# Duals of random vectors and processes with applications to prediction problems with missing values

Yukio Kasahara[a], Mohsen Pourahmadi[*,b], Akihiko Inoue[c]

[a]*Department of Mathematics, Hokkaido University, Sapporo 060-0810, Japan*
[b]*Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA*
[c]*Department of Mathematics, Hiroshima University, Higashi-Hiroshima 739-8526, Japan*

**Abstract**

Important results in prediction theory dealing with missing values have been obtained traditionally using difficult techniques based on duality in Hilbert spaces of analytic functions (Nakazi, 1984; Miamee and Pourahmadi, 1988). We obtain and unify these results using a simple finite-dimensional duality lemma which is essentially an abstraction of a regression property of a multivariate normal random vector (Rao, 1973, p. 524) or its inverse covariance matrix. The approach reveals the roles of duality and biorthogonality of random vectors in dealing with infinite-dimensional and difficult prediction problems. A novelty of this approach is its reliance on the explicit representation of the prediction error in terms of the data rather than the predictor itself as in the traditional techniques. In particular, we find a new and explicit formula for the dual of the semi-finite process $\{X_t; t \leq n\}$ for a fixed $n$, which does not seem to be possible using the existing techniques.

*Key words:* finite prediction problems, biorthogonality, duality, missing values, stationary time series, Wold decomposition
*2008 MSC:* 62M20, 60G10, 60G25

## 1. Introduction

Irregular observations, missing values and outliers are common in time series data (Box and Tiao, 1975, Brubacher and Wilson, 1976). A framework for dealing with such anomalies is that of $X = \{X_t\}_{t\in\mathbb{Z}}$ being a mean-zero, weakly stationary stochastic process with the autocovariance function $\gamma = \{\gamma_k\}_{k\in\mathbb{Z}}$ and the spectral density function $f$, where the problem can be formulated as that of predicting or approximating an unknown value $X_0$ based on the observed values $\{X_t; t \in S\}$ for a given index set $S \subset \mathbb{Z} \setminus \{0\}$ and the knowledge of the autocovariance of the process. Such a problem is quite important to applications in business, economics, engineering, physical and natural sciences, and belongs to the area of prediction theory of stationary stochastic processes developed by Wiener (1949) and Kolmogorov (1941). By restricting attention to linear predictors and using the least-squares criterion to assess the goodness of predictors, a successful solution seeks to address the following two goals:

($P_1$) Express the linear least-squares predictor of $X_0$, denoted by $\hat{X}_0(S)$, and the prediction error $X_0 - \hat{X}_0(S)$ in terms of the observable $\{X_t; t \in S\}$.

($P_2$) Express the prediction error variance $\sigma^2(S) = \sigma^2(f, S) := E|X_0 - \hat{X}_0(S)|^2$ in terms of $f$.

The focus in prediction theory has been more on the goal ($P_2$), and the celebrated Szegö–Kolmogorov–Wiener theorem gives the variance of the one-step ahead prediction error based on the *infinite* past or for the "half-line" index set $S_0 := \{\ldots, -2, -1\}$ by

$$\sigma^2(f, S_0) = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\lambda) d\lambda\right) > 0 \tag{1}$$

*Corresponding author
*Email address:* `pourahm@stat.tamu.edu` (Mohsen Pourahmadi)

if $\log f$ is integrable, and otherwise $\sigma^2(S_0) = 0$. However, when the first $n$ consecutive integers are removed from $S_0$ or for the index set $S_{-n} := \{\ldots, -n-2, -n-1\}$, $n \geq 0$, the formula for the $(n+1)$-step prediction error variance (Wold, 1954, Kolmogorov, 1941) is

$$\sigma^2(f, S_{-n}) = |b_0|^2 + |b_1|^2 + \cdots + |b_n|^2, \qquad n = 0, 1, \ldots, \tag{2}$$

where $\{b_j\}$, the moving average (MA) coefficients of the process, is related to the Fourier coefficients of $\log f$ and $|b_0|^2 = \sigma^2(S_0)$ (see Nakazi and Takahashi, 1980, and Pourahmadi, 1984; see also Section 3 below).

A result similar to (1) for the interpolation of a single missing value corresponding to the index set $S_\infty := \mathbb{Z} \setminus \{0\}$ was obtained by Kolmogorov (1941). Specifically, the interpolation error variance is given by

$$\sigma^2(f, S_\infty) = \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda)^{-1} d\lambda \right)^{-1} > 0 \tag{3}$$

if $f^{-1} \in L^1 := L^1([-\pi, \pi], d\lambda/(2\pi))$, and otherwise $\sigma^2(S_\infty) = 0$. The corresponding prediction problem for the smaller index set $S_n := \{\ldots, n-1, n\} \setminus \{0\}$, $n \geq 0$, was stated as open in Rozanov (1967, p. 107) and is perhaps one of the most challenging problems in prediction theory next to (1). The index set $S_n$ is, indeed, of special interest as it forms a bridge connecting $S_0$ and $S_\infty$; it reduces to $S_0$ when $n = 0$ and tends to $S_\infty$ as $n \to \infty$. In a remarkable paper, Nakazi (1984) using delicate, but complicated analytical techniques (and assuming that $f^{-1} \in L^1$) showed that

$$\sigma^2(f, S_n) = \left( |a_0|^2 + |a_1|^2 + \cdots + |a_n|^2 \right)^{-1}, \qquad n = 0, 1, \ldots, \tag{4}$$

where $\{a_j\}$ is related to the autoregressive (AR) parameters of the process (see Section 3 below). Comparing (2) and (4), it is natural to ask why there is such an "inverse-dual" relationship between them.

Concerning the question above, it is worth noting that Nakazi's technique, if viewed properly, reduces the computation of $\sigma^2(f, S_n)$ to that of the $(n+1)$-step prediction error variance of another stationary process $\{Y_t\}$ with the spectral density function $f^{-1}$, which we call the *dual* of $\{X_t\}$ (see Definition 1 and Section 4.3). His result and technique have spawned considerable research in this area in the last two decades; see Miamee and Pourahmadi (1988), Miamee (1993), Cheng *et al.* (1998), Frank and Klotz (2002), Klotz and Riedel (2002) and Bondon (2002). A unifying feature of most of the known results thus far seems to be a fundamental duality principle of the form

$$\sigma^2(f, S) \cdot \sigma^2(f^{-1}, S^c) = 1, \tag{5}$$

where $S^c$ is the complement of $S$ in $\mathbb{Z} \setminus \{0\}$ and $f^{-1} \in L^1$; see Cheng *et al.* (1998) and Urbanik (2000).

The first occurrence of (5) seems to be in the 1949 Russian version of Yaglom (1963) for the case of deleting finitely many points from $S_\infty$. Proof of (5), in general, like those of the main results in Nakazi (1984), Miamee and Pourahmadi (1988), Cheng *et al.* (1998), and Urbanik (2000), is long, unintuitive and relies on duality techniques from functional and harmonic analysis and requires $f^{-1} \in L^1$ which is not natural for the index set $S_n$. Surprisingly, an implicit version of (5) had been developed in Grenander and Rosenblatt (1954, Theorem 1) as the limit of a quadratic form involving Szegö's orthogonal polynomials on the unit circle; see also Simon (2005, p. 165). However, it had remained dormant and not used in the context of prediction theory, except in Pourahmadi (1993).

In this paper, we establish a finite-dimensional duality principle (Lemma 1) which involves the notion of dual of a random vector, and show that some prediction problems, including the above and some new ones, which are related to removing a finite number of indices from $S_n$ and $S_\infty$, can be solved in a unified manner. In Section 2, we present the main lemma, some auxiliary facts about dual of a random vector and their consequences for computing the prediction error variances and predictors. In Section 3, using the lemma we first solve three finite prediction problems for $X_0$ based on the knowledge of $\{X_t; t \in K\}$ with $K = \{-m, \ldots, n\} \setminus (M \cup \{0\})$, $m, n \geq 0$, where $M$, the index set of the missing values, is relatively small. Then, we obtain the solutions of Kolmogorov, Nakazi, and Yaglom's prediction problems in a unified manner by studying the limit of the solutions by letting $m \to \infty$, followed by $n \to \infty$. As a consequence, we find a new and explicit formula for the dual of the random process $\{X_t; t \leq n\}$ for a fixed $n$, which does not seem to be possible using the technique of Urbanik (2000), Klotz and Riedel (2002) and Frank and Klotz (2002). This is particularly useful in developing series representations for predictors and interpolators, and sheds light on the approaches of Bondon (2002) and Salehi (1979). In Section 5, we close the paper with some discussions.

## 2. A finite-dimensional duality principle

In this section, we introduce dual of a random vector, study its properties and state a finite-dimensional duality lemma related to it. The lemma will be used in Section 3 to obtain an explicit formula for the dual of the process $\{X_t : t \leq n\}$ for a fixed $n$, and to solve and unify various challenging prediction problems through the limits of the solutions of their finite past counterparts.

Let $H$ be the class of random variables with zero-mean and finite variance with the inner product $(Y, Z) := E[Y\overline{Z}]$ and norm $\|Y\| := E[|Y|^2]^{1/2}$. For a finite index set $N$, we put $H_N := \{X = (X_j)_{j\in N} : X_j \in H, j \in N\}$.

**Definition 1.** A random vector $Y \in H_N$ is called the *dual of* $X \in H_N$ if it satisfies the following conditions:

(i) The components $Y_j$, $j \in N$, belong to $\mathrm{sp}\{X_k; k \in N\}$.

(ii) $X$ and $Y$ are *biorthogonal*: $(X_i, Y_j) = \delta_{ij}$ for $i, j \in N$, or $\mathrm{Cov}(X, Y) = I$.

For $X \in H_N$, $l \in N$ and $K \subset N$, we write $\hat{X}_l(K)$ for the linear least squares predictor of $X_l$ based on $\{X_k; k \in K\}$, i.e., the orthogonal projection of $X_l$ onto $\mathrm{sp}\{X_k; k \in K\}$.

From the two representations in Proposition 3 (2), (3) in the Appendix for the dual $Y$, we find the following representation for the standardized interpolation error is immediate:

$$\frac{X_i - \hat{X}_i(N_i)}{\|X_i - \hat{X}_i(N_i)\|^2} = \sum\nolimits_{j\in N} \gamma^{i,j} X_j \quad \text{with} \quad N_i = N \setminus \{i\}.$$

In particular, $\gamma^{i,i} = 1/\|X_i - \hat{X}_i(N_i)\|^2$. Notice that these equalities hold even if $\Gamma$ is not a Toeplitz matrix or when $X$ is not a segment of a stationary process. For some statistical/physical interpretations of the entries of $\Gamma^{-1} = (\gamma^{i,j})_{i,j\in N}$, the inverse of a stationary covariance matrix, see Bhansali (1990) and references therein.

Now, we are ready to state the main duality lemma.

**Lemma 1.** *Let $N$ be a finite index set. Assume that $X \in H_N$ has the dual $Y \in H_N$ and that $K$, $M$ and a singleton $\{l\}$ partition $N$, i.e., $N = K \cup \{l\} \cup M$ (disjoint union).*

(a) *It holds that* $X_l - \hat{X}_l(K) = \dfrac{Y_l - \hat{Y}_l(M)}{\|Y_l - \hat{Y}_l(M)\|^2}$.

(b) *It holds that* $\|X_l - \hat{X}_l(K)\| = \dfrac{1}{\|Y_l - \hat{Y}_l(M)\|}$.

(c) *Let $\Gamma = (\gamma_{i,j})_{i,j\in N}$ be the covariance matrix of $X$ and write $\Gamma^{-1} = (\gamma^{i,j})_{i,j\in N}$. Then*

$$X_l - \hat{X}_l(K) = \sum\nolimits_{i\in M\cup\{l\}} \alpha'_i Y_i, \tag{6}$$

$$\|X_l - \hat{X}_l(K)\|^2 = \alpha'_l, \tag{7}$$

*where $(\alpha'_i)_{i\in M\cup\{l\}}$ is the solution to the following system of linear equations:*

$$\sum\nolimits_{i\in M\cup\{l\}} \alpha'_i \gamma^{i,j} = \delta_{lj}, \qquad j \in M \cup \{l\}. \tag{8}$$

*In particular, the prediction error variance $\sigma_l^2(K) = \|X_l - \hat{X}_l(K)\|^2$ is given by*

$$\sigma_l^2(K) = \textit{the } (l,l)\textit{-entry of the inverse of } (\gamma^{i,j})_{i,j\in M\cup\{l\}}, \tag{9}$$

*and the predictor coefficients $\alpha_k$ in $\hat{X}_l(K) = \sum_{k\in K} \alpha_k X_k$ are given by*

$$\alpha_k = -\sum\nolimits_{i\in M\cup\{l\}} \alpha'_i \gamma^{i,k}, \qquad k \in K, \tag{10}$$

*whence we have*

$$\hat{X}_l(K) = -\sum\nolimits_{k\in K} \left(\sum\nolimits_{i\in M\cup\{l\}} \alpha'_i \gamma^{i,k}\right) X_k. \tag{11}$$

Proof. Since $X$ and $Y$ are minimal and biorthogonal, $X_l - \hat{X}_l(K)$ and $Y_l - \hat{Y}_l(M)$ are nonzero and belong to the same one-dimensional space, that is, the orthogonal complement of $\mathrm{sp}\{X_j;\, j \in K\} \oplus \mathrm{sp}\{Y_j;\, j \in M\}$ in $\mathrm{sp}\{X_j;\, j \in N\}$. Therefore, one is a multiple of the other; for some $c \in \mathbb{C}$, $X_l - \hat{X}_l(K) = c\{Y_l - \hat{Y}_l(M)\}/\|Y_l - \hat{Y}_l(M)\|^2$. But, since $c$ is equal to

$$c\,\frac{(Y_l - \hat{Y}_l(M), Y_l - \hat{Y}_l(M))}{\|Y_l - \hat{Y}_l(M)\|^2} = c\,\frac{(Y_l - \hat{Y}_l(M), Y_l)}{\|Y_l - \hat{Y}_l(M)\|^2} = (X_l - \hat{X}_l(K), Y_l) = (X_l, Y_l) = 1,$$

we get the assertions in (a) and (b).

Next, we prove the assertions in (c). Since $Y_j$'s are linearly independent, (a) shows that $X_l - \hat{X}_l(K)$ is uniquely expressed in the form (6). Then $\alpha'_l = \|Y_l - \hat{Y}_l(M)\|^{-2}$, which, in view of (b), is equal to $\|X_l - \hat{X}_l(K)\|^2$, and (7) holds. From $(X_i, Y_j) = \delta_{ij}$ and $(Y_i, Y_j) = \gamma^{i,j}$, it follows that the predictor coefficients $\alpha_k$ in $\hat{X}_l(K) = \sum_{k\in K} \alpha_k X_k$ satisfy

$$\alpha_k = (\hat{X}_l(K), Y_k) = \left(X_l - \sum\nolimits_{i\in M\cup\{0\}} \alpha'_i Y_i, Y_k\right) = -\sum\nolimits_{i\in M\cup\{0\}} \alpha'_i \gamma^{i,k}.$$

Thus (10), whence (11). Similarly, for $j \in M \cup \{l\}$, we have $(\hat{X}_l(K), Y_j) = 0$ and

$$\sum\nolimits_{i\in M\cup\{l\}} \alpha'_i \gamma^{i,j} = \sum\nolimits_{i\in M\cup\{l\}} \alpha'_i (Y_i, Y_j) = (X_l - \hat{X}_l(K), Y_j) = \delta_{lj}.$$

Therefore, (8) follows. Finally, we obtain (9) from (7) and (8).

## 3. Applications to prediction problems

In this section, we illustrate the role of the Lemma 1(c) in unifying diverse prediction problems and finding an explicit formula for the dual of the random process $\{X_t;\, t \le n\}$ for a fixed $n$.

Throughout this section we assume that $\{X_j\}_{j\in\mathbb{Z}}$ is a mean zero, purely nondeterministic stationary process, so it admits the MA representation (Wold decomposition)

$$X_j = \sum\nolimits_{k=-\infty}^{j} b_{j-k}\varepsilon_k, \qquad j \in \mathbb{Z}, \tag{12}$$

where $\{\varepsilon_j\}_{j\in\mathbb{Z}}$ is the normalized innovation of $\{X_j\}_{j\in\mathbb{Z}}$ defined by

$$\varepsilon_j := \{X_j - \hat{X}_j(\{\dots, j-2,, j-1\})\}/\|X_j - \hat{X}_j(\{\dots, j-2,, j-1\})\|, \qquad j \in \mathbb{Z},$$

and $\{b_k\}_{k=0}^{\infty}$ is the MA coefficients given by $b_k := (X_0, \varepsilon_{-k})$. We define a sequence of complex numbers $\{a_k\}_{k=0}^{\infty}$ by

$$\sum\nolimits_{k=0}^{j} b_k a_{j-k} = \delta_{0j}, \qquad j \ge 0. \tag{13}$$

If the series $\sum_{j=0}^{\infty} a_j X_{-j}$ is mean-convergent, then (12) can be inverted as

$$\varepsilon_j = \sum\nolimits_{k=-\infty}^{j} a_{j-k}X_k, \qquad j \in \mathbb{Z}. \tag{14}$$

This is essentially the same as the AR representation (see Pourahmadi, 2001), and we call $\{a_k\}$ the AR coefficients of $\{X_j\}_{j\in\mathbb{Z}}$. As suggested in (2) and (4), these $\{b_k\}$ and $\{a_k\}$ play important roles in prediction theory.

### *3.1. Finite prediction problems with missing values*

Let $M$ be a finite set of integers that does not contain zero. Throughout this section, it represents the index set of missing (unknown) values when predicting $X_0$. For given $M$, we take the integers $m, n \ge 0$ so large that $M \subset N := \{-m, \dots, n\}$, and put $K = N \setminus (M \cup \{0\})$, which represents the index set of the observed values, then we have the partition $N = K \cup \{0\} \cup M$ as in Lemma 1. We start with the prediction problem for a finite index set $K$. Once the problem is solved for such a $K$, the solutions for infinite index sets $S_n \setminus M$ and $S_\infty \setminus M$ are obtained by taking the limit of the solutions, first as $m \to \infty$, and then $n \to \infty$.

Traditionally, the predictor coefficients $\alpha_k$ in $\hat{X}_0(K) = \sum_{k\in K} \alpha_k X_k$ and the prediction error variance $\sigma^2(K) = \|X_0 - \hat{X}_0(K)\|^2$ are computed from $(\gamma_{i,j})_{i,j\in K\cup\{0\}}$ by solving the normal equations:

$$\sum\nolimits_{k\in K} \alpha_k \gamma_{k,j} = \gamma_{0,j}, \quad j \in K, \qquad \sigma^2(K) = \gamma_{0,0} - \sum\nolimits_{k\in K} \alpha_k \gamma_{k,0}. \tag{15}$$

However, the results so obtained are not convenient for studying the asymptotic behaviors of the finite prediction error variance and the predictor coefficients as $m \to \infty$ and/or $n \to \infty$. The problem is made much simpler by using the Lemma 1 and some basic facts about the finite MA and AR representations, see (19)–(21) below.

For the (future) segment $\{X_j\}_{j=0}^{\infty}$ of a stationary process, define its normalized innovation $\{\varepsilon_{j,0}\}_{j=0}^{\infty}$ using the Gram–Schmidt method: Set $\varepsilon_{0,0} := X_0/\|X_0\|$ and

$$\varepsilon_{j,0} := \{X_j - \hat{X}_j(\{0,\dots,j-1\})/\|X_j - \hat{X}_j(\{0,\dots,j-1\})\|, \qquad j \geq 1.$$

Then $\{X_j\}$ and $\{\varepsilon_{j,0}\}$ admit the following finite MA and AR representations:

$$X_j = \sum\nolimits_{k=0}^{j} b_{j-k,j}\varepsilon_{k,0}, \qquad \varepsilon_{j,0} = \sum\nolimits_{k=0}^{j} a_{j-k,j} X_k, \qquad j \geq 0.$$

Here $\{b_{k,j}\}_{k=0}^{j}$ is defined by $b_{k,j} := (X_j, \varepsilon_{j-k,0})$ and $\{a_{k,j}\}_{k=0}^{j}$ by $\sum_{k=i}^{j} b_{j-k,j} a_{k-i,k} = \delta_{ij}$ or $\sum_{k=i}^{j} a_{j-k,j} b_{k-i,k} = \delta_{ij}$ for $i \leq j$. These finite MA and AR coefficients converge to their infinite counterparts:

$$\lim_{j\to\infty} b_{k,j} = b_k, \qquad \lim_{j\to\infty} a_{k,j} = a_k. \tag{16}$$

If we consider $\{X_j\}_{j=-m}^{\infty}$ instead of $\{X_j\}_{j=0}^{\infty}$, then by stationarity, it follows that

$$X_j = \sum\nolimits_{k=-m}^{j} b_{j-k,m+j}\varepsilon_{k,-m}, \qquad \varepsilon_{j,-m} = \sum\nolimits_{k=-m}^{j} a_{j-k,m+j} X_k, \qquad j \geq -m, \tag{17}$$

where $\{\varepsilon_{j,-m}\}_{j=-m}^{\infty}$ , the normalized innovation of $\{X_j\}_{j=-m}^{\infty}$, is defined in the usual manner. Note that

$$\varepsilon_j = \lim_{m\to\infty} \varepsilon_{j,-m}, \qquad j \in \mathbb{Z}, \tag{18}$$

so that the representations in (17) reduce to (12) and (14) as $m \to \infty$.

Now, we present the key ingredients for applying the Lemma 1 to some prediction problems. Recall that $N = \{-m,\dots,n\}$ and $X$ is the vector $(X_j)_{j\in N}$ with covariance matrix $\Gamma = (\gamma_{i-j})_{i,j\in N}$. From Proposition 3 (3), its dual $Y$ is given by $Y = \Gamma^{-1}X$. Let $\varepsilon$ be the normalized innovation vector of $X$, i.e., $\varepsilon := (\varepsilon_{j,-m})_{j\in N}$. Then from (17) we have $X = B\varepsilon$ and $\varepsilon = AX$, where $A$ and $B$ are the lower triangular matrices with $(i, j)$-entries $a_{i-j,m+i}$ and $b_{i-j,m+i}$ for $-m \leq j \leq i \leq n$, respectively. Since $A = B^{-1}$ and $\Gamma = BB^*$, we have $\Gamma^{-1} = A^*A$ and $Y = A^*\varepsilon$. Thus, the $(i, j)$-entry $\gamma^{i,j}$ of $\Gamma^{-1}$ and the $j$-th entry $Y_j$ of $Y$ have the representations

$$\gamma^{i,j} = \sum\nolimits_{k=i\vee j}^{n} \overline{a}_{k-i,m+k} a_{k-j,m+k}, \qquad Y_j = \sum\nolimits_{k=j}^{n} \overline{a}_{k-j,m+k}\varepsilon_{k,-m}, \tag{19}$$

which we show are suitable for studying their limits as first $m \to \infty$ and then $n \to \infty$; see (16) and (18).

Now, we are ready to express the predictor $\hat{X}_0(K)$, the prediction error $X_0 - \hat{X}_0(K)$ and its variance $\sigma^2(K)$ using Lemma 1 (c). In particular, it follows from (6), (9) and (19) that

$$\sigma^2(K) = \text{the } (0,0)\text{-entry of the inverse of } \left(\sum\nolimits_{k=i\vee j}^{n} \overline{a}_{k-i,m+k} a_{k-j,m+k}\right)_{i,j\in M\cup\{0\}} \tag{20}$$

and

$$X_0 - \hat{X}_0(K) = \sum\nolimits_{i\in M\cup\{0\}} \alpha'_i \left(\sum\nolimits_{k=i}^{n} \overline{a}_{k-i,m+k}\varepsilon_{k,-m}\right), \tag{21}$$

where $\alpha'_i$'s are as in Lemma 1 (c) with $l = 0$.

*3.2. Examples of finite prediction problems*

In this section, we highlight some important consequences of (20) and (21) by presenting a few special cases corresponding to the classical prediction problems of Kolmogorov (1941), Yaglom (1963) and Nakazi (1984). These are listed as examples next according to the cardinality of the index set $M$ of the missing values. They are useful in illustrating the procedure for obtaining prediction-theoretic results for the two infinite index sets $S = S_n \setminus M$ and $S = S_\infty \setminus M$ from their finite-dimensional counterparts.

Since most classic prediction results and conditions are stated in terms of the spectral density function, it is instructive to connect the time-domain results in (12)–(14) on the MA and AR coefficients to the spectral density of the process. To this end, we recall that when $\{X_j\}_{j\in\mathbb{Z}}$ is purely nondeterministic, it has the spectral density function $f$ with $\log f \in L^1$. Thus, there exists an outer function $h$ in the Hardy class $H^2$ such that $f = |h|^2$ and $h(0) > 0$, and we have

$$h(z) = \sum_{k=0}^{\infty} b_k z^k, \qquad \frac{1}{h(z)} = \sum_{k=0}^{\infty} a_k z^k \tag{22}$$

in the unit disc. In particular, this shows that $f^{-1} \in L^1$ if and only if $\{a_k\}$ is square summable, see Pourahmadi (2001, Chap. 8).

**Example 1 (The Finite Kolmogorov–Nakazi Problem).** This finite interpolation problem corresponds to $M = \phi$ (the empty set) and $K = \{-m, \ldots, n\} \setminus \{0\}$, so that the solution of (8) is given by $\alpha'_0 = 1/\gamma^{0,0}$. Consequently, from (19)–(21), we have

$$\sigma^2(K) = \left(\sum_{k=0}^{n} |a_{k,m+k}|^2\right)^{-1} \tag{23}$$

and

$$X_0 - \hat{X}_0(K) = \left(\sum_{k=0}^{n} |a_{k,m+k}|^2\right)^{-1} \sum_{k=0}^{n} \bar{a}_{k,m+k}\varepsilon_{k,-m}. \tag{24}$$

Next, we show that (23) and (24) are, indeed, precursors of some important results in prediction theory due to Kolmogorov (1941), Masani (1960), and Nakazi (1984). First, the result in (4) for the infinite index set $S_n = \{\ldots, n-1, n\} \setminus \{0\}$ which is due to Nakazi (1984) was obtained using tedious duality arguments from functional and harmonic analyses and under a very restrictive condition. Here we obtain it simply by taking the limit of (23) as $m \to \infty$ (without assuming $f^{-1} \in L^1$). Indeed, from (16) and (23) it is immediate that

$$\sigma^2(S_n) = \left(\sum_{k=0}^{n} |a_k|^2\right)^{-1}. \tag{25}$$

Furthermore, in view of (18), it follows from (24) that

$$X_0 - \hat{X}_0(S_n) = \left(\sum_{k=0}^{n} |a_k|^2\right)^{-1} \sum_{k=0}^{n} \bar{a}_k\varepsilon_k, \tag{26}$$

which provides an explicit formula for the dual of the semi-finite process $\{X_t; t \le n\}$ for a fixed $n$.

The solution (3) of the Kolmogorov (1941) interpolation problem with $S_\infty = \mathbb{Z} \setminus \{0\}$ also follows from (25) by simply taking the limit as $n \to \infty$, provided that $\{a_k\}$ is square summable. Thus, as in Kolmogorov (1941), assuming that $\{X_t\}$ is minimal or $f^{-1} \in L^1$, we obtain

$$\sigma^2(S_\infty) = \left(\sum_{k=0}^{\infty} |a_k|^2\right)^{-1} = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda)^{-1} d\lambda\right)^{-1}.$$

Under the same minimality condition, the limit of (26) as $n \to \infty$, leads to

$$X_0 - \hat{X}_0(S_\infty) = \left(\sum_{k=0}^{\infty} |a_k|^2\right)^{-1} \sum_{k=0}^{\infty} \bar{a}_k\varepsilon_k,$$

which is Masani's representation of the two-sided innovation of $\{X_j\}$ at time 0 (Masani, 1960), which is a moving average of the future innovations.

In fact, the origin of above moving average representation can be traced to (19) and (26). A version of (26) seems to have appeared first in Box and Tiao (1975) in the context of intervention analysis; see Pourahmadi (1989, and 2001, Sect. 8.4) for a more rigorous derivation, detailed discussion and connection with outlier detection.

Our second example corresponds to $M$ having one element and hence involves inversion of $2 \times 2$ matrices, no matter how large $K$ is.

**Example 2 (The Finite Past with a Single Missing Value).** This problem corresponds to $m > 0$, $n = 0$,

$$K = \{-m, \ldots, -1\} \setminus \{-u\}, \qquad M = \{-u\},$$

where $1 \le u \le m$, so that $X_{-u}$ from the finite past of length $m$ is missing. From (19), the $2 \times 2$ matrix for solving (8) is

$$\begin{pmatrix} \gamma^{-u,-u} & \gamma^{-u,0} \\ \gamma^{0,-u} & \gamma^{0,0} \end{pmatrix} = \begin{pmatrix} \sum_{k=0}^{u} |a_{u-k,m-k}|^2 & a_{0,m}\, \bar{a}_{u,m} \\ \bar{a}_{0,m}\, a_{u,m} & |a_{0,m}|^2 \end{pmatrix}.$$

Hence, using the subscript $m$ to emphasize the dependence on $m$, we have

$$\alpha'_{0,m} = \frac{1}{\Delta_m} \sum_{k=0}^{u} |a_{u-k,m-k}|^2, \qquad \alpha'_{-u,m} = -\frac{\bar{a}_{0,m} a_{u,m}}{\Delta_m},$$

where $\Delta_m = |a_{0,m}|^2 \sum_{k=1}^{u} |a_{u-k,m-k}|^2$. Thus, from (20) and (21),

$$\sigma^2(K) = \frac{\sum_{k=0}^{u} |a_{u-k,m-k}|^2}{|a_{0,m}|^2 \sum_{k=1}^{u} |a_{u-k,m-k}|^2}, \qquad X_0 - \hat{X}_0(K) = \alpha'_{0,m} \bar{a}_{0,m} \varepsilon_{0,-m} + \alpha'_{-u,m} \sum_{k=0}^{u} \bar{a}_{u-k,m-k} \varepsilon_{-k,-m}, \tag{27}$$

and, taking the limit as $m \to \infty$,

$$\sigma^2(S_0 \setminus \{-u\}) = |b_0|^2 \frac{\sum_{k=0}^{u} |a_k|^2}{\sum_{k=0}^{u-1} |a_k|^2}, \qquad X_0 - \hat{X}_0(S_0 \setminus \{-u\}) = \alpha'_0 \bar{a}_0 \varepsilon_0 + \alpha'_{-u} \sum_{k=0}^{u} \bar{a}_{u-k} \varepsilon_{-k}, \tag{28}$$

where $\alpha'_0$ and $\alpha'_{-u}$ are the limits of $\alpha'_{0,m}$ and $\alpha'_{-u,m}$, as $m \to \infty$, respectively. The expressions in (28) were obtained first in Pourahmadi (1992); see also Pourahmadi and Soofi (2000) and Pourahmadi (2001, Section 8.3). However, those in (27) are new and have not appeared before.

For $n > 0$, slightly more general calculations leading to analogues of (27) and (28) can be used to show in a more rigorous manner that the inverse autocorrelation function of $\{X_t\}$ at lag $u$ is the negative of the partial correlation between $X_0$ and $X_u$ after elimination of the effects of $X_t$, $t \neq 0, u$, as shown formally in Kanto (1984) for processes with strictly positive spectral density functions.

**Example 3 (The Finite Yaglom Problem).** There are many situations where the cardinality of $M$ is two or more; see Pourahmadi *et al.* (2007), Box and Tiao (1975), Brubacher and Wilson (1976), Damsleth (1980), Abraham (1981), and there are several ad hoc methods for interpolating the missing values. For example, Brubacher and Wilson (1976) minimizes $\sum_{-m}^{n} \varepsilon_j^2 = \sum_{-m}^{n} (\sum_{k=-\infty}^{j} a_{j-k} X_k)^2$ with respect to the unknown $X_j$, $j \in M \cup \{0\}$, and then study the solution of the normal equations as $m, n \to \infty$. Budinsky (1989) has shown that this approach under some conditions gives the same result as the more rigorous approach of Yaglom (1963). In applying Lemma 1 (c) to this problem, we first note that, due to the large cardinality of $M$, handling (20) and (21) via (8) does not lead to simple explicit formulas as in (27) and (28). Nevertheless, the limits of the expressions in (20) and (21) as first $m \to \infty$, and then as $n \to \infty$ (assuming $f^{-1} \in L^1$) have simple forms in terms of the AR parameters:

$$X_0 - \hat{X}_0(S) = \sum_{i \in M \cup \{0\}} \alpha'_i \sum_{k=i}^{\infty} \bar{a}_{k-i}\, \varepsilon_k, \qquad \sigma^2(S) = \text{the } (0,0)\text{-entry of the inverse of } \Big( \sum_{k=i \vee j}^{\infty} \bar{a}_{k-i}\, a_{k-j} \Big)_{i,j \in M \cup \{0\}}. \tag{29}$$

Using (22) and writing the entries of the above matrix, in terms of the Fourier coefficients of $f^{-1}$, it follows that (29) reduces to the results in Yaglom (1963); see also Salehi (1979).

## 4. Applications to Series Representations

The Wold decomposition (12) expresses predictors and prediction errors in terms of the innovation process $\{\varepsilon_t\}$. This works well for achieving the goal ($P_2$) in Section 1, but since the innovation $\varepsilon_t$ is not directly observable the resulting predictor formulas are not suitable for computation. To get around this difficulty, one must express the innovations or the predictors in terms of the past observations. In this section, we obtain series representations for predictors and interpolators in terms of the observed values.

### 4.1. *The Infinite past and the Wold decomposition*

An alternative method of solving prediction problems for $S = S_n \setminus M$ is to reduce them to a slightly different class of finite prediction problems than those in Section 3.2, using the Wold decomposition of a purely nondeterministic stationary process.

As in Section 3.1, write $N = \{-m, \dots, n\}$ and $N = K \cup \{0\} \cup M$ (disjoint union), so that

$$S = S_n \setminus M = \{\dots, -m-2, -m-1\} \cup K \quad \text{(disjoint union).}$$

For $j \geq -m$, let $\hat{X}_j$ be the linear least-squares predictor of $X_j$ based on the infinite past $\{X_k; k < -m\}$. Then, by (12), we have $X_j - \hat{X}_j = \sum_{k=-m}^{j} b_{j-k}\varepsilon_k$, $j \geq -m$, which are orthogonal to $\overline{\text{sp}}\{X_j;\, j < -m\}$, and it follows that

$$\overline{\text{sp}}\{X_j;\, j \in S\} = \text{sp}\{X_j - \hat{X}_j;\, j \in K\} \oplus \overline{\text{sp}}\{X_j;\, j < -m\}.$$

This equality plays the key role in finding the predictor of $X_0$ and its prediction error variance, based on $\{X_j;\, j \in S\}$. In fact, by using it, we only have to solve the problem of predicting $X_0 - \hat{X}_0$ based on $\{X_j - \hat{X}_j;\, j \in K\}$. More precisely, we consider the vector $X' := (X_j - \hat{X}_j)_{j \in N}$ with the covariance matrix $G = (g_{i,j})_{i,j \in N}$, where $g_{i,j} := \sum_{k=-m}^{i \wedge j} b_{i-k}\overline{b}_{j-k}$ (see Pourahmadi, 2001, p. 273). Then, writing $X_0 = \hat{X}_0 + (X_0 - \hat{X}_0)$, we get

$$\hat{X}_0(S) = \hat{X}_0 + \sum\nolimits_{k \in K} \alpha_k (X_k - \hat{X}_k), \qquad \sigma^2(S) = \left\| (X_0 - \hat{X}_0) - \sum\nolimits_{k \in K} \alpha_k (X_k - \hat{X}_k) \right\|^2, \tag{30}$$

where $\sum_{k \in K} \alpha_k (X_k - \hat{X}_k)$ is the predictor of $X_0 - \hat{X}_0$ based on $\{X_k - \hat{X}_k; k \in K\}$, and the predictor coefficients $\alpha_k$ and prediction error variance $\sigma^2(S)$ are usually obtained using (15) with $\gamma_{i,j}$ replaced by $g_{i,j}$. Here we apply Lemma 1 (c) to the above finite prediction problem for $X'$. In so doing, the following representations of the $(i, j)$-entry $g^{i,j}$ of $G^{-1}$ and the $j$-th entry $Y_j$ of the dual $Y$ of $X'$ are available:

$$g^{i,j} = \sum\nolimits_{k=i \vee j}^{n} \overline{a}_{k-i} a_{k-j}, \qquad Y_j = \sum\nolimits_{k=j}^{n} \overline{a}_{k-j} \varepsilon_k. \tag{31}$$

In fact, these are obtained from using (13) and Proposition 3 (3) or by letting $m \to \infty$ in (19). These explicit representations turn out to be crucial for finding series representations for certain predictors and interpolators discussed in the next two subsections.

### 4.2. *Series representation of the predictors based on incomplete infinite past*

In this section, we obtain series representations for the predictors in terms of the observed values from an incomplete past. A novelty of our approach is its reliance on the representation of the prediction error in terms of the dual of the random vector $Y$ in (31), hence the solution of the problem (P$_1$) for $S = S_n \setminus M$ is more direct and simpler than the procedures of Bondon (2002, Theorem 3.1) and Nikfar (2006).

Let $f_{j,k} := -\sum_{i=0}^{k} b_{k-i} a_{j+i}$ be the coefficients of the $(k+1)$-step ahead predictor based on the infinite past $S_0 = \{\dots, -2, -1\}$, i.e., $\hat{X}_k(S_0) = \sum_{j=1}^{\infty} f_{j,k} X_{-j}$ for $k = 0, 1, \dots$. Then, assuming that $\{X_j\}_{j \in \mathbb{Z}}$ has the mean-convergent AR representation (14), it follows from (30) with $S = \{\dots, -m-2, -m-1\} \cup K$ that

$$\hat{X}_0(S) = \sum\nolimits_{k \in K} \alpha_k X_k + \sum\nolimits_{j=1}^{\infty} \left( f_{j,m} - \sum\nolimits_{k \in K} \alpha_k f_{j,m+k} \right) X_{-m-j}.$$

On the other hand, from (6) and (31), we have

$$\hat{X}_0(S) = X_0 - \sum\nolimits_{i \in M \cup \{0\}} \alpha'_i \left( \sum\nolimits_{k=i}^{n} \overline{a}_{k-i} \varepsilon_k \right),$$

where replacing in for $\varepsilon_k$ from (14) and after some algebra, we get the following alternative series representation for the predictor of $X_0$ based on the incomplete past:

$$\hat{X}_0(S) = -\sum\nolimits_{j \in S} \left( \sum\nolimits_{i \in M \cup \{0\}} \alpha'_i \sum\nolimits_{k=i \vee j}^{n} \overline{a}_{k-i} a_{k-j} \right) X_j. \tag{32}$$

Then, the prediction error has the following representation:

$$X_0 - \hat{X}_0(S) = \sum_{i\in M\cup\{0\}} \alpha'_i \left(\sum_{k=i}^{n} \overline{a}_{k-i}\varepsilon_k\right), \tag{33}$$

in terms of the dual $Y$ in (31). Furthermore, it follows that the sequence $\{\sum_{k=j}^{n} \overline{a}_{k-j}\varepsilon_k\}_{j=-\infty}^{n}$ spans $\overline{\mathrm{sp}}\{X_j;\ j \leq n\}$, the infinite past up to $n$ of the process $\{X_t\}$.

The formulas (32) and (33) were obtained initially by Bondon (2002, Theorem 3.2) without using the notion of duality.

*4.3. Series representation of the interpolators*

Series representation for the interpolator of $X_0$ based on the observed values from the index set $S = S_\infty \setminus M = \mathbb{Z} \setminus (M \cup \{0\})$ was obtained by Salehi (1979). Here we obtain such representation using the idea of the dual process. Assuming $f^{-1} \in L^1$ or $\sum_{j=0}^{\infty} |a_j|^2 < \infty$, the process $\xi_j := \sum_{k=j}^{\infty} \overline{a}_{k-j}\varepsilon_k,\ j \in \mathbb{Z}$, is well-defined in the sense of mean-square convergence. The process $\{\xi_j\}$ has already appeared in prediction theory and time series analysis, and is called the *standardized two-sided innovation* (Masani, 1960) or the *inverse process* (Cleveland, 1972) of $\{X_t\}_{t\in\mathbb{Z}}$.

From (12), (13), and the above results, we have the following:

(i) $(X_i, \xi_j) = \delta_{ij}$ for $i, j \in \mathbb{Z}$.

(ii) $\xi_j = \{X_j - \hat{X}_j(\mathbb{Z} \setminus \{j\})\}/\|X_j - \hat{X}_j(\mathbb{Z} \setminus \{j\})\|^2$ for $j \in \mathbb{Z}$.

(iii) $\{\xi_j;\ j \in \mathbb{Z}\}$ spans the space $\overline{\mathrm{sp}}\{X_j;\ j \in \mathbb{Z}\}$.

(iv) $\{\xi_j;\ j \in \mathbb{Z}\}$ is a stationary process with the autocovariance function $\gamma^j := (2\pi)^{-1}\int_{-\pi}^{\pi} e^{-ij\lambda} f(\lambda)^{-1} d\lambda,\ j \in \mathbb{Z}$, i.e., $(\xi_i, \xi_j) = \gamma^{i-j} = \sum_{k=i\vee j}^{\infty} \overline{a}_{k-i}a_{k-j}$ for $i, j \in \mathbb{Z}$.

Now, for solving the interpolation problem with $S = \mathbb{Z} \setminus (M \cup \{0\})$, we need to show that $\{\xi_j;\ j \in M \cup \{0\}\}$ spans the orthogonal complement of $\overline{\mathrm{sp}}\{X_j;\ j \in S\}$ in $\overline{\mathrm{sp}}\{X_j;\ j \in \mathbb{Z}\}$. Then, it turns out that there is unique $(\alpha'_j)_{j\in M\cup\{0\}}$ satisfying

$$X_0 - \hat{X}_0(S) = \sum_{i\in M\cup\{0\}} \alpha'_i\xi_i = \sum_{i\in M\cup\{0\}} \alpha'_i \left(\sum_{k=i}^{\infty} \overline{a}_{k-i}\varepsilon_k\right)$$

(see (21) and (33)), and that $\sigma^2(S) = \alpha'_0$. Since $(X_0, \xi_j) - \sum_{i\in M\cup\{0\}} \alpha'_i(\xi_i, \xi_j) = (\hat{X}_0(S), \xi_j) = 0$ for $j \in M \cup \{0\}$, we can compute $(\alpha'_i)_{i\in M\cup\{0\}}$ by solving the following system of linear equations: $\sum_{i\in M\cup\{0\}} \alpha'_i\gamma^{i-j} = \delta_{j0},\ j \in M \cup \{0\}$. As for the interpolator, if $\sum_{j=-\infty}^{\infty} \gamma^j X_{-j}$ is mean-convergent, then $(\xi_j)_{j\in\mathbb{Z}}$ admits the series representation $\xi_i = \sum_{j=-\infty}^{\infty} \gamma^{i-j}X_j$, $i \in \mathbb{Z}$, and we have $\hat{X}_0(S) = -\sum_{j\in S}\left(\sum_{i\in M\cup\{0\}} \alpha_i\gamma^{i-j}\right)X_j$, which is the two-sided version of the formula (32).

## 5. Discussion and future work

We have established the central role of a basic property of the inverse of the covariance matrix of a random vector in providing a time-domain, geometric and finite-dimensional approach to a class of prediction problems for stationary stochastic processes. It brings considerable clarity and simplicity to this area of prediction theory as compared to the classical spectral-domain approach based on analytic function theory and duality in the infinite-dimensional spaces. Since our duality lemma is not confined to stationary processes or Toeplitz matrices, it has the potential of being useful in solving similar prediction problems for nonstationary processes, particularly those with low displacement ranks (Kailath and Sayed, 1995). However, the present form of the lemma does not seem to be useful for predicting infinite-variance- or $L^p$-processes (Cambanis and Soltani, 1984; Cheng *et al.*, 1998).

From application point of view, we note that the two simple formulas (2) and (4) and their extensions provide explicit and informative expressions for the prediction error variances. Like their predecessors (1) and (3), they serve as yardsticks to assess the impact (worth) of observations in predicting $X_0$ when they are added to or deleted from the infinite past and highlight the role of the autoregressive and moving-average parameters for this purpose; see Pourahmadi and Soofi (2000). In fact, Bondon (2002, Theorem 3.3, and 2005) show that a finite number of missing values do not affect the prediction of $X_0$ if and only if the AR parameters corresponding to the indices of those missing values are zero. Furthermore, the examples in Section 3 indicate how the interpolators of the missing values can be computed rigorously without resorting to formal derivations (Box and Tiao, 1975, Brubacher and Wilson, 1976, and Budinsky, 1989).

## A. Dual of a random vector

For the sake of completeness and ease of reference, in the next two propositions, we summarize the characterization, interpretation and other basic information about the dual of a random vector in terms of its covariance matrix and certain prediction errors. Let $N$, $H_N$ and $\hat{X}_l(K)$ for $X \in H_N$, $l \in N$ and $K \subset N$ be as in Section 3.

**Proposition 2.** *For any $X \in H_N$, the following conditions are equivalent:*

(1) *The components $X_j$, $j \in N$, of $X$ are linearly independent.*

(2) *The covariance matrix $\Gamma = (\gamma_{i,j})_{i,j \in N}$ of $X$ with $\gamma_{i,j} = (X_i, X_j)$ is nonsingular.*

(3) *$X$ is minimal: $X_j \notin \mathrm{sp}\{X_i; i \in N,\ i \neq j\}$ for $j \in N$.*

(4) *$X$ has a dual.*

Proof. Clearly, (1)–(3) are equivalent. Assume (3) and define $Y = (Y_j)_{j \in N} \in H_N$ by $Y_j = (X_j - \hat{X}_j(N_j))/\|X_j - \hat{X}_j(N_j)\|^2$, where $N_j := N \setminus \{j\}$. Then $Y_j$ belongs to $\mathrm{sp}\{X_k; k \in N\}$, and $(X_i, Y_j) = \delta_{ij}$ holds:

$$(X_j, Y_j) = \frac{(X_j, X_j - \hat{X}_j(N_j))}{\|X_j - \hat{X}_j(N_j)\|^2} = \frac{(X_j - \hat{X}_j(N_j), X_j - \hat{X}_j(N_j))}{\|X_j - \hat{X}_j(N_j)\|^2} = 1,$$

and for $i \neq j$, $(X_i, Y_j) = (X_i, X_j - \hat{X}_j(N_j)/\|X_j - \hat{X}_j(N_j)\|^2 = 0$. Thus $Y$ is a dual of $X$, and hence (4). Conversely, assume (4) and let $Y$ be a dual of $X$. If $X$ is not minimal, then there exists $j \in N$ such that $X_j \in \mathrm{sp}\{X_i; i \in N,\ i \neq j\}$, that is, $X_j = \sum_{i \neq j} c_i X_i$ for some $c_i \in \mathbb{C}$, and, since $(X_i, Y_j) = 0$ for $i \neq j$, we have $(X_j, Y_j) = \sum_{i \neq j} c_i (X_i, Y_j) = 0$. However, this contradicts $(X_j, Y_j) = 1$. Thus, $X$ is minimal, and (3) follows.

The proof above reveals the importance of the "standardized" interpolation errors of components of $X$ in defining its dual. More explicit representations and other properties of the dual are given next.

**Proposition 3.** *Let $X \in H_N$, with the covariance matrix $\Gamma$, have a dual $Y$. Then, the following assertions hold:*

(1) *The dual $Y$ is unique.*

(2) *The dual $Y$ is given by $Y_j = (X_j - \hat{X}_j(N_j))/\|X_j - \hat{X}_j(N_j)\|^2$ with $N_j := N \setminus \{j\}$ for $j \in N$.*

(3) *The dual $Y$ is also given by $Y = \Gamma^{-1} X$ or $Y_i = \sum_{j \in N} \gamma^{i,j} X_j$, $i \in N$, where $\Gamma^{-1} = (\gamma^{i,j})_{i,j \in N}$.*

(4) *The covariance matrix of $Y$ is equal to $\Gamma^{-1}$.*

(5) *The dual of $Y$ is $X$.*

(6) $\mathrm{sp}\{X_j; j \in N\} = \mathrm{sp}\{Y_j; j \in N\}$.

Proof. First, we prove (1). Let $Z$ be another dual of $X$ and $j \in N$ be fixed. Then $(X_i, Y_j - Z_j) = 0$ for all $i \in N$. However, since $Y_j - Z_j \in \mathrm{sp}\{X_k; k \in N\}$, it follows that $Y_j = Z_j$ and hence (1). (2) follows from the proof of Proposition 2. To prove (3) and (4), we put $Y = \Gamma^{-1} X$. Then $Y_j \in \mathrm{sp}\{X_k; k \in N\}$. Since $\Gamma^{-1}$ is Hermitian, we have

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, X)\, \Gamma^{-1} = \Gamma\Gamma^{-1} = I, \qquad \mathrm{Cov}(Y, Y) = \Gamma^{-1}\mathrm{Cov}(X, X)\, \Gamma^{-1} = \Gamma^{-1}\Gamma\, \Gamma^{-1} = \Gamma^{-1}.$$

Thus (3) and (4) follow. Finally, we obtain (5) and (6) from (3) and (4).

## References


Abraham, B., 1981. Missing observations in time series. Comm. StatistÅ—Theory Methods 10, 1643–1653.

Bhansali, R.J., 1990. On a relationship between the inverse of a stationary covariance matrix and the linear interpolator. J. Appl. Probab. 27, 156–170.

Bondon, P., 2002. Prediction with incomplete past of a stationary process. Stochastic Process. Appl. 98, 67–76.

Bondon, P., 2005. Influence of missing values on the prediction of a stationary time series. J. Time Ser. Anal. 26, 519–525.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. Time series analysis. Forecasting and control. Prentice Hall, Englewood Cliffs, 3rd ed.

Box, G.E.P., Tiao, G.C., 1975. Intervention analysis with applications to economic and environmental problems. J. Amer. Statist. Assoc. 70, 70–79.

Brockwell, P.J., Davis, R.A., 1991. Time series: theory and methods. Springer-Verlag, New York, 2nd ed.

Brubacher, S.R., Wilson, G.T., 1976. Interpolating time series with application to the estimation of holiday effects on electricity demand. Appl. Statist. 25, 107-116.

Budinsky, P., 1989. On the equivalence of two methods for interpolation. Kybernetika 25, 461–466.

Cambanis, S., Soltani, A.R., 1984. Prediction of stable processes: spectral and moving average representations. Z. Wahrsch. Verw. Gebiete 66, 593–612.

Cheng, R., Miamee, A.G., Pourahmadi, M., 1998. Some extremal problems in $L^p(w)$. Proc. Amer. Math. Soc. 126, 2333–2340.

Cleveland, W.S., 1972. The inverse autocorrelations of a time series and their applications. Technometrics 14, 277–293.

Damsleth, E., 1980. Interpolating missing values in a time series. Scand. J. Statist. 7, 33–39.

Frank, M., Klotz, L., 2002. A duality method in prediction theory of multivariate stationary sequences. Math. Nachr. 244, 64–77.

Grenander, U., Rosenblatt, M., 1954. An extension of a theorem of G. Szegö and its application to the study of stochastic processes. Trans. Amer. Math. Soc. 76, 112–126.

Kailath, T., Sayed, A.H., 1995. Displacement structure: theory and applications. SIAM Rev. 37, 297–386.

Kanto, A.J., 1984. A characterization of the inverse autocorrelation function. Comm. Statist. A—Theory Methods 13, 2503–2510.

Klotz, L. Riedel, M., 2002. Some remarks on duality of stationary sequences. Colloq. Math. 92, 225–228.

Kolmogorov, A.N., 1941. Stationary sequences in a Hilbert space. Bull. Univ. Moscow 2, 1–40.

Landau, H.J., 1987. Maximum entropy and the moment problem. Bull. Amer. Math. Soc. 16, 47–77.

Masani, P., 1960. The prediction theory of multivariate stochastic processes. III. Unbounded spectral densities. Acta Math. 104 141–162.

Miamee, A.G., 1993. On basicity of exponentials in $L^p(d\mu)$ and general prediction problems. Period. Math. Hungar. 26, 115–124.

Miamee, A.G., Pourahmadi, M., 1988. Best approximations in $L^p(d\mu)$ and prediction problems of Szeg0̆151, Kolmogorov, Yaglom, and Nakazi. J. London Math. Soc. 38, 133–145.

Nakazi, T., 1984. Two problems in prediction theory. Studia Math. 78, 7–14.

Nakazi, T., Takahashi, K., 1980. Prediction *n* units of time ahead. Proc. Amer. Math. Soc. 80, 658–659.

Nikfar, M., 2006. The generalization of Szegö's theorem, ??, ??–??.

Pourahmadi, M., 1984. Taylor expansion of $\exp(\sum_{k=0}^{\infty} a_k z^k)$ and some applications. Amer. Math. Monthly 91, 303–307.

Pourahmadi, M., 1989. Estimation and interpolation of missing values of a stationary time series. J. Time Ser. Anal. 10, 149–169.

Pourahmadi, M., 1992. Alternating projections and interpolation of stationary processes. J. Appl. Probab. 29, 921–931.

Pourahmadi, M., 1993. Two prediction problems and extensions of a theorem of Szegö. Bull. Iranian Math. Soc. 19, 1–12.

Pourahmadi, M., 2001. Foundations of time series analysis and prediction theory. Wiley, New York.

Pourahmadi, M., Soofi, E.S., 2000. Prediction variance and information worth of observations in time series. J. Time Ser. Anal. 21, 413–434.

Pourahmadi, M., Inoue, A., Kasahara, Y., 2007. A prediction problem in $L^2(w)$. Proc. Amer. Math. Soc. 135, 1233–1239

Rao, C.R., 1973. Linear statistical inference and its applications. Wiley, New York, 2nd ed.

Rozanov, Y.A., 1967. Stationary random processes. Holden-Day, San Francisco.

Salehi, H., 1979. Algorithms for linear interpolator and interpolation error for minimal stationary stochastic processes. Ann. Probab. 7, 840–846.

Simon, B., 2005. Orthogonal polynomials on the unit circle. Part 1. Classical theory. AMS, Providence.

Urbanik, K., 2000. A duality principle for stationary random sequences. Colloq. Math. 86, 153–162.

Wiener, N., 1949. Extrapolation, Interpolation, and Smoothing of Stationary Time Series. With Engineering Applications. The Technology Press of the Massachusetts Institute of Technology, Cambridge, Mass.

Wiener, N., Masani, P., 1958. The prediction theory of multivariate stochastic processes. II. The linear predictor. Acta Math. 99, 93–137.

Wold, H., 1954. A study in the analysis of stationary time series. Almqvist and Wiksell, Stockholm, 2nd ed. [1938, Uppsala]

Yaglom, A.M., 1963. On a problem of linear interpolation of stationary random sequences and processes. AMS. Select. Transl. Math. Statist. Probab. 4, 339–344. [1949, Uspehi Matem. Nauk 4, 173–178]