Speed of convergence for the blind deconvolution of a linear systems with discrete random input

E. Gassiat* E. Gautherat†

Abstract: 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. In this paper, we give the speed of convergence of the estimators in the parametric situation. With n noisy observations, the estimator satisfies a central limit theorem with speed $\sqrt{n}$ as usual, while with non noisy observations, the speed of convergence is governed by the l_1-tail of the inverse filter, which may have exponential decay. It appears that noisy and non noisy models are of different statistical nature. We also extend results concerning Hankel estimation to Toeplitz estimation, and prove a formula to compute Toeplitz forms that may have interest in itself.

A.M.S. Classification: 62G05, 62M09, 62M10

Key Words: Deconvolution, Contrast Function, Hankel matrix, Discrete Linear Systems.

*Laboratoire Analyse et Probabilité, Université d'Evry-Val d'Essonne, Boulevard des Coquibus, 91025 Evry

†Laboratoire Analyse et Probabilité, Université d'Evry-Val d'Essonne, Boulevard des Coquibus, 91025 Evry

1 Introduction

In a recent paper ([1]) 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. We further refined the method to take noise into account (see [2]) so that the estimated filter converges to the unknown filter whatever the signal to noise ratio was. We claimed in both papers that our estimator should behave asymptotically better than other estimators that apply to many input distributions, since our estimator takes the discreteness into account. This claim was supported by the result that, for autoregressive processess, the estimator achieved the exact value of the unknown filter with a finite number of observations.
The aim of this paper is to give theoretical results on the speed of convergence to make the claim proved.In a similar context but with known alphabet, Li ([4]) proposed an estimator to deal with multilevel inputs. In the non noisy situation and when the filter is parametrized with a finite dimensional parameter, the speed of convergence was upper bounded by the l_1-tail of the inverse of the unknown filter. We prove here a similar result for our estimator, and we also give partial results on the speed of convergence in the non parametric situation. Since our method may take noise into account, we give also results concerning the estimator when noise is present.

Let us now be more precise. We assume that the observed sequence $(Y_k)_{k\in\mathbb{Z}}$ is the output of an unknown linear time-invariant system $\mathcal{U}$ with impulse response $(u_k)_{k\in\mathbb{Z}}$ that is driven by an unobservable input sequence $(X_k)_{k\in\mathbb{Z}}$, corrupted or not with additive noise $(\sigma_0\epsilon_k)_{k\in\mathbb{Z}}$ where the level σ_0 is either 0, or known, or unknown:

$$Y_k = \sum_{j\in\mathbb{Z}} u_j X_{k-j} + \sigma_0 \cdot \epsilon_k \tag{1}$$

The input signal is known *a priori* to be discretely distributed with p different possible values. The linear system $u = (u_j)_{j\in\mathbb{Z}}$ is invertible; $\theta = (\theta_k)_{k\in\mathbb{Z}}$ is the inverse filter of u, that is :

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

where δ_k denotes the Kronecker symbol. Notice that we do not make any phase assumption on the system : the system $\mathcal{U}$ could have non minimum phase. Also, the input signal needs not to be independently distributed. To solve the problem of blind identification, we apply an adjustable linear time-invariant system $\mathcal{S}$: $s = (s_k)_{k\in\mathbb{Z}}$ to the output $(Y_k)_{k\in\mathbb{Z}}$ and work on the sequence $(Z(s)_k)_{k\in\mathbb{Z}}$:

$$Z(s)_k = \sum_j s_j Y_{k-j} \tag{2}$$

It is clear that the sequence $(Z(s)_k)_{k\in\mathbb{Z}}$ is just the sum of the sequence obtained as the result of the linear system $\mathcal{S} * \mathcal{U}$ applied to the input X and of the linear system $\mathcal{S}$ applied to the noise $\sigma_0 \cdot \epsilon$:

$$Z(s)_k = \sum_j (s * u)_j X_{k-j} + \sigma_0 \cdot \sum_j s_j \epsilon_{k-j} \tag{3}$$

Under very general assumptions, linear combinations of random variables lead to variables with strictly bigger support. Say that a distribution is p-concentrated if its support reduces to p different points. Our estimator relies on the quantification of the p-concentration of a probability distribution, in particular through the Hankel matrix of the first algebraic moments or Toeplitz matrix of the trigonometric moments of the distributions, and the fact that the inverse filter θ may be found as the particular filter s that leads to the only p-concentrated series $(Z(s)_k)_{k\in\mathbb{Z}}$ if there is no noise. In the presence of noise, divisibility of gaussian distributions allows a similar method to estimate simultaneously the inverse filter and the noise level σ_0. Let us now recall the definition of the estimators, and the convergence results we obtained.

General assumptions on the model (1) are the following :

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

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

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

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

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

(M5) $\epsilon = (\epsilon_k)_{k\in\mathbb{Z}}$ is a sequence of i.i.d. gaussian variables which are independent of the input signal; σ_0 is unknown; $E(\epsilon_1) = 0$; $E(\epsilon_1^2) = 1$.

Let us give simple examples where the assumptions hold:

- White input sequence: When the variables X_t, $t \in \mathbb{Z}$, are independent identically distributed, **(M1), (M3), (M4)** hold.

- Thresholded process: Let $(W_t)_{t\in\mathbb{Z}}$ be a stationary process such that the distribution of any finite marginal is continuous (for example, a Gaussian process). Let $m_1 < m_2 \cdots < m_{p-1}$ be real numbers, and set $m_0 = -\infty$, $m_p = +\infty$. Define the thresholded process $(X_t)_{t\in\mathbb{Z}}$ by: $X_t = x_j$ if and only if $W_t \in]m_j, m_{j+1}]$. Then **(M1), (M3), (M4)** hold.

- Aperiodic recurrent Markov chain: Let $(X_t)_{t \in \mathbb{Z}}$ be a Markov chain with state space of cardinality p and a transition matrix such that any transition probability is positive (all states communicate in one step). Then **(M1)** and **(M4)** hold. It is also easy to see that the chain is aperiodic and recurrent, so that **(M3)** holds.

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

In the next section, we shall recall the general estimation procedure proposed in [1] when there is no noise, Hankel estimation, Toeplitz estimation, and refinement to deal with the presence of noise, proposed in [2], together with their convergence Theorems. We also set new results when using Toeplitz forms in the estimation procedure. We refer to previous papers the interested readers for details and explanations on the procedures. Subsequent sections give the asymptotic speed of convergences in the parametric case when there is no noise and asymptotic results concerning the noisy situation. Numerical experiments to illustrate those theoretical results may be found in [2].

2 Estimation procedures and previous results.

2.1 T-system estimation procedure.

let $\Phi = (1, \Phi_1, \ldots, \Phi_{2p})$ be a Tchebytchev system (T-system) of functions on $[0,1]$ (for the definition of T-system see [3]). For any filter s, define:

$$\begin{aligned} c(s) &= (c^i(s))_{i=1,\ldots,2p} \\ c^i(s) &= E\left[\Phi_i\left(\varphi(Z(s)_1)\right)\right] \end{aligned}$$

where φ is a given continuous bijective function which maps $\mathbb{R}$ onto $]0,1[$. Notice that $\varphi(Z_1(s))$ is now a variable taking value in $]0,1[$. Now, a nice property of a T-system is the following: Let $\mathcal{P}$ be the set of all probability measures on $[0,1]$ and

$$\mathcal{K} := \{c \in \mathbb{R}^{2p} \;:\; \exists P \in \mathcal{P}, \int_0^1 \Phi dP = c\}.$$

Recall that if V is a random variable taking value in $]0,1[$ then: $E(\Phi(V))$ lies on the boundary of $\mathcal{K}$ if and only if V is discrete with at most p points of support. Let now h be a non negative and continuous function defined on $\overline{\mathcal{K}}$ such that:

$$h(c) = 0 \Longleftrightarrow c \in bd(\mathcal{K})$$

We then define a contrast function H by:

Definition 2.1

$$H(s) = h(c(s)), \quad s \in \Theta$$

Define now the parameter space Θ as a subset of $l_1(\mathbb{Z})$ which is unambiguous on scale and delay. The sequence H_n is defined as an empirical contrast function in the following way. To use only the observations $Y_1, \ldots, Y_n$, we need to truncate the filter s.
Let $k(n)$ be an increasing sequence of integers. Define:

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

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

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

We may now define:

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

We now define the estimator:

Definition 2.2 *$\hat{\theta}$ is any minimizer of H_n over Θ_n:*

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

We assume throughout the sequel that:

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

The following Theorems were proved in [1]

Theorem 2.3 *Assume that* ***(M1), (M2), (M3), (M4)*** *hold, and that $\sigma_0 = 0$. If Θ is compact, then $\hat{\theta}$ converges almost surely, in $l^1(\mathbb{Z})$, 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 set $\mathcal{S}$ of dimension q: $\xi = (\xi_j)_{j=1\ldots q}$:

$$\Theta := \{\theta(\xi), \ \xi \in \mathcal{S}\}$$

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 a given compact set K containing ξ^*.

Theorem 2.4 *Assume that the application $\xi \to \theta(\xi)$ from $\mathbb{R}^q$ to $l^1(\mathbb{Z})$ is continuous, and that assumptions* ***(M1)*** *to* ***(M4)*** *hold, and that $\sigma_0 = 0$. Assume the identifiability assumption:*

$$\theta_k(\xi) = r\theta_{k-K}(\xi'), \ \forall k \in \mathbb{Z} \Longleftrightarrow r = 1, \ K = 0 \ and \ \xi = \xi'.$$

Then, $\hat{\xi}$ converges, almost surely, as n approaches infinity, to ξ^*.

2.2 Hankel forms

We propose here a similar estimation procedure using Hankel forms based on the algebraic moments. Let Φ be the sequence of moment functions: $\Phi_j(x) = x^j$, $j = 1, \dots, 2p$. Let $M(s)$ be the $(p+1) \times (p+1)$ Hankel matrix given by: $M(s)_{i,j} = c^{i+j-2}(s)$, $i, j = 1, \dots, p+1$, where here $c(s) = E(\Phi(Z(s)_1))$ (Notice that here we do not need to map $\mathbb{R}$ on the interval $]0, 1[$). This matrix is non negative as soon as c is the begining of the moment sequence of a random variable, and degenerates if and only if this random variable is discrete with at most p points of support. Set: $h(c) = \det[M]$, and $H(s) = h(c(s))$. Then Theorems 2.3 and 2.4 hold.
The interest of this estimation procedure is because it may be used to handle with the presence of noise.
Define $M(s, \sigma)$ as the Hankel matrix built using the solutions $c^j(s, \sigma)$ of the triangular system :

$$E(Z(s)_1^j) = \sum_{i=0}^{j} C_j^i c^i(s, \sigma) \cdot v(s, \sigma)^{j-i} \mu_{j-i}, \; j = 0, \dots, 2p \tag{4}$$

where $v^2(s, \sigma) = \sigma^2 \cdot \|s\|_2^2$ and μ_{j-i} is the $j - i$-th moment of the standard gaussian. Define the fonction $H(s, \sigma)$ of the filter and the noise level as the value of the determinant of $M(s, \sigma)$. Define the estimators $\hat{c}^j(s, \sigma)$ of the pseudo-moments $c^j(s, \sigma)$ as the solutions of the triangular system:

$$\hat{c}^j(s) = \sum_{i=0}^{j} C_j^i \cdot \hat{c}^i(s, \sigma) \cdot (\sigma \cdot \|s\|_2)^{j-i} \cdot \mu_{j-i}, \quad j = 1, \dots, 2p$$

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

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

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

$$J_n(s, \sigma) = (H_n(s, \sigma))^2 + (\delta(n))^2 \cdot \sigma \tag{5}$$

We set:

Definition 2.5 *The estimator $(\hat{\theta}, \hat{\sigma})$ is any minimizer of J_n over $\Theta \times \mathbb{R}^+$.*

To have a good asymptotic behaviour 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 some more assumption on the processes :

Assumption (M6) : Assume that

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

and that:

$$\lim_{n\to\infty} (\delta(n))^{-1} \cdot \frac{1}{n} \sum_{t=1}^{n} \left([X_t + \sum_k \theta_k \epsilon_{t-k}]^j - m_j(\theta) \right) = 0$$

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

In [2], a convergence Theorem was proved for the estimator minimizing $|\det[M_n(s,\sigma)]| + \delta(n) \cdot \sigma$. Since we shall derive the speed of convergence via a Taylor expansion, we choosed here to work with $(\det[M_n(s,\sigma)])^2$. Following the same lines, we easily have:

Theorem 2.6 *Assume that (M1) to (M6) hold.*
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. Assume that the model is finitely parameterized by $\xi = (\xi_k)_{k=1,\ldots,q}$ in $\mathbb{R}^q$, such that the parametrizing function $\theta(\xi)$ is continuous, one to one, with continuous inverse for ξ in a compact set Ξ. Define the estimator $(\widehat{\xi}, \widehat{\sigma})$ as the minimizer of $J_n(\theta(\xi), \sigma)$ over $\Xi \times \mathbb{R}^+$. If the parametrizing function verifies assumptions (M1) to (M6), we obviously have :

Corollary 2.7 *$(\widehat{\xi}, \widehat{\sigma})$ converges in probability to the true value (ξ, σ_0) of the parameter.*

We shall also recall a useful formula given in Lindsay ([5]), which gives the value $M(W)$ of the determinant of the Hankel matrix based on the $2p$ first algebraic moments of a random variable W:

Proposition 2.8 *Let $W_0, \ldots, W_p$ be $p+1$ independent copies of W. We have:*

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

2.3 Toeplitz forms

In case some information is available concerning the magnitude of the X_t, Toeplitz forms may be used to handle with noisy observations in a similar manner as Hankel forms. We here assume that:

- **(T)** The unknown alphabet A is a subset of $]m, M[$, m and M are known.

Let $F = 2\pi(M-m)\|u\|_1 \sup_{s\in\Theta} \|s\|_1$. Let Φ be the sequence of moment functions: $\Phi_j(x) = \exp(\frac{i}{F}(j-p-1)x)$, $j = 1, \ldots, 2p$. Let $T(s)$ be the $(p+1)\times(p+1)$ Toeplitz matrix given by: $T(s)_{i,j} = c^{i-j}(s)$, $i, j = 1, \ldots, p+1$, where here $c(s) = E(\Phi(Z_1(s)))$ This matrix is non negative as soon as c is the begining of the Fourier coefficients of the distribution of a random variable, and degenerates if and only if this random variable is discrete with at most p points of support. Set: $h(c) = \det[T]$, and $H(s) = h(c(s))$. Then Theorems 2.3 and 2.4 hold if **(T)** is assumed.
Again, this procedure may be used to handle with the presence of noise.
Define $T(s,\sigma)$ as the Toeplitz matrix built using the $c^j(s,\sigma)$:

$$E(\exp(\frac{ij}{F}Z(s)_1)) = c^j(s,\sigma)\cdot \exp(-\frac{j^2\sigma^2\|s\|_2^2}{2F^2}) \quad j = -p, \ldots, p \tag{6}$$

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

$$\widehat{c}^j(s,\sigma) = \widehat{c}^j(s)\cdot \exp(\frac{j^2\sigma^2\|s\|_2^2}{2F^2}) \quad j = -p, \ldots, p$$

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

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

$J_n(s,\sigma)$ is defined again by (5), and $(\widehat{\theta}, \widehat{\sigma})$ as any minimizer of J_n over $\Theta \times \mathbb{R}^+$. Replace assumption **(M6)** by
Assumption (T6) : Assume that

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

and that:

$$\lim_{n\to\infty} (\delta(n))^{-1}\cdot \frac{1}{n}\sum_{t=1}^{n}\left(\exp(\frac{i}{F}j(X_t + \sum_k \theta_k\epsilon_{t-k})) - m_j(\theta)\right) = 0$$

in probability for $j = 1, \ldots, 2p$, where $m_j(\theta) = E(\exp(\frac{i}{F}j(X_t + \sum_k \theta_k\epsilon_{t-k})))$.

The following Theorem holds, its proof is analoguous to that of Theorem 2.6, and will be omitted.

Theorem 2.9 *Assume that (M1) to (M5), (T) and (T6) hold.*
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. With a parametric continuous parametrization, $(\widehat{\xi}, \widehat{\sigma})$ converges in probability to the true value (ξ, σ_0) of the parameter.

We shall now prove a useful formula, which gives the value $T(W)$ of the determinant of the Toeplitz matrix based on the p first complex exponential moments of a random variable W. It is proved in section 5.

Proposition 2.10 *Let $W_0, \ldots, W_p$ be $p+1$ independent copies of W. We have:*

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

We shall now study the speed of convergence of the proposed estimators in the parametric situation. The parameter will now be $\xi = (\xi_j)_{j=1\ldots q}$ in the parameter space $\mathcal{S}$.

$$\Theta := \{\theta(\xi),\ \xi \in \mathcal{S}\}$$

We assume that the model is identifiable. ξ^* is the true parameter value.

3 Speed of convergence: non noisy observations.

Throughout this section, the level noise σ_0 is set to 0.
From now on, the notation $D^r_x F(y)$ will designate the r-th derivative of F with respect to the variable x and evaluated at point y.
Let also Θ^* be the set of all elements of Θ except θ.
Since the speed of convergence is studied via Taylor expansions, we shall always assume:

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

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 $D^2_s H(\theta)$ is necessarily non negative. The assumption concerns the definiteness. Notice also that the gradient operator $D^1_s H$ of H is in L_∞ with no more assumptions.
In case H is the Hankel form, h is a multi polynomial and c is an algebraic power, so that they are twice continuously differentiable. It will be later proved that in this case $D^2_s H(\theta)$ is positive definite on Θ^*, so that **(D)** holds with no particular assumption. In case H is the Toeplitz form, h is a multi polynomial and c is a complex exponential, so that they are twice continuously differentiable. It will also be later proved that in this case $D^2_s H(\theta)$ is positive definite on Θ^*, so that again **(D)** holds with no particular assumption.
We shall also use the following assumptions:

- **(P)** The application $\xi \to \theta(\xi)$ is twice continuously differentiable.
  For any $i = 1, \ldots, q$, $(\frac{\partial \theta_k}{\partial \xi_i})_{k \in \mathbb{Z}}$ and $(\frac{\partial^2 \theta_k}{\partial \xi_i^2})_{k \in \mathbb{Z}}$ are in $l_1(\mathbb{Z})$.
  Moreover, $(\frac{\partial \theta_k}{\partial \xi_1}(\xi^*))_{k \in \mathbb{Z}}, \ldots, (\frac{\partial \theta_k}{\partial \xi_q}(\xi^*))_{k \in \mathbb{Z}}$ are linearly independent, and no linear combination of these derivatives equals θ.

3.1 Asymptotic result

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

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

The main result of this section is:

Theorem 3.2 *Assume that **(M1)**, **(M2)**, **(M3)**, **(M4)**, **(P)** hold. If the estimation method is not Hankel nor Toeplitz, assume moreover **(D)**. We have almost surely for big enough n:*

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

where C is a constant.

Proof.
Using Theorem 2.4, $\hat{\xi}$ is almost surely consistent. So that for big enough n, and for $i = 1, \ldots, q$:

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

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

$$0 = \frac{\partial}{\partial \xi_i} L_n(\xi^*) + \sum_{j=1}^{q} (\hat{\xi}_j - \xi_j^*) \frac{\partial^2}{\partial \xi_i \partial \xi_j} L_n(\tilde{\xi}^j) \tag{7}$$

where $\tilde{\xi}^j \in [\hat{\xi}, \xi^*)$. Usually, empirical 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 **non truncated** empirical moments. To do this, define for any filter s the empirical moments for the non truncated series:

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

and also

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

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:

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

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

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

for all k, and then:

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

This will allow to prove (see section 5)

$$|\frac{\partial}{\partial \xi_i} L_n(\xi^*)| \leq C_4 \cdot \sum_{|k| > k(n)} |\theta_k| \tag{8}$$

for some constant C_4. Now, for any $i, j = 1, \ldots, q$ 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 proven, 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,

$$\sum_{i,j=1,\ldots,q} \frac{\partial^2}{\partial \xi_i \partial \xi_j} L(\xi^*)(\hat{\xi}_i - \xi_i^*)(\hat{\xi}_j - \xi_j^*) \geq \lambda \|\hat{\xi} - \xi^*\|_2^2 \tag{9}$$

where λ is the smallest eigen value of $D_2 L(\xi^*)$.
The Theorem follows using (7), (8) and (9) and with $C = \frac{2C_4}{\lambda}$.

3.2 Definiteness of $D_s^2 H(\theta)$

Let us study the second derivative operator of H with respect to s when it is the Hankel or the Toeplitz form.
Let $(X_t^0)_{t\in\mathbb{Z}},\ldots,(X_t^p)_{t\in\mathbb{Z}}$ be $p+1$ independent copies of $(X_t)_{t\in\mathbb{Z}}$. Define for $i=0,\ldots,p$ and $t\in\mathbb{Z}$

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

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

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

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

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

$$v^T\cdot D_s^2 H(\theta)\cdot v = \frac{2}{(p+1)!}\sum_{i<j} E\left((Z_0^i(v)-Z_0^j(v))\prod_{i'<j',(i',j')\neq(i,j)}(X_0^{i'}-X_0^{j'})\right)^2$$

and for the Toeplitz procedure:

$$v^T\cdot D_s^2 H(\theta)\cdot v = \frac{2^{p(p+1)/2}}{(p+1)!F^2}\sum_{i<j} E\left((Z_0^i(v)-Z_0^j(v))^2\prod_{i'<j',(i',j')\neq(i,j)}(1-\cos(\frac{X_0^{i'}-X_0^{j'}}{F}))\right)$$

*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*

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

with

- *For the Hankel procedure*

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

- *For the Toeplitz procedure*

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

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

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

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. And if the X_k^i are i.i.d., for $k \neq l \in \mathbb{Z}$, $E(X_{-k}^0 - X_{-k}^1)(X_{-l}^0 - X_{-l}^1) \prod_{i<j,(i,j)\neq(0,1)} (X_0^i - X_0^j)^2) = 0$. For $k = l \neq 0$, $E(X_{-k}^0 - X_{-k}^1)(X_{-l}^0 - X_{-l}^1) \prod_{i<j,(i,j)\neq(0,1)} (X_0^i - X_0^j)^2 = (p-1)! C_H$. Same arguments lead to the result for the Toeplitz procedure.

4 Speed of convergence: noisy observations.

Now, the level of noise σ is unknown, and the estimator $(\hat{\xi}, \hat{\sigma})$ minimizes

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

defined in section 2 for Hankel or Toeplitz forms.
With Φ the algebraic powers if the Hankel procedure is in study, and Φ the complex exponentials if the Toeplitz procedure is in study define

$$m_j(\xi) = (E(\Phi_j(Z_t(\theta(\xi)))))_{j=1,\ldots,2p}$$

Define now the vectors

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

$$D_\xi^1 M_n(\xi) = ((\frac{\partial}{\partial \xi_i} M_n(\xi)))_{i=1,\ldots,q}$$

Let us introduce the following assumption:

- **(M8)**

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

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

The study of the asymptotic distribution proceeds as usual by a Taylor expansion of the contrast function. Using **(P)**, J_n is twice continuously differentiable. Under **(M8)**,

(M6) holds. Then $(\hat{\xi}, \hat{\sigma})$ is a zero of DJ_n, and converges in probability to (ξ, σ_0) by Corollary 2.7, so that the following Taylor expansion holds:

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

We have:

$$D^1 J_n(\theta, \sigma_0) = (2H_n(\theta, \sigma_0) D^1_\xi H_n(\theta, \sigma_0); 2H_n(\theta, \sigma_0) D^1_\sigma H_n(\theta, \sigma_0) + (\delta(n))^2)^T$$

Now, the matrix $D^2 J_n(\theta, \sigma_0)$ equals

$$2H_n(\theta, \sigma_0) D^2 H_n(\theta, \sigma_0) + 2D^1 H_n(\theta, \sigma_0) D^1 H_n(\theta, \sigma_0)^T$$

Notice that we have $H_n(\theta, \sigma_0) \neq 0$. Indeed, the pseudo-moments may not be moments of discrete random variables. We may then rewrite the Taylor expansion:

$$\begin{pmatrix} D^1_\xi H_n(\theta, \sigma_0) \\ \\ D^1_\sigma H_n(\theta, \sigma_0) + \frac{(\delta(n))^2}{2H_n(\theta, \sigma_0)} \end{pmatrix} + \left(D^2 H_n(\theta, \sigma_0) + \frac{D^1 H_n(\theta,\sigma_0) D^1 H_n(\theta,\sigma_0)^T}{H_n(\theta,\sigma_0)} \right) \cdot \begin{pmatrix} \hat{\xi} - \xi^* \\ \\ \hat{\sigma} - \sigma_0 \end{pmatrix} (1+o(1)) = 0 \qquad (10)$$

As usual, the asymptotic result comes from adequate central limit theorems and law of large numbers, together with the asymptotic definiteness of the second derivative matrix. Define $K(\theta(\xi), \sigma)$ the triangular matrix inverting the system (4) for the Hankel procedure, or inverting the system (6) for the Toeplitz procedure. In this last case, $K(\theta(\xi), \sigma)$ is a diagonal matrix. Notice that $K(\theta(\xi), \sigma)$ is differentiable with respect to ξ and with respect to σ. Define $D^1_{\xi_j} K$ the derivative matrix of K with respect to ξ_j, $D^1_{\xi_j} c$ be the derivative vector of $c(\theta(\xi), \sigma)$ with respect to ξ_j. Let V be the $q \times (2pq)$ matrix:

$$V = (V_1 V_2)$$

where V_1 is the $(q) \times (2p)$ matrix with lines

$$(V_1)_j = (D^1_{\xi_j} c)^T \cdot D^2_c h \cdot K + D^1_c h^T \cdot D^1_{\xi_j} K$$

where all functions or matrices or evaluated at $(\theta(\xi^*), \sigma_0)$, and V_2 is the $(q+1) \times (2pq)$ matrix which is block diagonal:

$$V_2 = Diag(V_2^1, \dots, V_2^q)$$

where for $j = 1, \dots, q$, the block V_2^j is the $2p$-dimensional vector

$$D^1_c h^T \cdot K$$

We now have:

Lemma 4.1 *Under the assumptions of Theorem 4.2 below,* $(\sqrt{n}D_\xi^1 H_n(\theta,\sigma_0), \sqrt{n}H_n(\theta,\sigma_0))$ *converges in distribution to a centered gaussian distribution, and* $D_\sigma^1 H_n(\theta,\sigma_0)$ *converges in probability to a positive constant. The asymptotic distribution of* $\sqrt{n}D_\xi^1 H_n(\theta,\sigma_0)$ *has variance* $V \cdot \Gamma \cdot V^T$.

Define C the $q \times q$ matrix given by:

$$C_{i,j} = v_i^T \cdot D_s^2 H(\theta) \cdot v_j$$

with v_i the filter $\frac{\partial \theta}{\partial \xi_i}(\xi^*)$, and $D_s^2 H(\theta)$ given in Proposition 3.3. We have

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:*

$$\Sigma = C^{-1} \cdot V \cdot \Gamma \cdot V^T \cdot (C^{-1})^T$$

Roughly speaking, the asymptotic result comes from the fact that asymptotically, the matrix involved in equation (10) has the bottom-right term tending to infinity, so that asymptotically, the inverse has only the up-left term as non zero term. It is proved in the last section.

5 Discussion

Theorems 3.2 and 4.2 allow to see that the model without noise and the model with noise are statistically of different nature. The model with noise is a regular parametric model, where the optimal speed of convergence for parametric estimation is $\sqrt{n}$. The model with no noise is not regular, it is not dominated. The parameter may here be estimated with speed the l_1 tail of the inverse filter, which may have exponential decay. This speed appears to be related to some truncation parameter. Since, with no truncation, the parameter could be perfectly estimated, the model appears in some sense non random. No lower bound is known for estimating a parameter in such non regular models, so that we do not know whether our estimator achieves the optimal rate or not.

Let us also make some remarks about non parametric estimation, that is the direct estimation of the inverse filter θ. With no noise, following the same lines for the study of the convergence rate, it appears that the gradient of H_n at point θ has an l_∞-norm upper bounded by the l_1-tail of the inverse filter. However, the remaining problem is to study the resting terms in th Taylor expansion, together with the convergence rate of the second derivative operator of H_n to the operator D^2H. The main difference with the parametric case is that all norms are not equivalent, so that the norm in which the remaining terms are studied has a great importance.

6 Proofs

Proof of Proposition 2.10.

Let MT be the Toeplitz matrix:

$$MT = \begin{pmatrix} 1 & E(e^{-i2\pi\frac{W_1}{F}}) & \cdots & E(e^{-i2\pi p\frac{W_p}{F}}) \\ E(e^{i2\pi\frac{W_0}{F}}) & 1 & \cdots & E(e^{-i2\pi(p-1)\frac{W_p}{F}}) \\ \vdots & \cdots & \ddots & \cdots \\ E(e^{i2\pi p\frac{W_0}{F}}) & E(e^{i2\pi(p-1)\frac{W_1}{F}}) & \cdots & 1 \end{pmatrix}$$

Let τ be a permutation of the $p+1$ indices.

$$\forall\tau, \quad MT = \begin{pmatrix} 1 & E(e^{-i2\pi\frac{W_{\tau(1)}}{F}}) & \cdots & E(e^{-i2\pi p\frac{W_{\tau(p)}}{F}}) \\ E(e^{i2\pi\frac{W_{\tau(0)}}{F}}) & 1 & \cdots & E(e^{-i2\pi(p-1)\frac{W_{\tau(p)}}{F}}) \\ \vdots & \cdots & \ddots & \cdots \\ E(e^{i2\pi p\frac{W_{\tau(0)}}{F}}) & E(e^{i2\pi(p-1)\frac{W_{\tau(1)}}{F}}) & \cdots & 1 \end{pmatrix}$$

Taking the determinant of $MT(W)$, we have a sum of products where, in each product, the involved variables are independent, so that we may write:

$$\forall\tau, \quad det(MT) = Edet \begin{pmatrix} 1 & e^{-i2\pi\frac{W_{\tau(1)}}{F}} & \cdots & e^{-i2\pi p\frac{W_{\tau(p)}}{F}} \\ e^{i2\pi\frac{W_{\tau(0)}}{F}} & 1 & \cdots & e^{-i2\pi(p-1)\frac{W_{\tau(p)}}{F}} \\ \vdots & \cdots & \ddots & \cdots \\ e^{i2\pi p\frac{W_{\tau(0)}}{F}} & e^{i2\pi(p-1)\frac{W_{\tau(1)}}{F}} & \cdots & 1 \end{pmatrix}$$

and also

$$\forall\tau, \quad det(MT) = E\prod_{j=0}^{p} e^{-i2\pi j\frac{W_{\tau(j)}}{F}}\ det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ e^{i2\pi\frac{W_{\tau(0)}}{F}} & e^{i2\pi\frac{W_{\tau(1)}}{F}} & \cdots & e^{i2\pi p\frac{W_{\tau(p)}}{F}} \\ \vdots & \cdots & \ddots & \cdots \\ e^{i2\pi p\frac{W_{\tau(0)}}{F}} & e^{i2\pi p\frac{W_{\tau(1)}}{F}} & \cdots & e^{i2\pi p\frac{W_{\tau(p)}}{F}} \end{pmatrix}$$

Let now $s(\tau)$ be the signature of the permutation τ. We have:

$$\forall\sigma, \quad det(MT) = E\prod_{j=0}^{p} e^{-i2\pi j\frac{W_{\tau(j)}}{F}}(-1)^{s(\tau)}\ det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ e^{i2\pi\frac{W_0}{F}} & e^{i2\pi\frac{W_1}{F}} & \cdots & e^{i2\pi p\frac{W_p}{F}} \\ \vdots & \cdots & \ddots & \cdots \\ e^{i2\pi p\frac{W_0}{F}} & e^{i2\pi p\frac{W_1}{F}} & \cdots & e^{i2\pi p\frac{W_p}{F}} \end{pmatrix}$$

But the distribution of the random variable involved in the expectation does not depend on the permutation. We thus have:

$$(p+1!)det(MT) = E\sum_{\tau}\prod_{j=0}^{p} e^{-i2\pi j\frac{W_{\tau(j)}}{F}}(-1)^{s(\tau)}\ det\begin{pmatrix} 1 & 1 & \cdots & 1 \\ e^{i2\pi\frac{W_0}{F}} & e^{i2\pi\frac{W_1}{F}} & \dots & e^{i2\pi p\frac{W_p}{F}} \\ \vdots & \dots & \ddots & \dots \\ e^{i2\pi p\frac{W_0}{F}} & e^{i2\pi p\frac{W_1}{F}} & \dots & e^{i2\pi p\frac{W_p}{F}} \end{pmatrix}$$

But $\sum_{\sigma}\prod_{j=0}^{p} e^{-i2\pi j\frac{W_{\tau(j)}}{F}}(-1)^{e(\tau)}$ is the exact expansion of the determinant of the matrix

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ e^{-i2\pi\frac{W_0}{F}} & e^{-i2\pi\frac{W_1}{F}} & \dots & e^{-i2\pi p\frac{W_p}{F}} \\ \vdots & \dots & \ddots & \dots \\ e^{-i2\pi p\frac{W_0}{F}} & e^{-i2\pi p\frac{W_1}{F}} & \dots & e^{-i2\pi p\frac{W_p}{F}} \end{pmatrix}$$

which is a Vandermonde matrix, with known determinant. We thus obtain:

$$(p+1!)det(MT) = E\left[\prod_{0\le k<j\le p}\left(e^{-i2\pi\frac{W_k}{F}} - e^{-i2\pi\frac{W_j}{F}}\right)\cdot\prod_{0\le k<j\le p}\left(e^{i2\pi\frac{W_k}{F}} - e^{i2\pi\frac{W_j}{F}}\right)\right]$$

which leads to the formula of the Proposition.

Proof of Proposition 3.1.
Differentiability of L and L_n easily come from that of H, H_n and θ. Now we have:

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

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

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

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

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

which, using **(D)**, is nonzero unless $y\cdot D_\xi^1\theta(\xi^*)$ is either the null series, which is impossible since by **(P)** $(\frac{\partial\theta_k}{\partial\xi_1}(\xi^*))_{k\in\mathbb{Z}},\dots,(\frac{\partial\theta_k}{\partial\xi_q}(\xi^*))_{k\in\mathbb{Z}}$ are linearly independent, or the filter $\theta(\xi^*)$, which is also impossible by **(P)**.

Proof of Theorem 3.2.
It remains to prove formula (8). Following the first formula, we have

$$|\frac{\partial}{\partial\xi_i}L_n(\xi^*)| \le \sup_{|k|\le k(n)}|\frac{\partial}{\partial s_k}H_n(\theta(\xi^*)) - \frac{\partial}{\partial s_k}\tilde{H}_n(\theta(\xi^*))|\cdot\|\frac{\partial}{\partial\xi_i}\theta(\xi^*)\|_1$$

Now for any k

$$\frac{\partial}{\partial s_k}H_n(\theta(\xi^*))-\frac{\partial}{\partial s_k}\tilde{H}_n(\theta(\xi^*)) = D_c^1 h(c_n(\theta(\xi^*)))\cdot\frac{\partial}{\partial s_k}c_n(\theta(\xi^*))-D_c^1 h(\tilde{c}_n(\theta(\xi^*)))\cdot\frac{\partial}{\partial s_k}\tilde{c}_n(\theta(\xi^*))$$

and then

$$\begin{aligned}\frac{\partial}{\partial s_k}H_n(\theta(\xi^*)) - \frac{\partial}{\partial s_k}\tilde{H}_n(\theta(\xi^*)) &= \left(D_c^1 h(c_n(\theta(\xi^*))) - Dh_c^1(\tilde{c}_n(\theta(\xi^*)))\right)\cdot\frac{\partial}{\partial s_k}c_n(\theta(\xi^*)) + \\ & D_c^1 h(\tilde{c}_n(\theta(\xi^*)))\left(\frac{\partial}{\partial s_k}c_n(\theta(\xi^*)) - \frac{\partial}{\partial s_k}\tilde{c}_n(\theta(\xi^*))\right)\end{aligned}$$

Moreover

$$\begin{aligned}\|D_c^1 h(c_n(\theta(\xi^*))) - D_c^1 h(\tilde{c}_n(\theta(\xi^*)))\|_\infty &\le 2pC_1\|c_n(\theta(\xi^*)) - \tilde{c}_n(\theta(\xi^*))\|_\infty \\ &\le 2pC_1\cdot C_2\|\hat{Z}(\theta(\xi^*))_t - Z(\theta(\xi^*))_t\|_\infty \\ &\le 2pC_1C_2\|X\|_\infty\|u\|_1\cdot\sum_{|k|>k(n)}|\theta_k|\end{aligned}$$

with $C_1 = \|D_c^2 h\|_\infty$, $C_2 = \|D(\Phi\circ\varphi)\|_\infty$ and $\|X\|_\infty$ the maximum possible absolute value in the alphabet where the variables X_t take value. Here, the norms $\|\cdot\|_\infty$ for functions are taken as the supremum value of the function on the space where the possible moments and there derivatives take value, which are compact since the X_t are bounded and the filters are summable.
Notice also that for any k

$$\|\frac{\partial}{\partial s_k}c_n(\theta(\xi^*))\|_\infty \le C_2\cdot\|X\|_\infty\cdot\|u\|_1$$

so that for any k

$$|D_c^1 h(c_n(\theta(\xi^*))) - Dh_c^1(\tilde{c}_n(\theta(\xi^*)))|\cdot|\frac{\partial}{\partial s_k}c_n(\theta(\xi^*))| \le 4p^2C_1C_2^2\|u\|_1^2\|X\|_\infty^2\cdot\sum_{|k|>k(n)}|\theta_k|$$

We also have:

$$\|D_c^1 h(\tilde{c}_n(\theta(\xi^*)))\| \le \|D_c^1 h\|_\infty$$

and for any $|k|\le k(n)$

$$\|\frac{\partial}{\partial s_k}c_n(\theta(\xi^*)) - \frac{\partial}{\partial s_k}\tilde{c}_n(\theta(\xi^*)\|_\infty \le C_3\|u\|_1^2\|X\|_\infty^2\cdot\sum_{|k|>k(n)}|\theta_k|$$

with $C_3 = \|D^2(\Phi\circ\varphi)\|_\infty$, so that

$$|D_c^1 h(\tilde{c}_n(\theta(\xi^*)))(\frac{\partial}{\partial s_k}c_n(\theta(\xi^*)) - \frac{\partial}{\partial s_k}\tilde{c}_n(\theta(\xi^*))| \le 2pC_3\|D_c^1 h\|_\infty\|u\|_1^2\|X\|_\infty^2\cdot\sum_{|k|>k(n)}|\theta_k|$$

We finally obtain (8) with $C_4 = 2p(2pC_1C_2^2 + C_3\|Dh\|_\infty)\|X\|_\infty^2\|(\frac{\partial}{\partial \xi_i}\theta_k(\xi^*))\|_1\|u\|_1^2$.

Proof of Proposition 3.3.

Let us first study the Hankel procedure. Applying Proposition 2.8 we have:

$$H(s) = \frac{1}{(p+1)!}E[\prod_{i<j}(Z_0^i(s) - Z_0^j(s))^2]$$

so that

$$\frac{\partial}{\partial s_k}H(s) = \frac{2}{(p+1)!}E\left(\sum_{i<j}(Y_{-k}^i - Y_{-k}^j)(Z_0^i(s) - Z_0^j(s)) \prod_{i'<j',(i',j')\neq(i,j)} (Z_0^{i'}(s) - Z_0^{j'}(s))^2\right)$$

and

$$\begin{aligned}
\frac{\partial^2}{\partial s_k \partial s_l}H(s) &= \frac{2}{(p+1)!}E(\sum_{i<j}(Y_{-k}^i - Y_{-k}^j)(Y_{-l}^i - Y_{-l}^j) \prod_{i'<j',(i',j')\neq(i,j)} (Z_0^{i'}(s) - Z_0^{j'}(s))^2) \\
&+ \frac{4}{(p+1)!}E(\sum_{i<j,i'<j',(i',j')\neq(i,j)} (Y_{-k}^i - Y_{-k}^j)(Z_0^i(s) - Z_0^j(s))(Y_{-l}^{i'} - Y_{-l}^{j'}) \\
&(Z_0^{i'}(s) - Z_0^{j'}(s)) \prod_{i"<j",(i",j")\neq(i',j') and \neq(i,j)} (Z_0^{i"}(s) - Z_0^{j"}(s))^2)
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial^2}{\partial s_k \partial s_l}H(\theta) &= \frac{2}{(p+1)!}E(\sum_{i<j}(Y_{-k}^i - Y_{-k}^j)(Y_{-l}^i - Y_{-l}^j) \prod_{i'<j',(i',j')\neq(i,j)} (X_0^{i'} - X_0^{j'})^2) \\
&+ \frac{4}{(p+1)!}E(\sum_{i<j,i'<j',(i',j')\neq(i,j)} (Y_{-k}^i - Y_{-k}^j)(Y_{-l}^{i'} - Y_{-l}^{j'})(X_0^{i'} - X_0^{j'})(X_0^i - X_0^j) \\
&\prod_{i"<j",(i",j")\neq(i',j') and \neq(i,j)} (X_0^{i"} - X_0^{j"})^2)
\end{aligned}$$

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

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

(which is also easily seen from the fact that the $p+1$ variables X_0^i take value in the same alphabet with p values). This leads to

$$\frac{\partial^2}{\partial s_k \partial s_l}H(\theta) = \frac{2}{(p+1)!}E(\sum_{i<j}(Y_{-k}^i - Y_{-k}^j)(Y_{-l}^i - Y_{-l}^j) \prod_{i'<j',(i',j')\neq(i,j)} (X_0^{i'} - X_0^{j'})^2)$$

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

Notice that for $v = \lambda \cdot \theta$, where λ is a real number, $v^T \cdot D_2H(\theta) \cdot v = 0$. However, the set Θ^* cannot contain any multiple of delayed θ except 0, either Θ would be ambiguous on scale and delay. Now, $v^T \cdot D_2 H(\theta) \cdot v = 0$ if and only if for any $i < j$, $(Z_0^i(v) - Z_0^j(v)) \prod_{i'<j',(i',j')\neq(i,j)}(X_0^{i'} - X_0^{j'}) = 0$ a.s. Now,

$$(Z_0^i(v) - Z_0^j(v)) \prod_{i'<j',(i',j')\neq(i,j)} (X_0^{i'} - X_0^{j'}) = \sum_{k\neq 0}(u * v)_k P_{-k}$$

where $P_{-k} = (X_{-k}^i - X_{-k}^j)\prod_{i'<j',(i',j')\neq(i,j)}(X_0^{i'} - X_0^{j'})$. The variables P_k have discrete distribution with at least 2 different points of support, and using **(M4)**, any finite trajectory has positive probability. But, 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 on 0, so that $D_2 H(\theta)$ is positive definite on Θ^*.

Let us now study the Toeplitz procedure. Applying Proposition 2.10 we have:

$$H(s) = \frac{2^{p(p+1)/2}}{(p+1)!} E[\prod_{j<k}(1 - \cos(\frac{Z_0^j(s) - Z_0^k(s)}{F}))]$$

so that

$$\frac{\partial}{\partial s_l} H(s) = \frac{2^{p(p+1)/2}}{(p+1)!F} E\left(\sum_{j<k}(Y_{-l}^j - Y_{-l}^k)\sin(\frac{Z_0^j(s) - Z_0^k(s)}{F})) \prod_{j'<k',(j',k')\neq(j,k)} (1 - \cos(\frac{Z_0^{j'}(s) - Z_0^{k'}(s)}{F}))\right)$$

and

$$\begin{aligned}\frac{\partial^2}{\partial s_l \partial s_m} H(s) &= \frac{2^{p(p+1)/2}}{(p+1)!F^2} E(\sum_{j<k}(Y_{-l}^j - Y_{-l}^k)(Y_{-m}^j - Y_{-m}^k)\cos(\frac{Z_0^j(s) - Z_0^k(s)}{F}) \\ &\quad \prod_{j'<k',(j',k')\neq(j,k)} (1 - \cos(\frac{Z_0^{j'}(s) - Z_0^{k'}(s)}{F}))) \\ &\quad + \frac{2^{p(p+1)/2}}{(p+1)!F^2} E(\sum_{j<k,j'<k',(j',k')\neq(j,k)} (Y_{-l}^j - Y_{-l}^k)(Y_{-l}^{j'} - Y_{-l}^{k'})\sin(\frac{Z_0^j(s) - Z_0^k(s)}{F}) \\ &\quad \sin(\frac{Z_0^{j'}(s) - Z_0^{k'}(s)}{F})) \prod_{j"<k",(j",k")\neq(j,k) and (i',j')} (1 - \cos(\frac{Z_0^{j"}(s) - Z_0^{k"}(s)}{F})))\end{aligned}$$

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

$$\begin{aligned}\frac{\partial^2}{\partial s_l \partial s_m} H(\theta) &= \frac{2^{p(p+1)/2}}{(p+1)!F^2} E(\sum_{j<k}(Y_{-l}^j - Y_{-l}^k)(Y_{-m}^j - Y_{-m}^k)\cos(\frac{X_0^j - X_0^k}{F}) \\ &\quad \prod_{j'<k',(j',k')\neq(j,k)} (1 - \cos(\frac{X_0^{j'} - X_0^{k'}}{F}))) \\ &\quad + \frac{2^{p(p+1)/2}}{(p+1)!F^2} E(\sum_{j<k,j'<k',(j',k')\neq(j,k)} (Y_{-l}^j - Y_{-l}^k)\sin(\frac{X_0^j - X_0^k}{F}) \\ &\quad (Y_{-l}^{j'} - Y_{-l}^{k'})\sin(\frac{X_0^{j'} - X_0^{k'}}{F}) \prod_{j"<k",(j",k")\neq(j,k) and (i',j')} (1 - \cos(\frac{X_0^{j"} - X_0^{k"}}{F})))\end{aligned}$$

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

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

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

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

This leads to

$$\begin{aligned} \frac{\partial^2}{\partial s_l \partial s_m} H(\theta) &= \frac{2^{p(p+1)/2}}{(p+1)!F^2} E(\sum_{j<k}(Y_{-l}^j - Y_{-l}^k)(Y_{-m}^j - Y_{-m}^k) & (11) \\ & \prod_{j'<k',(j',k')\neq(j,k)} (1-\cos(\frac{X_0^{j'} - X_0^{k'}}{F}))) & (12) \end{aligned}$$

which applied to v leads to the formula of Proposition 3.3.
Now, $v^T \cdot D_2 H(\theta) \cdot v = 0$ if and only if for any $i<j$,

$$(Z_0^j(v) - Z_0^k(v))^2 \prod_{j'<k',(j',k')\neq(j,k)} (1-\cos(\frac{X_0^{j'} - X_0^{k'}}{F})) = 0$$

a.s. But $Z_0^j(v) - Z_0^k(v) = \sum_l (u*v)_l (X_{-l}^j - X_0^{-l})$. Now, on the event where $\prod_{j'<k',(j',k')\neq(j,k)}(1-\cos(\frac{X_0^{j'}-X_0^{k'}}{F})) \neq 0$, $X_0^j = X_0^k$ 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 $Z_0^j(v) - Z_0^k(v)$ cannot be alwways 0, so that $D_2 H(\theta)$ is positive definite on Θ^*.

Proof of Lemma 4.1.
First of all, the ergodicity of (X_t) allows 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 $D^1 H(\theta,\sigma_0)$. To compute the derivatives of H with respect to ξ or σ, recall for the Hankel procedure the following formula, which may be found in [2]:

$$\forall \sigma \leq \sigma_0,\ \forall s,\ \forall i,\ \ c^i(s,\sigma) = E[(Y_0(s) + \sqrt{\sigma_0^2 - \sigma^2}\epsilon_0(s))^i]$$

We may now use formula (2.8) in the same way as when studying the definiteness of $D_s^2 H(\theta)$. Let for $i=0,\ldots,p$, ϵ^i be p independent random sequence of independent variables, independent from the $(X_t^i)_{i=0,\ldots,p,t\in\mathbb{Z}}$, with standard gaussian distribution. We then have, for all ξ and all $\sigma<\sigma_0$:

$$H(\theta(\xi),\sigma) = \frac{1}{(p+1)!} E\left[\prod_{i<j}(Y_0^i(\theta(\xi)) - Y_0^j(\theta(\xi)) + \sqrt{\sigma_0^2-\sigma^2}(\epsilon_0^i(\theta(\xi)) - \epsilon_0^j(\theta(\xi))))^2\right]$$

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 to 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

$$H(\theta(\xi), \sigma) = \frac{2^{p(p+1)/2}}{(p+1)!} E\left[\prod_{i<j} (1 - \cos(\frac{Y_0^i(\theta(\xi)) - Y_0^j(\theta(\xi)) + \sqrt{\sigma_0^2 - \sigma^2}(\epsilon_0^i(\theta(\xi)) - \epsilon_0^j(\theta(\xi))))}{F}))\right]$$

This leads to:

$$D_\xi^1 H(\theta, \sigma_0) = 0$$

$$D_\xi^2 H(\theta, \sigma_0) = C$$

and

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

for the Hankel procedure, and

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

for the Toeplitz procedure.
Now we have if x is any of the variables ξ_j:

$$D_x^1 H_n(\theta, \sigma_0) = (D_c^1 h(c_n(\theta, \sigma_0)) - D_c^1 h(c(\theta, \sigma_0))) \cdot D_x^1 c_n(\theta, \sigma_0) + D_c^1 h(c(\theta, \sigma_0)) \cdot (D_x^1 c_n(\theta, \sigma_0) - D_x^1 c(\theta, \sigma_0)) \tag{13}$$

and also

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

Notice now that

$$c_n(\theta, \sigma_0) = K(\theta, \sigma_0) \cdot m_n(\xi^*) \tag{14}$$

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

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

where the $o(1)$ are in probability. Using (14) we have also

$$D_x^1 c_n(\theta, \sigma_0) = K(\theta, \sigma_0) \cdot D_x^1 M_n(\xi) + D_x^1 K(\theta, \sigma_0) \cdot M_n(\xi) + o(\frac{1}{\sqrt{n}})$$

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 $D_x^1 c_n(\theta, \sigma_0)$. Applying those results to (13) and using the central limit theorem given by **(M8)** leads to the asymptotic distribution for $\sqrt{n} D_\xi^1 H_n(\theta, \sigma_0)$ given in Lemma 4.1.
Similarly, expanding also $H_n(\theta, \sigma_0)$ as

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

allows 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.

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

$$D_\xi^2 H(\theta, \sigma_0) = C,$$

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

converges in distribution to some random variable, and also $(D_n)_{12}$, 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 non zero 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 equation (10) lead to

$$(\hat{\xi}-\xi^*) = \left((D_n)^{11} D_\xi^1 H_n(\theta,\sigma_0) - (D_n)^{11} \cdot \frac{(D_n)_{12}}{(D_n)_{22}}(D_\sigma^1 H_n(\theta,\sigma_0) + \frac{\delta(n)^2}{2H_n(\theta,\sigma_0)}\right)(1+o(1)) \tag{15}$$

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

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

Moreover,

$$\frac{(D_n)_{12}}{(D_n)_{22}}$$

converges to 0 in probability, and

$$\frac{(D_n)_{12}}{(D_n)_{22}H_n(\theta,\sigma_0)} \cdot (\delta(n))^2$$

converges also to 0 in probability, so that equation (15) now leads to the Theorem.

References

[1] Gamboa F, Gassiat E. Blind deconvolution of discrete linear systems. *Annals of Stat.*, (to appear) 1996.

[2] Gassiat E, Gautherat E. Identification of noisy linear systems with discrete random input. *I.E.E.E. Trans. Info. Theory*, 1996 (To appear).

[3] Krein M G, Nudel'man AA. *The Markov moment problem and extremal problems.* Am. Math. Soc., 1977.

[4] Li T H. Blind deconvolution of linear systems with multilevel nonstationary inputs. *Annals of Stat.*, 23:690–704, 1995.

[5] Lindsay, B.G. On the determinants of moment matrices. *Annals of Stat.*, 17:711–721, 1989.