NEUVIEME COLLOQUE SUR LE TRAITEMENT DU
SIGNAL ET SES APPLICATIONS

NICE du 16 au 20 MAI 1983

New results on robust estimation of signals
Nouveaux résultats sur l'estimation robuste des signaux

Dimitri Kazakos

Electrical Engineering Department, University of Virginia
Charlottesville, VA 22901

RESUME

Dans cette communication nous présentons quelque nouveaux resultats sur l'elaboration des filtres robustes pour la prédiction linéaire, l'interpolation et la filtrage des signaux aléatoires. On a présumé qu'il existe quelque incertitude sur leurs spectres, et l'elaboration robuste est optimale pour les spectres les moins favorables.

SUMMARY

In this paper we present some new results on the design of robust filters for linear prediction, interpolation and filtering of random signals. It is assumed that some uncertainty about their spectra exists, and the robust design optimizes for the least favorable spectra.

New results on robust estimation of signals

Nouveaux résultats sur l'estimation robuste des signaux

Introduction

This paper considers the linear prediction, filtering and interpolation of second order random, stationary processes, under the condition that their spectral structure is vaguely specified. In order for the classical theory of linear prediction and filtering of Wiener [1] and Kolmogorov [2] to be effective, it is imperative that the spectra of the processes concerned be completely known. In this paper we replace this assumption by a weaker one; namely, we assume only that a certain neighborhood of a spectrum is given and we construct optimal in a specific sense linear filters, based on the knowledge that the spectrum of the process belongs to that neighborhood. We adopt the approach of minimax design, which follows Huber's [3] pioneering work on robust estimation, and we extend the more recent work of Hosoya [4] in several directions, as will be made clear in the sequel.

Preliminaries and review of known results

In this paper we will assume that a noisy version of a multivariate stationary process is observed. Both discrete-time and continuous-time cases will be considered. Let $\{Y(t);\ t\varepsilon I_0\}$ be the observation record, consisting of the d-dimensional process $Y(t)$. The observation interval I_0 will be either the continuous or the discrete time axis. The observation will be either noisy or noiseless. In general,

$$Y(t) = X(t) + N(t); \quad t\varepsilon I_0$$

where $X(t)$ is the signal component and $N(t)$ is the additive noise. Suppose now that a linear, time invariant filter with transfer function H operates on $\{Y(t)\}$ in order to produce an estimate $\hat{X}(t)$ of $X(t)$. Then, the covariance matrix of the error $X(t) - \hat{X}(t)$ has the form: [5]

$$P_d(H, S, N) = (2\pi)^{-1} \int_{-\pi}^{\pi} \{[H(e^{i\lambda}) - I]S(\lambda)[H(e^{i\lambda}) - I]^* + H(e^{i\lambda})N(\lambda)H^*(e^{i\lambda})\}d\lambda \qquad (1a)$$

for discrete time observations, and

$$P_c(H, S, N) = (2\pi)^{-1} \int_{-\infty}^{+\infty} \{[H(i\omega) - I]\ S(i\omega)[H(i\omega) - I]^* + H(i\omega)N(i\omega)H^*(i\omega)\}d\omega \qquad (1b)$$

for continuous time observations.

S, N denote the spectral density matrices of $X(t)$, $N(t)$ correspondingly. Whenver the argument $e^{i\lambda}$ $(i\omega)$ is used, we have a discrete (continuous) time case. By H^* we denote the transpose conjugate of the matrix H. Classical filtering theory has resolved the problem of minimizing the error co covariance matrix P_c, P_d through an appropriate choice of H, assuming N, S known and fixed.

For the discrete time case, if we let $N=0$ and the observation record consists of the past only, i.e. $I_0 = \{k;\ k<t\}$ is used for estimating $S(t)$, then we have the pure prediction problem. The natural criterion for choosing the optimal H which is here constrained to be causal $(H\varepsilon C)$ is the minimization of the trace of P_d. In classical prediction theory [5] it has been established that the resulting infimum is:

$$\inf_{H\varepsilon C} \text{trace } P_d(H, S, 0) = \text{trace } P_d(H_0, S, 0) = $$
$$= \exp\{(2\pi d)^{-1} \int_{-\pi}^{\pi} \log \det[2\pi S(\lambda)]d\lambda\}$$

The optimum causal $H_0(z)$ is found through the spectral factorization of $S(\lambda)$. Using the identity $\log \det A = \text{trace} \log A$, (for any positive definite matrix A), we have that:

$$\log \det P_d(H_0, S, 0) = (2\pi d)^{-1} \int_{-\pi}^{\pi} \log \det [2\pi S(\lambda)]d\lambda \qquad (2)$$

If the right hand side of (2) is finite, then $S(\lambda)$ factorizes as: [5]

$$S(\lambda) = (2\pi)^{-1}\phi(e^{i\lambda})\phi^*(e^{i\lambda})$$

where $\phi(z)$ is holomorphic within the unit circle $|z| = 1$, with $\phi(0) = I_d$. The transfer function of the filter giving the best linear one-step predictor, is

$$H_\theta(e^{i\lambda}) = e^{i\lambda}[e^{-1}\ \phi(e^{i\lambda})]_+\phi^{-1}(e^{i\lambda})$$

where $[A(z)]_+$ means the terms of positive powers of z only, in the Laurent series expansion of $A(z)$.

Another case of interest in the discrete time framework is the interpolation or smoothing problem. If we let $N=0$, and the set $I_0 = \{k;\ k\neq t\}$ is available for linear estimation of $S(t)$, we have the interpolation problem. If the trace of $P_d(H, S, 0)$ is the optimality criterion, then the resulting optimal covariance matrix for the interpolation error, is: [5]

$$P_d(H_0, S, 0) = 4\pi^2[\int_{-\pi}^{\pi} S^{-1}(\lambda)d\lambda]^{-1} \qquad (3)$$

The optimal transfer function H_0 has the form:

$$H_0 = I_d - (2\pi)^{-1}[\int_{-\pi}^{\pi} S^{-1}(\lambda)d\lambda]S^{-1}(\lambda)$$

For the case of nonzero noise and nonrealizable filters H that act upon a doubly infinite data record, we have a classical, tractable Wiener filtering problem, for both discrete and continuous time. We will not elaborate on this case, due to space limitations. Robust solutions have been obtained for the scalar $(d=1)$ case by other authors ([6] - [8]).

For the case of nonzero noise and under the constraint of causal or realizable filter H, the problem is highly intractable in general. There is one exception, namely the case of white noise, for which analytical expressions for the minimum error are available.

For continuous time observations, white noise $N(t)$ with a constant spectral density matrix N_0, and for optimal causal filtering, the following expression for the minimum achievable error covariance matrix is available: [9]

$$\text{trace } P_d(H_0, S, N_0) \cdot N_0^{-1} =$$

$$= (2\pi)^{-1} \int_{-\infty}^{+\infty} \log|I + S(i\omega)N_0^{-1}|d\omega \quad (4)$$

The corresponding expression for discrete time observations seems to have been derived in the literature only for scalar processes, and has the form: [10]

$$P_c(H_0, S, N_0) = N_0\{1 - \exp \frac{-1}{2\pi} \int_{-\pi}^{\pi} \log(1 +$$

$$+ N_0^{-1}S(\lambda))d\lambda\} \quad (5)$$

In the present paper we address the problem of designing linear filter functions H that provide robust filtering solutions in the presence of spectral uncertainty. The problem is stated in the next section.

As a preliminary step, we observe that if we define:

$$Q(H, S) \triangleq \text{trace } P(H, S, N) =$$

$$= (2\pi)^{-1} \int \text{trace}\{HH^*(S+N) - S(H+H^*) + S\}d\lambda \quad (6)$$

for either the discrete or continuous time case, the functional Q is convex in H and linear in S. This observation will be utilized in the robust design. If Q(H, S) is the criterion, a robust design seeks to minimize Q with respect to H and maximize it with respect to S.

Robust estimation

Let us denote by Q(H, S) the generic expression for the trace of the error covariance matrix for both continuous and discrete time problems. We will assume that N is known, hence constant, and that S, H are members of compact Hausdorff spaces F , H_1 respectively. Due to the convex-concave nature of Q(H, S), there is a minimax value:

$$\min_{H\varepsilon H_1} \max_{S\varepsilon F} Q(H, S) = \max_{S\varepsilon F} \min_{H\varepsilon H_1} Q(H, S) = Q(H^*, S^*)$$

$$= \max_{S\varepsilon F} Q(H^*, S) \quad (7)$$

The existence of the minimax or saddle point solution is a consequence of a theorem due to Ky Fan [11]. Thus, to find the robust filter $H^*\varepsilon H_1$, we have to minimize Q(H, S) over $H\varepsilon H_1$ and then maximize the resulting minimum. According to the previously mentioned results, we may use the existing formulas and maximize them over $S\varepsilon F$. The maximizing value S* will provide us with the robust filter H* matched to S*.

For the purpose of maximizing the minimum mean square error expressions over classes of spectra, we will develop two theorems.

Theorem 1

Let $\{G_i(f_i);\ i=1, \ldots, d\}$ be concave and differentiable functions of f_i. Let F_1 be a family of nonnegative functions $f_i(\lambda);\ i=1, \ldots d;$ $\lambda\varepsilon I$, defined by

$$F_1 = \{f;\ i=1, \ldots, d:\ u_i(\lambda) \le f_i(\lambda) \le$$

$$\le v_i(\lambda);\ \lambda\varepsilon I;\ \int_I f_i(\lambda)d\lambda = 1;\} \quad (8)$$

where $f = [f_1 \ldots f_d]$.

Define the functional:

$$G(f) = \int_I \sum_{i=1}^{d} G_i(f_i(\lambda))d\lambda \quad (9)$$

Then G(f) is maximized over $f\varepsilon F_1$, by the function $f^\circ = [f_1^\circ\ f_2^\circ \ldots f_d^\circ]$, where

$$f_i^\circ(\lambda) = \max[u_i(\lambda), \min[c_i, v_i(\lambda)]];$$

$$i = 1, \ldots, d \quad (10)$$

and the constant c_i is uniquely determined by the requirement that $f_i^\circ(\lambda)$ integrates to 1.

(Proof is deleted due to space limiations; it will appear in a journal version of this paper.)

Theorem 2

Let $\{G_i(f_i);\ i=1, \ldots, d\}$ be concave functions of f_i. Let F_2 be a family of nonnegative functions $f(\lambda) = [f_1(\lambda), f_2(\lambda), \ldots, f_d(\lambda)];\ \lambda\varepsilon I$, defined by:

$$F_2 = \{f(\lambda);\ \lambda\varepsilon I;\ \int_{I_k} \sum_{i=1}^{d} f_i(\lambda)d\lambda = p_k;$$

$$k=1, \ldots, m\} \quad (11)$$

where $\{I_1, I_2, \ldots, I_m\}$ is a partition of I. Then, the functional

$$G(f) = \int_I \sum_{i=1}^{d} G_i(f_i(\lambda))d\lambda$$

is maximized over $f\varepsilon F_2$ by the functions:

$$f_i(\lambda) = \bar{s}_k^{-1} p_k \alpha_i \text{ for } \lambda\varepsilon I_k;\ i=1, \ldots, d;$$

$$k=1, \ldots, m \quad (12)$$

where s_k is the measure of I_k, and the nonnegative constants $\{\alpha_i\}$ are determined from the set of equations:

$$\sum_{k=1}^{m} p_k G_i'(s_k^{-1}\, p_k \alpha_i) = \lambda;\ i=1, \ldots, d;$$

$$\sum_{i=1}^{d} \alpha_i = 1$$

(The proof is deleted due to space limitations; it will appear in a journal version of this paper.)

We are now in a position to seek the least favorable spectrum S by maximizing over S the previous minimum error expressions. The minimization will be performed for S belonging to each of two distinct spectral classes, F_1, F_2, defined as follows. Let $f_1(\lambda)$, $f_2(\lambda)$, ..., $f_d(\lambda)$ be the eigenvalues of the spectral density matrix $S(\lambda)$. Then, we define F_1 to be the set (8), and F_2 to be the set (11). We note that eq. (2) that gives the optimum prediction error, can be expressed in the form of eq. (9), with $G_i(f_i) = \log(2\pi f_i)$; i=1, ..., d, which is a concave function of f_i. Thus, through direct application of theorems 1 and 2 we can immediately identify the maximizing spectra of families F_1 and F_2, as specified in the theorem statements by equations (10) and (12) respectively.

The next case under consideration is for the causal filtering problem in white noise. We assume N_0, the white noise spectrum, to be constant and known. We redefine the functions $f_1(\omega)$, $f_2(\omega)$, ..., $f_d(\omega)$ as being the eigenvalues of the "whitened" matrix $S(i\omega)N_0^{-1}$, and we use eq. (4) as the optimality criterion. It is easy to see that eq. (4) can be expressed in the form of eq. (12), with $G_i(f_i) = \log(1 + f_i)$, which is a concave function. If we consider the spectral uncertainty classes F_1, F_2, where we use the re-defined eigenvalues $f_i(\omega)$, then Theorems 1 and 2 are immediately applicable for determining the maximizing spectra, given by equations (10) and (12). A similar approach yields the maximization over $S(\lambda)$ of eq. (5), for the discrete time case.

Conclusions

In this paper we have developed robust solutions for the filtering and prediction problem when there is uncertainty about the spectra. We treated the case of multivariate stationary processes, for discrete time and continuous time situations. We developed explicit robust solutions only for the cases for which the available error expressions are in closed form.

The spectral classes we considered are motivated by realistic uncertainty conditions. The first class F_1, specifies the spectral density matrix through an upper and a lower bound to each eigenvalue. Those bounds can be viewed as confidence intervals for an available estimate of the spectral density matrix.

The second class, F_2, characterizes the spectral that are known only through the total power of each eigenvalue in a given frequency interval, I_k. Thus, F_2 can be viewed as a "total power" constraint.

Finally, we observe that our theory deals only with linear robust filters. Some statisticians believe that truly robust filters must be nonlinear. However, the introduction of nonlinearieties would make the error expressions very cumbersome. The problem is still open, challenging, and of great interest to both theoreticians and practically motivated researchers.

_____._____._____.

Research supported by the Air Force Office of Scientific Research by Grant AFOSR 82-0030.

References

[1] N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series. MIT Press, 1949.

[2] A. Kolmogorov, "Interpolation and Extrapolation," Bull. Acad. Sci., USSR, Ser. Math. 5, pp. 3-14, 1941.

[3] P. Huber, "Robust estimation of a location parameter," Ann. Math. Statis., v. 35, pp. 73-101.

[4] Y. Hosoya, "Robust linear extrapolation of second order stationary processes," Ann. of Probability, v. 6, pp. 574-584, 1978.

[5] E.J. Hannan, Multiple Time Series, New York, Wiley, 1970.

[6] L.J. Cimini and S.A. Kassam, "Robust and quantized Wiener filters for p-point spectral classes," Proceedings of the 14th Conf. on Information Sciences and Systems, Princeton Univ., Princeton, N.J., March 1980.

[7] S.A. Kassam and T.L. Lim, "Robust Wiener filters," J. Franklin Inst., vol. 304, pp. 171-185, 1977.

[8] H.V. Poor, "On robust Wiener filtering," IEEE Trans. on Aut. Control, vol. AC-25, pp. 531-536, June 1980.

[9] J. Snyders, "On the error matrix in optimal linear filtering of stationary processes," IEEE Trans. on Infor. Theory, vol. IT-19, pp. 593-599, 1973.

[10] A.J. Viterbi, "On the minimum mean square error resulting from linear filtering of stationary signals in white noise," IEEE Trans. on Infor. Theory, vol. IT-11, pp. 594-595, 1965.

[11] Ky Fan, "Minimax Theorems," Proc. of National Acad. of Science, vol. 39, pp. 42-47, 1953.