Robust Time Series Estimation

Demetrios Kazakos, Kami S. Makki
College of Science and Technology
3100 Cleburne Str,Houston, Texas, 77004, U.S.A
Department of Electrical Engineering & Computer Science, University of Toledo
2801 W. Bancroft, Toledo, Ohio 43606, U.S.A

Abstract— In this paper we present some new results on the problem of robust estimation for stationary multiple time series processes. For these processes, we consider the prediction, smoothing and causal filtering problem in cases for which the minimum achievable mean square error is expressed in a closed form in terms of the spectral density matrix of the signal. We consider three convex classes of spectral uncertainties, and develop robust solutions for these cases.

I. Introduction

The subject of robust estimation of signals has received a lot of attention in the engineering literature [1, 2, 3]. In this paper we present some new robust estimation results for multiple stationary time series. We consider robust linear prediction, filtering and interpolation of second order random processes. For the stationary, we assume that the spectral density matrices are incompletely known, and incomplete knowledge is expressed as membership to a convex uncertainty class. We then seek minimax robust linear filters, in a game theoretic sense. We determine very general conditions for existence of minimax robust solutions.

II. Stationary Processes

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

$$Y(t) = X(t) + N(t);\ t\varepsilon\, I_O \qquad (1)$$

where X(t) is the signal component and N(t) 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 the signal X(t).

Then, the covariance matrix of the error X(t) - $\hat{X}(t)$ has the form: [5]

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

for discrete time observations, and [5]

$$P_c[H,S,N] = (2\pi)^{-1}\int_{-\infty}^{\infty}\{[H(iw) - I]S(iw)[H(iw) - I]^* + H(iw)N(iw)H^*(iw)\}dw \qquad (3)$$

for continuous time observations, where S, N denote the spectral density matrices of X(t), N(t) correspondingly, and H* is the transpose conjugate of the matrix H.

Classical Wiener filtering theory [5] has resolved the problem of minimizing the error covariance matrix for both continuous and discrete time, when S and N are known. In more recent work, Chen and Kassam [6] have determined robust solutions to the multivariate robust noncausal filtering problem (2), (3).

In the preceding formalism, the data record is doubly infinite, and the mean square error expressions P_d, P_c are simpler. When the causality assumption is imposed, the data record is semi-infinite, and the solutions are more complex. Hosoya [4] determined the robust solution to the linear prediction problem for ε-contaminated spectra and scalar processes, and Vastola and Poor [7], and Franke [8], found robust solutions for more general cases.

The minimum mean square error in one step prediction based on the infinite past, for the scalar process case is [5]

$$P_1(S) = \exp\{(2\pi)^{-1}\int_{-\pi}^{\pi}\log S(\lambda)d\lambda\} \qquad (4)$$

In this paper, we present some robust estimation results for the d-dimensional process case. For a d-dimensional process, with spectral density matrix S(λ), the minimum trace of the optimum prediction error covariance matrix, $P_d(S)$, is:

$$g_d(S) = \mathrm{trace}P_d(S) = \exp\{(2\pi d)^{-1}\int_{-\pi}^{\pi}\log|S(\lambda)|d\lambda\} \qquad (5)$$

The optimum causal filter $H_0(z)$ is found through the spectral factorization of S(λ): [5]

$$S(\lambda) = (2\pi)^{-1}\Phi(e^{i\lambda})\Phi^*(e^{i\lambda}) \qquad (6)$$

where Φ (z) is holomorphic within the unit circle |z| = 1, with $\Phi(0)=I_d$. The transfer function of the filter giving the best linear-step predictor, is: [5]

$$H_O(e^{i\lambda}) = e^{i\lambda}[e^{-i\lambda}\Phi(e^{i\lambda})]_+\,\Phi^{-1}(e^{i\lambda}) \qquad (7)$$

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

The functional

$$A(S) = \int_{-\pi}^{\pi}\log\det S(\lambda)d\lambda = \int_{-\pi}^{\pi}\mathrm{tr}\log S(\lambda)\,d\lambda \qquad (8)$$

is convex. Thus, maximizing $g_d(S)$ of eq. (5) over convex sets of S, is a convex optimization problem.

Another case of interest in discrete-time is the interpolation or smoothing problem where we let N=0 and the doubly infinite set $I_O = \{k, k \neq t\}$ is available for linear estimation of S(t) in the time-discrete case, we have the interpolation problem. The trace of the error covariance matrix is the optimality criterion. The resulting optimum error covariance matrix is: [5]

$$P_{dI}(S) = 4\pi^2 [\int_{-\pi}^{\pi} S^{-1}(\lambda) d\lambda]^{-1} \quad (9)$$

The optimal transfer function H_0 has the expression:

$$H_O = I_d - (2\pi)^{-1} [\int_{-\pi}^{\pi} S^{-1}(\lambda) d\lambda] S^{-1}(\lambda) \quad (10)$$

For a positive definite matrix A, we have the identity:

log det A = trace log A.

From eq. (9), using the previous identity, we find:

$$G(S) \stackrel{\Delta}{=} \log\det\{[P_{dI}(S)]^{-1} 4\pi^2\} = \log\det \int_{-\pi}^{\pi} S^{-1}(\lambda) d\lambda =$$

$$= \mathrm{tr} \log \int_{-\pi}^{\pi} S^{-1}(\lambda) d\lambda \quad (11)$$

As a measure of the error covariance matrix for the interpolation problem we use the functional –G(S). We will show that G(S) is concave with respect to S. Let

$$S_a(\lambda) = aS_1(\lambda) + (1-a)S_2(\lambda),$$

$$H(a) = \log \int_{-\pi}^{\pi} S_a^{-1}(\lambda) d\lambda$$

The second derivative of H(a) is:

$$\ddot{H}(a) = -[\int_{-\pi}^{\pi} S_a^{-1}(\lambda) d\lambda]^{-2} [-\int_{-\pi}^{\pi} S_a^{-2}(\lambda)[S_1(\lambda) - S_2(\lambda)] d\lambda]^2 -$$

$$-[\int_{-\pi}^{\pi} S_a^{-1}(\lambda) d\lambda]^{-1} \cdot 2 \int_{-\pi}^{\pi} S_a^{-3}(\lambda)[S_1(\lambda) - S_2(\lambda)]^2 d\lambda$$

hence $-\ddot{H}$ (a) is nonnegative definite matrix, and thus

$$-\ddot{G}(S_a) = \mathrm{tr}[-\ddot{H}(a)] \geq o$$

and in conclusion G(S) is a convex functional of S, because – G(S) is concave.

For the case of nonzero additive noise N(t), and under the constraint of realizable or causal filters H, there is no closed form expression of the resulting minimum error covariance matrix in the general case, and the problem is very intractable. There is one exception: when the noise is white, some analytical expressions for the minimum error are available.

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

$$h_d(S) \stackrel{\Delta}{=} trP_{d2}[S, N_O] \cdot N_O^{-1} =$$

$$= (2\pi)^{-1} \int_{-\infty}^{+\infty} \log | I_d + S(iw) N_O^{-1} | dw. \quad (12)$$

The corresponding expression for discrete time observations seems to have been derived only for the scalar (d = 1) process case, and has the form: [10]

$$q(S) \stackrel{\Delta}{=} N_O \{1 - \exp[-1/2\pi \int_{-\pi}^{\pi} \log(1 + N_O^{-1} S(\lambda)) d\lambda]\}. \quad (13)$$

No multidimensional extension of the above result seems available.

Due to the concavity of the log function, it is very easy to demonstrate that both $h_d(S)$ of eq. (12) and q(S) of eq. (13) are concave functionals of S, a result that will be used in the sequel to characterize minimax robust solutions.

III. Robust Estimation

Let us denote by W(H, S) the generic expression for the trace or determinant of the error covariance matrix in estimating X(t) from $\{Y(t); t\varepsilon I_O\}$ as in the preceding section, where H is the transfer function of the linear estimator filter, and S the signal spectrum. We assume fixed noise spectral density. Two cases for the observation set I_O are considered: $I_{01} = \{k; k \leq t\}; I_{02} = \{k; k \neq t\}$ where X(t) is the signal to be estimated at time t, for t either discrete or continuous. Thus, either causal filtering cases or noiseless smoothing in discrete time are considered [11, 12]. The functional W(H, S) is convex in H and concave in S. Due to the convex-concave nature of W(H, S), there is a minimax value, a conclusion that follows from the general minimax theorem of Ky Fan [13] and convex optimization theory, e.g.: (Rockafellar [14]).

$$\min_{H\varepsilon H_1} \max_{S\varepsilon F} W(H,S) = \max_{S\varepsilon F} \min_{H\varepsilon H_1} W(H,S) =$$

$$= W(H^*, S^*) = \max_{S\varepsilon F} W(H^*, S) = \max_{S\varepsilon F} V(S). \quad (14)$$

where H_1 is a convex class of filters, and F is a convex uncertainty class for the spectra S. Note that the procedure for finding a robust filter H* is to maximize over $S\varepsilon F$ the minimized over H error expression, denoted V(S).

We can now determine robust filters through the least favorable or maximizing spectra. Three convex classes of spectral density matrices will be considered, denoted as C_1, C_2, C_3.

$$C_1 = \{S(\lambda); \lambda\varepsilon[-\pi,\pi], 0 \leq a_i(\lambda) \leq s_i(\lambda) \leq b_i(\lambda);$$

$$\int_{-\pi}^{\pi} s_i(\lambda) d\lambda = p_i; i = 1, \cdots, d\} \quad (15)$$

where $s_1(\lambda), \ldots, s_d(\lambda)$ are the eigenvalues of $S(\lambda)$, $a_i(\lambda)$, $b_i(\lambda)$ are known lower and upper spectral bounds, and p_i are known constants. This class can be also defined for time continuous problems, substituting w for λ and the real axis $(-\infty, +\infty)$ for $[-\pi, \pi]$.

The second class is defined through fixing the total signal power in given frequency bands, i.e.:

Let $I_1 U I_2 U \cdots I_m = [-\pi, \pi], I_i \cap I_j = 0$ be a disjoint set of frequency bands covering the spectrum $[-\pi, \pi]$. We define the convex class:

$$C_2 = \{S(\lambda); \lambda \varepsilon [-\pi, \pi], \int_{I_k} \sum_{i=1}^{d} s_i(\lambda) d\lambda = p_k, k = 1, ..., m\} \quad (16)$$

where p_k is the total power of all d components in the spectral band I_k. The third class of spectra is defined as a convex combination of m known spectra:

$$C_3 = \{S(\lambda) = \sum_{i=1}^{m} q_i S_i(\lambda), \sum_{i=1}^{m} q_i = 1, q_i \geq o\} \quad (17)$$

We are now in a position to state and prove two theorems that will be used to determine robust solutions.

Theorem 1

Let $\{G_i[s_i(\lambda)]; i = 1, ..., d\}$ be concave and differentiable functions of $s_i(\lambda)$, let $s_1(\lambda), \ldots, s_d(\lambda)$ be the eigenvalues of $S(\lambda)$, and $S(\lambda) \varepsilon C_1$. Define the functional:

$$G[S] = \int_{-\pi}^{\pi} \sum_{i=1}^{d} G_i[s_i(\lambda)] d\lambda \quad (18)$$

Then, G[S] is maximized over $S \varepsilon C_1$ by the functions $[s_1{}^0(\lambda), \cdots, s_d{}^0(\lambda)]$, where:

$$s_k{}^0(\lambda) = \max[a_k(\lambda), \min[c_k, b_k(\lambda)]] \quad (19)$$

where c_k is uniquely chosen so that: $\int_{-\pi}^{\pi} s_k(\lambda) d\lambda = p_k$.

Theorem 2

Let G[S] be as in eq. (18). The maximum of G(S) over $S \varepsilon C_2$, where C_2 is defined by eq. (16), is achieved by the functions:

$$s_k{}^0(\lambda) = q_k{}^{-1} p_k a_i \ \text{for} \lambda \varepsilon I_k, \ i = 1, ..., d, k = 1, ..., m; \quad (20)$$

where q_k is the measure of I_k and the nonnegative constants $\{a_i\}$ sum to 1:

$$\sum_{i=1}^{d} a_i = 1, a_i \geq 0, a = (a_1 \cdots a_d). \quad (21)$$

Let

$$g(a) = \sum_{i=1}^{d} \sum_{k=1}^{m} q_k G_i(q_k{}^{-1} p_k a_i) \quad (22)$$

Then, the maximizing $a^0 = (a_1{}^0 \cdots a_d{}^0)$ is found as the vector that maximizes g(a) under the conditions (21).

(Note: complete proof for both Theorems can be found in the extended version of this paper.)

We can now apply Theorem 1 for robust estimation, based on the minimax saddle point equation (14), for various cases. Consider the prediction problem, for a time series based on the infinite past. For the case of scalar process, the minimum mean square error is given by eq. (4), and is directly related to the "spectral entropy" H(S) of the process:

$$H(S) = (2\pi)^{-1} \int_{-\pi}^{\pi} \log S(\lambda) d\lambda, \quad (23)$$

$$P_1(S) = \exp H(S). \quad (24)$$

More generally, for the d-dimensional process with spectral density matrix $S(\lambda)$, the entropy per dimension is:

$$H(S) = (2\pi d)^{-1} \int_{-\pi}^{\pi} \log \det S(\lambda) d\lambda \quad (25)$$

and the trace of the optimum prediction error matrix $P_d(S)$ is:

$$g_d(S) = \operatorname{tr} P_d(S) = \exp H(S). \quad (26)$$

According to the general result (14), and equations (23)-(26), the robust prediction filter corresponds to the maximum entropy process. This observation and connection was made in the survey paper of Poor and Kassam [3]. In this paper, we expand it for d-dimensional processes and new spectral uncertainty classes. We seek the maxentropic process that maximizes H(S). Consider the class C_1 of spectral density matrices, defined by eq. (15). We note that:

$$H(S) = (2\pi d)^{-1} \int_{-\pi}^{\pi} \sum_{k=1}^{d} \log s_k(\lambda) d\lambda \quad (27)$$

where $s_1(\lambda), ..., s_d(\lambda)$ are the eigenvalues of $S(\lambda)$. According to Theorem 1, H(S) is maximized over $S \varepsilon C_1$ by the spectral density matrix S with eigenvalues:

$$\{s_1{}^0(\lambda), s_2{}^0(\lambda), \cdots, s_d{}^0(\lambda)\}$$

expressed by equation (19).

Consider next the maximization of H(S) over the spectral class C_2, defined by eq. (16). We identify the concave functions $G_i(s_i(\lambda))$ of theorems 1, 2 by:

$$G_i(s_i(\lambda)) = \log s_i(\lambda) \quad (28)$$

Then, using Theorem 2, we find that the maximizing spectral density has eigenvalues

$$\{s_1{}^0(\lambda), s_2{}^0(\lambda), \cdots, s_d{}^0(\lambda)\}$$

expressed by equation (20), and parametrized by the nonnegative vector a defined by eq. (21). The optimum value of a is found as the argument maximizing the function:

$$g(a) = \sum_{i=1}^{d} \sum_{k=1}^{m} q_k \log[q_k{}^{-1} p_k a_i] \quad (29)$$

In Section II, we have also proven that H(S) is a concave in S functional. Therefore, the maximum entropy H(S) over the convex class C_3 defined eq. (17), is determined by a convex Kuhn-Tucker [13] maximization:

$$\max_{Q \varepsilon B} H(S) = \max_{Q \varepsilon B} \int_{-\pi}^{\pi} \log \det \sum_{i=1}^{m} q_i S_i(\lambda) d\lambda \quad (30)$$

where

$$B = \{Q = (q_1 \cdots q_m), q_i \geq 0, \sum_{i=1}^{m} q_i = 1\}. \quad (31)$$

We consider next the discrete time robust interpolation problem, for the three classes of spectra, C_1, C_2, C_3. The robust interpolation is achieved by an interpolation filter H_O of eq. (10), for a spectrum S that minimizes the criterion G(S) given by eq. (11). It is shown in Section II that G(S) is a convex functional of S, hence we have a convex minimization problem. However, we cannot find any more explicit solutions for the general classes of spectra C_1, C_2, C_3, as we could do for the prediction error. We will consider more restrictive classes of spectra. Let

$$C_O = \{S(\lambda); S(\lambda) = AF(\lambda)A^T\} \quad (32)$$

where A is an orthogonal constant matrix, and $F(\lambda) = \mathrm{diag}[s_1(\lambda),...,s_d(\lambda)]$ is the diagonal matrix consisting of the d eigenvalues $\{s_1(\lambda), s_2(\lambda),..., s_d(\lambda)\}$ of $S(\lambda)$. For $S(\lambda)$ in the class C_O, we have:

$$G(S) = \log\det\int_{-\pi}^{\pi} S^{-1}(\lambda)d\lambda = \log\det(A^T)^{-1}\int_{-\pi}^{\pi} F^{-1}(\lambda)d\lambda A^{-1} =$$

$$= \log\prod_{k=1}^{d}\int_{-\pi}^{\pi} s_k^{-1}(\lambda)d\lambda = \sum_{k=1}^{d}\log\int_{-\pi}^{\pi} s_k^{-1}(\lambda)d\lambda. \quad (33)$$

We are now in a position to seek the minimizing spectrum S of the class $C_O \cap C_1$. With a slight modification of Theorem 1, we observe that the minimum of the integral:

$$\int_{-\pi}^{\pi} s_k^{-1}(\lambda)d\lambda$$

for $s_k(\lambda)$ that satisfy the conditions:

$$a_k(\lambda) \le s_k(\lambda) \le b_k(\lambda), \int_{-\pi}^{\pi} s_k(\lambda)d\lambda = p_k$$

is achieved by $s_k^{\,o}(\lambda)$, defined by eq.(19).

Next, we consider the problem of robust causal filtering, for both discrete and continuous time. For the discrete time case, the minimum mean square error expression for filtering in white noise is available only for the scalar (d = 1) case, and expressed by q(S), of eq. (13).

The robust causal filter corresponds to the spectral density that maximizes q(s). We consider the case of known, fixed noise level N_o, and incompletely known signal spectrum $S(\lambda)$.

In this case, the robust filter corresponds to the spectral density S that maximizes the integral:

$$B(S) = \int_{-\pi}^{\pi} \log[1 + N_O^{-1}S(\lambda)]d\lambda \quad (34)$$

For d=1, the spectral density of the class C_1 that maximizes B(S), is:

$$N_O^{-1}S^0(\lambda) = \max[a(\lambda), \min[c, b(\lambda)]] \quad (35)$$

where c is the value that is chosen so that $S^0(\lambda)$ has given total power. This conclusion is a special case of Theorem 1.

The spectral density of class C_2 that maximizes B(S) is the piecewise constant:

$$N_O^{-1}S^0(\lambda) = q_k^{-1} \cdot p_k; k = 1,...,m \quad (36)$$

Finally, the maximization of B(S) over $S \varepsilon C_3$ is a Kuhn-Tucker convex optimization problem [13].

Consider the problem of robust causal filtering in continuous time, in the presence of white noise. The closed form expression for the minimum mean square error is eq. (12). Suppose N_0 is the known covariance matrix of the white noise, and $\{z_1(w), z_2(w), z_3(w), \ldots, z_d(w)\}$ are the eigenvalues of the matrix $Z(iw) = S(iw)N_0^{-1}$. We assume that Z(iw) is a member of the uncertainty class C_1 of eq. (15), but with $(-\infty, +\infty)$ replacing the integration limits $[-\pi, \pi]$. By an obvious modification of Theorem 1, we find that the maximum of $h_d(S)$ of eq. (12) is achieved by functions:

$$z_k^0(w) = \max[a_k(w), \min[c_k, b_k(w)]] \quad (37)$$

where c_k is chosen so that $z_k^0(w)$ integrates to a given value p_k.

Note that an additional constraint needs to be placed on the eigenvalues $\{z_k(w)\}$, namely that all of them are zero outside a finite interval. This is necessary in order to keep $h_d(S)$ finite.

If $[-W_0, W_0]$ is the w-interval outside of which all eigenvalues are zero, we can replace the limits $(-\infty, \infty)$ at the integral (12) by $[-W_0, W_0]$.

Under the finite support assumption, we can now proceed and solve the robust causal filtering problem for the class C_2 of eq. (16). We consider a subdivision $\{I_1, I_2, \ldots I_m\}$ of the spectral band $[-W_0, W_0]$, and then define C_2 accordingly. The uncertainty class is now chosen so as the matrix Z(w) is in C_2. Using Theorem 2, we find the set of eigenvalues $\{z_k(w)\}$ that maximize $h_d(S)$:

$$z_k^{\,O}(w) = q_k^{-1}p_k a_i \text{ for } w\varepsilon I_k, i = 1,...,d,\ k = 1,...,m. \quad (38)$$

where q_k is the measure I_k, and $\{\lambda_i\}$ are as in (21). The optimum a is found by maximizing the function

$$g(a) = \sum_{i=1}^{d}\sum_{k=1}^{m} q_k \log[1 + q_k^{-1}p_k a_i] \quad (39)$$

under the constraints (21).

IV. REFERENCES

[1] R. Mangoubi. Robust Estimation and Failure Detection, A Concise Treatment. Springer-Verlag, Berlin, Germany, 1998.

[2] L. Xie, Y. Chai Soh, "Robust Kalman filtering for uncertain systems", Systems and Control Letters, Vol. 22, 1994, pp 123-129.

[3] S. A. Kassam and H. Vincent Poor, "Robust Techniques for Signal Processing: A Survey", IEEE Trans. on Infor. Theory, Vol. 73, no. 3, March 1985.

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

[5] R. H. Shumway, David S. Stoffer, Time series analysis and its applications, New York: Springer, 2000.

[6] C. T. Chen and S. A. Kassam, "Robust Wiener filtering for multiple inputs with channel distortion", IEEE Trans. on Infor. Theory, vol. 1T-30, pp. 674-677, July 1984.

[7] K. S. Vastola, and H. V. Poor, "Robust Wiener- Kolmogorov theory", IEEE Trans. on infor. Theory, Vol 30, no. 2, pp. 316-327, March 1984.

[8] J. Franke, "Minimax robust prediction of discrete time series", Z. Wahr. Verw. Geb, Vol. 68, pp. 337-364, 1985.

[9] J. Snyders, "On the error matrix in optimal linear filtering of stationary processes", IEEE Trans. on Infor. Theory, Vol. 1T-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. 1T-11, pp. 594-595, 1965.

[11] K. Ohrn, A. Ahl'en and M. Sternad, "A probabilistic approach to multivariable robust filtering and open-loop control", IEEE Trans. on Automatic Control, Vol. 40, pp 405-418, 1995.

[12] P. Bolzern, P. Colaneri and G De Nicolao, "Optimal robust filtering for linear systems subject to time-varying parameter perturbations," Proc. 32nd Conf. on Decision and Control, San Antonio, Texas, pp 1018-1023, 1993.

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

[14] R. T. Rockafellar, Convex Analysis, Princeton University Press.