Consider now the introduction of a noise modulation $\psi(t)$ in addition to the periodic component. The new signal then becomes

$$v(t) = A_0 \cos [\omega_0 t + \mu \sin \omega_m t + \psi(t)]. \quad (3)$$

Middleton[1] has shown that the autocorrelation function, $\phi(\tau)$, of $v(t)$ is given by

$$\phi(\tau) = \frac{A_0^2}{2} e^{-D^2_0 \Omega(\tau)} J_0\left(2\mu \sin \frac{\omega_m \tau}{2}\right) \cos \omega_0 \tau, \quad (4)$$

where

$$\Omega(\tau) = \frac{1}{\pi} \int_0^\infty \frac{W_N(f) \sin^2 \frac{\omega\tau}{2}}{\omega^2} d\omega$$

$W_N(f)$ = Power density spectrum of the modulating noise voltage (watts per hertz).

A comparison of (2) and (4) shows that the noise modulation introduces a decreasing exponential as a factor into the autocorrelation function; hence the envelope of the autocorrelation function is no longer periodic. Only the peak at $\tau = 0$ has the value $A_0^2/2$. We are only interested in the amplitude of the correlation peaks at the values $\tau = 2\pi k/\omega_m$, since these are the largest peaks in the absence of noise.

Let $y = \exp [-D_0^2 \Omega(\tau)] = e^{-A}$. At $\tau = 0$, $\Omega(\tau) = 0$, and $y = 1$. For the case when the kth peak is down α dB from the peak at $\tau = 0$;

$$-\alpha = 20 \log_{10} \left(\frac{y}{1}\right) = 20 \log_{10} e^{-A} = -20A \log_{10} e$$

or

$$A = 0.1151\alpha$$

$$\therefore \frac{D_0^2}{\pi} \int_0^\infty \frac{W_N(f) \sin^2 \left(\frac{\pi k \omega}{\omega_m}\right) d\omega}{\omega^2} = 0.1151\alpha. \quad (5)$$

The following two power density spectra of the modulating noise voltage will be investigated:

1) the result of white Gaussian noise passed through an ideal low-pass filter with cutoff frequency ω_b

$$W_N(f) = \begin{cases} \frac{2\pi}{\omega_b} \text{ watts per hertz}, & 0 < \omega < \omega_b \\ 0, & \omega > \omega_b; \end{cases} \quad (6)$$

2) the result of white Gaussian noise passed through an RC low-pass filter whose 3 dB cutoff frequency is ω_b

$$W_N(f) = \frac{4\omega_b}{\omega^2 + \omega_b^2} \text{ watts per hertz}. \quad (7)$$

For the above two power spectra,

$$P = \int_0^\infty W_N(f)\, df = 1.$$

Substituting (6) into (5) and simplifying yields

$$\left(\frac{D_0}{\omega_b}\right)^2 [x \text{ Si } (x) + \cos x - 1] = 0.1151\alpha. \quad (8)$$

Substituting (7) into (5) and simplifying yields

$$\left(\frac{D_0}{\omega_b}\right)^2 [x - 1 + e^{-x}] = 0.1151\alpha \quad (9)$$

where $x = 2\pi k \omega_b/\omega_m$ and Si $(x) = \int_0^x \sin y/y\, dy$.

Figs. 1 and 2 are plots of (8) and (9), respectively, for $\alpha = 1$, 3, and 5 dB. For negative values of k simply use $|k|$. The case $x \leq 1$ and $x \geq 10$ are summarized in Table I. The maximum error introduced by these approximating equations is less than 5 percent.

H. J. Blinchikoff
Carlyle Barton Laboratory
The Johns Hopkins University
Baltimore, Md. 21218

[1] D. Middleton, *An Introduction to Statistical Communication Theory*. New York: McGraw-Hill, 1963, ch. 14.

Manuscript received February 16, 1967. This work was supported by the Air Force Office of Scientific Research under Contract F44620-67-C-0001.

On the Generation of All Spectral Factors

***Abstract*—Two solutions are presented to the problem of finding from one spectral factorization all spectral factorizations of a non-negative para-Hermitian matrix. The first solution results from the theory of equivalent networks while the second is given through a derivation from first principles.**

A problem of considerable interest in view of its applications to filtering theory, stability theory, network synthesis, and other areas of system theory is that of spectral factorization. Since one would like to investigate all possibilities, Kalman[1] recently has posed, and solved for small order, the problem of finding all spectral factorizations of nonminimal size. Here we present two slightly different but complete and simple solutions, the first following immediately from the known theory of equivalent circuits and the second following a development independent of network considerations for which we give the details.

One is given an $n \times n$ matrix $\mathbf{A}(p)$ each element being a rational function, with real coefficients, of the complex variable $p = \sigma + j\omega$ ($\mathbf{A}$ is called real-rational) with the properties that $\mathbf{A}$ is para-Hermitian, that is $\tilde{\mathbf{A}}_* = \mathbf{A}$ [where the tilde denotes matrix transposition and the subscript asterisk denotes replacement of p by $-p$ (Hurwitz conjugation)], and that $\mathbf{A}(j\omega)$ is positive semidefinite for almost all ω, written $\mathbf{A} \geq 0$ ($\mathbf{A} \geq 0$ is called non-negative). One wishes to find a (or all) factorization(s) $\mathbf{A} = \tilde{\mathbf{K}}_* \mathbf{K}$ such that $\mathbf{K}$ is real-rational (the factorization is then also called real-rational). Various methods are available for finding one particular factorization[2]–[5] (Also see Newcomb[6] p. 89.) and methods for passing among ones of minimal size, $\mathbf{A} = \tilde{\mathbf{J}}_* \mathbf{J}$ with dimension $\mathbf{J} = r \times n$, r = rank $\mathbf{A}$, are well known (Belevitch,[7] p. 307 and Youla,[2] p. 176).

Oono and Yasuura through the theory of equivalent networks (Oono and Yasuura,[8] p. 137) and Youla's subsequent amplification of their work (Youla,[2] p. 178) have put forward the following result which gives a method of finding all spectral factorizations given any one minimal factor $\mathbf{J}$. Here $\mathbf{1}_k$ is the $k \times k$ identity, $\mathbf{0}_{k-r,n}$ is the $(k - r) \times n$ zero matrix, and a $k \times k$ matrix $\mathbf{\Phi}(p)$ is called para-unitary if $\tilde{\mathbf{\Phi}}_* \mathbf{\Phi} = \mathbf{1}_k$.

Theorem 1

Let $\mathbf{A}(p)$ be an $n \times n$ real-rational matrix satisfying

$$\mathbf{A} = \tilde{\mathbf{A}}_* \geq 0 \quad (1)$$

and consider any two real-rational factorizations

$$\mathbf{A} = \tilde{\mathbf{J}}_* \mathbf{J} = \tilde{K}_* \mathbf{K} \quad (2)$$

having $\mathbf{J}$ and $\mathbf{K}$, respectively, $r \times n$ and $k \times n$ with $k \geq r$ = rank $\mathbf{A}$. Then there exists a $k \times k$ real-rational para-unitary $\mathbf{\Phi}$ such that

$$\mathbf{K} = \mathbf{\Phi} \begin{bmatrix} \mathbf{J} \\ \hline \mathbf{0}_{k-r,n} \end{bmatrix} \quad (3)$$

Proof: With slight changes in notation the proof is essentially given in Newcomb,[9] pp. 319, 186. The procedure can be outlined as a) forming $\mathbf{K}\mathbf{J}^{-1}$ by obtaining a right pseudo-inverse for $\mathbf{J}$, and b) extending the rectangular $\mathbf{K}\mathbf{J}^{-1}$ to the square $\mathbf{\Phi}$. Q.E.D.

The important conclusion to be reached from Theorem 1 is that given any one factorization of minimal size, $\mathbf{A} = \tilde{\mathbf{J}}_*\mathbf{J}$, we can find all others by forming

$$\mathbf{\Phi}\begin{bmatrix} \mathbf{J} \\ \hline \mathbf{0} \end{bmatrix}$$

where $\mathbf{0}$ ranges over all zero matrices of n columns and $\mathbf{\Phi}$ ranges over all (real-rational) para-unitary matrices. All $\mathbf{\Phi}$ can be readily generated through degree one or two factors (Newcomb,[9] p. 190) and a simple means of finding one $\mathbf{J}$ is through the Gauss factorization (such a $\mathbf{J}$ very simply yields $\mathbf{J}^{-1}$ mentioned in the theorem's proof). (See Newcomb,[9] p. 168.)

At this point we prove in detail an alternate result which is, however, a nontrivial but straightforward generalization of the result of Youla[2] (p. 176).

Lemma 1

Let $\mathbf{A}(p)$ be an $n \times n$ real-rational matrix of rank n almost everywhere, and satisfying (1). Let $\mathbf{J}(p)$ be an $n \times n$ real-rational matrix such that

$$\mathbf{A} = \tilde{\mathbf{J}}_*\mathbf{J}. \tag{4}$$

Then $\mathbf{K}(p)$ is a $k \times n$ matrix satisfying

$$\mathbf{A} = \tilde{\mathbf{K}}_*\mathbf{K} \tag{5}$$

if and only if there exists a $k \times n$ matrix $\mathbf{V}(p)$ with

$$\tilde{\mathbf{V}}_*\mathbf{V} = \mathbf{1}_n \tag{6}$$

and

$$\mathbf{K} = \mathbf{V}\mathbf{J}. \tag{7}$$

Proof: Suppose (4), (6), and (7) hold. That (5) is satisfied follows by direct calculation.

Conversely suppose (4) and (5) hold. Then it is easy to verify that a matrix $\mathbf{V}(p)$ satisfying (6) and (7) is given by

$$\mathbf{V} = \mathbf{K}\mathbf{J}^{-1}. \tag{8}$$

[Note that $J^{-1}(p)$ exists by the hypothesis on the rank of $A(p)$]. Q.E.D.

Lemma 1 may now be generalized to the situation where $\mathbf{A}(p)$ has rank $r < n$ almost everywhere.

Theorem 2

Let $\mathbf{A}(p)$ be an $n \times n$ real-rational matrix of rank r almost everywhere, and satisfying (1). Let $\mathbf{J}(p)$ be an $r \times n$ real-rational matrix such that

$$\mathbf{A} = \tilde{\mathbf{J}}_*\mathbf{J}. \tag{9}$$

Then the $k \times n$ matrix $\mathbf{K}(p)$ satisfies

$$\mathbf{A} = \tilde{\mathbf{K}}_*\mathbf{K} \tag{10}$$

if and only if there exists a $k \times r$ matrix $\mathbf{V}(p)$ such that

$$\tilde{\mathbf{V}}_*\mathbf{V} = \mathbf{1}_r \tag{11}$$

and

$$\mathbf{K} = \mathbf{V}\mathbf{J}. \tag{12}$$

Proof: Assuming that (9), (11), and (12) hold, (10) follows by direct calculation.

To prove the converse we assume that (9) and (10) hold. Let $\mathbf{X}(p)$ be a nonsingular $n \times n$ matrix such that

$$\mathbf{J}\mathbf{X} = [\hat{\mathbf{J}} \mid \mathbf{0}_{r,n-r}] \tag{13}$$

where $\hat{\mathbf{J}}$ is a nonsingular $r \times r$ matrix. Then immediately

$$\tilde{\mathbf{X}}_*\tilde{\mathbf{K}}_*\mathbf{K}\mathbf{X} = \begin{bmatrix} \tilde{\hat{\mathbf{J}}}_*\hat{\mathbf{J}} & \mathbf{0}_{r,n-r} \\ \hline \mathbf{0}_{n-r,r} & \mathbf{0}_{n-r,n-r} \end{bmatrix} \tag{14}$$

on using (9), (10), and (13). A simple argument using the nonnegativity of $\mathbf{A}(j\omega)$ then implies

$$\mathbf{K}\mathbf{X} = [\hat{\mathbf{K}} \mid \mathbf{0}_{k,n-r}] \tag{15}$$

for some $\hat{\mathbf{K}}(p)$. Using Lemma 1 we conclude the existence of a $k \times r$ matrix $\mathbf{V}(p)$ satisfying

$$\tilde{\mathbf{V}}_*\mathbf{V} = \mathbf{1}_r \tag{11}$$

and

$$\hat{\mathbf{K}} = \mathbf{V}\hat{\mathbf{J}}, \tag{16}$$

or, using (13) and (15)

$$\mathbf{K} = \mathbf{V}\mathbf{J} \tag{12}$$

which completes the proof. Q.E.D.

Notice that since $\mathbf{J}(p)$ has rank r (following from the rank of $\mathbf{A}(p)$), it has a right inverse satisfying

$$\mathbf{J}\mathbf{J}^{-1} = \mathbf{1}_r. \tag{17}$$

Thus the matrix $\mathbf{V}$ which relates any two matrices $\mathbf{J}$ and $\mathbf{K}$ satisfying (9) and (10) is simply, by (12)

$$\mathbf{V} = \mathbf{K}\mathbf{J}^{-1}. \tag{18}$$

It is not, however, immediate, but a nontrivial fact, that

$$\mathbf{V}\mathbf{J} = \mathbf{K}$$

if $\mathbf{V}$ is defined by (18). This is because $\mathbf{J}^{-1}$ is merely a right inverse of $\mathbf{J}$.

We point out that the difference between the two theorems is in the fact that $\mathbf{\Phi}$ is nonsingular while $\mathbf{V}$ generally is not. The two results can be reconciled by noting that any $\mathbf{V}$ can be extended to a (nonsingular) para-unitary $\mathbf{\Phi}$ by known means (Newcomb,[9] p. 186). Theorem 1 has the advantage that the structure of para-unitary matrices is well investigated (Oono and Yasuura,[8] pp. 138–142) while Theorem 2 has the advantage of working with smaller matrices.

For filtering purposes it is often of interest to find all state variable realizations for $(\mathbf{K}p)$ that is, all constant matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ and integers m such that

$$\mathbf{K}(p) = \mathbf{D} + \mathbf{C}(p\mathbf{1}_m - \mathbf{A})^{-1}\mathbf{B} \tag{19}$$

All realizations of $\mathbf{K}$ can be found using previous theories.[10]

R. W. Newcomb
Stanford Electronics Labs.
Stanford, Calif.
B. D. O. Anderson
Dept. Elec. Engrg.
Newcastle University
Newcastle, Australia

References

[1] R. E. Kalman, "On the realization of passive linear systems," presented at the 1966 Joint Berkeley-Stanford Systems Theory Seminar, Berkeley, Calif.
[2] D. C. Youla, "On the factorization of rational matrices," *IRE Trans. Information Theory*, vol. IT-7, pp. 172–189, July 1961.
[3] M. C. Davis, "Factoring the spectral matrix," *IEEE Trans. Automatic Control*, vol. AC-8, pp. 296–305, October 1963.

[4] B. D. O. Anderson, "Spectral factorization by algebra," Stanford Electronics Labs., Stanford, Calif., Tech. Rept. 6560-6, September 1966.
[5] R. W. Brockett, "Spectral factorization of rational matrices," *IEEE Trans. Information Theory* (to be published).
[6] R. W. Newcomb, "A Bayard-type nonreciprocal n-port synthesis," *IEEE Trans. Circuit Theory*, vol. CT-10, pp. 85–90 March 1963.
[7] V. Belevitch, "Synthèse des réseaux électriques passifs à n paires de bornes de matrice de répartition prédéterminée (in French), *Ann. Télécommun.*, vol. 6, pp. 302–312, November 1951.
[8] Y. Oono and K. Yasuura, "Synthesis of finite passive $2n$-terminal networks with prescribed scattering matrices," *Mem. Fac. of Engrg., Kyushu University*, vol. 14, pp. 125–177, May 1954.
[9] R. W. Newcomb, *Linear Multiport Synthesis*. New York: McGraw-Hill, 1966.
[10] B. D. O. Anderson, R. W. Newcomb, R. E. Kalman, and D. C. Youla, "Equivalence of linear time-invariant dynamical systems," *J. Franklin Inst.*, vol. 281, pp. 371–378, May 1966.

Recursive Computations for the Optimal Tracking of Time-Varying Parameters

***Abstract*—Recursive relations are given for updating the conditional density $p(\theta_k \mid X_{k-1}, \cdots X_1)$ (also for $p(\theta_k \mid X_k, \cdots, X_1)$), where θ_k is a parameter of the density of X_k. The observations $X_1, X_2, \cdots$ are assumed to be conditionally independent (i.e., for known parameters), and the sequence of time-varying parameters $\theta_1, \theta_2, \cdots$ constitutes a Markov-M sequence. The result requires the storage of an intermediate function of $(\theta_{k-1}, \cdots, \theta_{k-M})$.**

In the problem of Bayes optimal unsupervised pattern classification in the presence of an unknown time (or sample-to-sample) varying parameter, and many other adaptive problems, the need arises for repeatedly computing the conditional density of the current value of the parameter, conditioned on the sequence of past observations. Likelihood functions, or other decision statistics which depend on this time-varying parameter can then be averaged over the updated conditional density of the current value of the time-varying unknown parameter in order to make optimal decisions.

In the following, a recursive solution is given for updating the sequence of conditional densities as new data is received, under the assumptions that the parameters constitute a realization of a Markov-M sequence, and the observations are conditionally independent. The solution is also compared to one given by Fralick,[1] and shown to be a correction to his result for $M > 1$.

Let $X_1, X_2, \cdots$, be a sequence of conditionally independent samples from a density $p(X \mid \theta)$, where the characterizing parameter θ is time-varying as these samples are taken, that is, X_k is from $p(\cdot \mid \theta_k)$, $k = 1, 2, \cdots$. The conditional independence can be expressed in two equivalent ways, as follows. If $\lambda_k \triangleq (X_k, \cdots, X_1)$,

$$p(\lambda_k \mid \theta_k, \cdots, \theta_1) = \prod_{i=1}^{k} p(X_i \mid \theta_i), \tag{1a}$$

or

$$p(X_k \mid \theta_j, \cdots, \theta_1, \lambda_{k-1}) = p(X_k \mid \theta_k), \tag{1b}$$

for any $j \geq k$, where $p(X_1 \mid \theta_j, \cdots, \theta_1, \lambda_0) \triangleq p(X_1 \mid \theta_j, \cdots, \theta_1)$. The parameter sequence, $\theta_1, \theta_2, \cdots$, being Markov-$M$ means that whenever $k > M$,

$$p(\theta_k \mid \theta_{k-1}, \cdots, \theta_1) = p(\theta_k \mid \theta_{k-1}, \cdots, \theta_{k-M}). \tag{2}$$

Under assumptions (1)and (2) a recursive expression for $p(\theta_k \mid \lambda_{k-1})$ (or alternately for $p(\theta_k \mid \lambda_k)$) is desired. It does not seem possible to obtain an expression for $p(\theta_k|\lambda_{k-1})$ directly in terms of $p(\theta_{k-1}|\lambda_{k-2})$, except for $M = 1$. We can, however, find a direct recursion relation for the intermediate function $p(\theta_k \cdots, \theta_{k-M+1} \mid \lambda_{k-1})$, and then obtain $p(\theta_k \mid \lambda_{k-1})$ from it by "integrating out" the $M - 1$ other variables:

Manuscript received January 31, 1967; revised October 2, 1967. This work was supported by the Joint Services Electronics Program under Grant AF-AFOSR-766-67.

$$p(\theta_k \mid \lambda_{k-1}) = \int \cdots \int p(\theta_k, \cdots, \theta_{k-M+1} \mid \lambda_{k-1}) \cdot d\theta_{k-1} \cdots d\theta_{k-M+1}. \tag{3}$$

The recursive expression for $p(\theta_k, \cdots, \theta_{k-M+1} \mid \lambda_{k-1})$ is

$$p(\theta_k, \cdots, \theta_{k-M+1} \mid \lambda_{k-1}) = \frac{p(X_{k-1} \mid \theta_{k-1})}{p(X_{k-1} \mid \theta_{k-1})} \int p(\theta_{k-1}, \cdots, \theta_{k-M} \mid \lambda_{k-2}) \cdot p(\theta_k \mid \theta_{k-1}, \cdots, \theta_{k-M})\, d\theta_{k-M}. \tag{4}$$

Similarly, $p(\theta_k \mid \lambda_k)$ can be obtained from $p(\theta_k, \cdots, \theta_{k-M+1} \mid \lambda_k)$;

$$p(\theta_k \mid \lambda_k) = \int \cdots \int p(\theta_k, \cdots, \theta_{k-M+1} \mid \lambda_k) \cdot d\theta_{k-1} \cdots d\theta_{k-1} \cdots d\theta_{k-M+1}, \tag{5}$$

where

$$p(\theta_k, \cdots, \theta_{k-M+1} \mid \lambda_k) = \frac{p(X_k \mid \theta_k)}{p(X_k \mid \lambda_{k-1})} \int p(\theta_{k-1}, \cdots, \theta_{k-M} \mid \lambda_{k-1}) \cdot p(\theta_k \mid \theta_{k-1}, \cdots \theta_{k-M})\, d\theta_{k-M} \quad \text{(equation (1))} \tag{6}$$

The proof of (4) and (6) are similar, and are straightforward after three key facts are realized, namely that the conditional independence of the X_i, given the characterizing parameters (equation (1)), implies

$$p(X_i \mid \theta_k, \cdots, \theta_j, \lambda_{i-1}) = p(X_i \mid \theta_i), \tag{7}$$

and

$$p(\theta_k, \cdots, \theta_j \mid \lambda_i) = \frac{p(X_i \mid \theta_i)}{p(X_i \mid \lambda_{i-1})}\, p(\theta_k, \cdots, \theta_j \mid \lambda_{i-1}) \tag{8}$$

where $k \geq i \geq j$, and that (1) and (2) together imply

$$p(\theta_k \mid \theta_{k-1}, \cdots, \theta_{k-M}, \lambda_{k-1}) = p(\theta_k \mid \theta_{k-1}, \cdots, \theta_{k-M}). \tag{9}$$

Remarks: Equation (6) shows that $p(\theta_k, \cdots, \theta_{k-M+1} \mid \lambda_k)$ can be found from 1) $p(\theta_{k-1}, \cdots, \theta_{k-M} \mid \lambda_{k-1})$ (the function one step back) 2) the new data through $p(X_k \mid \theta_k)$, and 3) the transition law $p(\theta_k \mid \theta_{k-1}, \cdots, \theta_{k-M})$ of the Markov-M parameter sequence. $p(X_k \mid \lambda_{k-1})$ is simply a normalizing factor which makes the density integrate to one, and can be found by integrating the unnormalized function, or omitted in applications where the scale factor is irrelevant. Similar remarks apply to (4).

Note that in both of the above recursion formulas (4) and (6) a function of M variables ($p(\theta_k, \cdots, \theta_{k-M+1} \mid \lambda_{k-1})$ or $p(\theta_k, \cdots, \theta_{k-M+1} \mid \lambda_k)$, respectively) must be stored between steps. (It is not normally necessary to store $p(X_i \mid \theta_i)$ and $p(\theta_k \mid \theta_{k-1}, \cdots, \theta_{k-M})$ in the same sense because they may often be generated from some fixed rules such as exponentiation, etc.) It would be much simpler if, instead of having to store multidimensional functions, $p(\theta_k \mid \lambda_{k-1})$ could be found from $\{p(\theta_{k-1} \mid \lambda_{k-2}), \cdots, p(\theta_{k-M} \mid \lambda_{k-M-1})\}$. Fralick[1] has given an expression for doing just this, but unfortunately, it does not seem to be valid for $M > 1$. His expression is

$$p(\theta_k \mid \lambda_{k-1}) = \int \cdots \int \left\{ \prod_{i=1}^{M} \frac{p(X_{k-i} \mid \theta_{k-i})}{p(X_{k-i} \mid \lambda_{k-i-1})}\, p(\theta_{k-i} \mid \lambda_{k-i-1}) \right\} \cdot p(\theta_k \mid \theta_{k-1}, \cdots, \theta_{k-M})\, d\theta_{k-1} \cdots d\theta_{k-M}. \tag{10}$$