Reprinted by permission IEEE TRANSACTIONS ON INFORMATION THEORY
Vol. IT-19, No. 4, July 1973, pp. 561–564

On the Approximation of Optimal Realizable Linear Filters Using a Karhunen–Loève Expansion

T. E. FORTMANN AND B. D. O. ANDERSON

***Abstract*—The Karhunen–Loève expansion of a random process is used to derive the impulse response of the optimal *realizable* linear estimator for the process. The expansion is truncated to yield an approximate state-variable model of the process in terms of the first N eigenvalues**

Manuscript received July 26, 1971; revised January 30, 1973. This work was supported in part by the Australian Research Grants Committee.
The authors are with the Department of Electrical Engineering, University of Newcastle, New South Wales 2308, Australia.

and eigenfunctions. The Kalman–Bucy filter for this model provides an approximate realizable linear estimator which approaches the optimal one as $N \to \infty$. A bound on the truncation error is obtained.

I. Introduction

The problem to be considered here is a classic one: find the minimum mean-square realizable linear estimate of a random signal corrupted by additive white noise. The received data have the form

$$z(t) = y(t) + n(t), \qquad T_i \leq t \leq T_f \tag{1}$$

where the signal y has a known mean (assumed zero for convenience) and *covariance function*

$$R(t,\tau) \triangleq E[y(t)y(\tau)], \qquad T_i \leq t, \tau \leq T_f. \tag{2}$$

The noise n is white with (two-sided) spectral height N_0 and is uncorrelated with y. A *realizable* linear filter with impulse response h is required such that the estimate

$$\hat{y}(t) = \int_{T_i}^{t} h(t,\tau)z(\tau)\, d\tau \tag{3}$$

minimizes the mean-square error $E[y(t) - \hat{y}(t)]^2$ for each t in the interval $T_i \leq t \leq T_f$.

It is well known [1, sec. 6.1] that the required h satisfies the integral equation

$$N_0 h(t,\tau) + \int_{T_i}^{t} h(t,\sigma)R(\tau,\sigma)\, d\sigma = R(t,\tau), \quad T_i \leq \tau < t \leq T_f \tag{4}$$

and the minimum mean-square error is

$$E[y(t) - \hat{y}(t)]^2 = N_0 h(t,t), \qquad T_i \leq t \leq T_f. \tag{5}$$

Solving (4) can be quite difficult in general, particularly because of the realizability constraint. If the process y is stationary and $T_i = -\infty$, then the solution may be found using spectral factorization with the computations being greatly simplified if R has a rational Fourier transform.

Another approach to the estimation problem when T_i and T_f are finite is to use the fact that if the covariance function R has eigenvalues $\{\lambda_i, i = 1,2,\cdots,\infty\}$ and (orthonormal) eigenfunctions $\{\phi_i, i = 1,2,\cdots,\infty\}$, then it may be written as

$$R(t,\tau) = \sum_{i=1}^{\infty} \lambda_i \phi_i(t)\phi_i(\tau), \qquad T_i \leq t, \tau \leq T_f \tag{6}$$

and the signal process y may be expressed as

$$y(t) = \sum_{i=1}^{\infty} y_i \phi_i(t), \qquad T_i \leq t \leq T_f. \tag{7}$$

The random variables y_i are defined by

$$y_i \triangleq \int_{T_i}^{T_f} y(\tau)\phi_i(\tau)\, d\tau, \qquad i = 1,2,\cdots \tag{8}$$

and are uncorrelated

$$E[y_i y_j] = \lambda_i \delta_{ij}, \qquad i,j = 1,2,\cdots. \tag{9}$$

This characterization of the signal process, known as a *Karhunen–Loève expansion*, may be used [1, p. 198] to derive easily an expression for the impulse response of the optimal *unrealizable*[1] linear filter,

$$h_U(t,\tau) = \sum_{i=1}^{\infty} \frac{\lambda_i}{\lambda_i + N_0} \phi_i(t)\phi_i(\tau) \tag{10}$$

whose mean-square error is

$$E[y(t) - \hat{y}_U(t)]^2 = N_0 h_U(t,t) = N_0 \sum_{i=1}^{\infty} \frac{\lambda_i}{\lambda_i + N_0} \phi_i^2(t). \tag{11}$$

The objective here is to show how one can derive the optimal *realizable* (*i.e.*, *causal*) linear filter from the Karhunen–Loève expansion and then obtain a suboptimal approximation to it by truncation. To the authors' knowledge, most previous derivations (see, e.g., [1]) using eigenvalues and eigenfunctions have led to unrealizable filters.

II. Optimal Filter and Process Model

If the series (7) is truncated at N terms, the resulting process

$$y_N(t) = \sum_{i=1}^{N} y_i \phi_i(t) \tag{12}$$

approximates y and has a covariance

$$R_N(t,\tau) = \sum_{i=1}^{N} \lambda_i \phi_i(t)\phi_i(\tau) \tag{13}$$

that approximates R. The work of Swerling [5] suggests that one can causally estimate the finite set of coefficients $\{y_i, i = 1,2,\cdots,N\}$ with a set of readily derivable estimators. These can be found as the solutions of a set of integral equations that involve the *finite-rank* kernel R_N, and hence are easily explicitly solved.

In fact, one can even bypass estimation of the individual y_i, for the solution of (4) with R replaced by R_N and h by h_N is given by

$$h_N(t,\tau) = N_0^{-1}\boldsymbol{\Phi}_N'(t) \cdot \left[\boldsymbol{\Lambda}_N^{-1} + N_0^{-1}\int_{T_i}^{t} \boldsymbol{\Phi}_N(\tau)\boldsymbol{\Phi}_N'(\tau)\, d\tau\right]^{-1} \boldsymbol{\Phi}_N(\tau) \tag{14}$$

for $t > \tau$. Here $\boldsymbol{\Phi}_N(\cdot)$ is the N-vector with ith entry $\phi_i(\cdot)$, and $\boldsymbol{\Lambda}_N$ is a diagonal matrix with ith diagonal entry λ_i. The formula (14) is obtainable from standard results involving finite-rank kernels of integral equation theory, see, e.g., [6].

However, we shall now present a derivation of (14) based on the Kalman–Bucy filter which will probably appeal more to engineers on pedagogical grounds. A state-variable model for the truncated signal process (12) is provided by the time-varying N-dimensional linear system

$$\begin{aligned} \dot{\mathbf{x}}_N(t) &= \mathbf{0} \\ y_N(t) &= \boldsymbol{\Phi}_N'(t)\mathbf{x}_N(t) \\ z_N(t) &= y_N(t) + n(t) \end{aligned} \tag{15}$$

where the initial state $\mathbf{x}_N(T_i)$ is a zero-mean random variable with

$$E[\mathbf{x}_N(T_i)\mathbf{x}_N'(T_i)] = \boldsymbol{\Lambda}_N. \tag{16}$$

It is easy to verify that the covariance of y_N in this model is (13), as required.

The optimal realizable linear estimate $\hat{y}_N(t)$ of $y_N(t)$ may be expressed as the output of a Kalman–Bucy filter [1], [2]

$$\begin{aligned} \dot{\hat{\mathbf{x}}}_N &= \mathbf{0}\hat{\mathbf{x}}_N + N_0^{-1}\mathbf{P}_N(t)\boldsymbol{\Phi}_N(t)[z_n(t) - \hat{y}_N(t)] \\ \hat{\mathbf{x}}_N(T_i) &= \mathbf{0} \\ \hat{y}_N(t) &= \boldsymbol{\Phi}_N'(t)\hat{\mathbf{x}}_N(t) \end{aligned} \tag{17}$$

where $\mathbf{P}_N$ satisfies the Riccati equation

$$\begin{aligned} \dot{\mathbf{P}}_N(t) &= -N_0^{-1}\mathbf{P}_N(t)\boldsymbol{\Phi}_N(t)\boldsymbol{\Phi}_N'(t)\mathbf{P}_N(t) \\ \mathbf{P}_N(T_i) &= \boldsymbol{\Lambda}_N. \end{aligned} \tag{18}$$

[1] This means that the upper limit on the integral in (3) is replaced by T_f and $h(t,\tau) \neq 0$ for $\tau > t$.

The variance of the estimation error is

$$E[y_N(t) - \hat{y}_N(t)]^2 = \Phi_N'(t)E[\tilde{x}_N(t)\tilde{x}_N'(t)]\Phi_N(t) = \Phi_N'(t)P_N(t)\Phi_N(t) \quad (19)$$

where $\tilde{x}_N = x_N - \hat{x}_N$.

By forming the differential equation and boundary condition for P_N^{-1}, one obtains easily an explicit formula for P_N^{-1}, leading to

$$P_N(t) = \left[\Lambda_N^{-1} + N_0^{-1}\int_{T_i}^t \Phi_N(\tau)\Phi_N'(\tau)\,d\tau\right]^{-1} \quad (20)$$

and

$$P_N(T_f) = [\Lambda_N^{-1} + N_0^{-1}I]^{-1} = \operatorname{diag}\left[\frac{\lambda_i}{1 + \lambda_i/N_0}\right] \quad (21)$$

because the ϕ_i are orthonormal.

Next, from (17) and (18) one obtains easily

$$\frac{d}{dt}[P_N^{-1}\hat{x}_N] = N_0^{-1}\Phi_N z_N. \quad (22)$$

This leads to

$$P_N^{-1}(t)\hat{x}_N(t) = N_0^{-1}\int_{T_i}^t \Phi_N(\tau)z_N(\tau)\,d\tau \quad (23)$$

so that

$$\hat{y}_N(t) = N_0^{-1}\Phi_N'(t)P_N(t)\int_{T_i}^t \Phi_N(\tau)z_N(\tau)\,d\tau \quad (24)$$

is an explicit expression for the optimal realizable estimate of $y_N(t)$.

To summarize, the optimal realizable linear estimator *for the truncated process (12)* has impulse response

$$h_N(t,\tau) = N_0^{-1}\Phi_N'(t)P_N(t)\Phi_N(\tau) \quad (25)$$

where $P_N(t)$ is given by (19), and error variance

$$E[y_N(t) - \hat{y}_N(t)]^2 = N_0 h_N(t,t) = \Phi_N'(t)P_N(t)\Phi_N(t). \quad (26)$$

Under the additional assumption that R is continuous, it is not difficult to verify that $h_N(t,\cdot)$ approaches $h(t,\cdot)$ in $L^2[T_i,T_f]$ as $N \to \infty$, this convergence being uniform for $T_i \le t \le T_f$. More specifically[2]

$$\|h(t,\cdot) - h_N(t,\cdot)\| \le [\|R(t,\cdot) - R_N(t,\cdot)\|/N_0][1 + \|R_N(t,\cdot)\|/N_0] \quad (27)$$
$$\le (\alpha_N/N_0)(1 + \beta_N/N_0) \quad (28)$$

for $T_i \le t \le T_f$, where

$$\alpha_N = \max_{T_i \le t \le T_f} \|R(t,\cdot) - R_N(t,\cdot)\| \xrightarrow[N\to\infty]{} 0 \quad (29)$$

$$\beta_N \triangleq \max_{T_i \le t \le T_f} \|R_N(t,\cdot)\| \le \max_{T_i \le t \le T_f} \|R(t,\cdot)\| \triangleq \beta. \quad (30)$$

III. Suboptimal Filter

The impulse response h_N is optimal for the truncated process y_N and approximates the optimal realizable impulse response h for the original process y. Computation of h_N, however, involves only the first N terms of the Karhunen–Loève expansion for y. Thus it makes sense to use h_N in a filter for the original signal process y, producing a *suboptimal* estimate

$$\tilde{y}_N(t) = \int_0^t h_N(t,\tau)z(\tau)\,d\tau. \quad (31)$$

The error variance incurred by using this filter is, of course, greater than that incurred by the optimal filter, and the increase may be characterized as follows.

Proposition: If R is continuous, the error variance of the suboptimal filter (31) is bounded by

$$E[y(t) - \tilde{y}_N(t)]^2 \le E[y(t) - \hat{y}(t)]^2 + (\lambda_{\max} + N_0)(\alpha_N/N_0)^2(1 + \beta_N/N_0)^2 \quad (32)$$

where $\lambda_{\max}$ is the largest eigenvalue of R, α_N and β_N were defined in (30) and (31), and $E[y - \hat{y}]^2$ is the error variance of the optimal filter (3).

The suboptimal interval error is bounded by

$$\int_{T_i}^{T_f} E[y(t) - \tilde{y}_N(t)]^2\,dt \le \int_{T_i}^{T_f} E[y(t) - \hat{y}(t)]^2\,dt + (\lambda_{\max} + N_0)(1 + \beta_N/N_0)^2 \sum_{i=N+1}^{\infty} (\lambda_i/N_0)^2. \quad (33)$$

Proof: This follows easily by expanding the expression for the mean-square error, using the well-known fact that the optimal error is uncorrelated with past data, and substituting (27) and (28).

The penalty paid for suboptimality, in terms of error variance, is bounded by the right-hand term in (32). This bound vanishes, of course, as $N \to \infty$.

IV. Discussion

The Karhunen–Loève expansion of a random signal process y corrupted by white noise has been truncated at N terms and the resulting process y_N realized by a linear system (14) with an N-dimensional state vector. The optimal realizable impulse response h_N for filtering the truncated process was found from the Kalman–Bucy filter and was seen to converge to the corresponding impulse response for the original process.

It was then proposed to use h_N, which is computable from only the first N terms of the Karhunen–Loève expansion, in a suboptimal filter for the original process y, and the additional error variance incurred with this scheme was specified by (32). From the practical point of view the results are probably not extremely useful because the computation of eigenvalues and eigenfunctions is generally difficult.

After submitting this correspondence, the authors became aware of some work of Gardner [3], [4], in which a signal process is subjected to an arbitrary "disturbance transformation" and then a suboptimal estimator is derived by truncating a series expansion of the signal process. This is done by working directly with the integral equations rather than with a state-variable approach. Realizability is obtained essentially by using "sliding gate functions", but the resulting filters would appear to be more complicated than ours in principle.

Acknowledgment

The authors are indebted to the Associate Editor, Prof. Kailath, for a number of useful suggestions, particularly the relationship to Swerling [5].

[2] Both h and h_N satisfy integral equations of the form (4), which may be manipulated to yield an expression for the difference $h - h_N$. Taking the norm and establishing a bound on the norm of the inverse of the integral operator yields (27), and (28) follows from the continuity of R.

References

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Part I. New York: Wiley, 1968.
[2] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Trans. ASME, Ser. D: J. Basic Eng.*, vol. 83, pp. 95–108, 1961.
[3] W. A. Gardner and L. E. Franks, "An alternative approach to linear least squares estimation of continuous random processes," in *Proc. 5th Annual Princeton Conf. Inform. Sci. and Systems*, Princeton, N.J., pp. 267–75, Mar. 1971.
[4] W. A. Gardner, "Representation and estimation of cyclostationary processes," Ph.D. dissertation, Univ. Mass., Amherst, Mass., Aug. 1972.
[5] P. Swerling, "Modern state estimation methods from the viewpoint of the method of least squares," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 707–19, Dec. 1971.
[6] F. Smithies, *Integral Equations*. New York: Cambridge Univ. Press, 1965.