



HAL
open science

ARMA Lattice Identification: A New Hereditary Algorithm

André Monin, Gérard Salut

► **To cite this version:**

André Monin, Gérard Salut. ARMA Lattice Identification: A New Hereditary Algorithm. IEEE Transactions on Signal Processing, 1996, 4 (2), pp.360-370. 10.1109/78.485931 . hal-02916434

HAL Id: hal-02916434

<https://laas.hal.science/hal-02916434>

Submitted on 17 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ARMA Lattice Identification: A New Hereditary Algorithm.

André MONIN, Gérard SALUT.

Laboratoire d'Analyse et d'Architecture des Systèmes.

Centre National de la Recherche Scientifique.

7 Avenue du Colonel Roche 31077 TOULOUSE Cedex.

Tel: (33).61.33.62.95 / Fax: (33). 61.55.35.77 / email: monin@laas.laas.fr

EDICS: SP 3.9.1.

Abstract: This paper derives an identification solution of the optimal linear predictor of ARMA type, as a time varying lattice of arbitrarily fixed dimension, for a process whose output signal only is known. The projection technique introduced here leads to an hereditary algorithm which is the adaptive extension to raw data of previous results of the authors on lattice realization from given autocorrelation functions ([1]). It produces a minimum phase linear model of the signal whose n-th order “whiteness” of the associated innovation has the following restricted meaning: orthogonality to an n-dimensional subspace memory of the past in a suitable Hilbert sequence space. The L^2 metric of that sequence space leads to a least-squares Identification algorithm which possesses a “certainty equivalence principle” with respect to the corresponding Realization algorithm (i.e., sample correlation products replace true correlation terms). Due to the detailed state-space time-varying computations, this is possible here while avoiding the well known “side errors” from missing correlation products which usually occur in a blunt replacement of the output autocorrelation by averaged sample products. Application examples show the superiority of the hereditary algorithm over classical recursive and non-recursive algorithms, in terms of accuracy, adaptativity and order reduction capabilities.

1. Introduction.

The ARMA stochastic identification problem has been extensively investigated for many years and numerous algorithms have been proposed. Because of its numerical robustness, linear identification under lattice form is of special interest. Moreover, it is often desirable to obtain recursive algorithms on account of their ability to cope with adaptive filtering situations. Recall that the aim of ARMA identification is to fit a linear stochastic realization to an experimental set of measurements of the output $\{y_\tau, \tau = 1 \dots t\}$ (possibly of the input/output $\{u_\tau, y_\tau, \tau = 1 \dots t\}$) of some unknown system. For synthesis results on this subject, the reader is referred to our closely related paper [1], on L_p^2 optimal rational approximations, which is the “realization” counterpart of the present “identification” algorithm.

The common approach consists in deriving a one-step-ahead predictor \hat{y}_τ of the output and then evaluating the prediction performance through the experimental mean square error, i. e., in the stationary case, $J_t = (1/t) \sum_{\tau=1}^t (y_\tau - \hat{y}_\tau)^2$, in order to optimize the predictor parameters. This kind of formulation is useful to avoid the gaussian hypothesis under which the searched solution coincides with the maximum likelihood predictor.

Many different techniques have been proposed to achieve this goal and principally differ according to the class of predictors chosen as well as approximations made. One may first dis-

tinguish them according to the existence or not of feedback in the dynamic of the filter and then bring out two main classes of systems:

1.1. Transversal filters or Finite Impulse Response (FIR) filters.

The one-step-ahead predictor is represented as a moving average (MA) over the past output and takes the following canonical form:

$$\hat{y}_t = \sum_{i=1}^n a_i y_{t-i} \quad (1.1)$$

It is the most popular class of predictors because the computation of the optimal parameters $\{a_i\}$ can be straightforwardly achieved as a linear quadratic optimum in a one-step recursive form.

Noting $\tilde{y}_t = y_t - \hat{y}_t$ and substituting y_t to $\hat{y}_t + \tilde{y}_t$, it leads to an autoregressive (AR) model of the output. Many algorithms have been developed according to this formulation, particularly under various lattice forms ([2]-[10] for instance). Unfortunately, it is well known that if the observation signals y_t is corrupted by noise v_t , which is most often the case in practice, such models are biased [11].

1.2. Rational filters or Infinite Impulse Response (IIR) filters.

The one-step-ahead predictor is represented as an ARMA over the past output and takes the following canonical forms:

$$\hat{v}_t = \sum_{i=1}^n a_i \hat{y}_{t-i} + \sum_{j=1}^q b_j \tilde{y}_{t-j}. \quad (1.2)$$

This type of model is well known to be the most efficient linear form, with respect to the number of parameters, as it may represent IIR systems with only a finite set of parameters. Substituting y_t to $\tilde{y}_t + \hat{y}_t$ leads to an ARMA model:

$$v_t = \sum_{i=1}^n a_i y_{t-i} - \sum_{i=1}^n a_i \tilde{y}_{t-i} + \sum_{j=1}^q b_j \tilde{y}_{t-j} + \tilde{y}_t \quad (1.3)$$

whose general predictor is again of the type (1.2). This is why the general ARMA model is closed under additive measurement noises.

Computation of the optimal parameters is however a more difficult task since the criterium is no longer quadratic with respect to the parameters. Moreover, the non-linearity involved is not explicitly tractable.

1.2.1 Finitely Recursive algorithms:

Many algorithms have been developed using essentially two kinds of approximations:

- The Extended Least Square (ELS) algorithm ignores the dependence of \hat{y}_{t-i} and \tilde{y}_{t-i} with respect to the parameters $\{a_i, b_j\}$ when performing the minimization of the criterium J_t [12] [13]. The minimum is then achieved via classical linear regression algorithms or gradient type algorithms (Recursive Extended Least Square algorithm [14] for

instance).

- The innovation approach or prediction error methods leads to a prefiltered state vector $\{y_{t-1}, \dots, y_{t-n}, \tilde{y}_{t-1}, \dots, \tilde{y}_{t-n}\}$ which takes into account the implicit dependence neglected above. The optimal parameters are generally computed using local search methods such as stochastic gradient or Gauss-Newton algorithms[11] [12] [14] [15]. (Recursive Maximum Likelihood algorithm [14] for instance).

1.2.2 Hereditary algorithms:

In none of these approaches transient behavior is optimized. We are concerned in this paper with a time varying approach of IIR identification preserving a transient convergence similar to that of [1] for known auto correlations, which has proved to be useful in terms of convergence rate as well as final accuracy. The price to be paid by such algorithms is their hereditary nature (growing memory) in terms of time t , due to updating of the overall least-square criterium as data accumulate. We shall deal in the sequel with:

- Global hereditary criterium (stationary case):

$$E^t[(y. - \hat{y}.)^2] = \frac{1}{t} \sum_{\tau=1}^t (y_{\tau} - \hat{y}_{\tau})^2 \quad (1.4)$$

- Locally limited memory criterium (adaptive case):

$$E_T^t[(y. - \hat{y}.)^2] = \frac{1}{T} \sum_{\tau=t-T+1}^t (y_{\tau} - \hat{y}_{\tau})^2 \quad (1.5)$$

We develop the stationary case and will just underline the difference in the second case. We deal here with ARMA problems where the AR and MA orders are chosen equal for convenience. (Extension to the general case is easily obtained using representations such as in [12]).

Problem setting:

Suppose that the predictor which minimizes the criterium $E^{t-1}[(y. - \hat{y}.)^2]$ is available. Its parameters $\{a_i^{t-1}, b_i^{t-1}\}$ have been obtained from the knowledge of the data sequence $\{y_1, \dots, y_{t-1}\}$. We will write it \hat{y}_{τ}^{t-1} . When, at time t , the new data y_t is available, the n-th order linear predictor of y is computed as the projection:

$$\hat{y}_{\tau}^t = \sum_{i=1}^n a_i^t \hat{y}_{\tau-i}^{t-1} + b_i^t \tilde{y}_{\tau-i}^{t-1}, \quad \forall \tau = 1 \dots t+1 \quad (1.6)$$

In this time varying situation, the criterium $E^t[(y. - \hat{y}.)^2]$ is quadratic with respect to the parameters $\{a_i^t, b_i^t\}$ since it only depends on the past values of the previously computed parameters, that is $\{a_i^{\tau}, b_i^{\tau}, \tau < t\}$. On the other hand, it will lead to an hereditary computation. Note that although the representations (1.2) and (1.6) are different, they are asymptotically equiva-

lent in the steady state. Let us define the projection in that context.

The sequence space:

Let H_t be the Hilbert space spanned by sequences X_\bullet of length t of random vector realizations of L_P^2 (with finite variance). The quadratic norm of a sequence $X_\bullet = \{X_\tau, \tau = 1 \dots t\}$ of H_t is defined as:

$$\|X_\bullet\|_t^2 = \frac{1}{t} \sum_{\tau=1}^t \|X_\tau\|^2 \quad (1.7)$$

where $\|X\| = X^T X$ is the euclidean norm of real vectors.

Orthogonal projections:

The projection at time t of an element X_\bullet of H_t over another element Y_\bullet is defined as the linear transformation $\hat{X}_\tau^t = M^t Y_\tau$ which minimizes the norm $\|X_\bullet - \hat{X}_\tau^t\|_t^2$. The projector is the optimal matrix M^t of appropriate dimension. Minimizing the criterium with respect to M^t leads to the following well known projection equation:

$$E^t[(X_\bullet - M^t Y_\bullet) Y_\bullet^T] = 0 \quad (1.8)$$

When the process Y is not degenerated (non-singular autocorrelation), the orthogonal projection takes the following form:

$$\hat{X}_\tau^t = E^t[X_\bullet Y_\bullet^T] (E^t[Y_\bullet Y_\bullet^T])^{-1} Y_\tau, \forall \tau = 1 \dots t \quad (1.9)$$

Notations:

We will write $\hat{X}_\tau^t = P_\tau^t[X_\bullet | Y_\bullet]$ for the above projection and $\langle X | Y \rangle_t = E^t[X_\bullet (Y_\bullet)^T]$ for the extended inner product of sequences of H_t . Although this notation may seem a bit heavy, it is necessary because of the non stationarity of the extended inner product. It means precisely the projection of the trajectory $\tau \rightarrow X_\tau$ over the trajectory $\tau \rightarrow Y_\tau$, the optimization being made over the interval $[1, t]$.

With this definition, the optimal predictor of (1.6) can be rewritten as:

$$\hat{y}_\tau^t = P_\tau^t[y_\bullet | \tilde{y}_{\bullet-1}^{t-1}, \hat{y}_{\bullet-1}^{t-1}, \dots, \tilde{y}_{\bullet-n}^{t-n}, \hat{y}_{\bullet-n}^{t-n}], \forall \tau = 1 \dots t \quad (1.10)$$

Note that the above formulation is not complete as \hat{y}_τ and \tilde{y}_τ are not defined for $\tau \leq 0$. For instance, the projection of \hat{y}_1^1 cannot be computed from the formula (1.10) as \hat{y}_0^1, y_0 are not yet defined. Remember, however, that an unknown variable is equivalent to a degenerate random variable (infinite variance) so that the linear projection of any other variable over it is zero. In other words, the projection over an unknown variable is the same as the projection over zero. With this adding, the projector is then well defined and leads to consider the usual windowed case, that is $y_\tau = \hat{y}_\tau = 0, \forall \tau \leq 0$. A useful consequence of this assumption is:

$$(Y_0 = 0) \Rightarrow P_\tau^t[X_{\bullet-1} | Y_{\bullet-1}] = P_{\tau-1}^{t-1}[X_\bullet | Y_\bullet] \quad (1.11)$$

This is a direct consequence of the definition of the projector (1.9).

2. The Levinson formulas for ARMA process.

Our aim now is to recursively compute an orthogonal basis, according to the inner prod-

uct $\langle \cdot | \cdot \rangle_{t-1}$, of the space $\{\tilde{y}_{\cdot-1}^{t-1}, \hat{y}_{\cdot-1}^{t-1}, \dots, \tilde{y}_{\cdot-n}^{t-n}, \hat{y}_{\cdot-n}^{t-n}\}_{sp}$ over which the projection is made. This is commonly achieved using the classical Gram-Schmidt procedure. Note that strict application of this procedure, which run in a $2n$ -dimensional space H_{t-1}^{2n} , would lead to a rather involved computation [10]. It is convenient in our case to work with a 2-channel structure in order to preserve the recursive structure of the set with respect to time [12] [1]. In fact, we orthogonalize this set by 2-dimensional block corresponding to one step backward. The derived basis is therefore not really orthogonal (which would lead to a diagonal covariance matrix) but only block orthogonal (a 2-dimensional block diagonal covariance matrix). The space H_t^{2n} is here decomposed as $(H_t^2)^n$. It is defined below.

2.1. The unnormalized form.

Definition 1 : Let us define first the 2-dimensional vectors $Z_\tau^t \equiv [\tilde{y}_\tau^t, \hat{y}_\tau^t]^T$, $\tau = 1 \dots t-1$. We define the projection space H_{t-1}^{2n} as:

$$H_{t-1}^{2n} \equiv \{\tilde{y}_{\cdot-1}^{t-1}, \hat{y}_{\cdot-1}^{t-1}, \dots, \tilde{y}_{\cdot-n}^{t-n}, \hat{y}_{\cdot-n}^{t-n}\}_{sp} \equiv \{Z_{\cdot-1}^{t-1}, \dots, Z_{\cdot-n}^{t-n}\}_{sp} \quad (2.1)$$

The projection of y at time t can then be rewritten as:

$$\hat{y}_\tau^t = P_\tau^t[y \cdot | Z_{\cdot-1}^{t-1}, \dots, Z_{\cdot-n}^{t-n}] \quad (2.2)$$

Our goal is first to find a set of 2-dimensional vectors $\{X_{\cdot-1}^{i,t-1}\}_{i=0 \dots n-1}$ of H_{t-1}^2 which spans the same space, and such that $\forall i \neq j, \langle X_{\cdot-1}^{i,t-1} | X_{\cdot-1}^{j,t-1} \rangle_t = 0$. At time t , it is equivalent, according to (1.11), to make sure that $\langle X_{\cdot-1}^{i,t} | X_{\cdot-1}^{j,t} \rangle_t = 0$.

Definition 2 : Backward residuals obey the following defining relations:

$$\begin{cases} X_\tau^{0,t} = Z_\tau^t \\ X_\tau^{i,t} = Z_{\tau-i}^{t-i} - P_\tau^t[Z_{\cdot-i}^{t-i} | Z_{\cdot-i+1}^{t-i+1}, \dots, Z_{\cdot}^t], \forall i = 1 \dots n-1 \end{cases} \quad (2.3)$$

With this definition, it is clear that:

$$\{\tilde{y}_{\cdot-1}^{t-1}, \hat{y}_{\cdot-1}^{t-1}, \dots, \tilde{y}_{\cdot-n}^{t-n}, \hat{y}_{\cdot-n}^{t-n}\}_{sp} = \{X_{\cdot-1}^{0,t-1}\}_{sp} \oplus \dots \oplus \{X_{\cdot-1}^{n-1,t-1}\}_{sp} \quad (2.4)$$

and the one step ahead predictor can be computed as the sum of n 2-dimensional projections as:

$$\hat{y}_\tau^t = P_\tau^t[y \cdot | X_{\cdot-1}^{0,t-1}] + \dots + P_\tau^t[y \cdot | X_{\cdot-1}^{n-1,t-1}] \quad (2.5)$$

These are the 2-dimensional backward residuals of the Levinson formulas. Note that the above direct sum is related to the extended inner product of H_{t-1}^2 . It would be convenient to make the residuals computation recursive in the same way as in the classical case. To this aim, let us set the following:

Definition 3 : *Forward residuals obey the following defining relations:*

$$\begin{cases} Z_{\tau}^{0,t} = Z_{\tau}^t \\ Z_{\tau}^{i,t} = Z_{\tau}^t - P_{\tau}^t[Z_{\tau}^t | Z_{\tau}^{t-1}, \dots, Z_{\tau}^{t-i}], \forall i = 1 \dots n-1 \end{cases} \quad (2.6)$$

With this definition, for all $i = 1 \dots n-1$, we have the following decomposition:

$$\{Z_{\tau}^t, \dots, Z_{\tau}^{t-i}\}_{sp} = \{Z_{\tau}^{i,t}\}_{sp} \oplus \dots \oplus \{Z_{\tau}^{0,t-i}\}_{sp} \quad (2.7)$$

the orthogonality being related one more time to the inner product $\langle | \rangle_t$. With this decomposition, the backward residuals can be written as follows:

$$X_{\tau}^{i,t} = Z_{\tau-i}^{t-i} - P_{\tau}^t[Z_{\tau-i}^{t-i} | Z_{\tau-i}^{i-2,t-1}, \dots, Z_{\tau-i}^{0,t-i+1}] - P_{\tau}^t[Z_{\tau-i}^{t-i} | Z_{\tau}^{i-1,t}] \quad (2.8)$$

Using then that, at the previous time $t-1$, we had:

$$X_{\tau-1}^{i-1,t-1} = Z_{\tau-i}^{t-i} - P_{\tau-1}^{t-1}[Z_{\tau-i}^{t-i} | Z_{\tau-i}^{i-2,t-1}, \dots, Z_{\tau-i}^{0,t-i+1}] \quad (2.9)$$

we may rewrite (2.3), using formula (1.11), under the following recursive form:

$$X_{\tau}^{i,t} = X_{\tau}^{i-1,t-1} - P_{\tau}^t[Z_{\tau-i}^{t-i} | Z_{\tau}^{i-1,t}] \quad (2.10)$$

Carrying this expression into (2.3) and using the orthogonality of $Z_{\tau}^{i-1,t}$ within the space

$\{Z_{\tau-i}^{i-2,t-1}\}_{sp} \oplus \dots \oplus \{Z_{\tau-i}^{0,t-i+1}\}_{sp}$, we find that:

$$X_{\tau}^{i,t} = X_{\tau-1}^{i-1,t-1} - P_{\tau}^t[X_{\tau-1}^{i-1,t-1} | Z_{\tau}^{i-1,t}] \quad (2.11)$$

Note that the set $\{Z_{\tau}^{i,t}, \dots, Z_{\tau}^{0,t-i}\}_{sp}$ is degenerated since, according to (2.3), the first component of $Z_{\tau}^{i,t}$ is equal to \tilde{y}_{τ}^t , the first component of Z_{τ}^t . Therefore, because of the orthogonality $E^t[\tilde{y}_{\tau}^t \tilde{y}_{\tau}^t] = 0$, the second component is also orthogonal to \tilde{y}_{τ}^t . As a consequence, the projection of $X_{\tau}^{i,t-1}$ over \tilde{y}_{τ}^t being zero, one may simplify the projection term as:

$$P_{\tau}^t[X_{\tau-1}^{i-1,t-1} | Z_{\tau}^{i,t}] = P_{\tau}^t[X_{\tau-1}^{i-1,t-1} | z_{\tau}^{i,t}] \quad (2.12)$$

with the following definition:

Definition 4 : $z_{\tau}^{i,t} \equiv$ *second component of $Z_{\tau}^{i,t}$.*

The systems (2.3) and (2.6) are then finally equivalent to:

$$\begin{cases} X_{\tau}^{0,t} = Z_{\tau}^t \\ X_{\tau}^{i,t} = X_{\tau-l}^{i-1,t-1} - P_{\tau}^t[X_{\tau-l}^{i-1,t-1} | z_{\tau}^{i-1,t}], \forall i = 1 \dots n-l \end{cases} \quad (2.13)$$

$$\begin{cases} z_{\tau}^{0,t} = \hat{y}_{\tau}^t \\ z_{\tau}^{i,t} = z_{\tau}^{i-1,t} - P_{\tau}^t[z_{\tau}^{i-1,t} | X_{\tau-l}^{i-1,t-1}], \forall i = 1 \dots n-l \end{cases} \quad (2.14)$$

According to the orthogonal projection formula (1.9), if we define the projection coefficients as:

$$\begin{aligned} K_t^i &= E^t[X_{\tau-l}^{i,t-1} z_{\tau}^{i,t}] / E^t[(z_{\tau}^{i,t})^2] \\ Q_t^i &= E^t[z_{\tau}^{i,t} (X_{\tau-l}^{i,t-1})] (E^t[X_{\tau-l}^{i,t-1} (X_{\tau-l}^{i,t-1})^T])^{-1}, \end{aligned}$$

the systems (2.13) and (2.14) take the following form:

$$\begin{cases} X_{\tau}^{0,t} = [\tilde{y}_{\tau}^t \hat{y}_{\tau}^t]^T \\ X_{\tau}^{i,t} = X_{\tau-l}^{i-1,t-1} - K_t^{i-1} z_{\tau}^{i-1,t}, \forall i = 1 \dots n-l \end{cases} \quad (2.15)$$

$$\begin{cases} z_{\tau}^{0,t} = \hat{y}_{\tau}^t \\ z_{\tau}^{i,t} = z_{\tau}^{i-1,t} - Q_t^{i-1} X_{\tau-l}^{i-1,t}, \forall i = 1 \dots n-l \end{cases} \quad (2.16)$$

Remark: Note that the projection coefficients may seem to be implicitly defined as they depend on the forward residuals $\{z_{\tau}^{i,t}\}_{i=0 \dots n-l}$ which are not available at the step t . But, according to (2.16) and to the orthogonality properties (2.4), it is clear that:

$$\begin{aligned} E^t[z_{\tau}^{i,t} X_{\tau-l}^{i,t-1}] &= E^t[z_{\tau}^{0,t} X_{\tau-l}^{i,t-1}] \\ &= E^t[\hat{y}_{\tau}^t X_{\tau-l}^{i,t-1}] \end{aligned} \quad (2.17)$$

Moreover, as $\tilde{y}_{\tau}^t = y_{\tau} - \hat{y}_{\tau}^t$ is orthogonal to $\{X_{\tau-l}^{i,t-1}\}_{i=0}^{n-l}$, we have:

$$E^t[z_{\tau}^{i,t} X_{\tau-l}^{i,t-1}] = E^t[y_{\tau} X_{\tau-l}^{i,t-1}] \quad (2.18)$$

and the predictor parameters are well defined.

2.2. The normalized form.

The above orthogonalization implies that the magnitude of the forward residuals $z_{\tau}^{i,t}$ decreases quickly with i . As a consequence, it is interesting, from a numerical point of view, to normalize the forward and backward residuals in view to deal with quantities of similar magnitude. Concerning the forward residuals $z_{\tau}^{i,t}$, there is a unique way to make the normalization, that is to define the n variables $u_{\tau}^{i,t}$ as:

$$u_{\tau}^{i,t} = z_{\tau}^{i,t} / \|z_{\tau}^{i,t}\| \quad (2.19)$$

with $\|z_{\cdot}^{i,t}\|_t = \sqrt{E^t[(z_{\cdot}^{i,t})^2]}$ as a definition.

On the other hand, the normalization of the backward residuals is not uniquely defined as they are 2-dimensional vectors, but only up to any non singular orthogonal matrix. There is however a natural way to achieve this normalization:

Consider the n 2-dimensional normalized vectors of H_t^2 $\{V_{\cdot}^{i,t}\}_{i=0\dots n-1}$ such that $V_{\tau}^{i,t} = R_t^i X_{\tau}^{i,t}$, $\forall i = 0\dots n-1$ with $E^t[V_{\cdot}^{i,t-1}(V_{\cdot}^{i,t-1})^T] = I$, the identity matrix. At time t , it leads to make sure that $E^{t+1}[V_{\cdot}^{i,t}(V_{\cdot}^{i,t})^T] = I$ which is equivalent to make sure that $\frac{t}{t+1}E^t[V_{\cdot}^{i,t}(V_{\cdot}^{i,t})^T] = I$. We proceed to compute the normalization matrices recursively on the order $i = 0\dots n$.

Let us develop the normalized backward vectors. Equations (2.13) and (2.19) lead to:

$$V_{\tau}^{i,t} = R_t^i (R_{t-1}^{i-1})^{-1} (V_{\tau-1}^{i-1,t-1} - E^t[V_{\cdot}^{i-1,t-1} u_{\cdot}^{i-1,t}]) u_{\tau}^{i-1,t} \quad (2.20)$$

Let us then compute the autocorrelation matrix of $V_{\tau}^{i,t}$:

$$\begin{aligned} E^t[V_{\cdot}^{i,t}(V_{\cdot}^{i,t})^T] &= R_t^i (R_{t-1}^{i-1})^{-1} (E^t[V_{\cdot}^{i-1,t-1}(V_{\cdot}^{i-1,t-1})^T]) \\ &- E^t[V_{\cdot}^{i-1,t-1} u_{\cdot}^{i-1,t}] E^t[(u_{\cdot}^{i-1,t})^2] E^t[(V_{\cdot}^{i-1,t-1})^T u_{\cdot}^{i-1,t}] \\ &\times (R_t^i (R_{t-1}^{i-1})^{-1})^T \end{aligned} \quad (2.21)$$

At the previous step, one has:

$$E^t[V_{\cdot}^{i-1,t-1}(V_{\cdot}^{i-1,t-1})^T] = I \quad (2.22)$$

If one defines the projection coefficients as:

$$S_t^i = E^t[V_{\cdot}^{i,t-1} u_{\cdot}^{i,t}], \quad (2.23)$$

the relation (2.21) takes the following form:

$$I = R_t^i (R_{t-1}^{i-1})^{-1} \left(\frac{t}{t+1} \right) (I - S_t^{i-1} (S_t^{i-1})^T) (R_t^i (R_{t-1}^{i-1})^{-1})^T \quad (2.24)$$

As a consequence, it is natural to choose the normalization matrices R_t^i such that:

$$R_t^i (R_{t-1}^{i-1})^{-1} = \left(\frac{t}{t+1} (I - S_t^{i-1} (S_t^{i-1})^T) \right)^{-1/2} \quad (2.25)$$

A similar procedure, applied to the normalization coefficients of the forward residuals leads to the relation:

$$\frac{\|z_{\cdot}^{i,t}\|_t}{\|z_{\cdot}^{i-1,t}\|_t} = (I - (S_t^{i-1})^T S_t^{i-1})^{-1/2} \quad (2.26)$$

Finally, the orthonormalized form writes:

$$V_{\tau}^{0,t} = \begin{bmatrix} E^{t+1}[(y_{\cdot-t} - \hat{y}_{\cdot-t}^t)^2] & 0 \\ 0 & E^{t+1}[(\hat{y}_{\cdot-t}^t)^2] \end{bmatrix}^{-1/2} \begin{bmatrix} (y_{\tau} - \hat{y}_{\tau}^t) \\ \hat{y}_{\tau}^t \end{bmatrix} \quad (2.27)$$

$$V_{\tau}^{i,t} = \left[\left(\frac{t}{t+1} \right) (I - S_t^{i-1} (S_t^{i-1})^T) \right]^{-1/2} (V_{\tau-t}^{i-1,t-1} - S_t^{i-1} u_{\tau}^{i-1,t}), i = 1 \dots n-1$$

$$\begin{aligned} z_{\tau}^{0,t} &= \hat{y}_{\tau}^t / \|z_{\cdot}^{0,t}\|_t \\ u_{\tau}^{i,t} &= (I - (S_t^{i-1})^T S_t^{i-1})^{-1/2} (u_{\tau}^{i-1,t} - (S_t^{i-1})^T V_{\tau-t}^{i-1,t-1}) \forall i = 1 \dots n-1 \end{aligned} \quad (2.28)$$

Figure 1 illustrates the involved lattice structure with $N^i \equiv (I - S^{i-1} (S^{i-1})^T)^{-1/2}$ and $n^i \equiv (I - (S^{i-1})^T S^{i-1})^{1/2}$ as definitions.

3. The algorithm.

3.1. The recursive steps.

The aim of the algorithm is to compute the parameters of the predictor, that is S_t^i . As it has been set in our problem statement, they depend on the whole trajectories $\{y_{\tau}, \tau = 1 \dots t\}$ and $\{\hat{y}_{\tau}^{t-1}, \tau = 1 \dots t-1\}$ and are not actually finitely recursive. Let us examine more precisely what this means. According to (2.23), (2.18) and (2.19), the lattice parameters are defined as follows:

$$S_t^i = E^t[V_{\cdot-t}^{i,t-1} y_{\cdot}] / \|z_{\cdot}^{i,t}\|_t \quad (3.1)$$

Recall that the forward residuals obey the following recursion:

$$\begin{aligned} z_{\tau}^{i,t} &= z_{\tau}^{i+1,t} + P_{\tau}^t[y_{\cdot} | X_{\cdot-t}^{i,t-1}] \\ &= z_{\tau}^{i+1,t} + E^t[y_{\cdot} V_{\cdot-t}^{i,t-1}] V_{\tau-t}^{i,t-1} \end{aligned} \quad (3.2)$$

So, we have:

$$\|z_{\cdot}^{i,t}\|_t = \sqrt{\sum_{k=i}^{n-1} \|E^t[y_{\cdot} V_{\cdot-t}^{k,t-1}]\|^2}, \forall i = 0 \dots n-1 \quad (3.3)$$

and it appears that the parameters are directly linked to the quantities $E^t[V_{\cdot-t}^{i,t-1} y_{\cdot}]$. In fact, one has to remember the whole past trajectory $\{V_{\tau}^{i,t-1}, \tau = 1 \dots t-1\}_{i=0 \dots n-1}$ in order to compute the new trajectory $\{V_{\tau}^{i,t}, \tau = 1 \dots t\}_{i=0 \dots n-1}$ which allows to extract the expression $E^t[V_{\cdot-t}^{i,t-1} y_{\cdot}]$ computed as the sum over the interval $[1, t]$ and finally the parameters with

(3.1) and (3.3). It should be noted that square roots of the matrices $I - S_t^i (S_t^i)^T$ can be computed using the algebraic following identity:

$$(S \in R^n / \|S\| \neq I) \Rightarrow (I - SS^T)^{-1/2} = (I - SS^T)((I - S^T S)^{-1/2} - I) / S^T S \quad (3.4)$$

The algorithm is described by Figure 2.

3.2. The initial steps.

As was noticed in 1., the data prior to time 0 are assumed to be completely unknown, so that we cannot make any projection over them. As a consequence, the coefficients related to these variables are undetermined (or zero) and the algorithm established above doesn't run. Recall that the predictor is defined by (1.10). Because we deal with the one-step-ahead predictor, it is clear that the first estimator to be non singular is $\{\hat{y}_\tau^2, \tau = 2, 3\}$. The only variable over which the projection can be made is indeed y_1 . So, at this step (time $t = 2$), the predictor is defined by $\hat{y}_\tau^2 = P_\tau^2[y_\bullet | y_{\bullet-1}]$, for $\tau = 2, 3$ and the criterium to minimize is $(y_2 - \hat{y}_2^2)^2$. At the next step (time $t = 3$), \hat{y}_2^2 being available, one might project over the set $\{y_1, y_2, \hat{y}_2^2\}$ but it is easy to check that this set is singular due to the fact that the predictor \hat{y}_2^2 is proportional to y_1 . In other words, it is equivalent to project over the spaces $\{y_2, \hat{y}_2^2\}$ or $\{y_1, y_2\}$ minimizing the criterium $(y_2 - \hat{y}_2^2)^2 + (y_3 - \hat{y}_3^2)^2$. We have chosen the former and therefore define $\hat{y}_\tau^3 = P_\tau^3[y_\bullet | \hat{y}_{\tau-1}^2, \hat{y}_{\tau-1}^2]$. In fact, as soon as a new observation is available, only one new parameter is determined. So, during the first $2n$ steps of the algorithm (n being the system order), one has to distinguish two kinds of projections:

- $t = 2k$, then $\hat{y}_\tau^t = P_\tau^t[y_\bullet | Z_{\tau-1}^{t-1}, \dots, Z_{\tau-k+1}^{t-k+1}, \tilde{y}_{\tau-k}^{t-k}]$
- $t = 2k + 1$ then $\hat{y}_\tau^t = P_\tau^t[y_\bullet | Z_{\tau-1}^{t-1}, \dots, Z_{\tau-k}^{t-k}]$

Clearly, the odd steps are quite identical to the recursive steps. For the even steps, one has to modify the general procedure. Note first that, for odd steps $t = 2k - 1$, one has, $u_{\tau}^{k-2, t} = S_t^{k-2} V_{\tau-1}^{k-2, t-1}$. As a consequence, $E^t[(u_{\tau}^{k-2, t})^2] = I$ implies that $\|S_t^{k-2}\|^2 = I$. The normalization matrix of the backward residuals $X_{\tau}^{k-1, t}$ is then singular according to (3.4) as noticed above. So, define the unnormalized backwards residuals as follows:

$$\begin{cases} X_{\tau}^{0, t-1} = Z_{\tau}^{t-1} \\ X_{\tau}^{i, t-1} = X_{\tau-1}^{i-1, t-2} - P_{\tau}^{t-1} [X_{\tau-1}^{i-1, t-2} | Z_{\tau-1}^{i-1, t-1}], \forall i = 1 \dots k-2 \\ x_{\tau}^{k-1, t-1} = x_{\tau-1}^{k-2, t-2} - P_{\tau}^{t-1} [x_{\tau-1}^{k-2, t-2} | Z_{\tau-1}^{k-2, t-1}] \end{cases} \quad (3.5)$$

where

Definition 5 : $x_{\tau}^{k-2, t-2} \equiv$ first component of $X_{\tau}^{k-2, t-2}$.

Clearly we have the orthogonal decomposition:

$$\{Z_{\tau-1}^{t-1}, \dots, Z_{\tau-k+1}^{t-k+1}, \tilde{y}_{\tau-k}^{t-k}\}_{sp} = \{X_{\tau-1}^{0, t-1}\}_{sp} \oplus \dots \oplus \{X_{\tau-1}^{k-2, t-1}\}_{sp} \oplus \{x_{\tau-1}^{k-1, t-1}\}_{sp}$$

with respect to the inner product $\langle | \rangle_{t-1}$.

Let us now normalize these backward residuals. Formulas concerning the $k-1$ first residuals are obviously the same as for the recursive steps. Concerning the last one, that is

$x_{\tau}^{k-1, t-1}$, a particular step is necessary. Let $v_{\tau}^{k-1, t-1} = r_{t-1}^{k-1} x_{\tau}^{k-1, t-1}$ be the normalized residual. The last equation of (3.5) allows to compute the normalization coefficient r_{t-1}^{k-1} as:

$$v_{\tau-1}^{k-1, t-1} = r_{t-1}^{k-1} [I \ 0] (R_{t-1}^{k-2})^{-1} (V_{\tau-2}^{k-2, t-2} - S_{t-2}^{k-2} u_{\tau-1}^{k-2, t-1}) \quad (3.6)$$

Setting $\rho = r_{t-1}^{k-1} [I \ 0] (R_{t-1}^{k-2})^{-1}$ and assuming that $E^t[(x_{\tau-1}^{k-1, t-1})^2] = I$ lead to:

$$\rho \frac{t-1}{t} (I - S_{t-1}^{k-2} (S_{t-1}^{k-2})^T) \rho^T = I \quad (3.7)$$

Let us note the component of S as $S = [(S)_I, (S)_2]^T$. using the fact that:

$$\|S_{t-1}^{k-2}\|^2 = (S_{t-1}^{k-2})_I^2 + (S_{t-1}^{k-2})_2^2 = I \quad (3.8)$$

the matrix $I - SS^T$ may be written:

$$\begin{aligned} I - SS^T &= \begin{bmatrix} I - (S)_I^2 & -(S)_I(S)_2 \\ -(S)_I(S)_2 & I - (S)_2^2 \end{bmatrix} = \begin{bmatrix} (S)_2^2 & -(S)_I(S)_2 \\ -(S)_I(S)_2 & (S)_I^2 \end{bmatrix} \\ &= \begin{bmatrix} -(S)_2 \\ (S)_I \end{bmatrix} \begin{bmatrix} -(S)_2 & (S)_I \end{bmatrix} \end{aligned} \quad (3.9)$$

So, one natural solution of equation (3.7) is:

$$\rho = \sqrt{\frac{t}{t-1}} \begin{bmatrix} -(S_{t-2}^{k-2})_2 \\ (S_{t-2}^{k-2})_I \end{bmatrix} \quad (3.10)$$

Finally, we have:

$$v_{\tau-1}^{k-1, t-1} = \sqrt{\frac{t}{t-1}} \begin{bmatrix} -(S_{t-2}^{k-2})_2 \\ (S_{t-2}^{k-2})_I \end{bmatrix} (V_{\tau-2}^{k-2, t-2} - S_{t-2}^{k-2} u_{\tau-1}^{k-2, t-1}) \quad (3.11)$$

3.3. Adaptativity.

The hereditary nature of our algorithm, which preserves all correlation terms, proves to be essential for accuracy. However, for adaptive purposes, a limit must be imposed upon heredity in order to follow-up possible drifts with time. This is achieved by imposing a finite past time-horizon upon the hereditary past. Note N this horizon length. In this case, the hereditary part of the algorithm is modified as follows:

- While $t < N$, the algorithm is the same that the one described in Figure 2.

For $t \geq N$, the hereditary loop is computed from $\tau = t - N + 1$ to $t - 1$ and the formulas of the algorithm have to be modified as follows:

$$V_{\tau}^{0, t} = \begin{bmatrix} (y_{\tau} - \hat{y}_{\tau}^t) / (E^t[y_{\cdot}^2] - \|z_{\cdot}^{0, t}\|_t) \\ \hat{y}_{\tau}^t / \|z_{\cdot}^{0, t}\| \end{bmatrix} \quad (3.12)$$

$$V_{\tau}^{i, t} = [I - S_t^{i-1} (S_t^{i-1})^T]^{-1/2} (V_{\tau-1}^{i-1, t-1} - S_t^{i-1} u_{\tau}^{i-1, t}), i = 1 \dots n-1$$

$$\text{with } E^t[y^2] = \frac{1}{N} \sum_{\sigma=t-N+1}^t y_{\sigma}^2 \quad (3.13)$$

$$E^{t+1}[y \cdot V_{\cdot-1}^{i,t}] = \frac{1}{N} \sum_{\sigma=t-N+2}^{t+1} y_{\sigma} V_{\sigma-1}^{i,t} \quad (3.14)$$

4. Asymptotic convergence.

In a previous paper [1], the authors have established a realization algorithm of the exact ARMA lattice predictor of order n , from the autocorrelation function of the process. This is valid only when a precise model of the process exists (ideal spectra, fractional or distributed parameters systems,...). When only experimental data are available, constructing the ARMA lattice predictor of order n from those data solves in particular the identification problem of an n -th order linear system in ARMA form, when that is the real case.

To show that, it suffices to take a note of:

4.1. The certainty equivalence property.

The lattice realization algorithm of [1], and the lattice identification algorithm of the present paper are identical, up to the substitution:

$$E[y_t V_{t-1}^{i,t}] \rightarrow E^{t+1}[y \cdot V_{\cdot-1}^{i,t}] \quad (4.1)$$

Proof: It suffices to compare Figure 2 of the present paper with Figure 4 of [1] (together with its formula 4.11, 4.12 and 4.13).

4.2. Asymptotic convergence.

Proposition 1 . *Whenever $\{y_t\}$ is an ergodic process in the L^2 sense, the hereditary algorithm converges in probability to the unique stationary optimal predictor.*

Proof: When one compares the algorithms appearing in each of the papers mentioned above, with property 4.1 in mind, one immediately concludes by simple inspection that they both represent the same (state+parameters) hereditary transition map, with simply a different input. Indeed, we may write, on the one hand:

$$\begin{pmatrix} v_t \\ p_t \\ \|z_t^0\| \end{pmatrix} = r \left(\begin{pmatrix} v_{t-1} \\ p_{t-1} \\ \|z_{t-1}^0\| \end{pmatrix}, \{E[y_t y_{t-\tau}]\}_{0 \leq \tau < t} \right) \quad (4.2)$$

for the realization algorithm with $v \in R^{2n \times t}$, $p \in R^{2n}$, $\|z^0\| \in R$ and predictor \hat{y}_t as:

$$\hat{y}_t = l(v_t, p_t, \|z_t^0\|) \quad (4.3)$$

and, on the other hand:

$$\begin{bmatrix} V_t \\ S_t \\ \|Z_t^0\| \end{bmatrix} = r \left(\begin{bmatrix} V_{t-1} \\ S_{t-1} \\ \|Z_{t-1}^0\| \end{bmatrix}, \{E^t[y \cdot y_{t-\tau}]\}_{0 \leq \tau < t} \right) \quad (4.4)$$

for the identification algorithm with $V \in R^{2n \times t}$, $S \in R^{2n}$, $\|Z^0\| \in R$ and predictor \hat{Y}_t as:

$$\hat{Y}_t = l(V_t, S_t, \|Z_t^0\|) \quad (4.5)$$

In both cases, r and l are the same maps, from which it results that the input-output maps of both algorithms are the same. They are rational (containing only the four arithmetical operations) and continuous (all inversions are non-singular by definition). Note that the symbols \hat{Y} , V and Z have been used, instead of \hat{y} , v and z in both cases, to stress the fact that they are not numerically identical for finite t .

From [1], system (4.2) is the exact time-varying solution, for all initial states, of the optimal predictor. Since ergodicity implies stationarity, the asymptotic solution of (4.2) is, by definition, the unique value of the stationary optimal predictor.

System (4.4)-(4.5) is identical to system (4.2)-(4.3), except for its input signal $E^t[y \cdot y_{t-\tau}]$ instead of $E[y_t y_{t-\tau}]$. One has to prove that the asymptotic output of (4.4)-(4.5) tends to the asymptotic output of (4.2)-(4.3) in a suitable topology.

Let us choose the L^2 -ergodic topology for the input, i. e.:

$$E^t[y \cdot y_{t-\tau}] \xrightarrow{L^2\text{-erg}} E[y_t y_{t-\tau}] \quad \text{iff} \quad \frac{1}{T} \int_0^T \|E^t[y \cdot y_{t-\tau}] - E[y_t y_{t-\tau}]\|^2 d\tau \rightarrow 0.$$

Let Φ be the common input-output map of both algorithms. By continuity of Φ , we have the following diagram:

$$\begin{array}{ccc} E[y_t y_{t-\tau}] & \xrightarrow{\Phi} & \{\hat{y}_t\} \\ \uparrow L^2\text{-erg} & & \uparrow L^2\text{-erg} \\ E^t[y \cdot y_{t-\tau}] & \xrightarrow{\Phi} & \{\hat{Y}_t\} \end{array}$$

As L^2 -erg convergence implies convergence in probability, the proof is completed.

5. Examples and comparisons.

As was announced in the abstract, the algorithm proposed here is shown to be more accurate than classical solutions, in term of spectral resolution, especially in the transient phase.

5.1. Spectral resolution.

To illustrate this property, we have first chosen a simple 4-dimensional linear stochastic system whose poles and zeros are represented in Figure 7 (a). The module of the corresponding transfer function (square root of the spectral density) is represented from 0 to the half of the sampling frequency in Figure 7 (b).

Figure 4 shows the transient spectrum obtained with finitely recursive algorithms such as the “Extended Least Square” and “Recursive Maximum Likelihood” algorithms [14]-[15]. It appears clearly that these algorithms do not separate the two distinct poles of the system, even after 2000 iterations. Note that, for such low-order examples as above, the lattice character of our algorithm is of little concern. The differences mainly lie in frequency resolution.

In Figure 9, we exhibit the power spectra obtained with the classical Gauss-Newton Block Data algorithm (GNBD) [14], from the MATLAB identification tool-box, compared with the one obtained using the Hereditary Least Square algorithm (HLS). We have run the GNBD algorithm for successive data sequences of length 50, 150, 200... Although the GNBD algorithm converges to the real power spectrum of the signal (after a number of steps greater than 750), it clearly appears that the hereditary algorithm detects the two poles much faster (after 150 steps only). In order to underline this fact, we represent in Figure 6 all the power spectra (more precisely, their square roots) obtained after 150 iterations with all the methods we tested.

As is well known, it is the more difficult to distinguish two resonance modes as they are nearer to each other. We therefore have chosen, as a second example, the linear stochastic system described in Figure 7 through its pole/zero representation.

We see in Figure 9 that the GNBD algorithm does not separate the two frequencies for a data length lower than 2000, which is not the case with our algorithm which early detects them. (Note that the frequency scale has been expanded to make sure a better reading).

Concerning the finitely recursive algorithms (ELS and RML), it appears clearly in Figure 9 that they do not distinguish between the two frequencies. The lack of resolution on the part of ELS and RML algorithms is due to the approximations which are inherent to the usual finitely recursive algorithms, whose consequence is to skip higher interval correlation terms.

5.2. Adaptativity.

The signal to be identified consists of four frequencies with a drift along time which contains information. This is the case in the SONAR passive estimation, the drift being due to the Doppler effect. A simulation is presented in Figure 10 and shows the accuracy of our algorithm to match the emitted spectrum. The dotted line represents the spectrum to be identified and the continuous line the results of our HLS algorithm.

Due to the superior transient capabilities of the HLS algorithm, a local time-horizon of 500 appears to be sufficient to follow-up the drift while preserving accuracy on modes which are closed to each other. It should be noticed, in the above example, that the lattice approach is necessary from a numerical point of view.

5.3. Order reduction.

Finally, we exhibit the performance of our algorithm when order reduction is required. As is well known, a good test for spectral parameters identification is to choose a signal spectrum with abrupt slopes. This implies higher order whose low order approximations are difficult to get in an optimal way [1]. We simulated a low-pass filter of dimension 20 (synthesized with the algorithm of [1]) and use it to generate the signal to be identified with a 4-dimensional system. The results appear in Figure 11, Figure 12 and Figure 13 and confirm again the accuracy of the hereditary algorithm.

6. Conclusion.

We have derived in this work a lattice solution to the identification problem of the optimal linear predictor of ARMA type, which leads to a recursive algorithm with hereditary terms. These terms, which are inherent to the statistical optimization, are due to the increasing memory structure of higher interval correlation computations which are essential for an accurate spectral resolution. Heredity consists in preserving such terms which implies a growing memory algorithm as opposed to more usual approximations in view of finite recursiveness. The projection technique introduced produces a minimum phase linear model of the signal whose n -th order “whiteness” of the associated innovation has the following restricted meaning: orthogonality to an n -dimensional subspace memory of the past in a suitable Hilbert sequence space. This is suitable for stationary signals, in which case we have shown the optimality of the hereditary algorithm for fixed samples of data as well as its asymptotic convergence, for unlimited records. In non-stationary situations, the algorithm exhibits also superior adaptive properties, when a limited window over the past is considered. Finally, we give an example of the better parametric properties of the hereditary algorithm, when trying to fit a low order predictor to higher dimensional spectrum such as the ideal low-pass.

7. References

- [1] **A. Monin, G. Salut**, “Exact ARMA Lattice Predictor From Autocorrelation Functions”, IEEE Trans. on Signal Processing, vol SP-42, n°4, pp 877-886, 1994.
- [2] **D. T. L. Lee, B. Friedlander, M. Morf**, “Recursive Ladder Algorithms for ARMA Modeling”, IEEE Trans. on Automatic Control, vol. AC-27, n°4, pp. 753-764, 1982.
- [3] **F. Ling, J. G. Proakis**, “A Generalized Multichannel Least Squares Lattice Algorithm Based on Sequential Processing Stages”, IEEE Trans. on Acoust. Speech, Signal Processing, vol. ASSP-32, n°2, pp. 381-389, 1984.
- [4] **S. Li, B. W. Dickinson**, “An Efficient Method to Compute Consistent Estimates of the AR Parameters of an ARMA Model”, IEEE Trans. on Automatic Control, vol. AC-31, n°3, pp. 275-278, 1986.
- [5] **F. Ling, D. Manolakis, J. G. Proakis**, “A recursive Modified Gram-Schmidt Algorithm for Least-Squares Estimation”, IEEE Trans. on Acoust. Speech, Signal Processing, vol. ASSP-34, n°4, pp. 829-835, 1986.
- [6] **F. Ling, D. Manolakis, J. G. Proakis**, “Numerically Robust Least-Squares Lattice-Ladder Algorithms with Direct Updating of the Reflection Coefficients”, IEEE Trans. on Acoust. Speech, Signal Processing, vol. ASSP-34, n°4, pp. 837-845, 1986.
- [7] **S. Karaboyas, N. Kaloupsidis**, “Efficient Adaptive Algorithms for ARX Identification”, IEEE Trans. on Signal Processing, vol. SP-39, n°3, pp. 571-582, 1991.
- [8] **D. T. M. Slock, L. Chisci, H. Lev-Ari, T. Kailath**, “Modular and Numerically Stable Fast Transversal Filters for Multichannel and Multi-experiment RLS”, IEEE Trans. on Signal Processing, vol. SP-40, n°4, pp. 784-802, 1992.
- [9] **G. Glentis, N. Kaloupsidis**, “Fast Adaptive Algorithm for Multichannel Filtering and System Identification”, IEEE Trans. on Acoust. Speech, Signal Processing, vol. ASSP-40, n°10, 1992.
- [10] **E. Carlson, M. H. Hayes**, “Least Squares ARMA Modeling of Linear Time-Varying Systems: Lattice Filter Structures and Fast RLS Algorithm”, IEEE Trans. on Acoust. Speech, Signal Processing, vol. ASSP-35, n°7, pp. 994-1014, 1987.

- [11] **John J. Shynk**, “Adaptive IRR Filtering”, IEEE Acoust. Speech, Signal Processing Magazine, vol. 6, pp. 4-21, 1989.
- [12] **A. Benveniste, C. Chaure**, “AR and ARMA Algorithms of Levinson Type: An Innovations Approach”, IEEE Trans. on Automatic Control, vol. AC-26, n°6, pp. 1243-1261, 1981.
- [13] **A. Benveniste, M. Metivier, P. Priouret**, *Algorithmes Adaptatifs Et Approximations Stochastiques*, Edition Masson, Paris, 1987.
- [14] **L. Ljung**, “System Identification: Theory for the User”, Prentice-Hall, Inc. Englewood Cliffs, New Jersey 07632, 1987.
- [15] **F. Michaut**, *Méthodes Adaptatives pour le Signal.*, Edition Hermès, Paris, 1992.
- [16] **A. Monin, G. Salut**, “The Optimal Linear Predictor In Reduced Dimension”, Proceedings European Control Conference, Grenoble, France, 1991.
- [17] **T. Soderstrom, L. Ljung, I. Gustavson**, “A Theoretical Analysis Of Recursive Identification Methods”, Automatica, vol. 14, pp. 231-244, 1978.

Figure and table

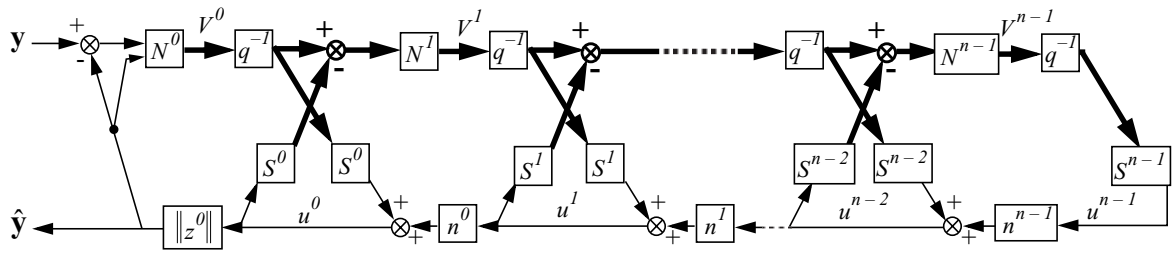


Figure 1 - The normalized lattice structure

For $t = 1$ to T

Computation of the backward residuals $V_{\tau}^{i,t}, \tau = 1 \dots t - 1$

For $\tau = 1$ to t

$$V_{\tau}^{0,t} = \begin{bmatrix} (y_{\tau} - u_{\tau}^{0,t} \|z_{\cdot}^{0,t}\|_{\tau}) / \left(\sqrt{\frac{t}{t+1}} (E^t[y_{\cdot}^2] - \|z_{\cdot}^{0,t}\|_{\tau}) \right) \\ u_{\tau}^{0,t} / \sqrt{\frac{t}{t+1}} \end{bmatrix}$$

For $i = 1$ to $n - 1$

Hereditary terms

$$V_{\tau}^{i,t} = \left[\left(\frac{t}{t+1} \right) (I - S_t^{i-1} (S_t^{i-1})^T)^{-1/2} (V_{\tau-1}^{i-1,t-1} - S_t^{i-1} u_{\tau}^{i-1,t}) \right]$$

$$u_{\tau}^{n-1,t} = (S_t^{n-1})^T V_{\tau-1}^{n-1,t-1}$$

For $i = n - 2$ down to 0

$$u_{\tau}^{i,t} = (I - (S_t^i)^T S_t^i)^{1/2} u_{\tau}^{i+1,t} + (S_t^i)^T V_{\tau-1}^{i,t-1}$$

Computation of the means $E^{t+1}[y_{\cdot} V_{\cdot-1}^{i,t}]$

For $i = 0$ to $n - 1$

$$E^{t+1}[y_{\cdot} V_{\cdot-1}^{i,t}] = \frac{1}{t+1} \sum_{\sigma=1}^{t+1} y_{\sigma} V_{\sigma-1}^{i,t}$$

Computation of the parameters S_{t+1}^i

For $i = 0$ to $n - 1$

$$\|z_{\cdot}^{i,t+1}\|_{t+1} = \sqrt{\sum_{k=i}^{n-1} \|E^{t+1}[y_{\cdot} V_{\cdot-1}^{k,t}]\|^2}$$

$$S_{t+1}^i = E^{t+1}[V_{\cdot-1}^{i,t} y_{\cdot}] / \|z_{\cdot}^{i,t+1}\|_{t+1}$$

Figure 2 - The hereditary algorithm.

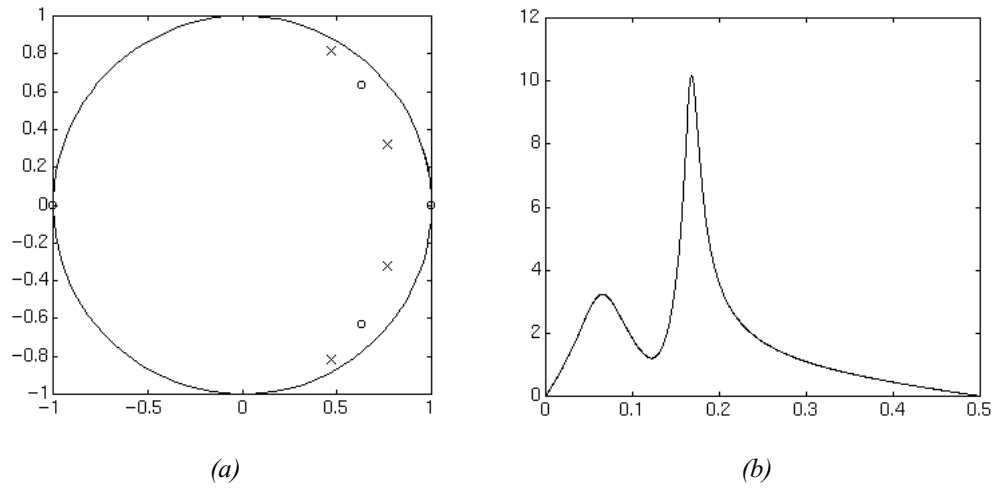


Figure 3 - Poles (x) and zeros (o) location (a) and power spectrum (b) (example 1).

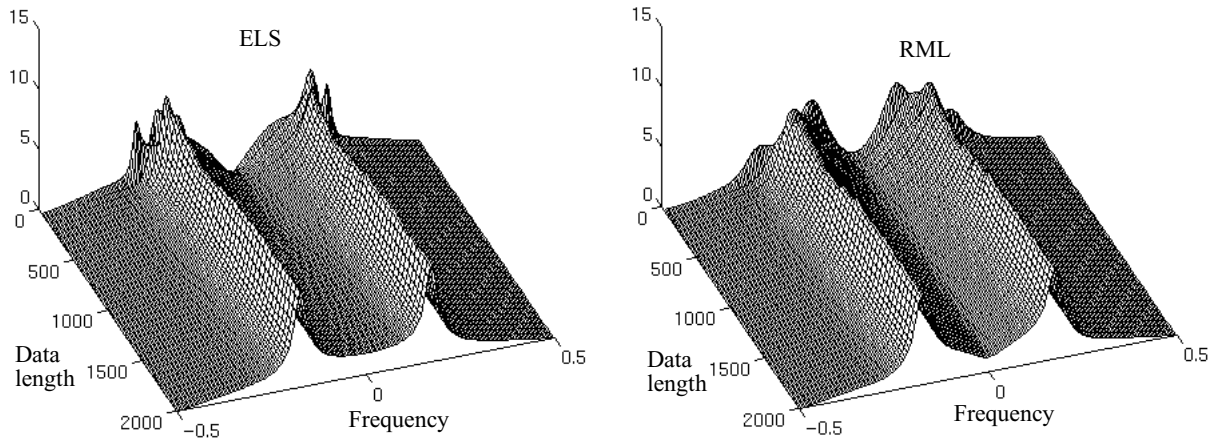


Figure 4 - Spectral density obtained with classical finitely recursive identification algorithms (example 1)

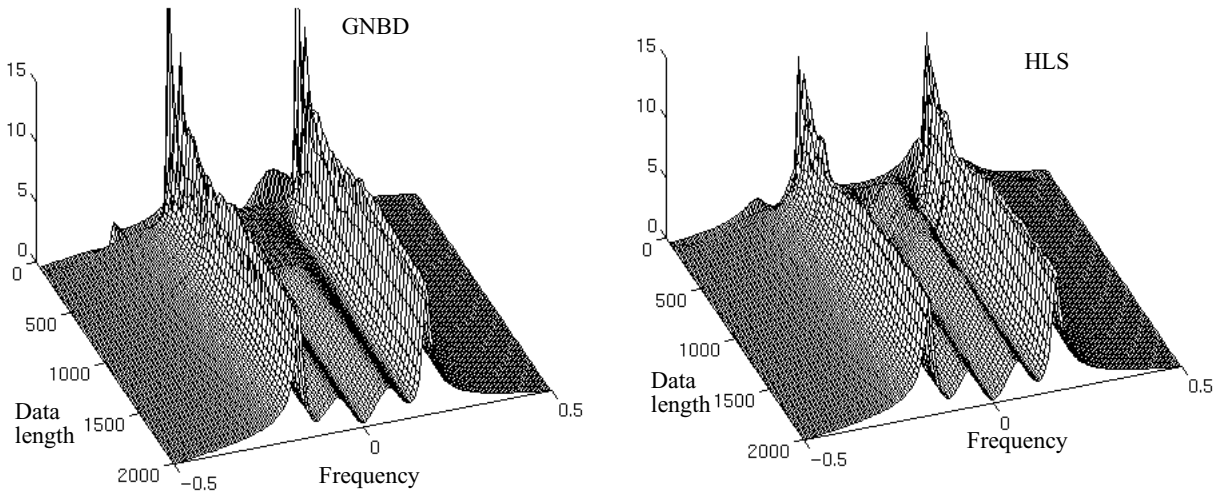


Figure 5 - Transient spectral densities obtained with Gauss-Newton and hereditary algorithms (example 1).

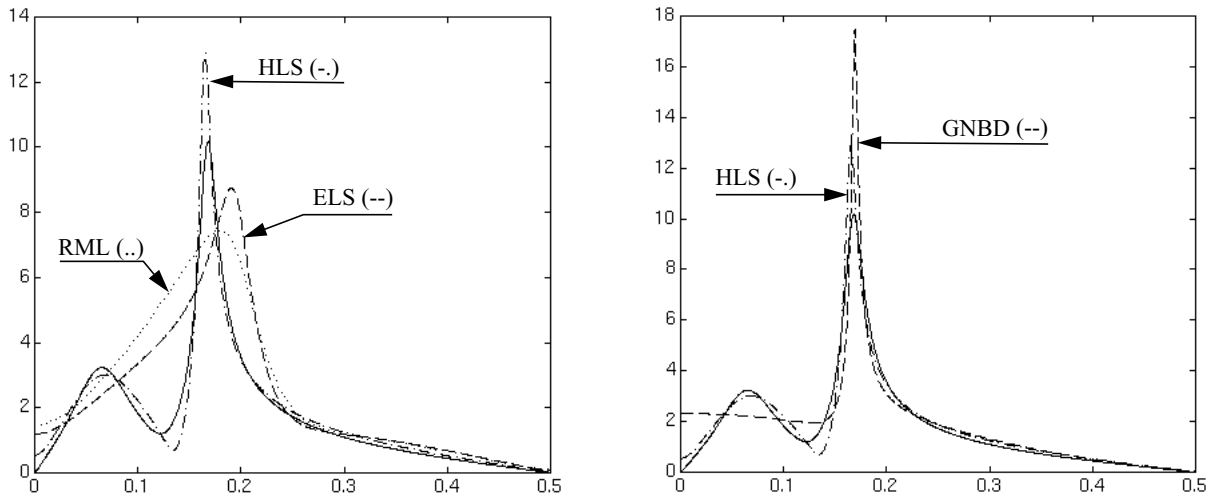


Figure 6 - Power spectra obtained for a 150 long sequence (example 1).

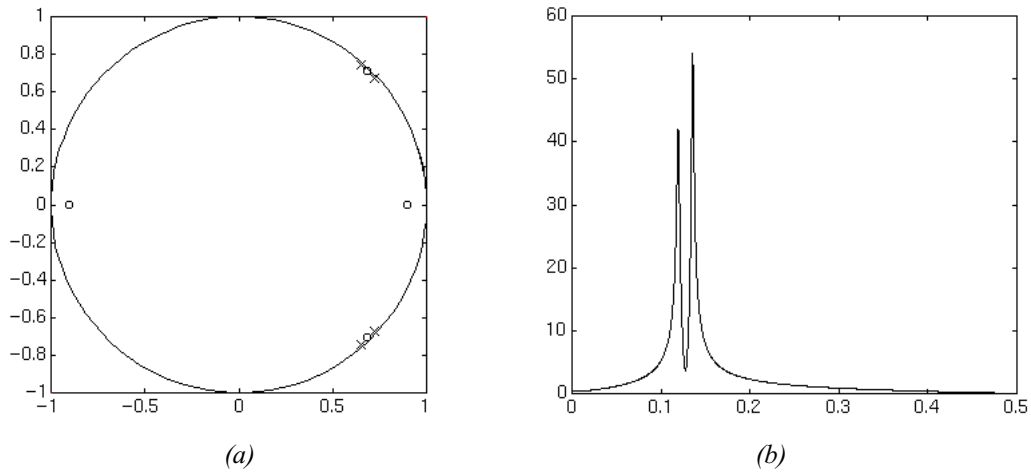


Figure 7 - Poles (x) and zeros (o) location (a) and power spectrum (b) (example 2).

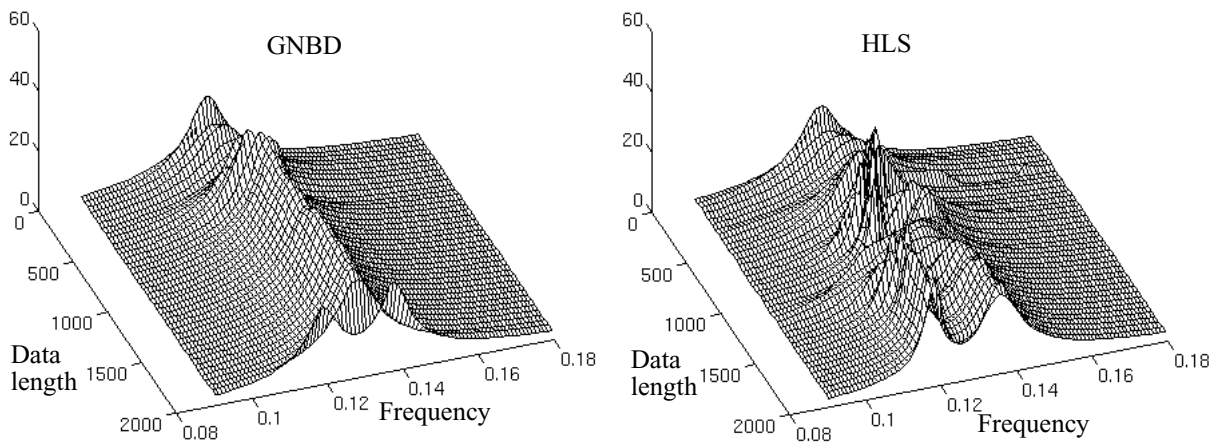


Figure 8 - Transient spectral densities obtained with Gauss-Newton and hereditary algorithms (example 2).

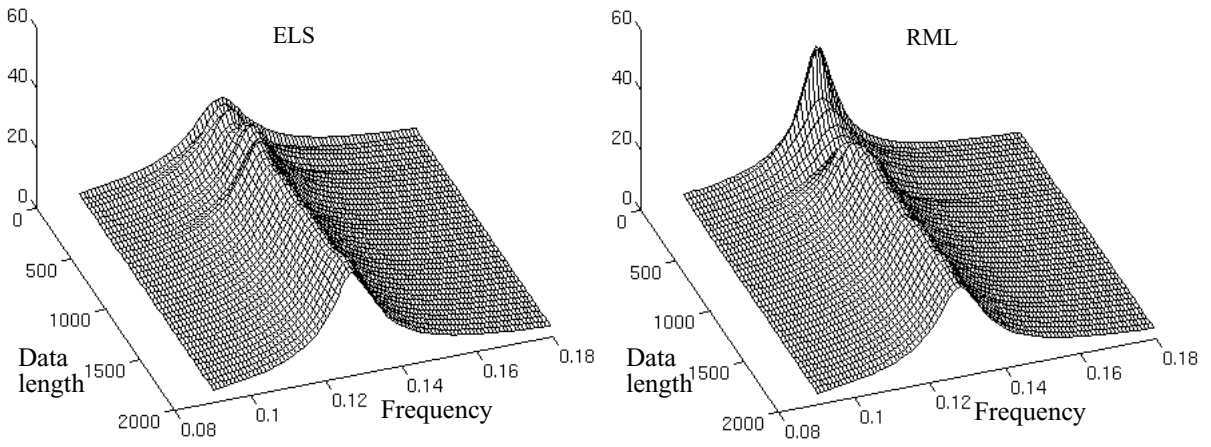


Figure 9 - Transient spectral densities obtained with finitely recursive algorithms (example 2).

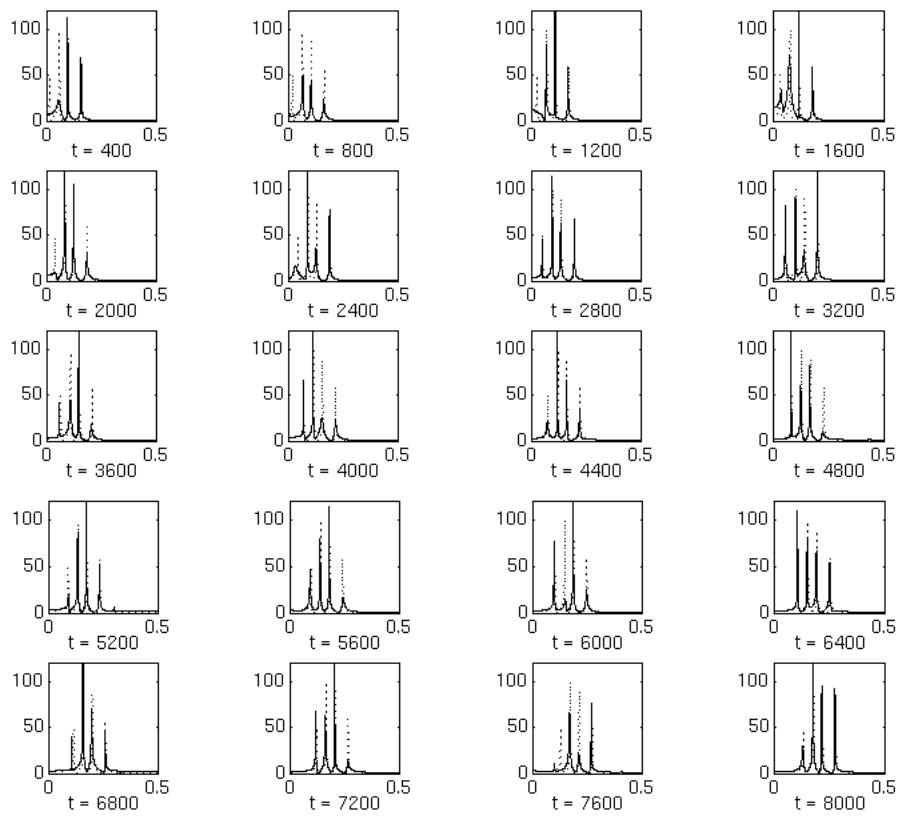


Figure 10 - Non stationary situation (example 3).

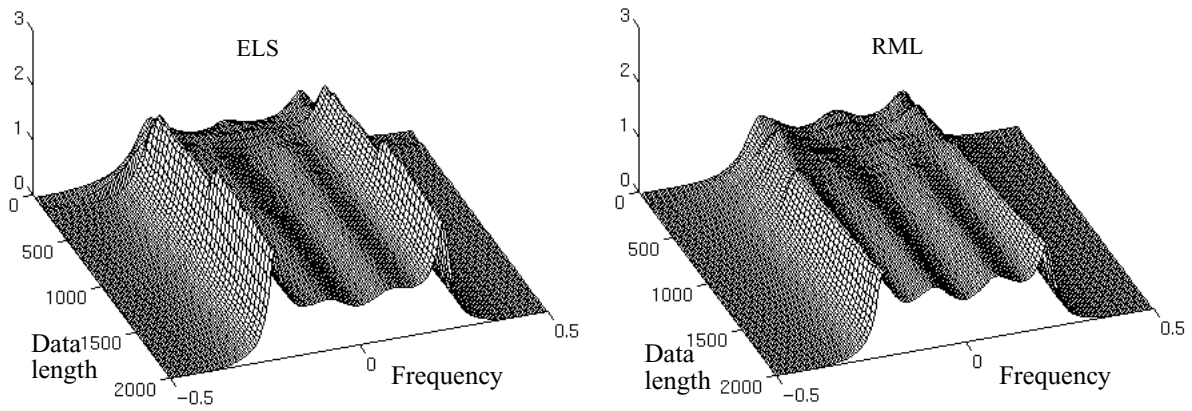


Figure 11 - The low-pass filter approximations with ELS and RML

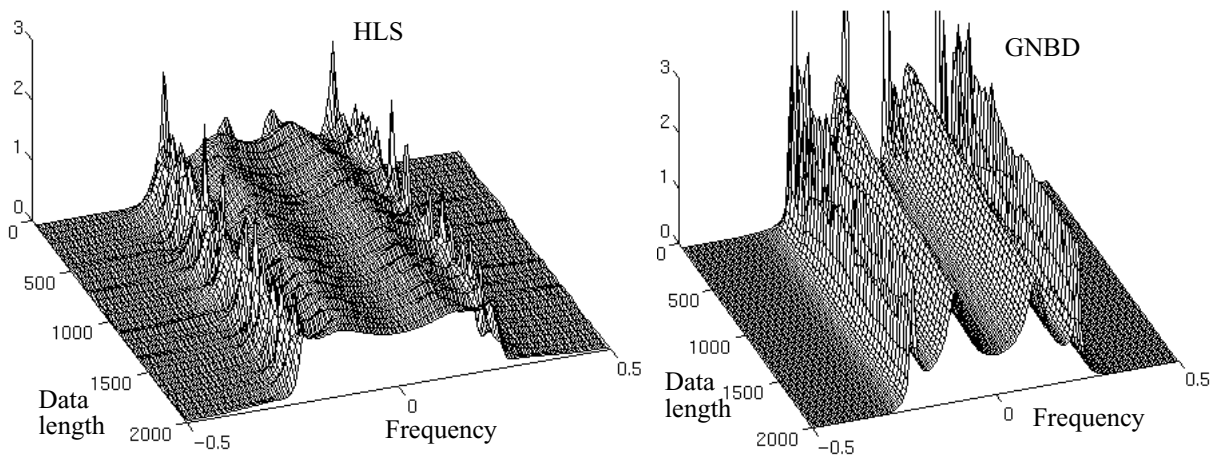


Figure 12 - The low-pass filter approximations with HLS and GNBD.

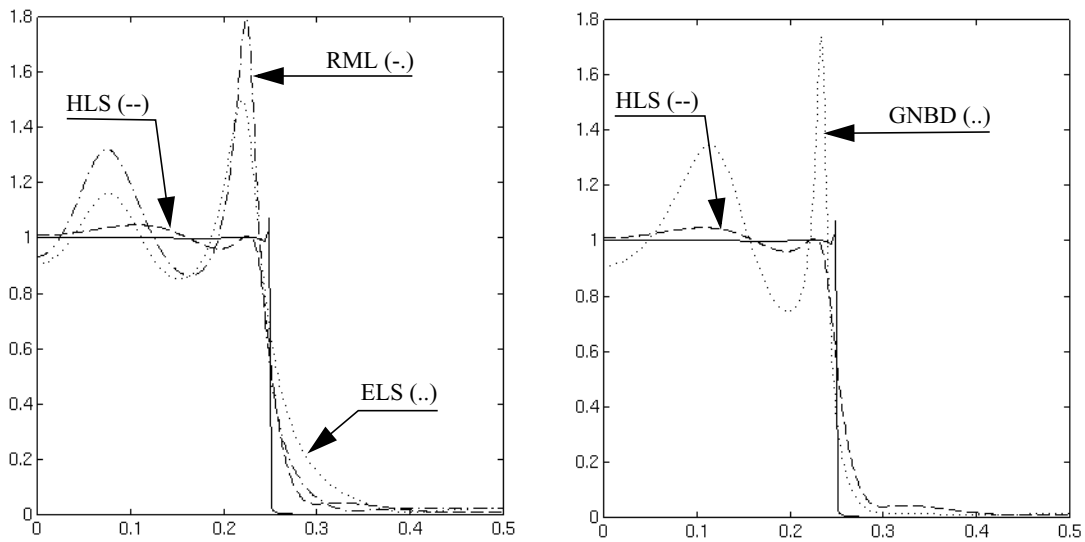


Figure 13 - Final results after 2000 steps.

List of figure/table

Figure 1 - The normalized lattice structure

Figure 2- The hereditary algorithm.

Figure 3 - Poles (x) and zeros (o) location (a) and power spectrum (b) (example 1).

Figure 4 - Spectral density obtained with classical finitely recursive identification algorithms (example 1)

Figure 5 - Transient spectral densities obtained with Gauss-Newton and hereditary algorithms (example 1).

Figure 6 - Power spectra obtained for a 150 long sequence (example 1).

Figure 7 - Poles (x) and zeros (o) location (a) and power spectrum (b) (example 2).

Figure 8 - Transient spectral densities obtained with Gauss-Newton and hereditary algorithms (example 2).

Figure 9 - Transient spectral densities obtained with finitely recursive algorithms (example 2).

Figure 10 - Non stationary situation (example 3).

Figure 11 - The low-pass filter approximations with ELS and RML

Figure 12 - The low-pass filter approximations with HLS and GNBD.

Figure 13 - Final results after 2000 steps.

Biographies

André Monin was born in Le Creusot, France in 1958. He graduated from the Ecole Nationale Supérieure d'Ingénieurs Electriciens de Grenoble in 1980. From 1981 to 1983, he was teaching assistant in the Ecole Normale Supérieure de Marrakech, Morocco. Since 1985, he has been with the Laboratoire d'Automatique et d'Analyse des Systèmes of the Centre Nationale de la Recherche Scientifique, France, as "Chargé de Recherche" from 1989.

After some work on non linear systems representation for his "Thèse de l' Université Paul Sabatier", Toulouse, France, 1987, his interests moved to the area of non linear filtering, sys-

tems realization and identification.

Gérard Salut was born in 1947 in Toulouse, France. He graduated as an Engineer from the Institut des Sciences Appliquées, Toulouse, in 1968. After a stay as a Research Associate with the Institute of Computer Science of the University of London, he was "Chargé de Mission" for the Scientific Affairs Department of the French Embassy in London, up to 1972. He obtained the degree of Docteur d'Etat from the University of Toulouse in 1976, and the "Médaille de Bronze" of CNRS in 1977. From 1977 to 1978, he was NSF-CNRS Exchange Fellow at the Electronics Research Laboratory of the University of California, Berkeley. He is now Director of Research at the Laboratoire d'Automatique et d'Analyse des Systèmes of CNRS, in Toulouse, France. His areas of research include non-linear signal processing, identification and model reduction, as well as particle resolution of filtering and optimization problems, with an emphasis on RADAR, SONAR and GPS applications.