RECURSIVE RELATIONS FOR MULTISTEP PREDICTION OF A STATIONARY TIME SERIES

By Pascal Bondon

Centre National de la Recherche Scientifique

First version received January 1999

Abstract. Recursive relations are established between the coefficients of the finite past multistep linear predictors of a stationary time series. These relations generalize known results when the prediction is based on infinite past and permit simplification of the numerical calculation of the finite past predictors.

Keywords. Stationary time series; linear prediction; multistep prediction; Levinson's algorithm.

1. INTRODUCTION

Multistep prediction problems appear in many time series applications including forecasting in economics and geophysics, electric power load scheduling, and predictive control of processes. Most often, the multistep prediction is accomplished with the 'plug-in' method in which the multistep forecasts are calculated from an initial model fitted to the available data by repeatedly using the model where the unknown future values are replaced by their own forecasts. This method uses the initial model as if it were the 'true' model generating the time series. As, in practice, all fitted models may be incorrect, a direct method for multistep prediction which consists in fitting a different model for each step may be more appropriate; see Findley (1984), Bhansali (1996) and references therein.

This paper concentrates on linear models. The algorithm of Levinson (1946) and Durbin (1960) allows us to calculate the parameters of autoregressive models of increasing orders that minimize any h-step ahead prediction error variance, $h \geqslant 1$. Therefore, it is possible to use the Levinson algorithm to fit the models, and, if the order of each autoregression is fixed at a given value p, the numerical complexity for fitting s models is proportional to s.

When the indices of the data to be predicted are consecutive integers, it is natural to ask whether the coefficients of the multistep predictors can be calculated more efficiently. The idea is to use some relations between the prediction coefficients which are not only order recursive like the Levinson recursions, but are also step recursive. Step recursive relations exist when the prediction is based on infinite past – see Box *et al.* (1994, Section 5.3) – but do not seem to be available when the past is finite.

The purpose of this paper is to establish such relations that will appear as being the finite past counterpart of the expressions in Box *et al*. (1994). Then, an algorithm will be proposed for computing the multistep predictors when the order of each predictor is fixed at a given value p. This algorithm will be shown to be numerically more efficient than the Levinson algorithm and than the innovations algorithm in Brockwell and Davis (1991, Section 5.2). A possible application of this algorithm is speech recognition based on multistep predictors. In speech analysis, the order p is a given integer in the range [8, 16] (Markel and Gray, 1976).

2. PRELIMINARIES

Let $(X_n)_{n\in\mathbb{Z}}$ be a zero-mean, weakly stationary, univariate real-valued time series with covariance function (γ_n). (X_n) is defined on a probability space $(\Omega, \mathscr{F}, \mathrm{P})$, and $L^2(\Omega, \mathscr{F}, \mathrm{P})$ denotes the Hilbert space with inner product $\langle X, Y\rangle = \mathrm{E}XY$ and norm $\|X\| = \sqrt{\mathrm{E}X^2}$. For any integers l, n satisfying $l \leqslant n$, we define the subspaces of $L^2(\Omega, \mathscr{F}, \mathrm{P})$,

$$H_{l,n} = \mathrm{sp}\{X_k;\ l \leqslant k \leqslant n\}$$

$$H_n = \overline{\mathrm{sp}}\{X_k;\ k \leqslant n\}$$

where $\overline{\mathrm{sp}}S$ is the closed linear span of elements of the set S in the norm of $L^2(\Omega, \mathscr{F}, \mathrm{P})$. We denote respectively by $P_{l,n}$ and P_n the orthogonal projection operator onto $H_{l,n}$ and H_n.

We assume that (X_n) is purely nondeterministic, and hence has the mean-square convergent moving average series representation

$$X_n = \sum_{i=0}^{\infty} c_i \epsilon_{n-i}$$

where $\epsilon_n = X_n - P_{n-1}X_n$, $c_0 = 1$, and $\sum_{i=0}^{\infty} c_i^2 < \infty$. Furthermore, we suppose that (X_n) has the mean-square convergent autoregressive series representation

$$X_n = \epsilon_n + \sum_{i=1}^{\infty} a_i X_{n-i}. \tag{2.1}$$

Sufficient conditions for the mean-square convergence of the series in (2.1) may be found in Bloomfield (1985). It is well known that the parameters (c_i) and (a_i) satisfy the relations

$$a_0 = -1$$

$$a_i = -\sum_{j=0}^{i-1} a_j c_{i-j} \qquad i \geqslant 1.$$

Since (X_n) has an autoregressive representation, it results from Bloomfield (1985, Thm 1) that, for any $h \geqslant 1$, the infinite past h-step predictor $P_n X_{n+h}$ has a mean-square convergent series representation

$$P_n X_{n+h} = \sum_{i=1}^{\infty} a_i^h X_{n+1-i}$$

where

$$a_i^h = \sum_{j=0}^{h-1} c_j a_{i+h-1-j} = -\sum_{j=h}^{i+h-1} c_j a_{i+h-1-j} \qquad i \geqslant 1 \tag{2.2}$$

see also Wiener and Masani (1958, Thm 5.7). It should be noted that $a_i^1 = a_i$ for all $i \geqslant 1$. The parameters (a_i^h), $h > 1$, can be calculated recursively using either Equation (5.3.9) or Equation (A5.2.4) in Box *et al.* (1994, Section 5.3). These relations are respectively

$$a_i^h = a_{i+h-1} + \sum_{j=1}^{h-1} a_j a_i^{h-j} \tag{2.3}$$

$$a_i^h = a_{i+1}^{h-1} + c_{h-1} a_i. \tag{2.4}$$

When the prediction is based on a finite number of observations, say $(X_k)_{1 \leqslant k \leqslant n}$, two recursive methods are presented in Brockwell and Davis (1991, Section 5.2) for computing the one-step predictor $P_{1,n} X_{n+1}$. These are the Levinson algorithm, which gives the coefficients of $(X_k)_{1 \leqslant k \leqslant n}$ in the representation

$$P_{1,n} X_{n+1} = \sum_{i=1}^{n} a_{n,i}^1 X_{n+1-i}$$

and the innovations algorithm, which gives the coefficients of $(X_k - P_{1,k-1} X_k)_{1 \leqslant k \leqslant n}$ in the orthogonal expansion

$$P_{1,n} X_{n+1} = \sum_{i=1}^{n} c_{n,i} (X_{n+1-i} - P_{1,n-i} X_{n+1-i}).$$

Moreover, it is shown how to use the innovations algorithm to compute the h-step predictor for any $h \geqslant 1$, see Remark 4.2. In the following, we use the representation

$$P_{1,n} X_{n+h} = \sum_{i=1}^{n} a_{n,i}^h X_{n+1-i}$$

and we denote by v_n^h the prediction error variance

$$v_n^h = \| X_{n+h} - P_{1,n} X_{n+h} \|^2.$$

3. RECURSIVE RELATIONS BETWEEN THE PREDICTION COEFFICIENTS

Propositions 3.1 and 3.2 derive, respectively, the finite past counterpart of relations (2.3) and (2.4).

PROPOSITION 3.1. For any step $h \geqslant 1$ and any order $n \geqslant 1$,

$$a_{n,i}^{h} = a_{n+h-1,i+h-1}^{1} + \sum_{j=1}^{h-1} a_{n+h-1,j}^{1} a_{n,i}^{h-j} \qquad i = 1, \ldots, n. \tag{3.1}$$

PROOF. Since $h \geqslant 1$

$$H_{1,n} \subset H_{1,n+h-1}$$

and then

$$P_{1,n}X_{n+h} = P_{1,n} \circ P_{1,n+h-1}X_{n+h}.$$

Therefore,

$$\begin{aligned} P_{1,n}X_{n+h} &= P_{1,n} \sum_{j=1}^{n+h-1} a_{n+h-1,j}^{1} X_{n+h-j} \\ &= \sum_{j=1}^{h-1} a_{n+h-1,j}^{1} P_{1,n}X_{n+h-j} + \sum_{j=h}^{n+h-1} a_{n+h-1,j}^{1} X_{n+h-j} \\ &= \sum_{j=1}^{h-1} a_{n+h-1,j}^{1} \sum_{i=1}^{n} a_{n,i}^{h-j} X_{n+1-i} + \sum_{i=1}^{n} a_{n+h-1,i+h-1}^{1} X_{n+1-i} \end{aligned}$$

which is equivalent to (3.1) because (X_n) is nondeterministic. ∎

PROPOSITION 3.2. For any step $h > 1$ and any order $n \geqslant 1$,

$$a_{n,i}^{h} = a_{n+1,i+1}^{h-1} + a_{n+1,1}^{h-1} a_{n,i}^{1} \qquad i = 1, \ldots, n \tag{3.2}$$

$$v_n^h = v_{n+1}^{h-1} + (a_{n+1,1}^{h-1})^2 v_n^1. \tag{3.3}$$

PROOF. Since

$$H_{1,n+1} = H_{1,n} \oplus \mathrm{sp}\{X_{n+1} - P_{1,n}X_{n+1}\}$$

we have

$$P_{1,n+1}X_{n+h} = P_{1,n}X_{n+h} + \lambda(X_{n+1} - P_{1,n}X_{n+1}) \tag{3.4}$$

where $\lambda \in \mathbb{R}$. Equation (3.4) is equivalent to

$$\sum_{i=1}^{n+1} a_{n+1,i}^{h-1} X_{n+2-i} = \sum_{i=1}^{n} a_{n,i}^{h} X_{n+1-i} + \lambda X_{n+1} - \lambda \sum_{i=1}^{n} a_{n,i}^{1} X_{n+1-i}$$

from which we deduce that

$$a_{n+1,1}^{h-1} = \lambda$$

and

$$a_{n+1,i+1}^{h-1} = a_{n,i}^{h} - \lambda a_{n,i}^{1} \qquad \text{for } 1 \leqslant i \leqslant n.$$

This proves (3.2). It results from (3.4) that

$$\|P_{1,n+1}X_{n+h}\|^2 = \|P_{1,n}X_{n+h}\|^2 + \lambda^2 v_n^1.$$

Furthermore,

$$v_n^h = \|X_{n+h}\|^2 - \|P_{1,n}X_{n+h}\|^2$$

and

$$v_{n+1}^{h-1} = \|X_{n+h}\|^2 - \|P_{1,n+1}X_{n+h}\|^2.$$

Hence,

$$v_n^h = v_{n+1}^{h-1} + \lambda^2 v_n^1$$

which gives (3.3). ■

To see the symmetry between (2.4) and (3.2), it should be observed that $c_h = a_1^h$ for any $h \geqslant 1$. This results from the second equality in (2.2) and $a_0 = -1$. On the other hand, the infinite past relation which corresponds to (3.3) is

$$v^h = v^{h-1} + c_{h-1}^2 \sigma_\epsilon^2$$

where

$$v^h = \|X_{n+h} - P_n X_{n+h}\|^2$$

and

$$\sigma_\epsilon^2 = \|\epsilon_n\|^2 = v^1.$$

This relation is the recursive form of

$$v^h = \sigma_\epsilon^2 \sum_{i=0}^{h-1} c_i^2.$$

REMARK 3.1. When (X_n) is a causal autoregressive process of order p, $a_i = 0$ for $i > p$ in (2.1) and the relations (2.3) and (2.4) can be used to calculate the finite past predictor $P_{1,n}X_{n+h}$ provided that the order n satisfies $n \geqslant p$. Indeed, according to (2.4), $a_i^h = 0$ for $h \geqslant 1$ and $i > p$, and then $P_n X_{n+h} \in H_{n-p+1,n}$. Assume that $n \geqslant p$. Then

$$H_{n-p+1,n} \subset H_{1,n} \subset H_n$$

and we have

$$P_{1,n}X_{n+h} = P_{n-p+1,n}X_{n+h} = P_nX_{n+h}.$$

Thus

$$a^h_{n,i} = a^h_{p,i} = a^h_i \qquad \text{for } 1 \leqslant i \leqslant p$$

$$a^h_{n,i} = a^h_i = 0 \qquad \text{for } i > p$$

and

$$v^h_n = v^h_p = v^h.$$

4. CALCULATION OF THE h-STEP PREDICTORS

We now show how the relations (3.1), (3.2) and (3.3) can be used for computing efficiently the h-step prediction coefficients $(a^h_{p,i})$ and the prediction error variances v^h_p for $1 \leqslant h \leqslant s$, where p is a given order and s is a given final step.

For $h = 1$, we can use the Levinson algorithm which calculates the coefficients $(a^1_{n,i})$ and the variances v^1_n for $1 \leqslant n \leqslant p$. For $h > 1$, $(a^h_{p,i})$ and v^h_p can also be computed with a Levinson type recursion by using also $(a^1_{n,i})$ and v^1_n for $1 \leqslant n \leqslant p-1$. The algorithm is given in the following proposition.

PROPOSITION 4.1. For any step $h \geqslant 1$ and any order $n \geqslant 1$, we have

$$a^h_{n,n} = \left[\gamma_{n+h-1} - \sum_{i=1}^{n-1} a^1_{n-1,i}\gamma_{n+h-i-1}\right](v^1_{n-1})^{-1} \tag{4.1}$$

$$a^h_{n,i} = a^h_{n-1,i} - a^h_{n,n}a^1_{n-1,n-i} \qquad i = 1, \ldots, n-1 \tag{4.2}$$

$$v^h_n = v^h_{n-1} - (a^h_{n,n})^2 v^1_{n-1} \tag{4.3}$$

with $v^h_0 = \gamma_0$.

PROOF. From the orthogonal decomposition

$$H_{1,n} = H_{2,n} \oplus \text{sp}\{X_1 - P_{2,n}X_1\},$$

we deduce that

$$P_{1,n}X_{n+h} = P_{2,n}X_{n+h} + \theta(X_1 - P_{2,n}X_1) \tag{4.4}$$

where

$$\theta = \frac{\langle X_{n+h}, X_1 - P_{2,n}X_1\rangle}{\|X_1 - P_{2,n}X_1\|^2}. \tag{4.5}$$

We note that $\|X_1 - P_{2,n}X_1\| \neq 0$ since (X_n) is nondeterministic. Now, we deduce from the stationarity of (X_n) that

$$P_{2,n}X_{n+h} = \sum_{i=1}^{n-1} a_{n-1,i}^{h} X_{n+1-i}$$

$$P_{2,n}X_{1} = \sum_{i=1}^{n-1} a_{n-1,i}^{1} X_{i+1}$$

$$\|X_1 - P_{2,n}X_1\|^2 = v_{n-1}^{1}.$$

Therefore, (4.4) can be written

$$\sum_{i=1}^{n} a_{n,i}^{h} X_{n+1-i} = \sum_{i=1}^{n-1} a_{n-1,i}^{h} X_{n+1-i} + \theta X_1 - \theta \sum_{i=1}^{n-1} a_{n-1,i}^{1} X_{i+1}$$

which is equivalent to $a_{n,n}^{h} = \theta$ and

$$a_{n,i}^{h} = a_{n-1,i}^{h} - \theta a_{n-1,n-i}^{1} \qquad \text{for } 1 \leqslant i \leqslant n-1.$$

Hence (4.2) holds. Now, we deduce from (4.5) that

$$a_{n,n}^{h} = \left\langle X_{n+h}, X_1 - \sum_{i=1}^{n-1} a_{n-1,i}^{1} X_{i+1} \right\rangle (v_{n-1}^{1})^{-1}$$

which is equivalent to (4.1). It results from (4.4) that

$$\|P_{1,n}X_{n+h}\|^2 = \|P_{2,n}X_{n+h}\|^2 + \theta^2 v_{n-1}^{1}.$$

On the other hand, we have

$$v_{n}^{h} = \|X_{n+h}\|^2 - \|P_{1,n}X_{n+h}\|^2$$

and

$$v_{n-1}^{h} = \|X_{n+h}\|^2 - \|P_{2,n}X_{n+h}\|^2.$$

Therefore,

$$v_{n}^{h} = v_{n-1}^{h} - \theta^2 v_{n-1}^{1}$$

from which (4.3) follows. ■

For each n and h, the calculation of $(a_{n,i}^{h})$ and v_{n}^{h} in (4.1)–(4.3) involves $2n+1$ products. Therefore, when $(a_{p,i}^{h})$ and v_{p}^{h} are computed for $1 \leqslant h \leqslant s$ using Proposition 4.1 (algorithm A_1), the numerical complexity is

$$N_1 = s \sum_{n=1}^{p} (2n+1) = p^2 s + 2ps.$$

An alternative to algorithm A_1 consists in computing $(a_{n,i}^{1})$ for $1 \leqslant n \leqslant p+s-1$ using (4.1)–(4.3), and then to compute for $2 \leqslant h \leqslant s$, $(a_{p,i}^{h})$ using (3.1) where $n = p$, and v_{p}^{h} with the relation

$$v_p^h = \langle X_{p+h} - P_{1,p}X_{p+h}, X_{p+h}\rangle = \gamma_0 - \sum_{i=1}^{p} a_{p,i}^h \gamma_{h+i-1}. \tag{4.6}$$

This is algorithm A_2, the numerical complexity of which is

$$N_2 = (p+s-1)^2 + 2(p+s-1) - 2 + p\sum_{h=2}^{s}(h-1) + p(s-1)$$

$$= p^2 + \frac{1}{2}p(s^2+5s-2) + s^2 - 3.$$

To compare N_1 and N_2, we set $\tau = s/p$. Then

$$N_2 - N_1 = p^3\tau(\tfrac{1}{2}\tau - 1) + p^2(\tau^2 + \tfrac{1}{2}\tau + 1) - p - 3.$$

If $\tau \geqslant 2$,

$$N_2 - N_1 \geqslant 6p^2 - p - 3 = 2p^2 + (p-1)(4p+3) > 0 \qquad \text{for } p \geqslant 1.$$

Now, if $\tau = \frac{1}{2}$,

$$N_2 - N_1 = -\tfrac{3}{8}(p-2)(p^2 - 2p - \tfrac{4}{3}) - 2 < 0 \qquad \text{for } p \geqslant 1.$$

Consequently, the choice of the method A_1 or A_2 depends on the pair (p, s).

We now propose an algorithm whose numerical complexity is less than N_1 and N_2 for any pair (p, s) with $p, s > 1$. This algorithm uses the order and step recursive relations given in the following proposition.

PROPOSITION 4.2. For any step $h > 1$ and any order $n > 1$, we have

$$a_{n,i}^h = a_{n,i+1}^{h-1} + a_{n,1}^{h-1}a_{n-1,i}^1 - a_{n,n}^h a_{n-1,n-i}^1 \qquad i = 1, \ldots, n-1 \tag{4.7}$$

$$v_n^h = v_n^{h-1} + [(a_{n,1}^{h-1})^2 - (a_{n,n}^h)^2]v_{n-1}^1. \tag{4.8}$$

PROOF. The relations (4.7) and (4.8) can be deduced from Propositions 3.2 and 4.1 by replacing n by $n-1$ in (3.2) and (3.3) and by inserting the corresponding expressions for $a_{n-1,i}^h$ and v_{n-1}^h in (4.2) and (4.3). Alternatively, (4.7) and (4.8) can be deduced directly from the relation

$$P_{1,n}X_{n+h} = P_{2,n}X_{n+h} + a_{n,n}^h(X_1 - P_{2,n}X_1)$$

$$= P_{2,n+1}X_{n+h} - a_{n,1}^{h-1}(X_{n+1} - P_{2,n}X_{n+1}) + a_{n,n}^h(X_1 - P_{2,n}X_1)$$

■

Algorithm A_3 consists in computing $(a_{n,i}^1)$ and v_n^1 for $1 \leqslant n \leqslant p$ using (4.1)–(4.3), and then to compute $(a_{p,i}^h)$ and v_p^h for $2 \leqslant h \leqslant s$ using (4.1), (4.7) and (4.8) where $n = p$. The numerical complexity of these calculations is

$$N_3 = p^2 + 2p + (s-1)(3p+1) = p^2 + p(3s-1) + s - 1.$$

We have

$$N_1 - N_3 = (s-1)(p^2 - p - 1) > 0 \qquad \text{if } p > 1$$

and

$$N_2 - N_3 = \tfrac{1}{2}ps(s-1) + (s-2)(s+1) > 0 \qquad \text{if } s > 1.$$

Therefore, $N_3 < \min(N_1, N_2)$ for any pair (p, s) with $p, s > 1$.

REMARK 4.1. An alternative to algorithm A_3 consists in computing $(a^1_{n,i})$ and v^1_n for $1 \leqslant n \leqslant p+s-1$ using (4.1)–(4.3), and $(a^h_{n,i})$ and v^h_n for $2 \leqslant h \leqslant s$ and $p \leqslant n \leqslant p+s-h$ using (3.2) and (3.3). The numerical complexity of this method (algorithm A_4) is

$$N_4 = (p+s-1)^2 + 2(p+s-1) + \sum_{h=2}^{s} \sum_{n=p}^{p+s-h} (n+2)$$

$$= p^2 + \tfrac{1}{2}ps(s+3) + \tfrac{1}{6}(s-1)(s^2 + 10s + 6).$$

We have

$$N_4 - N_3 = (s-1)[\tfrac{1}{2}p(s-2) + \tfrac{1}{6}s(s+10)] > 0 \qquad \text{if } s > 1$$

and therefore algorithm A_3 is preferable to algorithm A_4. On the other hand, algorithm A_4 calculates the coefficients $(a^h_{q-h+1,i})$ and the variances v^h_{q-h+1} for $1 \leqslant h \leqslant s$ where q is a given integer, $q \geqslant s$. The parameters $(a^h_{q-h+1,i})$ are the prediction coefficients of $P_{h,q}X_{q+h}$ and they involve only the covariances γ_k for $0 \leqslant k \leqslant q$. Furthermore, to calculate $(a^h_{q-h+1,i})$ and v^h_{q-h+1} for $2 \leqslant h \leqslant s$, it is sufficient to use (3.2) and (3.3) for $n = q - h + 1$.

REMARK 4.2. A different approach for computing the multistep predictors for $1 \leqslant h \leqslant s$ is to calculate $(a^1_{n,i})$ and v^1_n for $1 \leqslant n \leqslant p$ using (4.1)–(4.3), and then to decompose the observation space $H_{1,p}$ as the sum of p orthogonal subspaces for computing the h-step predictors for $2 \leqslant h \leqslant s$. For any step $h > 1$, $P_{1,p}X_{p+h}$ and v^h_p are computed by the relations

$$P_{1,p}X_{p+h} = \sum_{i=1}^{p} c^h_i (X_i - P_{1,i-1}X_i)$$

$$v^h_p = \gamma_0 - \sum_{i=1}^{p} (c^h_i)^2 v^1_{i-1}$$

where $P_{1,0}X_1 = 0$ and

$$c^h_i = \frac{\langle X_{p+h}, X_i - P_{1,i-1}X_i \rangle}{\|X_i - P_{1,i-1}X_i\|^2} = \left[\gamma_{p+h-i} - \sum_{j=1}^{i-1} a^1_{i-1,j}\gamma_{p+h-i+j} \right] (v^1_{i-1})^{-1}.$$

The number of products in these computations is $p^2 + 2p$ for the Levinson algorithm, $i - 1$ for $P_{1,i-1}X_i$, i for c_i^h, and $2p$ for v_p^h. The resulting numerical complexity is

$$N_5 = p^2 + 2p + \tfrac{1}{2}p(p-1) + (s-1)[\tfrac{1}{2}p(p+1) + 2p] = p^2(\tfrac{1}{2}s + 1) + p(\tfrac{5}{2}s - 1).$$

We have

$$N_5 - N_3 = \tfrac{1}{2}s(p-2)(p+1) + 1 > 0 \qquad \text{if } p > 1$$

and thus algorithm A_3 is less complex than this algorithm. On the other hand, we have

$$N_1 - N_5 = p(p-1)(\tfrac{1}{2}s - 1) > 0 \qquad \text{if } p > 1 \text{ and } s > 2.$$

REMARK 4.3. The Levinson algorithm gives the Cholesky decomposition of the inverse of the covariance matrix $[\Gamma_p(i, j) = \gamma_{i-j},\ i, j = 1, \ldots, p]$,

$$\Gamma_p^{-1} = A_p' \Sigma_p^{-2} A_p$$

where A_p is lower triangular with $A_p(i, i) = 1$ and $A_p(i, j) = -a_{i-1,i-j}^1$ for $i > j$, and Σ_p^2 is diagonal with $\Sigma_p^2(i, i) = v_{i-1}^1$. For any step $h > 1$, the coefficients $(a_{p,i}^h)$ can be computed directly by the relation

$$(a_{p,1}^h, \ldots, a_{p,p}^h)' = \Gamma_p^{-1}(\gamma_h, \ldots, \gamma_{h+p-1})' = A_p' \Sigma_p^{-2} A_p (\gamma_h, \ldots, \gamma_{h+p-1})'$$

and the variance v_p^h can be calculated with (4.6). Then the global numerical complexity for computing $(a_{p,i}^h)$ and v_p^h for $1 \leqslant h \leqslant s$ is

$$N_6 = p^2 + 2p + (s-1)(p^2 + p) = p^2 s + p(s+1).$$

We have

$$N_6 - N_3 = (s-1)(p^2 - 2p - 1) > 0 \qquad \text{if } s > 1 \text{ and } p > 2.$$

Hence, except when $p = 1$ or $p = 2$, using algorithm A_3 is preferable.

REMARK 4.4. It results from (4.3) that $v_n^h \leqslant v_{n-1}^h$. This inequality is also an immediate consequence of the stationarity of (X_n) and of the inclusion $H_{2,n} \subset H_{1,n}$. Furthermore, taking $h = 1$ in (4.3), we deduce that $|a_{n,n}^1| < 1$. The same inequality does not hold for $a_{n,n}^h$ when $h > 1$. Indeed, consider the causal autoregressive process

$$X_n = a_1 X_{n-1} + a_2 X_{n-2} + \epsilon_n \tag{4.9}$$

where $\epsilon_n = X_n - P_{n-1}X_n$, $\|\epsilon_n\|^2 = \sigma_\epsilon^2$. We have

$$P_{1,2}X_4 = (a_2 + a_1^2)X_2 + a_1 a_2 X_1.$$

Taking $a_2 = -\frac{1}{4}a_1^2$, (4.9) is a causal model if $|a_1| < 2$, and we get $a_{2,2}^2 = -\frac{1}{4}a_1^3$.

Hence, if $4 < |a_1^3| < 8$, $|a_{2,2}^2| > 1$, and if $|a_1^3| \leqslant 4$, $|a_{2,2}^2| \leqslant 1$. When the past is infinite, the inclusion $H_n \subset H_{n+1}$ implies that the h-step prediction error variance is a nondecreasing function of h. When the past is finite, no inequality exists between v_n^{h-1} and v_n^h. Indeed, when $a_2 = 0$ in (4.9), we have

$$P_{1,1}X_2 = a_1 X_1$$

$$v_1^1 = \sigma_\epsilon^2$$

$$P_{1,1}X_3 = a_1^2 X_1$$

$$v_1^2 = \sigma_\epsilon^2(1 + a_1^2)$$

and therefore $v_1^1 < v_1^2$. But, when $a_1 = 0$ in (4.9), we have

$$P_{1,1}X_2 = 0$$

$$v_1^1 = \sigma_\epsilon^2(1 - a_2^2)^{-1}$$

$$P_{1,1}X_3 = a_2 X_1$$

$$v_1^2 = \sigma_\epsilon^2$$

and thus $v_1^1 > v_1^2$. Lastly, we deduce from (3.3) that $v_{n+1}^{h-1} \leqslant v_n^h$. This inequality also results from $H_{1,n} \subset H_{1,n+1}$.

REMARK 4.5. The results of this paper can be easily generalized to the case where (X_n) is a m-variate time series as follows. The coefficients $(a_{n,i}^h)$ are now $m \times m$ matrices, and we define

$$\gamma_n = \mathrm{E}X_n X_0'$$

and

$$v_n^h = \mathrm{E}(X_{n+h} - P_{1,n}X_{n+h})(X_{n+h} - P_{1,n}X_{n+h})'.$$

Propositions 3.1 and 3.2 hold if (3.3) is replaced by

$$v_n^h = v_{n+1}^{h-1} + a_{n+1,1}^{h-1},\ v_n^1(a_{n+1,1}^{h-1})'.$$

In contrast to the univariate algorithm, the multivariate Levinson algorithm requires the solution of two sets of linear equations, one arising in the calculation of the forward predictor $P_{1,n}X_{n+1}$, and the other in the calculation of the backward predictor $P_{1,n}X_0$ (Whittle, 1963). Let $(\tilde{a}_{n,i}^1)$ be the matrices such that

$$P_{1,n}X_0 = \sum_{i=1}^{n} \tilde{a}_{n,i}^1 X_i$$

and $\tilde{v}_n^1$ be the covariance matrix of the one-step backward prediction error,

$$\tilde{v}_n^1 = \mathrm{E}(X_0 - P_{1,n}X_0)(X_0 - P_{1,n}X_0)'.$$

It can be shown that the relations in Proposition 4.1 become

$$a_{n,n}^h = \left[\gamma_{n+h-1} - \sum_{i=1}^{n-1} \gamma_{n+h-i-1}(\tilde{a}_{n-1,i}^1)'\right](\tilde{v}_{n-1}^1)^{-1}$$

$$a_{n,i}^h = a_{n-1,i}^h - a_{n,n}^h \tilde{a}_{n-1,n-i}^1 \qquad i = 1, \ldots, n-1$$

$$v_n^h = v_{n-1}^h - a_{n,n}^h \tilde{v}_{n-1}^1 (a_{n,n}^h)'.$$

On the other hand, the relations in Proposition 4.2 are now

$$a_{n,i}^h = a_{n,i+1}^{h-1} + a_{n,1}^{h-1} a_{n-1,i}^1 - a_{n,n}^h \tilde{a}_{n-1,n-i}^1 \qquad i = 1, \ldots, n-1$$

$$v_n^h = v_n^{h-1} + a_{n,1}^{h-1} v_{n-1}^1 (a_{n,1}^{h-1})' - a_{n,n}^h \tilde{v}_{n-1}^1 (a_{n,n}^h)'.$$

ACKNOWLEDGEMENT

The author is indebted to the referee for Remarks 4.2 and 4.3.

REFERENCES

BHANSALI, R. J. (1996) Asymptotically efficient autoregressive model selection for multistep prediction. *Ann. Inst. Statist. Math.* 48, 577–602.

BLOOMFIELD, P. (1985) On series representations for linear predictors. *Ann. Probab.* 13, 226–33.

BOX, G. E. P., JENKINS, G. M. and REINSEL, G. C. (1994) *Time Series Analysis, Forecasting and Control* (3rd edn). Englewood Cliffs: Prentice-Hall.

BROCKWELL, P. J. and DAVIS, R. A. (1991) *Time Series: Theory and Methods* (2nd edn). New York: Springer Verlag.

DURBIN, J. (1960) The fitting of time series models. *Rev. Int. Stat.* 28, 233–44.

FINDLEY, D. F. (1984) On some ambiguities associated with the fitting of ARMA models to time series. *J. Time Ser. Anal.* 5, 213–25.

LEVINSON, N. (1946) The Wiener RMS (root mean square) error criterion in filter design and prediction. *J. Math. Phys.* 25, 261–78.

MARKEL, J. D. and GRAY, A. H., Jr. (1976) *Linear Prediction of Speech*. Berlin: Springer Verlag.

WHITTLE, P. (1963) On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*. 40, 129–34.

WIENER, N. and MASANI, P. (1958) The prediction theory of multivariate stochastic processes, II. *Acta Math.* 99, 93–137.