On the Optimality of the Karhunen-Loève Expansion

**INTRODUCTION**

Consider a zero-mean random process $X(t)$ that is continuous-in-time on the interval $0 \leq t \leq T$. It has a Karhunen-Loève expansion

$$X(t) = \frac{1}{N} \sum_{1}^{N} \xi_{k} \phi_{k}(t) \quad 0 \leq t \leq T$$

(1)

in which the $\phi_{k}(t)$ are the eigenfunctions of the integral equation

$$\lambda_{k} \phi_{k}(t) = \int_{0}^{T} R_{x}(t, s) \phi_{k}(s) \, ds \quad 0 \leq t \leq T.$$  

(2)

The $\phi_{k}(t)$ are assumed to be normalized (and orthogonalized if there are repeated eigenvalues) and arranged in order of decreasing eigenvalues. Then the random coefficients $\xi_{k}$ have the property

$$E[\xi_{k} \xi_{l}] = \delta_{k+l}.$$  

(3)

Now consider a finite dimensional approximation to $X(t)$

$$X_{L}(t) = \sum_{1}^{L} X_{k} \phi_{k}(t) \quad 0 \leq t \leq T,$$

(4)

in which the $\phi_{k}(t)$, $k = 1, 2, \ldots$, are a complete orthonormal (CON) set and the $X_{k}$ are the corresponding Fourier coefficients of the process. Let the random variable $R_{L}$ be defined by

$$R_{L} = \int_{0}^{T} X^{2}(t) \, dt$$

(5)

and let the random variables $R_{l}$, $l = 1, 2, \ldots$, be the integral-square truncation error in the expansion of (4)

$$R_{l} = R_{L} - \sum_{1}^{l} X_{k}^{2} = \int_{0}^{T} \left[ X(t) - \sum_{1}^{l} X_{k} \phi_{k}(t) \right]^{2} \, dt.$$  

(6)

It is well known [1], [2] that for any fixed $L$ the mean truncation error, $E[R_{L}]$, is minimized if the $\phi_{k}$ of eq. (4) are chosen to be the first $M$ of the Karhunen-Loève expansion [1], [2]. Note that this solution depends only on the correlation function $R_{x}(t, s)$.

Rather than fixing $L$ and seeking to minimize $E[R_{L}]$, one might instead pose the following question. Suppose it is required that the integral-square truncation error always be less than or equal to some threshold level $\epsilon$, and we seek to minimize the number of coefficients required. Specifically, let $L$ be the random variable such that $L$ is the value of $l$ for which

$$R_{l-1} > \epsilon^2, \quad R_{l} \leq \epsilon^2.$$  

(7)

We then wish to choose the $\phi_{k}$ to be the CON set that minimizes $E[L]$. The set of $\phi_{k}$ that is optimum in this sense clearly depends in a detailed manner on the probability distribution of the process; in particular, it is not determined just by the correlation function. However, since the distribution of a Gaussian process depends only on its correlation function, one would conjecture that for a Gaussian process the Karhunen-Loève expansion is again valid under the criterion of minimizing $E[L]$ for a given threshold error $\epsilon$. The substance of this correspondence is to prove this conjecture under the assumption that the process $X(t)$ has finite power

$$E[R_{0}] = \sum_{k=1}^{\infty} \lambda_{k} < \infty.$$  

(8)

We note for future reference that since the $\lambda_{k}$ are monotonic this implies [3] that the $\lambda_{k}$ are $0(k^{-1})$; that is,

$$\lambda_{k} \leq C/k, \quad C < \infty.$$  

(9)

We now turn directly to the proof.

**Proof of the Optimality Property:** To begin, let us constrain our choice of the functions $\phi_{k}$. For an arbitrary choice of $M$ we allow the first $M \phi_{k}$ to be any orthonormal basis for the space spanned by the first $M \phi_{k}$; the remaining $\phi_{k}$ we set equal to $\phi_{k}$.

$$\phi_{k}(t) = \psi_{k}(t) \quad 0 \leq t \leq T$$

(10)

Let

$$E_{M}[L] = \frac{1}{N} \sum_{N=0}^{N} \frac{n P[R_{N} \leq \epsilon^2, R_{N-1} > \epsilon^2]}{1}$$

(11)

where

$$E_{M}[L] = \frac{1}{N} \sum_{N=0}^{N} \frac{n P[R_{N} \leq \epsilon^2, R_{N-1} > \epsilon^2]}{1}$$

(12)

At this point we need to take an aside from the main argument and develop a bound for $P[R_{N} > \epsilon^2]$, $n \geq M$. Using the Chernoff bound, we have

$$P[R_{N} > \epsilon^2] \leq E \{ \exp \{ \lambda (X_{N} - \epsilon^2) \} \}$$

(13)

$$= e^{-\lambda \epsilon^2} E \{ \exp \{ \lambda X_{N} \} \}$$

(14)

$$= e^{-\lambda \epsilon^2} \sum_{k=1}^{\infty} E \{ \exp \{ \lambda X_{k} \} \}$$

(15)

$$= e^{-\lambda \epsilon^2} \prod_{k=1}^{\infty} (1 - 2\lambda \lambda)$$

(16)

which holds for all $\lambda$ such that

$$0 < \lambda < 1/2\lambda_{k}.$$  

(17)
Taking the log of both sides of (11) and setting $\lambda = 1/3\alpha_{n+1}$ yields

$$\log P[R_n > \varepsilon^1] \leq -\frac{\varepsilon^2}{6\lambda_{n+1}} - \frac{1}{2} \sum_{k=1}^{n} \log \left[1 - \frac{1}{3\lambda_{n+1}}\right].$$

(13)

Now, for $0 < x < 1/3$, $-\log (1 - x) < 2x$. Thus, (13) becomes

$$\log P[R_n > \varepsilon^1] \leq -\frac{\varepsilon^2}{6\lambda_{n+1}} + \frac{1}{3\lambda_{n+1}} \sum_{k=1}^{n} \lambda_k.$$

(14)

The sequence $\lambda_k$ is summable; thus, for any value of $\varepsilon^1$ there exists a finite $N_0 = N_0(\varepsilon^1)$ such that

$$\sum_{k=N_0}^{\infty} \lambda_k < \frac{\varepsilon^2}{4}.$$

Then, for any $n \geq N_0$ we have

$$\log P[R_n > \varepsilon^1] \leq -\frac{\varepsilon^2}{12\lambda_{n+1}},$$

or

$$P[R_n > \varepsilon^1] \leq \exp \{-\varepsilon^2/(12\lambda_{n+1})\}.$$

(16)

We now make use of inequality (16) as follows. First, we note that since the sequence $\lambda_k$ is $0(n^{-1})$, then the limit of the second term in (10) is 0, so that

$$E_M[L] = \lim_{N \to \infty} \sum_{n=0}^{N} P[R_n > \varepsilon^1]$$

$$= \sum_{k=1}^{n} P[R_k > \varepsilon^1] \leq \lim_{N \to \infty} \sum_{n=0}^{N} P[R_n > \varepsilon^1].$$

(17)

Next, we note from inequality (16) that if $M > N_0(\varepsilon^2)$, the second term in (17) can be bounded by

$$\sum_{n=M}^{\infty} \exp \{-\varepsilon^2/(12\lambda_{n+1})\}$$

which, for $\lambda_n \sim 0(n^{-1})$, can be made as small as desired by making $M$ sufficiently large. We thus finally consider how to choose the $\phi_0, \ldots, \phi_M$ to minimize the first term in (17).

Let us set

$$R'_{n} = \sum_{k=1}^{n} (X_k)^2,$$

(18)

so that

$$R_n = R'_n + R_M.$$  

(19)

Then, letting $f_{RM}(\cdot)$ denote the probability density for $R_M$, we have

$$P[R_n > \varepsilon^1] = P[R_M > \varepsilon^1]$$

$$+ \int_{0}^{\varepsilon^1} P[R'_n > \varepsilon^1 - \alpha] f_{RM}(\alpha) \, d\alpha.$$  

(20)

To minimize $P[R_n > \varepsilon^1]$ it is sufficient to determine a set of $\phi$ that minimize $P[R'_n > \varepsilon^1 - \alpha]$ independently of $\alpha$.

The probability appearing in the integral of this expression is described in terms of an integral in the $M$-dimensional space spanned by the orthonormal basis $\phi_1, \ldots, \phi_M$. The corresponding coordinates, $\xi_1, \ldots, \xi_M$, of the process $X(t)$ have a Gaussian distribution with the surfaces of equiprobability density being ellipsoids whose axes are oriented along the $\phi_1$ in order of decreasing magnitude. In this space the $\phi_0, \ldots, \phi_M$ form a second orthonormal set with the corresponding coordinates of the process $X(t)$ being $\xi_0, \ldots, \xi_M$. The cylinder of cross-sectional dimension $M - n$ and radius $\varepsilon^1 - \alpha$ oriented parallel to the $\phi_0, \ldots, \phi_n$ axes, we denote by