Covariance Factorization via Newton–Raphson Iteration

BRIAN D. O. ANDERSON, FELLOW, IEEE

***Abstract*—The solution of an integral equation arising in a covariance factorization problem is obtained by a Newton–Raphson iteration that is almost always globally convergent. Interpretations of the iterates are given, and the result is shown to specialize to known algorithms when the covariance is stationary with a rational Fourier transform.**

I. INTRODUCTION

LET K: $[0,T] \times [0,T] \to R^{m\times m}$ be an $m \times m$ matrix of continuous functions of two real variables with the property that K is a self-adjoint integral kernel with $\delta I + K$ positive definite, i.e.,

$$K(t,s) = K'(s,t), \tag{1}$$

where the prime denotes transpose here and hereafter, and

$$\int_0^T u'(t)u(t)\,dt + \int_0^T \int_0^T u'(t)K(t,s)u(s)\,dt\,ds > 0 \tag{2}$$

for all continuous $u(\cdot)$ with $\int_0^T u'(t)u(t)\,dt > 0$. We write these conditions as $K = K^a$ and $\delta I + K > 0$.

It is sometimes of interest to obtain a Volterra kernel $k(\cdot,\cdot)$: $[0,T] \times [0,T] \to R^{m\times m}$, with $k(t,s)$ zero for $t < s$ and continuous for $0 \le s \le t \le T$, such that

$$\delta I + K = (\delta I + k)(\delta I + k^a). \tag{3}$$

This is shorthand for

$$K(t,s) = k(t,s) + k'(s,t) + \int_0^{\min(s,t)} k(t,\sigma)k'(s,\sigma)\,d\sigma \tag{4}$$

when $t \neq s$, with obvious modification at $t = s$. This decomposition of $\delta I + K$ always exists [1]–[5].

An application to stochastic processes of this decomposition was perceived some time ago, e.g., see Lévy [1] who noted that finding k from K is equivalent to the problem of finding the impulse response of a causal linear invertible system with a vector of unit intensity independent white noise inputs, given that the output covariance of the system is $\delta I + K$. The stochastic process overtones are also described in various papers by Kailath [2]–[4]. There are also applications to control theory, at least if one poses the problem of finding k from K with the order of the two factors on the right of (3) reversed, but we shall not dwell on these here.

Essentially the same technique for solving (4) is due to Lévy [1] and to Gohberg and Krein. See [5] for an accessible treatment of their work, which was actually done earlier than the publication date of [5] might suggest. This technique is described in Section II below.

The equation $\delta I + K = (\delta I + k^a)(\delta I + k)$, which can obviously be treated in much the same way, also arises in the inverse Sturm–Liouville problem [6], [7]. Here, much the same technique is suggested for its solution. In both the covariance factorization and inverse Sturm–Liouville problems, the fixed limit on the integral in (4), or its Sturm–Liouville equivalent, is taken as 0 for convenience in any finite interval problem; however, there do exist problems in which the range of integration is sensibly infinite, in which case the solution existence theory becomes more complex. For example, if $K(t,s)$ is stationary, i.e., $K(t,s) = K(t-s)$, then it is of interest to seek a stationary $k(t,s) = k(t-s)$. For such a $k(\cdot,\cdot)$ to exist, the lower integration limit in (4) must be replaced by $-\infty$, and it is also conventional to assume that $K(t)$ possesses a Fourier transform with certain specified behavior at infinity.

The purpose of this paper is to describe an alternative route to the computation of k: a sequence $k_0, k_1, k_2, \cdots$ is constructed via Newton–Raphson iteration, and convergence to k under the mildest of conditions on k_0 is established in the finite interval case. The details are set out in Section III.

We might first ask why this result should be expected. It is well-known (and easily checked) that a Newton–Raphson iteration can be used to find the square root of a positive number, i.e., to solve $x^2 = a$, and that this iteration converges for an arbitrary positive initial iterate. It is perhaps less well-known that similar results hold for

Manuscript received March 10, 1977; revised May 20, 1977. This work was supported by the Australian Research Grants Committee.

The author is with the Department of Electrical Engineering, University of Newcastle, New South Wales, 2308, Australia.

0018-9448/78/0300-0183$00.75 © 1978 IEEE

more sophisticated types of factorization of positive quantities. For example, Kleinman [8] has a procedure for finding a solution of an algebraic Riccati equation arising in linear optimal control theory by Newton–Raphson iteration. This equation is intimately bound up with a factorization of a (positive definite) power spectrum in that a solution of the equation defines a spectrum factorization, and conversely. Of course, power spectrum factorization is essentially equivalent to performing the factorization (4) with K and k stationary. Again, Tunnicliffe Wilson [9] has described the use of a Newton–Raphson iteration for the factoring of matrix spectral densities that is abstractly equivalent to Kleinman's method. The method for factoring (4) that we derive can in fact be specialized to obtain the iterative equations of Tunnicliffe Wilson when K and k are time-invariant and have rational Laplace transforms; this is shown later in the paper.

One might also ask why an iterative method theoretically requiring an infinite number of iterations to converge should be of interest when there is available a "finite" solution (as one might regard that of Lévy or Gohberg and Krein). The general answer to this question is that numerical analysts have often found that iterative methods with good convergence properties are preferable to "finite" methods for doing certain tasks. (Particular examples include, as noted above, two specializations of the covariance or spectral factorization task considered here.) A more specific answer may be provided here by considering certain properties of the two schemes. The iterative method requires the construction of resolvent kernels of only Volterra kernels, whereas the "finite" method involves the construction of resolvent kernels of non-Volterra kernels, i.e., the solution of Fredholm equations of the second kind; these latter are more difficult to solve than Volterra equations. Be that as it may, it would seem unwise in view of the sparsity of evidence to claim that either method is definitively superior. This is so especially since at the moment we have little feel for the numerical stability in implementations of either algorithm.

II. Lévy–Gohberg–Krein Solution

The method of [1], [5] proceeds in two steps. In the first, one computes, for each t and s, a Volterra kernel $h(\cdot,\cdot)$ via

$$h(t,s) + \int_0^t h(t,u)K(u,s)\,du = K(t,s), \qquad 0 \le s \le t \le T, \quad (5)$$

which we might write as

$$h + [hK]_+ = K_+ \quad (6)$$

where $K_+(t,s)$ is $K(t,s)$ for $t \ge s$ and is zero otherwise. The property $\delta I + K > 0$ ensures that $h(\cdot,\cdot)$ is well defined. The second step requires computation of the resolvent kernel of h and yields the desired kernel k; specifically,

$$k(t,s) - h(t,s) = \int_s^t h(t,u)k(u,s)\,du, \qquad 0 \le s \le t \le T \quad (7)$$

or

$$(\delta I - h)(\delta I + k) = \delta I. \quad (8)$$

For a full interpretation of these steps, the references should be consulted, see especially [4]. We can formally check the truth of the claim as follows. Equation (6) states that $[h - K + hK]_+ = 0$, or that $h - K + hK = m^a$ for some Volterra m. Accordingly,

$$\begin{aligned}(\delta I - h)(\delta I + K)(\delta - h^a) &= [\delta I - h + K - hK][\delta I - h^a] \\ &= [\delta I - m^a][\delta I - h^a] \\ &= \delta I - n^a\end{aligned}$$

for some Volterra n. Because $(\delta I - h)(\delta I + K)(\delta I - h^a)$ is self-adjoint, n^a must be zero. Use of (8) then establishes that $(\delta I + k)(\delta I + k^a) = \delta I + K$.

From the computational point of view, it should be recognized that (5) will be solved to determine a function of s for each fixed value of t, i.e., one solves the Fredholm equation

$$\phi_t(s) + \int_0^t \phi_t(u)K(u,s) = K(t,s)$$

for each fixed $t \in [0,T]$. A basic method of inverting this equation involves approximation of the integral by a sum, and replacement of the equation by a matrix equation $\boldsymbol{Ax} = \boldsymbol{b}$ where the unknown $\boldsymbol{x}$ contains values of $\phi_t(s)$ at various s in the interval $[0,t]$. Equation (7) on the other hand is solved to determine a function of t for each fixed s, i.e.,

$$\psi_s(t) - h(t,s) = \int_s^t h(t,u)\psi_s(u)\,du.$$

Approximation by a sum here is much more appealing, since the resulting matrix equation involves a triangular matrix and is thus readily solvable. One might also conceive of using a Neumann series to solve this equation, since the Volterra property guarantees convergence. For material on solving integral equations, see [10], [11] and, for the properties of the Neumann series, see, e.g., [12].

III. Newton–Raphson Solution

The idea behind a Newton–Raphson solution to (3) is the following. Suppose that k_i is an approximation to k. Then we seek Δk_i such that

$$(\delta I + k_i + \Delta k_i)(\delta I + k_i^a + \Delta k_i^a) = \delta I + K.$$

However, we agree to neglect second-order terms, and thus we select Δk_i to satisfy

$$(\delta I + k_i)(\delta I + k_i^a) + \Delta k_i(\delta I + k_i^a) + (\delta I + k_i)\Delta k_i^a = \delta I + K. \quad (9)$$

Rewriting this as

$$(\delta I + k_i)^{-1}\Delta k_i + \Delta k_i^a(\delta I + k_i^a)^{-1} = (\delta I + k_i)^{-1}(\delta I + K)(\delta I + k_i^a)^{-1} - \delta I,$$

we see that

$$\Delta k_i = (\delta I + k_i)L_{i+} \quad (10)$$

where

$$L_i = (\delta I + k_i)^{-1}(\delta I + K)(\delta I + k_i^a)^{-1} - \delta I. \quad (11)$$

The main result is then as follows.

Theorem: Let $\delta I + K$ be a positive definite continuous self-adjoint kernel on $[0,T] \times [0,T]$. Let $k_0(\cdot,\cdot)$ be an arbitrary Volterra kernel on $[0,T] \times [0,T]$ with $k_0(t,s)$ continuous for $0 \le s \le t \le T$. For each i, define Δk_i and L_i by (10) and (11), respectively, and define $k_{i+1} = k_i + \Delta k_i$. Then, for each t,s in $[0,T]$, $k_i(t,s) \to k(t,s)$.

The proof will proceed with several lemmas.

Lemma 1: Suppose that the hypotheses of the theorem hold. Then, for $i \ge 1$,

$$L_i = L_i^a \le 0. \quad (12)$$

(This means that $L_i(t,s) = L_i'(s,t)$ and that $\int_0^T \int_0^T u'(t)L_i(t,s)u(s)\,dt\,ds \le 0$ for all continuous $u(\cdot)$.)

Proof: From (9), we have by an easy calculation

$$(\delta I + k_{i+1})(\delta I + k_{i+1}^a) = (\delta I + k_i + \Delta k_i)(\delta I + k_i^a + \Delta k_i^a) = \delta I + K + \Delta k_i \Delta k_i^a. \quad (13)$$

Hence, for $i \ge 0$, definition (11) yields

$$L_{i+1} = (\delta I + k_{i+1})^{-1}(\delta I + K)(\delta I + k_{i+1}^a)^{-1} - \delta I = -(\delta I + k_{i+1})^{-1}\Delta k_i \Delta k_i^a (\delta I + k_{i+1}^a)^{-1},$$

and thus $L_{i+1} \le 0$, since it is of the form mm^a for some Volterra m.

Lemma 2: Suppose that the hypotheses of the theorem hold. Define a Volterra kernel h_i by $(\delta I + h_i) = (\delta I + k_i)^{-1}(\delta I + k)$. Then, for all t and $i \ge 1$,

$$h_i(t,t) + h_i'(t,t) \le h_{i+1}(t,t) + h_{i+1}'(t,t) \le 0. \quad (14)$$

Proof: Evidently, for $i \ge 1$,

$$(\delta I + h_i)(\delta I + h_i^a) = (\delta I + k_i)^{-1}(\delta I + K)(\delta I + k_i^a) = L_i + \delta I \le \delta I,$$

the inequality following from Lemma 1. Immediately, we have $h_i + h_i^a + h_i h_i^a \le 0$ or $h_i + h_i^a \le -h_i h_i^a \le 0$. This implies that $h_i(t,t) + h_i'(t,t) \le 0$ for all t.

Next, observe that the definition of h_i forces

$$h_i(t,s) + k_i(t,s) + \int_s^t h_i(t,u)k_i(u,s)\,du = k(t,s)$$

for $t \ge s$. Consequently, using the continuity of $h_i(\cdot,\cdot)$ and $k_i(\cdot,\cdot)$,

$$k_i(t,t) + h_i(t,t) = k(t,t),$$

so that

$$h_{i+1}(t,t) - h_i(t,t) = -\Delta k_i(t,t).$$

From (10), we see that $\Delta k_i(t,t) + \Delta k_i'(t,t) = 2L_i(t,t)$. Because $L_i \le 0$, $L_i(t,t) \le 0$. Therefore

$$h_{i+1}(t,t) + h_{i+1}'(t,t) - h_i(t,t) - h_i'(t,t) = -2L_i(t,t) \ge 0. \quad (15)$$

Lemma 3: Suppose that the hypotheses of the theorem hold. Then, for all $s,t \in [0,T]$, $\lim_{i\to\infty} L_i(t,s) = 0$.

Proof: From Lemma 2 we see that $\lim_{i\to\infty}[h_i(t,t) + h_i'(t,t)]$ exists, and from (15) it follows that $\lim_{i\to\infty} L_i(t,t) = 0$. Since $L_i \le 0$ by Lemma 1, we have for each t and s

$$\begin{bmatrix} L_i(t,t) & L_i(t,s) \\ L_i'(t,s) & L_i(s,s) \end{bmatrix} \le 0.$$

Now consider the 2×2 minor of this matrix formed from the $j-j$ element of $L_i(t,t)$, the $k-k$ element of $L_i(s,s)$, and the appropriate entries of $L_i(t,s)$ and $L_i'(t,s)$. Then, nonpositivity of the matrix implies that

$$L_i^{jj}(t,t)L_i^{kk}(s,s) - [L_i^{jk}(t,s)]^2 \ge 0.$$

The convergence of $L_i(t,t)$ and $L_i(s,s)$ as $i \to \infty$ shows that $L_i(t,s) \to 0$ as $i \to \infty$.

We can now complete the proof of the theorem. If we know that $k_i(t,s)$ is bounded uniformly for all i, t, and s, then (10) and Lemma 3 yield $\Delta k_i(t,s) \to 0$ for all t,s so that $k_i(t,s)$ approaches a limit $\bar{k}(t,s)$. It is then easily checked that one must have $\bar{k} = k$. Now $k(t,s)$ is bounded for all t and s, and thus the bound on $k_j(t,s)$ will follow if one can be demonstrated on $h_i(t,s)$. The technical details establishing this bound now follow.

Since $(\delta I + h_i)(\delta I + h_i^a) = L_i + \delta I$, it follows that $h_i + h_i^a + h_i h_i^a \to 0$ as $i \to \infty$. Examining this on the diagonal, we find that

$$\tfrac{1}{2}[h_i(t,t) + h_i'(t,t)] + \int_0^t h_i(t,\alpha)h_i'(t,\alpha)\,d\alpha \to 0,$$

and using (14), we conclude that for all suitably large i and all t,

$$\left|\left| \int_0^t h_i(t,\alpha)h_i'(t,\alpha)\,d\alpha \right|\right| < M$$

for some constant M. It follows by the Schwarz inequality that

$$h_i h_i^a = \int_0^{\min(t,s)} h_i(t,\sigma)h_i(s,\sigma)\,d\sigma$$

is bounded for all t, s, and thus, since $h_i + h_i^a + h_i h_i^a \to 0$, that h_i is bounded. This completes the proof of the theorem.

Several points should be noted.

1) Equation (13) illustrates the quadratic convergence property of the Newton–Raphson procedure.

2) It would seem that, by dropping the continuity property on K but imposing an $\mathcal{L}_2$ property (in which case k still exists [5]), variations on the above proof would again establish convergence of the method.

3) In [5], the factorization of nonself-adjoint kernels $\delta I + K$ is examined. Again, Newton–Raphson equations can be developed. However, global convergence from arbitrary initial iterates cannot be established, although local convergence can. Thus, suppose we seek Volterra k and l such that

$$(\delta I + k)(\delta I + l^a) = \delta I + K, \quad (16)$$

and suppose that ith approximants k_i and l_i are available. Then one finds

$$(\delta I + k_i)(\delta I + l_i^a) + \Delta k_i(\delta I + l_i^a) + (\delta I + k)\Delta l_i^a = \delta I + K,$$

or

$$(\delta I + k_i)^{-1}\Delta k_i + \Delta l_i^a(\delta I + l_i^a)^{-1} = (\delta I + k_i)^{-1}(\delta I + K)(\delta I + l_i^a)^{-1} - \delta I = L_i$$

and

$$\Delta k_i = (\delta I + k_i)L_{i+}, \qquad \Delta l_i^a = L_{i-}(\delta I + l_i^a).$$

Moreover,

$$L_{i+1} = -(\delta I + k_{i+1})^{-1}\Delta k_i \Delta l_i^a (\delta I + l_{i+1}^a)^{-1}.$$

One obtains a local convergence result by taking $\|\Delta k_i\|$ and $\|\Delta l_i^a\|$ suitably small in some norm and arguing then that $\|\Delta k_{i+r}\|$ and $\|\Delta l_{i+r}^a\|$ must decay to zero suitably fast as $r \to \infty$.

4) Certain of the quantities arising in the iteration described have interpretations in terms of stochastic process concepts. Let ϵ_t be a vector white noise process passing into a linear system with impulse response $\delta(t-s)I + k(t,s)$. If the output of this system is z_t, then $E[z_t z_s'] = \delta(t-s)I + K(t,s)$. Now suppose that z_t is the input process to a linear system with impulse response $(\delta I + k_i)^{-1}$. Denote the output of this system by w_t^i. Then $E[w_t^i w_s^{i'}] = \delta(t-s) + L_i(t,s)$. So the "error kernel" $L_i(t,s)$ encountered in the iteration is a measure of the departure from pure whiteness of the process obtained by using $(\delta I + k_i)^{-1}$ as a whitening filter for the process z_t rather than the correct whitening filter $(\delta I + k)^{-1}$. The iterative procedure causes the processes $w_t^i, w_t^{i+1}, \cdots$ to approach closer and closer to white processes. Such an idea lies behind a self-adjusting Kalman filter as described in [13].

The iterative equations themselves also have a stochastic interpretation. Suppose one wants to whiten w_t^i. Let $L_{i+} = l_i$. Then, with L_i suitably small in norm, one has $\delta I + L_i \approx (\delta I + l_i)(\delta I + l_i^a) \approx (\delta I - l_i)^{-1}(\delta I - l_i^a)^{-1}$, i.e., $(\delta I - l_i)^{-1}$ will be an approximate whitening filter for w_t^i. This means that $(\delta I + k_{i+1})^{-1} \triangleq (\delta I - l_i)^{-1}(\delta I + k_i)^{-1}$ will be an approximate whitening filter for z_t. This relation among k_i, l_i, and k_{i+1} can be verified to be the same as (10).

IV. Specializations of the Recursive Equations

In this section, we indicate the effect of assuming the kernel K to be either cyclostationary, i.e., $K(t,s) = K(t+\Delta, s+\Delta)$ for some fixed Δ and all t and s, or to be stationary, i.e., $K(t,s) = K(|t-s|)$. We assume that the lower limit of integrations in the basic factorization equation (4) becomes $-\infty$, and that the behavior of K, for large values of its argument, is such that there exist, respectively, cyclostationary and stationary $k(\cdot,\cdot)$ satisfying (4).

One can check that, in the Lévy–Gohberg–Krein method, the kernels k and h inherit the property of K, and that, in the Newton–Raphson method, the kernels k_i inherit the property *provided that k_0 possesses the property.*

Let us concentrate further on the stationary case. We suppose that two-sided Laplace transforms exist. Then the Newton–Raphson iteration leads to

$$[1 + \mathcal{L}(k_i)]^{-1}\mathcal{L}(\Delta k_i) = \{[I + \mathcal{L}(k_i)]^{-1}[I + \mathcal{L}(K)][I + \mathcal{L}(k_i^a)]^{-1}\}_+ - I \quad (17)$$

where $\{\cdot\}_+$ denotes the operation of $\mathcal{L}\{1(t)\mathcal{L}^{-1}(\cdot)\}$, i.e., taking the casual part; here, $1(t)$ is the unit step function. Of course, there is a technical question here of establishing the existence of the various transforms and their limiting properties. Now, in case $\mathcal{L}(K)$ is rational, we obtain a recursion due to Tunnicliffe Wilson [9]. To avoid confusion of notation, let K be scalar, with $\mathcal{L}[K] = m(s)/(n(s)n(-s))$; here, $m(s)$ is an even polynomial with, in order to ensure $\delta I + K > 0$, $n(j\omega)n(-j\omega) + m(j\omega) > 0$ for all real ω. Of course, $n(s)$ is also polynomial. Suppose that $n(s)$ has all roots in Re $[s] < 0$; an inductive argument will establish that $\mathcal{L}[k_i] = (p_i(s))/n(s)$ for some $p_i(s)$ if one takes $\mathcal{L}[k_0] = (p_0(s))/n(s)$.

The recursion for $p_{i+1}(s)$ is, with an asterisk subscript on a polynomial denoting negation of the indeterminate,

$$\left(1 + \frac{p_i}{n}\right)^{-1}\frac{\Delta p_i}{n} = \left\{\left(1 + \frac{p_i}{n}\right)^{-1}\left(1 + \frac{m}{nn_*}\right)\left(1 + \frac{p_{i*}}{n_*}\right)^{-1}\right\}_+ - 1,$$

or

$$\frac{\Delta p_i}{n + p_i} + \frac{\Delta p_{i*}}{n_* + p_{i*}} = \frac{m + nn_*}{(n + p_i)(n_* + p_{i*})} - 1,$$

or

$$p_{i+1}(n_* + p_{i*}) + p_{i+1*}(n + p_i) = m + p_i p_{i*}. \quad (18)$$

Equation (18) is equivalent to the recursion suggested by Tunnicliffe Wilson. In unpublished work, we have shown that this iteration is simply a frequency domain version of a better-known iteration scheme based on state-space ideas due to Kleinman [8].

V. Conclusions

We have stressed in the introduction that a cautious approach should be adopted in deciding whether the Lévy–Gohberg–Krein approach is to be preferred to the Newton–Raphson approach to spectral factorization. It is quite possible that the application and ultimate numerical form of the answer might dictate a preference, e.g., whether tables of numbers or weight coefficients for basis functions are required. Again, it is possible that particular properties of $K(\cdot,\cdot)$ might be relevant; if, for example, $K(\cdot,\cdot)$ were stationary (or cyclostationary), one method might prove preferable to the other.

References

[1] P. Lévy, *Processus Stochastiques et Mouvement Brownien*, 2nd ed. Paris: Gauthier-Villars, 1964.
[2] T. Kailath, "Application of a resolvent identity to a linear smoothing problem," *SIAM J. Contr.*, vol. 7, pp. 68–74, Feb. 1969.
[3] T. Kailath, "Fredholm resolvents, Wiener–Hopf equations, and

Riccati differential equations," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 665–672, Nov. 1969.
[4] T. Kailath, "A view of three decades of filtering theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 146–181, Mar. 1974.
[5] I. C. Gohberg and M. G. Krein, *Theory and Applications of Volterra Operators in Hilbert Space*, (in Russian). Moscow: Nauka, 1967 (Transl.: *Amer. Math. Soc.*, Providence, RI, 1970).
[6] I. M. Gel'fand and B. M. Levitan, "On the determination of a differential equation from its spectral function," *Amer. Math. Soc. Transl.*, (2), vol. 1, pp. 253–304, 1955.
[7] M. R. Wohlers, *Lumped and Distributed Passive Networks*. New York: Academic, 1969.
[8] D. L. Kleinman, "On the iterative technique for Riccati equation computation," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 114–115, Feb. 1968.
[9] G. Tunnicliffe Wilson, "The factorisation of matrical spectral densities," *SIAM J. Appl. Math.*, vol. 23, pp. 420–426.
[10] B. Noble, "The numerical solution of nonlinear integral equations and related topics," in *Nonlinear Integral Equations*, P. M. Anselone, Ed. Madison, WI: Univ. Wisconsin Press, 1964, pp. 215–318.
[11] L. Fox and E. T. Goodwin, "The numerical solution of nonsingular linear integral equations," *Philosophical Trans. Royal Soc.*, vol. 245, pp. 501–534, 1953.
[12] F. Riesz and B. Sz-Nagy, *Functional Analysis*. New York: Unger, 1955.
[13] B. Carew and P. R. Belanger, "Identification of optimum filter steady-state gain for systems with unknown noise covariance," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 582–587, Dec. 1973.