The Annals of Statistics
1999, Vol. 27, No. 5, 1684–1705

SPEED OF CONVERGENCE FOR THE BLIND DECONVOLUTION OF A LINEAR SYSTEM WITH DISCRETE RANDOM INPUT

BY E. GASSIAT AND E. GAUTHERAT

Université Paris–Sud and Université de Reims

In a recent paper, we proposed a new estimation method for the blind deconvolution of a linear system with discrete random input, when the observations may be noise perturbed. We give here asymptotic properties of the estimators in the parametric situation. With nonnoisy observations, the speed of convergence is governed by the l_1-tail of the inverse filter, which may have an exponential decrease. With noisy observations, the estimator satisfies a limit theorem with known distribution, which allows for the construction of confidence regions. To our knowledge, this is the first precise asymptotic result in the noisy blind deconvolution problem with an unknown level of noise. We also extend results concerning Hankel's estimation to Toeplitz's estimation and prove a formula to compute Toeplitz forms that may have interest in itself.

1. Introduction. Let us consider an observed sequence $(Y_k)_{k\in\mathbf{Z}}$ which is the output of an unknown linear time-invariant system $\mathscr{U}$ with impulse response $(u_k)_{k\in\mathbf{Z}}$ that is driven by an unobservable input sequence $(X_k)_{k\in\mathbf{Z}}$, corrupted or not with additive noise $(\sigma_0\varepsilon_k)_{k\in\mathbf{Z}}$ where the level σ_0 is either 0, or known, or unknown,

$$Y_k = \sum_{j\in\mathbf{Z}} u_j X_{k-j} + \sigma_0\varepsilon_k. \tag{1.1}$$

The linear system $u = (u_j)_{j\in\mathbf{Z}}$ is invertible; $\theta = (\theta_k)_{k\in\mathbf{Z}}$ is the inverse filter of u, that is,

$$\sum_j \theta_j u_{k-j} = \delta_k, \qquad k \in \mathbf{Z},$$

in which δ_k denotes the Kronecker symbol. When u has finite length, the system is a noisy moving-average (MA), when θ has finite length, the system is autoregressive (AR).

The estimation of the parameters of the linear system u without observing the variables (X_t) is known as "blind identification" and has a long history. The case in which the input signal is discretely distributed has received considerable interest over the past few years. Indeed, this happens to be the situation in important applications in digital signal processing (cellular telephone, high-definition satellite,...) and it appears that, in the noiseless case, it is possible to propose estimators that converge very fast; see Li (1995), Sashadri (1994), van der Veen, Talwar and Paulraj (1997), Gamboa and Gassiat (1996, 1997).

Received January 1997; revised Februry 1999.
AMS 1991 *subject classifications*. Primary 62G05; secondary 62M09.
Key words and phrases. Deconvolution, contrast function, Hankel matrix, discrete linear systems.

All these authors (except the last ones) assume the knowledge of the finite alphabet in which the input series take value. In the other direction, there are very few results concerning noisy observations. Li (1993) gives, for a known level of noise σ_0 and an AR(p) linear system, an estimator which is proven to be $\sqrt{n}$-consistent, but the asymptotic distribution is not given. Other papers concern numerical strategies to find presumably consistent estimators. Liu and Chen (1995) propose an algorithm to compute a Bayesian estimator, Cappé, Doucet, Lavielle and Moulines (1999) survey stochastic EM algorithms to approach the maximum likelihood estimate, but no consistency results are given. For noisy MA processes, Bermond and Kéribin (1998) give asymptotic results and numerical issues for a split-likelihood estimator. The consistency of maximum-likelihood estimators and their asymptotic distribution is still an open problem in the general case. In particular MA cases, (1.1) may be modelled as a hidden Markov model (HMM), asymptotic results of Bickel, Ritov and Ryden (1998) may be used, numerical strategies are proposed, for instance, in Anton-Haro, Fonollosa and Fonollosa (1997).

In a recent paper [Gamboa and Gassiat (1996)], one of the authors proposed a new method for the estimation of a linear filter when the input series takes value in an unknown finite alphabet with known cardinality and when the filtered output is noiseless. To solve the problem of blind identification, they apply an adjustable linear time-invariant system $\mathscr{S}$: $s = (s_k)_{k\in\mathbf{Z}}$ to the output $(Y_k)_{k\in\mathbf{Z}}$ and work on the sequence $(Z(s)_k)_{k\in\mathbf{Z}}$,

$$Z(s)_k = \sum_j s_j Y_{k-j}. \tag{1.2}$$

In the absence of noise, the sequence $(Z(s)_k)_{k\in\mathbf{Z}}$ is the result of the linear system $\mathscr{S} * \mathscr{U}$ applied to the input X. The estimator relies on the quantification of the fact that a variable is discrete, after noticing that $Z(s)_1$ takes at most the same number of values as the input does iff s equals θ up to scale and delay. This quantification may be made, for instance, by using the Hankel matrix of the first algebraic moments or the Toeplitz matrix of the trigonometric moments of the empirical distribution. We further refined the method [Gassiat and Gautherat (1998)] to take noise into account. The advantages of our estimator are the following: the estimated filter converges to the unknown filter whatever the signal to noise ratio is; the assumptions on the input series are very weak, the variables may be dependent; the precise values of the alphabet need not to be known; this may be of importance for noncooperative digital communications when one's aim is to discover confidential communications; numerical implementation is easy.

In this paper, we give theoretical results on the speed of convergence both in the nonnoisy and in the noisy situation.

1. In the nonnoisy situation and when the filter is parametrized with a finite dimensional parameter, the speed of convergence is upper bounded by the l_1-tail of the inverse of the unknown filter; see Theorem 3.2. Though

the result is analogous to that of Li (1995), the proof takes advantage of discreteness also for empirical distributions of the input series.
2. In the noisy situation, we prove a central limit theorem which allows the construction of asymptotic confidence regions; see Theorem 4.2. This holds for parametric filters that may have infinite length and the inverse of infinite length. To our knowledge, this is the first result of this kind for such models.

In the next section, we shall briefly recall the general estimation procedure proposed by Gamboa and Gassiat (1996) when there is no noise and the refinement to deal with the presence of noise proposed by Gassiat and Gautherat (1998), together with their convergence theorems. We also get new results when using Toeplitz forms in the estimation procedure. We refer interested readers to previous papers for details and explanations of the procedures. Subsequent sections give the asymptotic speed of convergence in the parametric case when there is no noise and the asymptotic distribution in the noisy situation. Numerical experiments to illustrate these theoretical results may be found in Gassiat and Gautherat (1998). Technical parts of the proofs are given in the last section.

2. Estimation procedures and previous results. We make the following assumptions on model (1.1):

(M1) The input signal consists of discrete real random variables X_k with unknown common support $A := \{x_1, \dots, x_p\}$ of known cardinality p.

(M2) $U(x) := \sum_k u_k e^{ikx}$ is a continuous function that does not vanish on $[0, 2\pi]$.

(M3) $X = (X_k)_{k\in\mathbf{Z}}$ is a stationary ergodic process.

(M4) For any integer n and for any integers $j_1, \dots, j_n$ in $\{1, \dots, p\}$,

$$P(X_1 = x_{j_1}, \dots, X_n = x_{j_n}) > 0.$$

(M5) $\varepsilon = (\varepsilon_k)_{k\in\mathbf{Z}}$ is a sequence of i.i.d. Gaussian variables which are independent of the input signal; σ_0 is unknown; $E(\varepsilon_1) = 0$; $E(\varepsilon_1^2) = 1$.

The Gaussian distribution for the noise ε has been chosen for the sake of simplicity. However, all the probabilistic results of the section remain true with a noise of the form $\sigma_0\eta_k$, when the scale σ_0 is unknown, and η_k has an infinitely divisible distribution of class L; see Petrov (1975).

Under (M1), (M2) and (M4), it is proved by Gamboa and Gassiat (1996) that the variable $Z(s)_1$ takes at most p different values if and only if s equals θ up to scale and delay, that is, iff there exists $r \in \mathbf{R}$, $\ell \in \mathbf{Z}$ such that $\forall\, k \in \mathbf{Z},\ s_k = r\theta_{k-\ell}$. We thus define the parameter space Θ as a subset of $l_1(\mathbf{Z})$ which is unambiguous in scale and delay, that is, so that whenever two elements s and s' in Θ satisfy $\forall\, k \in \mathbf{Z},\ s_k = rs'_{k-\ell}$, then $r = 1$ and $\ell = 0$.

The procedure relies on a function that discriminates between variables having at most p points of support and the others, using only a finite number of moments of the variable. Let $\Phi = (1, \Phi_1, \dots, \Phi_{2p})$ be a set of real (or complex) functions. Let h be a continuous real function on $\mathbf{R}^{2p+1}$ (or $\mathbf{C}^{2p+1}$). Here h is said to discriminate between variables having at most p points of support and the others if the following holds. For any real random variable Z, define $c_Z = E[\Phi(Z)]$. Then $h(c_Z) \geq 0$, and $h(c_Z) = 0$ if and only if Z has at most p points of support. Examples of such functions h may be found in Gamboa and Gassiat (1996). In this paper, we shall use the following ones:

1. *Hankel forms.* Here h is the determinant of the Hankel matrix of the algebraic moments, that is, $\Phi_j(x) = x^j$, $j = 0, \dots, 2p$, and for any $c = (1, c^1, \dots, c^{2p})$ in $\{1\} \times \mathbf{R}^{2p}$, $h(c) = \det[M]$ where M is the $(p+1) \times (p+1)$ Hankel matrix given by $M_{i,j} = c^{i+j-2}$, $i, j = 1, \dots, p+1$.
2. *Toeplitz forms.* Here h is the determinant of the Toeplitz matrix of the trigonometric moments. In this case we assume

   (T) $2\pi \|u\|_1 \sup_{s \in \Theta} \|s\|_1 |\max_j x_j - \min_j x_j|$ is known to be strictly upper bounded by F.

   Here $\Phi_j(x) = \exp(\frac{i}{F}(j - p - 1)x)$, $j = 1, \dots, 2p$, and for any $c = (1, c^1, \dots, c^{2p})$ in $\{1\} \times \mathbf{C}^{2p}$, $h(c) = \det[T]$ where T is the $(p+1) \times (p+1)$ Toeplitz matrix given by $T_{i,j} = c^{i-j}$, $i, j = 1, \dots, p+1$.

For any filter s, define

$$c(s) = (c^i(s))_{i=1,\dots,2p} = \left(E\left[\Phi_i(Z(s)_1)\right]\right)_{i=1,\dots,2p}. \tag{2.1}$$

We then define a contrast function H by

$$H(s) = h(c(s)), \qquad s \in \Theta. \tag{2.2}$$

Obviously, for any $s \in \Theta$, $H(s) \geq 0$, and $H(s) = 0$ if and only if $s = \theta$. The sequence H_n is defined as an empirical contrast function in the following way. To use only the observations $Y_1, \dots, Y_n$, we need to truncate the filter s.

Let $k(n)$ be an increasing sequence of integers. Define

$$\hat{Z}(s)_t = \sum_{k=-k(n)}^{+k(n)} s_k Y_{t-k}$$

for $t = 1 + k(n), \dots, n - k(n)$, and

$$c_n(s) := \frac{1}{n - 2k(n)} \sum_{t=1+k(n)}^{n-k(n)} \Phi\left(\hat{Z}(s)_t\right).$$

We may now define

$$H_n(s) := h(c_n(s)).$$

DEFINITION 2.1. The estimator $\hat{\theta}$ is any minimizer of H_n over Θ_n where

$$\Theta_n = \Theta \cap \{s\colon s_k = 0 \text{ for } |k| > k(n)\}.$$

We assume throughout the sequel that

$$\lim_{n \leftarrow +\infty} k(n) = +\infty \quad \text{and} \quad \lim_{n \leftarrow +\infty} \frac{k(n)}{n} = 0.$$

The following theorems were proved by Gamboa and Gassiat (1996).

THEOREM 2.2. *Assume that* (M1), (M2), (M3), (M4) *hold and that* $\sigma_0 = 0$. *If* Θ *is compact, then* $\hat{\theta}$ *converges almost surely, in* $l^1(\mathbf{Z})$, *to* θ *as* n *tends to infinity.*

Suppose that the set Θ can be represented as a parametric model with real-valued parameter vector ξ in a compact set $\mathscr{K}$ of dimension q, $\xi = (\xi_j)_{j=1\dots q}$,

$$\Theta := \{\theta(\xi),\ \xi \in \mathscr{K}\}.$$

Let ξ^* be the true parameter value. To estimate ξ^*, we minimize $L_n(\xi) := H_n(\theta(\xi))$. Let $\hat{\xi}$ be any minimizer of L_n over $\mathscr{K}$. Assume the identifiability,

$$\theta_k(\xi) = r\theta_{k-\ell}(\xi'), \qquad \forall k \in \mathbf{Z} \quad \Longleftrightarrow \quad r = 1,\ \ell = 0 \text{ and } \xi = \xi'.$$

THEOREM 2.3. *Assume that the application* $\xi \to \theta(\xi)$ *from* $\mathbf{R}^q$ *to* $l^1(\mathbf{Z})$ *is continuous, that assumptions* (M1) *to* (M4) *hold, and that* $\sigma_0 = 0$.
Then, $\hat{\xi}$ *converges, almost surely, as* n *approaches infinity, to* ξ^*.

Using Hankel forms or Toeplitz forms, the method extends to noisy observations.

Hankel forms. Define $M(s, \sigma)$ as the $(p+1) \times (p+1)$ Hankel matrix built using the solutions $c^j(s, \sigma)$ of the triangular system ($M(s, \sigma)_{i,j} = c^{i+j-2}(s, \sigma)$),

$$E(Z(s)_1^j) = \sum_{i=0}^{j} C_j^i c^i(s, \sigma) v(s, \sigma)^{j-i} \mu_{j-i}, \qquad j = 0, \dots, 2p, \tag{2.3}$$

where $C_j^i = i!(j-i)!/j!$ is the binomial coefficient, $v^2(s, \sigma) = \sigma^2 \|s\|_2^2$ and μ_{j-i} is the $j-i$th moment of the standard Gaussian distribution. Define the function $H(s, \sigma)$ of the filter and the noise level as the value of the determinant of $M(s, \sigma)$. Define the estimators $c_n^j(s, \sigma)$ of the pseudo-moments $c^j(s, \sigma)$ as the solutions of the triangular system

$$c_n^j(s) = \sum_{i=0}^{j} C_j^i c_n^i(s, \sigma)(\sigma \|s\|_2)^{j-i} \mu_{j-i}, \qquad j = 1, \dots, 2p.$$

Let $M_n(s, \sigma)$ be the Hankel matrix built using the $c_n^j(s, \sigma)$, and let $H_n(s, \sigma)$ be the estimator of the function H,

$$H_n(s, \sigma) = \det[M_n(s, \sigma)].$$

Toeplitz forms. Define $T(s,\sigma)$ as the $(p+1)\times(p+1)$-Toeplitz matrix built using the $c^j(s,\sigma)$ $(T(s,\sigma)_{i,j}=c^{i-j}(s,\sigma))$,

$$E\biggl(\exp\biggl(\frac{ij}{F}Z(s)_1\biggr)\biggr)=c^j(s,\sigma)\exp\biggl(-\frac{j^2\sigma^2\|s\|_2^2}{2F^2}\biggr),\qquad j=-p,\dots,p. \tag{2.4}$$

Define the function $H(s,\sigma)$ of the filter and the noise level as the value of the determinant of $T(s,\sigma)$. Define the estimators $c_n^j(s,\sigma)$ of the pseudo-moments $c^j(s,\sigma)$ by

$$c_n^j(s,\sigma)=c_n^j(s)\exp\biggl(\frac{j^2\sigma^2\|s\|_2^2}{2F^2}\biggr),\qquad j=-p,\dots,p.$$

Let $T_n(s,\sigma)$ be the Toeplitz matrix built using the $c_n^j(s,\sigma)$, and let $H_n(s,\sigma)$ be the estimator of the function H,

$$H_n(s,\sigma)=\det[T_n(s,\sigma)].$$

Let $\delta(n)$ be a sequence of positive real numbers with limit 0 as n tends to infinity. Define in both cases:

$$J_n(s,\sigma)=(H_n(s,\sigma))^2+(\delta(n))^2\sigma. \tag{2.5}$$

We set the following definition.

DEFINITION 2.4. The estimator $(\widehat{\theta},\widehat{\sigma})$ is any minimizer of J_n over $\Theta\times\mathbf{R}^+$.

To have good asymptotic behavior of the estimator, the speed $\delta(n)$ has to be related to the stochastic variation of the empirical moments and to the truncation parameter $k(n)$. We then need a slightly stronger assumption concerning the following processes.

ASSUMPTION (M6). Assume that

$$\sum_{|k|>k(n)}|\theta_k|=o(\delta(n))$$

and that

$$\lim_{n\to\infty}(\delta(n))^{-1}\frac{1}{n}\sum_{t=1}^{n}\biggl(\Phi_j[X_t+\sum_k\theta_k\varepsilon_{t-k}]-m_j(\theta)\biggr)=0$$

in probability for $j=1,\dots,2p$, where $m_j(\theta)=E[\Phi_j(X_t+\sum_k\theta_k\varepsilon_{t-k})]$.

In Gassiat and Gautherat (1998), a convergence theorem was proved for the estimator minimizing $|H_n(s,\sigma)|+\delta(n)\sigma$. Since we shall derive the speed of convergence via a Taylor expansion, we chose here to work with $(H_n(s,\sigma))^2$. Following the same lines, we easily have the theorem.

THEOREM 2.5. *Assume that* (M1) – (M6) *hold. When using Toeplitz forms, assume in addition* (T). *Then, as n tends to infinity,* $\widehat{\theta}$ *converges in* l_1 *in probability to* θ *and* $\widehat{\sigma}$ *converges in probability to* σ_0.

An immediate corollary of this theorem is that the method leads to consistent estimation in the parametric case. With the same assumptions as in Theorem 2.3, with (M6) [and (T) if Toeplitz forms are used], define the estimator $(\widehat{\xi}, \widehat{\sigma})$ as the minimizer of $J_n(\theta(\xi), \sigma)$ over $\mathscr{K} \times \mathbf{R}^+$. Then we have the corollary.

COROLLARY 2.6. *$(\widehat{\xi}, \widehat{\sigma})$ converges in probability towards the true value (ξ^*, σ_0) of the parameter.*

We shall also recall a useful formula given by Lindsay (1989), which gives the value $M(W)$ of the determinant of the Hankel matrix based on the first $2p$ algebraic moments of a random variable W.

PROPOSITION 2.7. *Let $W_0, \dots, W_p$ be $p+1$ independent copies of W. We have*

$$M(W) = \frac{1}{(p+1)!} E\left[\prod_{i<j}\left(W_i - W_j\right)^2\right].$$

A similar formula holds for the value $T(W)$ of the determinant of the Toeplitz matrix based on the first p complex exponential moments of a random variable W. Its proof follows Lindsay's (1989) and will be omitted.

PROPOSITION 2.8. *Let $W_0, \dots, W_p$ be $p+1$ independent copies of W. We have*

$$T(W) = \frac{2^{p(p+1)/2}}{(p+1)!} E \prod_{j<k}\left(1 - \cos\left(\frac{W_j - W_k}{F}\right)\right)$$

REMARKS.

(i) When p is not small, algebraic moments have large asymptotic variance, and Toeplitz forms should be preferred after a rough estimate of F.

(ii) The numerical implementation of the method is easy. It requires only a minimization procedure of a function which is computed using only empirical moments of the observations. There is no need of simulations as required by the Bayesian methods [van der Veen, Talwar and Paulraj (1997)] or stochastic EM algorithms [Cappé, Doucet, Lavielle and Moulines (1999)]. Like the other methods, the convergence is sensitive to the choice of the initial guess; a data driven method for the choice of the initial values and for the choice of the penalization term $\delta(n)$ is given by Gassiat and Gautherat (1998).

From now on, the parameter will be $\xi = (\xi_j)_{j=1\dots q}$ in $\mathscr{K}$ so that $\Theta := \{\theta(\xi),\ \xi \in \mathscr{K}\}$, with the associated identifiability and continuity assumptions.

3. Speed of convergence: nonnoisy observations. Throughout this section, the level noise σ_0 is set at 0. The notation $D_x^r F(y)$ will designate the rth derivative of F with respect to the variable x and will be evaluated at point y.

Let Θ^* be the set of all elements of Θ except θ. Let us introduce the assumptions:

(D) The functions $h(\cdot)$ and $\Phi(\cdot)$ are twice continuously differentiable. Let $D_s^2 H(s) = ((\partial^2/\partial s_k \partial s_l)H(s))_{k,l\in\mathbf{Z}}$. Then $D_s^2 H(\theta)$ is positive definite on the set Θ^*.

(P) The application $\xi \to \theta(\xi)$ is twice continuously differentiable. For any $i = 1, \dots, q$, $(\partial\theta_k/\partial\xi_i)_{k\in\mathbf{Z}}$ and $(\partial^2\theta_k/\partial\xi_i^2)_{k\in\mathbf{Z}}$ are in $l_1(\mathbf{Z})$. Moreover, $((\partial\theta_k/\partial\xi_1)(\xi^*))_{k\in\mathbf{Z}}, \dots, (\partial\theta_k/\partial\xi_q(\xi^*))_{k\in\mathbf{Z}}$ and $(\theta_k)_{k\in\mathbf{Z}}$ are linearly independent.

It should be seen that $H(\cdot)$ and H_n are twice continuously differentiable and that since θ is a minimizer of H, the operator $D_s^2 H(\theta)$ is necessarily nonnegative. The assumption concerns the definiteness. Notice also that the gradient operator $D_s^1 H$ of H is in L_∞ with no more assumptions. In case H is the Hankel form, h is a multipolynomial and c is an algebraic power, so that they are twice continuously differentiable. It will be later proved that in this case $D_s^2 H(\theta)$ is positive definite on Θ^*, so that (D) holds with no particular assumption. In case H is the Toeplitz form, h is a multipolynomial and c is a complex exponential, so that they are twice continuously differentiable. It will also be later proved that in this case $D_s^2 H(\theta)$ is positive definite on Θ^*, so that again (D) holds with no particular assumption.

To estimate ξ^*, we minimize $L_n(\xi)$ as described in Section 2. Let $L(\xi) = H(\theta(\xi))$. A useful result will be the following.

PROPOSITION 3.1. *Under* (P) *and* (D), *the functions* $L(\xi)$ *and* $L_n(\xi)$ *are twice continuously differentiable. Let* $D_\xi^2 L(\xi) = (\partial^2 L(\xi)/\partial\xi_k \partial\xi_l)_{k,\,l=1,\dots,q}$. *Then* $D_\xi^2 L(\xi^*)$ *is positive definite on* $\mathscr{S} - \xi^*$.

The main result of this section follows.

THEOREM 3.2. *Assume that* (M1), (M2), (M3), (M4), (P) *hold. If the estimation method is not Hankel or Toeplitz, assume moreover* (D). *We have almost surely for big enough* n,

$$\|\hat{\xi} - \xi^*\|_2 \le C \sum_{|k|>k(n)} |\theta_k(\xi^*)|,$$

where C *is a constant.*

REMARKS.

(i) The constant C is explicitly given in the proof; it may be estimated by a plug-in method after having estimated the distribution of X_1 (i.e., the set of possible values and the respective weights); this may be done with Hankel or Toeplitz forms. See Gautherat (1997).

(ii) Usually, empirically based estimators are related to the speed of convergence of the empirical functions to the expectation of the functions, so that $\sqrt{n}$ speed of convergence is obtained. The key idea here will be to relate not to the expectation of the functions but to the nontruncated empirical moments and to notice that, at the true parameter value, the random variable is discrete, and its empirical distribution is also that of a discrete variable. Though the method may look like Li's (1995), this last point is the main different idea that exploits the discreteness assumption.

PROOF OF THEOREM 3.2. Using Theorem 2.3, $\hat{\xi}$ is almost surely consistent. So that for big enough n, and for $i = 1, \dots, q$,

$$\frac{\partial}{\partial \xi_i} L_n(\hat{\xi}) = 0.$$

Using a first order Taylor expansion, we obtain, for all $i = 1, \dots, q$,

$$0 = \frac{\partial}{\partial \xi_i} L_n(\xi^*) + \sum_{j=1}^{q} (\hat{\xi}_j - \xi_j^*) \frac{\partial^2}{\partial \xi_i \partial \xi_j} L_n(\tilde{\xi}^j), \tag{3.1}$$

where $\tilde{\xi}^j \in [\hat{\xi}, \xi^*)$. Define for any filter s the empirical moments for the non truncated series

$$\tilde{c}_n(s) = \frac{1}{n - 2k(n)} \sum_{t=1+k(n)}^{n-k(n)} \Phi\big(Z(s)_t\big)$$

and also

$$\tilde{H}_n(s) = h(\tilde{c}_n(s)), \qquad \tilde{L}_n(\xi) = \tilde{H}_n(\theta(\xi)).$$

Since $Z(\theta(\xi^*))_t = X_t$ for all t, $\tilde{c}_n(\theta(\xi^*))$ is the expectation of a random variable taking at most p distinct values, so that $\theta(\xi^*)$ is a minimum point of $\tilde{H}_n$. Now

$$\frac{\partial}{\partial \xi_i} L_n(\xi^*) = \sum_{|k| \le k(n)} \frac{\partial}{\partial s_k} H_n(\theta(\xi^*)) \frac{\partial}{\partial \xi_i} \theta_k(\xi^*)$$

since H_n depends only on s_k for $|k| \le k(n)$. Using the previous remark,

$$\frac{\partial}{\partial s_k} \tilde{H}_n(\theta(\xi^*)) = 0$$

for all k, and then,

$$\frac{\partial}{\partial \xi_i} L_n(\xi^*) = \sum_{|k| \le k(n)} \left(\frac{\partial}{\partial s_k} H_n(\theta(\xi^*)) - \frac{\partial}{\partial s_k} \tilde{H}_n(\theta(\xi^*)) \right) \frac{\partial}{\partial \xi_i} \theta_k(\xi^*). \tag{3.2}$$

This will allow us to prove (see Section 5)

$$\left| \frac{\partial}{\partial \xi_i} L_n(\xi^*) \right| \le \tilde{C} \sum_{|k| > k(n)} |\theta_k| \tag{3.3}$$

for some constant $\tilde{C}$. Now, for any $i, j = 1, \dots, q$ we have

$$\frac{\partial^2}{\partial\xi_i\partial\xi_j}L_n(\tilde{\xi}^j) = \frac{\partial^2}{\partial\xi_i\partial\xi_j}L(\xi^*) + \left(\frac{\partial^2}{\partial\xi_i\partial\xi_j}L_n(\tilde{\xi}^j) - \frac{\partial^2}{\partial\xi_i\partial\xi_j}L(\xi^*)\right).$$

It is easily proved, using the ergodicity of the process (X_t) and the fact that it is bounded, that

$$\left(\frac{\partial^2}{\partial\xi_i\partial\xi_j}L_n(\tilde{\xi}^j) - \frac{\partial^2}{\partial\xi_i\partial\xi_j}L(\xi^*)\right)$$

tends to 0 a.s. and uniformly in i and j since there are a finite number of them.

Now, using Proposition 3.1,

$$\sum_{i,j=1,\dots,q}\frac{\partial^2}{\partial\xi_i\partial\xi_j}L(\xi^*)(\hat{\xi}_i - \xi_i^*)(\hat{\xi}_j - \xi_j^*) \geq \lambda\|\hat{\xi} - \xi^*\|_2^2, \tag{3.4}$$

where λ is the smallest eigenvalue of $D_2L(\xi^*)$.

The theorem follows using (3.1), (3.3) and (3.4) and with $C = (\tilde{C}/\lambda)\sqrt{q}$. □

Let us study the second derivative operator of H with respect to s when it is the Hankel or the Toeplitz form.

Let $(X_t^0)_{t\in\mathbf{Z}}, \dots, (X_t^p)_{t\in\mathbf{Z}}$ be $p+1$ independent copies of $(X_t)_{t\in\mathbf{Z}}$. Define for $i = 0, \dots, p$ and $t \in \mathbf{Z}$,

$$Y_t^i = \sum_{k\in\mathbf{Z}} u_k X_{t-k}^i.$$

$(Y_t^0)_{t\in\mathbf{Z}}, \dots, (Y_t^p)_{t\in\mathbf{Z}}$ are $p+1$ independent copies of $(Y_t)_{t\in\mathbf{Z}}$. In the same way, define for $i = 0, \dots, p$, $t \in \mathbf{Z}$ and any filter s,

$$Z^i(s)_t = \sum_{k\in\mathbf{Z}} s_k Y_{t-k}^i.$$

$(Z^0(s)_t)_{t\in\mathbf{Z}}, \dots, (Z^p(s)_t)_{t\in\mathbf{Z}}$ are $p+1$ independent copies of $(Z(s)_t)_{t\in\mathbf{Z}}$. We have the following result.

PROPOSITION 3.3. *For any filter $v = (v_k)_{k\in\mathbf{Z}}$, we have for the Hankel procedure,*

$$v^T D_s^2 H(\theta)v = \frac{2}{(p+1)!}\sum_{i<j} E\left((Z^i(v)_0 - Z^j(v)_0)\prod_{i'<j',\,(i',j')\neq(i,j)}(X_0^{i'} - X_0^{j'})\right)^2$$

and for the Toeplitz procedure,

$$v^T D_s^2 H(\theta) v = \frac{2^{p(p+1)/2}}{(p+1)!F^2} \sum_{i<j} E\Bigg((Z^i(v)_0 - Z^j(v)_0)^2 \times \prod_{i'<j',\,(i',\,j')\neq(i,j)} \Bigg(1 - \cos\bigg(\frac{X_0^{i'} - X_0^{j'}}{F}\bigg)\Bigg)\Bigg)$$

In particular, $D_s^2 H(\theta)$ *is positive definite on* Θ^* *under assumption* (M4).

To finish the study of $D_s^2 H(\theta)$, let us mention a result.

PROPOSITION 3.4. *For any filter* v, *and if the variables* X_t *are i.i.d., then*

$$v^T D_s^2 H(\theta) v = C_H \sum_{k\neq 0} (v * u)_k^2$$

with:

For the Hankel procedure,

$$C_H = \frac{2\mathrm{Var}(X_0)}{(p-1)!} E\Bigg(\prod_{i<j,\,(i,\,j)\neq(0,\,1)} (X_0^i - X_0^j)^2 \Bigg).$$

For the Toeplitz procedure,

$$C_H = \frac{2^{1+p(p+1)/2}\mathrm{Var}(X_0)}{(p-1)!F^2} E\Bigg(\prod_{i<j,\,(i,\,j)\neq(0,\,1)} \Bigg(1 - \cos\bigg(\frac{X_0^i - X_0^j}{F}\bigg)\Bigg)\Bigg).$$

Indeed, by easy computation and applying Proposition 3.3, we have for the Hankel procedure,

$$v^T D_s^2 H(\theta) v = \frac{1}{(p-1)!} \sum_{k,\,l\in\mathbf{Z}^*} (v*u)_k (v*u)_l \times E\Bigg[(X_{-k}^0 - X_{-k}^1)(X_{-l}^0 - X_{-l}^1) \prod_{i<j,\,(i,\,j)\neq(0,\,1)} (X_0^i - X_0^j)^2 \Bigg]$$

Now, if $k = 0$ or $l = 0$, $(X_{-k}^0 - X_{-k}^1)(X_{-l}^0 - X_{-l}^1) \prod_{i<j,\,(i,\,j)\neq(0,\,1)} (X_0^i - X_0^j)^2 = 0$ a.s. If the X_k^i are i.i.d., for $k \neq l \in \mathbf{Z}$, $E(X_{-k}^0 - X_{-k}^1)(X_{-l}^0 - X_{-l}^1) \prod_{i<j,\,(i,\,j)\neq(0,\,1)} (X_0^i - X_0^j)^2) = 0$. For $k = l \neq 0$, $E(X_{-k}^0 - X_{-k}^1)(X_{-l}^0 - X_{-l}^1) \prod_{i<j,\,(i,\,j)\neq(0,\,1)} (X_0^i - X_0^j)^2 = (p-1)!C_H$.

The same arguments lead to the result for the Toeplitz procedure.

4. Asymptotic distribution with noisy observations. Now, the level of noise σ is unknown, and the estimator $(\hat{\xi}, \hat{\sigma})$ minimizes

$$J_n(\xi, \sigma) = (H_n(\theta(\xi), \sigma))^2 + (\delta(n))^2 \sigma$$

defined in Section 2 for Hankel or Toeplitz forms. Define

$$\mu_j(\xi) = \Big(E(\Phi_j(Z_t(\theta(\xi)))) \Big)_{j=1,\dots,2p},$$

$$M_n(\xi) = \Big(\frac{1}{n} \sum_{t=1}^{n} (\Phi_j(Z_t(\theta(\xi)))) \Big)_{j=1,\dots,2p},$$

$$D^1_\xi M_n(\xi) = \Big(\Big(\frac{\partial}{\partial \xi_i} M_n(\xi) \Big) \Big)_{i=1,\dots,q}.$$

Let us introduce the following assumption:

$$\lim_{n\to\infty} \frac{k(n)}{\sqrt{n}} = 0, \qquad \lim_{n\to\infty} \sqrt{n} \sum_{|k|>k(n)} |\theta_k| = 0, \qquad \lim_{n\to\infty} \sqrt{n}\delta(n) = +\infty,$$

(M8)

$$\lim_{n\to\infty} \sqrt{n}\delta(n)^2 = 0.$$

The vector $\sqrt{n}(M_n(\xi^*) - m(\xi^*), D^1_\xi M_n(\xi^*) - D^1_\xi m(\xi^*))$ converges in distribution to $\mathscr{N}(0, \Gamma)$

Notice that under (M8), (M6) holds. Define $R(\theta(\xi), \sigma)$ the triangular matrix inverting the system (2.3) for the Hankel procedure, or inverting the system (2.4) for the Toeplitz procedure. In this last case, $R(\theta(\xi), \sigma)$ is a diagonal matrix. Notice that $R(\theta(\xi), \sigma)$ is differentiable with respect to ξ and with respect to σ. Let Γ_1 be the asymptotic variance of $\sqrt{n}(M_n(\xi^*) - m(\xi^*))$. Define

$$V = D^1_c h \cdot R \cdot \Gamma_1 \cdot R^T \cdot (D^1_c h)^T.$$

We now have the lemma.

LEMMA 4.1. *Under the assumptions of Theorem* 4.2 *below*,

$$(\sqrt{n} D^1_\xi H_n(\theta, \sigma_0), \sqrt{n} H_n(\theta, \sigma_0))$$

converges in distribution to a centered Gaussian distribution, $D^1_\sigma H_n(\theta, \sigma_0)$ *converges in probability to a negative constant and* $\sqrt{n} H_n(\theta, \sigma_0)$ *has asymptotic variance* V.

Define G the $q \times q$ matrix given by

$$G_{i,j} = \Big(\frac{\partial \theta}{\partial \xi_i}(\xi^*) \Big)^T D^2_s H(\theta) \Big(\frac{\partial \theta}{\partial \xi_j}(\xi^*) \Big).$$

We have the theorem.

THEOREM 4.2. *Assume that* (M1), (M2), (M3), (M4), (M5), (M8), (P) *hold. Then, as n tends to infinity,* $\sqrt{n}(\hat{\xi}-\xi^*)$ *converges in distribution to the centered Gaussian distribution with variance* Σ *given by*

$$\Sigma = \left(\frac{D^2_{\xi,\sigma}H_n(\theta,\sigma_0)}{D^1_\sigma H_n(\theta,\sigma_0)}\right)^2 G^{-1}V(G^{-1})^T.$$

Let us give an outline of the proof (the complete proof is given in Section 5). Using (P), J_n is twice continuously differentiable. Then $(\hat{\xi},\hat{\sigma})$ is a zero of D^1J_n and converges in probability to (ξ^*,σ_0) by Corollary 2.6, so that the following Taylor expansion holds:

$$D^1J_n(\xi^*,\sigma_0) + D^2J_n(\xi^*,\sigma_0)(\hat{\xi}-\xi^*;\hat{\sigma}-\sigma_0)^T(1+o(1)) = 0.$$

Direct computation, noticing that $H_n(\theta,\sigma_0)\neq 0$ (indeed, the pseudo-moments may not be moments of discrete random variables), leads to

$$\begin{aligned}
&\begin{pmatrix} D^1_\xi H_n(\theta,\sigma_0) \\ D^1_\sigma H_n(\theta,\sigma_0) + \dfrac{(\delta(n))^2}{2H_n(\theta,\sigma_0)} \end{pmatrix} \\
&\quad + \left(D^2H_n(\theta,\sigma_0) + \frac{D^1H_n(\theta,\sigma_0)D^1H_n(\theta,\sigma_0)^T}{H_n(\theta,\sigma_0)}\right)\begin{pmatrix}\hat{\xi}-\xi^* \\ \hat{\sigma}-\sigma_0\end{pmatrix} \\
&\quad \times(1+o(1)) = 0.
\end{aligned} \tag{4.1}$$

Roughly speaking, the asymptotic result comes from the fact that asymptotically, the matrix involved in equation (4.1) has the bottom-right term tending to infinity, so that asymptotically, the inverse has only the up-left term as a nonzero term.

REMARKS.

(i) Hankel or Toeplitz forms may be used to obtain estimators of the distribution of X_1 (possible values with corresponding weights); see Gautherat (1997). In case the X_t are i.i.d., this also allows us to estimate Γ. All other matrices and constants may be estimated by a plug-in method, and thus Σ may be consistently estimated. This allows us to construct a confidence region for ξ^*.

(ii) The estimator is not efficient. In general, roughly speaking, a lower bound for the asymptotic variance is given via the maximum likelihood estimator (m.l.e.) Asymptotic general results for the m.l.e. are still open problems; after the first submission of the paper, the result of Bickel, Ritov and Ryden (1998) for the m.l.e. in HMMs allow one to obtain a central limit theorem for the m.l.e. for noisy MA processes with discrete inputs. However, the asymptotic variance (and Fisher information) is defined as the asymptotic variance of the derivative of conditional likelihoods and is not directly computed.

5. Proofs.

PROOF OF PROPOSITION 3.1. Differentiability of L and L_n easily comes from that of H, H_n and θ. Now we have

$$D_\xi^2 L(\xi) = D_\xi^1\theta(\xi)^T D_s^2 H(\theta(\xi)) D_\xi^1\theta(\xi) + D_s^1 H(\theta(\xi)) D_\xi^2\theta(\xi).$$

Since H is minimum at point $\theta(\xi^*)$, $D_s^1 H(\theta(\xi^*)) = 0$. We then have

$$D_\xi^2 L(\xi^*) = D_\xi^1\theta(\xi^*)^T D_s^2 H(\theta(\xi^*)) D_\xi^1\theta(\xi^*).$$

Taking the associated quadratic form at some nonzero point $y = (\xi_i - \xi_i^*)_{i=1,\dots,q}$,

$$y^T D_\xi^2 L(\xi^*) y = (y D_\xi^1\theta(\xi^*))^T D_s^2 H(\theta(\xi^*))(y D_\xi^1\theta(\xi^*)),$$

which, using (D), is nonzero unless $y D_\xi^1\theta(\xi^*)$ is either the null series or the filter $\theta(\xi^*)$, which is impossible by (P). □

PROOF OF THEOREM 3.2. Formula (3.3) remains to be proved. Using (3.2), we have

$$\left|\frac{\partial}{\partial\xi_i} L_n(\xi^*)\right| \le A_n \left\|\frac{\partial}{\partial\xi_i}\theta(\xi^*)\right\|_1 \tag{5.1}$$

with

$$A_n = \sup_{|k|\le k(n)} \left|\frac{\partial}{\partial s_k} H_n(\theta(\xi^*)) - \frac{\partial}{\partial s_k}\tilde{H}_n(\theta(\xi^*))\right|.$$

By the chaining rule,

$$\frac{\partial}{\partial s_k} H_n(\theta) = D_c^1 h(c_n(\theta)) \frac{\partial}{\partial s_k} c_n(\theta) \quad \text{and} \quad \frac{\partial}{\partial s_k}\tilde{H}_n(\theta) = D_c^1 h(\tilde{c}_n(\theta)) \frac{\partial}{\partial s_k}\tilde{c}_n(\theta).$$

Notice that

$$\sup_i \sup_k \left|\frac{\partial}{\partial s_k} c_n^i(\theta)\right| \le C_1$$

with $C_1 = \|D^1\Phi\|_\infty \|X\|_\infty \|u\|_1$. Here $\|X\|_\infty$ is the maximum possible absolute value in the alphabet in which the variables X_t take value. The norms $\|\cdot\|_\infty$ for functions are taken as the supremum value of the function on the space where the possible moments and their derivatives take value, which are compact since the X_t are bounded and the filters are summable. Then

$$A_n \le 2pC_1 B_n + \|D_c^1 h\|_\infty E_n \tag{5.2}$$

with

$$B_n = \sup_i \left|D_c^1 h(c_n(\theta))_i - D_c^1 h(\tilde{c}_n(\theta))_i\right|$$

and

$$E_n = \left|\frac{\partial}{\partial s_k} c_n(\theta) - \frac{\partial}{\partial s_k}\tilde{c}_n(\theta)\right|.$$

However, easily,

$$B_n \leq \|D_c^2 h\|_\infty \|D^1\Phi\|_\infty \|X\|_\infty \|u\|_1 \sum_{|k|>k(n)} |\theta_k|, \tag{5.3}$$

and for any $|k| \leq k(n)$,

$$\sup_i \left| \frac{\partial}{\partial s_k} c_n^i(\theta) - \frac{\partial}{\partial s_k} \tilde{c}_n^i(\theta \right| \leq \|D^2\Phi\|_\infty \|u\|_1^2 \|X\|_\infty^2 \sum_{|k|>k(n)} |\theta_k|. \tag{5.4}$$

Taking into account (5.1), (5.2), (5.3) and (5.4), we finally obtain (3.3) with

$$\tilde{C} = \|u\|_1^2 \|X\|_\infty^2 \sup_i \left\| \left(\frac{\partial}{\partial \xi_i} \theta_k(\xi^*) \right) \right\|_1 (2p\|D^1\Phi\|_\infty^2 \|D_c^2 h\|_\infty + \|D^2\Phi\|_\infty \|D_c^1 h\|_\infty). \quad \square$$

PROOF OF PROPOSITION 3.3. Let us first study the Hankel procedure. Applying Proposition 2.7, we have

$$H(s) = \frac{1}{(p+1)!} E\left[\prod_{i<j} (Z^i(s)_0 - Z^j(s)_0)^2 \right]$$

so that

$$\frac{\partial}{\partial s_k} H(s) = \frac{2}{(p+1)!} E\Bigg(\sum_{i<j} (Y_{-k}^i - Y_{-k}^j)(Z^i(s)_0 - Z^j(s)_0) \times \prod_{i'<j',\,(i',j')\neq(i,j)} (Z^{i'}(s)_0 - Z^{j'}(s)_0)^2 \Bigg)$$

and

$$\begin{aligned} \frac{\partial^2}{\partial s_k \partial s_l} H(s) = {} & \frac{2}{(p+1)!} E\Bigg(\sum_{i<j} (Y_{-k}^i - Y_{-k}^j)(Y_{-l}^i - Y_{-l}^j) \\ & \qquad \times \prod_{i'<j',\,(i',j')\neq(i,j)} (Z^{i'}(s)_0 - Z^{j'}(s)_0)^2 \Bigg) \\ & + \frac{4}{(p+1)!} E\Bigg(\sum_{i<j,\,i'<j',\,(i',j')\neq(i,j)} (Y_{-k}^i - Y_{-k}^j)(Z^i(s)_0 - Z^j(s)_0) \\ & \qquad \times (Y_{-l}^{i'} - Y_{-l}^{j'})(Z^{i'}(s)_0 - Z^{j'}(s)_0) \\ & \qquad \times \prod_{i''<j'',\,(i'',j'')\neq(i',j') \text{ and } \neq(i,j)} (Z^{i''}(s)_0 - Z^{j''}(s)_0)^2 \Bigg). \end{aligned}$$

At the point $s = \theta$, this leads to

$$\frac{\partial^2}{\partial s_k \partial s_l} H(\theta) = \frac{2}{(p+1)!} E\Bigg(\sum_{i<j} (Y^i_{-k} - Y^j_{-k})(Y^i_{-l} - Y^j_{-l}) \times \prod_{i'<j',(i',j')\neq(i,j)} (X^{i'}_0 - X^{j'}_0)^2 \Bigg)$$
$$+ \frac{4}{(p+1)!} E\Bigg(\sum_{i<j,i'<j',(i',j')\neq(i,j)} (Y^i_{-k} - Y^j_{-k})(Y^{i'}_{-l} - Y^{j'}_{-l}) \cdot (X^{i'}_0 - X^{j'}_0)(X^i_0 - X^j_0) \times \prod_{i''<j'',(i'',j'')\neq(i',j') and \neq (i,j)} (X^{i''}_0 - X^{j''}_0)^2 \Bigg).$$

However $H(\theta) = 0$ says that a.s.,

$$\prod_{i<j} (X^i_0 - X^j_0) = 0$$

(which is also easily seen from the fact that the $p+1$ variables X^i_0 take values in the same alphabet with p values). This leads to

$$\frac{\partial^2}{\partial s_k \partial s_l} H(\theta) = \frac{2}{(p+1)!} E\Big(\sum_{i<j} (Y^i_{-k} - Y^j_{-k})(Y^i_{-l} - Y^j_{-l}) \prod_{i'<j',\,(i',\,j')\neq(i,\,j)} (X^{i'}_0 - X^{j'}_0)^2\Big),$$

which applied to v leads to the formula of Proposition 3.3.

Notice that for $v = \lambda\theta$, where λ is a real number, $v^T D_2 H(\theta) v = 0$. However, the set Θ^* cannot contain any multiple of delayed θ except 0; either Θ would be ambiguous on scale and delay. Now, $v^T D_2 H(\theta) v = 0$ if and only if for any $i < j$, $(Z^i(v)_0 - Z^j(v)_0) \prod_{i'<j',\,(i',\,j')\neq(i,\,j)} (X^{i'}_0 - X^{j'}_0) = 0$ a.s. Now,

$$(Z^i(v)_0 - Z^j(v)_0) \prod_{i'<j',\,(i',\,j')\neq(i,\,j)} (X^{i'}_0 - X^{j'}_0) = \sum_{k\neq 0} (u * v)_k P_{-k},$$

where $P_{-k} = (X^i_{-k} - X^j_{-k}) \prod_{i'<j',\,(i',\,j')\neq(i,\,j)} (X^{i'}_0 - X^{j'}_0)$. The variables P_k have discrete distribution with at least two different points of support, and, using (M4), any finite trajectory has positive probability. However, as soon as v is not a multiple of θ, there is at least one $k \neq 0$ such that $(u * v)_k \neq 0$. We may conclude that the distribution of $\sum_{k\neq 0} (u * v)_k P_{-k}$ may not be degenerate with support $\{0\}$, so that $D_2 H(\theta)$ is positive definite on Θ^*.

Let us now study the Toeplitz procedure. Applying Proposition 2.8, we have

$$H(s) = \frac{2^{p(p+1)/2}}{(p+1)!} E\Bigg[\prod_{j<k} \Bigg(1 - \cos\Bigg(\frac{Z^j(s)_0 - Z^k(s)_0}{F} \Bigg) \Bigg) \Bigg]$$

so that

$$\frac{\partial}{\partial s_l}H(s) = \frac{2^{p(p+1)/2}}{(p+1)!F}E\biggl(\sum_{j<k}(Y^j_{-l} - Y^k_{-l})\sin\biggl(\frac{Z^j(s)_0 - Z^k(s)_0}{F}\biggr) \times \prod_{j'<k',\,(j',\,k')\neq(j,\,k)}\biggl(1-\cos\biggl(\frac{Z^{j'}(s)_0 - Z^{k'}(s)_0}{F}\biggr)\biggr)\biggr)$$

and

$$\begin{aligned}\frac{\partial^2}{\partial s_l \partial s_m}&H(s)\\ &= \frac{2^{p(p+1)/2}}{(p+1)!F^2}E\biggl(\sum_{j<k}(Y^j_{-l} - Y^k_{-l})(Y^j_{-m} - Y^k_{-m})\cos\biggl(\frac{Z^j(s)_0 - Z^k(s)_0}{F}\biggr)\\ &\qquad \times \prod_{j'<k',(j',k')\neq(j,k)}\biggl(1-\cos\biggl(\frac{Z^{j'}(s)_0 - Z^{k'}(s)_0}{F}\biggr)\biggr)\biggr)\\ &\quad + \frac{2^{p(p+1)/2}}{(p+1)!F^2}E\biggl(\sum_{j<k,\,j'<k',(j',k')\neq(j,k)}(Y^j_{-l} - Y^k_{-l})(Y^{j'}_{-l} - Y^{k'}_{-l})\\ &\qquad \times \sin\biggl(\frac{Z^j(s)_0 - Z^k(s)_0}{F}\biggr)\sin\biggl(\frac{Z^{j'}(s)_0 - Z^{k'}(s)_0}{F}\biggr)\\ &\qquad \times \prod_{j''<k'',(j'',k'')\neq(j,k)\text{ and }(i',j')}\biggl(1-\cos\biggl(\frac{Z^{j''}(s)_0 - Z^{k''}(s)_0}{F}\biggr)\biggr)\biggr)\end{aligned}$$

At the point $s = \theta$, this leads to

$$\begin{aligned}\frac{\partial^2}{\partial s_l \partial s_m}&H(\theta)\\ &= \frac{2^{p(p+1)/2}}{(p+1)!F^2}E\biggl(\sum_{j<k}(Y^j_{-l} - Y^k_{-l})(Y^j_{-m} - Y^k_{-m})\cos\biggl(\frac{X^j_0 - X^k_0}{F}\biggr)\\ &\qquad \times \prod_{j'<k',(j',k')\neq(j,k)}\biggl(1-\cos\biggl(\frac{X^{j'}_0 - X^{k'}_0}{F}\biggr)\biggr)\biggr)\\ &\quad + \frac{2^{p(p+1)/2}}{(p+1)!F^2}E\biggl(\sum_{j<k,\,j'<k',(j',k')\neq(j,k)}(Y^j_{-l} - Y^k_{-l})\sin\biggl(\frac{X^j_0 - X^k_0}{F}\biggr)\\ &\qquad \times (Y^{j'}_{-l} - Y^{k'}_{-l})\sin\biggl(\frac{X^{j'}_0 - X^{k'}_0}{F}\biggr)\\ &\qquad \times \prod_{j''<k'',(j'',k'')\neq(j,k)\text{ and }(i',j')}\biggl(1-\cos\biggl(\frac{X^{j''}_0 - X^{k''}_0}{F}\biggr)\biggr)\biggr).\end{aligned}$$

But $H(\theta) = 0$ says that a.s.,

$$\prod_{j<k} (X_0^j - X_0^k) = 0$$

so that as soon as $\cos((X_0^j - X_0^k)/F) \neq 1$,

$$\prod_{j'<k',\,(j',\,k')\neq(j,\,k)} \left(1 - \cos\left(\frac{X_0^{j'} - X_0^{k'}}{F}\right)\right) = 0.$$

This leads to

$$\begin{aligned} \frac{\partial^2}{\partial s_l \partial s_m} H(\theta) = \frac{2^{p(p+1)/2}}{(p+1)!F^2} E\Bigg(\sum_{j<k} (Y_{-l}^j - Y_{-l}^k)(Y_{-m}^j - Y_{-m}^k) \\ \times \prod_{j'<k',\,(j',\,k')\neq(j,\,k)} \left(1 - \cos\left(\frac{X_0^{j'} - X_0^{k'}}{F}\right)\right)\Bigg), \end{aligned} \tag{5.5}$$

which applied to v leads to the formula of Proposition 3.3.

Now, $v^T D_2 H(\theta) v = 0$ if and only if for any $i < j$,

$$(Z^j(v)_0 - Z^k(v)_0)^2 \prod_{j'<k',\,(j',\,k')\neq(j,\,k)} \left(1 - \cos\left(\frac{X_0^{j'} - X_0^{k'}}{F}\right)\right) = 0$$

a.s. But $Z^j(v)_0 - Z^k(v)_0 = \sum_l (u * v)_l (X_{-l}^j - X_0^{-l})$. Now, on the event where $\prod_{j'<k',\,(j',\,k')\neq(j,\,k)} (1 - \cos((X_0^{j'} - X_0^{k'})/F)) \neq 0$, $X_0^j = X_0^k$ and as soon as v is not a multiple of θ, there is at least one $l \neq 0$ such that $(u * v)_l \neq 0$. We may conclude that the variable $Z^j(v)_0 - Z^k(v)_0$ cannot be always 0, so that $D_2 H(\theta)$ is positive definite on Θ^*. □

PROOF OF LEMMA 4.1. First, the ergodicity of (X_t) allows us to prove that $H_n(\theta, \sigma_0)$ converges in probability to $H(\theta, \sigma_0) = 0$, and that $D^1 H_n(\theta, \sigma_0)$ converges in probability to $D^1 H(\theta, \sigma_0)$. In order to compute the derivatives of H with respect to ξ or σ let us recall the following formula, which may be found in Gassiat and Gautherat (1998):

$$\forall\ \sigma \leq \sigma_0,\ \forall\ s,\ \forall\ i, \qquad c^i(s, \sigma) = E\left[\Phi_i\left(Y(s)_0 + \sqrt{\sigma_0^2 - \sigma^2}\,\varepsilon(s)_0\right)\right]$$

in which for any k,

$$Y(s)_k = \sum_j (s * u)_j X_{k-j}$$

and

$$\varepsilon(s)_k = \sum_j s_j \varepsilon_{k-j}.$$

For the Hankel procedure, we may now use Proposition 2.7 in the same way as when studying the definiteness of $D_s^2 H(\theta)$. Let, for $i = 0, \dots, p$, ε^i be p independent random sequences of independent variables, independent from the $(X_t^i)_{i=0,\dots,p,t\in\mathbf{Z}}$, with standard Gaussian distribution. We then have, for all ξ and all $\sigma < \sigma_0$,

$$H(\theta(\xi),\sigma) = \frac{1}{(p+1)!}E\bigg[\prod_{i<j}(Y^i(\theta)_0 - Y^j(\theta)_0 + \sqrt{\sigma_0^2-\sigma^2}(\varepsilon^i(\theta)_0 - \varepsilon^j(\theta)_0))^2\bigg].$$

It is now possible to take derivatives of this expression with respect to ξ and/or σ. We then take the value at $\xi = \xi^*$, and then look at the terms as polynomials in $\sqrt{\sigma_0^2-\sigma^2}$, beginning with negative exponents. Since it is already known that H is infinitely differentiable with respect to ξ and σ, the coefficients of terms with negative exponents are 0, and the terms with positive exponents will be set at 0 when letting σ tend to σ_0, so that we just have to compute the constant term in the polynomials.

For the Toeplitz procedure we may do the same thing with the formula

$$\begin{aligned} H(\theta(\xi),\sigma) &= \frac{2^{p(p+1)/2}}{(p+1)!}E \\ &\quad\times\bigg[\prod_{i<j}\bigg(1-\cos\bigg(\frac{Y^i(\theta)_0 - Y^j(\theta)_0 + \sqrt{\sigma_0^2-\sigma^2}(\varepsilon^i(\theta)_0 - \varepsilon^j(\theta)_0))}{F}\bigg)\bigg)\bigg].\end{aligned}$$

This leads to $D_\xi^1 H(\theta,\sigma_0) = 0$ and $D_\xi^2 H(\theta,\sigma_0) =$ a positive matrix,

$$D_\sigma^1 H(\theta,\sigma_0) = -\frac{2\sigma_0\|\theta\|_2^2}{(p+1)!}\sum_{i<j}E\bigg[\prod_{i'<j',\,(i'j')\neq(i,j)}(X_0^{i'} - X_0^{j'})^2\bigg]$$

for the Hankel procedure, and

$$D_\sigma^1 H(\theta,\sigma_0) = -\frac{2^{p(p+1)/2+1}\sigma_0\|\theta\|_2^2}{(p+1)!F}\sum_{i<j}E\bigg[\prod_{i'<j',\,(i'j')\neq(i,j)}\bigg(1-\cos\bigg(\frac{X_0^{i'} - X_0^{j'}}{F}\bigg)\bigg)\bigg]$$

for the Toeplitz procedure.

Now we have, if x is any of the variables ξ_j,

$$\begin{aligned} D_x^1 H_n(\theta,\sigma_0) &= (D_c^1 h(c_n(\theta,\sigma_0)) - D_c^1 h(c(\theta,\sigma_0)))D_x^1 c_n(\theta,\sigma_0) \\ &\quad + D_c^1 h(c(\theta,\sigma_0))(D_x^1 c_n(\theta,\sigma_0) - D_x^1 c(\theta,\sigma_0)) \end{aligned} \tag{5.6}$$

and also

$$D_c^1 h(c_n(\theta)) - D_c^1 h(c(\theta,\sigma_0)) = (c_n(\theta,\sigma_0) - c(\theta,\sigma_0))^T (D_c^2 h(c(\theta,\sigma_0)))(1+o(1))$$

Notice now that

$$c_n(\theta,\sigma_0) = R(\theta,\sigma_0)m_n(\xi^*) \tag{5.7}$$

in which $(m_n(\xi)) = (m_n^j(\theta(\xi)))_{j=1,\dots,2p}$ with

$$m_n^j(\theta(\xi)) = \frac{1}{n-2k(n)} \sum_{t=1+k(n)}^{n-k(n)} \Phi_j(\hat{Z}_t(\theta(\xi))).$$

We then easily obtain, using (M8), the ergodicity of (X_t) and (M5), that

$$m_n(\xi^*) = M_n(\xi^*) + o\left(\frac{1}{\sqrt{n}}\right), \qquad D_\xi^1 m_n(\xi^*) = D_\xi^1 M_n(\xi^*) + o\left(\frac{1}{\sqrt{n}}\right),$$

in which the $o(1)$ are in probability. Using (5.7) we have also

$$D_x^1 c_n(\theta, \sigma_0) = R(\theta, \sigma_0) D_x^1 M_n(\xi) + D_x^1 R(\theta, \sigma_0) M_n(\xi) + o\left(\frac{1}{\sqrt{n}}\right).$$

In particular, using (M8), $c_n(\theta, \sigma_0)$ converges in probability to $c(\theta, \sigma_0)$ and $D_x^1 c_n(\theta, \sigma_0)$ converges in probability to $D_x^1 c_n(\theta, \sigma_0)$.

Similarly, expanding also $H_n(\theta, \sigma_0)$ as

$$\begin{aligned} H_n(\theta, \sigma_0) &= h(c_n(\theta, \sigma_0)) - h(c(\theta, \sigma_0)) \\ &= D_c^1 h(c(\theta, \sigma_0))(c_n(\theta, \sigma_0) - c(\theta, \sigma_0))(1 + o(1)) \end{aligned}$$

allows us to prove that jointly $(\sqrt{n} D_\xi^1 H_n(\theta, \sigma_0), \sqrt{n} H_n(\theta, \sigma_0))$ converges in distribution to a centered Gaussian distribution and to compute the asymptotic variance. □

PROOF OF THEOREM 4.2. Let D_n be the matrix

$$D^2 H_n(\theta, \sigma_0) + \frac{D^1 H_n(\theta, \sigma_0) D^1 H_n(\theta, \sigma_0)^T}{H_n(\theta, \sigma_0)}.$$

Write

$$D_n = \begin{pmatrix} (D_n)_{11} & (D_n)_{12} \\ (D_n)_{21} & (D_n)_{22} \end{pmatrix}.$$

We have, using Lemma 4.1,

$$(D_n)_{11} = D_\xi^2 H_n(\theta, \sigma_0) + \frac{D_\xi^1 H_n(\theta, \sigma_0) D_\xi^1 H_n(\theta, \sigma_0)^T}{H_n(\theta, \sigma_0)}$$

converges in probability to $D_\xi^2 H(\theta, \sigma_0)$,

$$(D_n)_{12} = D_{\xi,\sigma}^2 H_n(\theta, \sigma_0) + \frac{D_\xi^1 H_n(\theta, \sigma_0) D_\sigma^1 H_n(\theta, \sigma_0)^T}{H_n(\theta, \sigma_0)}$$

converges in distribution to some random variable, as does $(D_n)_{12}$, and

$$(D_n)_{22} = D_\sigma^2 H_n(\theta, \sigma_0) + \frac{(D_\sigma^1 H_n(\theta, \sigma_0))^2}{H_n(\theta, \sigma_0)}$$

converges to $+\infty$. It follows that in probability for big enough n, the matrix D_n has nonzero determinant and is invertible. Let

$$D_n^{-1} = \begin{pmatrix} (D_n)^{11} & (D_n)^{12} \\ (D_n)^{21} & (D_n)^{22} \end{pmatrix}$$

be its inverse. Usual linear computations together with (4.1) lead to

$$\begin{aligned}(\hat{\xi} - \xi^*) = \Bigg(&(D_n)^{11} D_\xi^1 H_n(\theta, \sigma_0) - (D_n)^{11} \\ &\times \frac{(D_n)_{12}}{(D_n)_{22}} \left(D_\sigma^1 H_n(\theta, \sigma_0) + \frac{\delta(n)^2}{2H_n(\theta, \sigma_0)} \right) \Bigg) (1 + o(1))\end{aligned} \tag{5.8}$$

and

$$((D_n)^{11})^{-1} = \left((D_n)_{11} - \frac{(D_n)_{12}(D_n)_{21}^T}{(D_n)_{22}} \right).$$

We now have that

$$\frac{(D_n)_{12}(D_n)_{21}^T}{(D_n)_{22}}$$

converges to 0 in probability, so that

$$((D_n)^{11}) = G^{-1}(1 + o(1)).$$

We thus obtain

$$\sqrt{n}(\hat{\xi} - \xi^*) = -\frac{D_{\xi,\sigma}^2 H_n(\theta, \sigma_0)}{D_\sigma^1 H_n(\theta, \sigma_0)} \sqrt{n} H_n(\theta, \sigma_0)(1 + o(1))$$

which leads to the theorem. □

Acknowledgment. The authors thank the referees for their helpful comments.

REFERENCES

ANTON-HARO, C., FONOLLOSA, J. A. and FONOLLOSA, J. R. (1997). Blind channel estimation and data detection using hidden Markov models. *IEEE Trans. Signal Process.* **45**.

BERMOND, O. and KÉRIBIN, C. (1998). *M*-Likelihood estimation for noisy MA processes. Unpublished manuscript.

BICKEL, P. J., RITOV, Y. and RYDEN, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Statist.* **26** 1614–1635.

CAPPÉ, O., DOUCET, A., LAVIELLE, M. and MOULINES, E. (1999). Methods for blind maximum-likelihood linear system identification. *Signal Process.* **73** 3–25.

GAMBOA, F. and GASSIAT, E. (1996). Blind deconvolution of discrete linear systems. *Ann. Statist.* **24** 1964–1981.

GAMBOA, F. and GASSIAT, E. (1997). Source separation when the input sources are discrete or have constant modulus. *IEEE Trans. Signal Process.* **45** 3062–3072.

GASSIAT, E. and GAUTHERAT, E. (1998). Identification of noisy linear systems with discrete random input. *IEEE Trans. Inform. Theory* **44** 1941–1952.

GAUTHERAT, E. (1997). Déconvolution aveugle des systèmes linéaires aléatoires bruités ou non. Thèse de l'Université d'Evry-Val d'Essonne.

LI, T. H. (1993). Estimation and blind deconvolution of autoregressive systems with nonstationary binary inputs. *J. Time Ser. Anal.* **14** 575–588.

LI, T. H. (1995). Blind deconvolution of linear systems with multilevel nonstationary inputs. *Ann. Statist.* **23** 690–704.

LINDSAY, B. G. (1989). On the determinants of moment matrices. *Ann. Statist.* **17** 711–721.

LIU, J. and CHEN, R. (1995). Blind deconvolution via sequential imputations. *J. Amer. Statist. Assoc.* **90** 567–576.

PETROV, V. (1975). *Sums of Independent Variables*. Springer, Berlin.

SASHADRI, X. (1994). Joint data and channel estimation using blind trellis search techniques. In *Blind Deconvolution* (S. Hawking, ed.) 259–286. Prentice Hall, Englewood Cliffs, NJ.

VAN DER VEEN, A. J., TALWAR, S. and PAULRAJ, A. (1997). A subspace approach to blind space–time signal processing for wireless communication systems. *IEEE Trans. Signal Process.* **45** 173-190.

LABORATOIRE DE MATHÉMATIQUE
U.M.R. C 8628
UNIVERSITÉ PARIS-SUD
BAT. 425
91405 ORSAY CEDEX
FRANCE
E-MAIL: elisabeth.gassiat@math.u-psud.fr

LABORATOIRE DE MATHÉMATIQUE
UPRESA 6056
UNIVERSITÉ DE REIMS
MOULIN DE LA HOUSSE, B.P. 1039
51687 REIMS CEDEX 2
FRANCE