A New Algorithm for Optimal Interpolation of Discrete-Time Stationary Processes\textsuperscript{1}

by

Michele Pavon

Department of Statistics
The Florida State University
Tallahassee, Florida 32306

and

LADSEB-CNR
Corso Stati Uniti 4
35100 Padova, Italy

March, 1982
FSU Statistics Report No. M616

\textsuperscript{1}Research supported by CNR, Italy.
A New Algorithm for Optimal Interpolation of Discrete-Time Stationary Processes

by

Michele Paven

Abstract

A new approach to the interpolation problem for multivariate stationary Gaussian processes is presented. This approach hinges on the recently developed stochastic realization theory. New representations for the optimal interpolator and interpolation error variance are derived. In particular we show that the interpolation estimate is characterized by two steady state Kalman filters, one evolving forward and one backward in time. The derivation is simple and enlightening. The results appear to be of computational interest.
1. Introduction

The problem of linear interpolation of stationary processes is of considerable importance in the engineering and social sciences and has generated a rather extensive literature beginning with Kolmogorov's fundamental article [1-12].

The purpose of this paper is to present a new method to compute the optimal interpolator and interpolator error variance for a multivariate stationary discrete-time Gaussian process. Our derivation relies on concepts and techniques from the recently developed stochastic realization theory [13-16]. We show that the optimal interpolator admits a decomposition into the sum of the optimal predictor plus a vector orthogonal to the past of the process. This vector is written in terms of two Kalman estimates, the first generated by a steady-state forward filter and the second by a steady-state backward filter. This is the key representation from which we immediately derive the interpolation error intensity and several other expressions for the interpolator. These are phrased in terms of the available values of the process, in terms of the forward and backward innovations, and in terms of the spectral measure of the process. In particular, we rederive the classical spectral domain results [5, p. 102], gaining new insight into their stochastic significance. We study the case when the process has a rational spectral density and only the value at time zero is missing.
We do so since our chief aim here is to convey the basic ideas underlying our approach, rather than discussing the most general problem to which it can be applied. Indeed, this method appears to be quite powerful and we have already succeeded in obtaining the first explicit solution to the interpolation problem for continuous-time stationary increments processes [19]. Moreover, we feel that the case of non-rational spectral density can be treated in a similar fashion employing the infinite-dimensional stochastic realization theory [20-24]. Extensions of this theory to the nonstationary case are also possible in contrast with the existing spectral techniques. These extensions will be the subject of forthcoming papers.

Our results seem to be of computational interest since they are compact and they only involve quantities which can be efficiently computed from the spectral density via the deterministic and stochastic realization algorithms [25, 13-15].

The paper is organized as follows. In the next section we formulate the problem and introduce the relevant mathematical notation. In Section 3 we recall some basic results from stochastic realization theory. These are then applied in Section 4 to the interpolation problem in order to derive the main representation for the estimate. In Section 5 we present various time-domain expressions for the optimal interpolator, whereas in Section 6 we derive spectral-domain formulae and make contact with the classical theory of stationary processes.
2. Basic Notation and Problem Formulation.

We denote transposition by a prime. All vectors without a prime are column vectors. The identity matrix is indicated by I. If \( R \) is a positive (nonnegative) definite matrix we write \( R > 0 (R \geq 0) \). The nonnegative square root of the symmetric nonnegative matrix \( R \) is indicated by \( R^{\frac{1}{2}} \). Let \( Z \) denote the integers and let \( \{ \xi(t); t \in Z \} \) be a centered \( p \)-dimensional Gaussian process. Then \( H_{t}^{-}(\xi), H_{t}^{+}(\xi), H(\xi(t)) \) and \( H(\xi) \) indicate the Gaussian space [26] generated by \( \{ \xi^{i}(s); i = 1, \ldots, p; s \in A \} \) where \( A \) is the set \( \{ s \in Z|s \leq t \}, \{ s \in Z|s \geq t \}, \{ t \} \) and \( Z \), respectively. Let \( H \) be a Gaussian space. Then \( E(\cdot|H) \) denotes the orthogonal projection operator onto \( H \). By orthogonal projection of a vector is to be intended the vector obtained projecting its components.

If \( K \) is another Gaussian space, \( \overline{E}(H|K) \) denotes the closure in the mean square norm of the projections of elements in \( H \) onto \( K \). Let \( K \subset H \). Then \( H \ominus K \) denotes the orthogonal complement of \( K \) in \( H \).

We shall be concerned with a centered, Gaussian, \( \mathbb{R}^{m} \)-valued stochastic process \( \{ y(t); t \in Z \} \) defined on an underlying probability space \( (\Omega, F, P) \). We assume that \( y \) is stationary, purely nondeterministic of full rank and with rational spectral density \( \Phi(\cdot) \) satisfying \( \Phi(e^{i\omega}) > 0 \) for all \( \omega \). The last assumption is very common both in the stochastic realization theory and in the interpolation theory since it considerably simplifies matters. It implies that \( y \) is a minimal process [5].
We shall study the following interpolation problem: given \( \{y(t); t \neq 0\} \) determine \( \hat{y}(0) = \text{E}(y(0)|H_{-1}^-(y)vH_1^+(y)) \) where \( H_{-1}^-(y)vH_1^+(y) \) denotes the smallest Gaussian space containing \( H_{-1}^-(y) \) and \( H_1^+(y) \).

3. Preliminaries

Let us decompose the spectral density as

\[
\phi(z) = S(z) + S(1/z)'
\]

with \( S \) discrete positive real [15] and let \([F, G, H, J]\) be a minimal realization [25] of \( S \) so that \( S(z) = H(zI - F)^{-1}G + J \). There exist well-known computational methods for determining such a quadruplet from \( \phi \) [25], and therefore in the sequel we shall regard it as given. Notice that under our assumptions \( |\lambda(F)| < 1 \), i.e. the \( F \) matrix has all its eigenvalues in the open unit disc. Then it can be shown [14, 15] that \( y \) admits two Markovian representations (stochastic realizations). The first

\[
x_*(t + 1) = Fx_*(t) + B_*u_*(t), \tag{3.1a}
\]

\[
y(t) = Hx_*(t) + {\hat{R}}_*u_*(t), \tag{3.1b}
\]

evolves forward in time and is a steady-state Kalman filter. The matrices \( B_*, R_* \) and \( P_* = \text{E}(x_*(t)x_*(t)') \) satisfy the equations

\[
FPF' + BB' = P, \tag{3.2a}
\]

\[
FPH' + BR' = G, \tag{3.2b}
\]

\[
HPH' + R = \Lambda(0) \tag{3.2c}
\]
where $\Lambda(0) = E[y(t)y(t)']$. The innovations $u_\phi$ form a Gaussian normalized white noise sequence characterized by the property $H^\phi_\epsilon(u_\phi) = H^\phi_\epsilon(y)$ for all $t \in Z$. Such a property becomes apparent when one inverts (3.1) using the fact that $R_\phi > 0$ [15] and gets

$$x_\phi(t + 1) = \Gamma_\phi x_\phi(t) + B_\phi R_\phi^{-1/2} y(t)$$  (3.3)

where the feedback matrix $\Gamma_\phi = F - B_\phi R_\phi^{-1/2} H$ satisfies $|\lambda(\Gamma_\phi)| < 1$. The second realization

$$\bar{x}_\phi(t - 1) = F^\phi \bar{x}_\phi(t) + B_\phi \bar{u}_\phi(t),$$  (3.4a)

$$y(t) = G^\phi \bar{x}_\phi(t) + R_\phi^{1/2} \bar{u}_\phi(t),$$  (3.4b)

is a backward steady-state Kalman filter. The matrices $\bar{R}_\phi$, $\bar{R}_\phi$ and $\bar{F}_\phi = E(\bar{x}_\phi(t)\bar{x}_\phi(t)')$ solve the system

$$F^\phi \bar{F} + \bar{B}\bar{R}^{1/2} = \bar{F},$$  (3.5a)

$$F^\phi \bar{G} + \bar{B}R_\phi^{-1/2} = \bar{H}^{1/2},$$  (3.5b)

$$G^\phi \bar{G} + \bar{R} = \Lambda(0)$$  (3.5c)

The backward innovations $\bar{u}_\phi$ is a Gaussian normalized white noise satisfying $H^\phi_\epsilon(\bar{u}_\phi) = H^\phi_\epsilon(y)$ for all $t \in Z$. Since $\bar{R}_\phi$ is nonsingular [15] we can invert (3.4) and get

$$\bar{x}_\phi(t - 1) = \bar{F}_\phi \bar{x}_\phi(t) + \bar{B}_\phi \bar{R}_\phi^{-1/2} y(t)$$  (3.6)
where the backward feedback matrix \( \mathbf{F}_* = \mathbf{F} - \mathbf{B}_* \mathbf{R}_*^{-1} \mathbf{G} \) satisfies \( |\lambda(\mathbf{F}_*)| < 1 \). Two other Markovian representations of \( y \) can be constructed from (3.1) and (3.4). In fact [14, 15, 17] there exists a forward model

\[
\begin{align*}
x^\star(t + 1) &= \mathbf{F}x^\star(t) + \mathbf{B}^* u^\star(t), \quad (3.7a) \\
y(t) &= \mathbf{H}x^\star(t) + (\mathbf{R}^*)^{-\frac{1}{2}} u^\star(t), \quad (3.7b)
\end{align*}
\]

corresponding to (3.4) and a backward one

\[
\begin{align*}
\mathbf{x}^\star(t - 1) &= \mathbf{F}^\star \mathbf{x}^\star(t) + \mathbf{B}^\star u^\star(t), \quad (3.8a) \\
y(t) &= \mathbf{G}^\star \mathbf{x}^\star(t) + (\mathbf{R}^*)^{-\frac{1}{2}} u^\star(t), \quad (3.8b)
\end{align*}
\]

corresponding to (3.1). Let \( P^* = \mathbb{E}(x^\star(t) x^\star(t)') \) and \( \mathbf{F}^* = \mathbb{E}(x^\star(t) x^\star(t)') \).

Then [13-15, 17]

\[
P^* - P^* > 0, \quad (3.9)
\]

\[
\begin{align*}
\mathbf{x}^\star(t) &= (P^*)^{-1} x^\star(t + 1), \quad \mathbf{F}^* (P^*)^{-1}, \quad (3.10) \\
\mathbf{x}^\star(t) &= (P^*)^{-1} x^\star(t + 1), \quad \mathbf{F}^* = (P^*)^{-1}. \quad (3.11)
\end{align*}
\]

Moreover \( (P^*, \mathbf{B}^*, \mathbf{R}^*) \) and \( (\mathbf{F}^*, \mathbf{B}^*, \mathbf{R}^*) \) satisfy (3.2) and (3.5), respectively. For further details on the correspondence between forward and backward realizations we refer the reader to [17].

We now record two projection results which will play an important role in the next section.
\[ E(x^*(t)|H_{t-1}^-) = x_*(t) \quad (3.12) \]
\[ E(x^*(t)|H_{t+1}^+) = \bar{x}_*(t) \quad (3.13) \]

These formulae follow from the filter property of (3.1) and (3.4).

4. Derivation of the Main Representation.

We shall now demonstrate that the two vectors \( x_*(0) \) and \( \bar{x}_*(0) \) contain all the relevant information on \( \{y(t), \ t \neq 0\} \) needed in estimating \( y(0) \). To this end first introduce the spaces \( N^- : = H_{-1}^-(y) \cap H(x_*(0)) \) and \( N^+ : = H_1^+(y) \cap H(\bar{x}_*(0)) \). Now note that \( N^- \) is orthogonal to \( H_1^+(y) \). In fact \( N^- H_0^+(y) \supseteq H_1^+(y) \) since
\[ E(H_0^+(y)|H_{-1}^-(y)) = H(x_*(0)), \text{ cf. e.g. } [14]. \] Similarly it is seen that \( N^+ \) is orthogonal to \( H_{-1}^-(y) \). Then we have the following orthogonal decomposition

\[ H_{-1}^-(y) \cap H_1^+(y) = N^- \oplus H(x_*(0)) \cap H(\bar{x}_*(0)) \oplus N^+ \quad (4.1) \]

**Lemma 4.1.** \( H(\hat{y}(0)) \subseteq H(x_*(0)) \cap H(\bar{x}_*(0)). \)

**Proof.** Clearly \( H(\hat{y}(0)) \subseteq N^- \). To see this note that
\[ E(\hat{y}(0)|H_{-1}^-(y)) = E(y(0)|H_{-1}^-(y)) = Hx_*(0), \] where the first equality results from the law of iterated conditioning. Hence the components of \( \hat{y}(0) - Hx_*(0) \) are orthogonal to \( H_{-1}^-(y) \) which contains \( N^- \). Analogously, we can show that \( H(\hat{y}(0)) \subseteq N^+ \). The result now follows from (4.1). \( \square \)
Theorem 4.2. The optimal interpolator admits the following orthogonal decomposition.

$$\hat{y}(0) = Hx_*(0) + K^\top\Pi^{-1}[x^*(1) - Fx_*(0)]$$ (4.2)

where $K = B_\ast R_\ast^{-1/2}$ is the Kalman gain and $\Pi = P^* - FP_\ast F^*$. The variance of the error $\hat{y}(0) = y(0) - \hat{y}(1)$ is given by

$$E(\hat{y}(0)\hat{y}(0)^\top) = R_\ast - K^\top\Pi^{-1}K$$ (4.3)

Proof. By Lemma 4.1 and the law of iterated conditioning we have

$$\hat{y}(0) = E(y(0)|H(x_*(0))\cap H(x_*(0)))$$

Next observe that the vectors $x_*(0)$ and $x^*(1) - Fx_*(0)$ are orthogonal because of (3.7) and (3.12). Moreover, in view of (3.10), their components form a basis for $H(x_*(0))\cap H(x_*(0))$. Thus,

$$\hat{y}(0) = E(y(0)|x_*(0)) + E(y(0)|x^*(1) - Fx_*(0))$$ (4.4)

The first term on the right-hand side of (4.4) is $Hx_*(0)$. The second can be computed via a standard projection formula. Let $\Pi = P^* - FP_\ast F^*$ be the variance of $x^*(1) - Fx_*(0)$. Then (3.2a) and (3.9) imply that $\Pi > 0$, and we get

$$E(y(0)|x^*(1) - Fx_*(0)) = E(y(0)[x^*(1) - Fx_*(0)]^\top)\Pi^{-1}[x^*(1) - Fx_*(0)].$$

Finally (3.7) gives $E(y(0)|x^*(1)) = HP^*F^* + (R^*)^{1/2}b^*$ and (3.1b)
gives $E(y(0)x_*(0)^\mathrm{T}) = HP_\alpha E^*$. Formula (4.2) now follows from the fact that both the maximum and the minimum variance model satisfy (3.2b). To prove (4.3) note that (4.2) yields

$$E(\hat{y}(0)y(0)^\mathrm{T}) = HP_\alpha H^* + K^\mathrm{T} \Pi^{-1} K.$$  

It remains to observe that the error is orthogonal to the interpolation estimate, and therefore its variance is given by $\Lambda(0) = [HP_\alpha H^* + K^\mathrm{T} \Pi^{-1} K]$, which is equal to (4.3) because of (3.2c) $\square$.

**Remark 4.5.** Relations (4.2) and (4.3) allow for an immediate comparison between the optimal interpolator and the optimal one-step predictor, namely $E(y(0)|H^-_1(y)) = Hx_*(0)$. The vector $K^\mathrm{T} \Pi^{-1} \{x^*(1) - Fx_*(0)\}$ is orthogonal to $H^-_1(y)$ and represents the modification of the estimate due to the extra data $\{y(t), t \geq 1\}$. Since the variance of the prediction error is $R_\alpha$, we see that the positive definite matrix $K^\mathrm{T} \Pi^{-1} K$ gives a measure of the improvement of our knowledge on $y(0)$ $\square$.

**Remark 4.6.** An alternative, but less direct, proof for Theorem 4.2 can be constructed relying on a theorem of N. Aronszajn [27], first applied to this problem by Salehi [8]. Let $T_1 = E\{\cdot | H^-_1(y)\}$, $T_2 = E\{\cdot | H^-_1(y)\} \Pi H^*_1(y)$, and $T = E\{\cdot | H^-_1(y)\} \Pi H^*_1(y)$. Then the theorem asserts that the sequence $S_1 = T_1$, $S_2 = T_1 + T_2 - T_2 T_1$, $S_3 = T_1 + T_2 - T_2 T_1 - T_1 T_2 + T_1 T_2 T_1$, ... converges strongly to $T$. Therefore, in particular, $S_N y(0)$ converges in the $L^2$ norm to $\hat{y}(0)$. From here Salehi proceeds in [8] to derive a rather involved expansion for the optimal interpolator in terms of the forward and
backward innovations. However, using stochastic realization, we immediately get (4.2). In fact (3.12) and (3.13) readily give
\[
S_{2n} y(0) = H x_\ast(0) + [G^\ast - \text{HP}_\ast F^\ast] \left[ \sum_{i=0}^{n-2} ((P^\ast)^{-1} F P_\ast F^\ast)^i (\bar{x}_\ast(0) - (P^\ast)^{-1} F x_\ast(0)) \right] + ((P^\ast)^{-1} F P_\ast F^\ast)^n \bar{x}_\ast(0),
\]
\[
S_{2n+1} y(0) = H x_\ast(0) + [G^\ast - \text{HP}_\ast F^\ast] \left[ \sum_{i=0}^{n-1} ((P^\ast)^{-1} F P_\ast F^\ast)^i (\bar{x}_\ast(0) - (P^\ast)^{-1} F x_\ast(0)) \right]
\]
for \( n \geq 1 \). It only remains to observe that \((P^\ast)^{-1} F P_\ast F^\ast\) has its eigenvalues inside the unit circle. In fact both \((P^\ast)^{-1} F P_\ast F^\ast\) and \( I - (P^\ast)^{-1} F P_\ast F^\ast = (P^\ast)^{-1} (P^\ast - F P_\ast F^\ast) \) have positive eigenvalues as product of positive definite matrices [28, p. 92]. Then a standard formula for geometric series, see viz. [28, p. 113], (3.2b) and (3.10) give (4.2). □

5. Time Domain Representations

Clearly, instead of (4.2) and (4.3), we can derive expressions in terms of the backward quantities. Indeed, if we use the components of \( \{ \bar{x}_\ast(0), \bar{x}^\ast(-1) - F \bar{x}_\ast(0) \} \) as a basis for \( H(x_\ast(0)) \cap H(\bar{x}_\ast(0)) \) and carry out a similar analysis we get
\[
\hat{\gamma}(0) = G^\ast \bar{x}_\ast(0) + \bar{K}^\ast \bar{P}^{-1} [\bar{x}^\ast(-1) - F^\ast \bar{x}_\ast(0)],
\]
and
\[
E(\tilde{y}(0)\tilde{y}(0)^\ast) = \bar{x}_\ast - \bar{K}^\ast \bar{P}^{-1} \bar{x},
\]
where \( \bar{x} = \bar{H}_\ast \hat{P}_\ast \) and \( \bar{P} = P^\ast - F^\ast (P^\ast)^{-1} F \).
It is also possible to derive a symmetric expression in terms of the two Kalman filter estimates. A simple calculation using (4.2), (3.5b) and (3.10) yields the formula

\[
\hat{y}(0) = K^{-1}\Pi^{-1}P_s^{-1}x_s(0) + K^{-1}p^{-1}\bar{\Gamma}_s^{-1}y_s(0).
\]

From (4.2) and (5.1) we shall now derive several other interesting representations.

**Theorem 5.1.** The optimal interpolator can be expressed as follows:

(i) \[
\hat{y}(0) = \sum_{i=1}^{\infty} K^{-1}P^{-1}F_i^{-1}B_s^{-1}R_s^{-1}y(-i) + \sum_{i=1}^{\infty} K^{-1}P_s^{-1}F_i^{-1}B_s^{-1}R_s^{-1}y(i);
\]

(ii) \[
\hat{y}(0) = \sum_{i=1}^{\infty} H^{-1}B_s^{-1}u_s(-i) + K^{-1}B_s^{-1}u_s(0) + \sum_{i=1}^{\infty} K^{-1}P^{-1}F_i^{-1}H^{-1}R_s^{-1}u_s(i),
\]

where \(\sum_p = p^* - P_s^*;\)

(iii) \[
\hat{y}(0) = \sum_{i=1}^{\infty} K^{-1}P^{-1}F_i^{-1}B_s^{-1}R_s^{-1}y(-i) + \sum_{i=1}^{\infty} K^{-1}P^{-1}F_i^{-1}H^{-1}R_s^{-1}u_s(i).
\]

**Proof.** Formula (i) results from (5.1), (3.3) and (3.6) in view of the asymptotic stability of \(\Gamma_s^*\) and \(\bar{\Gamma}_s^*\). To prove (ii) notice that (3.1a) allows to rewrite (4.2) as

\[
\hat{y}(0) = Hx_s(0) + K^{-1}B_s^{-1}u_s(0) + K^{-1}\Pi^{-1}\sum_{i=1}^{\infty} [x_s(i) - x_s(-i)].
\]

In [17, formula (2.52)] it was shown that \(z(t) = \sum_{i=1}^{\infty} [x_s(t) - x_s(t-i)]\) satisfies the recursion

\[
z(t) = \Gamma_0 z(t - 1) + H^*R_s^{-1}u_s(t)
\]

An integration of (3.1a) over the past and of (5.2) over the future
of the innovations now yields the result. Finally (iii) follows directly from (5.1), (3.1a) and (3.4a) □.

**Remark 5.2.** Notice that (i) in Theorem 5.1 is a representation of \( \hat{y}(0) \) in terms of the data \( y(t), t \neq 0 \). Also in (iii) \( \hat{y}(0) \) is phrased in terms of forward and backward innovations which can be computed from \( \{y(t), t \neq 0\} \). On the contrary, the forward innovations \( \{u_\alpha(t), t \geq 0\} \) appearing in (ii) cannot be calculated from the data \( \{y(t), t \neq 0\} \). However, such a representation will lead us to an important spectral domain result, cf. (ii) in Theorem 6.1 □.


Consider the spectral representation of \( y \) [5]

\[
y(t) = \int_{-\pi}^{\pi} e^{i\omega t} \, du(\omega)
\]

where \( u \) is a vector orthogonal stochastic measure such that

\[
\mathbb{E}\{du(\omega)du(\omega)\dagger\} = \frac{\Phi(e^{i\omega})}{2\pi} \, d\omega
\]

the symbol \( \dagger \) denoting complex conjugation and transposition.

**Theorem 6.1.** The optimal interpolator can be represented as follows.

(i) \( \hat{y}(0) = \int_{-\pi}^{\pi} (K^*p^{-1}p_\ast^{-1}(e^{-i\omega}I - \Gamma_\ast)^{-1}B_\ast^*B_\ast^{-1} + K^*p^{-1}p_\ast^{-1}(e^{-i\omega}I - \Gamma_\ast)^{-1}B_\ast^*B_\ast^{-1}d\mu(\omega)) \);