$

where

$$\rho^2 = (t_1-\tau_1)^2 + (t_2-\tau_2)^2 + ... + (t_D-\tau_D)^2$$

then it is clear that the corresponding eigenfunctions are also functions of $\rho$ alone, i.e.

$$\phi_n(t_1, ..., t_D) = \phi_n(\rho)$$

as this could be seen from the expansion (Mercer's):

$$R(|t-\tau|) = \sum_n \lambda_n^2 \phi_n(t) \phi_n(\tau)$$

$$= \sum_n \lambda_n^2 \phi_n(|t-\tau|) \phi_n(\tau)$$
The determination of the eigenfunctions becomes therefore a lot easier. The practical importance of the isotropic case would be seen by analogy to the results obtained by P. Faure in optical experiments [10]. As an example of this situation, we note that the theory of isotropic turbulence is used to describe the state of the atmosphere and to some extent that of the ocean. In particular, if \( \mu(x,y,z,t) \) denotes the refraction index, it is found that its autocorrelation

\[
E \left\{ k(x_1,y_1,z_1,t) \kappa(x_2,y_2,z_2,t) \right\} = \mathcal{N}(\sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2})
\]

is well approximated in some conditions by

\[
\mathcal{N}(\rho) = e^{-\rho^2/\alpha^2} \quad \text{or} \quad \mathcal{N}(\rho) = e^{-\rho/\alpha}
\]

where \( \alpha \) is an autocorrelation distance, cf[53, chapt. I].

b) Up to now, we have been considering only the case where

\( \mu_{\varphi} = \text{usual Lebesgue measure} \)

But Schmidt's theorem shows the results to be valid if we take

\( \mu_\chi = \text{Lebesgue-Stieltjes measure} = \int_0^t \beta(\tau) \, d\tau , \quad \varphi(t) > 0 \)

\( T \) closed and bounded

Then

\[
\chi(t) = \sum \lambda_i \chi_i \varphi_i(t) , \quad t \in T
\]
Let us note that here we have the following defining relations

\[ \lambda_i \Phi_i(t) = \int_\tau \mathcal{R}(t, \tau) \Phi_i(\tau) d\tau \]

and the orthogonality relations

\[ \int_\tau \Phi_i(t) \Phi_j(t) \mathcal{P}(t) dt = \delta_{ij} \]

\[ E\{ \varepsilon_i \varepsilon_j \} = \delta_{ij} \]

The presence of the weighting function \( \mathcal{P}(t) \) allows to extend the definition of a K.L. for an infinite interval, and to study the variations of the usual K.L. as \( T \rightarrow \infty \). Moreover, it is needed whenever we want to focus our attention around a specific region and give it more importance.

As an example, we consider the random process \( X(t) \) whose autocorrelation function is

\[ \mathcal{R}(t, s) = \exp(-\alpha |t-s|) \]

and expand it into the generalized K.L. over \( [0, \infty) \), with weight

\[ \mathcal{P}(t) = \exp\left[-\varepsilon \mu t\right] \]
Then the \( \Phi_j(t) \) are the Bessel functions [34]

\[
\Phi_j(t) = J_\mu(\lambda_j \exp[-\mu t])
\]

the \( \lambda_j \) in this equation being the positive roots of

\[
J_{\mu-1}(\lambda_j) = 0
\]

c) If we now apply (2-10) to the case where \( \Omega \) is still the probability space, and

\( \Omega_2 = \{1, 2, \ldots, m, \ldots\} = \text{finite or infinite set of integers} \)

\( \mu_e = \text{the counting measure} \)

we would be considering the random vector \( \vec{X}(\omega) = \vec{X} \) defined by its components

\[
\vec{X}^\tau = \{X_1, X_2, \ldots, X_m, \ldots\}
\]

where the superscript \( T \) stands for the transpose.

This we get the discrete version of the usual K.L., assuming of course that conditions (2-1), (2-2) and (2-3) to be fulfilled here,

\[
X_i = \sum_n \lambda_n \alpha_n \beta_n^i
\]

\( i = 1, 2, \ldots, m, \ldots \)

\( n = 1, 2, \ldots \)

where the random variables \( \alpha_n \) and the matrix elements \( \beta_n^i \) are given by

\[
\alpha_n \beta_n^i = \sum_j R_{ij} \beta_n^j
\]
\[ \lambda_n x_n = \sum x_i \beta_n^i \]  

(2-13)

with

\[ R_{ij} = E \{ x_i x_j \} \]

Moreover, the orthogonality conditions here yield

\[ E \{ \alpha_n \alpha_m \} = \delta_{nm} \]

\[ \sum \beta_n^i \beta_m^i = \delta_{nm} \]

Let us note that if we let \( \mathbf{R} \) to be the covariance matrix of the vector, i.e.

\[ \mathbf{R} = (R_{ij}) = \begin{bmatrix} E(x_1 x_1) & E(x_1 x_2) & \ldots \\ E(x_2 x_1) & E(x_2 x_2) & \ldots \\ \vdots & \vdots & \ddots \end{bmatrix} \]

and \( \hat{\beta}_n \) to be the vector defined by

\[ \hat{\beta}_n^T = (\beta_1^1, \beta_1^2, \ldots, \beta_1^n, \ldots) \]

equation (2-12) is to be rewritten as

\[ \lambda_n \hat{\beta}_n = \mathbf{R} \hat{\beta}_n \]

This clearly shows that if \( \mathbf{x} \) is a (finite) \( M \)-dimensional random vector, the deterministic orthonormal vectors \( \hat{\beta}_n \) are the eigenvectors of the \( M \times M \) matrix \( \mathbf{R} \) and the \( \lambda_n^2 \) are the corresponding eigenvalues. In this case \( n \leq M \).

A slightly modified form of the discrete K. L. for the finite dimensional vector was introduced by Hotelling [11] [64] in the case
where the \( X_i \) are uniformly spaced time-samples of the process \( X(t) \).

Hotelling called \( \alpha_n \)'s "the principal components" of \( \tilde{X} \), with the important practical constraint that \( n < M \); usually, the number \( n \) of K.L. coefficients \( \alpha_n \) taken in the Hotelling technique is much smaller than the number of samples \( M \). The rigorous justification of the use of this technique to characterize \( X(t) \) over an interval \([0, T]\) is that, though \( n < M \), the K.L. coefficients allow information compression, as will be seen in Section III; moreover the \( X_i \) are uncorrelated, though the \( X_i \) may not be (in fact, we might have oversampling). The \( \alpha_n \) become independent whenever \( X(t) \) is Gaussian.

Hotelling principal components, and therefore the discrete K.L., can be implemented by digital filters. This therefore allows an approximate implementation of the usual K.L. [11].

d) One of the most interesting generalizations of the K.L. is obtained by applying theorem 2.2 to the case where \( \Omega \), is the random space and

\[
\Omega_2 = \text{product space of } \{1, 2, \ldots, n, \ldots\} \otimes \mathcal{T}
\]

\( \mu_2 = \text{product measure of (counting measure)} \otimes \mathcal{T} \)

With these definitions, we immediately get a theorem first proved by Kelly and Root [12] when \( n \) is infinite.

**Theorem 2.3.** Let \( \tilde{X}(t) \) be a vector-valued random variable, with components \( \{X_i(t), X_i(t), \ldots\} \) defined in a finite interval \( T = [0, T] \).

Suppose the correlation functions

\[
R^{i, j}(s, t) = E \left\{ \tilde{X}^i(s) \tilde{X}^j(t) \right\}
\]
all exist and are continuous in \( T \times T \) and are uniformly bounded.

Suppose further that
\[
\sum_{i,j} \int_0^T \int_0^T |R^{i,j}(s,t)|^2 \, ds \, dt < \infty
\]
and that there exist real constants \( C_j; j = 1, 2, \ldots \) for which
\[
\int_0^T |R^{i,j}(s,t)|^2 \, dt \leq C_j^2, \quad i = 1, 2, \ldots
\]
and such that
\[
\sum_{j=1}^{\infty} C_j^2 < \infty
\]
Then the random process (vector) \( \tilde{X}(t), t \in T \) has the representation
\[
\tilde{X}^i(t) = \sum_{k} \lambda_k \tilde{x}_k \phi^i_k(t)
\]
with
\[
\tilde{x}_k = \sum_{j=1}^{\infty} \int_0^T \tilde{X}^j(t) \phi^j_k(t) \, dt
\]
and the \( \phi^i_k(t) \) are eigenfunctions given by
\[
\sum_{j=1}^{\infty} \int_0^T R^{i,j}(t,s) \phi^j_k(s) \, ds = \lambda_k^2 \phi^i_k(t)
\]
Moreover, we have the orthogonality relations
\[
\mathbb{E} \{ \tilde{x}_k \bar{\tilde{x}}_j \} = \delta_{kj}
\]
\[
\sum_{i=1}^{\infty} \int_0^T \phi^i_k(t) \bar{\phi}^i_k(t) \, dt = \delta_{kj}
\]
In vector notation theorem 2.3 can be written as:

\[ \tilde{X}(t) = \sum_{k} \lambda_k \tilde{X}_k \tilde{\Phi}_k(t) \]

with

\[ \lambda_n \tilde{\Phi}_k(t) = \int_{T} R(t,s) \tilde{\Phi}_k(s) \, ds \]

where \( R(t,s) = (R^{ij}(t,s)) \) = correlation function matrix. We see the advantage of the enunciation of theorem 2.2 within the general frame of measure theory, as it gives the conditions and proof of theorem 2.3 naturally and immediately. (Compare with [12]).

Let us remark at this point that M. Loeve [13] seems to have been the first one to use the technique of considering a vector random process to be just like a random function of two variables, a technique that was systematized in the engineering literature by V. Pugachev [14].

\[ e) \] But as we have considered \( \Omega_2 \) to be a product space, while \( \Omega_1 \) was the simple random space, that same technique yields

\[ \Omega_1 = \Omega \otimes \Omega' = \text{Product space of usual and discrete probability spaces} \]

\[ \Omega_2 = \text{usual space T} \]

with \( \mu_1 = (\text{probability measure}) \otimes (\text{discrete probability measure}) \)

\( \mu_2 = \text{ordinary Lebesque measure} \)

the following result due first to Chien & Fu [15].

**Theorem 2.4.** Let \( \{ \tilde{X}_v(t), 0 \leq t \leq T \}, v = 1, 2, \ldots, m \) be the m stochastic processes generating the zero-mean random functions \( \tilde{X}_v(t) \).
Let $p^*_i$ be the probability of occurrence of the $i^{th}$ process.

Then we have the representation

$$X_i^*(t) = \sum_{k=1}^\infty \chi_k(t) \phi_k(t), \quad i = 1, 2, \ldots, m$$

where we have the defining equations

$$\lambda_k \phi_k(t) = \int_0^T R(t, s) \phi_k(s) \, ds$$

$$\sum_{i=1}^m p_i \mathbb{E}\{X_i(t) \overline{X}_i(s)\} = R(t, s)$$

and the orthogonality relations

$$\int_0^T \phi_k(t) \overline{\phi}_l(t) \, dt = \delta_{kl}$$

$$\sum_{i=1}^m p_i \mathbb{E}\{\chi_k \overline{\chi}_l\} = \lambda_k \delta_{kl}$$

We sketch the proofs of theorems 2.3 and 2.4 from theorem 2.2.

For theorem 2.3, we have

$$X(t, \omega) = X(\omega, t, \omega) = \sum_k \chi_k \phi_k(t, \omega)$$

$\mathcal{A} \in [0, T]$

$\omega \in [1, 2, \ldots, \iota, \ldots]$
we get
\[ \hat{\chi}^i(t) = \sum_k x_k^i \phi_k^i(t) \]

with the orthogonality relations enunciated in the theorem.

Similarly, we have the dual of theorem 2.3, namely theorem 2.4.

We take
\[ \hat{\chi}(t) = \chi(\omega, \omega'; t) \]
\[ t \in [0, T] \]
\[ \omega \in \Omega \]
\[ \omega' \in \Omega' \]
\[ \mu(\omega), \mu(\omega') \]
\[ \nu_1, \nu_2, \ldots, \nu_m \]
\[ \hat{\chi}(t) = \sum_k x_k \chi(\omega, \omega') \phi_k(t) \]

with
\[ \int_T \phi_k(t) \phi_l(t) \, dt = \delta_{kl} \]

Also
\[ E \left\{ x_k(\omega, \omega') x_l(\omega, \omega') \right\} = \int_{\Omega} \int_{\Omega'} x_k(\omega, \omega') x_l(\omega, \omega') \, d\mu(\omega) \, d\mu(\omega') \]

But
\[ \int_{\Omega'} x_k(\omega, \omega') x_l(\omega, \omega') \, d\mu(\omega) = \sum_{\alpha=1}^m \nu_\alpha x_k(\omega) x_l(\omega) \]
and finally

\[
\mathbb{E} \left\{ x_k^l (\omega, \omega') \right\} = \int_{\Omega} \sum_{i} \mathcal{P}^i \ x_k^l (\omega, \omega') \ d\mu (\omega)
\]

\[
= \sum_{i=1}^{m} \mathcal{P}^i \ \mathbb{E} \left\{ x_k^l \mathcal{N} \right\} = \lambda^2 \delta_{k,l}
\]

and this proves the result. If \( m \) is infinite, the theorem still holds.

An interesting application of theorem 2.4 is found in pattern recognition. Whenever we have a random function which is the outcome of one of several different stochastic processes of which we know the relative probability, this expansion is useful in pre-weighting the features according to their relative importance.

Let us note that we could also use the K.L. expansion in the following form

\[
\phi_k (t, \omega) = \sum_{i=1}^{\infty} \phi_i (t) \phi_i (\omega; t), \quad \omega \in [0, a]
\]

where \( t \) is a parameter which could take any value. Then

\[
\mathbb{E} \left\{ \phi_k (t) \phi_j (t) \right\} = \lambda^2 (t) \delta_{ij}
\]

and

\[
\int_{0}^{a} \phi_i (x; t) \phi_j (x; t) \ dx = \delta_{ij}, \text{ for any } t!
\]
For $t$ given, we have of course

$$\lambda_i^\prime(t) \geq \lambda_{i+1}^\prime(t)$$

and

$$\lambda_i^\prime(t) \phi_i(x; t) = \int_0^a E \left\{ \frac{f_i}{y_i} (t, x) \rho_i (t, x') \right\} \phi_i (x'; t) \, dx'$$

This kind of expansion has been found useful [56]; and this illustrates again the great versatility of the K.L. expansion.

Let us conclude this section by remarking that we have not exhausted all possible extensions of the K.L. In fact, by selecting and combining the spaces $\Omega_1$ and $\Omega_2$ and their respective measures $\mu_1$ and $\mu_2$ in different ways, we will obtain other natural generalizations when needed. For the sake of clarity, we summarize the above results in Table I.
TABLE 1
GENERALIZED FORMS OF THE K.L. EXPANSION BASED ON THE SCHMIDT'S EXPANSION THEOREM

<table>
<thead>
<tr>
<th>$\Omega_1$ Space</th>
<th>$\Omega_2$ Space</th>
<th>$\mu_1$ Measure</th>
<th>$\mu_2$ Measure</th>
<th>K.L. Expansion</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Space $\Omega$</td>
<td>Closed Bounded Interval $\mathcal{T}$</td>
<td>Probability Measure $\mu(\omega)$</td>
<td>Ordinary Lebesque Measure $\mu$</td>
<td>$\chi(t) = \sum_{i=1}^{\infty} \xi_i \phi_i(t)$</td>
<td>This is the usual K.L.</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>Real Line $\mathbb{R}$ or any Compact Subset $\mathcal{T}'$ of it</td>
<td>Probability Measure $\mu(\omega)$</td>
<td>Lebesgue-Stieltje's Measure $\mu'$</td>
<td>$\chi(t) = \sum_{i} \xi_i \phi_i(t)$</td>
<td>The $\phi_i(t)$ are orthogonal with a weight $\int_{\Omega_2} \phi_i(t) p(t) dt = \delta_{ij}$</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>Multi-dimensional Euclidian space $\prod_{i=1}^{n} \mathcal{T}_i$ closed and bounded</td>
<td>Probability Measure $\mu(\omega)$</td>
<td>Usual Lebesque Measure $\mu$</td>
<td>$\chi(t_1, \ldots, t_n) = \sum_{i} \xi_i \phi_i(t_1, \ldots, t_n)$</td>
<td>Useful in the study of space-time varying fields, etc.,. Computational difficulties. Simple case of homogeneous isotropic fields.</td>
</tr>
<tr>
<td>Space</td>
<td>Space</td>
<td>Measure</td>
<td>Measure</td>
<td>K.L. Expansion</td>
<td>Remarks</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>---------</td>
<td>---------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>Random Space</td>
<td>set ((1, 2, \ldots, N)) or ((1, 2, \ldots, N, \ldots))</td>
<td>(\mu(\omega))</td>
<td>Counting Measure</td>
<td>A random vector variable (\hat{X} = \sum_{i} \lambda_i \beta_i \alpha_i)</td>
<td>Especially useful for (N) finite. (X', \ldots, X'') may be real-time samples of the process (X(t)). (X_i' = X(t_i), \ldots, X'_N = X(t_N)). Replaces correlated samples by &quot;equivalent&quot; uncorrelated variables. Very helpful in approximation methods. A modified version is in the Hotelling technique.</td>
</tr>
<tr>
<td>(\Omega)</td>
<td>(T \times (1, 2, \ldots, N)) or (T \times (1, 2, \ldots, N, \ldots))</td>
<td>Probability Measure (\mu(\omega))</td>
<td>Product Measure &amp; Counting measure (\hat{X}(t) = \sum_{i} \lambda_i \phi_i(t)) (E{x_i' x_j} = \delta_{ij}) (\lambda_i \beta_i = \sum_{k} R_{ij} \beta_k)</td>
<td>This is the K.L. expansion for the vector random process. (X'(t) \rightarrow (x_i', \ldots, \ldots)) (X''(t) \rightarrow (x_i, \ldots, \ldots)) Especially useful for detection model in seismic, antenna, etc., arrays.</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 1 (continued)

<table>
<thead>
<tr>
<th>$\Omega_1$ Space</th>
<th>$\Omega_2$ Space</th>
<th>$\mu_1$ Measure</th>
<th>$\mu_2$ Measure</th>
<th>K.L. Expansion</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product of Spaces $\Omega \times \Omega'$</td>
<td>$\mathcal{T}^1$</td>
<td>Produce Measures $\mu(\omega) \otimes \mu'(\omega')$</td>
<td>Usual Lebesque Measure $\mu$</td>
<td>$X_k^2(t) = \sum_{k=1}^{\infty} X_k \phi_k(t)$</td>
<td>Useful in pattern recognition for instance. Allows pre-weighting of the features according to their relative importance. Structure is the dual of the preceding case.</td>
</tr>
<tr>
<td>$\Omega'$ = discrete probability space</td>
<td>$\mathcal{T}^1$</td>
<td>$\mu(\omega')$ = discrete probability measure</td>
<td></td>
<td>$\lambda_k^2 = \int \mathcal{R}(t,z) \phi_k(z) dz$</td>
<td></td>
</tr>
</tbody>
</table>

Remark: The above cases are not exhaustive.
SECTION III

GENERAL PROPERTIES OF THE K.L. EXPANSION

Having recognized the underlying structural unity of the different generalized K.L. expansions, namely that they constitute but the concrete aspect of Schmidt's expansion theorem in the special case where \( \Omega_1 \) is the random space, it becomes an easy matter to realize that Schmidt's approximation theorem which we state below applies to them in the most straightforward manner. In particular the properties of corollary 3:1 and theorem 3.2 have not been explicitly recognized before for most of the generalized forms of the K.L.

A. Schmidt's approximation theorem and minimization of average truncation error

**Theorem 3.1.** (Schmidt) With the same notations as in Section II, and \( (\mu_n, \nu_n; \lambda_n) \) being a singular system for the \( \mathcal{L}^2 \)-kernel \( K(s,t) \), we let

\[
K_N(s,t) = \sum_{n=1}^{N} \lambda_n \mu_n(s) \nu_n(t)
\]

(3-1)

If \( a_1, a_2, \ldots, a_N \) and \( b_1, b_2, \ldots, b_N \) are arbitrary \( L_2(\mu_1) \) and \( L_2(\mu_2) \) functions respectively, then

\[
\int_{\Omega_1 \times \Omega_2} \left| K(s,t) - \sum_{n=1}^{N} a_n(s) b_n(t) \right|^2 d(\mu_1 \times \mu_2)(s,t) \geq \sum_{n=N+1}^{\infty} \lambda_n^2
\]

(3-2)
The proof is found in appendix A-4.

Let us now consider a zero-mean continuous (in q.m.) second-order stochastic process \( \tilde{X}(t) \) on the closed bounded interval \( T \). If \( a_1, a_2, \ldots, a_N \); and \( b_1(t), b_2(t), \ldots, b_N(t) \) are \( N \) arbitrary random variables and \( N, L^2 \) -functions respectively, we can state immediately the following corollary:

**Corollary 3.1.** Given the autocorrelation function of a zero-mean second order process \( \tilde{X}(t) \), among all the approximations of \( \tilde{X}(t) \) the form

\[
\sum_{n=1}^{N} a_n b_n(t), \quad N \text{ given}, \quad t \in T
\]  

(3-3)

the K.L. minimizes the average truncation-error on the energy of the process, i.e.

\[
\int_T E \left| \tilde{X}(t) - \sum_{n=1}^{N} a_n b_n(t) \right|^2 dt \geq \int_T E \left| \tilde{X}(t) - \sum_{n=1}^{N} \lambda_n \phi_n(t) \right|^2 dt = \sum_{n=N+1}^{\infty} \lambda_n = R_N
\]  

(3-4)

This property was discovered by Koschmann, Jordan and Brown in a more restricted context; and finally brought to this more general version by H. P. Kramer [9].

\( \alpha \)** Clearly, the property stated in corollary 3.1 applies to all the cases (a) through (e) of section II. More precisely, in the case
where we have a random vector

\[ \mathbf{\bar{X}}^T = (X_\mathbf{i}, X_\mathbf{e}, \ldots, X_\mathbf{i}, \ldots) \]

and in particular when the vector is a finite dimensional one (e.g. \(X_\mathbf{i}, X_\mathbf{e}, \ldots, X_\mathbf{i}\) represent samples of a process taken at uniform intervals), then the best approximation to that vector, in the sense that

\[
\sum_{i=1}^{N} E_i |X_i - \sum_{n=1}^{N} a_n b_n^i|^2 = \text{minimum}
\]

over all sets \((a_n)\) and \((b_n^i)\), for \(N\) given, is the associated "discrete" K.L. In other words, \((3-5)\) represents the minimization of a global energy error. Let us look to the problem more in detail; we have:

\[ X_\mathbf{i} = a_1 b_1^i + a_2 b_2^i + \ldots + a_N b_N^i + \mathbf{\bar{\varepsilon}}^i \]

\[ X_\mathbf{e} = a_1 b_1^e + a_2 b_2^e + \ldots + a_N b_N^e + \mathbf{\bar{\varepsilon}}^e \]

\[ X_\mathbf{i} = a_1 b_1^i + a_2 b_2^i + \ldots + a_N b_N^i + \mathbf{\bar{\varepsilon}}^i \]

or

\[ \mathbf{\bar{X}} = \mathbf{B} \mathbf{\bar{A}} + \mathbf{\bar{\varepsilon}} \]

Let us note that \(N\) may be \(\geq I\).

What the theorem says is that, to achieve condition \((3-5)\), one needs to.
consider the \((I \times I)\) covariance matrix

\[
\begin{bmatrix}
E(X_1 X_1) & E(X_1 X_2) & \cdots & E(X_1 X_I) \\
E(X_2 X_1) & E(X_2 X_2) & \cdots & E(X_2 X_I) \\
\vdots & \vdots & \ddots & \vdots \\
E(X_I X_1) & E(X_I X_2) & \cdots & E(X_I X_I)
\end{bmatrix}
\]

and to find its eigenvalues \((\lambda_1^2, \lambda_2^2, \ldots, \lambda_I^2)\) and its eigenvectors \(\beta^2_n, n = 1, 2, \ldots, I\). Then this is the coordinate set that will minimize \(\sum_{i=1}^{I} E[(e_i)^2]\).

\(\beta\) The case where we want a minimum biased average-truncation error on the energy of the process, i.e. when we want to weight the error in a certain time interval more heavily with respect to the rest of it, we use theorem 3.1 with \(\mu_2(t)\) equal to the Lebesgue-Stieltjes measure, i.e.

\[
\mu_2(t) = \int_{-t}^{t} \varphi(t) \, dt, \quad \varphi(t) > 0
\]

In that case (3-2) is written as

\[
\int_{T'} \left| X_n(t) - \sum_{n=1}^{N} a_n \cdot b_n(t) \right|^2 \varphi(t) \, dt \\
\int_{T'} \left| X_n(t) - \sum_{n=1}^{N} \lambda_n \cdot x_n(t) \right|^2 \varphi(t) \, dt = \sum_{n=N}^{\infty} \lambda_n^2
\]

(3-6)
The global truncation-error over the energy for an array (finite or infinite, of random processes, is minimized also by the use of the K.L. expansion. Precisely, in that case we have:

\[
\int_T \sum_{a} E \left| X^\pi(t) - \sum_{n=1}^{N} a_n \phi_n^\pi(t) \right|^2 dt \\
\int_T \sum_{a} E \left| X^\pi(t) - \sum_{n=1}^{N} \lambda_n \phi_n^\pi(t) \right|^2 dt = \sum_{n=N+1}^{\infty} \lambda_n^2
\]  

Similarly, for the case encountered in pattern recognition [15] where we have \( m \) preweighted stochastic processes \( X^\pi(t) \), \( \pi = 1, ..., m \), and already mentioned in section II, among all approximations

\[
X(t) = \sum_{n} a_n \phi_n(t)
\]

the corresponding K.L. minimizes the average global truncation-error

\[
\int_T \sum_{\pi=1}^{m} p_\pi E \left| \xi^\pi(t) \right|^2 dt
\]

where \( p_\pi \) is the probability of occurrence of process \( X^\pi(t) \).

Cases (\( \alpha \)) through (\( \delta \)) are meant to be illustrative only; and the scope of application of theorem 3.1 is much wider.

B. Minimization of an entropy function

Another property of the K.L. has been discovered [3] which is a particular, but important, consequence of theorem 3.1. In all subsec-
tion (B), we shall assume that \( X(t) \) is normalized, i.e.

\[
\int_{\Omega} |X(t)|^2 \, d\mu_z(t) = 1 \tag{3-10}
\]

and this is possible in view of the requirement that \( X(t) \) be \( L^2 \).

Condition (3-10) is to be assumed for all cases of application of theorem 2.2 and theorem 3.1. Then we can state

**Theorem 3.2.** Of all representations of \( X(t) \) of the form

\[
\tilde{X}(t) = \sum_{n} a_n b_n(t) , \quad t \in T
\tag{3-11}
\]

where

\[
\int_{T} b_n(t) b_m(t) \, dt = \delta_{nm}; \quad E\{ |a_n|^2 \} = \mu_n^2
\tag{3-12}
\]

and the \( a_n \)'s are the Fourier coefficients corresponding to the \( b_n(t) \)'s, the Karhunen-Loève expansion is the one which minimizes the entropy function \( H(\{ b_n(t) \}) \) defined by

\[
H(\{ b_n(t) \}) = -\sum_n \mu_n^2 \ln \mu_n^2 \tag{3-13}
\]

i.e.

\[
H(\{ \tilde{b}_n(t) \}) = \min_{\{ b_n(t) \}} H(\{ b_n(t) \})
\tag{3-14}
\]

The proof is based on the following lemma which is proved in Appendix A-5.
Lemma 3.1. For the inequality

$$-\sum_{n=1}^{\infty} \lambda_n^2 \ln \lambda_n \leq -\sum_{n=1}^{\infty} \mu_n^2 \ln \mu_n$$

(3-15)

to hold, it is sufficient that

$$\sum_{n=1}^{N} \mu_n^2 \leq \sum_{n=1}^{N} \lambda_n^2, \quad \forall \ N$$

(3-16)

Proof of (3-16). We can write, with the notations of A-4

$$\int_{T} E \left| X(t) - \sum_{n=1}^{N} a_n \ b_n(t) \right|^2 dt = \left\| X \right\|^2 + \sum_{n=1}^{N} \left( a_n - X \ b_n, a_n - X \ b_n \right)$$

$$- \sum_{n=1}^{N} \left( X \ b_n, X \ b_n \right)$$

Since

$$a_n = X \ b_n = \int_{T} X(t) \ b_n(t) \ dt \quad (i.e. \ a_n \ is \ a \ Fourier \ coefficient)$$

and because of Ky Fan's theorem which we quote in section VII (cf. [8], p. 134), we see that (3-16) is verified.

Let us note that lemma 3.1 is general; and therefore the application of theorem 3.2 extends immediately to all the generalized forms of the K.L. we have been considering.

We will not go into the details of the concrete applications. Let us, however, remark that the property stated in theorem 3.2 was intuitively to be expected. What it says roughly is
that the K.L. expansion is the ordering map of a stochastic process into its "natural" (usually infinite) system of coordinate functions, because it takes into consideration first the random coefficient with the largest variance, in other words, the one with the greatest uncertainty or equivalently, the greatest amount of information. This explains the useful role of the K.L. in the ranking of feature measurements and pattern recognition.

C. Remark.

One might ask whether the K.L. minimizes the average truncation error on the power of a random process. Straightforward calculations using the methods of the calculus of variations show the answer to be negative. But as the number of terms taken from the K.L. increases, we approach optimality.

C. Application to the autocorrelation functions; extension of Mercer's theorem

Here we are concerned with the application of Schmidt's theorems 2.2 and 3.1 to autocorrelation functions $R(s,t) = E\{X(s)\overline{X(t)}\}$ which are $\mathcal{L}^2$ and more particularly continuous and over a compact space. This, not only because of the role of the resulting extended Mercer's theorem, (in the foregoing results the proof of lemma 2.1 is based on appendix A-II, and there is no circularity in the reasoning), but also because of the intrinsic value of Mercer's theorem itself. Theorem 2.2, ap-
plied to $R(t,s)$ which is assumed Hermitian, yields immediately

$$R(t,s) = \sum_{n=1}^{\infty} \lambda_n \phi_n(t) \phi_n(s) \left( \xi^2 \right) \quad t, s \in \Omega_2$$

(3-17)

The convergence is shown to be uniform in $(s \times t)$ in $A - \mathcal{L}$ when $R(s,t)$ is continuous, a property which we will assume to be verified.

Mercer's theorem applies also when $\Omega_2 = \prod_i \left[ \alpha_i, \beta_i \right]$, i.e.

$$R(t_1, \ldots, t_D; \tau_1, \ldots, \tau_D) = \sum_n \lambda_n \phi_n(t_1, \ldots, t_D) \phi_n(\tau_1, \ldots, \tau_D)$$

It applies when a positive weighting factor $p(t)$ is such that

$$\int_\mathcal{T} \phi_i(t) \phi_j(t) p(t) \, dt = \delta_{i,j}$$

and where the $\phi_i(t)$'s are the solutions to

$$\lambda_i \phi_i(t) = \int_\mathcal{T} R(t, s) \phi_i(s) p(s) \, ds$$

It applies equally well to the vector case (cf. [12]), i.e. to the case where we have an autocorrelation function matrix $(R(s,t))$ (possibly infinite), i.e.

$$R^{i,j}(s,t) = \sum_n \lambda_n \phi^{i,n}(s) \phi^{j,n}(t) \quad s, t \in \mathcal{T}$$

where

$$\lambda_n \phi^{i,n}(t) = \sum_{j=1}^{\infty} \int_\mathcal{T} R^{i,j}(t,s) \phi^{j,n}(s) \, ds$$
Moreover, theorem 3.1 shows the optimal character of expansion (3-17) in the sense that

\[
\int_T |R(t,s) - \sum_{n=1}^N a_n(t) b_n(s)|^2 \, dt \geq \\
\int_T |R(t,s) - \sum_{n=1}^N \lambda_n \varphi_n(t) \phi_n(s)|^2 \, dt \\
= \sum_{n=N+1}^\infty \lambda_n^2
\]

(3-18)
SECTION IV
THE K.L. EXPANSION, GAUSSIAN CHARACTER AND INDEPENDENCE

A. Properties of the K.L. expansion in the Gaussian case.

When a random function is Gaussian, its K.L. expansion exhibits very remarkable properties. In particular:

Theorem 4.1. If \( X(t) \) is a Gaussian process, then the coefficients of its K.L. expansion are independent random variables. Moreover, for each \( t \), the series

\[
\sum \xi_i \phi_i(t)
\]

(4-1)

of the K.L. converges almost surely to \( \tilde{X}(t) \); in other words, it converges pointwise to \( \tilde{X}(t) = X(t,\omega) \) for almost all points \( \omega \) of the probability space \( \Omega \).

Proof. The \( \xi_i \) are independent because they are Gaussian and uncorrelated; and (4-1) converges almost surely to \( \tilde{X}(t) \) because Levy's theorem [7, p. 251] states that for independent random variables, almost sure convergence and convergence in the mean are equivalent.

This independence property of the \( \tilde{X}_i \)'s when \( X(t) \) is Gaussian makes the K.L. a very useful tool in detection and estimation, as it attaches to the sum of the first \( N \) terms in (4-1) a probability distribution that can be determined quite easily (cf. Chapter III). Let us remark, however, that the proof of Theorem 4.1 shows that the latter applies equally well to all canonical expansion as they are defined in
section V.

We now consider $X(t)$ and all associated finite-dimensional approximations

$$\sum_1^N \alpha_i \phi_i(t)$$

where the $\{\phi_i(t)\}$ constitute a complete orthonormal set in $L^2[0,T]$; and the $\alpha_i$ are the corresponding Fourier coefficients

$$\alpha_i = \int_T X(t) \phi_i(t) \, dt,$$

and we call $\alpha_i$ and $\alpha_i$ the random variables determined by

$$R_o = \int_T X^2(t) \, dt, \quad R_N = R_o - \sum_1^N \alpha_i^2 = \int_T [X(t) - \sum_1^N \alpha_i \phi_i(t)]^2 \, dt \quad (4-2)$$

The integral-square truncation-error is $R_N$, and we want to minimize the number of coefficients $\alpha_i$ required to keep $R_N$ less than or equal to some threshold level $\epsilon^2$, we then have:

**Theorem 4.2.** (Algazi and Sakrison) When $X(t)$ is Gaussian, the K.L. expansion is of all expansions orthogonal in the system of coordinate functions, the one that minimizes the average number $E(N)$ so that, for a given $\epsilon^2$, we have

$$R_{N-1} > \epsilon^2, \quad R_N \leq \epsilon^2 \quad (4-3)$$

The proof is given in A-6. Let us remark the potential usefulness of
Theorem 4.2 in techniques like Wald's sequential detection, but this problem remains open.

B. Independent Expansions, the K.L. and the Gaussian Character

We will now see that Theorem 4.1 has already found a kind of a loose converse, which has been arising in the researches about expansions into independent terms of random variables and processes. Let us first state a very important theorem.

Theorem 4.3. (Skitovich-Ramachandran) Consider the expansions

\[ X = \sum_{i=1}^{\infty} a_i \xi_i \]
\[ Y = \sum_{i=1}^{\infty} b_i \xi_i \]

and let the \( \{\xi_i\} \) be a sequence of independent random variables; let \( \{a_i\} \) and \( \{b_i\} \) be two sequences of real constants such that:

(i) the subsequences \( \{a_i/b_i : a_i b_i \neq 0\} \) and \( \{b_i/a_i : a_i b_i \neq 0\} \) are bounded

(ii) \( \sum_{i=1}^{\infty} a_i \xi_i \) and \( \sum_{i=1}^{\infty} b_i \xi_i \) converge with probability one to the random variables \( \tilde{X} \) and \( \tilde{Y} \) respectively; and

(iii) \( \tilde{X} \) and \( \tilde{Y} \) are independent.
Then, for every \( i \) such that \( a_i \neq b_i \neq 0 \), \( z_\omega \) is Gaussian.

For the proof, (cf. [16], [17], [63]). This theorem has been used extensively by P. Pierre [4] in connection with the study of representation of stochastic processes by expansions into independent terms. In particular, it allowed him to make a somewhat detailed investigation of the independence of the K.L. coefficients. For, let us note first that condition (ii) of Skitovich-Ramachandran's theorem is satisfied automatically if the coefficients \( \{ a_i \} \) on one side, \( \{ b_i \} \) on the other are independent; in that case, convergence in the mean is equivalent to convergence almost everywhere. We may consider \( X \) and \( Y \) as being the expressions of the random process \( X(t) \) at times \( t_1 \) and \( t_2 \) respectively with \( t_1, t_2 \in T \). We may also consider one interval \( T_1 \) imbedded into another \( T_2 \), and get an expression for the coefficients of the K.L. corresponding to \( T_1 \), in terms of a series of the coefficients of the K.L. corresponding to \( T_2 \), and apply the above theorem. These two techniques, which presuppose the knowledge of an analytic expression for the eigenvalues of the processes, have in particular yielded the two following results:

(a) "If \( X(t) \) is a zero-mean, independent-increment process, and if any two of its K.L. coefficients are independent, then \( X(t) \) is Gaussian" (i.e. Brownian motion).

(b) "Let \( X(t) \) have covariance function \( R(t,s) = \min(t,s) \); and let \( \{ a_i(T_1) \} \) and \( \{ a_i(T_2) \} \) be the set of K.L. coefficients for the intervals \([0,T_1]\) and \([0,T_2]\) respectively, with \( T_1 < T_2 \). Let \( nT_1 \neq mT_2 \), where \( n \) and \( m \) are odd integers. If the \( \{ a_i(T_2) \} \) are all independent, then \( a_i(T_1) \) and \( a_i(T_1) \) are independent if and only if \( X(t) \) is Gaus-
As an illustration of the techniques used to apply theorem 4.3, we give a proof of the last result (b) [4].

We can write

\[ X(t) = \sum \hat{a}_{x_j}^{(T_2)} \phi_j(t; \tau_2), \quad t \in T_2 \]  
\[ (4-5) \]

\[ \hat{X}(t) = \sum \hat{a}_{x_i}^{(T_1)} \phi_i(t; \tau_1), \quad t \in T_1 \]  
\[ (4-6) \]

Then

\[ \hat{a}_{x_i}^{(T_1)} = \int_{T_1} \hat{X}(t) \phi_i(t; \tau_1) \, dt \]
\[ = \int \sum \hat{a}_{x_j}^{(T_2)} \phi_j(t; \tau_2) \phi_i(t; \tau_1) \, dt \]
\[ = \sum \hat{I}_{j:i}^{(T_2, T_1)} \hat{a}_{x_j}^{(T_2)} \]  
\[ (4-7) \]

where

\[ \hat{I}_{j:i}^{(T_2, T_1)} = \int_{T_1} \phi_j(t; \tau_2) \phi_i(t; \tau_1) \, dt \]

and term by term integration of the series for \( \hat{X}(t) \) is justified, because of the uniform convergence of the series expansion of \( R(t,s) \) (Mercer's). Here, it is known that for \( R(t,s) = \min(t,s) \)
\[ \lambda_i^2 = \frac{4T^2}{\pi^2 (2i+1)^2}, \quad i = 0, 1, 2, \ldots \]

\[ \phi_n(t, \tau_i) = \sqrt{2/T_i} \sin \left( \frac{t}{\lambda_i} \right) \quad (4-8) \]

Therefore

\[ I_{j'i} (T_1, \tau_j) = \frac{2}{\sqrt{T_1 T_2}} \left[ \frac{\lambda_i^2 (\tau_j) \lambda_j^2 (\tau_i)}{\lambda_i^2 (\tau_i) + \lambda_j^2 (\tau_j)} \right] \sin \left[ \frac{T_1}{\lambda_i (\tau_j)} \right] \cos \left[ \frac{T_1}{\lambda_j (\tau_i)} \right] \quad (4-9) \]

If \( nT_i \neq mT_j \) where \( m \) and \( n \) are odd integers, then

\[ \cos \left[ \frac{T_1}{\lambda_i (\tau_j)} \right] \neq 0, \quad I_{j'i} (T_1, \tau_j) \neq 0 \]

\[ I_{j'i} (T_1, \tau_j) / I_{j'i} (T_1, \tau_j) \rightarrow \sin \left( \frac{T_1}{\lambda_i (\tau_j)} \right) / \sin \left( \frac{T_1}{\lambda_i (\tau_i)} \right) \neq 0 \]

and theorem 4.3 applies, which gives result (b).

Loosely speaking, Ramachandran's theorem implies that "sufficient regularity" of the coordinate functions of the K.L. and independence of the coefficients insures the Gaussian character of the process. For more completeness of presentation in this difficult area of the relationship of independence and Gaussianess, we quote other results [4], [18].

(c) "If \( X(t) \) has a rational spectral density, then its K.L. coefficients are independent for each of a sequence of intervals, \([-T_k, T_k]\) with \( T_k \rightarrow \infty \), if and only if \( X(t) \) is Gaussian."
(d) "If $X(t)$ is an $L_2$-martingale, and all its K.L. coefficients are independent, then $\hat{X}(t)$ is Gaussian." Let us recall that a martingale is a stochastic process $\hat{X}(t)$ such that:

$$E\left[|\hat{X}(t)|^2\right] < \infty \quad \text{and} \quad E\left\{\hat{X}(t_{n+1})/\hat{X}(t_i), \ldots, \hat{X}(t_{n})\right\} = \hat{X}(t_n)$$

with probability one; with $t_1 < t_2 < \ldots < t_{n+1}$.

Remark: Though the above general ideas and results do not seem to be valid only for the K.L., they, nevertheless, exhibit of it a very "tight structure." In particular, we cannot make any important assumption of independence of some K.L. coefficients without implying at the same time that the process (and the coefficients) are Gaussian. This is therefore a limitation on the K.L. As we have seen, the K.L. is mostly useful when the coefficients are independent.

SECTION V

THE K.L. AND CANONICAL REPRESENTATIONS

A deterministic linear operator, acting on a random process expanded into its K.L., does not yield another K.L., but a canonical expansion, as defined below. Moreover, we will see that the K.L. is just a very special type of canonical expansion. Therefore, studying canonical expansions allows us to better understand the structure of the K.L., and may suggest to us, for practical purposes, approximations to it. We shall also take a look at integral canonical
expansions, because of their intimate relation to the usual canonical expansions, and because this will give us some interesting new results. To our knowledge, the most thorough study of canonical expansions has been made by V. S. Pugachev [14].

A. Two Forms of Canonical Expansions; An integral form of the K.L.:

1. Let us consider a second-order random process $X(t)$, $A \in \mathbb{R}$. Then the Hilbert space induced by $X(t)$, i.e., $L_2 \{ X(t), t \in \mathbb{R} \}$ is separable. Using the Gram-Schmidt method of orthogonalization (and normalization), we can transform any series representation of $X(t)$ into one with orthogonal random coefficients

$$X(t) = \sum_{i} x_i \phi_i(t)$$  \hspace{1cm} (5-1)

where $E \left\{ x_i x_j \right\} = \lambda_i \delta_{ij}$.  \hspace{1cm} (5-2)

Suppose that we are given the system of covariances $\lambda_i$, and that we want to find the coordinate function system $\phi_i(t)$ which optimizes the series representation in (5-1) in the mean-square sense. We have to take the $\left\{ \frac{\phi_i(t)}{A_i} \right\}$ to be the Fourier coefficients of $X(t)$ with respect to the orthogonal system $\left\{ \tilde{x}_i \right\}$, i.e.,

$$\phi_i(t) = \frac{1}{\lambda_i} E \left\{ X(t) \tilde{x}_i \right\}$$  \hspace{1cm} (5-3)

Expansion (5-1), with conditions (5-2) and (5-3) is said to be canonical. We have so far assumed the existence of one expansion at least.
This is always possible and in an infinite number of ways as we shall see below, when we get a method of construction for canonical expansions. Let us remark that canonical expansions have orthogonal coefficients like the K.L. expansion. Their coordinate functions, however, are not necessarily orthogonal. In fact, the only canonical expansion whose coordinate functions \( \{ \phi_n(t) \} \) are orthogonal is the K.L. Canonical expansions, therefore, maintain the main attributes of the K.L.; namely the separation of time dependence from the random space one; the orthogonality of the coefficients; the Fourier-type character of the expansion. They allow time operations to be performed on processes easily and when orthogonality leads to the independence of the coefficients, they are widely used in models for detection, estimation and filtering.

Let us consider now \( R(t,s) = \mathbb{E} \left\{ X(t) X(s) \right\} \).

Then (5-1), (5-2) yield:

\[
R(t,s) = \sum N \phi_n(t) \phi_n(s)
\]

Therefore (5-4) is a consequence of (5-1). It is also proven that (5-1) (cf. [14] Chapter 9). In other words, for second order random processes, there is one-to-one correspondence between the canonical representation of \( X(t) \) and the associated representation of its covariance function.

To get results in their generality, let us recall that the Hilbert space \( L_2 \left( X(t), t \in T \right) \) spanned by a second-order random process \( X(t) \) consists of all random variables which may be obtained by means of linear operations on the random variables \( X(t) \). In other words, \( L_2 \left( X(t), t \in T \right) \) constitutes the set of all possible linear function-
als over the random variables $X(t)$, $t \in T$. These functionals comprise, not only $L_2$-functions, but also the Dirac delta function and its derivatives, etc. It is shown (cf. [14] p 262) that for the canonical expansion (5-1) to converge in the mean, it is necessary and sufficient that the set of linear functionals $\mathcal{F}_n$ such that $\mathcal{F}_n(X(t)) = \tilde{x}_n$ be complete with respect to the random function $X(t)$.

A system of functions \( \{a_n(t)\} \) is complete with respect to $X(t)$ if there is no function $\psi(t)$ such that, for every $n$,

$$
\int_{T} \int_{T} \psi(t) a_n(s) R(t,s) \, dt \, ds = 0
$$

while

$$
\int_{T} \int_{T} \psi(t) \psi(s) R(t,s) \, dt \, ds > 0
$$

(strictly) (5-6)

The definition is similar for functionals in general. This is why the most general expression of the coefficients $\phi_n$, in (5-1) is given by

$$
\tilde{x}_n = \mathcal{F}_n(X(t))
$$

and if we restrict ourselves to a complete system of $L^2$-functions $\{a_n(t)\}$, then,

$$
\tilde{x}_n = \int_{T} a_n(t) X(t) \, dt
$$

Then, formulas (5-2) and (5-3) yield:

$$
\phi_n(t) = \frac{1}{\lambda_n} \int_{T} a_n(s) R(t,s) \, ds
$$

and

$$
\int_{T} a_n(t) \phi_m(t) \, dt = \delta_{nm}
$$
Since it is well known that if a certain system \( \{ f_n(t) \} \) of \( L^k \) functions is complete relative to all functions of another system \( \{ a_n(t) \} \) which is known to be complete in \( L^k \), then the system \( \{ f_n(t) \} \) which is known to be complete in \( L^k \), then the system \( \{ f_n(t) \} \) is also complete in \( L^k \) (cf. [39], p. 93). We can then start from an arbitrary complete system \( \{ f_n(t) \} \) and by the Gram-Schmidt process build an \( \{ a_n(t) \} \) from it. The conditions (5-9) and (5-10) are necessary and sufficient conditions for the existence of a canonical expansion (5-1). We easily see that if we take, \( a_n(t) = \phi_n(t) \) where

\[
\lambda_n \phi_n(t) = \int_T K(t, s) \phi_n(s) \, ds
\]

we get the K.L. expansion.

2. Canonical expansions are but a particular case of integral canonical expansions defined as follows. If the zero-mean, second-order random process \( X(t) \) can be represented by:

\[
X(t) = \int_A \mathcal{V}(\lambda) \lambda_n(t, \lambda) \, d\lambda ; \quad t \in T, \quad \lambda \in \Lambda
\]

(5-11)

where:

\[
\mathbb{E} \left\{ \mathcal{V}(\lambda) \mathcal{V}(\lambda') \right\} = G(\lambda) \delta(\lambda - \lambda')
\]

(5-12)

then we say that we have an integral canonical representation of \( X(t) \) expressed in terms of uncorrelated random process \( \mathcal{V}(\lambda) \). More specifically, \( X(t) \) is represented in terms of generalized white noise, (5-12). The process \( X(t) \) is seen to be the output of a linear, time-varying filter \( h(t, \lambda) \) whose input is a generalized white random process, \( \mathcal{V}(t) \).

The possibility of such an expression (5-11) with condition (5-12), in the general case, is investigated by Pugachev [14] (cf. Chapter 13, par. 97) and H. Korezlioglu [19]. This is also a general case of
the method of shaping filters (cf. Laning & Battin [58]). It is
clear that, if in (5-11), $A = \{l, r, \ldots, m, \ldots\}$ we get the canonical ex-
ansion $X(t) = \sum \lambda_i \mathcal{L}(t, \lambda) = \sum \mathcal{L}(t, \lambda) \{h_i(t)\}$, with

$$E \{\lambda_i \mathcal{L}(t, \lambda)\} = \lambda_i \delta_{i,j}.$$  

By analogy with (5-4), there is a result due to K. Karhunen.

**Theorem 5-1**: The necessary and sufficient condition for the re-
presentation of a random function $X(t)$ by (5-11) is that its covariance

can be represented by:

$$R(t, s) = \int_D \mathcal{G}(\lambda) \mathcal{L}(t, \lambda) \mathcal{L}(t', \lambda) \ d\lambda$$  

(5-13)


These conditions of possibility of (5-13) are analogous to some extent
to the discrete case of canonical expansions where

$$\mathcal{L}(t, \lambda) = \frac{1}{\mathcal{G}(\lambda)} \int_T a(s, \lambda) R(t, s) ds$$  

(5-14)

and

$$\int_T \mathcal{L}(t, \lambda) \mathcal{L}(t, \lambda') dt = \delta(\lambda - \lambda')$$  

(5-15)

and

$$\int_D \mathcal{H}(t, \lambda) \mathcal{H}(t', \lambda) d\lambda = \delta(t-t')$$  

(5-16)

i.e. the conditions are analogous to (5-9) and 5-10) respectively.

3. Let us now consider an integral form of the K.L. expansion. For
it to exist, it is necessary only that in (5-14) and 5-15)

$$\mathcal{L}(t, \lambda) = a(t, \lambda)$$

Then $h(t, \lambda)$ would be defined by the eigenfunction equation

$$\mathcal{G}(\lambda) \mathcal{L}(t, \lambda) = \int_T R(t, \omega) \mathcal{L}(t, \lambda) d\omega$$
where however the eigenfunctions are not denumerable. We are not studying the problem in general here. We will give just a very simple illustration of the possibility of such an integral form of the K.L. Let us therefore consider the zero-mean process $X(t)$ with autocorrelation function

$$R(t, s) = e^{-|t-s|}; t, s \in (-\infty, +\infty)$$

Then we have the following associated (singular) integral equation

$$\alpha \Phi(t) = \int_{-\infty}^{+\infty} e^{-|t-s|} \Phi(s) \, ds$$

It is known that [39]

$$\Phi(t) = e^{\varepsilon \lambda t}, \quad \lambda = \frac{\varepsilon}{4 \pi^2}$$

i.e.

$$G(\lambda) = \frac{\varepsilon}{4 \pi^2}, \quad \xi(t, \lambda) = e^{\varepsilon \lambda t}$$

and it is clear that

$$\begin{align*}
\int_{-\infty}^{+\infty} e^{-\varepsilon \lambda t} e^{\varepsilon \lambda t'} \, dt &= \delta(\lambda - \lambda') \\
\int_{-\infty}^{+\infty} e^{-\varepsilon \lambda t} e^{\varepsilon \lambda t'} \, d\lambda &= \delta(t - t') \quad \Delta = T = (+\infty, +\infty)
\end{align*}$$

Therefore

$$X(t) = \int_{-\infty}^{+\infty} \xi(t, \lambda) \chi(\lambda) \, d\lambda = \int_{-\infty}^{+\infty} e^{\varepsilon \lambda t} \chi(\lambda) \, d\lambda$$

with

$$E\left\{ \chi(\lambda) \, \chi(\lambda') \right\} = \frac{2}{4 \pi^2 \varepsilon^2} \delta(\lambda - \lambda')$$

which is but the spectral representation of the stationary random process $X(t)$ defined by

$$E\{ X(t) \} = 0, \quad E\{ X(t) \, X(s) \} = R(t, s) = e^{-|t-s|}.$$
The remark is general, and any spectral representation of a second order stationary process can be considered as the integral form of the K.L. expansion.

B. Relationship between the two forms of canonical expansions.

In a recent paper [20], a class of series expansions of second order wide-sense stationary processes was derived, which was based on the spectral representation of such processes. We are going to see how these results extend to the non-stationary case.

Let us use an integral canonical representation of \( \mathcal{R}(t, s) \); and we do not discuss here the general problem of finding such a representation. We know from the works quoted above that for large classes of random functions, we can construct an integral representation. Then

\[
\mathcal{R}(t, s) = \int_{A} G(\lambda) \, \mathcal{H}(t, \lambda) \, \mathcal{K}(s, \lambda) \, d\lambda
\]

where \( A \) is a certain domain, left unprecised for the time being. We want to get, from the integral representation, a canonical expansion such that:

\[
\begin{align*}
\hat{x}(t) &= \sum_{n} \hat{v}_{n} \int_{t} \hat{x}_{n}(l) \, dl, \\
\hat{v}_{n}(l) &= \frac{1}{A_{n}} \int_{t} a_{n}(s) \, \mathcal{R}(l, s) \, ds
\end{align*}
\]

\[(5-17)\]

and

\[
\int_{t} \int_{s} \hat{a}_{n}(t) \, a_{m}(s) \, \mathcal{R}(t, s) \, dt \, ds = D_{n} \delta_{n,m}
\]

\[(5-18)\]

It is sufficient to find a complete set \( \{a_{n}(t)\} \) which satisfies the last two conditions (5-17) and (5-18) for the expansion to be valid in the mean.

Equation (5-16) yields:

\[
\int_{t} \int_{s} \hat{a}_{n}(t) \, a_{m}(s) \, dt \, ds \left[ \int_{A} d\lambda \, G(\lambda) \, \mathcal{H}(t, \lambda) \, \mathcal{K}(s, \lambda) \right] = D_{n} \hat{a}_{n,m}
\]
Interchanging the order of integrations, and this is legitimate by Fubini's theorem, we obtain:
\[
\int_\Delta d\lambda \, G_\lambda (\lambda) \left[ \int_T dt \, \overline{a}_n (t) \, \overline{\ell} (t, \lambda) \int_T du \, a_n (s) \, \ell (s, \lambda) \right] = \mathcal{D}_n \, \mathcal{D}_{nm}
\]

Letting
\[
\overline{A}_n (\lambda) = \int_T dt \, \overline{a}_n (t) \, \ell (t, \lambda)
\]

the orthogonality conditions (5-16) can now be written as:
\[
\int_\Delta d\lambda \, G_\lambda (\lambda) \, \overline{A}_n (\lambda) \, \overline{A}_m (\lambda) = \mathcal{D}_n \, \mathcal{D}_{nm}
\]

Let us suppose that the domain \( \Delta \) coincides with \( \mathcal{T} \), i.e. \( \mathcal{T} = \Delta \); and let the system \( \{ \overline{A}_n (\lambda) \} \) be complete and orthonormal in \( L_2 (G(\lambda) \, d\lambda, \mathcal{T}) \)

Let us also assume \( \mathcal{D}_1 = \mathcal{D}_2 = \cdots = \mathcal{D}_n = \cdots = \mathcal{I} \). (This is an assumption which we can make without any loss of generality.) Then:
\[
\int_\Delta d\lambda \, G_\lambda (\lambda) \, \overline{A}_n (\lambda) \, \overline{A}_m (\lambda) = \mathcal{D}_n \, \mathcal{D}_{nm}
\]

After interchange of integrations, we obtain:
\[
\int_\Delta d\lambda \, G_\lambda (\lambda) \, \overline{A}_n (\lambda) \, \overline{A}_m (\lambda) = \mathcal{D}_n \, \mathcal{D}_{nm}
\]

or
\[
\int_\Delta d\lambda \, G_\lambda (\lambda) \, \overline{A}_n (\lambda) \, \overline{A}_m (\lambda) \, d\lambda
\]

(5-19)

Also
\[
\overline{\chi}_n = \int_T \overline{a}_n (t) \, \chi (t) \, dt = \int_T \overline{a}_n (t) \left[ \int_T \chi (\lambda) \, \overline{\ell} (t, \lambda) \, d\lambda \right] \, dt
\]

Interchanging integrations, we get
\[
\overline{\chi}_n = \int_T \overline{\ell} (t, \lambda) \, \overline{a}_n (t) \, \chi (\lambda) \, d\lambda \]

i.e.
\[
\overline{\chi}_n = \int_T \overline{\ell} (t, \lambda) \, \overline{a}_n (t) \, \chi (\lambda) \, d\lambda
\]

(5-20)
Therefore, from the knowledge of an integral canonical representation of \( X(t) \), we are able to deduce a whole class of expansions (5-1) in the same interval. What is remarkable is that such a representation

\[
X(t) = \sum_{n} \eta_n \int_{-\infty}^{\infty} f_n(t) \, dt , \quad \text{with} \quad \eta_n \quad \text{and} \quad f_n(t) \quad \text{given by (5-20) and (5-19)}
\]

and (5-19) does not assume stationarity of the process, nor is its observation based on a very particular isomorphism, like the one that exists between \( X(t, \omega) \) and \( \mathcal{E}^{t \lambda} \) in the stationary case). All that is needed, besides the existence of an integral canonical representation with \( T \subset \Delta \) is a set \( \left\{ A_n \right\} \) which is complete (and orthonormal) relative to the functions \( X(t) \). It follows immediately from this that there is a corresponding canonical expansion for the auto-correlation function of the process:

\[
R(t,s) = \mathbb{E} \left\{ X(t) X(s) \right\} = \sum_{n} f_n(t) f_n(s)
\]

Application: Formulas (5-19) and (5-20) applied to a wide-sense stationary process yield the results of the first part of the article of Masry, Liu & Steiglitz [20]. We have:

\[
R(t,s) = \int_{-\infty}^{\infty} e^{i t \lambda} A(t, \lambda) \, dS(\lambda)
\]

or equivalently

\[
X(t, \omega) = \int_{-\infty}^{\infty} e^{i t \lambda} d \xi(\lambda, \omega) , \quad \mathbb{E} \left( \left| d \xi(\lambda, \omega) \right|^2 \right) = dS(\lambda)
\]

This shows that

\[
\mathcal{G}(\lambda) \, d\lambda = \mathcal{E}^{t \lambda} \xi(\lambda, \omega) \quad \text{and} \quad T = (-\infty, +\infty), \quad \text{since} \quad \Delta = (-\infty, +\infty)
\]
Therefore:
\[
\begin{align*}
\mathbf{f}_n(t) &= \int_{-\infty}^{t} d\mathcal{S}(\lambda) e^{i\lambda t} \ A_n(\lambda) \\
\mathbf{G}_n(t) &= \int_{-\infty}^{t} d\mathcal{G}(\lambda) \ A_n(\lambda)
\end{align*}
\]

By varying the $A_n(\lambda)$, we get - as the quoted article amply shows - a great variety of the most useful canonical representations.

Let us remark that the use of the integral representation to obtain a discrete representation in the case of deterministic functions is a classical well trodden subject (cf. e.g. Titchmarsh [21]).

C. Operations on the K.L. expansion

1. Let us recall that if $K(s,t)$ is an $L^2$ kernel, i.e.

\[
\int_\mathcal{D} \int_{\mathcal{T}} \left| K(s,t) \right|^2 \, ds \, dt < \infty
\]

and if $x(t)$ is also $L^2$, then the application of Schwarz inequality to $y = Kx$, i.e.

\[
\langle y(t) \rangle = \int_{\mathcal{T}} K(t,s) x(s) \, ds
\]

shows that $y(t)$ is also $L^2$.

If we assume that $K(s,t)$, which needs not by symmetric, to be continuous, then $y(t)$ is also continuous. Moreover, Hilbert-Schmidt's theorem (cf. e.g. [8]) shows that $y(t)$ is then developable in a uniformly convergent series of functions $\Psi_i(t)$, where the $\Psi_i(t)$'s are the eigenfunctions of

\[
\int_{\mathcal{T}} K(t,u) \overline{K(s,u)} \, du = K K^*(t,s).
\]

This implies that in linear transformations of the K.L. of the process $X(t)$ there is in general no problem of convergence with respect to $t$, i.e. $\gamma(t)$ generally converges in $t$. 
Of all linear operations, integration and differentiation are the most important ones. As to integration

$$\gamma(t) = \int_0^t X(u) \, du = \int_0^T U(t-u) \, X(u) \, du$$

where $U(t)$ is the step-function

and $\gamma(t)$ is mean-square continuous if $X(t)$ is mean-square continuous.

We know that we can integrate term by term and get

$$\gamma(t) = \sum \phi_i(t') \xi_i \, du = \sum \phi_i(t') \xi_i$$

(5-19)

where the canonical expansion in (5-19) converges uniformly in $t$. But when the linear operation is differentiation, i.e.

$$X'(t) = - \int_0^T \delta'(t-u) \, X(u) \, du$$

then

$$\delta'(t-u) = K(t, u)$$

is not $L^2$.

And the above results are not applicable.

To obtain some useful results, we overturn the difficulty by means of a sufficient condition that is proved in ([7], p. 520):

**Lemma 5.1.** If the derivative \( \frac{\partial^{n+1}}{\partial t^{n+1} \partial s^n} R(t,s) \) exists and is finite, then $X(t)$ is $n$ times sample differentiable.

This is the basis of the theorem of T. Kadota [5] which follows:

**Theorem 5.1.** If $X(t)$ is a separable and measurable stochastic process with zero-mean and covariance $R(t,s)$, if \( \frac{\partial^{n} R(t,s)}{\partial t^{n} \partial s^n} \) exists
and is continuous, and if $X(t)$ (the $n^{th}$ derivative) exists almost surely, then

$$\mathcal{X}^{(n)}(t) = \sum_{i} \mathcal{Z}_i \Phi_i^{(n)}(t)$$

(5-20)

where

$$\mathcal{Z}_i = \int_{0}^{T} X(t) \Phi_i(t) \, dt$$

and the series converges in the stochastic mean, uniformly in $t$. The proof is given in Appendix A-8.

2. Linear transformations act only on the deterministic parts of the K.L., i.e. on the $\Phi_i(t)'s$, and not on the $\mathcal{Z}_i's$. In this lies part of its importance. Moreover, in non-linear transformations, the K.L. provides an efficient method of linearization, since the first elements of the expansion carry the bulk of information associated with the process. In particular, when $X(t)$ is Gaussian, all the statistics of the different powers of the K.L. coefficients $\mathcal{Z}_i$ are determined very easily (e.g. cf. VanTrees [22]).

3. The problem of invariance of the K.L. expansion in a transformation.

What we are interested in is the question of whether the eigenfunctions of $X(t)$ for instance, are conserved in the transformation

$$\mathcal{X}'(t) = \int_{0}^{T} \Phi'(t,s) \mathcal{X}(s) \, ds$$

because of the impracticality of calculating the eigenfunctions $\Phi_i(t)$ after each transformation. As we shall see, the K.L. does not manifest any invariance character in simple transformations, in particular not in the linear ones, which are the easiest to handle and which represent many of the tractable
models of physical systems we usually deal with. From this there results largely the "impracticality" of the K.L. But to get some feeling for the difficulties, we take a look at the figure below:

![Diagram](image)

Fig. 5-1

If we want a simultaneous representation like

\[ x(t) = \sum \alpha_i \phi_i(t) \]
\[ y(t) = \sum \beta_i \phi_i(t) \]

where

\[ E \{ \alpha_i \alpha_j \} = \delta_{ij} = E \{ \beta_i \beta_j \} \]

we would be solving the problem of simultaneous diagonalization (cf. [23], [24]), which does not always admit a solution. And in our case, we want the \( \{ \phi_i(t) \} \) to be also orthogonal!

Let us consider however a simple case where a simultaneous K.L. expansion for \( x(t) \) and \( y(t) = \int_\gamma K(t, \gamma) x(\gamma) d\gamma \) is possible.

For this, it is necessary and sufficient that \( R_x(t, s) \) and \( R_y(t, s) \) have the same eigenfunctions. For instance:

\[ R_x(t, s) \text{ and } \left\{ R_x(t, s) + \frac{N_x}{\hbar} \delta(t-s) \right\} \]

have the same eigenfunctions, though \( \delta(t-s) \) is not \( \mathcal{L}^2 \).
Let us consider now the optimum filter equation (cf. e.g. [22]), which can be written, in the case of white additive noise, as:

\[
\int_0^T \mathcal{R}_x(t, u) \left[ R_x(u, s) + \frac{N_0}{\Delta} \delta(u-s) \right] du = R_x(t, s)
\]

(5-22)

Our purpose is to find an explicit expression for the optimum filter impulse response \( h_o(t, s) \) in terms of the eigenfunctions of the covariance kernel \( R_x(t, s) \).

We are assured of the existence of the inverse filter of \( [ R_x(t, s) + \frac{N_0}{\Delta} \delta(t-s) ] \) (cf. [59]). Let us denote it by \( [ R_x(t, s) + \frac{N_0}{\Delta} \delta(t-s) ]^{-1} \). Let us denote it by \( [ R_x(t, s) + \frac{N_0}{\Delta} \delta(t-s) ]^{-1} \).

Then (5-22) can be rewritten as

\[
\mathcal{R}_o(t, s) = \int_0^T R_x(t, u) \left[ R_x(u, s) + \frac{N_0}{\Delta} \delta(u-s) \right]^{-1} du; \quad t, s \in [0, T].
\]

We can express the inverse filter impulse response in a series of simpler filters, in a way very similar to a geometric series, if we see the iterated kernels of \( R_x(t, s) \). These iterated kernels are defined as:

\[
\begin{align*}
R_x^{(0)}(t, s) &= R_x(t, s) \\
R_x^{(1)}(t, s) &= \int_0^T R_x(t, u) R_x(u, s) du \\
R_x^{(2)}(t, s) &= \int_0^T R_x(t, u) R_x^{(1)}(u, s) du \\
&\vdots \\
R_x^{(n)}(t, s) &= \int_0^T R_x(t, u) R_x^{(n-1)}(u, s) du
\end{align*}
\]

and

\[
R_x^{(\infty)}(t, s) = \delta(t-s)
\]

Then in (5-9) it is shown that:

\[
\left[ R_x(t, s) + \frac{N_0}{\Delta} \delta(t-s) \right]^{-1} = \frac{2}{N_0} \cdot \frac{1}{\delta(t-s) + \frac{2}{N_0} R_x(t, s)}
\]

\[
= \frac{2}{N_0} \left\{ \delta(t-s) + (-1)^\frac{2}{N_0} R_x(t, s) + (-1)^2 \frac{2}{N_0} R_x^{(2)}(t, s) + \cdots \right\}
\]
Therefore

\[ \mathcal{L}_0(t,s) = \frac{\varphi}{N_0} \int_0^\infty R_X(t,u) \left[ \delta(u-s) + \sum_{n=1}^{\infty} (-1)^n \left( \frac{u}{N_0} \right)^n R_X^{(n)}(u,s) \right] du \]

\[ = \sum_{n=1}^{\infty} (-1)^n \left( \frac{u}{N_0} \right)^n R_X^{(n)}(t,s) \]

But we know that whenever \( R_X(t,s) \) is continuous, Mercer's theorem applies for all \( n, n = 1, 2, \ldots \). Therefore \( \mathcal{L}_0(t,s) \) can be further written as

\[ \mathcal{L}_0(t,s) = \sum_{n=1}^{\infty} (-1)^{n+1} \left( \frac{u}{N_0} \right)^n \sum_{i=1}^{\infty} \lambda_i^n \phi_i(t) \phi_i(s) \]

\[ = \sum_{n=1}^{\infty} \left( \sum_{i=1}^{\infty} (-1)^{n+1} \left( \frac{u}{N_0} \right)^n \lambda_i^n \right) \phi_i(t) \phi_i(s) \]

But

\[ \sum_{n=1}^{\infty} (-1)^{n+1} \left( \frac{u}{N_0} \right)^n \lambda_i^n = \frac{\varphi}{N_0} \lambda_i \cdot \frac{1}{1 + \frac{u}{N_0} \lambda_i} = \frac{\lambda_i}{\lambda_i + N_0/2} \]

i.e.

\[ \mathcal{L}_0(t,s) = \sum_{i=1}^{\infty} \left( \frac{\lambda_i}{\lambda_i + N_0/2} \right) \phi_i(t) \phi_i(s) \]

(5-23)

where the series on the right of (5-23) is convergent, since \( N_0 > 0 \). Expression (5-23) is therefore one case of invariance of the eigenfunctions in a linear transformation.

Another case is obtained as follows. We want that, if

\[ X(t) \rightarrow \mathcal{Y}(t) = \int_T \mathcal{L}(t,s) \mathcal{X}(s) ds \]

then also

\[ \phi_i(t) \rightarrow \mathcal{Y}_i(t) = \int_T \mathcal{L}_i(t,s) \phi_i(s) ds \]
where the $\Phi_i(t)$, being eigenfunctions of $X(t)$, are orthonormal, and the $\Psi_j(t)$ are required to be orthogonal, i.e.

$$\int_0^T \int_0^T h(t,\lambda) \Phi_i(\lambda) d\lambda \left[ \int_0^T h(t,\mu) \Phi_j(\mu) d\mu \right] = 0, \quad i \neq j$$

(5-24)

Assuming the interchange of integrations possible, and letting

$$\mathcal{R}(\lambda, \mu) = \int_0^T dt \cdot h(t, \lambda) h(t, \mu)$$

(5-25)

then (5-24) becomes, after noticing that $\mathcal{R}(\lambda, \mu) = \mathcal{R}(\mu, \lambda)$,

$$\int_0^T \int_0^T \Phi_i(\lambda) \Phi_j(\mu) \mathcal{R}(\lambda, \mu) = 0$$

for all $i \neq j$.

Being symmetric, $\mathcal{R}(\lambda, \mu)$ does have eigenfunctions. If we make the further assumption that it is positive-definite, then its system of eigenfunctions is complete, and writing (5-24) as

$$\int_0^T \int_0^T \mathcal{R}(\lambda, \mu) \Phi_j(\mu) \Phi_i(\lambda) = 0$$

we see that it is necessary that

$$\mathcal{D}_j \Phi_i(\mu) = \int_0^T \mathcal{R}(\mu, \lambda) \Phi_j(\lambda) d\lambda$$

(5-26)

We are therefore not solving (5-24) in its generality; and yet we see that (5-26), if we are given the filter $h(t,2)$, determines the (narrow) class of admissible inputs $X(t)$, i.e. those for which,
\[ \chi(t) = \sum_{i} x_i \Phi_i(t) \]

where the \( \{ \Phi_i(t) \} \) are given by (5-26), and \( E \{ x_i x_j \} = \Lambda_{ij} \delta_{ij} \).

In other words, for the above defined invariance to hold, it is necessary and sufficient that

\[
E \{ \chi(t) \chi(s) \} = R_{\chi}(t,s) = R(t,s)
\]

\[
= \int_{T} \hat{h}(\lambda,t) \hat{h}(\lambda,s) d\lambda
\]

(5-27).
It is well known that all (separable) Hilbert spaces are isomorphic. This is at heart the reason why the formulation of problems in statistical inference for instance over a second-order process \( X(t) \) can be made in different ways, through the K.L. expansion (as will be seen in Chapter III) as well as the reproducing kernel Hilbert space concepts which we review briefly. (For a thorough account, cf. [3]). The situation is very much similar to the equivalence of matrix and wave formulations in quantum mechanics.

Let us define first what is meant by a representation of \( X(t) \).

**Definition 6.1.** A Hilbert space \( \mathcal{H} \) is said to be a representation of a random process \( \tilde{X}(t) \) if \( \mathcal{H} \) is congruent to the Hilbert space \( L_2(\mathbb{R}) \), induced by \( \tilde{X}(t) \), i.e., if \( \mathcal{H} \) and \( L_2(\mathbb{R}) \) are isomorphic and the inner product is conserved.

It is shown ([3], p. 277) that a family of function vectors \( \{f(t), t \in T\} \) in a Hilbert space \( \mathcal{H} \) is a representation of a random process \( \{X(t), t \in T\} \) if, for every \( s \) and \( t \) in \( T \), we have:

\[
\left( f(s), f(t) \right)_{\mathcal{H}} = R(s,t) = E \left\{ X(t)X(s) \right\}
\]

(6-1)

Let us see how the K.L. is a representation of \( X(t) \), in the sense defined above. From Mercer's theorem

\[
R(s,t) = \sum_{\lambda} \lambda \phi(s) \phi(t)
\]

(6-2)
with our usual notations. Let $H$ be the space of all (here we are limiting ourselves to real) sequences $\{a_n, n = 1, 2, \ldots\}$ such that

$$\sum \lambda_n a_n^2 < \infty \quad (6-3)$$

with inner product for 2 elements $\{a\}$ and $\{b\}$ of $H$ defined by

$$(a, b) = \sum \lambda_n a_n b_n \quad (6-4)$$

Then, it is clear that the sequence

$$\phi(t) = \{\phi_n(t), n = 1, 2, \ldots\} \quad (6-5)$$

belongs to $H$. Moreover

$$(\phi(t), \phi(s)) = \sum \lambda_n \phi_n(t) \phi_n(s) = K(t, s) \quad (6-6)$$

which shows that $L_2 (X(t), t \in T)$ is congruent to $L_2 (\mathbb{H}, t \in \mathbb{T})$ and that the K.L. is a representation of $X(t)$.

We now define a reproducing kernel Hilbert space (R.K.H.S.)

**Definition 6.2.** A Hilbert space $H$ is said to be an R.K.H.S. with reproducing kernel $K$, if the members of $H$ are functions on some set $T$, and if there is a kernel $K$ on $T \times T$ having the following properties:

$$K(\cdot, t) \in H, \quad \text{for every } t \in T \quad (6-7)$$

$$(g, K(\cdot, t)) = g(t), \quad \text{for every } g \in H \quad (6-8)$$
where \((g, K(\cdot, t))\) represents the inner product, for \(t\) fixed, of the elements \(g\) and \(K\) of \(H\).

This definition is justified in part by the following existence and uniqueness theorem:\[3\]:

**Theorem 6.1.** A symmetric non-negative kernel \(K\) generates a unique Hilbert space, denoted by \(H(K)\), of which \(K\) is the reproducing kernel.

Two basic properties from our point of view are given by the following two theorems (cf. \[3\] for their proof).

**Theorem 6.2.** If \(K\) is a reproducing kernel for the Hilbert space \(H\), then the family of functions \(K(\cdot, t)\) span \(H\).

**Theorem 6.3.** (Loeve)[62]. Any random function of second-order possesses a representation by a reproducing kernel Hilbert space.

Theorem 6.3, is the key to the equivalence of the K.L. and R.K.H.S. formulations.

We will now give an example of the application of the R.K.H.S. theory to detection theory, that will make more explicit that equivalence (cf. \[25\]).

Let us consider the problem of detecting a sure signal \(m(t)\) in colored noise. We have the two hypotheses:

\[
X(t) = m(t) + \tilde{N}(t) : H^1
\]

\[
\tilde{X}(t) = \tilde{N}(t) : H^0
\]

(6-9)
where \( m(t) \) is the sure signal to be detected, and \( N(t) \) is the noise with autocorrelation \( R(t,s) \). Let us, after Kailath [25], project \( \tilde{X}(t) \), \( m(t) \), and \( N(t) \) on the signal itself; and let us assume that we have normalized \( m(t) \), i.e.,

\[
\int_0^T m^2(t) \, dt = 1
\]

Then, the projections are \( \tilde{X}_1 \), \( \tilde{X}_2 \), and \( \tilde{N} \), respectively:

\[
\tilde{X}(t) = \tilde{X}_1 \cdot m(t) + \tilde{X}_{\text{rem}}(t)
\]

\[
\tilde{N}(t) = \tilde{N}_1 \cdot m(t) + \tilde{N}_{\text{rem}}(t)
\]

with the constraints that

\[
E \{ \tilde{N}_{\text{rem}}(t) \tilde{N}_1(t) \} = 0
\]

(i.e. we tune out the component of \( \tilde{N}(t) \) that is outside the space of \( \tilde{m}(t) \) and that \( \tilde{N}_1 \) be a linear functional of \( \tilde{N}(t) \)). Therefore, (6-9) becomes equivalent to

\[
\tilde{X}_1 \equiv 1 + \tilde{N}_1 : H_1
\]

\[
\tilde{X}_2 = \tilde{N}_1 : H_2
\]

(6-12)

since \( \tilde{X}_{\text{rem}}(t) = \tilde{N}_{\text{rem}}(t) \) under both hypotheses. This means that for detection purposes, the component \( \tilde{N}_{\text{rem}}(t) \) is irrelevant.

The problem therefore is that of solving (6-11) with respect to \( \tilde{N} \) or equivalently

\[
E \{ \tilde{N}_{\text{rem}}(t) \tilde{N}_1(t) \} = 0
\]
Assuming there is a finite covariance solution to (6-13) and letting
\[ \mathcal{U} = \frac{\mathcal{N}}{\sigma^2} \text{, where } \sigma^2 = \mathbb{E}\left\{ \mathcal{N}^2 \right\}, \]
we are thus let to solve for \( \mathcal{U} \) in the equation:
\[ \mathbb{E}\left\{ \mathcal{N}(t) \mathcal{U} \right\} = m(t) \] (6-14)

\( \mathcal{U} \) is therefore the sufficient statistic which is found by means of the K.L. expansion (cf. Chapter III, or any book on detection theory) to be
\[ \mathcal{U} = \sum \alpha_i \mathcal{N}(t_i) \] (6-15)

But this sufficient statistic is very remarkably related to the reproducing kernel Hilbert space theory by the following [25]:

**Theorem 6.4.** A finite variance solution to \( \mathbb{E}\left\{ \mathcal{N}(t) \mathcal{U} \right\} = m(t) \) exists if and only if, \( m(t) \) belongs to the R.K.H.S. of \( \mathcal{N}(t) \).

**Proof.** Let us define on the space of \( m(t) \) an inner product as follows:
\[ \left( m(t), m'(t) \right) \triangleq \left\{ \mathbb{E}\left( \mathcal{N}(t) m(t) \right), \mathbb{E}\left( \mathcal{N}(t) m'(t) \right) \right\} = \mathbb{E}\left( \mathcal{U}, \mathcal{U}' \right) \] (6-16)

Then the space of \( m(t) \) so defined forms a Hilbert space, because the random variables \( \mathcal{U} \) (which are linear functionals of \( \mathcal{N}(t), t \in T \)) form a Hilbert space [3].

Let us now verify that this is the reproducing kernel Hilbert space \( H(R, T) \):
\[ R(t, s_0) = E \{ N(t), N(s_0) \} \]

is one of the \( m(t) \) since \( N(s_0) \) has finite variance.

\[ (R(t, s_0), m(t)) = \{E(N(t)N(s_0)), E(N(t)U)\} = E(N(s_0)U) = m(s_0) \]

therefore the signal space \( m(t) \), with inner product (6-16) is truly the R,K,H,S. of \( R(t, s) \).

This concludes the illustration of the relationship of functionals over a K.L. expansion (like the one in (6-15)), and the corresponding R,K,H,S.
CHAPTER II
COVARIANCE KERNELS AND INTEGRAL EQUATIONS

The main difficulty in the use of the K.L. expansion lies in the computation of the eigenvalues and eigenfunctions \( \lambda \) and \( \phi \) associated with it. In this chapter we will review results and methods for the solution of this problem, with emphasis on the more recent ones and suggest slight improvements.

SECTION VII
PROPERTIES OF COVARIANCE KERNELS

A. Properties of the eigenvalues and eigenfunctions

We are dealing with covariance kernels, i.e. the class of function such that

\[
\mathcal{R}(s, t) = \overline{\mathcal{R}(t, s)} , \quad \int_0^T \int_0^T \phi(s) \mathcal{R}(s, t) \phi(t) \, dt \, ds > 0 \quad (7-1)
\]

for any function \( \phi(t) \). It is of great interest to know how the eigenvalues and eigenfunctions of such kernels behave; in particular, how do they vary with \( T \), the interval of excitation and observation; and what happens when \( n \to \infty \), \( n \) being the ordering index of the sequence

\[
\lambda_1 \gg \lambda_2 \gg \ldots \gg \lambda_n \gg \ldots \quad (7-2)
\]

To emphasize the dependence on \( T \), we will often write

\[
\lambda_i (T) \quad \text{and} \quad \phi_i (t; T)
\]
We will review a few of the results; for their proof, we will refer to different authors.

1. Let us consider the integral equation

\[ \lambda \phi(t; \mathcal{T}) = \int_0^T R(t, s) \phi(s; \mathcal{T}) \, ds \]  

(7-3)

Then every \( \lambda \phi(t; \mathcal{T}) \) is a monotone increasing function of the interval \( [22] \). This property gives in particular a bound for the \( \lambda \)'s

\[ \lambda \phi(t; \mathcal{T}) < \lambda \phi(t; \infty) \]  

(7-4)

2. Each of the eigenfunctions \( \phi(t; \mathcal{T}) \) of a stationary random process has either odd or even symmetry about the point \( t = \mathcal{T}/2 \) [26]. We give a general simple proof in section IX. This allows a simplification of the calculations of the eigenfunctions.

3. Stationary random processes can be expanded on a finite interval \( T \) in a harmonic Fourier series with uncorrelated coefficients if and only if the autocorrelation function is periodic, and of period \( T \). (This result is due to Root and Pitcher [27].)

4. The largest eigenvalue \( \lambda \phi(t) \), is equal to the maximum of the functional

\[ J(\phi, \phi) = \int_0^T \phi(t) R(t, s) \phi(s) \, dt \, ds \]  

(7-5)

over all \( \phi(t) \) such that

\[ \int_0^T \phi(t) \, dt = 1 \]  

(7-6)

This maximum is reached for

\[ \phi(t) = \phi(t) \]  

(7-7)

This property is extremely useful in optimal signal design, in both deterministic and random cases [28], [29]. For the proof, one may refer to any book on integral equations. Closely related to this is the following theorem due to Ky Fan [8]:

\[ \sum_{k=1}^{n} \lambda_k = \sup \sum_{k=1}^{n} (R f_k, f_k) = \sup \sum_{k=1}^{n} \int_0^T \int_0^T f_k(t) R(t, s) f_k(s) \, dt \, ds \]  

(7-8)
the upper bound being taken over all orthonormal systems \( \left\{ f_1, f_2, \cdots, f_n \right\} \) containing \( n \) elements.

5. If \( R(t,s) = R(|t-\lambda|) \) has a rational spectral density (sometimes, it is called a "rational" kernel) \( S(\omega) \)

\[
S(\omega) = \frac{N(-\omega^2)}{D(-\omega^2)}
\]  

(7-9)

with

\[
N(-\omega^2) = \sum_{k=0}^{n} a_{2k} \omega^{2k}
\]  

(7-10)

\[
D(-\omega^2) = \sum_{k=0}^{n} b_{2k} \omega^{2k}
\]  

(7-11)

\[
r = n - m > 0
\]  

(7-12)

then, when \( n \to \infty \), we have the asymptotic expressions

\[
\lambda_i \sim \frac{a_{2m_i}}{b_{2m_i}} \left[ \frac{\pi i + \frac{1}{2}}{T} \right]^{-2r}
\]  

(7-13)

\[
\Phi_{\omega} (t) \sim \left( \frac{2}{T} \right)^{1/2} \cos \left[ \frac{\pi i + \frac{1}{2}}{T} t + \frac{\pi (r-1)}{4} \right]
\]  

(7-14)

for \( r \) even;

and, for \( r \) odd

\[
\lambda_i \sim \frac{a_{2m_i}}{b_{2m_i}} \left[ \frac{\pi i}{T} \right]^{-2r}
\]  

(7-15)

\[
\Phi_{\omega} (t) \sim \left( \frac{2}{T} \right)^{1/2} \cos \left[ \pi i t + \pi \frac{(r-1)}{4} \right]
\]  

(7-16)
The proof is found in Capon's article [30].

These asymptotic expressions are not only useful for exhibiting the behavior of the eigenvalues and eigenfunctions for large $i$ (sinusoids); but also for the determination of the truncation-error in the K.L. expansion quite conveniently.

6. With the notations as before, let $R(t,s) = R(t-\lambda)$ again and

$$G(\omega) = \left[ S(\omega) \right]^{\frac{1}{2}} \tag{7-17}$$

with corresponding inverse Fourier transform $g(u)$

$$g(u) \leftrightarrow G(\omega) \tag{7-18}$$

We define $\epsilon_n(t:T)$ by the equation

$$\left[ \lambda_n(T) \right]^{\frac{1}{2}} \epsilon_n(t:T) = \int_t^T g(t-\lambda) \phi_n(s:T) \, ds \tag{7-19}$$

for all $t \in (-\infty, +\infty)$.

Then, for each $T$

$$\int_{-\infty}^{+\infty} \epsilon_n(t:T) \epsilon_m(t:T) \, dt = \delta_{nm} \tag{7-20}$$

i.e the system $\{\epsilon_n(t:T)\}$ is orthonormal - though not necessarily complete - over $(-\infty, +\infty)$. For the proof, cf. [33] or [33].

B. Strictly positive definite kernels

Covariance kernels are all non-negative definite. In fact, the class of non-negative definite kernels is the class of covariance kernels [7].

From the theory of integral equations, we know that strict positive definiteness of kernels is equivalent to strict positiveness of all its
eigenvalues, which in turn is equivalent to the completeness of the corresponding eigenfunctions in the given $L_2$ space, e.g. $L_2 \left[ \sigma, T \right]$. In some problems, the assumption of completeness is crucial, e.g. cf. Chapt. 4 in [23]. We therefore need to be at rest as much as possible concerning this condition in the problems we have to handle by using the $\{ \phi_n(t) \}$ system. The available results, those easy to use, are unfortunately limited to stationary kernels, where they then apply to large and important subclasses, as shown by the following two theorems.

**Theorem 7.1** If $R_T(t-s) = R(t)$ is a stationary covariance kernel over the interval $T$, and if: a) the set on which $S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} R(t)e^{i\omega t} dt = c$ has measure zero; or if b) the covariance function $R(\tau)$ belongs to $L_2$ on $(-\infty, +\infty)$ and the set on which $S(\omega) > 0$ is of positive measure and is contained in a finite interval, then 0 is not an eigenvalue of $-\alpha \lambda_n(t-n)$. The proof is found in [32]. In particular, sufficient condition (a) of this theorem implies that if $S(\omega)$ is rational, then 0 is not an eigenvalue. This particular result is very useful. For instance, the set of eigenfunctions of $R(t-s) = e^{-\alpha|t-s|}$ or $R(t,s) = e^{-\alpha|t-s|} \cos \omega(t-s)$ which are of great importance in noise problems associated with lumped electrical networks, is complete.

**Theorem 7.2** Let $R(t-s)$ be a continuous autocorrelation function. If the points of increase of

$$G(\omega) = \int_{-\infty}^{\omega} S(\omega) d\omega$$

contain a convergent set (i.e. a set with an accumulation point), then $R_T$ is strictly positive definite.

Let us remark that this is a very weak condition; for instance, it is
sufficient that $G(\omega)$ have any continuous part at all. A fortiori, it is sufficient that $G$ have a non-zero density function. For the proof of theorem 7.2, cf [23, p 189].

Let us apply Theorem 7.2 to the band-limited white noise kernel

$$R(t, \lambda) = \frac{\sin \Omega |t-\lambda|}{\pi |t-\lambda|}$$

$$\mathcal{S}(\omega) = \mathcal{P}_\Omega(\omega) = \left\{ \begin{array}{ll}
1 & \omega \leq \Omega \\
0 & \omega > \Omega
\end{array} \right.$$  

The eigenfunctions of $R$ are the prolate spheroidal functions, studied extensively by Slepian [60]. We deduce that they form a complete set, since $G(\omega)$ obeys the conditions of the theorem and $R(t,s)$ is strictly positive definite.

SECTION VIII

SOLUTION OF THE INTEGRAL EQUATIONS ASSOCIATED WITH THE K.L.

We are directly concerned with the solution of the homogeneous Fredholm equation of the second kind; but often also with the Fredholm equation the first kind, as found in problems of detection for instance.

A. Analytical Techniques.

1. Case when the spectral density is rational. Here a general technique is available, but the problem remains of determining coefficients in as easy a way as possible. Equations of the second and first kinds can be solved, the solution of the latter one usually containing delta functions and their derivatives. The procedure is straightforward but tedious (cf. [36], appendix; or [37], Appendix and chapter VI). These techniques have been extended to the vector case [38]. An alternative approach for the latter case is given in (B) also. General methods applicable to both "rational" and
"non-rational" kernels are:

(a) In the case of degenerate kernels \( R(t, s) \),

\[
R(t, s) = \sum_{i=i}^{N} \lambda_i \Phi_x(t) \Phi_i(s), \quad N \text{ finite}
\]

there is obviously no problem for the equation of the second kind; and equations of the first kind are easily solved [22, p. 323].

Degenerate kernels are encountered in radar problems dealing with multiple targets.

B. Fredholm's technique [39, p. 151], allows the reduction of the finite dimensional vector-integral equation to the scalar one. It consists of the following:

Suppose we have to solve (as for the case of the vector K.L. expansion)

the vector equation

\[
\lambda \Phi(t) = \int_{a}^{T} R(t, \tau) \Phi(\tau) d\tau
\]

where

\[
\begin{bmatrix}
\Phi(t)
\end{bmatrix}^T = \begin{bmatrix}
\phi^1(t), \phi^2(t), ..., \phi^n(t)
\end{bmatrix}
\]

\[
R(t, \tau) = \begin{pmatrix}
R_{ij}(t, \tau)
\end{pmatrix}, \quad i, j = 1, 2, ...
\]

and

\[
R_{ij}(t, \tau) = \begin{bmatrix}
X_j(t)
\end{bmatrix} \begin{bmatrix}
X_i(\tau)
\end{bmatrix}
\]

To give a clearer picture of Fredholm's method, we consider the case \( N = 2 \).
Then we have
\[
\lambda \phi'(t) = \int_0^T R''(t, \tau) \phi'(\tau) d\tau + \int_0^T R_{12}(t, \tau) \phi^2(\tau) d\tau
\]
\[
\lambda \phi^2(t) = \int_0^T R_{21}(t, \tau) \phi'(\tau) d\tau + \int_0^T R_{22}(t, \tau) \phi^2(\tau) d\tau
\]
\[
(8-6)
\]

Let \( \mu(t) \) be the step function
\[
\mu(t) = \begin{cases} 
1 & t \geq 0 \\
0 & t < 0 
\end{cases}
\]

and let
\[
\Psi(t) = [\mu(t) - \mu(t-T)] \phi'(t) + [\mu(t-T) - \mu(t-2T)] \phi^2(t-T)
\]
\[
\rho(t, \tau) = [\mu(t) - \mu(t-T)] [\mu(\tau) - \mu(\tau-T)] R''(t, \tau)
\]
\[
+ [\mu(t-T) - \mu(t-2T)] [\mu(\tau) - \mu(\tau-T)] R_{12}(t, \tau-T)
\]
\[
+ [\mu(t-T) - \mu(t-2T)] [\mu(\tau) - \mu(\tau-T)] R_{21}(t-T, \tau)
\]
\[
+ [\mu(t-T) - \mu(t-2T)] [\mu(\tau) - \mu(\tau-T)] R_{22}(t-T, \tau-T)
\]

Then the system (8-6) is equivalent to
\[
\lambda \Psi(t) = \int_0^T \rho(t, \tau) \Psi(\tau) d\tau
\]

i.e. the system is reduced to the solution of a single integral equation. The method is general and could be helpful in particular when solving numerically. We then let
\[
\Psi(t) = \begin{cases} 
\phi'(t) & 0 \leq t \leq T \\
\phi^2(t-T) & T < t \leq 2T \\
\vdots & \\
\phi^N(t-(N-1)T) & (N-1)T < t \leq NT 
\end{cases}
\]
\[
(8-7)
\]

and \( \rho(t, \tau) \) to be equal to
\[ R^n(\tau, \tau'), R^{12}(\tau, \tau - \tau'), \ldots \]
\[ R^{21}(\tau - \tau', \tau), R^{NN}[\tau - (N-1)\tau', \tau - (N-1)\tau'] \]

respectively in the squares \[ (0 \leq \tau \leq \tau', 0 \leq \tau' \leq \tau), (0 \leq \tau < \tau', \tau < \tau' \leq 2\tau), \ldots \]
\[ (\tau < \tau' \leq 2\tau, 0 \leq \tau' \leq \tau), (N-1)\tau < \tau' \leq N\tau', (N-1)\tau < \tau < N\tau] \]

Therefore, whenever we can find analytically the eigenfunctions for all kernels \( R^{ij}(\tau, \tau) \) we are able to solve the vector equation (8-2) automatically by analytical methods. And this provides an alternative to the method already referred to for "rational" kernels [38].

2. Case when the kernel is not the Fourier transform of a rational power spectral density. ("Non rational" kernel)

Here we do not know of any general analytical method for the solution of Fredholm equations of both first and second kinds. We suggest however some guidelines for the approach to the problem.

a) As much as possible we should try to reduce the integral equation to a differential one with associated boundary conditions. Though the new system does usually have a closed form solution in terms of already well known functions, it may be simpler and is more amenable to computer solutions.

b) It would certainly be of help to set up tables any time the solution of a particular integral equation is found, of those equations that are reducible to the latter in one way or another. It is known, for instance, that the class of covariance functions is closed under integration. Therefore if a certain equation with covariance kernel \( R_o(\tau, \tau) \) can be reduced to a differential equation which we can solve,
then the integral equation with kernel $R_1(t, s) = \int_t^s R_0(x, \xi) \, dx \, d\xi$
can also be reduced to the same differential equation with two more
differentiations (with adjunction of the appropriate boundary values
of course).

c) Let $R_1(s, t) = f(s) \frac{\partial}{\partial s} R_0(s, t) \frac{d}{ds} f(t)$ (8-9)

Then the solution of all integral equations with kernels $R_1(s, t)$
is reduced to that of $R_0(s, t)$, whose solution is assumed to be
known. This last remark is important, as the case arises often in
engineering problems. It has been used [40] to solve the integral equation
associated with the general Gaussian-Markov process

$$R_c(t, s) = \int_t^s f(s) g(s) \quad 0 \leq t \leq T \leq T$$

$$= \frac{f(t)}{g(t)} \quad 0 \leq T \leq T$$ (8-10)

where $f(t)$ and $g(t)$ are continuous, and

$$h(t) = \frac{f(t)}{g(t)}$$ (8-11)

is continuous and strictly increasing in $[0, T]$.

Such processes transform, it is shown [41], to the Wiener process

$W^h(t)$, whose autocorrelation is

$$\min \left[ h(t), h(s) \right]$$, by the transformation

$$X(t) = g(t) W^h(t)$$ (8-12)

d) There is the special case where the integral equation is of
the form

$$m(t) = \int_0^t R(t, s) a(s) \, w(s) \, ds$$ (8-13)
where $\omega(t) > 0$ is known. It is immediately seen that (8-13) reduces to the ordinary case by writing it

$$\left[ \sqrt{\omega(t)} m(t) \right] = \int_0^T \left[ \sqrt{\omega(t)} R(t,s) \sqrt{\omega(s)} \right] \left[ \sqrt{\omega(s)} a(s) \right] ds \quad (8-14)$$

There are but meager analytical results available for the solution of integral equations with non-rational kernels. For more details, cf. [40], [42].

B. Computational techniques.

1. The state-variable approach (Baggeroer [43]. This is not a purely numerical technique, as it does yield results analytically also. But it is mostly helpful when used on a computer. Let us therefore consider a vector random process $\hat{\mathbf{Y}}(t)$ that is obtained by passing white noise $\hat{\mathbf{U}}(t)$ through a linear system described by the following N-dimensional equations:

$$\frac{d\hat{\mathbf{X}}(t)}{dt} = \mathbf{F}(t) \hat{\mathbf{X}}(t) + \mathbf{G}(t) \hat{\mathbf{U}}(t) \quad \text{(linear state equation)} \quad (8-15)$$

$$\hat{\mathbf{Y}}(t) = \mathbf{C}(t) \hat{\mathbf{X}}(t) \quad \text{(linear observation equation)} \quad (8-16)$$

Let the covariance matrix of the zero-mean white noise $\hat{\mathbf{U}}(t)$ be

$$\mathbb{E} \left[ \hat{\mathbf{U}}(t) \hat{\mathbf{U}}^T(\tau) \right] = \mathbf{Q} \delta(t-\tau) \quad (8-17)$$

where the superscript (T) indicates the transpose.

and the initial state of the system to be described by

$$\mathbb{E} \left[ \hat{\mathbf{X}}(T_o) \hat{\mathbf{X}}^T(T_o) \right] = \mathbf{P}_o \quad (8-18)$$
In these equations, $F$, $G$, $C$, $Q$, and $P_o$ are matrixes.

Then let us consider the equation

$$\lambda_i \hat{\phi}_i(t) = C(t) \int_{T_0}^{T} R_{\tilde{x}x}(t, \tau) C^T(\tau) \hat{\phi}_i(\tau) d\tau$$  \hspace{1cm} (8-19)

and

$$\tilde{\xi}_x(t) = \int_{T_0}^{T} R_{\tilde{x}x}(t, \tau) C^T(\tau) \hat{\phi}_i(\tau) d\tau$$  \hspace{1cm} (8-20)

With these notations and definitions, it is shown that we have the equivalent system (cf [143], or Appendix A-8).

$$\frac{d}{dt} \tilde{\xi}_{\alpha}(t) = F^T(t) \tilde{\xi}_{\alpha}(t) + G(t) Q C^T(t) \tilde{\eta}_{\alpha}(t)$$  \hspace{1cm} (8-21)

$$\frac{d}{dt} \tilde{\eta}_{\alpha}(t) = - \frac{C^T(t) C(t)}{\lambda_i} \tilde{\xi}_{\alpha}(t) - F^T(t) \tilde{\eta}_{\alpha}(t)$$  \hspace{1cm} (8-22)

with associated boundary conditions

$$\tilde{\xi}_{\alpha}(T_{\alpha}) = 0$$  \hspace{1cm} (8-23)

$$\tilde{\xi}_{\alpha}(T_0) = P_o \tilde{\eta}_{\alpha}(T_0)$$  \hspace{1cm} (8-24)

Therefore the homogeneous Fredholm integral equation is transformed into a set of differential equations (8-21) - (8-24) whose coefficients are expressed simply in terms of those of the state equations that are used to generate $\tilde{\gamma}(t)$. If we define the $(2N \times 2N)$ matrix $W(t; \lambda)$

$$W(t; \lambda) = \begin{bmatrix}
F^T(t) & GC(t)Q C^T(t) \\
\frac{C^T(t)C(t)}{\lambda} & -F^T(t)
\end{bmatrix}$$  \hspace{1cm} (8-25)
so that equations (8-21) and (8-22) are now written

\[
\begin{bmatrix}
\frac{d}{dt} \xi_1(t) \\
\vdots \\
\frac{d}{dt} \eta_i(t)
\end{bmatrix} = \mathcal{W}(t; \lambda_i)
\begin{bmatrix}
\xi_1(t) \\
\vdots \\
\eta_i(t)
\end{bmatrix}
\]

(8-26)

and if \( \Psi(t, T_0; \lambda) \) is the transition matrix associated with \( \mathcal{W}(t; \lambda_i) \), we have the solution

\[
\begin{bmatrix}
\xi_1(t) \\
\vdots \\
\eta_i(t)
\end{bmatrix} = \Psi(t, T_0; \lambda_i)
\begin{bmatrix}
P_0 \\
\vdots \\
I
\end{bmatrix}
\tilde{\eta}_i(T_0)
\]

(8-27)

Upon partitioning \( \Psi(t, T_0; \lambda_i) \) into four \( N \) by \( N \) matrixes,

\[
\Psi(t, T_0; \lambda) \triangleq \begin{bmatrix}
\psi_{\xi \xi}(t, T_0; \lambda) & \psi_{\xi \eta}(t, T_0; \lambda) \\
\psi_{\eta \xi}(t, T_0; \lambda) & \psi_{\eta \eta}(t, T_0; \lambda)
\end{bmatrix}
\]

(8-28)

we get

\[
\begin{bmatrix}
\xi_1(t) \\
\vdots \\
\eta_i(t)
\end{bmatrix} = \begin{bmatrix}
\psi_{\xi \xi}(t, T_0; \lambda)P_0\psi_{\xi \eta}(\cdot) \\
\vdots \\
\psi_{\eta \xi}(\cdot)P_0\psi_{\eta \eta}(\cdot)
\end{bmatrix}
\tilde{\eta}_i(T_0)
\]

(8-29)

while the boundary conditions (8-23) yields
\[ \tilde{\eta}_i(T_f) = \left[ \Psi_{\eta \xi}^0(T_f, T_o; \lambda_i) \right] \tilde{\eta}_i(T_o) = 0 \quad (8-30) \]

Equation (8-30) is a set of homogeneous equations, and has non-trivial solutions if and only if

\[ \det \mathbf{A} = \det \left[ \Psi_{\eta \xi}^0(T_f, T_o; \lambda_i) \mathbf{P}_0 + \Psi_{\eta \eta}^0(T_f, T_o; \lambda_i) \right] = 0 \quad (8-31) \]

This equation is the eigenvalue equation; whose roots are the eigenvalues \( \lambda_i \).

Then the corresponding eigenfunctions are immediately given by

\[ \hat{\phi}_i(t) = \frac{1}{\lambda_i} \mathbf{C}_i(t) \hat{\xi}_i(t) \quad (8-32) \]

from (8-19) and (8-20). Therefore

\[ \hat{\phi}_i(t) = \frac{\mathbf{C}_i(t)}{\lambda_i} \left[ \Psi_{\xi \xi}^0(t, T_o; \lambda_i) \mathbf{P}_0 + \Psi_{\eta \eta}^0(t, T_o; \lambda_i) \right] \tilde{\eta}_i(T_o) \quad (8-33) \]

where \( \tilde{\eta}_i(T_o) \) satisfies the orthogonality relation derived from (8-24)

\[ \left[ \Psi_{\eta \xi}^0(T_f, T_o; \lambda_i) \mathbf{P}_0 + \Psi_{\eta \eta}^0(T_f, T_o; \lambda_i) \right] \tilde{\eta}_i(T_o) = 0 \quad (8-34) \]
To summarize, once we have found a state-variable representation (8-15), (8-16) for the process $\mathcal{X}(t)$ whose covariance is given; and once we have determined the transition matrix associated with the matrix $\mathcal{W}(t, \lambda t)$, then we have immediately the two equations (8-31) and (8-33) to give the eigenvalues and eigenfunctions of equation (8-19).

Let us remark the advantages of the method. It can handle the vector case without any difficulty; the eigenfunctions are determined independently one from another, but above all, even when it is not possible to determine the transition matrix analytically (which is generally the case for time-varying systems), still it can be calculated by numerical methods, and the solution to (8-31) and (8-33) can be determined by computer. One drawback is that it is not always possible to determine the system (8-15), (8-16).

As an illustration of the method, we will solve the simple case of the Wiener process. We are therefore given that

$$R_{yy}(t, \tau) = \mu^2 \min(t, \tau), \quad 0 \leq t, \tau \leq T$$

(8-35)

A state-variable representation of a system that generates $\mathcal{X}(t)$ is

$$\frac{d}{dt} X(t) = \mu^2 \mathcal{U}(t) \quad , \quad 0 \leq t \leq T$$

(8-36)

$$\mathcal{X}(t) = \mathcal{X}(t)$$

(8-37)
with
\[
E \left\{ \mu(t) \cdot \mu(t') \right\} = \delta(t-t')
\]
(8-38)

\[
E \left\{ \chi^2(\sigma) \right\} = 0
\]
(8-39)

From which the matrix \( \mathbf{W} \) is obtained
\[
\mathbf{W}(\lambda) = \begin{bmatrix}
0 & 1 \\
-\lambda & 0 \\
-\mu & 0 \\
\end{bmatrix}
\]
(8-40)

The transition matrix is given by
\[
\mathcal{P}(t; T_0; A_0) = e^{\mathbf{W}(\lambda)(t-T_0)}
\]
(8-41)

i.e. here
\[
\mathcal{P}(t; 0; A) = \begin{bmatrix}
\alpha_0 \left( \frac{\mu}{\sqrt{\lambda}} t \right) & \sqrt{\lambda} \sin \left( \frac{\mu}{\sqrt{\lambda}} t \right) \\
-\frac{\mu}{\sqrt{\lambda}} \sin \left( \frac{\mu}{\sqrt{\lambda}} t \right) & \cos \left( \frac{\mu}{\sqrt{\lambda}} t \right)
\end{bmatrix}
\]
(8-42)

The eigenvalue equation is now, from (8-31),
\[
\alpha_0 \left( \frac{\mu}{\sqrt{\lambda}} T \right) = 0, \text{ i.e. } \lambda i = \left[ \frac{2\mu^T}{(2i-1)\pi} \right]^2
\]
(8-43)

for \( i = 1, 2, \ldots \).
2. The matrix-approximation method. This is a very general and very unsophisticated method, which ties together the usual K.L. and its discrete form as studied in sections II and III.

Our problem is to solve

\[
\lambda_i \mathbf{x}_i = \int_0^T R(t,s) \phi_i(s) \, \mathrm{d}s
\]

which corresponds, as we know, to

\[
\lambda_i \mathbf{x}_i = \int_0^T \mathbf{X}(t) \phi_i(t) \, \mathrm{d}t
\]

Let us approximate \(\mathbf{x}_i\) by \(\mathbf{x}_i'\) given by

\[
\lambda_i \mathbf{x}_i' = \frac{T}{N} \sum_{j=1}^{N} \mathbf{X}(t_j) \phi_i(t_j) , \quad j = 1, 2, \ldots, N
\]

or

\[
\mathbf{x}_i' = \sum_{j=1}^{N} \mathbf{X}(t_j) \phi_{ji} , \quad \text{with}
\]

\[
\phi_{ji} = \frac{T}{\lambda_i N} \psi_i(t_{ij})
\]

\[(8-44)\]
where the $\mathcal{X}(t_j)$ are time samples taken every $\frac{T}{N}$ seconds.

Let us suppose we will consider only $n$ eigenvalues and eigenfunctions $\lambda_i$ and $\phi_i(t)$ of (8-45), where in general $n < N$.

Therefore we can write

$$\tilde{\mathcal{X}}' = \Phi^T \tilde{\mathcal{X}}$$

(8-48)

where

$$\Phi = (\Phi_i(t)) ; \quad \tilde{\mathcal{X}}^T = [\mathcal{X}(t_1), \mathcal{X}(t_2), \ldots, \mathcal{X}(t_N)]$$

and

$$\tilde{\mathcal{X}}' = [\mathcal{X}'_1, \mathcal{X}'_2, \ldots, \mathcal{X}'_N]$$

Because of the uniqueness properties of the spectral representation of non-negative definite matrixes, $\lambda_i$ and $\phi_i(t)$ are well approximated for $N$ large by the eigenvalues and eigenvectors of the matrix equation

$$\mu_i \beta_i = \mathcal{R} \beta_i$$

(8-49)

where

$$\mathcal{R} = (R_{i,k}) = \mathbb{E} \{ \mathcal{X}(t_i) \mathcal{X}(t_k) \} ; \quad i, k = 1, 2, \ldots, N$$

i.e.

$$\mu_i \sim \lambda_i$$

$$(\beta_1, \beta_2, \ldots, \beta_N) \rightarrow [\phi_i(\frac{T}{N}), \phi_i(\frac{2T}{N}), \ldots, \phi_i(T)]$$

(8-50)

An application of the method is found in [11]. It is shown that it leads to a very satisfactory approximation on $\mathcal{R}(t, \tau) = e^{-\frac{\|t-\tau\|}{\tau^2}}$. 

The advantages of this method is that it is applicable to all cases; that today algorithms and computers are ready for the solution of eigenvalues and eigenvectors of large matrixes, and that it does not require any preliminary setup like the state-variable approach.
A. A model for the random linear time-varying channel:

a) The search for efficient, practical models for the random linear time-varying channel has been going on for some time. To our knowledge, the only general and widely used model so far is the so-called canonic or tapped delay-line one. As a quick review we will consider the case of the low-pass canonic model, valid only for the class of signals $s(t)$ of bandwidth $(-W_s, W_s)$:

With reference to Fig. 9-1,

$$ f_h(t, \tau) = \frac{\sin 2\pi W_s \tau}{2\pi W_s \tau} \times \sum_{\alpha=0}^{n-1} f_h(t, \alpha, \tau) \delta (\tau - \frac{\alpha}{2W_s}) \quad (9-1) $$
and

\[ a_{\alpha+1}(t) = \sum_{\alpha=0}^{n-1} L(t, \alpha, \frac{\alpha}{2W_s}), \alpha = 0, 1, \ldots, (n-1) \]

(9-2)

where * represents convolution, \( L(t, \alpha, \frac{\alpha}{2W_s}) \) is the impulse response of the random channel, and \( n \) is the number of taps (or samples on the channel function) needed to extract sufficient knowledge about the channel under the given assumptions.

Usually it is assumed that the tap functions \( a_{\alpha+1}(t) \) are themselves all band-limited to \( (-W_c, W_c) \), that the channel has a finite memory of \( L \) sec., and that the input signal \( s(t) \) is also time-limited to \( [0, T'] \). All these assumptions have a lot of physical appeal, but they are not very rigorous.

If we let \( W = (W_c + W_s) \) be the output bandwidth, then approximately \( n = \left( \frac{2W}{L} + 1 \right) \) and \( T = (T' + L) \) gives the number of taps sufficient for the model. The time \( T \) is the interval of channel sampling. We can then write

\[ g_i(t) = \int_{-\infty}^{\infty} L(t, \tau) s(\tau) d\tau \]

or

\[ \tilde{g}(t) = \sum_{\alpha=1}^{n} a_{\alpha}(t) \lambda(t - \frac{\alpha-1}{2W}) \]

(9-3)

The input \( s(t) \) needs to be sampled \( 2W'T' = \tilde{W} \) times.

Let us define \( \tilde{a}_{\lambda_i} = a_{\lambda_i}(t) \), where \( \lambda_i \) is one of the sampling times of the signal; and the vector \( \tilde{\lambda} \) and matrix \( \tilde{S} \) by:

\[ \tilde{\lambda}^T = \left( \tilde{a}_n, \tilde{a}_{n+1}, \ldots, \tilde{a}_{n+k}, \tilde{a}_{1}, \tilde{a}_{2}, \ldots, \tilde{a}_{k} \right) \]

(9-4)
Then the vector \( \tilde{z} \) of the samples of \( z(t) \) is given by

\[
\tilde{z} = S \tilde{A}
\]  

(9-6)

And if

\[
\Phi_{aa} \triangleq E \left[ (\tilde{A} - E(\tilde{A})) (\tilde{A} - E(\tilde{A}))^T \right]
\]

(9-7)

then

\[
\Phi_{zz} = E \left[ (\tilde{z} - E(\tilde{z})) (\tilde{z} - E(\tilde{z}))^T \right]
\]

\[
\Phi_{zz} = S \Phi_{aa} S^T
\]

(9-8)

With this formulation, one can handle usefully simply idealized cases.

For example, for a very-fast fading channel, we assume that \( \Phi_{aa} \) is diagonal; for a time-invariant channel whose tap gains are uncorrelated and have the same Gaussian distribution, \( \Phi_{aa} \) reduces to a matrix consisting of \( n \times k \) by \( k \) submatrixes on the diagonal.
One disadvantage however is that the \( \Omega_\lambda(t) \) samples are not statistically independent, since there is an oversampling of the channel-tap signals. In any event, uniform time-sampling does not provide independent samples except for very few cases, like the band-limited white process. For more details, cf. [44], or [45].

b) In what follows, we suggest another kind of model and we suppose we have a **second-order characterization** of the channel. Our purpose is to attempt to use the K.L. expansion to obtain a model for the random time-varying channel that is, in some sense, better than the tapped-delay-line one.

The ideal would be to find such an expansion for \( \mathcal{H}(t,\tau) \)

\[
\mathcal{H}(t,\tau) = \sum_i a_i \psi_i(t,\tau)
\]

as to produce the K.L. expansion for the output \( \mathcal{Z}(t) \) itself, i.e.,

\[
\mathcal{Z}(t) = \int_\tau \sum_i a_i \psi_i(t,\tau) \phi(\tau) d\tau = \sum_i a_i \phi_i(t)
\]

with

\[
E(a_i a_j) = 0 \quad \text{for} \quad i \neq j
\]

\[
\int_T \phi_i(t) \phi_j(t) dt = 0 \quad \text{for} \quad i \neq j.
\]

It is easy to see from this that to obtain a reasonable general expansion of \( \mathcal{H}(t,\tau) \) that would fulfill these two conditions is hopelessly difficult, as the second orthogonality condition requires that

\[
\int_T dt \left[ \int_\tau \psi_i(t,\tau) \phi(\tau) d\tau \int_\tau \psi_j(t,\tau') \phi(\tau') d\tau' \right] = 0 \quad \text{for} \quad i \neq j.
\]

i.e. the expansion of \( \mathcal{H}(t,\tau) \) must depend on the particular signal \( s(t) \).

Let us note that if we had a deterministic linear time-varying channel
with a random input \( x(t) \), we could easily enough obtain the appropriate expansion of \( x(t) \) that leads to the K.L. expansion of \( x(t) \) [54].

In the following, our approach will consist of starting from a K.L. expansion of the channel itself

\[
\phi_\lambda(t, \tau) = \sum_{i=1}^{\infty} \lambda_i \phi_i(t, \tau) \quad \tau \in [-T_\lambda, T_\lambda] (9-9)
\]

where

\[
E \left\{ \frac{\phi_i(t, \tau) \phi_j(t', \tau')}{\tau} \right\} = \delta_{ij}
\]

\[
\lambda_i \phi_i(t, \tau) = \int \int_{T} E \left\{ \phi_i(t, \tau) \frac{\phi_j(t', \tau')}{\tau} \right\} \phi_j(t', \tau') \, dt' \, d\tau' \int_{T} \phi_i(t, \tau) \phi_j(t, \tau) \, dt \, d\tau = \delta_{ij}
\]

1. The output \( x(t) \) can be expressed as

\[
x(t) = \sum_{i=1}^{\infty} \lambda_i \int_{-T_\lambda}^{T_\lambda} \phi_i(t, \tau) s(\tau) \, d\tau
\]

or

\[
\begin{cases}
x(t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(t) \\
\phi_i(t) = \int_{-T_\lambda}^{T_\lambda} \phi_i(t, \tau) s(\tau) \, d\tau
\end{cases}
\]

(9-10)

We have in (9-10) a canonical expansion for \( x(t) \), but the system \( \{\phi_i(t)\} \) is generally not orthogonal. Expansion (9-10) could be mostly useful when the shape of \( s(t) \) is very simple, e.g. rectangular, triangular, etc...
2. We can go a step further and get a channel model that is adapted to any signal which is band-limited to \((-W_s, W_s)\).

Letting

\[ S(t) = \sum_{n=-\infty}^{+\infty} S \left( \frac{n}{\pi W_s} \right) \frac{\sin \pi (2 W_s t - n)}{\pi (2 W_s t - n)} , \quad t \in (-\infty, +\infty) \]

then (9-10) becomes

\[ \tilde{z}(t) = \sum_{n} \sum_{i} \tilde{h}_i \left( \frac{n}{\pi W_s} \right) \psi_{in}(t) , \quad t \in [-T_z, T_z] \]

(9-11)

where

\[ \psi_{in}(t) = \int_{-T_z}^{T_z} \phi_i(t, \tau) \frac{\sin \pi (2 W_s \tau - n)}{\pi (2 W_s \tau - n)} d\tau \]

We represent (9-11) by the following figure

![Diagram](image)

**Fig. 9-2**

Let us note that the implementation of such a model requires a delay of at least \(T\) seconds.

At this point we remark that the evaluation of the \(\psi_{in}(t)\) functions is a complicated task. However it has been established by L. Zadeh [55]
that if one is observing the input and output of a time-varying black box B and it is not known whether B is linear or not, then there is no way of deciding between the t alternatives if B is non-stationary. This is why we will assume that the channel is observation-time stationary (o.t.s.). Then

\[ E \left\{ \frac{P(t, \tau)}{\sigma(t', \tau')} \right\} = R \left( t - t', \tau, \tau' \right) \]

and from section II,

\[ R \left( t - t', \tau, \tau' \right) = \sum_{\lambda} \lambda \phi_{\lambda} (t, \tau) \phi_{\lambda} (t', \tau') \quad \forall \tau, t, t', \tau' \in [-T_{\Delta}, T_{\Delta}] \]

which is also equal to

\[ \sum_{\lambda} \lambda \phi_{\lambda} (-t, \tau) \phi_{\lambda} (-t', \tau') = R \left( t - t', \tau, \tau' \right) \]

(9-12)

and this shows that the non-degenerate eigenfunctions \( \phi_{\lambda} (t, \tau) \) are either even or odd in t.

This property allows the reduction of the amount of computation by approximately \( \frac{1}{2} \). And if the channel is W.S.S. in the excitation time \( \tau \) also, since \( \sum_{\lambda} \frac{\pi \left( \mathcal{E}_{\lambda} - \eta \right)}{\pi \left( \mathcal{E}_{\lambda} - \mathcal{E}_{\lambda} \right)} \) is even, the integration can also be reduced to approximately \( \frac{1}{2} \) of the interval \( [-T_{\Delta}, T_{\Delta}] \).

Up to now we have not indicated the domains of variation of the indices \( i \) and \( n \). For \( n \), it is known that we can take its range from \( -\mathcal{W} \) to \( \mathcal{W} \). As for \( i \), \( i=1, 2, \ldots \), the number I will depend on the rate of convergence of the series \( \sum_{\lambda} \lambda = \sum_{\lambda} E \left( \mathcal{R}_{\lambda} \right) \), itself will in turn depend on
the autocorrelation of the channel. As some experiments show [11], it
will be quite good to limit $I$ to 10 to 20. It should be pointed out
that the "canonic" expansion $\sum_n S \left( \frac{\alpha}{\xi \sqrt{\lambda}} \right) \sin \pi \left( \frac{1}{\xi} \frac{\alpha}{\sqrt{\lambda}} \right)$ of $S(t)$ is not optimum in
the finite interval $[0, T]$, as it does not yield the highest rate of con­
vergence, though it is most simple, and very suitable for $(-\infty, +\infty)$.
The model should therefore be improved by considering other kinds of
expansions for $S(t)$, for instance, with Legendre polynomials. There
also remains the problem of evaluating the mean-square truncation error
in (9-11).

3: In the tapped-delay-line model, it is assumed that the fast-fading
channel is one for which $\Phi$ is diagonal, and the slow-fading one is
characterized by $\left( \Phi_{i,j} \right)$ constant for all $i$ and all $j$ [45]. It is
easy to see that the first assumption is usually incorrect, because it
is shown, at least for the wide-sense-stationary process which is band­
limited, that it can have its real time-samples (Nyquist rate) uncorrela­
ted if and only if the process has a flat spectrum [51]. In fact in the
two above considered cases, the tapped-delay line model loses its useful­
ness, as we could obtain the same degree of approximation by considering
basic definitions:

slowly fluctuating (stationary) process: $\frac{dR(t)}{dt}$ small (for small $t$)
fast-fluctuating (stationary) process: $\frac{dR(t)}{dt}$ large (for small $t$)

We suggest it would be better to assume (and justify experimentally for
a given channel) autocorrelations of the form

$$\mathcal{E} \frac{|t-\tau|}{a} \quad \text{or} \quad \mathcal{E} \frac{|t-\tau|^{2}}{a^{2}}$$

, or a sum of such terms.
**TABLE 2**

**COMPARISON OF THE TWO CHANNEL MODELS**

<table>
<thead>
<tr>
<th>Tapped-delay-line model</th>
<th>Model based on the K.L. expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need to know only general parameters of the channel, i.e. bandwidth $W_c$ and memory $L$.</td>
<td>Need to know precisely, or to assume, the autocorrelation function of the channel $\mathbb{E}{h(t,\tau)h(t',\tau')}$</td>
</tr>
<tr>
<td>A rather general model</td>
<td>Much more specific.</td>
</tr>
<tr>
<td>No complicated calculations. Based on direct sampling in the time domain.</td>
<td>The eigenfunctions $\phi_i(t,\tau)$ have to be determined, and this is a rather complicated task.</td>
</tr>
<tr>
<td>Valid for $t, \tau$ varying in $(-\infty, +\infty)$. The assumption that the signal $s(t)$ is both time and band-limited is an approximation.</td>
<td>Valid for $T$ ($t, \tau \in T$) finite or infinite.</td>
</tr>
<tr>
<td>The major problem is the correlatedness (and dependence) of the samples. Not only because of over-sampling, but because of ad hoc assumptions of independence of sample-independence (e.g. fast fading channel).</td>
<td>Provides with a systematic way of yielding independent coefficients in the expansion. Very important for further operations like detection, estimation, etc...</td>
</tr>
<tr>
<td>Slow convergence. $h_c(t,\tau) = \frac{\sin \pi \frac{\tau}{W_c}}{\pi \frac{\tau}{W_c}} \times \sum_{\alpha \neq 0} h_{\alpha}(t, \frac{\alpha}{2W_c}) \delta(t-\frac{\alpha}{2W_c})$ (for $s(t)$ low pass).</td>
<td>Usually rapid convergence of the expansion of the model. $h_c(t,\tau) = \sum_{i \epsilon i} H_i(t,\tau) \phi_i(t,\tau)$</td>
</tr>
</tbody>
</table>
Then we may assume "a" to be small for the fast fading channel, and "a" large for the slow-fading channel, where "a" is an autocorrelation distance.

Let us give an example for a particular channel model.

Example. We assume that we have acausal time-invariant channel

\[ \delta_{\tilde{H}}(t, \tau) = \delta_{\tilde{H}}(t-\tau) \quad \tau, t \in (-\infty, +\infty) \]

For a given input \( A(t) \) the output of such a channel is given by

\[ z_{\tilde{H}}(t) = \int_{-\infty}^{t} \delta_{\tilde{H}}(t-\tau) A(\tau) \, d\tau \]

Let \( R_{\tilde{H}}(\omega, \omega') \) denote the two-dimensional Fourier transform of \( E\{ \tilde{H}(\omega) \tilde{H}(\omega') \} \) and \( S(\omega) \) the Fourier transform of \( A(t) \). Let us also assume that

\[ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |R_{\tilde{H}}(\omega, \omega')|^2 |S(\omega)|^2 |S(\omega')|^2 \, d\omega \, d\omega' < \infty \]

Then the results of section II allow us to write:

\[ Z(\omega) = \sum_{k} H_{i} \tilde{F}_{k}(\omega) S(\omega) \]

with \( Z(\omega) \) defined as the stochastic Fourier transform of \( Z(t) \),

\[ \int_{-\infty}^{+\infty} R_{\tilde{H}}(\omega, \omega') \tilde{F}_{k}(\omega') |S(\omega)|^2 \, d\omega' = \lambda\phi_{k}(\omega) \]
Therefore (9-13) yields the K.L. expansion of the output of the channel in the frequency domain.

The same technique can be used to get a K.L. representation in the complex-frequency domain by use of the Laplace transform, with the domain of variation of \( t \) and \( \tau \) being \([0, T]\).

**Remark:** If we assume the channel to be time-invariant and stationary, then we can write:

\[
E \left\{ H_i(t) H_j(t') \right\} = \lambda_c \delta_{i,j}
\]

Let \( X(t) \) be the output of the channel for a given input \( X(t) \).
Assuming moreover the \( \{ \Phi_i(t) \} \) system to be complete with respect to \( X(t) \), we have:

\[
X(t) = \sum_{i} x_i \Phi_i(t)
\]

where

\[
x_i = \int_{0}^{T} X(t) \Phi_i(t) dt
\]

Denoting by \( \xi_i(t) \) the random process

\[
\xi_i(t) = \int_{0}^{T} X(t - \tau) \Phi_i(\tau) d\tau
\]

we get

\[
\int_{-\infty}^{+\infty} \Phi_i(\omega) \overline{\Phi_j(\omega)} |S(\omega)|^2 d\omega = \delta_{i,j}
\]
This means that, for any fixed value $t_0$ of $t$, the random variables $\xi_i(t_0)$, $i = 1, 2, \ldots$ are uncorrelated, and we can write:

$$E \left\{ \xi_i(t_0) \xi_j(t_0) \right\} = \int_0^\tau \int_0^\tau R(\tau - \tau') \phi_i(\tau) \phi_j(\tau') d\tau d\tau' = \Lambda_i \delta_{i,j}$$

In particular when the channel in Gaussian, the $\xi_i(t_0)$ become independent for a given value of $t$, with Gaussian distributions having $\Lambda_i$ as covariances, $i = 1, 2, \ldots$.

B. Problem of the implementation of the eigenfunctions

The concept of frequency is being generalized to new sets of complete orthogonal functions, and devices and systems are being investigated from this point of view. The practical use of the generalized frequency concept, (e.g. for the case of the Walsh functions), indicates a need to explore the possibility of implementing the eigenfunctions of a K.L. From what we have seen in Chapter II, i.e. because of the meager results obtained so far concerning the detailed structure of the eigenfunctions of covariance kernels, it becomes clear that such an implementation is an arduous problem. Y. Lee [46] has shown the way by designing filters to realize physically some important complete orthonormal systems, like Laguerre functions and this has been followed recently for the Walsh functions. In the case of rational spectra, the
eigenfunctions satisfy a differential equation with constant coefficients. They are therefore exponentials. We are sure in that case that the $\phi_i(t)$, when $i$ becomes large, are closely approximated by sinusoids, whose parameters are easily determined [30]. There are therefore, already cases where the implementation would seem not very difficult. For simplicity, we consider the Wiener process

$$R(t,s) = \min(t,s) \quad t,s \in [0,1]$$

Then, as we already have seen,

$$\phi_i(t) = \sqrt{2} \sin((i-1/2)\pi t)$$

The Laplace transform is

$$\mathcal{L} \phi_i(t) = \mathcal{L} \sin((i-1/2)\pi t) \cdot \frac{s^{i-1/2}}{s^{i-1/2} + (i-1/2)^2 \pi^2}$$

Such a circuit yields an output over $(0, \infty)$. We may restrict the output to $[0, 1]$ by using switches. To get the corresponding $X^t$, we connect the output to a multiplier which receives the random processes $X(t)$. We then have the output of the multiplier pass through an integrator:

$$X^t = \int_0^T X(t) \phi_i(t) \, dt$$

This method is in principle the same for $\mathcal{L}(t,s) = \mathcal{L} e^{-|t-s|}$, but the eigenfunctions are more difficult to obtain.

The general problem of practical implementation remains open.
The most important and most thorough applications of the K.L. so far have been in the problems of detection, estimation and recognition in the presence of random signals. Almost invariably it is assumed that the random signals are Gaussian so as to enable making a full probabilistic characterization of their K.L. coefficients, which then become independent. The most fundamental work in this area is U. Grenander's [8]. In the following we are not reviewing the whole theory of detection and estimation. Rather we are concerned with slight extensions of the results or improvement of the techniques through the use of the generalized K.L.

A. Detection

1. Sure signals in noise: The role of the K.L. here is first that it transforms a continuous signal into an (infinite dimensional) vector:

\[ \mathbf{x}(t) \rightarrow (x_1, x_2, \ldots, x_N, \ldots) \]

Therefore, when the matter is only one of comparing the probability attached to two such random signals:

\[ \mathbf{x}^o(t) \rightarrow (x_1^o, x_2^o, \ldots, x_N^o, \ldots) \]
\[ \mathbf{x}^1(t) \rightarrow (x_1^1, x_2^1, \ldots, x_N^1, \ldots) \]

where the vectors are represented by projections on the same set of coordinate functions \[ \Phi(l) \], then the ratio of the probability density functions corresponding to \( (x_1^o, x_2^o, \ldots, x_N^o) \) and \( (x_1^1, x_2^1, \ldots, x_N^1) \) tends to a limit as \( N \to \infty \).

There is the problem of mathematical rigor in the proof, which can be seen in the references e.g. [2]. This approach therefore, reduces the problem,
for statistical inference on a continuous process through such tests like the Bayesian test and Wald's sequential test, to one on finite dimensional vectors.

The second important role of the K.L., is that the projections \( Z_i \) are uncorrelated, therefore independent when \( Z_i(t) \) and \( Z_i'(t) \) are Gaussian processes. These two qualities are however common to all canonical expansions when the underlying statistics are Gaussian. But the fact that the K.L. has orthogonal coordinate functions makes it unique and very useful in theoretical calculations. To be more precise, suppose we want to detect a sure signal \( s(t) \) in noise \( n(t) \) from the observed signal \( \tilde{X}(t) \), and we base our decision on the maximum likelihood ratio test.

Therefore, let:

\[
X_i(t) = S(t) + N(t) \quad ; \quad 0 \leq t \leq T; \quad H_1
\]

\[
X_i(t) = N(t) \quad ; \quad 0 \leq t \leq T; \quad H_0
\]

and, with:

\[
E \left\{ N(t) N(t) \right\} = R(t, t)
\]

\[
\phi_{i}(t) = \int_{0}^{T} R(t, \tau) \phi_{i}(\tau) d\tau
\]

and

\[
X_i = \int_{0}^{T} X_i(t) \phi_{i}(t) dt = S_i + N_i
\]

with

\[
S_i = \int_{0}^{T} S(t) \phi_{i}(t) dt; \quad N_i = \int_{0}^{T} N(t) \phi_{i}(t) dt
\]

Then, in the Gaussian case, the likelihood ratio \( \Lambda \) is calculated

\[
\Lambda \left[ X_1, X_2, \ldots, X_n \right] = \frac{\prod_{i=1}^{N} \int \frac{1}{\sqrt{2\pi \lambda_i}} \exp \left[ -\frac{1}{2} \left( \frac{X_i - S_i}{\lambda_i} \right)^2 \right]}{\prod_{i=1}^{N} \int \frac{1}{\sqrt{2\pi \lambda_i}} \exp \left[ -\frac{1}{2} \left( \frac{X_i}{\lambda_i} \right)^2 \right]}\]

(10-2)

and then is compared to a preset threshold (cf. any testbook on Detection for the above results). Because of the monotonicity of the logarithmic
function, the decision is unchanged if we base it on $\ln \Lambda$ rather than $-\Lambda$.

And we see easily that:

$$\ln \Lambda \left[ x(t) \right] = \sum_{\lambda_i \neq \lambda} \frac{\sum s_i}{\lambda_i} + \frac{1}{2} \sum_{\lambda_i = \lambda} \frac{s_i^2}{\lambda_i} = \ln \Lambda \left[ x_1, x_2, \ldots, x_N, \ldots \right]$$

(10-3)

which is itself a Gaussian random variable. Since $1/2 \sum s_i^2 / \lambda_i$ is a constant, independent of the random signal, we can translate the decision threshold to get the new statistic:

$$\frac{1}{\Lambda} = \sum \frac{s_i}{\lambda_i}$$

(10-4)

It is seen that $\frac{1}{\Lambda}$ is very simply expressed in terms of the K.L. coefficients of $x(t)$ and the associated eigenvalues. Using operator formalism, we will get for $\frac{1}{\Lambda}$ an expression that does not require the knowledge of the $\lambda_i$'s, $\lambda_i$'s and $\lambda_i$'s individually. Writing

$$\int_0^T R(t, \tau) g(\tau) d\tau = \Lambda(t) a$$

$$R(t, \tau) g(\tau) = s$$

We then have

$$\frac{1}{\Lambda} = \int_0^T g(t) x(t) dt$$

(10-5)

In other words, to make the decision for detection, it is sufficient only to know the solution to an integral equation of the first kind and to correlate this solution with the random signal. This can be rigorously justified if that integral equation has a solution that is square integrable, for then:

$$g(t) = \sum \phi(t) \phi(t) ; \quad \sum \phi(t) = \int \frac{1}{2} \sum \lambda_i \phi(t) \phi(t) g(t) s_i^2 d\tau$$
Therefore:
\[ \sum \gamma_i \phi_i(t) = \sum \lambda_i g_i \phi_i(t) \ldots \]
\[ g_i = \gamma_i / \lambda_i \]

which yields immediately the above result. This shows why in many cases one is not interested in the K.L. itself, but in a functional over the K.L. and illustrates how many problems may be solved by other techniques like the R.K.H.S. method.

Remark: These results and techniques extend directly to the generalized forms of the K.L.; for example, to the "weighted" K.L.; or to the finite or (infinite) vectorial case. Because of the practical importance of the latter as a very handy tool for detection and estimation for radar systems (phased arrays), telemetry, sonar and seismic arrays, we will make a few comments. The K.L. technique applies here almost trivially, because, by its very definition, we have the map:

\[ \begin{array}{l}
\chi^1(t) \\
\chi^2(t) \\
\vdots \\
\chi^n(t)
\end{array} \rightarrow \left( z_1, z_2, \ldots, z_N, \ldots \right) \]

array of random processes Corresponding sequence of random coefficients of the K.L. expansion associated with the array of random processes (of section II.)

i.e. to a vector random process, we associate a series not of vector random variables, but of scalar random variables. We are not losing information by substituting the series to the vector process. Yet we keep the same simplicity of the likelihood ratio as for the case of the scalar random process. This makes the K.L. of great benefit, if we disregard the problems associated with the solution of the integral equations,
### TABLE 3

**DETECTION SUFFICIENT STATISTIC (SURE SIGNAL IN NOISE)**

<table>
<thead>
<tr>
<th>Single Channel</th>
<th>Sufficient Statistic</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X(t) = \sum \mathbf{x}_i \mathbf{\Phi}_i(t)$</td>
<td>$L = \frac{1}{\mathbf{\lambda}_0} \sum_i \mathbf{s}_i \mathbf{\Phi}_i(t)$</td>
<td>$H_0$</td>
</tr>
<tr>
<td>$X(t) = \mathbf{\Phi}_0(t)$</td>
<td>$L = \int X(t) dt$, with $\mathbf{g}(t) = \mathbf{s}(t)$</td>
<td>$H_1$</td>
</tr>
</tbody>
</table>

$\mathbf{s}(t)$ is the noise autocorrelation function

<table>
<thead>
<tr>
<th>Single Channel</th>
<th>Sufficient Statistic</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{X}(\mathbf{t}) = \mathbf{\sum} \mathbf{X}_i \mathbf{\Phi}_i(\mathbf{t})$</td>
<td>$L = \frac{1}{\mathbf{\lambda}_0} \sum_i \mathbf{s}_i \mathbf{\Phi}_i(\mathbf{t})$</td>
<td>$H_0$</td>
</tr>
<tr>
<td>$\mathbf{X}(\mathbf{t}) = \mathbf{\Phi}_0(\mathbf{t})$</td>
<td>$L = \int \mathbf{g}(\mathbf{t}) \mathbf{X}(\mathbf{t}) dt$, with $\int \mathbf{R}_n(\mathbf{t}, \mathbf{\tau}) \mathbf{g}(\mathbf{\tau}) d\mathbf{\tau} = \mathbf{s}(\mathbf{t})$</td>
<td>$H_1$</td>
</tr>
</tbody>
</table>

$\mathbf{s}(\mathbf{t})$ is the noise autocorrelation function

<table>
<thead>
<tr>
<th>Multiple Channel</th>
<th>Sufficient Statistic</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{X}(\mathbf{t}) = \mathbf{\sum} \mathbf{X}_i \mathbf{\Phi}_i(\mathbf{t})$</td>
<td>$L = \frac{1}{\mathbf{\lambda}_0} \sum_i \mathbf{s}_i \mathbf{\Phi}_i(\mathbf{t})$</td>
<td>$H_0$</td>
</tr>
<tr>
<td>$\mathbf{X}(\mathbf{t}) = \mathbf{\Phi}_0(\mathbf{t})$</td>
<td>$L = \int \mathbf{g}(\mathbf{t}) \mathbf{X}(\mathbf{t}) dt$, with $\int \mathbf{R}_n(\mathbf{t}, \mathbf{\tau}) \mathbf{g}(\mathbf{\tau}) d\mathbf{\tau} = \mathbf{s}(\mathbf{t})$</td>
<td>$H_1$</td>
</tr>
</tbody>
</table>

$\mathbf{s}(\mathbf{t})$ is the noise autocorrelation function

Array of receivers
Using the remarks, and notations of section II, and letting:

\[ \sqrt{\mathcal{C}} \mathcal{X} = \int_{\mathcal{E}} \Phi_{\mathcal{C}}(\mathcal{X}) \, d\mathcal{E} ; \quad \mathcal{S} = \int_{\mathcal{T}} \Phi_{\mathcal{S}}(\mathcal{T}) \cdot \mathcal{S}(\mathcal{T}) \, d\mathcal{T} \]

we will still have formula (10-4) valid:

\[ \mathcal{L} = \sum_{\mathcal{Y}} \mathcal{X} \cdot \frac{\mathcal{S}}{\mathcal{A}} \]

then defining:

\[ \mathcal{g}(\mathcal{T}) \quad \mathcal{h}(\mathcal{E}) \]

\[ \mathcal{S}(\mathcal{T}) = \int_{\mathcal{E}} \mathcal{R}(\mathcal{T}, \mathcal{E}) \cdot \mathcal{g}(\mathcal{E}) \, d\mathcal{E} \]

we will still have for the detection statistic the very simple expression:

\[ \mathcal{L} = \int_{\mathcal{T}} \mathcal{g}(\mathcal{T}) \cdot \mathcal{X}(\mathcal{T}) \, d\mathcal{T} \]  

(10-6)

Let us note that a similar result is immediately obtained for the "discrete" case. We have collected the results about the sufficient statistics for different cases in Table 3; and this illustrates the usefulness of the structure seen in section II.

2. Detection of Gaussian signals in Gaussian noise is also approached by the use of the K.L. But the problem is here more difficult since, in general, we will have two different orthogonal systems of coordinates. The whole problem is intimately connected with that of simultaneous diagonalization of two covariance kernels. cf. [47], [48] and [23].

The theory is too lengthy to be reviewed meaningfully here. However, we will mention one extension of the K.L. expansion of an entirely different structure from the ones we have studied in section II. Let us therefore consider two second-order
stochastic processes, a signal $s(t)$ and a noise $n(t)$ of given autocorrelation functions $R_s(t,\tau)$ and $R_n(t,\tau)$ respectively. Let us consider the integral equation

$$2_k \int_0^\tau R_n(t,\tau) \phi_k(\tau) d\tau = \int_0^\tau R_s(t,\tau) \phi_k(\tau) d\tau$$

which reduces to the usual K.L. eigenfunction integral equation when one of the processes, e.g. $n(t)$, is white. Then we can write [61]

$$s(t) = \sum_{k=1}^\infty s_k \theta_k(t), \quad t \in [0,T]$$

$$n(t) = \sum_{k=1}^\infty n_k \theta_k(t) + \xi(t), \quad t \in [0,T]$$

where

$$\theta_k(t) = \int_0^\tau R_n(t,\tau) \phi_k(\tau) d\tau$$

The $\{\theta_k(t)\}$ form a complete system in the signal space and moreover, with the usual definitions of $s_k$ and $n_k$ as Fourier coefficients,

$$E\{s_k s_j\} = \delta_{k,j}$$

$$E\{n_k n_j\} = \delta_{k,j}$$

$$E\{\xi_k \xi_j\} = 0$$

The above expansion has been found very useful in many problems, in particular in the determination of the mutual amount of information between two segments of continuous processes in the same time interval.
B. Estimation: Let us consider the following system: a signal \( a(t) \) which is to be estimated is sent through a linear channel \( h(t,\tau) \), then modulated by a zero-memory filter \( s(b,t) \). The modulated signal then passes through another linear channel \( h_2(t,\tau) \) whose output \( x(t) \) is corrupted by additive Gaussian noise \( n(t) \) of auto-correlation function \( R_n(t,\tau) \).

![Diagram](image.png)

From the observation of \( y(t) = x(t) + n(t) \), and assuming the signal \( a(t) \) to be taken from a set with Gaussian distribution of autocorrelation \( R_a(t,\tau) \) and that \( n(t) \) is statistically independent from \( a(t) \), we want to determine the maximum a posteriori probability (M.A.P.) estimate \( \hat{a}(t) \) over the observation interval \([0, T]\).

Since we are basing our decision on the observed output sample \( y(t) \) -- which is therefore fixed -- we can write

\[
P(a|y) = \frac{p(y|a)p(a)}{p(y)} = \frac{p(x+n|a)p(a)}{p(y)}
\]

\[
= A \, p(n) \, p(a)
\]

(10-7)

where \( A \) is a constant,

since \( n(t) \) is independent of \( a(t) \).

To take into account the dispersive character of the two linear channels, we suppose that for:

- \( a(t) \), \( t \in [-T_1, T_1] \)
- \( s(b,t) \), \( t \in [-T_2, T_2] \)
- \( n(t) \) and \( x(t) \), \( t \in [0, T] \)
Assuming \[ E \left\{ a_i(t) \right\} = E \left\{ n_i(t) \right\} = 0 \], and defining

\[ \lambda_i \] and \[ \mu_i \] by the integral equations

\[ \lambda_i \phi_i(t) = \int_{\bar{T}} R_\alpha(t, \tau) \phi_i(\tau) d\tau; \mu_i \psi_i(t) = \int_{\bar{T}} R_n(t, \tau) \psi_i(\tau) d\tau \]

then

\[ a_i(t) = \sum_k a_i \phi_i(t) \]
\[ n_i(t) = \sum_k n_i \psi_i(t) \]

and we get the classical form of result,

\[ p(a/y) = \lim_{K \to \infty} \frac{C}{(2\pi)^K} e^{-\sum_{k=1}^{K} \left( \frac{a_i \lambda_i}{\lambda_i} + \frac{n_i \mu_i}{\mu_i} \right)} \]

(10-8)

where

\[ a_i = (a_i \phi_i) = \int_{-\bar{T}}^{\bar{T}} a_i(t) \phi_i(t) dt \]

(10-9)
\[ n_i = (n_i \psi_i) = \int_{0}^{\bar{T}} n_i(t) \psi_i(t) dt \]

(10-10)

The problem of estimation of a waveform \( a(t) \) is reduced to estimating the \( \{a_i\}, i = 1, 2, ..., K \), then making \( K \to \infty \). Though at first the equation giving \( p(a/y) \) seems to carry very little meaning \( a_i, K \to \infty, p(a/y) \to 0 \) as expected), the estimation method at heart is one of comparison and we can compare sometimes, "indefinite" quantities by studying their ratio. The M.A.P. estimate of \( a_i(t) \) obeys the condition that the exponential coefficient in (10-8) is maximized, i.e.
\[
\dot{\alpha}_i = -\lambda i \sum_{j=1}^{K} \frac{\eta_j}{\mu_j} \frac{\partial \eta_j}{\partial \alpha_i}, \quad i = 1, 2, \ldots, K
\]  
(10-11)

But

\[
\eta_j = \int_{0}^{\tau} \frac{y(t) - x(t)}{2} \Psi_j(t) \, dt
\]  
(10-12)

\[
X(t) = \int_{-\tau}^{\tau} f_k(t, \tau) s \left[ -b(\tau), \tau \right] \, d\tau
\]  
(10-13)

\[
\frac{\partial x(t)}{\partial \alpha_i} = \int_{-\tau}^{\tau} f_k(t, \tau) \frac{\partial s_i}{\partial \alpha_i} \frac{\partial b}{\partial \alpha_i} \, d\tau
\]  
(10-14)

or

\[
\frac{\partial x(t)}{\partial \alpha_i} = \int_{-\tau}^{\tau} f_k(t, \tau) \frac{\partial s_i}{\partial \alpha_i} \left[ \int_{-\tau}^{\tau} f_k(\tau, \nu) \Phi_i(\nu) \, d\nu \right]
\]  
(10-15)

But

\[
\frac{\partial b(\cdot)}{\partial \alpha_i} = \frac{\partial}{\partial \alpha_i} \int_{-\tau}^{\tau} h(\tau, \nu) a(\nu) \, d\nu
\]  
(10-16)

\[
\frac{\partial}{\partial \alpha_i} \int_{-\tau}^{\tau} h(\tau, \nu) \sum_{k} a_k \lambda_k(\nu) \, d\nu = \int_{-\tau}^{\tau} h(\tau, \nu) \Phi_i(\nu) \, d\nu
\]  
(10-17)

Then

\[
\frac{\partial \eta_j}{\partial \alpha_i} = -\int_{0}^{\tau} \frac{d\eta_j}{dt} \left[ \int_{-\tau}^{\tau} f_k(t, \tau) \frac{\partial s_i}{\partial \alpha_i} \frac{\partial b}{\partial \alpha_i} \right] \int_{-\tau}^{\tau} f_k(\tau, \nu) \Phi_i(\nu) \, d\nu
\]  
(10-18)
It is proved [22] that

$$\hat{\alpha}_n(t) = \sum_{\lambda=1}^{\infty} \hat{\alpha}_{\lambda} \varphi_{\lambda}(t)$$  \hspace{1cm} (10-19)

Then

$$\hat{\alpha}(t) = \sum_{\lambda=1}^{\infty} \eta_{\lambda} \sum_{\nu=1}^{\infty} \eta_{\nu} \int_{0}^{T} dt' \psi(t') \int_{0}^{T} \partial_{t'} \varphi_{\nu}(t') \varphi_{\lambda}(t') \partial_{b}^{2} \int_{-T_i}^{T} dw \varphi_{\lambda}(w) \varphi_{\nu}(w)$$  \hspace{1cm} (10-20)

Permuting summations and integrations and putting into evidence Mercer's expansion, we get:

$$\hat{\alpha}(t) = \int_{0}^{T} dt' g(t') \int_{0}^{T} \partial_{t'} \varphi_{\nu}(t') \varphi_{\lambda}(t') \partial_{b}^{2} \int_{-T_i}^{T} dw \varphi_{\lambda}(w) \varphi_{\nu}(w)$$  \hspace{1cm} (10-21)

with $g(t)$ given by the first kind integral equation

$$g(t) = \int_{0}^{T} \varphi_{\lambda}(t) \varphi_{\nu}(t) dt$$  \hspace{1cm} (10-22)

(where the kernel $\varphi_{\lambda}(t,s)$ is assumed to be strictly positive-definite, so as to insure the existence of a solution for $g(t)$).

since

$$\left(\hat{\alpha}_{\lambda} - \hat{\alpha}_{\nu}\right) \left(t'\right) = \hat{\alpha}_{\lambda} \left(t'\right) = \sum_{\lambda} \hat{\alpha}_{\lambda} \varphi_{\lambda}(t)$$

and

$$\varphi_{\lambda}(t) = \sum_{\nu} \lambda_{\nu} \varphi_{\nu}(t) \psi_{\nu}(t)$$
The last two integral equations (10-21) and (10-22) give the M.A.P. estimate for the signal $a(t)$. It is clear that the solution is usually hopelessly difficult, except for a few cases like the linear modulation one. Let us note that this estimate is optimum for the total interval of observation, not for a specific time $t$. The M.A.P. estimation problem for the vector case needs a straightforward application of the vector K.L. expansion. Because there seems to be no particularly new phenomenon we will not go into the matter further.
CHAPTER IV
SUMMARY AND AREAS FOR FURTHER RESEARCH

In this research it was found that the use of basic concepts of measure theory brought up a unification in the theory of the K.L. expansion. The unity of structure has in turn led to the extension of the optimality properties to all forms of the K.L. These properties and the orthogonality of the terms of the K.L. expansion are at the heart of our interest in the problem. A study of some particular forms of the K.L. and their applications, in particular the discrete one, is carried in some detail.

A formula is found which generalized to non-stationary processes a technique for deriving canonical expansions from corresponding integral canonical expansions.

The K.L. of a two-dimensional process is applied to the linear time-varying random channel to yield a model that is compared to the usual tapped delay line one; and found to be theoretically more satisfactory.

Application of the generalized K.L. to classical problems yields some results like the detection statistic for complex systems of receivers. It was felt that there are connected with the present research very interesting areas for investigation. For example:

a) What are other applications of the generalized K.L. to physical problems? In what cases do these generalized K.L. simplify as far as calculation are concerned?
b) What use could be made of theorem 4.2, in particular in problems connected with sequential ratio testing? Moreover is not its proof as given so far somewhat artificial?

c) The problem of actual determination of the eigenfunctions of covariance kernels remains wide open. Probably the state space approach requires more investigation. The conjecture that the eigenfunctions associated with rational spectra are sinusoids is false in general; but it is interesting to study system artically the cases where it is verified.

d) More experimental and theoretical work could be done to find out how well the discrete K.L. and Hotelling representation approximate the continuous K.L. The K.L. gives a bound for efficiency of representation; then, how well do other representations, and in particular the canonical ones, come close to it?

e) The channel model that has been proposed needs more elaboration. It is to be notices that it is still suboptimal, and the deterministic signals have to be represented more efficiently.
Theorem: If \((\nu_n, \nu_n; \lambda_n)\) is a singular system of the non-null \(L^2\) kernel \(K(s,t)\), then

\[
K(s,t) = \lim_{N \to \infty} \sum_{n=1}^{N} \lambda_n \nu_n(s) \nu_n(t)
\]

a.e.

Proof. We follow the one given by F. Smithies [8, p. 148], except that we consider here the general case of finite-measure spaces.

The Fourier coefficients of \(K(s,t)\) considered as a function of \(t\), for fixed \(s\), are:

\[
\lambda_n \nu_n(t) = \int_{\Omega_x} K(s,t) \nu_n (t) d\mu_x (t)
\]  

(A-1-1)

in view of the definition of a singular system associated with \(K(s,t)\) given in Chapter I, Section II. (For a thorough background, one may refer to Chapter VIII of the above quoted book.) Then, applying Bessel's inequality to \(K(s,t)\), we obtain:

\[
\sum_{n=1}^{N} |\lambda_n \nu_n(s)|^2 \leq \int_{\Omega_x} |K(s,t)|^2 d\mu_x (t) < \infty
\]  

(A-1-2)

Integrating with respect of \(s\), we get:

\[
\sum_{n=1}^{N} \lambda_n^2 \leq \| K \|_2^2
\]  

(A-1-3)

where \(\| K \|\) is the \(L^2\)-norm of \(K(s,t)\);
and this shows that the series

\[ \sum_{n=1}^{N} \lambda_n \]

is convergent. Let us denote

\[ K_n(s, t) = \sum_{n=1}^{N} \lambda_n u_n(s) v_n(t) \]  

\[(A-1.1b)\]

If \( N < M \), we have

\[ \| K_N - K_M \|_2^2 = \| \sum_{n=N+1}^{M} \lambda_n u_n(s) v_n(t) \|_2^2 \]

\[ = \int_{\Omega_1 \times \Omega_2} \sum_{n=N+1}^{M} \lambda_n^2 u_n^2(s) v_n^2(t) \, d\mu_1(s) \, d\mu_2(t) = \sum_{n=N+1}^{M} \lambda_n^2 \to 0 \]

\[ N, M \to \infty \]  

\[(A-1.5)\]

because of \((A-1.3)\). And since the \( L^2 \)-space is complete, the sequence \( \{K_n(s, t)\} \) converges in mean-square to an \( L^2 \) kernel \( K_0(s, t) \). If \( x \) is an arbitrary \( L \)-function, we can write

\[ \| K_N x - K x \| = \| \sum_{n=1}^{N} (x, u_n) u_n - K x \| \]

\[ = \| \sum_{n=1}^{N} \lambda_n (x, v_n) u_n - K x \| \to 0 \]

\[ N \to \infty \]  

\[(A-1.6)\]

by Hilbert-Schmidt theorem (cf. Smithies [2, p. 147] or Tricomi [39, p. 110]).

But by the definition of \( K \) given above and Schwarz inequality, we also have

\[ \| K_N x - K_0 x \| \leq \| K_N - K_0 \| \cdot \| x \| \to 0 \]

\[ N \to \infty \]  

\[(A-1.7)\]
From this it results that

\[
\| K_x - K_0 x \| = \| K_x - K_N x + K_N x - K_0 x \| \\
\leq \| K_x - K_N x \| + \| K_0 x - K_N x \| \to 0 \quad N \to \infty
\]

therefore

\[
\| K_x - K_0 x \| = 0 \quad (A-1-8)
\]

for all \( x \in L^2 \)

But (A-18) implies that

\[
\| K - K_0 \| = 0 \quad (A-1-9)
\]

since it holds for all \( x \).
APPENDIX A-2

Derivation of Mercer's theorem from Schmidt's expansion theorem.

Theorem: Given the continuous L -covariance kernel \( R(t,s) \), then the series

\[
\sum_{\lambda_i} \lambda_i \phi_i(t) \phi_i(s)
\]

where

\[
\lambda_i \phi_i(t) = \int_{\Omega_i} R(t, s) \phi_i(s) d\mu(s)
\]

and \( \Omega_i \) is a finite-measure, compact space, converges absolutely and uniformly, and the bilinear formula

\[
R(t, s) = \sum_{\lambda_i} \lambda_i \phi_i(t) \phi_i(s)
\]

holds.

Proof. Here we are taking \( R(t,s) \) to be real, but the results are still valid if it is only Hermitian. Then \( R(t,s) \), being a covariance kernel, is non-negative definite; and we have in particular (cf.[39], p. 125),

\[
R(t, t) > 0, \ t \in \Omega_i
\]
Also, \( R(t, s) \) being continuous, the eigenfunctions \( \phi_i(t) \) are also continuous.

Therefore

\[
\mathcal{R}(t,s) - \sum_{i=1}^{N} \lambda_i \phi_i(t) \phi_i(s) = R_{N+1}(t,s)
\]

is continuous, and non-negative definite which implies among other results, that

\[
R_{N+1}(t,t) = R(t,t) - \sum_{i=1}^{N} \lambda_i \phi_i^2(t) > 0
\]

Therefore, the series with positive terms \( \sum_{i=1}^{\infty} \lambda_i \phi_i^2(t) \) converges. Applying Schwarz inequality to partial sums in (A-2-1), one gets

\[
\left| \sum_{i=N+1}^{N+h} \lambda_i \phi_i(t) \phi_i(s) \right|^2 \leq \sum_{i=N+1}^{N+h} \lambda_i \phi_i^2(t) \sum_{i=N+1}^{N+h} \lambda_i \phi_i^2(s)
\]

Letting \( s = s_0 \), and given an arbitrarily small number \( \epsilon \), we can find a corresponding integer \( N \) such that, the left member of (A-2-7) is less than \( \epsilon \). This is so because,

\[
\sum_{i=N+1}^{N+h} \lambda_i \phi_i^2(t) \leq \sum_{i=1}^{\infty} \lambda_i \phi_i^2(t) \leq R(t,t) \leq M
\]

where \( M \) is the maximum of \( R(t,t) \) for \( t \in \Omega \), and because (A-2-6) implies that there exists a number \( N_0(s_0, \epsilon) \) such that

\[
\sum_{i=N+1}^{N+h} \lambda_i \phi_i^2(s_0) < \frac{\epsilon}{M} , \quad \text{for} \quad N \geq N_0
\]
This means that the series $\sum_{i,} \lambda_i \phi_i(t) \phi_i(s)$ converges absolutely and uniformly in $t$ given $s$; and for a given $t$, absolutely and uniformly in $s$. Therefore, the bilinear formula (A-2-3) holds. And in particular we are allowed to write:

$$R(t, l) = \sum_{i=1}^{\infty} \lambda_i \phi_i^2(t), \; t \in \Omega_t$$  \hspace{1cm} (A-2-10)

But by Dini's [57] theorem, the convergence in (A-2-10) is uniform. Therefore, (A-2-7) implies that the convergence in (A-2-3) is uniform even when $t_*$ and $s$ vary simultaneously.
APPENDIX A-3

Proof of lemma 2.1.

"If $x(t)$ is continuous in q.m., where $t$ belongs to the compact
finite-measure space $\Omega$, then the series

$$\sum_{i} \mu_i x_i \phi_i(t)$$

converges uniformly in $t$ to $x(t)$, i.e.

$$E \left[ (x(t) - \sum_{i=1}^{N} \mu_i x_i \phi_i(t))^2 \right] \rightarrow 0$$

uniformly in $t$.

Proof. As we already know that the convergence in the mean in (A-3-2)
is valid and that the $\mu_i$'s and $\phi_i(t)$'s are the eigenvalues and eigenfunctions of
$R(t,s) = E \{ x(t) x(s) \}$, to prove the uniform convergence
in $t$ of (A-3-2) it suffices to notice that

$$E \left[ (x(t) - \sum_{i=1}^{N} \mu_i x_i \phi_i(t))^2 \right] = E \left[ x^2(t) - \sum_{i=1}^{N} \mu_i x_i \phi_i(t) x_i(t) \right]$$

$$+ \sum_{i,j=1}^{N} \mu_i \mu_j x_i x_j \phi_i(t) \phi_j(t)$$

$$R(t,t) - 2 \sum_{i=1}^{N} \mu_i \phi_i(t) \int_{\Omega} R(t,s) \phi_i(s) d\mu(s) + \sum_{i=1}^{N} \mu_i^2 \phi_i^2(t)$$

$$= R(t,t) - \sum_{i=1}^{N} \mu_i^2 \phi_i^2(t)$$

(A-3-3)
and that \((A-3-3)\) converges uniformly to zero by the result proved in A-II, which establishes the Lemma.
Theorem: Let \((a_n, v_n, \lambda_n)\) be a singular system for the \(\mathcal{L}^2\)-kernel \(K(s,t)\), and let
\[
K_N(t,s) = \sum_{n=1}^{N} \lambda_n a_n(s) v_n(t)
\]

If \(a_1(s), \ldots, a_N(s); b_1(t), \ldots, b_N(t)\) are arbitrary \(\mathcal{L}^2\)-functions, then
\[
\| K(s,t) - \sum_{n=1}^{N} a_n b_n(t) \|^2 \geq \| K(s,t) - K_N(s,t) \|^2 = \sum_{n=N+1}^{\infty} \lambda_n^2
\]

(A-h-1)

Proof. Again we follow the one given in reference ([8], p.149), except that we are considering general finite-measure spaces. Let us note that, since
\[
\| K \|^2 = \int_{\Omega_1 \times \Omega_2} | \lambda_n a_n(s) v_n(t) |^2 d\mu_1(s) d\mu_2(t) = \sum_{n=1}^{\infty} \lambda_n^2
\]

(A-h-2)

it is sufficient to prove that
\[
\| K - \sum_{n=1}^{N} a_n b_n \|^2 \geq \| K \|^2 - \sum_{n=1}^{N} \lambda_n^2
\]

We may suppose that the set \((b_1, b_2, \ldots, b_N)\) is orthonormal, for it is always possible to transform \(\sum_{n=1}^{N} a_n b_n\) by the Gram-Schmidt process so that
\[
\sum_{n=1}^{N} a_n' b_n' = \sum_{n=1}^{N} a_n b_n
\]
where the set \((b'_1, b'_2, ... b'_N)\) is now orthonormal. We then have

\[
\| K - \sum_{n=1}^{N} a_n b_n \| ^2 = \left( K - \sum_{n=1}^{N} a_n b_n, K - \sum_{n=1}^{N} a_n b_n \right)
\]

\[
= \| K \| ^2 - \sum_{n=1}^{N} (a_n, Kb_m) - \sum_{n=1}^{N} (Kb_n, a_n) + \sum_{n=1}^{N} (a_n, a_n)(b_n, b_m)
\]

\[
= \| K \| ^2 - \sum_{n=1}^{N} (a_n, Kb_n) - \sum_{n=1}^{N} (Kb_n, a_n) + \sum_{n=1}^{N} (a_n, a_n)
\]

\[(A-4.3)\]

Since \((b_n, b_m) = \delta_{nm}\).

This expression is itself equal to

\[
\| K \| ^2 + \sum_{n=1}^{N} (a_n - Kb_n, a_n - Kb_n) - \sum_{n=1}^{N} (Kb_n, Kb_n)
\]

\[
\geq \| K \| ^2 - \sum_{n=1}^{N} \| Kb_n \| ^2
\]

\[(A-4.4)\]

By \((A-4.2)\) and \((A-4.4)\), it is therefore sufficient to prove that

\[
\sum_{n=1}^{N} \| Kb_n \| ^2 \leq \sum_{n=1}^{N} \lambda_n^2
\]

But \[\sum_{n=1}^{N} \| Kb_n \| ^2 = \sum_{n=1}^{N} (K^*K b_n, b_n)\]

where \(K^*(s,t) = \overline{K(t,s)}\) i.e.
$K^*$ is the adjoint of operator $K$. Now $K^*K$ is itself Hermitian, since 

$$(K^*K)^* = K^*(K^*)^* = K^*K;$$

and we can apply to it Ky Fan's theorem (cf. [8], p. 134), which says that the supremum of $\sum_{n=1}^{N} (K^*K b_n, b_n)$ is equal to the sum $\xi_n$ of the first $N$ positive eigenvalues of $K^*K$ arranged in the decreasing order. Therefore

$$\sum_{n=1}^{N} \|Kb_n\|^{2} \leq \sum_{n=1}^{N} \lambda_n^{2}$$

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$$

(A-4-5)

and this completes the proof.
Lemma. If two probability distributions \( \lambda_i, \ i = 1, 2, \ldots \) and \( \mu_j, \ j = 1, 2, \ldots \) are such that
\[
\lambda_i \geq \lambda_{i+1}, \quad \mu_j \geq \mu_{j+1}, \quad \text{for all } i \text{ and all } j
\]
(A-5-1)

and
\[
\sum_{i=1}^{N} \lambda_i \geq \sum_{j=1}^{N} \mu_j
\]
(A-5-2)

for every integer \( N \geq 0 \), then
\[
- \sum_{i=1}^{\infty} \lambda_i \ln \lambda_i \leq - \sum_{j=1}^{\infty} \mu_j \ln \mu_j
\]
(A-5-3)

with the equality holding in (A-5-3) if, and only if,
\[
\sum_{i=1}^{N} \lambda_i = \sum_{j=1}^{N} \mu_j
\]
(A-5-4)

for every integer \( N \geq 0 \).

Proof. We follow here the one given in [31, p. 635-669]. Let us consider the entropy function \( H \) defined by the probability distribution \( \lambda_i, \ i = 1, 2, \ldots \)
\[
H = - \sum_{i=1}^{\infty} \lambda_i \ln \lambda_i
\]
(A-5-5)
with
\[ \sum_{i=1}^{\infty} \lambda_i = 1 \]  
(A-5-6)

and
\[ \lambda_i \geq \lambda_{i+1} \]  
(A-5-7)

Let us make the change of variables
\[ \sigma_k = \sum_{i=1}^{k} \lambda_i \]  
(A-5-8)

Then \( H \) is now expressed as
\[ H = \sum_{k=1}^{\infty} (\sigma_k - \sigma_{k-1}) \ln (\sigma_k - \sigma_{k-1}) \]  
(A-5-9)

with the constraints being
\[ \sigma_0 = 0, \quad \lim_{k \to \infty} \sigma_k = 1 \]  
(A-5-10)

and
\[ \sigma_{k+1} - \sigma_k \geq 0 \quad (k = 1, 2, \ldots) \]  
(A-5-11)

Moreover, condition (A-5-7) is rewritten as
\[ \sigma_{k+1} = 2 \sigma_k + \sigma_{k-1} \leq 0 \quad \text{for all } k \geq 1 \]  
(A-5-12)

A difficulty here is that \( H \) is a function of an infinite number of variables \( \sigma_k \). Let us remark that \( H \) is well defined and that the series
\[ H_N = -\sum_{k=1}^{N} (\sigma_k - \sigma_{k-1}) \ln (\sigma_k - \sigma_{k-1}) = -\sum_{i=1}^{N} \lambda_i \ln \lambda_i \]
is monotonous non-decreasing and positive. We then have

\[ H = H_N + \epsilon_N, \quad \epsilon_N \to 0 \quad \text{as} \quad N \to \infty \]

and

\[ \frac{\partial H}{\partial \sigma_k} = \frac{\partial H^N}{\partial \sigma_k}, \quad k = 1, 2, \ldots (N-1). \]

\[ = \ln \frac{\sigma_{k+1} - \sigma_k}{\sigma_k - \sigma_{k-1}} \quad (A-5-13) \]

Therefore if we are given the two sets \( \{ \sigma_k \} \) and \( \{ \sigma_k' \} \) corresponding to the probability distributions \( \{ p_i \} \) and \( \{ p_i' \} \) such that

\[ \sigma_k > \sigma_k', \quad k = 1, 2, \ldots \quad (A-5-14) \]

we can conclude that

\[ H(\{ \sigma_k \}) \leq H(\{ \sigma_k' \}) \quad (A-5-15) \]

where we have indicated by \( H(\{ \sigma_k \}) \) the entropy for the set of variables \( (\sigma_k) \).
Theorem. Let \( \tilde{X}_L(t) \) be a finite-dimensional approximation to the second-order Gaussian random function \( \chi(t) \), i.e.

\[
\tilde{X}_L(t) = \sum_{\ell=1}^{L} \tilde{X}_\ell \phi_\ell(t) \quad 0 \leq t \leq T
\]

where the \( \phi_\ell(t) \), \( \ell = 1, 2, \ldots \), are a complete orthonormal system and

\[
\tilde{X}_\ell = \int_0^T \chi(t) \phi_\ell(t) \, dt
\]

Let \( \tilde{R}_o \) and \( \tilde{R}_L \) be defined as

\[
\tilde{R}_o = \int_0^T \chi^2(t) \, dt
\]

\[
\tilde{R}_L = \tilde{R}_o - \sum_{\ell=1}^{L} \tilde{X}_\ell^2 = \int_0^T \left[ \chi(t) - \sum_{\ell=1}^{L} \tilde{X}_\ell \phi_\ell(t) \right]^2 \, dt
\]

Then, among all expansions of \( \chi(t) \) of the form (A-6-1), the K. L. representation is the one that minimizes the average number of coefficients \( E \{ L \} \) such that, for a given threshold level \( \varepsilon^2 \), we have:

\[
\tilde{R}_{L-1} > \varepsilon^2, \quad \tilde{R}_L \leq \varepsilon^2 \quad (A-6-h)
\]
Proof. Let $\Psi_k(t)$ be the set of (orthonormalized) eigenfunctions of the integral equation.

$$\lambda_k \Psi_k(t) = \int_0^T R(t, \tau) \Psi_k(\tau) \, d\tau, \quad \lambda \geq \lambda_{k+1} \tag{A-6-5}$$

Since the process is of second-order, we can write

$$E \{ \mathcal{R}_0 \} = \sum_{k=1}^{\infty} \lambda_k < \infty$$

and this in turn implies that [50, p. 178]

$$\lambda_k \rightarrow 0 \quad \lambda \rightarrow \infty \tag{A-6-6}$$

(a) Let us first require from the system $\phi_k(t)$ to be such that, $M$ be an arbitrary positive integer,

$$\left\{ \phi_1, \phi_2, \ldots, \phi_M \right\} = \text{orthonormal basis for the space} \left\{ \Psi_1, \Psi_2, \ldots, \Psi_M \right\}$$

$$\phi_k = \Psi_k, \quad k = M+1, M+2 \tag{A-6-7}$$

for all $t \in [0, T]$.

We will first find the choice of the $\phi_k$, $k = 1, \ldots, M$ that minimizes $E \{ L_M^2 \}$ for $M$ given, then will let $M$ to become infinite.

Let

$$E \{ L_M^2 \} = J_0 (\phi_1, \ldots, \phi_M) \tag{A-6-8}$$
represent the average value of $L$ under the constraint (A-6-7). We have to evaluate the probability

$$P \left\{ R_n \leq \varepsilon^2, R_{n-1} > \varepsilon^2 \right\} = P \left\{ R_n \leq \varepsilon^2 \right\} - P \left\{ R_{n-1} \leq \varepsilon^2 \right\}$$

$$= P \left\{ R_{n-1} > \varepsilon^2 \right\} - P \left\{ R_n > \varepsilon^2 \right\}$$

(A-6-9)

since it is needed for the calculation of

$$E_{\mu_1} \left( \mu_1 \right) = \lim_{N \to \infty} \sum_{n=1}^{N} n \left[ P \left\{ R_n \leq \varepsilon^2, R_{n-1} > \varepsilon^2 \right\} \right]$$

$$= \lim_{N \to \infty} \sum_{n=1}^{N} n \left( P \left\{ R_{n-1} > \varepsilon^2 \right\} - P \left\{ R_n > \varepsilon^2 \right\} \right)$$

$$= \lim_{N \to \infty} \left[ \sum_{n=0}^{N-1} P \left\{ R_n > \varepsilon^2 \right\} - NP \left\{ R_n > \varepsilon^2 \right\} \right]$$

(A-6-10)

(b) We need to find a bound for $P \left\{ R_n > \varepsilon^2 \right\}$, $n \geq M$, in order to make (A-6-10) more explicit. We will see the Chernoff bound [32, p. 77]. The positive number $\lambda$ being for the time being arbitrary, we can write therefore

$$P \left\{ R_n > \varepsilon^2 \right\} \leq E \left\{ \exp \left[ \lambda \left( R_n - \varepsilon^2 \right) \right] \right\} = e^{-\lambda \varepsilon^2} \exp \left\{ \lambda \lambda \varepsilon^2 \right\}$$

$$= e^{-\lambda \varepsilon^2} \exp \left\{ \lambda \sum_{k=n+1}^{\infty} (X_k) \right\}$$

$$= e^{-\lambda \varepsilon^2} \prod_{k=n+1}^{\infty} E \left\{ \exp \left[ \lambda (X_k)^2 \right] \right\}$$

$$= e^{-\lambda \varepsilon^2} \prod_{k=n+1}^{\infty} \left( 1 - 2 \lambda \lambda k \right)^{-\frac{1}{2}}$$

(A-6-11)
which holds for all $\lambda$ such that

$$0 < \lambda < \frac{1}{2\lambda_{n+1}}, \quad (A-6-12)$$

(The last equality could be seen derived in [51] for instance.)

Taking the logarithm of both sides in (A-6-11) and setting $\lambda = \frac{1}{6\lambda_{n+1}}$, we obtain

$$\log \mathbb{P} \{ R_n > \varepsilon^2 \} \leq - \frac{\varepsilon^2}{6\lambda_{n+1}} - \frac{1}{2} \sum_{k=n+1}^{\infty} \lambda_k \left[ 1 - \frac{\lambda_k}{3\lambda_{n+1}} \right] \quad (A-6-13)$$

Let us make use of the inequality

$$\ln (1-x) < 2x \quad \text{for} \quad 0 \leq x \leq \frac{1}{3}.$$

Then (A-6-13) becomes

$$\ln \mathbb{P} \{ R_n > \varepsilon^2 \} \leq - \frac{\varepsilon^2}{6\lambda_{n+1}} + \frac{1}{3\lambda_{n+1}} \sum_{k=n+1}^{\infty} \lambda_k \quad (A-6-14)$$

Since $\sum_{k=1}^{\infty} \lambda_k < \infty$, there exists, for any value of $\varepsilon^2$, a finite integer $N_0 = N_0(\varepsilon^2)$ such that

$$\sum_{k=1}^{N_0} \lambda_k < \frac{\varepsilon^2}{4}.$$

Therefore for any $n \geq N_0$, we can write

$$\ln \mathbb{P} \{ R_n > \varepsilon^2 \} \leq - \frac{\varepsilon^2}{12\lambda_{n+1}}$$

or

$$\mathbb{P} \{ R_n > \varepsilon^2 \} \leq \exp \left\{ - \frac{\varepsilon^2}{12\lambda_{n+1}} \right\} \quad (A-6-16)$$
Since \( n \lambda_n \to 0 \) as \( n \to \infty \), then the limit of the second term \( NP \{ R_N > \varepsilon \} \) in (A-6-10) is 0. Then we get

\[
E_{\mathcal{M}} \{ L \} = \lim_{N \to \infty} \sum_{n=0}^{N} \mathbb{P} \{ R_n > \varepsilon \} = \sum_{n=1}^{M-1} \mathbb{P} \{ R_n > \varepsilon \} + \lim_{N \to \infty} \sum_{n=M}^{N} \mathbb{P} \{ R_n > \varepsilon \}.
\]

(A-6-17)

But if we take \( M > N_0 (\varepsilon) \), the second term in (A-6-17) can be bounded by

\[
\sum_{n=M}^{\infty} -\varepsilon, 1 \leq \varepsilon / 12 \lambda_n, 1
\]

which goes to zero as \( M \) tends to infinity.

(c) Finally let us consider how to choose the set \( \Phi_1, \Phi_2, \ldots, \Phi_M \) so as to minimize the first term in (A-6-17). If we let

\[
R'_n = \sum_{k=n+1}^{M} (X_k)^2,
\]

(A-6-18)

then

\[
R_n = R'_n + R_{pl}
\]

(A-6-19)

since

\[
R_n = R_o - \sum_{k=1}^{n} (X_k)^2, \quad R_{pl} = R_o - \sum_{k=1}^{M} (X_k)^2, \quad n, M.
\]
Let us also denote by \( \mathcal{R}_M (\cdot) \) the probability density for \( \mathcal{R}_M \).

We then have

\[
\mathbb{P} \{ R_n > \varepsilon^n \} = \mathbb{P} \{ R_M > \varepsilon^M \} + \int_0^{\varepsilon^M} \mathbb{P} \{ R'_n > \varepsilon^n \} \mathcal{R}_M (\alpha) d\alpha
\]

\( (A-6-20) \)

To minimize \( \mathbb{P} \{ R_n > \varepsilon^n \} \) it is sufficient to determine a set of \( \phi'_i \) that minimize \( \mathbb{P} \{ R'_n > \varepsilon^n \} \) independently of \( \alpha \). Let us remark that the probability appearing in the integrand of \( (A-6-20) \) is described in terms of an integral in the \( M \)-dimensional space spanned by \( (\psi_1, \psi_2, \ldots, \psi_M) \). The corresponding coordinates \( \xi_1, \xi_2, \ldots, \xi_M \) of \( X(t) \) have a Gaussian distribution with the surfaces of equiprobability being ellipsoids whose axes are oriented along the \( \psi_k \) in order of decreasing magnitude.

\[
\frac{\xi_1^2}{\lambda_1} + \frac{\xi_2^2}{\lambda_2} + \ldots + \frac{\xi_M^2}{\lambda_M} = \text{Constant}.
\]

In this space the \( \phi'_1, \ldots, \phi'_M \) form a second orthonormal set with corresponding coordinates \( X_1, X_2, \ldots, X_M \) for \( X(t) \). Let us denote the cylinder (volume) of cross-sectional dimension \( (M-n) \) and radius \( (\varepsilon^n - \alpha) \) oriented parallel to the \( \phi'_1, \ldots, \phi'_n \) axes by

\[
A (\varepsilon^n - \alpha, \phi'_1, \ldots, \phi'_n) = \left\{ \chi: \sum_{m=1}^n (X_m)^2 \leq \varepsilon^n - \alpha \right\}
\]

\( (A-6-21) \)

It is therefore clear that
We wish to maximize the probability
\[ P \left\{ \xi \in A \left( \epsilon^2 - \alpha; \phi_1, \ldots, \phi_n \right) \right\} \]
over all sets of \( \phi'_s \). We are going to show that this is accomplished if
\[ \phi'_k = \psi_k, \quad K = 1, e, \ldots, n. \]

Let us also denote by \( A_\cap \) the common volume to the two "cylinders"
\[ A_\cap = A \left( \epsilon^2 - \alpha; \psi_1, \ldots, \psi_n \right) \cap A \left( \epsilon^2 - \alpha; \phi_1, \ldots, \phi_n \right) \]
and
\[ A'_\left( \epsilon^2 - \alpha; \psi_1, \ldots, \psi_n \right) = A \left( \epsilon^2 - \alpha; \psi_1, \ldots, \psi_n \right) - A_\cap \]
\[ A'_\left( \epsilon^2 - \alpha; \phi_1, \ldots, \phi_n \right) = A \left( \epsilon^2 - \alpha; \phi_1, \ldots, \phi_n \right) - A_\cap \]
Then
\[
\mathbb{P}\{ \xi \in A(\varepsilon^2 - \alpha; \psi_1, \psi_2, \ldots, \psi_n) \} = \mathbb{P}\{ \xi \in A(\varepsilon^2 - \alpha; \phi_1, \ldots, \phi_n) \}
\]
\[
= \mathbb{P}\{ \xi \in A'(\varepsilon^2 - \alpha; \psi_1, \ldots, \psi_n) \} = \mathbb{P}\{ \xi \in A'(\varepsilon^2 - \alpha; \phi_1, \ldots, \phi_n) \}
\]
\[\tag{A-6-27}\]

But this term is positive, since a volume element in \( A'(\varepsilon^2 - \alpha; \psi_1, \ldots, \psi_n) \), is congruent to a volume element in \( A(\varepsilon^2 - \alpha; \psi_1, \ldots, \psi_n) \), which lies within an equiprobability density ellipsoid of higher value. This is true independently of the value of \( \varepsilon^2 - \alpha \) and the index \( n, n = 0, \ldots, M \). From (A-6-20) it appears clear that the choice

\[
\phi_k = \psi_k, \quad k = 1, \ldots, M
\]
\[\tag{A-6-28}\]
minimizes the value \( \mathbb{P}\{ R \geq \varepsilon^2 \} \) for any value of \( \varepsilon^2 \), and for all \( n = 0, \ldots, M \). And this choice minimizes the first term in the expression (A-6-17) for \( E_M(\varepsilon^2) \). As \( M \to \infty \), the second term in (A-6-17) goes to 0 and we then have the solution to the problem

\[
\phi_k(t) = \psi_k(t) \quad 0 \leq t \leq T, \quad k = 1, \ldots
\]
\[\tag{A-6-29}\]

(d) This was the proof given in [52]. We can make some straightforward extensions of the theorem. We remark first that, if the \( \chi^2_l \) are not given by (A-6-2) the theorem is still valid as the Fourier coefficients insure most rapid convergence. If the given set \( \phi_k \) is not
orthogonal, we can always reduce the expansion into an orthogonal one and the theorem would still apply. Finally, we remark that the theorem does apply for the generalized expansions defined in section II.
APPENDIX A-7

Differentiation of the K.L. expansion

Theorem: (Kadota [5]). Let $X(t)$ be a separable and measurable stochastic process with zero-mean and covariance $R(t,s)$. If 
\[ \left( \frac{\partial^2}{\partial t^n \partial s^n} \right) R(t,s) \]
exists and is continuous and if $X^{(n)}(t)$, the $n^{th}$ almost sure derivative of $X(t)$ exists, then

\[ X^{(n)}(t) = \sum \mu_i \chi_i \phi_i^{(n)}(t) \quad t \in \Omega \]  
(A-7-1)

where the $\phi_i(t)$ are the eigenfunctions of $R(t,s)$, $\phi_i^{(n)}(t)$ their $n^{th}$ derivative, and

\[ \mu_i \chi_i = \int_{\Omega} X(t) \phi_i(t) d\mu(t) \]  
(almost surely)

and the series in (A-7-1) converges in the stochastic mean, uniformly in $t$.

Proof. We need to use the following result due to Kadota [49].

Lemma. With the same assumptions about $R(t,s)$ and its derivative

\[ \frac{\partial^2}{\partial t^n \partial s^n} \frac{R(t,s)}{R(t,s)} \]
then it is possible to write

\[ \frac{\partial^2}{\partial t^n \partial s^n} R(t,s) = \sum \lambda_i \phi_i^{(n)}(t) \phi_i^{(n)}(s) \]

where the series converges uniformly in $t$ and $s$. 
Then: \((\infty)\) Since \(\frac{d^{n}}{dt^{n}} R(t,s)\) exists, it is shown that the \(n^{th}\) stochastic mean derivative of \(X(t)\) exists. Let us denote it by \(\hat{X}'(t)\). Also \(E\left\{\hat{X}'(t) \hat{X}'(s)\right\} \leq E\left\{|X'(t)|^2\right\} < \infty\) as by Schwarz's inequality. If we are given that the \(n^{th}\) almost sure derivative exists, then:

\[X^{(n)}(t) = \hat{X}^{(n)}(t)\]  
(almost everywhere (A-7-2))

\[E\left\{\hat{X}^{(n)}(t) \hat{X}^{(n)}(s)\right\} = E\left\{X^{(n)}(t) X^{(n)}(s)\right\}\]  
(A-7-3)

We can also write, letting \(n = 1\),

\[E\left\{\hat{X}'(t) \hat{X}'(s)\right\} = E\left\{\lim_{\Delta t \to 0} \frac{X(t+\Delta t) - X(t)}{\Delta t} \cdot \lim_{\Delta s \to 0} \frac{X(s+\Delta s) - X(s)}{\Delta s}\right\}\]

\[= \lim_{\Delta t \to 0} \lim_{\Delta s \to 0} E\left\{\frac{X(t+\Delta t) - X(t)}{\Delta t} \cdot \frac{X(s+\Delta s) - X(s)}{\Delta s}\right\}\]

\[= \lim_{\Delta t \to 0} \lim_{\Delta s \to 0} \frac{1}{\Delta t} \cdot \frac{1}{\Delta s} \left[\frac{R(t+\Delta t, s+\Delta s) - R(t, s)}{\Delta t} - \frac{R(t, s+\Delta s) - R(t, s)}{\Delta s}\right]\]

\[= \frac{d^2}{ds \, dt} R(s, t)\]  
(A-7-4)

These equations are legitimate because the stochastic mean derivative exists, and because weak convergence is implied by strong convergence [3]. Letting now \(n = k + 1\), and assuming we have established (A-7-4) for \(n = k\), the same reasoning leads to equality

\[E\left\{\hat{X}^{(k)}(t) \hat{X}^{(k)}(s)\right\} = \frac{\delta^{(k)}}{\partial t^{k} \partial s^{k}} R(t,s)\]  
(A-7-5)
and, by induction, we establish (A-7-5) to be true for all n.

\( \beta \) To establish that in (A-7-1) the convergence is uniform in \( t \), we note that

\[
\lim_{I \to \infty} \mathbb{E} \left\{ \left| X^{(n)}(t) - \sum_{i=1}^{I} \mu_i \chi_i \phi_i^{(n)}(t) \right|^2 \right\} = \\
\lim_{I \to \infty} \mathbb{E} \left\{ \left| X^{(n)}(t) - \sum_{i=1}^{I} \mu_i \phi_i^{(n)}(t) \chi_i \chi_i^* \right|^2 \right\} + \sum_{i=1}^{I} \sum_{j=1}^{I} \phi_i^{(n)}(t) \phi_j^{(n)}(t) \chi_i \chi_j^* \mu_i \mu_j \}
\]

\[
= \frac{\partial^2}{\partial t^2} R(t,s) \bigg|_{t=s} - 2 \sum_{i=1}^{I} \mu_i \phi_i^{(n)}(t) \mathbb{E} \{ \chi_i X^{(n)}(t) \} \\
+ \sum_{i=1}^{I} \mu_i^2 \phi_i^2(t)
\]

But \( \mathbb{E} \{ \chi_i X^{(n)}(t) \} = \mathbb{E} \left\{ \int_{\Omega} \chi_i X(s) \phi_i(s) \, ds \right\} = \mu_i \), i.e.

\[
\lim_{I \to \infty} \left[ \frac{\partial^2}{\partial t^2} R(t,s) \bigg|_{t=s} - \sum_{i=1}^{I} \mu_i^2 \phi_i^2(t) \right] = 0
\]

(A-7-6)

by the lemma, and since \( \lambda_i = \mu_i^2 \).
APPENDIX A-8

Let \( \hat{\xi}_x(t) \) be defined by the integral

\[
\hat{\xi}_x(t) = \int_{T_0}^{T_0} R_x(t, \tau) C^\tau(\tau) \hat{\Phi}_x(\tau) \, d\tau
\]  

(A-8-1)

where the notations are the same as in section VIII.

Then we can write [22]

\[
R_x(t, \tau) = \begin{cases} 
\Theta(t, \tau) R_x(\tau, \tau) & \tau > t \\
R_x(t, \tau) \Theta(\tau, t) & \tau < t 
\end{cases}
\]  

(A-8-2)

where \( \Theta(t, \tau) \) is the transition matrix associated with \( \Phi(t) \), i.e.

\[
\begin{cases} 
\frac{\partial}{\partial t} \Theta(t, \tau) = \Phi(t) \Theta(t, \tau) \\
\Theta(\tau, \tau) = I
\end{cases}
\]  

(A-8-3)

Using (A-8-2), equation (A-8-1) will be expressed as

\[
\hat{\xi}_x(t) = \int_{T_0}^{T_0} \Theta(t, \tau) R_x(\tau, \tau) C^\tau(\tau) \hat{\Phi}_x(\tau) \, d\tau + R_x(t, t) \int_{T_0}^{T_0} \Theta(\tau, t) C^\tau(\tau) \hat{\Phi}_x(\tau) \, d\tau
\]  

(A-8-4)

Differentiating with respect to \( t \), we get
\[
\frac{d \tilde{\xi}}{dt} = \int_{T_0}^t \frac{d \Theta(t, \tau)}{dt} \mathcal{R}_x(t, \tau) \mathcal{C}_T^\tau(\tau) \mathcal{P}^\tau_c(\tau) \, d\tau
\]
\[
+ \frac{d \mathcal{R}_x(t, t)}{dt} \int_t^T \Theta(\tau, t) \mathcal{C}_T^\tau(\tau) \mathcal{P}^\tau_c(\tau) \, d\tau
\]
\[
+ \mathcal{R}_x(t, t) \int_t^T \frac{d \Theta(\tau, t)}{dt} \mathcal{C}_T^\tau(\tau) \mathcal{P}^\tau_c(\tau) \, d\tau
\]
(A-8-5)

After using (A-8-3) and the fact that
\[
\frac{d}{dt} \Theta^\tau(\tau, t) = -\mathcal{F}^\tau(t) \Theta^\tau(\tau, t)
\]
(A-8-6)

we obtain
\[
\frac{d \tilde{\xi}}{dt} = \mathcal{F}^\tau(t) \int_{T_0}^t \Theta(t, \tau) \mathcal{R}_x(t, \tau) \mathcal{C}_T^\tau(\tau) \mathcal{P}^\tau_c(\tau) \, d\tau
\]
\[
+ \left[ \frac{d \mathcal{R}_x(t, t)}{dt} - \mathcal{R}_x(t, t) \mathcal{F}^\tau(t) \right] \int_t^T \Theta(\tau, t) \mathcal{C}_T^\tau(\tau) \mathcal{P}^\tau_c(\tau) \, d\tau
\]
(A-8-7)

But it is shown [22] that
\[
\frac{d}{dt} \mathcal{R}_x(t, t) = \mathcal{F}^\tau(t) \mathcal{R}_x(t, t) + \mathcal{R}_x(t, t) \mathcal{F}^\tau(t) + \mathcal{Q} \mathcal{Q} \mathcal{C}_T^\tau(t)
\]
(A-8-8)

with the notations of section VIII. Therefore (A-8-8) applied to (A-8-7) yields:
\[
\frac{d \tilde{\xi}}{dt} = \mathcal{F}^\tau(t) \int_{T_0}^t \Theta(t, \tau) \mathcal{R}_x(t, \tau) \mathcal{C}_T^\tau(\tau) \mathcal{P}^\tau_c(\tau) \, d\tau
\]
\[
+ \left[ \mathcal{F}^\tau(t) \mathcal{R}_x(t, t) + \mathcal{Q} \mathcal{Q} \mathcal{C}_T^\tau(t) \right] \int_t^T \Theta(\tau, t) \mathcal{C}_T^\tau(\tau) \mathcal{P}^\tau_c(\tau) \, d\tau
\]
(A-8-9)
Using now (A-8-4) we get

\[
\frac{d\bar{\xi}_i}{dt} = \bar{p}_i(t) \bar{\xi}_i + \bar{X}(s) \mathcal{G} \mathcal{O}^\eta(t) \int_{T_0}^{T_0} \mathcal{G}(\tau, t) \mathcal{O}^\eta(\tau) \mathcal{G} \mathcal{O}^\eta(t) \bar{\phi}_i(d\tau)
\]  

(A-8-10)

We would have reduced the problem to the solution of a differential equation system if we were able to get rid of the integral contained in (A-8-10). To this effect, let us define

\[
\bar{\eta}_i(t) = \int_{T_0}^{t} \mathcal{G}(\tau, \tau) \mathcal{G} \mathcal{O}^\eta(t) \bar{\phi}_i(d\tau), \quad T_0 \leq t \leq T_p
\]  

(A-8-11)

Then (A-8-10) becomes

\[
\frac{d\bar{\xi}_i}{dt} = \bar{p}_i(t) \bar{\xi}_i + \bar{X}(s) \mathcal{G} \mathcal{O}^\eta(t) \bar{\eta}_i(t)
\]  

(A-8-12)

Differentiating equation (A-8-11), we get:

\[
\frac{d\bar{\eta}_i(t)}{dt} = - \mathcal{G}(\tau, \tau) \mathcal{G} \mathcal{O}^\eta(t) \bar{\phi}_i(t) + \int_{T_0}^{T_0} \mathcal{G}(\tau, \tau) \mathcal{G} \mathcal{O}^\eta(t) \bar{\phi}_i(d\tau)
\]

\[
= - \mathcal{G}(\tau, \tau) \bar{\phi}_i(t) - \mathcal{G}(\tau, \tau) \bar{\eta}_i(t)
\]  

(A-8-13)

Therefore we have the set of simultaneous differential equations (A-8-12) and (A-8-13) for which we need to find the associated boundary conditions.

Let us remark that \(\bar{\phi}_i(t), i = 1, 2, \ldots\), must be bounded at the end points \(t = T_0\) and \(t = T_p\). Therefore, letting \(t = T\) in (A-8-11),
we obtain

\[ \hat{\eta}_i (t_f) = 0 \]  

(A-8-14)

If in (A-8-14) we let \( t = T_\circ \), the first integral vanishes and we are left with:

\[ \hat{\xi} (t_\circ) = R_x (T_\circ, T_0) \int_{T_0}^{T_\circ} \Theta^{\gamma} (\tau, t) \xi_{\gamma}(\tau) \Phi_i (\tau) \, \text{d}\tau \]

\[ = R_x (T_0, T_0) \hat{\eta} (T_0) = P_0 \hat{\eta} (T_0) \]  

(A-8-15)

And (A-8-14) and (A-8-15) constitute the set of boundary conditions associated with the differential equations (A-8-12) and (A-8-13), and these are exactly equations (8-21) - (8-24) of section VIII.
BIBLIOGRAPHY


