Projecting Filters for Recursive Prediction of Discrete-Time Processes

By ALLEN GERSHO and DAVID J. GOODMAN

(Manuscript received June 19, 1970)

We consider the design of time-invariant recursive filters of constrained order for one-step prediction of discrete-time stationary processes. For this purpose, we introduce the projecting-filter concept. An nth-order projecting filter for a given process has the characterizing property that with the process as input, the output at each instant is the optimal linear combination of the n previous output and n latest input samples. This definition implies that (i) the filter is stable, (ii) any $n + 1$ consecutive samples of the prediction error sequence are mutually uncorrelated, (iii) the mean-square prediction error is at least as low as that of the best nth order nonrecursive predictor, and (iv) if the spectral density of the process is rational of order 2n or less, then the nth-order projecting filter coincides with the optimal (unconstrained) linear predictor.

A design algorithm for nth-order projecting filters iteratively generates successive sets of coefficients of a time-varying nth-order recursive filter which asymptotically approaches the desired time-invariant filter. The only input data needed for the algorithm are the autocovariance coefficients of the process to be predicted. When the order of the filter is matched to the order of the process, the time-varying filter is the same as the Kalman predictor. The algorithm has yielded effective projecting filters for several specific processes. Our results indicate that near optimal prediction may often be obtained with filters of order lower than that of the optimal unconstrained predictor.

I. INTRODUCTION

Although the optimal linear predictor of a random process must make use of the entire past of the process, any practical predictor can store only a finite number of data. One way to design a finite storage predictor is to determine the best linear combination of the n latest sample values of the process. However, for many processes, a large value of n is required to achieve a performance quality approaching that of the unconstrained optimal linear predictor. An alternate approach is to find the best recursive predictor constrained to operate only on the n latest data samples and the n latest predictions. This approach has the advantage of using condensed information from the entire past of the process with the consequence that optimal or near optimal prediction can often be achieved with a relatively small amount of storage.

The purpose of this paper is to introduce the projecting-filter approach to recursive prediction and to present an algorithm for the design of projecting filters that has yielded effective low-order predictors not otherwise attainable. So far, a complete theory of projecting filters has not been established. We do not yet know how broad is the class of processes which possess projecting filters of a given order; nor have we determined the class of processes for which our design algorithm is effective. However, we can report very favorable experience in the design of projecting filters for a variety of specific processes. We have also established some important theoretical properties of projecting filters.

1.1 *Optimal and Finite Memory Predictors*

In certain special cases the optimal (least mean-square error) unconstrained predictor is realizable with a finite-storage filter.[1] In particular, for an nth-order autoregressive, or wide-sense Markov, process the optimal unconstrained predictor is a finite-memory nonrecursive filter operating only on the n latest data samples. More generally, the optimal unconstrained predictor of any stationary process whose spectral density is rational of order $2n$ may be implemented as an nth-order recursive filter. The characteristics of the optimal filter may be determined by applying the discrete-time form of Wiener's spectral factorization technique. Even more generally, consider any nonstationary process which can be modeled as the response of an nth-order linear time-varying recursive filter to an uncorrelated noise input. The optimal unconstrained predictor is an nth-order time-varying recursive filter[2] which may be determined by use of the Kalman filtering equations,[3] or, more efficiently, by a generalization of the approach taken in Section VI of this paper.

If a random process cannot be modeled as the response of an nth-order recursive filter to an uncorrelated input, then the optimal unconstrained one-step linear predictor cannot be realized by an nth-order filter. Nevertheless, it is realistic to preselect the desired order,

n, of the predictor and to seek the best recursive filter of this order. In this way the structure of the predictor is conveniently specified for digital filter implementation while only the $2n$ parameter values need be supplied according to the process to be predicted. Unfortunately, with the least-mean-square error criterion, the constrained-order prediction problem is a special case of the unsolved problem of L_2 rational approximation on the unit circle.[4] No analytical solution is known and optimization search techniques are severely hampered by the multimodal nature of the error surface.*

1.2 *Projecting Filters*

In this paper we introduce the projecting filter principle of recursive prediction. Although the projecting filter is not a solution of the L_2 rational approximation problem, it has the local optimality property that at each step it forms the best linear combination of the available data. The term "projection" alludes to the geometrical interpretation of random variables as vectors in Hilbert space.[6, 7] Each prediction error of the projecting filter is a vector orthogonal to the n most recent inputs and the n previous errors. Hence the projecting filter performs a partial whitening of the input process. In this sense it approximates the action of the optimum unconstrained predictor, the error of which is a white-noise process—the innovations process of the input. If the input can be represented as the response of an nth-order filter to white noise, the nth-order projecting filter is the optimum unconstrained predictor. For any process, the mean-square error of a projecting filter is never greater than the mean-square error of the optimum nonrecursive filter of the same order. Projecting filters are stable.

1.3 *An Example*

These properties of projecting filters are observed in the example of the eighth-order process $\{x_k\}$ represented by

$$x_k = \epsilon_k - 0.8\epsilon_{k-1} + 0.5\epsilon_{k-2} + 0.25\epsilon_{k-3} - 0.6\epsilon_{k-4} - 0.2\epsilon_{k-5} + 0.1\epsilon_{k-6} + 0.4\epsilon_{k-7} - 0.08\epsilon_{k-8}$$

in which $\{\epsilon_k\}$ is a stationary white-noise process with zero mean and unit variance. The power spectral density function of $\{x_k\}$ has zeros at the 16 points in the z-plane indicated in Fig. 1. The eighth-order projecting filter for $\{x_k\}$, which is the optimum unconstrained predic-

* The complexity of the error as a function of the parameters is evidenced by the work of R. S. Phillips[5] on the corresponding continuous-time problem.

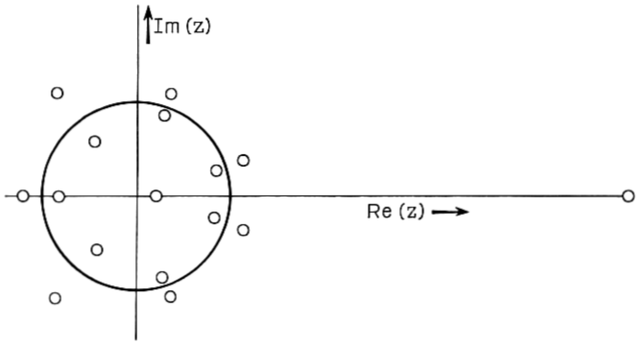

Fig. 1—Locations of zeros of the spectral density function of an eighth-order process. The eighth-order projecting filter has poles at the zero locations that are outside the unit circle.

tor, has poles at the eight locations indicated in Fig. 1 that are outside the unit circle. The pole positions of a seventh-order projecting filter are shown in Fig. 2. There are poles extremely close to all of the locations outside the unit circle indicated in Fig. 1, except the one furthest from the origin. Figures 3, 4, and 5 indicate the pole locations of the sixth-, third-, and first-order projecting filters, respectively. The poles of these filters do not coincide with zeros of the power spectral density function of $\{x_k\}$.

Figure 6 demonstrates the projecting-filter mean-square-error performance for this process. Here the horizontal base line is the optimal unconstrained prediction error. The white bars indicate errors of optimal constrained order nonrecursive predictors and the shaded bars are

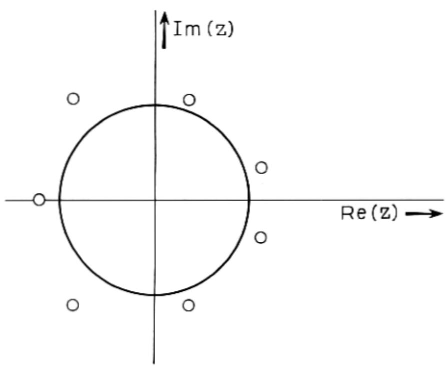

Fig. 2—Pole locations of seventh-order projecting filter.

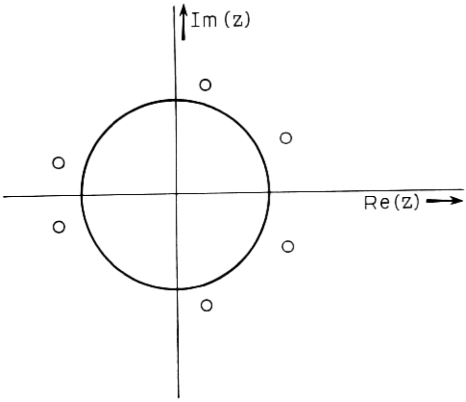

Fig. 3—Pole locations of sixth-order projecting filter.

the errors of the projecting filters. It is significant that the error of the seventh-order projecting filter is extremely close to the optimum linear-prediction error; the ratio of the two errors is approximately $1 + 10^{-7}$. By using the projecting filter approach to prediction, we have discovered a means of reducing predictor complexity with virtually no loss in accuracy. In addition, Fig. 6 shows the error resulting from low-order recursive filters and the advantages relative to nonrecursive prediction.

1.4 *Organization of the Paper*

The content of the paper falls into two categories. Some sections contain descriptive and analytic material relevant to predictors and

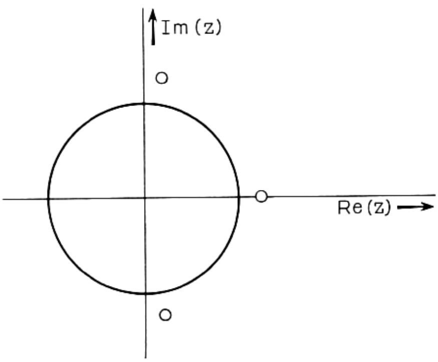

Fig. 4—Pole locations of third-order projecting filter.

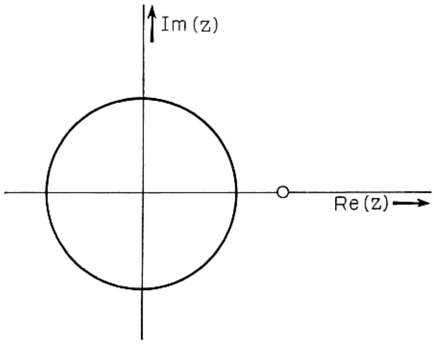

Fig. 5—Pole location of first-order projecting filter.

projecting filters in general and other sections pertain to the particular design method that has been used in synthesizing the predictors described in Section 1.3. Sections II, III and IV are in the first category; they define the prediction problem and the projecting-filter principle and focus attention on the essential properties of unconstrained predictors and projecting filters. Section V introduces the design method, an iterative scheme based upon successive projections in Hilbert space. This technique leads to a time-varying filter that asymptotically tends towards the desired projecting filter. Section VI shows that when the

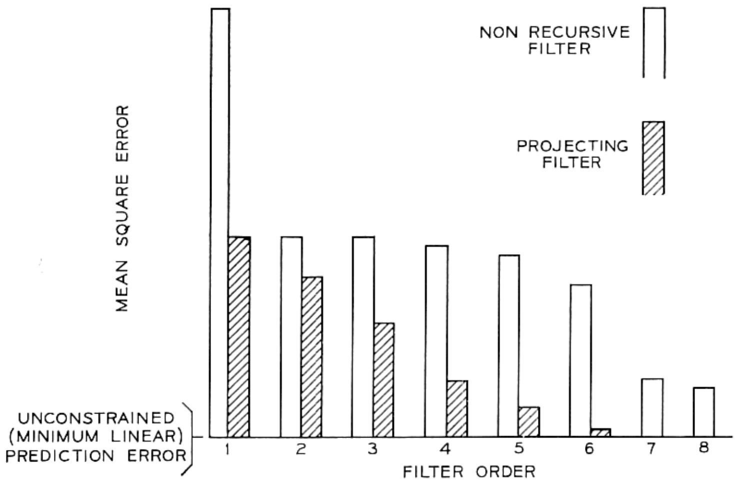

Fig. 6—Mean-square errors of projecting filters and optimal nonrecursive filters of orders 1 through 8.

order of the filter is matched to that of the process, the design algorithm converges and the projecting-filter approach results in an efficient analysis and design (equivalent to but simpler than the Kalman filtering equations) for the unconstrained optimum time-varying filter with a given initial state. Section VII presents a derivation of the design algorithm.

II. PROBLEM STATEMENT

We consider a purely-nondeterministic* stationary process $\{x_k\}$ with known covariance function, $r_k = Ex_ix_{i+k}$. We assume that the spectral density function of the process $f(z) = \Sigma r_k z^k$ has no zeros on the unit circle, $|z| = 1$. The purpose of this paper is to describe a new approach to the design of a stable one-step predicting filter with the nth-order recursive structure

$$y_k = \sum_{i=0}^{n-1} a_i x_{k-i} + \sum_{i=1}^{n} b_i y_{k-i}. \tag{1}$$

A natural measure of the performance of the predictor is the mean-square value of the prediction error

$$e_{k+1} = x_{k+1} - y_k. \tag{2}$$

Because the determination of the optimum filter coefficients with respect to this criterion is an intractable problem of approximation theory, our design method is based on a different performance objective. Rather than synthesize the least-squares nth-order recursive filter, we seek a stable time-invariant filter with the following

Projecting property: With input $\{x_k\}$, the output, y_k, is, at each instant k, the least mean-square linear combination of the data $\{x_k, x_{k-1}, \cdots, x_{k-n+1}, y_{k-1}, \cdots, y_{k-n}\}$ currently in the filter memory.

This implies that the filter coefficients a_i and b_i satisfy a set of linear equations involving the covariance functions of $\{x_k\}$ and $\{y_k\}$. The autocovariance of $\{x_k\}$ corresponds to the given data of the prediction problem but the cross-covariance between $\{x_k\}$ and $\{y_k\}$ and the autocovariance of $\{y_k\}$ are transcendental functions of a_i and b_i. It follows that an explicit solution for the coefficients from the constraints imposed by the projecting property is not possible. An algorithmic solution is presented in Section VII.

---

* See Ref. 1, p. 23.

III. UNCONSTRAINED PREDICTION

We refer to the problem defined in Section II as a constrained-order prediction problem because the order, n, of the predictor is prespecified. Another problem, which we refer to as unconstrained linear prediction, has received considerable attention in the literature of stochastic processes.[1,8] The optimum unconstrained prediction, $\hat{x}_{k+1}$, of x_{k+1} is the least mean-square linear combination of the entire past, $x_k, x_{k-1}, \cdots$ of $\{x_k\}$. In the terminology of the Hilbert space description of random variables, $\hat{x}_{k+1}$ is called the projection of x_{k+1} into the past of $\{x_k\}$, and we thus adopt the following convenient notation:

$$\hat{x}_{k+1} = P\{x_{k+1} \mid x_k, x_{k-1}, \cdots\}. \tag{3}$$

When $\{x_k\}$ is gaussian, the projection coincides with the conditional expectation.

3.1 *The Error Process*

The error process $\{v_k\}$, defined by

$$v_{k+1} = x_{k+1} - \hat{x}_{k+1}, \tag{4}$$

is the innovations process of $\{x_k\}$. It has the key orthogonality properties:

$$Ev_{k+1}x_{k-i} = 0, \qquad i = 0, 1, 2, \cdots; \tag{5}$$

$$Ev_{k+1}v_{k-i} = 0, \qquad i = 0, 1, 2, \cdots. \tag{6}$$

Equation (5), which characterizes the projection operation, indicates that the best linear predictor cannot make better use of the past of $\{x_k\}$. Equation (6), a direct consequence of equation (5), shows that the error process is white noise.

3.2 *Stability*

The optimal unconstrained prediction, $\hat{x}_{k+1}$, may be characterized as the limit of an infinite sequence of constrained-order nonrecursive predictions:

$$\hat{x}_{k+1} = \lim_{n \to \infty} \sum_{i=0}^{n-1} h_{in} x_{k-i} \tag{7}$$

where $h_{in}(i = 0, 1, \cdots, n-1)$ are the coefficients of the optimum nth-order nonrecursive predictor which may be calculated by means of well-known quadratic minimization techniques. The unconstrained predictor is a stable function of the data in the sense that

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} h_{in}^2 < \infty . \tag{8}$$

This is proved in Section IV.

3.3 *Process Representation*

We say $\{x_k\}$ is of nth-order if it can be represented as the response of a stable recursive nth-order filter to white noise so that

$$\sum_{i=0}^{n} \alpha_i x_{k-i} = \sum_{i=0}^{n} \beta_i \epsilon_{k-i} \tag{9}$$

in which α_n or β_n is nonzero, $\{\epsilon_k\}$ is a white-noise process, and $\Sigma\alpha_i z^i$ has no zeros in $|z| \leqq 1$. If $\{x_k\}$ is of order n, it is known that there exists an nth-order recursive filter which generates $\{\hat{x}_k\}$ in response to $\{x_k\}$. The error process of this filter is $\{v_k\}$, the innovations process of $\{x_k\}$. If $\Sigma\beta_i z^i \neq 0$ for $|z| \leqq 1$, then $v_k = \epsilon_k$.

Conversely, if $\{x_k\}$ does not possess an nth-order representation of the form of equation (9), the best unconstrained predictor cannot be realized by an nth-order filter. To prove this we assume that such a realization does exist. That is, we assume

$$\hat{x}_{k+1} = \sum_{i=0}^{n-1} d_i x_{k-i} + \sum_{i=1}^{n} c_i \hat{x}_{k+1-i}. \tag{10}$$

This combined with equation (4) implies

$$x_{k+1} - \sum_{i=0}^{n-1} (d_i + c_{i+1}) x_{k-i} = v_{k+1} + \sum_{i=0}^{n-1} c_{i+1} v_{k-i} \tag{11}$$

which shows that $\{x_k\}$ is in fact the response of an nth-order filter to the white-noise process $\{v_k\}$, which is a contradiction.

IV. PROJECTING FILTERS

4.1 *Orthogonality Properties*

We have shown that an nth-order recursive filter cannot perform optimal unconstrained linear prediction of a process of order greater than n. With such a process as input, the error process $\{e_k\}$, of an nth-order filter will necessarily have a higher mean-square value than that of the innovations process and $\{e_k\}$ will fail to meet the orthogonality conditions of equations (5) and (6). However, when the nth-order predictor possesses the projecting property defined in Section II, its error process satisfies some but not all of the orthogonality conditions met by innovations process. In particular, the projecting property requires that

$$y_k = P\{x_{k+1} \mid x_k, x_{k-1}, \cdots, x_{k-n+1}, y_{k-1}, \cdots, y_{k-n}\} \tag{12}$$

which is characterized by the orthogonality conditions

$$Ee_{k+1}x_{k-i} = 0, \qquad i = 0, 1, \cdots, n-1; \tag{13}$$

$$Ee_{k+1}e_{k-i} = 0, \qquad i = 0, 1, \cdots, n-1. \tag{14}$$

Note that in this case, equation (14) is not a direct consequence of equation (13). In fact equation (13) is satisfied by the error of the optimum nth-order nonrecursive filter, while equation (14) is not satisfied by this error unless $\{x_k\}$ is an nth-order autoregression, that is, an nth-order process with $\beta_i = 0$ for $i > 0$.

4.2 *Stability*

Projecting filters are inherently stable. In fact, some kind of stability property is implicit in any statement of steady-state properties of a time-invariant filter. In this paper we say that a filter is stable if its impulse response is square summable, which implies if the spectrum is rational, that the filter transfer function is analytic on and in the unit circle. We assume that the predicting filter has zero in each memory element prior to $k = 0$ at which time $\{x_k\}$ is applied to the input. The projecting property stated in Section II implies that in the limit as $k \to \infty$, y_k tends toward the projection indicated in equation (12). Thus in the limit, the orthogonality conditions of equations (13) and (14) are satisfied from which it follows that $Ee_{k+1}y_k \to 0$ and since $y_k + e_{k+1} = x_{k+1}$,

$$\lim_{k\to\infty} [Ey_k^2 + Ee_{k+1}^2] = Ex_{k+1}^2 = r_0$$

from which we infer

$$\limsup_{k\to\infty} Ey_k^2 < r_0. \tag{15}$$

We also know that the filter output for each $k \geqq 0$ is the finite sum

$$y_k = \sum_{i=0}^{k} g_i x_{k-i} \tag{16}$$

in which g_i is the filter impulse response. Equations (15) and (16) imply the existence of a positive number c, which bounds the mean-square output:

$$Ey_k^2 < c, \qquad \text{for all } k. \tag{17}$$

The existence of this bound leads to the following

Theorem: If a filter with impulse response g_i is a projecting filter, it is stable in the sense that

$$\sum_{i=0}^{\infty} g_i^2 < \infty. \tag{18}$$

Proof: In terms of $f(z)$, the power spectral density function of $\{x_k\}$, and the frequency transfer function of the filter we have

$$Ey_k^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{m=0}^{k} g_m e^{j\omega m} \right|^2 f(e^{j\omega})\, d\omega \geqq \lambda \sum_{m=0}^{k} g_m^2, \tag{19}$$

in which $\lambda = \min_{|z|=1} f(z) > 0$ according to the assumption stated in Section II. Equations (17) and (19) may be combined in the expression

$$\sum_{m=0}^{k} g_m^2 < c/\lambda, \quad \text{for all } k, \tag{20}$$

from which equation (18) follows.

The same reasoning leads to a proof of the stability of the unconstrained predictor. Replacing g_i is h_{in}, the impulse response of the nonrecursive predictor described in Section 3.3.

V. PROJECTING-FILTER DESIGN APPROACH

As we stated in Section II, an attempt to determine the filter coefficients by directly combining equation (1) and equations (13) and (14) leads to an intractable set of transcendental equations relating the coefficients and the autocovariance function of $\{x_k\}$. On the other hand, the iterative approach introduced in this paper leads to the computation of the desired coefficients by means of standard operations of arithmetic and matrix algebra.

Our design method results in a time-varying filter which, starting with zero in all memory elements, sequentially predicts $x_1, x_2, \cdots$ according to the projecting principle. At each step the filter forms the optimum linear combination of the available data.

Thus we define the process $\{x'_k\}$ such that

$$\begin{aligned} x'_k &= 0, \qquad k < 0; \\ x'_k &= x_k, \qquad k \geqq 0; \end{aligned} \tag{21}$$

and we adopt as our prediction of x_{k+1},

$$\begin{aligned} y_k &= 0, && k < 0; \\ y_k &= P\{x_{k+1} \,|x'_k, x'_{k-1}, \cdots, x'_{k-n+1}, y_{k-1}, \cdots, y_{k-n}\}, && k \geqq 0. \end{aligned} \tag{22}$$

Equation (22) uniquely defines the time-varying linear transformation which generates the nonstationary process $\{y_k\}$ from the stationary process $\{x_k\}$.

At each step the prediction error of the time-varying filter meets the orthogonality conditions of equations (13) and (14) so that $Ee_{k+1}y_k = 0$ and therefore $Ey_k^2 < r_0$ for all k. Following the proof of the theorem in Section 4.2 we can show that with the filter output represented by

$$y_k = \sum_{i=0}^{k} g_{ik}x_{k-i}, \qquad k \geqq 0, \tag{23}$$

the time-varying filter possesses the stability property

$$\limsup_{k\to\infty} \sum_{i=0}^{k} g_{ik}^2 < \infty. \tag{24}$$

Furthermore, if this filter approaches the time-invariant projecting filter with impulse response g_i in the sense that

$$\lim_{k\to\infty} \sum_{i=0}^{k} (g_{ik} - g_i)^2 = 0, \tag{25}$$

we are assured that this filter is stable and that it has the desired nth-order recursive structure. Hence if we determine, for each k, a_{ik} and b_{ik} such that

$$y_k = \sum_{i=0}^{n-1} a_{ik}x'_{k-i} + \sum_{i=1}^{n} b_{ik}y_{k-i} \tag{26}$$

is equivalent to equation (22), then successive computation of these coefficients leads to the desired time-invariant projecting filter.

Note that although y_k is uniquely determined by equation (22), the coefficients a_{ik} and b_{ik} in the representation of equation (26) are not unique when the set of stored data is linearly dependent. This situation is analyzed in Section 7.4.

VI. MATCHED-ORDER PROCESSES

We prove in Section 6.1 that when $\{x_k\}$ is of order n, the projecting-filter design technique results in least mean-square time-varying prediction in the sense that each output y_k is the optimum linear combination of the entire observed past of $\{x_k\}$. Thus y_k is equal to the output of the optimal nonrecursive filter of order $k + 1$ as described in Section 3.2 so that

$$\lim_{k\to\infty} (y_k - \hat{x}_{k+1})^2 = 0,$$

indicating that the design algorithm converges to the optimal unconstrained predictor. In Section 6.2, we derive simple formulas for the filter coefficients generated by the design procedure.

6.1 *Optimality*

We denote by $\mathcal{H}_k$ the subspace spanned by the random variables in the filter memory at time k: $x'_k, x'_{k-1}, \cdots, x'_{k-n+1}, y_{k-1}, \cdots, y_{k-n}$; and we denote by $\mathcal{R}_k$ the subspace spanned by the observed past of $\{x_k\}$: $x_k, x_{k-1}, \cdots, x_0$. Note that another spanning set of $\mathcal{R}_k$ is e_k, $e_{k-1}, \cdots, e_1, x_0$, where $\{e_k\}$ is the error sequence of the projecting filter. This statement follows by induction since x_0 spans $\mathcal{R}_0$ and if $\{e_j, e_{j-1}, \cdots, e_1, x_0\}$ spans $\mathcal{R}_j$ then $\{e_{j+1}, e_j, \cdots, e_1, x_0\}$ spans $\mathcal{R}_{j+1}$ because $e_{j+1} = x_{j+1} - y_j$ with y_j in $\mathcal{R}_j$.

In this section we assume that $\{x_k\}$ is an nth-order process represented by equation (9) with $\alpha_0 = 1$ so that

$$x_{j+1} = u_{j+1} - \sum_{i=1}^{n} \alpha_i x_{j+1-i} \tag{27}$$

in which $\{u_k\}$ is the moving average process with

$$u_{j+1} = \sum_{i=0}^{n} \beta_i \epsilon_{j+1-i} \tag{28}$$

and $\{\epsilon_k\}$ is a unit-mean-square white-noise process. Because $\Sigma\alpha_i z^i \neq 0$ for $|z| \leqq 1$, equation (27) may be expressed in the form

$$x_{k+1} = \sum_{i=0}^{\infty} h_i \epsilon_{k+1-i}, \tag{29}$$

in which $\{h_i\}$ is square summable. Equation (29) shows that

$$E\epsilon_{k+1}x_{k-i} = 0, \qquad i \geqq 0, \tag{30}$$

and equations (28) and (30) imply

$$Eu_{k+1}x_{k-i} = 0, \qquad i \geqq n. \tag{31}$$

If we let x^*_{k+1} denote the optimal "growing-memory" prediction of x_{k+1} with the projection characteristic

$$x_{k+1}^* = P\{x_{k+1} \mid \mathcal{R}_k\},$$

we have the following

Theorem: At each instant k, the time-varying filter output defined by

$$y_k = P\{x_{k+1} \mid \mathcal{H}_k\}$$

is the optimal growing-memory predictor in the sense that

$$y_k = x_{k+1}^* . \tag{32}$$

Proof: We will show that $x_{k+1}^* \in \mathcal{H}_k$ which implies equation (32) because $\mathcal{H}_k \subset \mathcal{R}_k$. Clearly, for $0 \leqq k < n$, $\mathcal{H}_k = \mathcal{R}_k$ so that $y_k = x_{k+1}^*$. We assume $y_k = x_{k+1}^*$ for all $k < j$ and show that this implies $y_j = x_{j+1}^*$. Hence, by induction, equation (32) is valid for all k.

Let $j \geqq n$ and assume equation (32) holds for all $k < j$. Then

$$Ee_{k+1}x_{k-i} = 0, \qquad \text{for} \quad k = 0, 1, \cdots, j-1; \quad i = 0, 1, \cdots, k. \tag{33}$$

This implies that the vectors $e_j, e_{j-1}, \cdots, e_1, x_0$, which span $\mathcal{R}_j$ are mutually orthogonal. Thus a projection into $\mathcal{R}_j$ is the sum of the projections into each of these basis vectors. In particular

$$P\{u_{j+1} \mid \mathcal{R}_j\} = P\{u_{j+1} \mid x_0\} + \sum_{i=0}^{j-1} P\{u_{j+1} \mid e_{j-i}\}. \tag{34}$$

Now note that $e_{j-i} \in \mathcal{R}_{j-i}$ and that equation (31) states that $u_{j+1} \perp \mathcal{R}_{j-i}$ for $i \geqq n$. Thus the first term in equation (34) and all but the first n terms of the summation are zero so that

$$P\{u_{j+1} \mid \mathcal{R}_j\} = \sum_{i=0}^{n-1} P\{u_{j+1} \mid e_{j-i}\}. \tag{35}$$

We now consider x_{j+1}^* by noting that the projection operator is linear and that $P\{x_{k-i} \mid \mathcal{R}_k\} = x_{k-i}$ for $i = 0, 1, \cdots, k$. Thus equation (27) implies

$$x_{j+1}^* = P\{x_{j+1} \mid \mathcal{R}_j\} = P\{u_{j+1} \mid \mathcal{R}_j\} - \sum_{i=1}^{n} \alpha_i x_{j+1-i} \tag{36}$$

or, from equation (35)

$$x_{j+1}^* = \sum_{i=0}^{n-1} P\{u_{j+1} \mid e_{j-i}\} - \sum_{i=1}^{n} \alpha_i x_{j+1-i}. \tag{37}$$

Note that the ith term in the first summation is proportional to e_{j-i}

so that x_{j+1}^* is a linear combination of $x_j, x_{j-1}, \cdots, x_{j-n+1}, y_{j-1}, \cdots, y_{j-n}$, the basis vectors of $\mathcal{H}_j$. Thus $x_{j+1}^* \in \mathcal{H}_j$ and

$$x_{j+1}^* = P\{x_{j+1} \mid \mathcal{H}_j\} = y_j .$$

Hence $x_{k+1}^* = y_k$ for all k. Q.E.D.

6.2 *Filter Coefficients*

In this section we derive explicit recursions for the coefficients and mean-square error of the optimal growing-memory predictor of a stationary nth-order process. We begin with equation (37) for the optimal prediction and observe that the projections have the form

$$P\{u_{k+1} \mid e_{k-i}\} = \gamma_{ik} e_{k-i} , \qquad i = 0, 1, \cdots, n-1, \tag{38}$$

where the coefficients are ratios of two expectations,

$$\gamma_{ik} = Eu_{k+1}e_{k-i}/Ee_{k-i}^2 . \tag{39}$$

These expectations may be expressed as functions of the auto-covariance coefficients,

$$\varphi_i = Eu_k u_{k-i} , \tag{40}$$

of the stationary moving average process $\{u_k\}$.

Our derivation begins with the expression of the error at step k, $e_{k+1} = x_{k+1} - x_{k+1}^*$, as the difference between equation (27) for x_{k+1} and equation (37) for x_{k+1}^*:

$$e_{k+1} = u_{k+1} - \sum_{i=0}^{n-1} \gamma_{ik} e_{k-i}. \tag{41}$$

Squaring equation (41) and taking the expectation we obtain

$$Ee_{k+1}^2 = \varphi_0 - \sum_{i=0}^{n-1} \gamma_{ik}^2 Ee_{k-i}^2 \tag{42}$$

which gives the mean-square error at step k in terms of current filter coefficients and past errors. To find the next set of coefficients, $\gamma_{i,k+1}$, we express e_{k+1-i} as in equation (41) and we find the expected product of this random variable and u_{k+2}. Then we divide by the mean-square indicated in equation (39) with the result

$$\gamma_{n-1,k+1} = \varphi_n / Ee_{k+2-n}^2 ,$$

$$\gamma_{i,k+1} = \left[\varphi_{i+1} - \sum_{j=0}^{n-i-2} \gamma_{j,k-i} \gamma_{i+j+1,k+1} Ee_{k-i-j}^2 \right] \Big/ Ee_{k+1-i}^2 ,$$

$$i = n-2, n-3, \cdots, 0, \tag{43}$$

where the upper limit on the sum is a consequence of the property, $Eu_{k+2}\, e_{k-i-j} = 0$ for $j \geqq n - i - 1$. [See equation (31)].

The filter coefficients a_{ik} and b_{ik} of equation (26) are related to the projection coefficients γ_{ik} and the autoregressive coefficients, α_i, of the process representation by

$$\begin{aligned} a_{ik} &= \gamma_{ik} - \alpha_{i+1} , \\ b_{ik} &= -\gamma_{i-1,k} , \end{aligned} \tag{44}$$

because equations (37) and (38) combine to form

$$x^*_{k+1} = \sum_{i=0}^{n-1} (\gamma_{ik} - \alpha_{i+1}) x_{k-i} - \sum_{i=1}^{n} \gamma_{i-1,k} x^*_{k+1-i} . \tag{45}$$

Our recursive technique for finding the characteristics of the optimal nth-order growing memory predictor thus consists of alternately performing the calculations of equations (42) and (43) and of obtaining the filter coefficients at each step by means of equation (45).

6.3 *Convergence of Filter Coefficients*

Since the time-varying filter output y_k converges to the optimal unconstrained predictor $\hat{x}_{k+1}$, one would expect that the time-varying coefficients a_{ik} and b_{ik} will converge to constant coefficients a_i and b_i. Since we have excluded processes with zeros on the unit circle, an nth-order recursive structure for the optimal predictor is known to exist.[1] But this is not sufficient. It is also necessary to exclude the possibility that the intrinsic order of the process is less than n. Then the coefficients of the nth-order recursive equation for the optimal predictor are unique and the time-varying coefficients a_{ik} and b_{ik} will in fact converge to these constant coefficients.

6.4 *Relation to Kalman Filtering*

In addition to proving convergence of our design approach, we have shown for the matched order case that the time-varying filter generated by the design procedure is the optimal growing-memory predictor. At each instant, k, the $2n$ stored data samples contain all the needed information about the observed past of the process, $x_0, x_1, \cdots, x_k$. It follows that the time-varying filter must be identical to the Kalman predictor[3] which is obtained by expressing the process model in state equation form. However, the Kalman development is computationally less efficient as may be seen by comparing the Ricatti equations with the simpler recursions given in Section 6.2.

In recent months recursions similar to ours have been published in various contexts. They appear in a paper by J. Rissanen and L. Barbosa[9] as steps in the factorization of the covariance matrix of $\{u_k\}$, the nth-order moving average, and Kailath[10] has indicated that such recursions follow from an innovations approach to prediction. Related formulas also appear in R. L. Kashyap's[11] derivation of predictor characteristics in terms of the parameters α_i and β_i of the process representation. In our derivation, as in Refs. 9 and 10, the basic data are the set of α_i and the autocovariance function of $\{u_k\}$. In contrast, the new design algorithm presented in Section VII uses only the covariances of the process to be predicted, quantities that are often more accessible in practice than the process parameters.

VII. SYNTHESIS TECHNIQUE

In this section we apply the projecting-filter design approach of Section V to obtain a computational algorithm for the general case in which the order of the process may differ from the order of the filter. The basic idea of the approach is to compute successive sets of weighting coefficients for an nth-order *time-varying* recursive filter which asymptotically approaches the desired *time-invariant* projecting filter.

As discussed in Section V, the time-varying projecting filter of interest is characterized by the input-output relationship

$$y_k = P\{x_{k+1} \mid \mathcal{H}_k\} \tag{46}$$

where $\mathcal{H}_k$ denotes the subspace spanned by the $2n$ variates

$$x'_k , x'_{k-1} , \cdots , x'_{k-n+1} , y_{k-1} , y_{k-2} , \cdots , y_{k-n} .$$

Equation (46) uniquely defines y_k as the projection of x_{k+1} into $\mathcal{H}_k$. This projection can be expressed explicitly as a linear combination of the $2n$ variates; that is,

$$y_k = \sum_{i=0}^{n-1} a_{ik}x'_{k-i} + \sum_{i=1}^{n} b_{ik}y_{k-i} . \tag{47}$$

Let $d(\mathcal{H}_k)$ denote the dimension of the subspace $\mathcal{H}_k$, i.e., $d(\mathcal{H}_k)$ is the minimum number of variates needed to span $\mathcal{H}_k$. If $d(\mathcal{H}_k) = 2n$ then the $2n$ spanning variates are linearly independent and the coefficient set used in equation (47) is unique. On the other hand if $d(\mathcal{H}_k) < 2n$, the $2n$ spanning variates are linearly dependent and consequently there is an infinite number of possible choices for the coefficient set. This situation always occurs in the first $2n - 1$ iterations ($0 \leqq k <$

$2n - 1$) and it may occur as well in subsequent iterations. To overcome this difficulty, we adopt a consistent procedure for selecting a linearly independent subset of the $2n$ spanning variates for each k. Variates are eliminated by setting appropriate coefficients to zero in equation (47). The remaining coefficients are then uniquely determined from the covariance matrix of the remaining variates and the cross-covariances between the remaining variates and x_{k+1}.

The algorithm is initialized with $y_0 = a_{00}x_0$ and all of the other coefficients a_{i0} and b_{i0} ($i \neq 0$) set to zero. Then each iteration consists of the following steps: (*i*) solving for the appropriate coefficient values, (*ii*) computing the needed covariances for the following iteration, (*iii*) determining an independent set of variates for the next prediction.

7.1 *Reduced Representation*

The procedure for eliminating dependent variates from the set of available data at time k leads to the following expression [equivalent to equation (47)] for the kth prediction

$$y_k = \sum_{i=0}^{p-1} a_{ik}x_{k-i} + \sum_{i=1}^{q} b_{ik}y_{k-i} \tag{48}$$

with $p \leqq n$, $q \leqq n$.* The coefficients that do not appear in equation (48) are all set to zero in the process of eliminating dependent variates; that is,

$$a_{ik} = 0, \qquad i = p, p+1, \cdots, n-1;$$

$$b_{ik} = 0, \qquad i = q+1, q+2, \cdots, n.$$

Note that x_{k-i} rather than x'_{k-i} appears in equation (48). This is so because $x'_{k-i} = 0$ for $i > k$ so that any set containing this variate is necessarily dependent. Hence, in the initial n steps, $p \leqq k - 1$. Section 7.4 presents the general method by which a set of independent variates is determined.

7.2 *The Filter Equations*

With the prediction error defined as $e_{k+1} = x_{k+1} - y_k$, the projecting property implies the following orthogonality conditions

$$\begin{aligned} Ee_{k+1}x_{k-i} &= 0, \qquad i = 0, 1, \cdots, p-1; \\ Ee_{k+1}y_{k-i} &= 0, \qquad i = 0, 1, \cdots, q. \end{aligned} \tag{49}$$

---

* Note that p and q depend on k. They will be denoted $p(k)$ and $q(k)$ when ambiguity might otherwise arise.

By substituting equation (48) for y_k into equation (49), we obtain the following set of $d(\mathcal{H}_k)$ linear equations in the $d(\mathcal{H}_k)$ coefficients:

$$r_{j+1} = \sum_{i=0}^{p-1} a_{ik} r_{i-j} + \sum_{i=1}^{q} b_{ik} w(k-j, k-i), \quad j = 0, 1, \cdots, p-1;$$

$$w(k+1, k-j) = \sum_{i=0}^{p-1} a_{ik} w(k-i, k-j) + \sum_{i=1}^{q} b_{ik} v(k-i, k-j), \quad j = 1, 2, \cdots, q; \tag{50}$$

in which we have adopted the notation:

$$r_i = Ex_k x_{k-i} = r_{-i} ,$$

$$w(k, j) = Ex_k y_j ,$$

$$v(k, j) = Ey_k y_j .$$

The function r_i comprises the given statistical information of the prediction problem and w and v must be expressed as functions of r_i and previously computed filter coefficients.

Equations (50) have the following partitioned matrix form

$$\begin{bmatrix} T_p & X_k \\ X_k' & V_k \end{bmatrix} \begin{bmatrix} A_k \\ B_k \end{bmatrix} = \begin{bmatrix} R_p \\ W_k \end{bmatrix} \tag{51}$$

with

T_p the $p \times p$ autocovariance matrix of $\{x_k\}$,
X_k the $p \times q$ cross-covariance matrix of $\{x_k\}$ and $\{y_k\}$,
V_k the $p \times p$ autocovariance matrix of $\{y_k\}$,
$A_k = [a_{0k}, a_{1k}, \cdots, a_{p-1,k}]'$,
$B_k = [b_{1k}, b_{2k}, \cdots, b_{qk}]'$,
$R_p = [r_1, r_2, \cdots, r_p]'$,
$W_k = [w(k+1, k-1), \cdots, w(k+1, k-q)]'$.

Note that T_p and R_p depend only on the given autocovariance function r_i and on p, the number of forward coefficients to be computed. They are independent of previously computed coefficients.

If we perform the multiplication indicated in equation (51) and then solve for A_k and B_k we derive

$$\begin{aligned} B_k &= [V_k - X_k' U_p X_k]^{-1} [W_k - X_k' C_p], \\ A_k &= C_p - U_p X_k B_k , \end{aligned} \tag{52}$$

where $U_p = T_p^{-1}$ and $C_p = U_p R_p$, the column matrix of weights corresponding to the optimum pth-order nonrecursive predictor. By using efficient algorithms developed for the analysis of nonrecursive predictors,[12,13] one may successively calculate $U_0, U_1, \cdots, U_{n-1}, C_0, C_1, \cdots, C_{n-1}$ before the start of the synthesis procedure so that at the kth step, only a $q \times q$ matrix inversion [rather than one of order $(p + q)$] is required. We are assured that the matrix to be inverted is nonsingular because we have eliminated dependent variates by reducing the number of unknowns from $2n$ to $p + q$. Note that A_k consists of the coefficients of the optimum pth-order nonrecursive predictor modified by $U_p X_k B_k$ which indicates the effect of the feedback section of the predicting filter.

7.3 *Obtaining Successive Covariance Statistics*

The nature of $w(k, j)$ depends on which time index is the greater. If $j \geqq k$ we observe that the projection property of the jth estimate implies that $Ex_k e_j = 0$ for $k = j - 1, j - 2, \cdots, j - n$. Thus if we substitute $x_{j+1} - e_{j+1}$ for y_j in the definition of $w(k, j)$, we obtain

$$\begin{aligned} w(k, j) &= E[(x_{j+1} - e_{j+1})x_k], \\ &= r_{j+1-k}, \qquad j = k, k+1, \cdots, k+n-1. \end{aligned} \tag{53}$$

For $j < k$, we substitute equation (26) for y_j in the definition of $w(k, j)$, with the result

$$w(k, j) = \sum_{i=0}^{n-1} a_{ij} r_{j-k-i} + \sum_{i=1}^{n} b_{ij} w(k, j - i), \qquad j = 0, 1, \cdots, k - 1. \tag{54}$$

Equation (54) indicates that $\{w(k, 0), w(k, 1), \cdots, w(k, k - 1)\}$ is the sequence of filter outputs when $\{r_{-k}, r_{-k+1}, \cdots, r_{-1}\}$ is the sequence of inputs. This is an example of the property of linear filters that the cross-covariance between input and output is the correlation of the filter impulse response with the input autocovariance function. Using the initial conditions $w(k, j) = 0$ for $j < 0$, we may iteratively apply equation (54) in order to compute the required values of $w(k, j)$ for $j < k$.

The autocovariance coefficients of $\{y_k\}$ may be determined from the orthogonality conditions. With $k - n \leqq j \leqq k$, we have $Ee_{k+1}y_j = 0$ so that

$$v(k, j) = E[(x_{k+1} - e_{k+1})y_j] = w(k + 1, j), \qquad j = k - n, \cdots, k, \tag{55}$$

and of course $v(j, k) = v(k, j)$. Thus, equations (53), (54), and (55) express, in terms of known quantities, the covariance coefficients that appear in equation (51).

7.4 *The Number of Independent Variates*

In Section 7.2 we have assumed that p and q, the number of forward coefficients and the number of feedback coefficients to be computed at time k are determined in a manner that assures the linear independence of the $p + q$ variates that appear in equation (48) and therefore, the existence of the inverse matrix of equation (52). In many instances $p = q = n$ so that all of the data in the predictor memory are linearly independent. On the other hand, there are two conditions under which the data are dependent. The first is called an initialization condition and this arises in the course of every synthesis procedure because the predictor begins to operate at $k = 0$ with zero in all memory elements except one. The initialization condition obtains for the first $2n - 2$ iterations of the design procedure during which $d(\mathcal{H}_k) \leqq k + 1 < 2n$ because $\mathcal{H}_k \subset \mathcal{R}_k$ and $d(\mathcal{R}_k) = k + 1$. The other condition under which $d(\mathcal{H}_k) < 2n$ is called a reduced order condition, which arises when certain of the final feedback coefficients and/or final forward coefficients are zero. A reduced-order condition arises for all processes of order less than n.

7.4.1 *Initialization*

In this section we assume that no reduced order condition arises during the first $2n - 1$ steps of the predictor synthesis. This implies that $d(\mathcal{H}_k) = k + 1$ so that $p + q$, the number of coefficients determined by orthogonality conditions, increases by one at each iteration. At $k = 0$, the predictor estimates x_1 given x_0 which implies $p = 1$, $q = 0$. For increasing k, we alternately increase q and p by one so that for $0 \leqq k \leqq 2n - 2$

$$\begin{aligned} p &= 1 + \tfrac{1}{2}k, \qquad q = \tfrac{1}{2}k, \qquad k \text{ even}; \\ p &= \tfrac{1}{2}(k + 1) = q, \qquad\qquad k \text{ odd}; \end{aligned} \tag{56}$$

when no reduced order condition arises. Table I shows the variates that appear in equation (48) during the initial design stages of a second-order predictor.

7.4.2 *Reduced-Order Condition*

At time $k + 1$, the dependency of the data in storage can be deduced by observation of the coefficients computed at time k. In this

TABLE I—STEPS IN PREDICTOR DESIGN

Time	Predicted Variate	Independent Data				Projection
0	x_1	x_0				y_0
1	x_2	x_1		y_0		y_1
2	x_3	x_2	x_1	y_1		y_2
3	x_4	x_3	x_2	y_2	y_1	y_3
k	x_{k+1}	x_k	x_{k-1}	y_{k-1}	y_{k-2}	y_k

section we show how the values of certain coefficients, in particular whether or not they are zero, determine the relationship between $d(\mathcal{H}_k)$ and $d(\mathcal{H}_{k+1})$, the numbers of linearly independent variates in storage at time k and at time $k + 1$. In the next section we present the algorithm for determining the number of forward coefficients and the number of feedback coefficients to be computed at each step of the design.

The following theorem states that there is a dependence among the variates in storage at time $k + 1$ if and only if the coefficients determined at time k correspond to a filter of order less than n.

Theorem: With $d(\mathcal{H}_k) = 2n$, $d(\mathcal{H}_{k+1}) = 2n - 1$ *if and only if* $a_{n-1,k} = b_{n,k} = 0$. *Otherwise* $d(\mathcal{H}_k) = 2n$.

Proof: Assume $a_{n-1,k} = b_{n,k} = 0$. Then

$$y_k = \sum_{i=0}^{n-2} a_{ik} x_{k-i} + \sum_{i=1}^{n-1} b_{ik} y_{k-i}$$

which shows the linear dependency of the following variates in storage at time $k + 1$: $x_k, x_{k-1}, \cdots, x_{k-n+2}, y_k, \cdots, y_{k-n+1}$. Thus $d(\mathcal{H}_{k+1}) < 2n$. On the other hand, the $2n - 1$ variates: $x_{k+1}, x_k, \cdots, x_{k-n+2}, y_{k-1}, \cdots, y_{k-n}$ are linearly independent. All except x_{k+1} are independent because they are in storage at time k and $d(\mathcal{H}_k) = 2n$. In addition, the assumption that $\{x_k\}$ is nondeterministic implies that x_{k+1} cannot be expressed as a linear combination of the other stored variates because each of these is in $\mathcal{R}_k$. It follows that $d(\mathcal{H}_{k+1}) = 2n - 1$.

To prove the converse, assume $d(\mathcal{H}_{k+1}) = 2n - 1$. It follows that there exists a linearly dependent set of stored data. By the reasoning given above this set does not include x_{k+1} because all of the other stored variates are in $\mathcal{R}_k$. However the set does include y_k because all of the other variates are independent. Hence y_k can be represented as a linear combination of $x_k, x_{k-1}, \cdots, x_{k-n+2}, y_{k-1}, \cdots, y_{k-n+1}$. But the data in storage at time k also includes x_{k-n+1} and y_{k-n} and the fact that

$d(\mathcal{H}_k) = 2n$ implies that the representation of y_k is unique. Therefore we have the coefficients of x_{k-n+1} and y_{k-n}, $a_{n-1,k} = b_{n,k} = 0$. Q.E.D.

By reasoning similar to that used to prove this theorem we may establish the dimensionality of the data in storage at time $k + 1$ when $d(\mathcal{H}_k) < 2n$. Thus we have the following corollaries which apply for all k including the initial steps of the predictor design.

Corollary 1: With $d(\mathcal{H}_k) = p + q$ *and* $p = q < n$, $d(\mathcal{H}_{k+1}) = p + q - 1$ *if and only if* $a_{p-1,k} = b_{q,k} = 0$. *Otherwise* $d(\mathcal{H}_{k+1}) = p + q + 1$.
Corollary 2: With $d(\mathcal{H}_k) = p + q$ *and* $n \geqq p = q + 1$, $d(\mathcal{H}_{k+1}) = p + q$ *if and only if* $a_{p-1,k} = 0$. *Otherwise* $d(\mathcal{H}_{k+1}) = p + q + 1$.
Corollary 3: With $d(\mathcal{H}_k) = p + q$ *and* $p = q - 1 < n$, $d(\mathcal{H}_{k+1}) = p + q$ *if and only if* $b_{qk} = 0$. *Otherwise* $d(\mathcal{H}_{k+1}) = p + q + 1$.

7.4.3 *The Number of Computed Coefficients*

On the basis of the theorem and corollaries of Section 7.4.2, we establish the procedure shown in Table II for determining the numbers of forward and feedback coefficients $p(k + 1)$ and $q(k + 1)$ to be computed at time $k + 1$. The table indicates that $p(k + 1)$ and $q(k + 1)$ may be determined from $p = p(k)$ and $q = q(k)$ (shown in the left column) and from the final two feedback coefficients and the final

TABLE II—THE NUMBER OF COEFFICIENTS COMPUTED

Number of Coefficients Computed at Time k		Final Coefficients Computed at Time k				Number of Coefficients Computed at Time $k+1$	
		$b_{q,k}$	$b_{q-1,k}$	$a_{p-1,k}$	$a_{p-2,k}$	$p(k+1)$	$q(k+1)$
1	$p = q = n$	$\neq 0$				n	n
2				$\neq 0$		n	n
3	$p = q$	0	$\neq 0$	0		p	$q - 1$
4		0		0	$\neq 0$	$p - 1$	q
5	$p = q < n$	$\neq 0$				$p + 1$	q
6				$\neq 0$		p	$q + 1$
7	$p > q$			$\neq 0$		p	$q + 1$
8		$\neq 0$		0		p	q
9		0		0	$\neq 0$	$p - 1$	$q + 1$
10	$p < q$	$\neq 0$				$p + 1$	p
11		0		$\neq 0$		p	q
12		0	$\neq 0$	0		$p + 1$	$q - 1$
13	any p, q	0	0	0	0	irregular	

two forward coefficients (shown in the central four columns) computed at time k. If there is no entry for one of the coefficients, the indicated relationship between $p(k+1)$, $q(k+1)$ and p, q is independent of that coefficient. The other symbols indicate that a coefficient must necessarily be zero or nonzero for a relationship to be valid.

If, at time k, $p + q = d(\mathcal{H}_k)$, the variates $x_k, x_{k-1}, \cdots, x_{k-p+1}, y_{k-1}, \cdots, y_{k-q}$ are independent. This condition and the theorem and corollaries imply that the set $\{x_{k+1}, x_k, \cdots, x_{k-p(k+1)+2}, y_k, \cdots, y_{k-q(k+1)+1}\}$ is independent and spans $\mathcal{H}_{k+1}$. Thus lines 1 and 2 of Table II follow from the theorem; lines 3 through 6, from the theorem and Corollary 1; lines 7 through 9, from Corollary 2; and lines 10 through 12 from Corollary 3.

The table accounts for all possible combinations of computed coefficient values except those in which the last two forward coefficients and the last two feedback coefficients are all zero. This situation arises during the initial design stages whenever the input process is partially decorrelated. The manner in which independent variates are chosen for such a process is described in Section 7.4.5. When the irregularity arises in the design of predictors for other processes, there is no independent basis of $\mathcal{H}_{k+1}$ that is the union of consecutive members of $\{x_k\}$ beginning with x_{k+1} and consecutive members of $\{y_k\}$ beginning with y_k. Thus it is impossible to represent $P\{y_k \mid \mathcal{H}_k\}$ in the concise form of equation (48). Nor is it possible in general to determine at all times subsequent to k an independent set of stored data solely by considering p, q and the previously computed coefficients. All this serves to complicate quite substantially the representation of y_k, the equations which determine the coefficients, and the algorithm for determining the numbers of coefficients to be computed after the occurrence of the irregular condition indicated on the last line of Table II.

Rather than add substantially to the size of this paper by presenting a general technique for treating this situation, we simply note that except for partially decorrelated processes, it has never arisen in our experience of designing projecting filters and that in fact it appears to represent a pathological case. We have not discovered an example of a process for which four projection coefficients are simultaneously zero after one or more of their counterparts is nonzero at the previous time instant.

7.4.4 *Low Order Processes*

When $\{x_k\}$ is the response of an mth-order filter to white noise and m is no greater than n, the order of the predictor, the synthesis method leads to the mth-order form of the optimum unconstrained predictor. Section 6.1 contains a proof of this statement for $m = n$ and in this section we show that if $m < n$, a reduced-order situation arises and the effective order of the predictor does not grow beyond n.

Let a_{ik} and b_{ik} be the coefficients of the optimal growing-memory mth-order predictor, determined in the manner indicated in Section 6.2. Thus

$$x_{k+1}^* = \sum_{i=0}^{m-1} a_{ik} x_{k-i} + \sum_{i=1}^{m} b_{ik} x_{k+1-i}^* . \tag{57}$$

Note that for all $k \leqq 2m - 1$, y_k, the output of the nth-order predictor is identical to x_{k+1}^* because the design proceeds as for a predictor of order m.

Equation (56) indicates that at step $2m$ the initialization procedure leads to $p = m + 1$, $q = m$ and

$$y_{2m} = \sum_{i=0}^{m} a'_{i,2m} x_{2m-i} + \sum_{i=1}^{m} b'_{i,2m} y_{2m-i} \tag{58}$$

where $a'_{i,2m}$ and $b'_{i,2m}$ are determined uniquely by the orthogonality conditions. Hence it follows from the optimality of equation (57) that $a'_{m,2m} = 0$ and that the other coefficients are equal to the ones in equation (57) with $k = 2m$. Line 8 of Table I indicates that $p(2m + 1) = m + 1$ and $q(2m + 1) = m$ and once again we have $a'_{m,2m+1} = 0$ and the other coefficients equal to those in equation (57) for the optimal mth-order predictor. It is clear that for all $k \geqq 2m$ this sequence is repeated with $p(k) = m + 1$, $q(k) = m$ and $a'_{m,k} = 0$. Hence the algorithm converges to the unique mth-order form of the unconstrained optimal predictor.

7.4.5 *Partially Decorrelated Input Process*

A partially decorrelated process is a nonwhite process for which every set of $j + 1$ ($j > 0$) adjacent samples is uncorrelated. In other words, $\{x_k\}$ is partially decorrelated if for some $j > 0$, $r_1 = r_2 = \cdots = r_j = 0$ and $r_{j+1} \neq 0$. For example the error process of an nth-order projecting filter is partially decorrelated with $j = n$.

Note that with a partially decorrelated input, the initial j generating filter outputs (corresponding to optimal nonrecursive predictions) are zero. Thus

$$y_k = x_{k+1}^* = 0 = a_{ik} = b_{ik}, \text{ for } 0 \leqq k < j \text{ and all } i. \tag{59}$$

This is a reduced-order situation conforming to line 13 of Table II (if we assume $b_{0k} = 0$ and $a_{ik} = b_{ik} = 0$ for $i < 0$). For this irregular case we adopt the following initialization procedure as an alternative to equation (56).

(*i*) All coefficients are 0 for $k < j$.
(*ii*) $p(j) = j + 1$, $q(j) = 0$.
(*iii*) $p(k)$, $q(k)$ according to Table II for $k > j$.

VIII. CONCLUSIONS

This paper introduces the projecting-filter principle of constrained-order recursive prediction and presents one technique of projecting filter synthesis. This technique has led to the design of the predictors described in Section 1.3 and to several other successful designs for a variety of random processes. However, the class of processes for which the technique is valid (that is, for which the algorithm converges to a time-invariant filter) and indeed the class for which a projecting filter of a given order exists have not as yet been determined. These questions are the subject of current research. Another important area of investigation involves the numerical aspect of the synthesis—the study of the sensitivity of this or any other design method to round-off in the calculation of coefficients.

Our studies to date indicate that the projecting filter is valuable in that it predicts many processes more accurately than other known devices of equal complexity. Our results are readily extended to vector-valued processes. Finally, we note that the projecting filter principle is applicable to a large class of estimation problems of which prediction one unit of time in the future is but a single example.

REFERENCES

1. Whittle, P., *Prediction and Regulation by Linear Least-Square Methods,* Princeton, New Jersey: D. Van Nostrand Company, Inc., 1963, pp. 31–35.
2. Whittle, P., "Recursive Relations for Predictors of Nonstationary Processes," J. Roy. Statist. Soc., Ser. B, *27,* No. 3 (1965), pp. 523–532.
3. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME, J. Basic Engrg., Ser. D, *82,* No. 1 (March 1960), pp. 35–45.
4. Walsh, J. L., *Interpolation and Approximation by Rational Functions in the Complex Domain,* Amer. Math. Soc. Colloquium Publ. *XX,* Providence, 1960.
5. Phillips, R. S., "RMS-error Criterion in Servomechanism Design," Chap. 7 in H. M. James, N. B. Nichols, and R. S. Phillips, *Theory of Servomechanisms,* MIT Rad. Lab. Series, *25,* McGraw-Hill Book Co., New York 1947.
6. Kolmogorov, A. N., "Interpolation and Extrapolation of Stationary Random Sequences," Bull. Acad. Sci., USSR, Ser. Math., *5* (1941); transl.: Memorandum RM-3090-PR, Rand Corp., Santa Monica, California, April 1962.

7. Luenberger, D. G., *Optimization by Vector Space Methods,* New York: John Wiley and Sons, Inc., 1969, Chapter 4.
8. Doob, J. L., *Stochastic Processes,* New York: John Wiley and Sons, Inc., 1969, pp. 569–581.
9. Rissanen, J., and Barbosa, L., "Properties of Infinite Covariance Matrices and Stability of Optimum Predictors," Information Sciences, *1,* No. 2 (April 1969), pp. 221–236.
10. Kailath, T., "The Innovations Approach to Detection and Estimation Theory," Proc. IEEE, *58,* No. 5 (May 1970), pp. 680–695.
11. Kashyap, R. L., "A New Method of Recursive Estimation in Discrete Linear Systems," IEEE Trans. Aut. Cont., *AC-15,* No. 1 (February 1970), pp. 18–24.
12. Trench, W. F., "An Algorithm for the Inversion of Finite Toeplitz Matrices," J. Soc. Indust. Appl. Math., *12,* No. 3 (September 1964), pp. 515–522.
13. Trench, W. F., "Weighting Coefficients for the Prediction of Stationary Time Series from the Finite Past," J. Soc. Indust. Appl. Math., *15,* No. 6 (November 1967), pp. 1502–1510.