total variation classes, if the nominal $m_n^0 \ll \lambda$, and for the band class, if $m_n^0 \ll \lambda$ and $m_n^1 \ll \lambda$. Then we have

$$\int_0^T \hat{h}^2(t)n(t)\,dt \le \int_{[0,T]} \hat{h}^2\,dm_n = \int_{[0,T]} s^2\hat{n}^{-2}\,dm_n$$
$$\le \int_{[0,T]} s^2\hat{n}^{-2}\,d\hat{m}_n = \int_0^T \hat{h}^2(t)\hat{n}(t)\,dt. \quad (22)$$

To prove the first inequality in (22), we used the fact that the absolutely continuous part of a measure is no larger than the measure itself; to prove the second we used the fact that $\hat{n}$ is stochastically smallest under $\hat{m}_n$ over all other elements in the capacity class.

Case 2: Here we make use of the discrete-set version of Lemma 1. The optimal filter is $\hat{h}_i = s_i/\hat{n}_i$, where $\hat{n}_i$ is the pmf corresponding to the measure $\hat{m}_n$ singled out by Lemma 1 when applied to this case. The equivalent form to (21) is

$$\sum_{i=1}^k \hat{h}_i^2 n_i = \sum_{i=1}^k s_i^2\hat{n}_i^{-2}n_i \le \sum_{i=1}^k s_i^2\hat{n}_i^{-2}\hat{n}_i = \sum_{i=1}^k \hat{h}_i^2\hat{n}_i. \quad (23)$$

To prove the inequality in (23) we used the fact that $\hat{n}$ becomes stochastically smallest under $\hat{m}_n$, singled out by Lemma 1.

Cases 3 and 4: The optimal filter here is $\hat{H}(\omega) = S(\omega)/\hat{N}(\omega)$, where $\hat{N}(\omega)$ is the R–N derivative of the least favorable spectral measure $\hat{m}_N$, singled out by Lemma 1, with respect to λ. Both the continuous-time and the discrete-time problems are represented here with common notation. The equivalent form to (21) is

$$\int |\hat{H}|^2\,dm_N = \int |S|^2\hat{N}^{-2}\,dm_N \le \int |S|^2\hat{N}^{-2}\,d\hat{m}_N = \int |\hat{H}|^2\,d\hat{m}_N. \quad (24)$$

To prove the inequality in (24), we used the fact that $\hat{N}$ becomes stochastically smallest under $\hat{m}_N$ singled out by Lemma 1. The integrals in (24) are over $[-\omega_0, \omega_0]$ or over $[-\pi, \pi]$ for the continuous- and discrete-time cases, respectively.

V. Conclusion

The robust matched filter for uncertainty in the noise autocorrelation function or the noise spectral measure is derived for both continuous and discrete-time problems, when the uncertainty classes are generated by two-alternating capacities. In all cases the maximin robust matched filter depends on the inverse of the worst case noise statistic, which is obtained as the Huber–Strassen derivative of the capacity generating the uncertainty class with respect to the Lebesgue (or other equivalent measure) on a suitable interval.

References

[1] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing," *Proc. IEEE*, vol. 73, Mar. 1985, pp. 433–481.

[2] H. V. Poor, "Robust matched filters," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 677–687, Sept. 1983.

[3] S. Verdu and H. V. Poor, "Minimax robust discrete-time matched filters," *IEEE Trans. Commun.*, vol. COM-31, pp. 208–215, Feb. 1983.

[4] P. J. Huber, "A robust version of the probability ratio test," *Ann. Math. Statist.*, vol. 36, pp. 1753–1758, 1965.

[5] S. A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 242–247, 1981.

[6] K. S. Vastola and H. V. Poor, "On the p-point uncertainty class," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 374–376, 1984.

[7] P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," *Ann. Statist.*, vol. 1, pp. 251–265, 1973.

Estimating the Order of a FIR Filter for Noise Cancellation

PIERRE COMON, MEMBER IEEE, AND DINH TUAN PHAM, MEMBER IEEE

Abstract—**The problem of designing a finite impulse response (FIR) filter devoted to the joint process problem, and in particular to noise cancellation, is investigated. The goal is restricted to how to choose an optimal order of the FIR filter. The analysis is based upon the maximization of a noise reduction criterion, which is in fact perfectly matched to the desired performance, namely the minimization of residuals. The criterion includes estimation errors introduced by the substitution of the true covariance matrices for estimates. Since true covariance matrices again enter the noise reduction criterion, they must be replaced by estimates. The criterion is then modified in order to take this fact into account, but the scheme is not iterated any further. Simulations are presented in various simple cases.**

I. Introduction

Noise cancellation has been studied for more than 20 years. Most essential results are now well known. On the other hand, there are always problems raised by the application of optimal Bayesian solutions (namely solutions that need the knowledge of the *a posteriori* probability density function) in the adaptive case. The second order moments are for instance required in the Gaussian context, and must actually be computed from the observations. This is rarely taken into account when estimating the optimal solutions. As it has been early pointed out [6], works on this topic are published regularly. In [2] a simple noise reduction principle has been used to evaluate the performance and limits of a spectral Wiener filter. Based on the same principle, we propose here to compute the optimal order of a FIR filter for noise cancellation, or more generally for the joint-process regression problem [10]. The idea of using estimation errors to estimate the order of a filter is not new. It has been already used, for instance, in [1] for autoregressive models fitting, and in [7] for continuous-time noise cancelling. The same approximation philosophy is also adopted in [12]. In the context of noise cancelling, the observation model is constituted of a signal s_t additively corrupted by a noise x_t, and of a vector-valued process y_t standing for the noise references. So the observations (u_t, v_t) can be written as

$$u_t = s_t + x_t$$
$$v_t = y_t. \quad (1)$$

u_t and v_t are jointly stationary up to second order; x_t and y_t denote noises. For each t, y_t is a $n \times 1$ vector in which the entries are each issuing from sensors that receive noise alone. Furthermore, signal and noises satisfy the following properties of independence:

$$E\{s_t x_{t+k}\} = 0, \qquad \forall\, k \in \mathbb{Z}$$
$$E\{s_t y_{t+k}\} = 0, \qquad \forall\, k \in \mathbb{Z}. \quad (2)$$

The goal in the so-called joint-process problem [10], is to extract the process s_t from u_t. The optimal linear estimate of the signal s_t at time t, given the past values of the observation process

Manuscript received November 4, 1986; revised July 12, 1989.
P. Comon is with Thomson-Sintra, B.P. 53, F-06801 Cagnes sur mer, France.
D. T. Pham is with Tim3-Imag, B.P. 68, F38402 St. Martin d'Heres, France.
IEEE Log Number 8933848.

0018-9448/90/0300-0429$01.00 ©1990 IEEE

$(u_t', v_t')'$ is the orthogonal projection of s_t onto the space $\mathscr{W}_t$ spanned by the components of $\{u_{t-k}, v_{t-k}, k \in \mathbb{N}\}$, according to the L^2 scalar product; prime denotes here transposition. This signal estimate could be obtained from the Wiener theory [8]. However, the computation of the optimal filter requires the knowledge of the spectral densities of the processes (u_t', v_t') and s_t, and their factorization. Since the factorization of a spectral density is a complicated problem [5], and since there is no simple procedure to estimate the moments of the unobserved signal process, s_t, the general Wiener theory is not adequate for practical use [3].

Let f_i be the coefficients vectors of the linear regression of x_t on y_t:

$$x_t = \sum_{i=0}^{p} f_i' y_{t-i} + z_t = F'Y_t + z_t, \quad E\{z_t y_{t-k}\} = 0, \qquad \forall\, k \geq 0 \tag{3}$$

where

$$F = (f_0', f_1', \cdots, f_p')' \quad \text{and} \quad Y_t = (y_t', y_{t-1}', \cdots, y_{t-p}')'.$$

The best linear unbiased estimate [4] (BLUE) of x, which is consistent when z_t and y_{t-k} are independent for all $k \in \mathbb{Z}$ (strong independence assumption), requires the additional knowledge of the covariance function of the regression residual process $\zeta_t = u_t - F'Y_t$. Moreover, it can be shown that the BLUE is not consistent when z_t and y_{t-k} are independent only for $k \geq 0$ (weak independence assumption). Thus, we shall restrict our attention merely to the regression of x_t on y_t, which is eventually the only feasible way of solving the problem under the weak independence assumption. The purpose of the paper will be to design a strategy for choosing the number of coefficients to be used.

II. Regression Formulas

A. LLMS Estimate

The best linear least-mean-squares (LLMS) estimate of x is then given by

$$x_t^0 = F'Y_t, \quad F = R^{-1}C, \quad C = E\{Y_t u_t\}, \quad R = E\{Y_t Y_t'\} \tag{4a}$$

and the best estimate of s_t amounts to

$$\zeta_t = u_t - x_t^0. \tag{4b}$$

To summarize the notations, we have

$$x_t = F'Y_t + z_t \quad \text{and} \quad u_t = F'Y_t - \zeta_t. \tag{4c}$$

The last relation may be interpreted as the regression of u_t onto Y_t.

Notice that our observation model contains a particular case: the q-step prediction problem. In fact, this is obtained if $s_t = 0$, $u_t = x_t = y_{t+q}$, and $n = 1$. Furthermore, if ζ_t is a white noise, then y_t is an autoregressive (AR) process of order p.

B. Square Window

The covariance matrices are actually unknown, and the adaptive regression formulas are obtained by using optimal regression solutions with estimated covariance matrices

$$\hat{x}_t = \hat{F}'Y_t \tag{5}$$

with

$$\hat{F} = \hat{R}_t^{-1}\hat{C}_t; \quad \hat{C}_t = \sum_{k=\delta}^{N-1+\delta} Y_{t-k} u_{t-k}, \quad \hat{R}_t = \sum_{k=\delta}^{N-1+\delta} Y_{t-k} Y_{t-k}' \tag{6}$$

with δ being an integer *a priori* chosen ($\delta \geq 0$). The interest of such a delay will be emphasized later in Section III-A. The adaptive LLMS estimate of the signal, $\hat{s}_t$, is defined by

$$\hat{s}_t = u_t - \hat{x}_t = (s_t + z_t) + (x_t^0 - \hat{x}_t) = \zeta_t + \epsilon_t \tag{7}$$

where

$\zeta_t = s_t + z_t$ is also given by (4b) and $\epsilon_t = x_t^0 - \hat{x}_t$.

or, from (4a) and (5)

$$\hat{s}_t = \zeta_t + (F - \hat{F})'Y_t. \tag{8}$$

It is clear from (8) and (4b) that the estimate $\hat{s}_t$ splits into a minimal error ζ_t, which is the estimate of s_t, and an extraneous error ϵ_t, due exclusively to estimation errors upon the correlation matrices, achieving zero in the nonadaptive optimal case. Using (4c), estimator (6) can be rewritten as

$$\hat{C}_t = \sum_k Y_{t-k}\zeta_{t-k} + \sum_k Y_{t-k}Y_{t-k}'F \tag{9}$$

yielding finally:

$$(\hat{F} - F)' = \sum_k \zeta_{t-k} Y_{t-k}' \hat{R}_t^{-1}. \tag{10}$$

C. Exponential Window

Another estimate of interest is obtained with an exponential averaging of the following sample correlations:

$$\begin{aligned} \overline{C}_t &= \alpha \overline{C}_{t-1} + (1-\alpha) Y_{t-\delta} u_{t-\delta} \\ \overline{R}_t &= \alpha \overline{R}_{t-1} + (1-\alpha) Y_{t-\delta} Y_{t-\delta}' \end{aligned} \tag{11}$$

with

$$\alpha \in [0, 1].$$

The observation window is equivalent in that case to $]-\infty, t-\delta]$. Similarly to (10), we have for estimator (12):

$$\overline{F} = \overline{R}_t^{-1}\overline{C}_t$$

$$(\overline{F} - F)' = \sum \alpha^k \zeta_{t-k} Y_{t-k}' \left[\sum \alpha^k Y_{t-k} Y_{t-k}'\right]^{-1}. \tag{12}$$

This corresponds to a weighted LLS estimator with weights α^k assigned to the square of the residual at time $t-k$. Extension to this case will be addressed in the Appendix.

III. Order Estimation

A. Optimization Criterion

The purpose of this section is to define an optimal order, p^0, for the filter (3). Since we are dealing here with noise cancellation problems, it turns out to be the most natural to consider that the best value of p is the one yielding the greatest noise reduction. In other words, the adaptive FIR filter is said to perform the best if the output noise power (ONP) is minimum. The ONP can be expressed as

$$\text{ONP} = E\{(z_t + \epsilon_t)^2\}. \tag{25}$$

Since the quantity $E\{s_t^2\}$ does not depend on p, the following minimization problems are equivalent:

$$\begin{aligned} p^0 &= \arg\min\,[\text{ONP}] \\ p^0 &= \arg\min\,[\text{ONP} + E\{s_t^2\}]. \end{aligned} \tag{26}$$

Let us go back to the formulas of adaptive regression. We have introduced in Section II a delay δ in the expressions of the sample correlations (6) or (11); the interest of such a procedure will appear now more clearly. For sufficiently large values of N or δ, we know that the estimated covariance matrices $\hat{C}_t, \hat{R}_t, \overline{C}_t, \overline{R}_t$ are approximately independent of the samples s_t, x_t, Y_t. Then, the processes $z_t = (x_t - F'Y_t)$ and $\zeta_t = (s_t + x_t - F'Y_t)$ are also approximately independent. We can easily see that this amounts to the asymptotic independence between ϵ_t and $\{s_t, x_t, Y_t, z_t, \zeta_t\}$; this property will be necessary in the sequel. With regard to the practice, it is not always possible to choose N large enough to achieve a reasonable independence; on the other hand, the additional use of factor δ helps a lot. However, the choice of δ relies upon the user's experience and is essentially based on heuristics: The value of δ should aim to be longer than the correlation length of $E\{z_t \epsilon_{t+\tau}\}$. As pointed out in Section IV-B, the optimal choice of δ is out of the scope of the present correspondence.

If we utilize the previously mentioned independence assumption, the optimization problem (26) becomes

$$p^0 = \arg\min \psi(p)$$
$$\psi(p) = E\{(\zeta_t + \epsilon_t)^2\} = E\{\zeta_t^2\} + E\{\epsilon_t^2\}. \qquad (27)$$

Denote the covariance functions of u, Y, ζ by Γ_u, Γ_Y, and Γ_ζ, respectively. Let us now express the two terms in (27). First, by definition of u_t in (1) and (2), we get

$$\Gamma_u(\tau) = F\Gamma_Y(\tau)F' + \Gamma_\zeta(\tau).$$

This yields

$$E\{\zeta_t^2\} = \Gamma_u(0) - F\Gamma_Y(0)F'. \qquad (28)$$

Secondly, the variance of ϵ_t is

$$E\{\epsilon_t^2\} = \operatorname{trace}\{\Gamma_Y(0)E[(\hat{F} - F)(\hat{F} - F)']\}.$$

Yet, according to [4], $(\hat{F} - F)$ is asymptotically normally distributed as

$$(\hat{F} - F) \sim N\left(0, \frac{1}{N}\Gamma_Y^{-1}(0)\left[\sum_{\tau=-\infty}^{\infty}\Gamma_\zeta(\tau)\Gamma_Y(\tau)\right]\Gamma_Y^{-1}(0)\right).$$

Then asymptotically

$$E\{\epsilon_t^2\} \simeq \operatorname{trace}\left\{1/N\left[\sum_{k=-\infty}^{\infty}\Gamma_\zeta(k)\Gamma_Y(k)\right]\Gamma_Y^{-1}(0)\right\}. \qquad (29)$$

On the other hand, because of (2), (4), and the weak assumption of independence, for all $\tau \geq 0$, the relation next holds:

$$\Gamma_\zeta(\tau) = E\{\zeta_t\zeta_{t+\tau}\} = E\{u_t\zeta_{t+\tau}\} = \Gamma_u(\tau) - \Gamma_{uY}(\tau)F', \qquad (30)$$

where

$$\Gamma_{uY}(\tau) \triangleq E\{u_t Y_{t+\tau}\}.$$

Thus, from (28), (29), and (30), we obtain the new expression

$$\psi(p) = 2\sum_{k=1}^{\infty}[\Gamma_u(k) - \Gamma_{uY}(k)F']\operatorname{trace}\{\Gamma_Y(k)\Gamma_Y^{-1}(0)\}/N$$
$$+[\Gamma_u(0) - \Gamma_{uY}(0)F']n(p+1)/N + \Gamma_u(0) - F\Gamma_Y(0)F'. \qquad (31)$$

A more practical form can be deduced using the fact that $F' = \Gamma_Y(0)^{-1}\Gamma_{Yu}(0)$:

$$\psi(p) = 2\sum_{k=1}^{\infty}\left[\Gamma_u(k) - \Gamma_{uY}(k)\Gamma_Y(0)^{-1}\Gamma_{Yu}(0)\right]$$
$$\cdot \operatorname{trace}\{\Gamma_Y(k)\Gamma_Y^{-1}(0)\}/N$$
$$+\left[\Gamma_u(0) - \Gamma_{uY}(0)\Gamma_Y(0)^{-1}\Gamma_{Yu}(0)\right](1 + n(p+1)/N). \qquad (32)$$

This function is generally unimodal, and provides a single solution p^0. However, there is no theoretical proof that $\psi(p)$ admits a single minimum.

Example: Let us look at the form of (32) in the case of white processes s_t, x_t and Y_t. It turns out that, for any $k \neq 0$:

$$\Gamma_u(k) = 0; \quad \Gamma_Y(k) = 0, \quad \Gamma_{uY}(k) = 0.$$

Thus (32) simplifies into

$$\psi(p) = \left[\Gamma_u(0) - \Gamma_{uY}(0)\Gamma_Y(0)^{-1}\Gamma_{Yu}(0)\right](1 + n(p+1)/N). \qquad (33)$$

This kind of result has been already used in [3] for a robustness evaluation. Expression (32) is much more complicated than (33) and will need to resort to a numerical solution.

B. Practical Criterion

Since covariance functions $\Gamma_{uY}(k)$, $\Gamma_u(k)$, and $\Gamma_Y(k)$ entering expression (32) are unknown, they must be replaced by estimates, namely $C_{uY}(k)/N$, $C_u(k)/N$ and $C_y(k)/N$:

$$C_u(k) = \sum_{i=\delta+k}^{N-1+\delta+k} u_{t-i}u_{t-i+k}$$
$$C_{uY}(k) = \sum_{i=\delta+k}^{N-1+\delta+k} u_{t-i}Y'_{t-i+k}$$
$$C_Y(k) = \sum_{i=\delta+k}^{N-1+\delta+k} Y_{t-i}Y'_{t-i+k}. \qquad (34)$$

Note that, for the sake of simplicity, variable t has been omitted in these expressions. For instance $C_{uY}(0) = \hat{C}_t$ and $C_Y(0) = \hat{R}_t$. The substitution of the true values $\Gamma_{uY}(k), \Gamma_u(k), \Gamma_Y(k)$ for their estimates $C_{uY}(k)/N, C_u(k)/N, C_Y(k)/N$ involves a small extraneous error that is not negligible. Indeed, from the central limit theorem, it is known that the error is of order $1/\sqrt{N}$ [11], but the expression for $\psi(p)$ as given in (32) contains terms of order $1/N$. Therefore, we need to assess the magnitude of errors in $\psi(p)$ introduced by the parameter estimation up to this precision level. Since the first term of (32) is already divided by $1/N$, replacing the unknown covariances by their estimates in this term will introduce errors of higher order than $1/N$. This may seem unclear since (32) includes an infinite sum, but it can be justified as follows. Assume that the quantity

$$\Phi(p) = \sum_{k=1}^{\infty}\left[\Gamma_u(k) - \Gamma_{uY}(k)\Gamma_Y^{-1}(0)\Gamma_{Yu}(0)\right]$$
$$\cdot \operatorname{trace}\{\Gamma_Y(k)\Gamma_Y^{-1}(0)\}$$

is estimated by

$$\hat{\Phi}(p) = \sum_{k=1}^{M}\left[C_u(k) - C_{uY}(k)C_Y^{-1}(0)C_{Yu}(0)\right]$$
$$\cdot \operatorname{trace}\{C_Y(k)C_Y^{-1}(0)\}.$$

Then one may admit that $\hat{\Phi}(p)$ is a consistent estimate of $\Phi(p)$ when $n \to \infty$ and $M \to \infty$ provided $M/N \to 0$ (the rigorous proof is skipped for reasons of space). In fact, the same assumptions suffice for deriving consistent estimates of spectra from correlation lags [4]. The first term of $\psi(p)$ being $2\hat{\Phi}(p)/N$, it is indeed of order $1/N$. As a conclusion we shall not worry about the first term. On the other hand, the errors in the second term are, up to the first order in $1/N$:

$$1/N\left[C_u(0)-C_{uY}(0)C_Y(0)^{-1}C_{Yu}(0)\right] - \Gamma_u(0)+\Gamma_{uY}(0)\Gamma_Y(0)^{-1}\Gamma_{Yu}(0). \quad (35)$$

Since, from (6) and (34), $\hat{F}=C_Y(0)^{-1}C_{Yu}(0)$, the first bracket

$$1/N\sum_{i=\delta+k}^{N-1+\delta+k}\left(u_{t-i}-\hat{F}'Y_{t-i}\right)^2=1/N\sum_{i=\delta+k}^{N-1+\delta+k}\hat{s}_{t-i}^2$$

where $\hat{s}(t)$ is given by (7). Then using (8), the last expression equals

$$1/N\sum_{i=\delta+k}^{N-1+\delta+k}\zeta_{t-i}^2+(\hat{F}-F)'[C_Y(0)/N](\hat{F}-F) -2/N(\hat{F}-F)'\sum_{i=\delta+k}^{N-1+\delta+k}\zeta_{t-i}Y_{t-i}.$$

Using (9b), it can be seen that the second term is the same as the last term, up to a factor -2. Thus, noting that $\Gamma_u(0)-\Gamma_{uY}(0)\Gamma_Y(0)^{-1}\Gamma_{Yu}(0)$ is $E\{\zeta_t^2\}$, the estimation error (35) is

$$\left[1/N\sum_{i=\delta+k}^{N-1+\delta+k}\zeta_{t-i}^2-E\{\zeta_t^2\}\right]+(F-\hat{F})'[C_Y(0)/N](F-\hat{F}). \quad (36)$$

The first part in brackets in (36) has zero mean, and depends little on the chosen order p (this may be justified by heuristical arguments by assuming that there exists a true order, and that we are in a reasonable neighborhood of it). Therefore, it can be dropped.

The second part however is always negative and depends on p. Since $C_Y(0)/N$ tends to $\Gamma_Y(0)$ as N tends to infinity, result (14) shows that $N(\hat{F}-F)'[C_Y(0)/N](\hat{F}-F)$ converges in distribution to $\operatorname{trace}(\Gamma_Y(0)W)$, where W is a Wishart matrix with one degree of freedom and mean $\Gamma_Y(0)^{-1}\Sigma_k\Gamma_\zeta(k)\Gamma_Y(k)\Gamma_Y(0)^{-1}$. In order to correct the effect of the estimation errors in $\psi(p)$, one should add to it the mean of $\operatorname{trace}(\Gamma_Y(0)W)/N$, which is precisely from (29) $E\{\epsilon_t^2\}=\psi(p)-E\{\zeta_t^2\}$. Taking into account this correction, the criterion (32) becomes

$$\hat{\psi}(p)=\left[C_u(0)-C_{uY}(0)C_Y(0)^{-1}C_{Yu}(0)\right](1+2n(p+1)/N) +4/N\sum_{k=1}^{\infty}\left[C_u(k)-C_{uY}(k)C_Y(0)^{-1}C_{Yu}(0)\right] \cdot\operatorname{trace}\left\{C_Y(k)C_Y(0)^{-1}\right\}. \quad (37)$$

Remark: Our approach is somewhat similar to that of Akaike [1] (but also much more complicated) in the problem of choosing the order of an AR process for prediction. Expression (37) of $\hat{\psi}(p)$ reduces to Akaike criterion when ζ_t is a white noise sequence (cf. formula (30)).

Further, the summation in (37) must again be extended only to a finite set of lags k. So, the set of lags where $C_Y(k)$ vanishes may be neglected.

IV. Applications

A. Computational Remarks

The computation of $\hat{\psi}(p)$ via formula (37) is time consuming. However, a drastic reduction in the computational load can be made by resorting to the well-known Levison–Durbin recursions (that we do not detail), and by some manipulations that allow a recursive computation of $\hat{\psi}(p)$ explained in the following. Instead of $C_Y(k)/N$ we use a block Toeplitz approximate $\hat{\Gamma}_Y(0)$ with blocks $\hat{\gamma}_y(i-j)$ at the (i,j) place, $\{i,j=0,\cdots,p\}$:

$$\hat{\gamma}_y(k)=1/N\sum_{i=\delta+k}^{N-1+\delta}y_{t-i}y'_{t-i+k}. \quad (38a)$$

Here, matrix $\hat{\gamma}_y(k)$ is $n\times n$. Similarly, define

$$\hat{\gamma}_u(k)=1/N\sum_{i=\delta+k}^{N-1+\delta}u_{t-i}u_{t-i+k} \quad (38b)$$

and

$$\hat{\gamma}_{uy}(k)=1/N\sum_{i=\delta+k}^{N-1+\delta}u_{t-i}y'_{t-i+k},$$

so that matrices $C_u(k)/N$ and $C_{uY}(k)/N$ are replaced by $\hat{\gamma}_u(k)$ and $\hat{\Gamma}_{uY}(k)=[\hat{\gamma}_{uy}(k)',\cdots,\hat{\gamma}_{uy}(k+p)']'$. Note that the use of (34) requires $N+K$ max samples whereas (38a) and (38b) require only N. On the other hand, with the same sample size, estimator (38a–38b) differs from (34) only by some end effects that are actually of order $1/N$. Thus, the considerations done in Section III keep available asymptotically, though they are more difficult to prove. The benefit to use (38a–38b) is that we have less covariances values to compute, and mostly, the use of block Toeplitz matrices permits easy inversion by the Levinson–Durbin–Wittle algorithm [9]. Indeed, let $[\hat{a}(0,p),\cdots,\hat{a}(p,p)]$ be the estimated coefficients of the forward autoregressive filter of order p, and $\hat{G}(p)$ be the estimated innovation covariance matrix. Then

$$\sum_{j=0}^{p}\hat{a}(j;p)\hat{\gamma}_y(k-j)=0_n, \qquad \text{for } k=1,\cdots,p$$
$$=\hat{G}(p), \qquad \text{for } k=0. \quad (39)$$

Then, it is well known that the inverse of the matrix $\hat{\Gamma}_Y(0)$ may be written as $T_p'\operatorname{Diag}[\hat{G}(0)^{-1},\cdots,\hat{G}(p)^{-1}]T_p$, where $\operatorname{Diag}(\cdots)$ stands for a block-diagonal matrix, and T_p is lower triangular with entries $\hat{a}(j;p)$. Additionally, coefficients $\hat{a}(j;p)$ and $\hat{G}(p)$ can be quickly computed through the Whittle algorithm [9]. Since the $\hat{a}(j;p)$'s and the $\hat{G}(p)$'s can be interpreted as estimates of the coefficients in the forward prediction error $\eta_t(p)=\Sigma a(j;p)y_{t+j}$ and its covariance matrix $G(p)=E\{\eta_t(p)\eta_t(p)'\}$, respectively, the quantities

$$\hat{\gamma}_{u\eta(p)}(k)=\sum_{j=0}^{p}\hat{a}(j;p)\hat{\gamma}_{uy}(k+j)$$

and

$$\hat{\gamma}_{\eta(p)}(k)=\sum_{j=0}^{p}\sum_{i=0}^{p}\hat{a}(i;p)\hat{\gamma}_y(k+j-i)\hat{a}(j;p)'$$

are estimates of the cross-covariance between u_t and $\eta_t(p)$, and the autocovariance of $\eta_t(p)$, respectively. Note that the last expression may be simplified with the help of (38). Now it can be shown that

$$\hat{\Gamma}_{uY}(0)\hat{\Gamma}_Y(0)^{-1}\hat{\Gamma}_{Yu}(0)=\sum_{i=0}^{p}\hat{\gamma}_{u'\eta(i)}(i)\hat{G}(i)^{-1}\hat{\gamma}_{u\eta(i)}(-i)$$

$$\hat{\Gamma}_{uY}(k)\hat{\Gamma}_Y(0)^{-1}\hat{\Gamma}_{Yu}(0)=\sum_{i=0}^{p}\hat{\gamma}_{u'\eta(i)}(k-i)\hat{G}(i)^{-1}\hat{\gamma}_{u\eta(i)}(-i)$$

$$\operatorname{trace}\left[\hat{\Gamma}_Y(k)\hat{\Gamma}_Y(0)^{-1}\right]=\sum_{i=0}^{p}\operatorname{trace}\left[\hat{\gamma}_{\eta(i)}(0)\hat{G}(i)^{-1}\right].$$

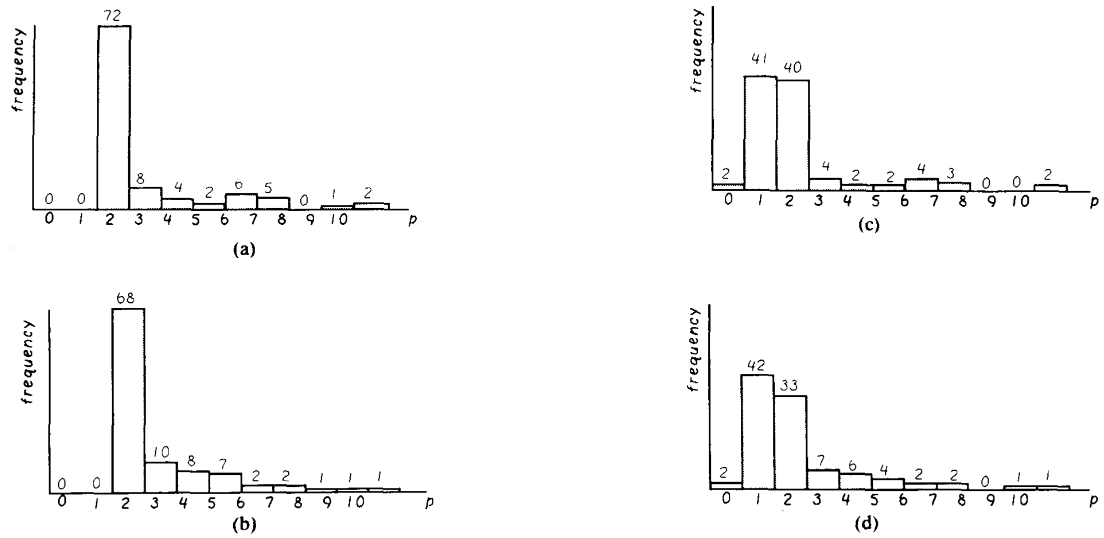

Fig. 1. Histograms of the orders obtained for two different second-order filters and two different noise coloration levels. (a) $F'=(1.0,0.5,0.3)$; $\alpha=0.7$; $N=200$. (b) $F'=(1.0,0.5,0.3)$; $\alpha=0.1$; $N=200$. (c) $F'=(0.5,0.2,0.1)$; $\alpha=0.7$; $N=200$. (d) $F'=(0.5,0.2,0.1)$; $\alpha=0.1$; $N=200$.

Therefore, $\hat{\psi}(p)$ can be computed as follows:

$$\begin{aligned}\hat{\psi}(p)=&\left[\hat{\gamma}_u(0)-\sum_{i=0}^{p}\hat{\gamma}_{u'\eta(i)}(i)\hat{G}(i)^{-1}\hat{\gamma}_{u\eta(i)}(-i)\right]\\&\cdot(1+2n(p+1)/N)\\&+4/N\sum_{k=1}^{\infty}\left[\hat{\gamma}_u(k)-\sum_{i=0}^{p}\hat{\gamma}_{u'\eta(i)}(k-i)\right.\\&\left.\cdot\hat{G}(i)^{-1}\hat{\gamma}_{u\eta(i)}(-i)\right]\sum_{j=0}^{p}\mathrm{Tr}\left[\hat{G}(j)^{-1}\hat{\gamma}_{\eta(j)}(0)\right].\end{aligned}\tag{40}$$

The interesting feature of this formula is that it permits a recursive computation of $\hat{\psi}(p)$. Indeed

$$\begin{aligned}\hat{\psi}(p)-&\left[\hat{\gamma}_u(0)-\hat{\psi}_p'(0)\right](1+2n(p+1)/N)\\&+4/N\sum\left[\hat{\gamma}_u(k)+\hat{\psi}_p'(k)\right]\hat{\psi}_p'',\end{aligned}\tag{41}$$

where

$$\begin{aligned}\hat{\psi}_p'(k)&\triangleq\sum_{i=0}^{p}\hat{\gamma}'_{u\eta(i)}(k-i)\hat{G}(i)^{-1}\hat{\gamma}_{u\eta(i)}(-i)\\&=\hat{\psi}_{p-1}(k)+\hat{\gamma}'_{u\eta(p)}(k-p)\hat{G}(p)^{-1}\gamma_{u\eta(p)}(-p)\end{aligned}\tag{42}$$

$$\begin{aligned}\hat{\psi}_p''&\triangleq\sum_{i=0}^{p}\mathrm{Trace}\left[\hat{\gamma}_{\eta(i)}(0)\hat{G}(i)^{-1}\right]\\&=\hat{\psi}_{p-1}''+\mathrm{trace}\left[\hat{\gamma}_{\eta(p)}(0)\hat{G}(p)^{-1}\right].\end{aligned}\tag{43}$$

B. Simulation Results

The goal of this section is to investigate the validity of criterion (41), but the complete analysis of the performances of the filter obtained is out of the scope of this correspondence. Thus, we shall focus our attention only on the values of order obtained, and the application of the estimated filter to actual data will not be reported. Consequently, the role of the time-shift parameter δ essential in order to maintain an acceptable independence between the data processed and the estimated filter (see Section III-A), is of no influence; here we set $\delta=0$. The various parameters assumed in this section are as follows:

- scalar noise reference: $n=1$,
- sample size $=N=200$,
- zero time shift δ,
- zero signal s_t,
- autoregressive noise $z_t=\zeta_t$ with reflection coefficient α and normalized energy (Two whiteness levels are investigated: $\alpha=0.7$ or $\alpha=0.1$),
- F is of order $p=2$. (Two kinds of impulse response are envisaged: $F'=(1.,0.5,0.3)$ and $F'=(0.5,0.2,0.1)$. Note that vector F is not normalized.)

The combination of these values of α and F lead to four examples. We give here the four corresponding order histograms, each performed with 100 independent experiments. Features of the examples are recalled on each figure.

Figs. 1(a) and (b) show excellent results either in the case of a near white noise (Fig. 1(b)) or a more strongly coloured noise (Fig. 1(a)). This is due to the fact that the entries of F are sufficiently large: $F'=(1.,0.5,0.3)$. For smaller values of f_i, the number p of coefficients is more difficult to identify. This is shown in Figs. 1(c) and (d), where the last coefficient is 0.1. Actually, the greater the power ratio

$$\rho=\frac{\mathrm{power}\{F'Y\}}{\mathrm{power}\{\zeta_t\}},$$

the easier p to identify. With $F'=(0.5,0.2,0.1)$, ratio ρ is 4.5 times smaller than in the precedent cases.

Moreover, if the signal s_t is not zero, the ratio ρ decreases. So, the smaller the signal-to-noise ratio, the better performs our criterion (this fact is common to all noise cancellation procedures that work with noise subtracting).

Appendix

The case of exponential averaging evoked in formulas (11) and (12) has not been investigated yet. Since the previous results are of asymptotic nature, i.e., limiting results as N tends to infinity, they will not apply in the case of exponential averaging. However, an heuristic argument can be developed to obtain results for α close to 1. From (12), the estimate $\bar{F}$ can be written as

$$\bar{F} = F + \left[(1-\alpha)\sum_{k=0}^{\infty}\alpha^k Y_{t-k}Y_{t-k}'\right]^{-1}\left[(1-\alpha)\sum_{k=0}^{\infty}\alpha^k Y_{t-k}\zeta_{t-k}'\right]. \tag{44}$$

The second bracket in this expression has zero mean and covariance matrix

$$V = (1-\alpha)^2\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\alpha^{k+j}\Gamma_\zeta(k-j)\Gamma_Y(k-j)$$
$$= (1-\alpha)^2\sum_{i=-\infty}^{\infty}\left\{\sum_{j=\max(-i,0)}\alpha^{2j+i}\right\}\Gamma_\zeta(i)\Gamma_Y(i).$$

Or after some manipulations

$$V = \frac{1-\alpha}{1+\alpha}\sum_{i=-\infty}^{\infty}\alpha^{|i|}\Gamma_\zeta(i)\Gamma_Y(i). \tag{45}$$

On the other hand, the first factor in brackets in (44), $\Sigma\alpha^k Y_{t-k}Y_{t-k}'$, obviously has mean $\Gamma_Y(0)$ and covariance matrix W in which the generic term W^{ij} is

$$W^{ij} = (1-\alpha)^2\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\alpha^{k+l}\{\Gamma_Y^{ii}(k-l)\Gamma_Y^{jj}(k-l)$$
$$+\Gamma_Y^{ij}(k-l)\Gamma_Y^{ji}(k-l)$$
$$+\mathrm{Cum}\left(Y_t^i,Y_t^j,Y_{t+k-l}^i,Y_{t+k-l}^j\right)\}.$$

By the same argument as previously used, this equals

$$W^{ij} = \frac{1-\alpha}{1+\alpha}\sum_{l=-\infty}^{\infty}\alpha^{|l|}\left[\Gamma_Y^{ii}(l)\Gamma_Y^{jj}(l)+\Gamma_Y^{ij}(l)\Gamma_Y^{ji}(l)\right.$$
$$\left.+\mathrm{cum}\left(Y_t^i,Y_t^j,Y_{t+l}^i,Y_{t+l}^j\right)\right]. \tag{46}$$

Assume that the covariance function and the 4th order cumulant function of the process Y_t are absolutely summable. Then this expression can be bounded by constant $*(1-\alpha)/(1+\alpha)$ and, hence, tends to 0 as α tends to 1. Thus, $(1-\alpha)\Sigma\alpha^k Y_{t-k}Y_{t-k}'$ converges in mean square to $\Gamma_Y(0)$ as α tends to 1. Assimilating this latter expression as equal to $\Gamma_Y(0)$ yields estimator $\bar{F}$ to have covariance matrix

$$\Gamma(\bar{F}) = \frac{1-\alpha}{1+\alpha}\Gamma_Y^{-1}(0)\left[\sum_{i=-\infty}^{\infty}\alpha^{|i|}\Gamma_\zeta(i)\Gamma_Y(i)\right]\Gamma_Y^{-1}(0). \tag{47}$$

Thus, the same computations as in Section III might be carried out in the case of exponential averaging, by replacing the matrix in (14) by expression (47). Further, note that as α tends to 1:

$$\sum_{i=-\infty}^{\infty}\alpha^{|i|}\Gamma_\zeta(i)\Gamma_Y(i) \text{ tends to } \sum_{i=-\infty}^{\infty}\Gamma_\zeta(i)\Gamma_Y(i)$$

and thence for $\alpha \simeq 1$, the use of exponential averaging is roughly equivalent to the use of a rectangular window of size $N = 2/(1-\alpha)$ in the LLSE.

Acknowledgment

The authors wish to thank the referees for their helpful comments.

References

[1] H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst Statist. Math.*, vol. 21, pp. 243–247, 1969.

[2] P. Comon and J. L. Lacoume, "Noise reduction for an estimated wiener filter using noise references," *IEEE Trans. Inform. Theory*, vol IT-32, no. 2, pp. 310–313, Mar. 1986.

[3] P. Comon and D. T. Pham, "An error bound for a noise canceller," *IEEE Trans. Acoust., Speech Signal Processing*, pp. 1513–1517, Oct. 1989.

[4] E. J. Hannan, *Multiple Time Series.* New York: Wiley, 1970.

[5] G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications.* San Francisco: Holden Day, 1968.

[6] S. A. Kassam and T. L. Lim, "Robust wiener filters," *J. of Franklin Institute*, vol. 304, pp. 172–185, Oct./Nov. 1977.

[7] W. Kofman, A. Silvent, and J. Lienard, "Etude theorique et experimentale du systeme correlofiltre," *Annales des Telecom.*, vol. 37, no. 3-4, pp. 115–122, Mar. 1982.

[8] H. L. Van Trees, *Detection, Estimation and Modulation Theory.* Vol. I, New York: Wiley, 1968.

[9] P. Whittle, "On the fitting of multivariate autoregression and approximate factorization of a spectral density matrix," *Biometrika*, vol. 50, 1963, pp. 129–134.

[10] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1986.

[11] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems.* New York: Wiley, 1988.

[12] R. Shibata, "Selection of the number of regression variables: a minimax choice of generalized FPE," *Ann. Inst. Math.*, vol. 38, pp. 459–474, 1986.

A Nonlinear Optimum-Detection Problem—II: Simple Numerical Examples

T. T. KADOTA, FELLOW, IEEE

***Abstract*—Simple numerical examples are presented to illustrate the effect of the previously derived nonlinear filters for combating nonlinear Gaussian noise in detecting deterministic signals. The nonlinear Gaussian noise is expressed as a quadratic form in stationary Gaussian noise that is also present in the data together with white Gaussian noise. Thus the nonlinear noise is referred to as the quadratic noise and the stationary noise as the linear noise. The former is assumed to be an order of magnitude smaller than the latter. When the signal overlaps with both the linear and the quadratic noise, use of both nonlinear filters for the small quadratic-noise region improves the detection performance well beyond the optimum level achievable in the absence of the quadratic noise. As the quadratic noise increases, this improvement diminishes and the performance deteriorates eventually below the level achievable by the linear and the first nonlinear filter combination. These conclusions are drawn from the results of the Monte Carlo simulation with 10 000 random samples. Although the examples used here are extremely simple, these findings should be a useful guide in designing the nonlinear filters to enhance the detection performance of the conventional linear filter in the presence of the quadratic noise while remaining cost-effective.**

I. Introduction

In the companion paper [1] an approximately optimum detection statistic was derived for detecting a deterministic signal in the presence of linear and quadratic noise in addition to white background noise. By the quadratic noise, we mean a stochastic process which takes a quadratic form in the spectral elements (or frequency components) of a stationary Gaussian process. To distinguish from the quadratic noise, we call the stationary Gaussian process the linear noise. Obviously, the quadratic

Manuscript received October 7, 1988; revised August 14, 1989.

The author is with AT&T Bell Laboratories (2C356), Murray Hill, NJ 07974.

IEEE Log Number 8933847.

0018-9448/90/0300-0434$01.00 ©1990 IEEE