

Adaptive estimation of the spectral density of a weakly or strongly dependent Gaussian process

Philippe Soulier

Département de Mathématiques,
Université d'Evry Val d'Essonne,
91025 Evry cedex, France

Keywords : Gaussian processes, Long range dependence, Spectral density, Adaptive estimation.

AMS Subject Classification 62G05 62G08 62M15

Abstract

This paper deals with estimation of the spectral density $f(x) = |1 - e^{ix}|^{-2d} f^*(x)$, of a stationary fractional Gaussian process, where $-1/2 < d < 1/2$ and f^* is positive. The optimal rate of convergence of an estimate of f is shown not to depend on d but only on the smoothness of f^* , and thus is the same for a long range ($d > 0$) and a short range dependent ($d = 0$) process. When the Fourier coefficients of f^* decrease exponentially fast, the exact asymptotic behaviour of the minimax risk is obtained. The log-periodogram regression estimate is shown to achieve the best possible rate of convergence when the smoothness of f^* is known, and to have adaptivity property when this smoothness is unknown.

1 Introduction

Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary Gaussian process with covariance function $\gamma_X(t) = \mathbb{E}[X_0 X_t]$. The spectral density f_X of the process X is characterized, if it exists, by the relation

$$\hat{f}_X(t) = \int_{-\pi}^{\pi} f(x) e^{itx} dx = \gamma_X(t).$$

A Gaussian process is usually said weakly dependent when its covariance function is absolutely summable. This implies that the spectral density is continuous and bounded.

The probabilistic and statistical theory for such weakly dependent Gaussian processes is well established. More recently, models with non summable covariance function have been considered. Such processes are called by contrast strongly dependent processes. The best known class of fractional processes is the class of fractionally integrated ARMA processes, usually referred to as ARFIMA(p, d, q) processes, obtained by fractional differentiation of a causally invertible ARMA process. It was first introduced by Adenstedt (1974), and popularized by Granger and Joyeux (1980) and Hosking (1981). Recall that an ARMA(p, q) process Y is defined by the recurrence equation

$$Q(B)Y = P(B)\epsilon, \quad (1.1)$$

where B is the backshift operator, ϵ is a white noise sequence with zero mean and variance σ^2 , and P and Q are mutually prime polynomials such that $P(0) = Q(0) = 1$. If Q has no roots inside the closed unit disk, then (1.1) has a stationary solution Y whose spectral density f_Y can be expressed as

$$f_Y(x) = \frac{\sigma^2}{2\pi} |P(e^{ix})/Q(e^{ix})|^2.$$

If moreover P has no roots inside the closed unit disk, then Y is causal and invertible. Given a causal and invertible ARMA(p, q) process Y , and a real number $d \in (-1/2, 1/2)$, an ARFIMA(p, d, q) process X is defined by

$$X = (I - B)^{-d}Y,$$

where the fractional differencing operator $(I - B)^{-d}$ is defined for $d \neq 0$ by the infinite series

$$(I - B)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(d+j)}{j! \Gamma(d)} B^j.$$

This series is summable for $d < 0$ and square summable for $d \in (0, 1/2)$. The spectral density of X is given by

$$f_X(x) = |1 - e^{ix}|^{-2d} f_Y(x) = \frac{\sigma^2}{2\pi} |1 - e^{ix}|^{-2d} |P(e^{ix})/Q(e^{ix})|^2.$$

A natural extension of the class of ARFIMA processes consists in assuming that the spectral density of the process X can be expressed as

$$f_X(x) = |1 - e^{ix}|^{-2d} f^*(x), \quad (1.2)$$

where $d \in]-1/2, 1/2[$ and no constraint is set on f^* apart from positivity and a certain degree of smoothness on $[-\pi, \pi]$. Such processes are often called fractional processes since they can be obtained by fractional differentiation of a weakly dependent process with spectral density f^* . They are strongly dependent when $d > 0$. When $d < 0$, then f is bounded but vanishes at zero, and this significantly changes the nature of the process

X . In both cases, the autocovariance coefficients $\gamma_X(\tau)$ decay hyperbolically at the rate τ^{-1+2d} instead of decaying exponentially fast as is the case for the autocovariance of an ARMA process. Hence, if $d \neq 0$, the rate of decay of the autocovariance function is entirely determined by d . Nevertheless, the smoothness of f^* still determine the main statistical properties of the process X . In particular, Giraitis, Robinson and Samarov (1997) and Iouditsky, Moulines and Soulier (2001) have shown that the rate of convergence of estimates of d depends only on the smoothness of f^* .

This paper is concerned with minimax and adaptive nonparametric estimation of the spectral density of a fractional Gaussian process with spectral density given by (1.2). In the weak dependence context (*i.e.* $d = 0$), these problems have been considered by several authors, see for instance Efremovich and Pinsker (1982) Golubev (1993), Efremovich (1998) and Comte (1999). In this context, the measure of accuracy of estimation is the L^2 risk. In the context of long range dependence, this risk cannot be considered, since the spectral density is not square integrable if $d > 1/4$. Hence an alternate measure of accuracy of estimation is needed. In the signal processing literature, it is customary to consider the log-spectrum (often referred to as the *cepstrum*) instead of the spectral density, and to use the logarithmic risk function, defined for positive functions f and h by

$$R(f, h) = \int_{-\pi}^{\pi} \{\log(f(x)) - \log(h(x))\}^2 dx = \|\log(f) - \log(h)\|^2.$$

To assess the performance of the estimate, we now define relevant functional classes and give lower bounds for the estimation risk over these classes. The most natural family of functional classes in time series analysis is the class of spectral density functions with exponentially decaying Fourier coefficients. Since spectral densities are even functions, it is enough to consider the normalized cosine basis defined as

$$h_0 = 1/\sqrt{2\pi}, \quad h_j(x) = \cos(jx)/\sqrt{\pi}, \quad j \geq 1.$$

For $\beta > 0$ and $L > 0$, define

$$\mathcal{A}(\beta, L) = \{h = \sum_{j=0}^{\infty} \theta_j h_j, \sum_{j=0}^{\infty} e^{2\beta j} \theta_j^2 \leq L^2\}.$$

This class is important since it is related to the ARMA and ARFIMA processes. If Y is an ARMA(p, q) process given by (1.1) and such that P and Q have no roots inside a closed disk centered at zero and with radius β , then its spectral density f_Y belongs to $\mathcal{A}(\beta, L)$ for a relevant choice of the constant L , and so does $\log(f_Y)$ since f_Y is strictly positive. The spectral density of an ARFIMA process X can then be expressed as $f_X = |1 - e^{ix}|^{-2d} f^*$ where f^* and $\log(f^*)$ belong to $\mathcal{A}(\beta, L)$ for some $\beta > 0$ and $L > 0$.

In minimax estimation, it is also customary to consider Sobolev ellipsoids, that is, classes defined, for $\beta > 0$ and $L > 0$, as follows,

$$\mathcal{S}(\beta, L) = \{h = \sum_{j=0}^{\infty} \theta_j h_j, \theta_0^2 + \sum_{j=1}^{\infty} j^{2\beta} \theta_j^2 \leq L^2\}.$$

In the context of spectral density estimation for fractional processes these classes are illustrated by the following example. Let X be a stationary invertible ARFIMA(p, d, q) process with $d > 0$ and spectral density $f_X(x) = |1 - e^{ix}|^{-2d} f_X^*(x)$, and let ϵ be a Gaussian white noise with variance σ^2 . Assume that the observed process is actually $\xi = X + \epsilon$. In that case, the spectral density of ξ is $f_\xi(x) = |1 - e^{ix}|^{-2d} f_\xi^*(x)$ with

$$f_\xi^*(x) = f_X^*(x) + \frac{\sigma^2}{2\pi} |1 - e^{ix}|^{2d},$$

and the regularity of f_ξ^* then depends on d . More precisely, if f_X^* is bounded away from zero, then $\log(f_\xi^*) \in \mathcal{S}(\beta, L)$, for any $\beta < 1/2 + 2d$, and for a suitable choice of L .

For these functional classes, the following lower bounds hold, which are proved in section 6

Theorem 1 For any $\gamma > 0$, $\beta > 1/2$, $L > 0$ and $\delta \in [0, 1/2[$,

$$\lim_{n \rightarrow \infty} \inf_{\hat{l}_n} \sup_{|d| \leq \delta} \sup_{l^* \in \mathcal{A}(\gamma, L)} \frac{n}{\log(n)} \mathbb{E}[\|\hat{l}_n - l\|^2] \geq 2\pi/\gamma, \quad (1.3)$$

$$\liminf_n \inf_{\hat{l}_n} \sup_{|d| \leq \delta} \sup_{l^* \in \mathcal{S}(\beta, L)} n^{\frac{2\beta}{2\beta+1}} \mathbb{E}[\|\hat{l}_n - l\|^2] \geq \mathcal{P}(\beta, L), \quad (1.4)$$

where the infimum is evaluated for all estimates \hat{l}_n of the log-spectrum $l = dg + l^*$ of a stationary process with spectral density $f = e^l$, based on n observations and $\mathcal{P}(\beta, L) = (4\pi\beta/(\beta+1))^{2\beta/(2\beta+1)} \{L^2(2\beta+1)\}^{1/(2\beta+1)}$.

Remark These lower bounds are actually proved in the case $d = 0$, and the proof is similar to the proof of the lower bound for the minimax risk over an ellipsoid in the following regression problem

$$\xi_{n,t} = l^*(t/n) + \eta_{n,t}, \quad 1 \leq t \leq n,$$

where $(\eta_{n,t})_{1 \leq t \leq n}$ is a triangular array of i.i.d. r.v.'s with zero mean and variance 2π and l^* is an even function on $[0, 1]$. For this problem, $\mathcal{P}(\beta, L)$ is known as Pinsker's constant (cf. Belitser, 2000).

2 Log-periodogram regression

In this section, we present an estimate of the log-spectrum $l = \log(f)$ of a fractional Gaussian process. Let $(h_{n,k})$, $1 \leq k \leq n$, be a sequence of complex numbers to be precised

later, and define $H_n^2 = \sum_{k=1}^n |h_{n,k}|^2$. The tapered discrete Fourier transform (DFT) and periodogram are defined as

$$d_n(x) = (2\pi H_n^2)^{-1/2} \sum_{k=1}^n h_{n,k} X_k e^{ikx}, \quad I_n(x) = |d_n(x)|^2.$$

Choosing $h_k = 1$, $1 \leq k \leq n$, yields $H_n^2 = n$ and $I_n(x)$ is the ordinary periodogram. Let $x_k = 2k\pi/n$, $1 \leq k \leq \tilde{n} := [(n-1)/2]$ be the so-called Fourier frequencies. The ordinary periodogram ordinates $(I_n(x_k))_{1 \leq k \leq \tilde{n}}$ have some well known properties which we recall now. The most elementary one is that they are shift invariant, since $\sum_{t=1}^n e^{itx_k} = 0$ for $k = 1, \dots, \tilde{n}$. If the process X is Gaussian white noise, then the ordinary periodogram ordinates are \tilde{n} i.i.d standard exponential random variables. If X is a Gaussian process with smooth spectral density f , then for a fixed number u of Fourier frequencies, the vector of ordinary periodogram ordinates $I_n(x_k)/f(x_{k_1}), \dots, I_n(x_u)/f(x_{k_u})$ converge in distribution to a vector whose components are u i.i.d standard exponential random variables. In this sense, it is often said that periodogram ordinates at Fourier frequencies are asymptotically i.i.d standard exponential random variables. In the case of a fractional process, this last property no longer holds and it has been shown that the vector $I_n(x_k)/f(x_{k_1}), \dots, I_n(x_u)/f(x_{k_u})$ converges in distribution to a vector whose components are neither independent nor identically distributed. More precisely, the asymptotic distribution can be represented as $Z_1^2 + Z_2^2, \dots, Z_{2u-1}^2 + Z_{2u}^2$, where Z_1, \dots, Z_{2u} are $2u$ correlated jointly Gaussian random variables : see for instance Hurvich and Beltrao (1993) and Terrin and Hurvich (1994).

Data tapering has been used for quite a long time in time series analysis (see Tukey (1967) for an early reference). In the long range dependence context, it was introduced by Velasco (1999a,b) to deal with non stationary time series and by Giraitis, Robinson and Samarov (2000) to reduce correlation between discrete Fourier transforms computed at Fourier frequencies. Hurvich and Chen (2000) have introduced a family of complex valued tapers which are well suited to the analysis of fractional processes. We will consider here the simplest of these, defined by

$$h_{n,t} = 1 - e^{2i\pi t/n}, \quad 1 \leq t \leq n. \quad (2.1)$$

This data taper has the desirable property of being shift invariant, contrarily to the Kolmogorov tapers used by Velasco (1999a,b). An adverse effect of tapering is that consecutive tapered DFT ordinates are correlated, even if the observed process is a Gaussian white noise, and this correlation does not vanish asymptotically. The taper used here nevertheless preserves this orthogonality property of non consecutive tapered DFT ordinates, and non consecutive tapered DFT ordinates are significantly less correlated than in the non tapered case (see section 5). This is very important when the observed process is a fractional process. Hence, when tapering is performed, only half of the periodogram ordinates will be used to define the estimates. This results in a twofold increase of the asymptotic variance of these estimates. This is better than when using the cosine bell taper as in

Giraitis, Robinson and Samarov (2000), which results in a threefold increase, because in that case only one out of three Fourier frequencies is used. This efficiency loss can be partially be compensated by aggregation or pooling of periodogram ordinates. This idea was first introduced in the context of long range dependence by Robinson (1995). For a fixed integer $m \geq 1$, define

$$\bar{I}_{n,k} = \begin{cases} \sum_{t=m(k-1)+1}^{mk} I_n(x_t), & 1 \leq k \leq [(n-m)/2m] \quad (\text{ordinary periodogram}), \\ \sum_{t=m(k-1)+1}^{mk} I_n(x_{2t}), & 1 \leq k \leq [(n-2m)/4m] \quad (\text{tapered periodogram}). \end{cases}$$

In order to simplify the notations and unify the presentation of the estimates, we will denote $K_n = [(n-m)/2m]$ in the case of the ordinary (non-tapered) periodogram and $K_n = [(n-2m)/4m]$ in the case of the tapered periodogram.

If X were a Gaussian white noise, then f_X would be a constant and for any sequence of frequencies $(y_k)_{1 \leq k \leq K_n}$, the distribution of $\bar{I}_{n,k}/f(y_k)$ would be the $\Gamma(m, 1)$ distribution, that is the distribution of the sum of m i.i.d. standard exponential r.v.'s. Recall that if Y is a $\Gamma(m, 1)$ random variable, then $\mathbb{E}[\log(Y)] = \psi(m)$ and $\text{var}(\log(Y)) = \psi'(m)$, where $\psi(z) = \Gamma'(z)/\Gamma(z)$ is the digamma function (see Johnson, Kotz and Balakrishnan (1995)). For instance, $-\psi(1)$ is Euler's constant and $\psi'(1) = \pi^2/6$. Define $g(x) = -2\log(|1 - e^{ix}|)$, and for $k = 1, \dots, K_n$, $y_k = (2k-1)\pi/2K_n$. For any process X , denote $Y_{n,k} = \log(\bar{I}_{n,k}) - \psi(m)$, and $\epsilon_{n,k} = \log(\bar{I}_{n,k}/f(y_k)) - \psi(m)$. If X is a fractional process with spectral density as in (1.2) : $f(x) = |1 - e^{ix}|^{-2d} f^*(x)$, where f^* is a positive function on $[-\pi, \pi]$, define $l = \log(f)$ and $l^* = \log(f^*)$. With these notations, we can write

$$Y_{n,k} = \log(f(y_k)) + \epsilon_{n,k} = dg(y_k) + l^*(y_k) + \epsilon_{n,k}, \quad 1 \leq k \leq K_n. \quad (2.2)$$

As already mentionned, the heuristic claim behind spectral methods is that the renormalized periodogram ordinates can be replaced by i.i.d. standard exponential random variables, hence the sequence $(\epsilon_{n,k})_{1 \leq k \leq K_n}$ can be considered as an i.i.d. sequence with zero mean and variance $\psi'(m)$. Even though this is far from being true, the conclusion derived from this heuristic claim can be justified rigorously when f^* , or equivalently l^* , is smooth enough. More precisely it will be assumed that $l^* \in \mathcal{A}(\gamma, L)$ for some $\gamma > 0$ or $l^* \in \mathcal{S}(\beta, L)$ for some $\beta > 3/2$. To deal with less smooth spectral densities, we need to introduce the following functional class. For $M \geq 0$, let $\mathcal{L}^*(M)$ be the class of even periodic functions l^* on $[-\pi, \pi]$ such that

$$\sup_{x \in [-\pi, \pi]} |l^*(x)| \leq M \quad \text{and} \quad \sup_{x, y \in [-\pi, \pi] \setminus \{0\}} |l^*(x) - l^*(y)| \leq M \frac{|x - y|}{|x| \wedge |y|}.$$

When a given function l^* belongs to $\mathcal{L}^*(M)$, then suitable bounds can be obtained. Unfortunately, the class $\mathcal{L}^*(M)$ is not compact since it is not equicontinuous at zero. Hence, to obtain uniform bounds, necessary in the context of minimax estimation, attention must be restricted to compact subclasses of $\mathcal{L}^*(M)$. This is not a practical restriction since the classes of interest are $\mathcal{S}(\beta, L) \cap \mathcal{L}^*(M)$ which are compact.

The interest of these classes in the context of spectral density estimation is that they contain the smooth part of the spectral density of an ARFIMA process X possibly observed with noise. Recall that the spectral density of such a process can be expressed as $f_\xi^*(x) = f^*(x) + \sigma^2/2\pi|1 - e^{ix}|^{2d}$, where f^* is a positive and smooth function, say at least twice continuously differentiable. It is then easily seen that $\log(f_\xi^*) \in \mathcal{L}^*(M)$ for a suitable M . Since moreover $\log(f_\xi^*) \in \mathcal{S}(\beta, L)$ for any $\beta < 1/2 + 2d < 3/2$, it is clear that this functional class is well suited to our purpose. Finally, note that for $\beta > 3/2$, $\mathcal{S}(\beta, L) \subset \mathcal{L}^*(M)$ for a relevant choice of L .

We now return to the log-periodogram regression. Whether valid or not, the approximation of $(\epsilon_{n,k})_{1 \leq k \leq K_n}$ by an i.i.d. sequence is the ground for log-periodogram regression which consists in estimating the coefficients $d, \theta_0, \dots, \theta_{p-1}$ by ordinary least squares :

$$(\hat{d}, \hat{\theta}_0, \dots, \hat{\theta}_{p-1}) = \arg \min_{\bar{d}, \bar{\theta}_0, \dots, \bar{\theta}_{p-1}} \sum_{k=1}^{K_n} \left(\log(\bar{I}_{n,k}) - \psi(m) - \bar{d}g(y_k) - \sum_{j=0}^{p-1} \bar{\theta}_j h_j(y_k) \right)^2.$$

For a given p , an estimate of the log-spectrum is then defined as

$$\hat{l}_p = \hat{d}_p g + \sum_{j=0}^{p-1} \hat{\theta}_j h_j.$$

For convenience, the performance of the estimates will be measured with respect to a discretized L^2 norm. For any $u \in \mathbb{R}^{K_n}$, define

$$\|u\|_n^2 = \frac{2\pi}{K_n} \sum_{k=1}^{K_n} u_k^2,$$

and a function ϕ will be identified with the vector $(\phi(y_1), \dots, \phi(y_{K_n}))^T$.

Theorem 2 *Let $\gamma > 0$, $\beta > 1/2$, $L > 0$, $M > 0$ and $\delta \in [0, 1/2[$. Denote $p_n(\gamma) = \lfloor \log(n)/2\gamma \rfloor$ and $p_n(L, \beta) = \lfloor L^{\frac{2}{2\beta+1}} n^{\frac{1}{2\beta+1}} \rfloor$. Let X be a stationary Gaussian process with log-spectrum $l = dg + l^*$ and let \hat{l}_p is the log-periodogram regression estimate of the l based either on the ordinary or the tapered periodogram. Then,*

$$\lim_{n \rightarrow \infty} \sup_{|d| \leq \delta} \sup_{l^* \in \mathcal{A}(\gamma, L)} \frac{n}{\log(n)} \mathbb{E}[\|\hat{l}_{p_n(\gamma)} - l\|_n^2] = 2\pi \varrho m \psi'(m)/\gamma, \quad (2.3)$$

$$\lim_{n \rightarrow \infty} \sup_{|d| \leq \delta} \sup_{l^* \in \mathcal{S}(\beta, L) \cap \mathcal{L}^*(M)} L^{-\frac{2}{2\beta+1}} n^{\frac{2\beta}{2\beta+1}} \mathbb{E}[\|\hat{l}_{p_n(L, \beta)} - l\|_n^2] \leq 4\pi \varrho m \psi'(m) + 1, \quad (2.4)$$

where $\varrho = 1$ when the ordinary periodogram is used and $\varrho = 2$ when the tapered periodogram is used.

Remarks

- As already mentioned, for $\beta > 3/2$, the condition $l^* \in \mathcal{L}^*(M)$ is not a restriction.
- Even though the estimate based on the tapered periodogram is inefficient in the minimax sense, we mention its performance here since it is used in the adaptive framework.
- As can be seen in the proof of Theorem 1 the bounds (1.3) and (1.4) hold for the norm $\|\cdot\|_n$ as well as for the L^2 -norm. Hence the log-periodogram estimate, whether based on the ordinary or tapered periodogram, is minimax rate optimal in the classes $\mathcal{A}(\gamma, L)$ for $\gamma > 0$ and $\mathcal{S}(\beta, L)$ for $\beta > 3/2$. For $\beta \in (1/2, 3/2]$, the log-periodogram estimate is rate optimal over the class $\mathcal{S}(\beta, L) \cap \mathcal{L}^*(M)$, since it is easily seen that the lower bound obtained for $\mathcal{S}(\beta, L)$ holds for $\mathcal{S}(\beta, L) \cap \mathcal{L}^*(M)$. Moreover, when based on the ordinary periodogram, it is efficient up to a multiplicative constant $m\psi'(m)$ over the analytic class $\mathcal{A}(\gamma, L)$. Since m can be chosen so that $m\psi'(m)$ is arbitrarily close to 1, we obtain as a corollary the exact asymptotic behaviour of the minimax risk for the analytic class $\mathcal{A}(\gamma, L)$.

Corollary 2.1 *Let $\gamma > 0$, $L > 0$ and $\delta \in [0, 1/2[$.*

$$\lim_{n \rightarrow \infty} \inf_{\hat{l}_n} \sup_{|d| \leq \delta} \sup_{l^* \in \mathcal{A}(\beta, L)} \frac{n}{\log(n)} \mathbb{E}[\|\hat{l}_n - l\|^2] = 2\pi/\gamma.$$

The situation for the Sobolev class is not clear. If short range dependence is assumed, *i.e.* if it is known that $d = 0$, then the projection estimator can be modified as in the case of the regression with i.i.d. noise, to obtain an efficient estimator. Defining

$$\tilde{l}_n^* = \sum_{j=0}^{q_n} \lambda_j \hat{\theta}_j h_j,$$

where $\lambda_j = 1 - q_n^{-\beta} j^\beta$, $0 \leq j \leq q_n := (L^2(\beta+1)(2\beta+1)n/4\pi\beta)^{\frac{1}{2\beta+1}}$, it can be proved that

$$\lim_{n \rightarrow \infty} \sup_{l^* \in \mathcal{S}(\beta, L) \cap \mathcal{L}^*(M)} n^{\frac{2\beta}{2\beta+1}} \mathbb{E}[\|\tilde{l}_n^* - l^*\|^2] = m\psi'(m)\mathcal{P}(\beta, L). \quad (2.5)$$

This upper bound can be proved as Theorem 3.4 in Belitser (2000), with a few additional technicalities due to the fact that the noise here is not an i.i.d. sequence. Hence we obtain the exact asymptotic minimax risk for the estimation of the log-spectrum of a shortly dependent Gaussian process.

$$\lim_{n \rightarrow \infty} \inf_{\tilde{l}_n^*} \sup_{l^* \in \mathcal{S}(\beta, L) \cap \mathcal{L}^*(M)} n^{\frac{2\beta}{2\beta+1}} \mathbb{E}[\|\tilde{l}_n^* - l^*\|^2] = \mathcal{P}(\beta, L).$$

If d is not assumed to be known and equal to zero, the above construction does not yield an efficient estimator, because of the effect of the function g in the regression. Since our original purpose is to consider only this latter case, we omit the proof of (2.5) and only prove Theorem 2 in section 5.1.

3 Adaptive estimation

The minimax estimates are unfeasible since they depend on the smoothness of the function l^* , which cannot be assumed known, especially in the case of the signal observed in noise where it depends on the parameter d , which is obviously unknown. Hence a data-driven criterion is needed to select the truncation order p which defines the spectral density estimate. Since the problem at hand is very similar to the problem of regression with i.i.d. noise with known variance, it is natural to investigate model selection criterion commonly used in this framework. For a given non decreasing sequence of integers $p_n^* \leq K_n$ and a positive real κ , define

$$S_{p,n}(\kappa) = \|Y_n - \hat{l}_p\|_n^2 + \kappa 2\psi'(m)pK_n^{-1}, \quad (3.1)$$

$$\hat{p}(\kappa) = \arg \min_{1 \leq p \leq p_n^*} S_{p,n}(\kappa). \quad (3.2)$$

The choice $p_n^* = K_n$ and $\kappa = 2$ yields the so-called Mallows' C_L criterion. It has been shown in Moulines et Soulier (2000) that if the ordinary periodogram is used, $S_{p,n}(2)$ is an asymptotically unbiased estimate of the quadratic risk $R_n(p) = \mathbb{E}[\|l - \hat{l}_p\|_n^2]$. If the number of non zero Fourier coefficients of the true l^* is infinite, then $R_n(\hat{p}(2))/\inf_{1 \leq p \leq K_n} R_n(p)$ converges in probability to 1. In the context of regression with i.i.d. noise, this result was proved by Polyak and Tsybakov (1989). This kind of optimality is rather weak, and falls short of providing a risk bound for the data-driven estimate $\hat{l}_{\hat{p}(2)}$. In the context of regression with i.i.d. Gaussian noise, a modification of Mallows C_L has recently been investigated (see Barron, Birge, Massart (1999) for a comprehensive presentation of this method). In these papers, a conservative penalization is used and risk bounds are obtained. The main result of this section is an adaptation to the context of log-periodogram regression of the model selection theory for regression with i.i.d. noise.

Theorem 3 *Let $\delta \in [0, 1/2)$ and $M > 0$. Let \mathcal{F}^* be a compact subclass of $\mathcal{L}^*(M)$. Let X be a stationary Gaussian process with spectral density $f = e^{dg+l^*}$ with $|d| \leq \delta$ and $l^* \in \mathcal{F}^*$. Define $p_n^* = \sqrt{K_n}/\log(K_n)$ and for $\kappa > 0$, define $\hat{p}(\kappa)$ as in (3.2) where \hat{l}_p is the log-periodogram regression estimate of the log-spectrum $l = dg + l^*$ based on the tapered periodogram. There exists a constant $\kappa_m > 0$ (which depends only on m) and a constant $C(\delta, \mathcal{F}^*)$, such that*

$$\mathbb{E}[\|\hat{l}_{\hat{p}(\kappa_m)} - l\|_n^2] \leq 4 \inf_{1 \leq p \leq p_n^*} \{\|l - \Pi_p l\|_n^2 + \kappa_m 2\pi\psi'(m)pK_n^{-1}\} + C(\delta, \mathcal{F}^*)K_n^{-1}.$$

The value of the constant κ_m is a crucial point to implement this method. It appears from the proof of Theorem 3 that a suitable value is $\kappa_m = \{4096(1 + \sqrt{24\pi\psi'(m-1/2)})^2 + (\sqrt{2\pi} + \sqrt{24\pi\psi'(m-1/2)})^2\}/(\pi\psi'(m))$. This is far from being satisfactory, from the theoretical as well as from the practical point of view. Nevertheless, minimax adaptive properties of the data-driven estimate can be inferred from Theorem 3.

Corollary 3.1 *Let $0 < \gamma_* < \gamma^*$, $1/2 < \beta_* < \beta^*$, $L^* > 0$, $M > 0$ and $\delta \in [0, 1/2[$.*

$$\limsup_{n \rightarrow \infty} \sup_{\beta_* < \beta < \beta^*} \sup_{0 < L < L^*} \sup_{|d| \leq \delta} \sup_{l^* \in \mathcal{A}(\gamma, L)} \frac{n}{\log(n)} \mathbb{E}[\|\hat{l}_{\hat{p}(\kappa_m)} - l\|_n^2] < \infty,$$

$$\limsup_{n \rightarrow \infty} \sup_{\gamma_* < \beta < \gamma^*} \sup_{0 < L < L^*} \sup_{|d| \leq \delta} \sup_{l^* \in \mathcal{S}(\beta, L) \cap \mathcal{L}^*(M)} n^{2\beta/(2\beta+1)} \mathbb{E}[\|\hat{l}_{\hat{p}(\kappa_m)} - l\|_n^2] < \infty.$$

We have proved that the adaptive estimator is rate optimal, but we have failed to prove exact adaptation up to the constant, at least in the case of the analytic class. In the weak dependence context, this could probably be done using classical method such as the Stein method. In the strong dependence context, it is unclear that these methods work, because of the technical problems described after theorem 2. Hence we have chosen the model selection approach, which on one hand is inefficient because of the constant κ , but on the other hand is easy to prove since it only involves a standard chaining argument.

4 Conclusion

The objective of this paper was to obtain a data-driven estimator of the spectral density $f = e^{dg+l^*}$ of a Gaussian stationary process, where $d \in (-1/2, 1/2)$ and l^* belongs to a certain functional class \mathcal{F}^* . The case $d = 0$ corresponds to weak dependence and $d > 0$ to strong dependence. We have shown that the log-periodogram regression estimator is rate optimal in various classes, and that its adaptive version is also rate optimal. We have also found the exact asymptotic minimax risk when $l^* \in \mathcal{A}(\gamma, L)$ for some $\gamma > 0$ and $L > 0$, and shown that the log-periodogram regression estimator is quasi-efficient. Some problems remain open.

- The exact asymptotic minimax risk must be computed for the Sobolev ellipsoids $\mathcal{S}(\beta, L)$ when $d \neq 0$. The standard techniques have failed to produce this because of the special role of the function g in the regression.
- Is exact adaptation up to the constant possible when $d \neq 0$?

In order to answer to these questions, it might be of interest to consider first the following regression problem

$$\xi_{n,t} = dg(t/n) + l^*(t/n) + \eta_{n,t}, \quad 1 \leq t \leq n,$$

where $(\eta_{n,t})_{1 \leq t \leq n}$ is a triangular array of i.i.d. r.v.'s with zero mean and known variance and l^* is an even smooth function on $[0, 1]$. The conclusions derived from this model could probably be applied to the problem of log-periodogram regression, even though they are not equivalent since the noise sequence $(\epsilon_{n,k})$ is not asymptotically uncorrelated in the strong dependence context.

5 Technical material

Functionals of the periodogram (tapered or not) of a Gaussian process are functionals of the discrete Fourier transforms which are jointly complex Gaussian vectors. Hence the basic tools to study these functionals are covariance bounds for the discrete Fourier transforms computed at the Fourier frequencies. Such bounds were first rigorously obtained by Robinson (1995a,b) and Giraitis, Robinson and Samarov (1997,2000) in the context of local estimation of the memory parameter d . Under the assumptions of this paper, they were obtained by Moulines and Soulier (1999) in the non tapered case and Hurvich, Moulines and Soulier (2000) in the tapered case. These bounds are then used in connection with general moment bounds for functions of Gaussian vectors. Such bounds have been investigated for a long time, but their interest was revived by the study of long range dependent Gaussian processes, in the time domain as well as in the frequency domain. Moment bounds for functions of Gaussian variables were obtained by Taqqu (1977), and generalizations to Gaussian vectors were obtained by Arcones (1994) and Soulier (1998). In the context of adaptive estimation, exponential inequalities are a fundamental tool. Such inequalities for tapered log-periodogram ordinates were first obtained by Giraitis, Robinson and Samarov (2000) under different assumptions related to local estimation of the memory parameter d , and adapted to the present context by Iouditsky, Moulines and Soulier (2001) and Hurvich, Moulines and Soulier (2000). We summarize the tools used in this paper in the following Theorem, adapted from the previous references.

Theorem 4 *Let $M > 0$, and $\delta \in [0, 1/2)$. Let \mathcal{F}^* be a compact subclass of $\mathcal{L}^*(M)$. Let $(X_t)_{t \in \mathbb{Z}}$ be a stationary Gaussian process with spectral density given $f = e^{dg+l^*}$ with $|d| \leq \delta$ and $l^* \in \mathcal{F}^*$. There exists a constant $C(\delta, \mathcal{F}^*)$ such that for all $1 \leq k < j \leq K_n$,*

$$|\text{cov}(\epsilon_k, \epsilon_j)| \leq C(\delta, \mathcal{F}^*) r_d^2(k, j), \quad (5.1)$$

$$|\mathbb{E}[\epsilon_k]| + |\mathbb{E}[\epsilon_k^2] - \psi'(m)| \leq C(\delta, \mathcal{F}^*) r_d(k, k+1) \quad (5.2)$$

with

$$r_d(j, k) = \begin{cases} \log(j) k^{-|d|} j^{|d|-1} & (\text{ordinary periodogram}) \\ k^{-1} (j-k)^{-2} (j/k)^{|d|}, & (\text{tapered periodogram}) \end{cases}$$

In the case of the tapered periodogram, if $\alpha, u_1, \dots, u_{K_n}$ are real numbers verifying $\sum_{k=1}^{K_n} u_k^2 \leq 1$ and $\alpha \max_{1 \leq k \leq K_n} |u_k| \leq 1/6$, then

$$\mathbb{E}[\exp\{\alpha \sum_{k=1}^{K_n} u_k \epsilon_k\}] \leq C(\delta, \mathcal{F}^*) e^{3\psi'(m-1/2)\alpha^2}. \quad (5.3)$$

5.1 Proof of Theorem 2

Let L_p denote the linear span in \mathbb{R}_n^K of the vectors g, h_0, \dots, h_{p-1} and let Π_p denote the orthogonal projection operator on L_p . Define $Y_n = (Y_{n,1}, \dots, Y_{n,K_n})^T$ and $\epsilon_n =$

$(\epsilon_{n,1}, \dots, \epsilon_{n,K_n})^T$. With these notations, we obtain $Y_n = l + \epsilon_n$ and

$$\hat{l}_p = \Pi_p Y_n = \Pi_p l + \Pi_p \epsilon_n,$$

and, Π_p being an orthogonal projector,

$$\mathbb{E}[\|\hat{l}_p - l\|_n^2] = \|l - \Pi_p l\|_n^2 + \mathbb{E}[\|\Pi_p \epsilon_n\|_n^2].$$

We now bound separately the bias and variance terms above.

Proposition 5.1 *Let $\delta \in [0, 1/2)$ and $M > 0$. Let X be a stationary Gaussian process with spectral density $f = e^{dg+l^*}$ with $|d| \leq \delta$ and $l^* \in \mathcal{L}^*(M)$. There exists a constant $C(\delta, \mathcal{F}^*)$ such that for all $p = 1, \dots, K_n/6$,*

$$|\mathbb{E}[\|\Pi_p \epsilon_n\|_n^2] - 2\pi\psi'(m)pK_n^{-1}| \leq C(\delta, \mathcal{F}^*)p \log^5(K_n)K_n^{-2}.$$

Proof of Proposition 5.1 For a positive integer $p < K_n$, denote $h_{-1}^{(p)} = \|g - \sum_{j=0}^{p-1} \langle g, h_j \rangle h_j\|_n^{-1} \{g - \sum_{j=0}^{p-1} \langle g, h_j \rangle h_j\}$. Note that $h_{-1}^{(p)}$ depends on n and p , but for short, the dependence in n is omitted in the notation. Then $\{h_{-1}^{(p)}, h_0, \dots, h_{p-1}\}$ is an orthonormal basis of L_p . Hence we get

$$\|\Pi_p \epsilon_n\|_n^2 = \sum_{j=-1}^{p-1} \langle h_j, \epsilon_n \rangle^2.$$

For any function u , denote $|u|_\infty = \max_{1 \leq k \leq K_n} |u(y_k)|$. For $j = -1, \dots, p-1$, applying Theorem 4, we get

$$|\mathbb{E}[\langle h_j, \epsilon_n \rangle^2] - 2\pi\psi'(m)K_n^{-1}| \leq C(\delta, \mathcal{F}^*)|h_j|_\infty^2 \log^r(K_n)K_n^{-2},$$

with $r = 3$ in the case of the ordinary periodogram and $r = 1$ in the case of the tapered periodogram. By definition, for $j \geq 0$, $|h_j|_\infty \leq \pi^{-1/2}$, and using the bounds obtained in the proof of Proposition 3.1 in Iouditsky Moulines and Soulier (2001), it is easily seen that for $p \leq K_n/6$,

$$|h_{-1}^{(p)}|_\infty \leq p^{1/2} \log(K). \quad (5.4)$$

The proof of Proposition 5.1 is then concluded by summing the previous bounds.

Proposition 5.2 *For any $L > 0$, $\beta > 1/2$ and $\gamma > 0$,*

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{l^* \in \mathcal{S}(\beta, L)} n^{\frac{2\beta}{2\beta+1}} \|l - \Pi_{p_n(\beta, L)} l\|_n^2 &\leq L^{\frac{2}{2\beta+1}}, \\ \lim_{n \rightarrow \infty} \sup_{l^* \in \mathcal{A}(\gamma, L)} n \log^{-1}(n) \|l - \Pi_{p_n(\gamma)} l\|_n^2 &= 0. \end{aligned}$$

Proof of proposition 5.2 Recall that $\tilde{l}_p = \Pi_p l$ and denote $l_p^* = \sum_{j=p}^{\infty} \theta_j h_j$. By definition, for any $p < K$, $\Pi_p(l - l_p^*) = l - l_p^*$, hence

$$\|l - \tilde{l}_p\|_n = \|l - \Pi_p l\|_n = \|l_p^* - \Pi_p l_p^*\|_n \leq \|l_p^*\|_n \leq \|l_p^*\| + \|\|l_p^*\|_n - \|l_p^*\|\|.$$

If h is an even function defined on $[-\pi, \pi]$, let \bar{h}_n be the even step function defined on $[0, \pi]$ as

$$\bar{h}_n(x) = \sum_{k=1}^{K_n} h(y_k) \mathbf{1}_{\{y_{k-1} \leq x < y_k\}},$$

where we have defined $y_0 = 0$. With this notation, we get $\|h\|_n = \|\bar{h}_n\|$, and

$$\|\|h\|_n - \|h\|\| = \|\|h_n\| - \|h\|\| \leq \|\bar{h}_n - h\|.$$

If $h \in \mathcal{S}(\beta, L)$ for some $\beta > 1/2$ and $L > 0$, then it is known that

$$\|\bar{h}_n - h\|^2 \leq C(\beta, L)n^{-1}.$$

If $l^* \in \mathcal{S}(\beta, L)$, this implies that

$$\|l - \tilde{l}_p\|_n \leq \|l_p^*\| + C(\beta, L)n^{-1} \leq L^2 p^{-2\beta} + C(\beta, L)n^{-1}.$$

Setting $p = \lfloor L^{\frac{2}{2\beta+1}} n^{\frac{1}{2\beta+1}} \rfloor$ yields

$$\|l - \tilde{l}_p\|_n \leq L^{\frac{2}{2\beta+1}} n^{\frac{-2\beta}{2\beta+1}} + C(\beta, L)n^{-1} = L^{\frac{2}{2\beta+1}} n^{\frac{-2\beta}{2\beta+1}} (1 + o(1)).$$

If $l^* \in \mathcal{A}(\gamma, L)$, setting $p = \lfloor \log(n)/2\gamma \rfloor$ yields

$$\|l - \tilde{l}_p\|_n \leq \|l_p^*\| + C(\beta, L)n^{-1} = O(n^{-1}).$$

This concludes the proof of Proposition 5.2. Theorem 2 is a straightforward consequence of Propositions 5.1 and 5.2.

5.2 An exponential inequality

As usual in adaptive estimation, the main tool to the proof of Theorem 3 is an exponential inequality. In the sequel, we consider only the *tapered* periodogram, since the basic tool is the exponential inequality (5.5). Recall that $Z(u) = \langle u, \epsilon_n \rangle = 2\pi K_n^{-1} \sum_{k=1}^{K_n} \epsilon_{n,k}$.

Proposition 5.3 *Let $\delta \in [0, 1/2)$ and $M > 0$. Let \mathcal{F}^* be a compact subclass of $\mathcal{L}^*(M)$. Let X be a stationary Gaussian process with spectral density $f = e^{dg+l^*}$ with $|d| \leq \delta$ and $l^* \in \mathcal{F}^*$. There exists a constant $C(\delta, \mathcal{F}^*)$ such that for all $x > 0$, all $1 \leq q \leq K_n/6$ and all $u \in B_q$, it holds that*

$$\mathbb{P}(Z(u) > \sigma_m \sqrt{x} + r_{q, K_n} x) \leq C(\delta, \mathcal{F}^*) e^{-K_n x}, \quad (5.5)$$

with $\sigma_m^2 = 24\pi\psi'(m-1/2)$ and $r_{q, K} = 4\pi(q+1)^{1/2} \log(K)$.

Proof of Proposition 5.3 Set $\tau_m^2 = 3\psi'(m - 1/2)$. If $\alpha < 1/6|u|_\infty$, (5.3) obviously implies that

$$\mathbb{E}[\exp\{\alpha \sum_{k=1}^{K_n} u_k \epsilon_k\}] \leq C(\delta, \mathcal{F}^*) e^{\tau_m^2 \alpha^2 / (1 - 3\alpha|u|_\infty)}.$$

Hence, for any $\alpha < 1/6|u|_\infty$ and any $\lambda > 0$, it holds that

$$\mathbb{P}\left(\sum_{k=1}^{K_n} u_k \epsilon_k > \lambda\right) \leq C(\delta, \mathcal{F}^*) e^{-\alpha\lambda + \tau_m^2 \alpha^2 / (1 - 3\alpha|u|_\infty)}.$$

It is possible to choose $\alpha = \lambda / (2\tau_m^2 + 6|u|_\infty \lambda) < 1/6|u|_\infty$, and this choice yields, for any $\lambda > 0$,

$$\mathbb{P}\left(\sum_{k=1}^{K_n} u_k \epsilon_k > \lambda\right) \leq C(\delta, \mathcal{F}^*) e^{-\lambda^2 / (4\tau_m^2 + 3|u|_\infty \lambda)}.$$

For $x > 0$, let $\lambda = 2\tau_m \sqrt{x} + 3|u|_\infty x$. The previous bound becomes

$$\mathbb{P}\left(\sum_{k=1}^{K_n} u_k \epsilon_k > 2\tau_m \sqrt{x} + 3|u|_\infty x\right) \leq C(\delta, \mathcal{F}^*) e^{-x}.$$

If $\|v\|_n = 1$, then $\sum_{k=1}^{K_n} v_k^2 = K_n/2\pi$, hence, defining $u = \sqrt{2\pi/K_n}v$, we get $\sum_{k=1}^{K_n} u_k^2 = 1$, $\sqrt{2\pi/K_n}|v|_\infty = |u|_\infty$ and $Z(v) = \sqrt{2\pi/K_n} \sum_{k=1}^{K_n} u_k \epsilon_{n,k}$. Hence

$$\mathbb{P}\left(Z(v) > 2\tau_m \sqrt{2\pi x/K_n} + 6\pi|v|_\infty x/K_n\right) = \mathbb{P}\left(\sum_{k=1}^{K_n} u_k \epsilon_k > 2\tau_m \sqrt{x} + 3|u|_\infty x\right) \leq C(\delta, \mathcal{F}^*) e^{-x}.$$

Setting $y = x/K_n$ yields, for all v such that $\|v\|_n = 1$,

$$\mathbb{P}\left(Z(v) > 2\tau_m \sqrt{2\pi y} + 4\pi|v|_\infty y\right) \leq C(\delta, \mathcal{F}^*) e^{-K_n y}. \quad (5.6)$$

Recall now that $v \in B_q$ means that v is a function in L_q , i.e. v can be expressed as $\sum_{j=-1}^{q-1} \theta_j h_j$, and such that

$$\|v\|_n = \frac{2\pi}{K} \sum_{k=1}^K v^2(y_k) = 1.$$

The orthogonality properties of the functions $h_{-1}^{(q)}, h_0, \dots, h_{q-1}$ imply that $\|v\|_n^2$ can also be expressed as

$$\|v\|_n^2 = \sum_{j=-1}^{q-1} \theta_j^2.$$

Thus, $v \in B_q$ implies that $\sum_{j=-1}^{q-1} \theta_j^2 = 1$, and by Hölder inequality, $\sum_{j=-1}^{q-1} |\theta_j| \leq (q+1)^{1/2}$. Applying (5.4) and the fact that if $K_n \geq 2$, then $\log(K_n) \geq \pi^{-1/2}$, we now obtain

$$\begin{aligned} |v|_\infty &\leq \sum_{j=-1}^{q-1} |\theta_j| |h_j|_\infty \leq |\theta_{-1}| |h_{-1}^{(q)}|_\infty + \pi^{-1/2} \sum_{j=0}^{q-1} |\theta_j| \\ &\leq \log(K_n) \sum_{j=-1}^{q-1} |\theta_j| \leq \log(K_n) (q+1)^{1/2}. \end{aligned}$$

Recalling that $\sigma_m^2 = 24\pi\psi'(m-1/2) = 8\pi\tau_m^2$ and $r_{q,K} = 6\pi(q+1)^{1/2} \log(K)$, we get for all $v \in B_q$,

$$\mathbb{P}(Z(v) > \sigma_m \sqrt{y} + r_{q,K_n} y) \leq \mathbb{P}\left(Z(v) > 2\tau_m \sqrt{2\pi y} + 6\pi |v|_\infty y\right) \leq C(\delta, \mathcal{F}^*) e^{-K_n y}.$$

This concludes the proof of Proposition 5.3.

5.3 Proof of Theorem 3

The proof of Theorem 3 is adapted from the proof of similar results in the context of regression with i.i.d. noise (see for instance the proof of Theorem 3.1 in Baraud *et al.* (1999)). For short, denote $\hat{p} = \hat{p}(\kappa)$, $\text{pen}(p) = 2\pi\kappa\psi'(m)p/K_n$, and $\tilde{l}_p = \Pi_p l$. Denote \langle, \rangle the scalar product associated to the norm $\|\cdot\|_n$ (we omit the index n in the notation). Since for any fixed p , \hat{l}_p is the OLS estimate of $dg + \sum_{j=0}^{p-1} \theta_j h_j$, by definition of \hat{p} , for all $1 \leq p \leq p_n^*$, it holds that

$$\|Y_n - \hat{l}_{\hat{p}}\|_n^2 + \text{pen}(\hat{p}) \leq \|Y_n - \hat{l}_p\|_n^2 + \text{pen}(p) \leq \|Y_n - \tilde{l}_p\|_n^2 + \text{pen}(p).$$

Since $Y_n = l + \epsilon_n$, it follows that

$$\begin{aligned} \|l - \hat{l}_{\hat{p}}\|_n^2 &\leq \|l - \tilde{l}_p\|_n^2 + 2\langle \epsilon_n, \tilde{l}_p - \hat{l}_{\hat{p}} \rangle + \text{pen}(p) - \text{pen}(\hat{p}), \\ &= \|l - \tilde{l}_p\|_n^2 + 2\langle \epsilon_n, l - \hat{l}_{\hat{p}} \rangle + 2\langle \epsilon_n, \tilde{l}_p - \hat{l}_{\hat{p}} \rangle + 2\langle \epsilon_n, \tilde{l}_p - l \rangle + \text{pen}(p) - \text{pen}(\hat{p}) \\ &\leq \|l - \tilde{l}_p\|_n^2 + 2\|\hat{l}_{\hat{p}} - \tilde{l}_p\|_n Z(\hat{u}) + 2Z(l - \hat{l}_{\hat{p}}) - 2Z(l - \tilde{l}_p) + \text{pen}(p) - \text{pen}(\hat{p}), \end{aligned}$$

where we have defined for all $u \in \mathbb{R}^{K_n}$ (or any function u identified to a K_n -dimensional vector as explained above) $Z(u) = \langle u, \epsilon_n \rangle$ and $w(\hat{p}) = \|l_{\hat{p}} - \hat{l}_{\hat{p}}\|_n^{-1} (l_{\hat{p}} - \hat{l}_{\hat{p}})$. Note that by definition, $\|w(\hat{p})\|_n = 1$. For any $\alpha > 0$, it holds that $2ab \leq \alpha a^2 + \alpha^{-1} b^2$. Applying this inequality to $2\|\hat{l}_{\hat{p}} - \tilde{l}_p\|_n Z(w(\hat{p}))$ yields

$$\begin{aligned} \|l - \hat{l}_{\hat{p}}\|_n^2 &\leq \|l - \tilde{l}_p\|_n^2 + \alpha \|l_{\hat{p}} - \hat{l}_{\hat{p}}\|_n^2 + \alpha^{-1} Z^2(w(\hat{p})) \\ &\quad + 2Z(l - \hat{l}_{\hat{p}}) - 2Z(l - \tilde{l}_p) + \text{pen}(p) - \text{pen}(\hat{p}) \\ &= \|l - \tilde{l}_p\|_n^2 + \alpha \|l - \hat{l}_{\hat{p}}\|_n^2 - \alpha \|l - \tilde{l}_p\|_n^2 + \alpha^{-1} Z^2(w(\hat{p})) \\ &\quad + 2Z(l - \hat{l}_{\hat{p}}) - 2Z(l - \tilde{l}_p) + \text{pen}(p) - \text{pen}(\hat{p}), \end{aligned}$$

whence

$$(1 - \alpha)\|l - \hat{l}_{\hat{p}}\|_n^2 \leq \|l - \tilde{l}_p\|_n^2 + \text{pen}(p) + \alpha^{-1}Z^2(w(\hat{p})) \\ + 2Z(l - l_{\hat{p}}) - \alpha\|l - l_{\hat{p}}\|_n^2 - 2Z(l - \tilde{l}_p) - \text{pen}(\hat{p}). \quad (5.7)$$

We will prove in Lemmas 5.1, 5.2 and 5.3 below that there exist constants $C(\delta, \mathcal{F}^*)$, C_m^* and D_m^* (which depend only on their arguments) such that

$$\mathbb{E}[\{Z^2(w(\hat{p})) - C_m^*\hat{p}K_n^{-1}\}^+] \leq C(\delta, \mathcal{F}^*)K_n^{-1}, \\ \mathbb{E}\left[\{Z(l - l_{\hat{p}}) - (\alpha/2)\|l - l_{\hat{p}}\|_n^2 - D_m^*\hat{p}/(\alpha K_n)\}^+\right] \leq C(\delta, \mathcal{F}^*)K_n^{-1}, \\ 2\mathbb{E}[Z(l - \tilde{l}_p)] \leq \|l - \tilde{l}_p\|_n^2 + C(\delta, \mathcal{F}^*)K_n^{-1}.$$

Using these bounds, Eq. (5.7) becomes

$$(1 - \alpha)\mathbb{E}[\|l - \hat{l}_{\hat{p}}\|_n^2] \leq \|l - \tilde{l}_p\|_n^2 + \text{pen}(p) \\ + 2\alpha^{-1}\mathbb{E}[Z(w(\hat{p}))^2 - C_m^*\hat{p}K_n^{-1}] + \mathbb{E}[2Z(l - l_{\hat{p}}) - \alpha\|l - l_{\hat{p}}\|_n^2 - D_m^*\hat{p}/(\alpha K_n)] \\ - 2\mathbb{E}[Z(l - \tilde{l}_p)] + \mathbb{E}[-\text{pen}(\hat{p}) + \alpha^{-1}(2C_m^* + D_m^*)\hat{p}K_n^{-1}] + C(\delta, \mathcal{F}^*)K_n^{-1} \\ \leq 2\|l - \tilde{l}_p\|_n^2 + \text{pen}(p) + C(\delta, \mathcal{F}^*)K_n^{-1} + \mathbb{E}[-\text{pen}(\hat{p}) + \alpha^{-1}(2C_m^* + D_m^*)\hat{p}K_n^{-1}].$$

Choosing $\alpha = 1/2$ and $\kappa = (2C_m^* + D_m^*)/(\pi\psi'(m))$ yields,

$$\mathbb{E}[\|l - \hat{l}_{\hat{p}}\|_n^2] \leq 4\|l - \tilde{l}_p\|_n^2 + 2\text{pen}(p) + C(\delta, \mathcal{F}^*)K_n^{-1}.$$

Since the left hand side above is independent of p , the right hand side can be minimized in p , which concludes the proof of Theorem 3.

Lemma 5.1 *Let $\delta \in [0, 1/2)$ and $M > 0$. Let \mathcal{F}^* be a compact subclass of $\mathcal{L}^*(M)$. Let X be a stationary Gaussian process with spectral density $f = e^{dg+l^*}$ with $|d| \leq \delta$ and $l^* \in \mathcal{F}^*$. There exists a constant $C(\delta, \mathcal{F}^*)$, and a constant C_m^* such that*

$$\mathbb{E}[\{Z^2(w(\hat{p})) - C_m^*\hat{p}K_n^{-1}\}^+] \leq C(\delta, \mathcal{F}^*)K_n^{-1}.$$

Proof of Lemma 5.1 For $q \leq K_n$, let B_q denote the unit ball of L_q for the norm $\|\cdot\|_n$. Denote

$$\bar{Z}(q) = \sup_{u \in B_q} Z(u).$$

Since by definition $\|l_q - \hat{l}_q\|_n^{-1}(l_q - \hat{l}_q) \in B_q$, then for all $q \leq K_n$, $Z(\|l_q - \hat{l}_q\|_n^{-1}(l_q - \hat{l}_q)) \leq \bar{Z}(q)$. We now use a chaining argument sometimes referred to as the “peeling device” (cf. Van de Geer, 2000) to obtain an exponential inequality for $\bar{Z}(q)$. For any $\delta_0 \in]0, 1]$ and any $k \geq 0$, set $\delta_k = \delta_0 2^{-k}$. For all $k \geq 0$, B_q can be covered by a collection of at most

$(3/\delta_k)^q$ balls $\{B(u, \delta_k), u \in T_k\}$ with center $u \in B_q$ and radius δ_k (see Van de Geer (2000), Lemma 2.5). Any $u \in B_q$ can then be expressed as

$$u = u_0 + \sum_{k \geq 1} (u_k - u_{k-1}),$$

with $u_k \in T_k$ and $\|u - u_k\|_n \leq \delta_k$, which implies that $\delta_k \leq \|u_k - u_{k-1}\|_n \leq \delta_k + \delta_{k-1} = 3\delta_k$. Let $x = (x_k)_{k \in \mathbb{N}}$ be a sequence of real numbers to be precised later and define $\eta_k = \sigma_m \sqrt{x_k} + r_{q, K_n} x_k$ and $v = (\eta_0 + \sum_{k \geq 1} 3\delta_k \eta_k)$.

$$\begin{aligned} \mathbb{P}(\bar{Z}(q) > v) &\leq \mathbb{P}\left(\forall k \in \mathbb{N}, \exists u_k \in T_k, Z(u_0) + \sum_{k \geq 1} Z(u_k - u_{k-1}) > v(\eta)\right) \\ &\leq \sum_{u \in T_0} \mathbb{P}(Z(u) > \eta_0) + \sum_{k \geq 1} \sum_{u \in T_k, v \in T_{k-1}, \|u-v\|_n \leq 3\delta_k} \mathbb{P}(Z(u-v) > 3\delta_k \eta_k). \end{aligned}$$

Applying Proposition 5.3, we get

$$\forall u \in T_0, \mathbb{P}[Z(u) > \eta_0] \leq C(\delta, \mathcal{F}^*) e^{-K_n x_0},$$

and for all $u \in T_k$ and $w \in T_{k-1}$ such that $\|u - w\|_n \leq 3\delta_k$,

$$\mathbb{P}[Z(u - w) > 3\delta_k \eta_k] \leq \mathbb{P}[Z(\|u - w\|_n^{-1}(u - w)) > \eta_k] \leq C(\delta, \mathcal{F}^*) e^{-K_n x_k}.$$

Since T_k has at most $(3/\delta_k)$ elements, denoting $H_k = q \log(3/\delta_k)$, we obtain

$$\mathbb{P}[\bar{Z}(q) > v] \leq C(\delta, \mathcal{F}^*) e^{H_0 - K_n x_0} + C(\delta, \mathcal{F}^*) \sum_{k \geq 1} e^{H_k + H_{k-1} - K_n x_k}.$$

For $\xi \in \mathbb{R}_+$, $k \in \mathbb{N}^*$ and $q \in \mathbb{N}^*$, define $K_n x_0 = H_0 + q + \xi$, $K_n x_k = H_k + H_{k-1} + (k+1)(q + \xi)$. We now obtain

$$\begin{aligned} \mathbb{P}[\bar{Z}(q) > v] &\leq C(\delta, \mathcal{F}^*) e^{H_0 - K_n x_k} + C(\delta, \mathcal{F}^*) \sum_{k=1}^{\infty} e^{H_k + H_{k-1} - K_n x_k} \\ &\leq C(\delta, \mathcal{F}^*) \sum_{k=0}^{\infty} e^{-k(q+\xi)} = C(\delta, \mathcal{F}^*) e^{-(q+\xi)} / (1 - e^{-(q+\xi)}) \leq C(\delta, \mathcal{F}^*) e^{-(q+\xi)} / (1 - e^{-1}). \end{aligned}$$

We must now evaluate v in terms of k and q . By definition, $H_k = q \log(3/\delta_0) + kq \log(2)$, and the series $\sum_{k=1}^{\infty} k\delta_k$ is summable, hence we can write

$$v(\eta) \leq A(\delta_0)(\sigma_m \sqrt{q/K_n} + r_{q, K_n} q/K_n) + B(\delta_0)(\sigma_m \sqrt{\xi/K_n} + r_{q, K_n} \xi/K_n),$$

where $A(\delta_0)$ and $B(\delta_0)$ can be computed explicitly and minimized with respect to δ_0 . For short, we just choose $\delta_0 = 1$ and a rough evaluation yields

$$v(\eta) \leq 32\{\sigma_m \sqrt{q/K_n} + r_{q, K_n} q/K_n\} + 15\{\sigma_m \sqrt{\xi/K_n} + r_{q, K_n} \xi/K_n\}.$$

By definition of $r_{q, K}$, and since it is assumed that $q < \sqrt{K_n}/\log(K_n)$, we further get

$$v(\eta) \leq 32(\sigma_m + 1)\sqrt{q/K_n} + 15\{\sigma_m \sqrt{\xi/K_n} + r_{q, K_n} \xi/K_n\}.$$

We have obtained

$$\mathbb{P}(\bar{Z}(q) > 32(1 + \sigma_m)\sqrt{q/K_n} + 15\{\sigma_m\sqrt{\xi/K_n} + r_{q,K_n}\xi/K_n\}) \leq C(\delta, \mathcal{F}^*)e^{-q-\xi}. \quad (5.8)$$

Denote $C_m^* = 2048(1 + \sigma_m)^2$. By linearity and symmetry, $\bar{Z}(q)$ is a non negative random variable, hence

$$\begin{aligned} \mathbb{E}[\{\bar{Z}^2(q) - C_m^*qK_n^{-1}\}^+] &= \int_0^\infty \mathbb{P}(\bar{Z}^2(q) - C_m^*qK_n^{-1} > v)dv \\ &= \int_0^\infty \mathbb{P}(\bar{Z}(q) > \sqrt{C_m^*qK_n^{-1} + v})dv \\ &\leq \int_0^\infty \mathbb{P}(\bar{Z}(q) > 32(1 + \sigma_m)\sqrt{q/K_n} + \sqrt{v/2})dv. \end{aligned}$$

Set $a = 15\sigma_m/\sqrt{K_n}$ and $b = 15r_{q,K_n}/K_n$, and $v = 4a^2\xi + 4b^2\xi$ in the last integral above. Applying (5.8), we get

$$\begin{aligned} \mathbb{E}[\{\bar{Z}^2(q) - C_m^*qK_n^{-1}\}^+] &\leq \int_0^\infty \mathbb{P}(\bar{Z}(q) > 32(1 + \sigma_m)\sqrt{q/K_n} + \sqrt{v/2})dv \\ &\leq \int_0^\infty \mathbb{P}(\bar{Z}(q) > 32(1 + \sigma_m)\sqrt{q/K_n} + a\sqrt{\xi} + b\xi)(4a^2 + 8b^2\xi)d\xi \\ &\leq C(\delta, \mathcal{F}^*)e^{-q} \int_0^\infty e^{-\xi}(4a^2 + 8b^2\xi)d\xi. \end{aligned}$$

By assumption, if $q \leq p_n^*$, then $r_{q,K_n}^2 \leq K_n$, hence we conclude that

$$\mathbb{E}[\{\bar{Z}^2(q) - C_m^*qK_n^{-1}\}^+] \leq C(\delta, \mathcal{F}^*)K_n^{-1}e^{-q}.$$

Summing these bounds over q yields

$$\begin{aligned} \mathbb{E}[\{Z^2(w(\hat{p})) - C_m^*\hat{p}K_n^{-1}\}^+] &\leq \sum_{q=1}^{p_n^*} \mathbb{E}[\{\bar{Z}^2(q) - C_m^*qK_n^{-1}\}^+] \\ &\leq C(\delta, \mathcal{F}^*)K_n^{-1} \sum_{q=1}^{\infty} e^{-q} \leq C(\delta, \mathcal{F}^*)K_n^{-1}. \end{aligned}$$

Lemma 5.2 *Let $\delta \in [0, 1/2)$ and $M > 0$. Let \mathcal{F}^* be a compact subclass of $\mathcal{L}^*(M)$. Let X be a stationary Gaussian process with spectral density $f = e^{dg+l^*}$ with $|d| \leq \delta$ and $l^* \in \mathcal{F}^*$. There exists a constant $C(\delta, \mathcal{F}^*)$, and a constant D_m^* which depends only on m such that*

$$\mathbb{E}\left[\{Z(l - l_{\hat{p}}) - (\alpha/2)\|l - l_{\hat{p}}\|_n^2 - D_m^*\hat{p}/(\alpha K_n)\}^+\right] \leq C(\delta, \mathcal{F}^*)K_n^{-1}.$$

Proof of Lemma 5.2 For any non negative integer q , define $v_q = \|l - l_q\|_n^{-1}(l - l_q)$. For $x > 0$ and $\epsilon > 0$, it holds that

$$\mathbb{P}\left(Z(l - l_q) > \epsilon \frac{\|l - l_q\|_n^2 + x^2}{2x}\right) \leq \mathbb{P}\left(Z(v_q) > \epsilon \frac{\|l - l_q\|_n^2 + x^2}{2x\|l - l_q\|_n}\right) \leq \mathbb{P}(Z(v_q) > \epsilon).$$

Choosing $x = \epsilon/\alpha$ and $\epsilon = \sigma_m \sqrt{y/2\pi} + r_{q,K_n} y$ for some $y > 0$ and applying (5.6) yields

$$\mathbb{P}\left(Z(l - l_q) - (\alpha/2)\|l - l_q\|_n^2 > \alpha^{-1}(\sigma_m \sqrt{y/2\pi} + |v_q|_\infty y)^2\right) \leq C(\delta, \mathcal{F}^*)e^{-K_n y},$$

where $|v|_\infty = \max_{1 \leq k \leq K_n} |v(y_k)|$. If $\|v\|_n = 1$, then it necessarily holds that $|v|_\infty \leq \sqrt{K_n/2\pi}$. Choosing $K_n y = 2\pi\sqrt{q+\xi}$ for $q \geq 1$ and $\xi \geq 0$ yields

$$\mathbb{P}\left(Z(l - l_q) - (\alpha/2)\|l - l_q\|_n^2 > \alpha^{-1}(\sigma_m(q+\xi)^{1/4} + \sqrt{2\pi(q+\xi)})^2 K_n^{-1}\right) \leq C(\delta, \mathcal{F}^*)e^{-\sqrt{q+\xi}},$$

Define now $D_m^* = (\sqrt{2\pi} + \sigma_m)^2$. Since $q + \xi \geq 1$, the previous inequality implies that

$$\mathbb{P}\left(Z(l - l_q) - (\alpha/2)\|l - l_q\|_n^2 > \alpha^{-1}D_m^*(q+\xi)K_n^{-1}\right) \leq C(\delta, \mathcal{F}^*)e^{-\sqrt{q+\xi}},$$

Integrating and summing this bound finally yields

$$\mathbb{E}\left[\left\{Z(\hat{p}) - (\alpha/2)\|l - \hat{p}\|_n^2 - D_m^*\hat{p}/(\alpha K_n)\right\}^+\right] \leq C(\delta, \mathcal{F}^*)K_n^{-1}.$$

Lemma 5.3 Let $\delta \in [0, 1/2)$ and $M > 0$. Let \mathcal{F}^* be a compact subclass of $\mathcal{L}^*(M)$. Let X be a stationary Gaussian process with spectral density $f = e^{dg+l^*}$ with $|d| \leq \delta$ and $l^* \in \mathcal{F}^*$. There exists a constant $C(\delta, \mathcal{F}^*)$,

$$2|\mathbb{E}[Z(l - \tilde{l}_p)]| \leq \|l - \tilde{l}_p\|_n^2 + C(\delta, \mathcal{F}^*)K_n^{-1}.$$

Proof of Lemma 5.3 Applying Cauchy-Schwarz inequality and the bound (5.2),

$$|2\mathbb{E}[Z(p)]| = 2|\mathbb{E}[\langle \epsilon_n, l - l_p \rangle]| \leq \|l - l_p\|_n^2 + \frac{2\pi}{K_n} \sum_{k=1}^{K_n} \mathbb{E}^2[\epsilon_{n,k}] \leq \|l - l_p\|_n^2 + C(\delta, \mathcal{F}^*)K_n^{-1}.$$

6 Proof of Theorem 1

We follow here the proof of Theorem 3.2 in Belitser (2000). Apart from Lemma 6.1 below, there are no significant differences, so we omit most technicalities.

Let λ be a continuously differentiable probability density function on $[-1, 1]$, such that $\lambda(-1) = \lambda(1) = \lambda'(-1) = \lambda'(1) = 0$ and with finite information :

$$I_\lambda := \int_{-1}^1 (\lambda'(x))^2 / \lambda(x) dx < \infty.$$

Let $(p_n)_{n \geq 1}$ be a non decreasing sequence of integers, such that for all n , $p_n \leq K$, let m_j $1 \leq j \leq p_n$ and R be positive real numbers and let $\Theta_n = \prod_{j=1}^{p_n} [-Rm_j, Rm_j]$.

Let Λ be the probability density function on Θ_n defined as

$$\Lambda(x) = \prod_{i=1}^{p_n} (Rm_i)^{-1} \lambda(Rm_i x_i).$$

Let Θ denote either $\mathcal{S}(\beta, L)$ or $\mathcal{A}(\gamma, L)$ and $r_n = \inf_{\hat{\theta}_n} \sup_{\theta \in \Theta} \mathbb{E}_\theta [\|\theta - \hat{\theta}\|^2]$. In the sequel, we identify a sequence $\theta = (\theta_j)_{j \geq 0}$ with the function $\sum_{j \geq 0} \theta_j h_j$. The orthonormality properties of the h_j 's yield that for any

$$\|\theta - \theta'\|^2 = \sum_{j=0}^{\infty} (\theta_j - \theta'_j)^2,$$

and if $\theta_j = \theta'_j = 0$ if $j \geq K_n$, then it also holds that $\|\theta - \theta'\|_n^2 = \sum_{j=0}^{\infty} (\theta_j - \hat{\theta}_j)^2$. Hence we have

$$\begin{aligned} r_n &\geq \inf_{\hat{\theta}_n} \int_{\Theta} \sum_{j=0}^{\infty} \mathbb{E}_\theta (\hat{\theta}_j - \theta_j)^2 \Lambda(d\theta) = \inf_{\hat{\theta}_n \in \Theta_n} \int_{\Theta} \sum_{j=0}^{\infty} \mathbb{E}_\theta (\hat{\theta}_j - \theta_j)^2 \Lambda(d\theta) \\ &\geq \inf_{\hat{\theta}_n \in \Theta_n} \int_{\Theta_n} \sum_{j=0}^{\infty} \mathbb{E}_\theta (\hat{\theta}_j - \theta_j)^2 \Lambda(d\theta) - \sup_{\hat{\theta}_n \in \Theta_n} \int_{\Theta_n \setminus \Theta} \sum_{j=0}^{\infty} \mathbb{E}_\theta (\hat{\theta}_j - \theta_j)^2 \Lambda(d\theta) \\ &= \inf_{\hat{\theta}_n \in \Theta_n} \sum_{j=1}^{p_n} \mathbb{E}(\hat{\theta}_j - \theta_j)^2 - 4R^2 \Lambda(\Theta_n \setminus \Theta) \sum_{j=1}^{p_n} m_j^2, \end{aligned}$$

where the infimum $\inf_{\hat{\theta}}$ is taken over all possible estimates of θ based on a realization $\{X_1, \dots, X_n\}$ of a stationary Gaussian process $(X_t)_{t \in \mathbb{Z}}$ with spectral density e^{l_θ} , \mathbb{E}_θ denotes the distribution of this process, and \mathbb{E} denotes the joint distribution of (θ, X) when the distribution of θ is Λ .

Applying the Van Trees inequality (cf. Gill et Levit (1995)), we get

$$\mathbb{E}(\hat{\theta}_j - \theta_j)^2 \geq \frac{1}{\mathbb{E}[I_n(\theta_j)] + m_j^{-2} I_\lambda},$$

where $I_n(\theta_j)$ is the j -th diagonal element of the Fisher information matrix of the distribution of the vector (X_1, \dots, X_n) .

The only difference with the proof of Theorem 3.2 of Belitser (2000) lies in the evaluation of $I_n(\theta_j)$. We state it as a Lemma whose proof is postponed.

Lemma 6.1 *Assume that there exists a integer s such that $\lim_{n \rightarrow \infty} n^{2/s} p_n \sum_{j=1}^{p_n} m_j^2 = 0$. Then, uniformly with respect to $\theta \in \Theta_n$ and $1 \leq j \leq p_n$, it holds that*

$$I_n(\theta_j) = (4\pi)^{-1} n(1 + o(1)).$$

Using this bound, we now obtain

$$r_n \geq \frac{4\pi}{n} \sum_{j=1}^{p_n} \frac{m_j^2}{m_j^2(1 + o(1)) + 4\pi I_\lambda/n} - 4R^2 \Lambda(\Theta_n \setminus \Theta) \sum_{j=1}^{p_n} m_j^2. \quad (6.1)$$

If $\Theta = \mathcal{A}(\gamma, L)$, then assumption \mathcal{F}_2 of Theorem 3.2 of Belitser (2000) holds. Hence we can choose $R = 1$, $p_n = \{\log(n) - \log \log(n)\}/2\gamma$ and

$$m_j^2 = \frac{4\pi(1 - e^{\gamma j}/\sqrt{n})}{\sqrt{n}e^{\gamma j}}, \quad 1 \leq j \leq p_n.$$

Then the assumption of Lemma 6.1 holds, $\Theta_n \subset \Theta$ for large enough n and

$$\frac{4\pi}{n} \sum_{j=1}^{p_n} \frac{m_j^2}{m_j^2(1 + o(1)) + 4\pi I_\lambda/n} = \frac{4\pi p_n}{n} - \frac{4\pi}{n} \sum_{j=1}^{p_n} \frac{4\pi I_\lambda}{n m_j^2(1 + o(1)) + 4\pi I_\lambda} = \frac{4\pi p_n}{n} + O(1/n).$$

Altogether, this yields $r_n = \frac{2\pi \log(n)}{\gamma n} (1 + o(1))$.

If $\Theta = \mathcal{S}(\beta, L)$, then the assumption of Lemma 6.1 holds. Assumption \mathcal{F}_1 of Theorem 3.2 of Belitser (2000) also holds, and this allows to choose $p_n = \{L^2(\beta + 1)(2\beta + 1)n/(4\pi\beta)\}^{\frac{1}{2\beta+1}}$,

$$m_j^2 = \frac{4\pi(1 - p_n^{-\beta} j^\beta)}{n p_n^{-\beta} j^\beta}, \quad 1 \leq j \leq p_n,$$

and for any $\epsilon > 0$, we can choose $R = R_\epsilon$ and the measure λ in such a way that $I_\lambda \leq 1 + \epsilon$ and

$$\Lambda(\Theta_n \setminus \Theta) \sum_{j=1}^{p_n} m_j^2 = o(n^{-2\beta/(2\beta+1)}),$$

(cf. Belitser (2000), pp. 71-72 for details). Similar computations also yields that

$$\begin{aligned} \frac{4\pi}{n} \sum_{j=1}^{p_n} \frac{m_j^2}{m_j^2(1 + o(1)) + 4\pi I_\lambda/n} &\geq \frac{4\pi}{n(1 + \epsilon)} \sum_{j=1}^{p_n} \frac{m_j^2}{m_j^2(1 + o(1)) + 4\pi/n} \\ &\geq \frac{4\pi}{n(1 + \epsilon)} \sum_{j=1}^{p_n} \frac{1 - p_n^{-\beta} j^\beta}{1 + o(1)} = \frac{\mathcal{P}(\beta, L)}{1 + \epsilon} n^{-2\beta/(2\beta+1)} (1 + o(1)). \end{aligned}$$

Since ϵ is arbitrary, we conclude that $r_n \geq \mathcal{P}(\beta, L) n^{-2\beta/(2\beta+1)} (1 + o(1))$. This concludes the proof of Theorem 1.

Proof of Lemma 6.1 Let $\Sigma(\theta)$ denote the covariance matrix of the Gaussian vector $(X_1, \dots, X_n)^T$. The Fisher information matrix $I_n(\theta)$ has diagonal elements $I_n(\theta_j)$ given by

$$I_n(\theta_j) = \frac{1}{2} \text{tr} \left\{ \left(\Sigma(\theta)^{-1} \partial_{\theta_j} \Sigma(\theta) \right)^2 \right\}.$$

For a given integrable function ϕ , let $T_n(\phi)$ denote the n -th order Toeplitz matrix of ϕ , with entries

$$T_n(\phi)_{u,v} = \int_{-\pi}^{\pi} \phi(x) e^{i(u-v)x} dx, \quad 1 \leq u, v \leq n.$$

Let J_n denote the n -dimensional identity matrix and $h_\theta = e^\theta - 1 - \theta$. With these notations, we can write

$$\Sigma(\theta) = T_n(e^\theta) = T_n(1 + \theta + h_\theta) = 2\pi J_n + T_n(\theta) + T_n(h_\theta).$$

Recall that the spectral radius $\rho(A)$ of a symmetric matrix $A = (A_{i,j})_{1 \leq i,j \leq q}$ is the greatest eigenvalue of AA^T , and that it is bounded by $\max_{1 \leq i \leq q} \sum_{j=1}^q |A_{i,j}|$. We will also use the following properties : for all symmetrix matrices A and B , it holds that

$$\rho(AB) \leq \rho(A)\rho(B), \quad \text{tr}(AB) = \text{tr}(BA), \quad |\text{tr}(AB^2)| \leq \rho(A)\text{tr}(B^2). \quad (6.2)$$

Let $\rho_n(\theta)$ denote the spectral radius of $W_n(\theta) := (2\pi)^{-1} \Sigma(\theta) - J_n = (2\pi)^{-1} T_n(e^\theta - 1)$. If $\rho_n(\theta) < 1$, then $\Sigma^{-1}(\theta) = (2\pi)^{-1} (J_n + V_n(\theta))$, with $V_n(\theta) = \sum_{k=1}^{\infty} (-1)^k W_n^k(\theta)$ and the spectral radius of $V_n(\theta)$ verifies $\rho(V_n(\theta)) \leq \rho_n(\theta)/(1 - \rho_n(\theta))$. Hence, using the properties (6.2)

$$\begin{aligned} 4\pi^2 \text{tr} \left\{ (\Sigma^{-1}(\theta) \partial_{\theta_j} \Sigma(\theta))^2 \right\} &= \text{tr} \left\{ (\partial_{\theta_j} \Sigma(\theta))^2 \right\} + 2\text{tr} \left\{ V_\theta(\partial_{\theta_j} \Sigma(\theta))^2 \right\} + \text{tr} \left\{ V_\theta^2 (\partial_{\theta_j} \Sigma(\theta))^2 \right\} \\ &= \text{tr} \left\{ (\partial_{\theta_j} \Sigma(\theta))^2 \right\} (1 + O(\rho_n(\theta))). \end{aligned} \quad (6.3)$$

Note now that $\partial_{\theta_j} \Sigma(\theta) = \partial_{\theta_j} T_n(e^\theta) = T_n(h_j e^\theta)$. Applying the Parseval-Bessel Theorem, we get

$$\text{tr}(T_n^2(h_j e^\theta)) = 2\pi n(1 + o(1)) \int_{-\pi}^{\pi} h_j^2(x) e^{2\theta(x)} dx = n(1 + o(1)) \int_{-\pi}^{\pi} e^{2\theta(x)} dx.$$

By definition, if $\theta \in \Theta_n$, then

$$\|\theta\|_\infty \leq \pi^{-1/2} \sum_{j=1}^{p_n} m_j = o(1),$$

uniformly with respect to $\theta \in \Theta_n$ under the assumptions of Lemma 6.1. Hence $\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} e^{2\theta(x)} dx = 2\pi$, and $\partial_{\theta_j} \Sigma(\theta) = 2\pi n(1 + o(1))$.

We must now prove that $\lim_{n \rightarrow \infty} \rho_n(\theta) = 0$, uniformly with respect to $\theta \in \Theta_n$. Denote $\gamma_k(\theta) = (2\pi)^{-1} \int_{-\pi}^{\pi} (e^\theta - 1) e^{ikx} dx$. With this notation, we get

$$\rho_n(\theta) \leq \sum_{k=0}^{n-1} |\gamma_k(\theta)|.$$

Let $s \geq 3$ be an integer. A Taylor expansion of the function $x \rightarrow (e^x - 1)$ up to the s -th order yields

$$\gamma_k(\theta) = \sqrt{\pi}\theta_k + \sum_{r=2}^{s-1} \frac{1}{2\pi r!} \int_{-\pi}^{\pi} \theta^r(x) e^{ikx} dx + \frac{1}{2\pi s!} \int_{-\pi}^{\pi} r_{\theta,s}(x) e^{ikx} dx,$$

where $|r_{\theta,s}| \leq |\theta|^s e^{|\theta|}$. If $\theta \in \Theta_n$, then

$$\|\theta\|_{\infty} \leq \pi^{-1/2} \sum_{j=1}^{p_n} m_j \leq p_n^{1/2} \eta_n,$$

where we have defined $\eta_n^2 := \sum_{j=1}^{p_n} m_j^2$. Also, for $r \geq 1$, θ^r is a trigonometric polynomial of degree at most rp_n , thus $\int_{-\pi}^{\pi} \theta^r(x) e^{ikx} dx = 0$ if $k > rp_n$. Let C denote a generic numerical constant whose value can change upon each appearance. By the Parseval-Bessel inequality,

$$\sum_{k=1}^n \left| \int_{-\pi}^{\pi} \theta^r(x) e^{ikx} dx \right| \leq Cp_n^{1/2} \|\theta^{2r}\| \leq Cp_n^{1/2} \epsilon_n^{r-1} \|\theta^2\| \leq Cp_n^{1/2} \eta_n.$$

The last term in the expansion is bounded by

$$\sum_{k=1}^n \left| \int_{-\pi}^{\pi} r_{\theta,s}(x) e^{ikx} dx \right| \leq Cn \sum_{j=1}^{p_n} m_j \leq Cnp_n^{s/2} \eta_n^s$$

Altogether, we get,

$$\sum_{k=0}^{n-1} |\gamma_k(\theta)| \leq Cp_n^{1/2} \eta_n + Cnp_n^{s/2} \eta_n^s = o(1),$$

uniformly with respect to $\theta \in \Theta_n$ under the assumptions of Lemma 6.1. This concludes the proof of Lemma 6.1.

Acknowledgement I thank the anonymous referee for his very valuable comments that greatly improved this paper. All remaining errors and shortcomings of the paper are obviously mine.

References

- [1] R.K. Adenstedt. On large-sample estimation for the mean of a stationary random sequence. *Annals of Statistics*, 2:1095–1107, 1974.
- [2] M.A. Arcones. Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Annals of Probability*, 15(4):2243–2274, 1994.

- [3] Y. Baraud, F. Comte, and G. Viennet. Model selection for (auto)-regression with dependent data.
- [4] A. Barron, L. Birge, and P. Massard. Risks bounds for model selection via penalization. *Probability Theory and Related Fields*, 113:301–415, 1999.
- [5] E.N. Belitser. *Minimax estimation in regression and random censorship models*. CWI Tracts. 127. Amsterdam: CWI, 2000.
- [6] F. Comte. Adaptive estimation of the spectrum of a stationary Gaussian sequence. *Bernoulli*, 7(2), 2001.
- [7] S. Efroimovich and M.S. Pinsker. Estimation of square integrable density on the basis of a sequence of observations. *Problems of Information Transmission*, 17(3):182–196, 1982.
- [8] S. Efroimovich. Data-driven efficient estimation of the spectral density. *J. Amer. Statist. Assoc.*, 93(442):762–769, 1998.
- [9] R. Gill and B. Levit. Applications of the van Trees inequality: A bayesian Cramer-Rao bound. *Bernoulli*, 1-2:59–79, 1995.
- [10] L. Giraitis, P.M. Robinson, and A. Samarov. Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. *Journal of Time Series Analysis*, 18:49–61, 1997.
- [11] L. Giraitis, P.M. Robinson, and A. Samarov. Adaptive rate optimal estimation of the long memory parameter. *J. Multivariate Analysis*, 72(2):183–207, 2000.
- [12] G.K. Golubev. Nonparametric estimation of smooth spectral densities of gaussian stationary sequences. *Probab. Theor. and Applications*, 38:630–639, 1993.
- [13] C.W.J. Granger and R. Joyeux. An introduction to long memory time series and fractional differencing. *Journal of Time Series Analysis*, 1:15–30, 1980.
- [14] J.R.M. Hosking. Fractional differencing. *Biometrika*, 60:165–176, 1991.
- [15] C.M. Hurvich and K. Beltrao. Asymptotics of the low-frequency ordinates of the periodogram of a long-memory time series. *Journal of Time Series Analysis*, 14:455–472, 1993.
- [16] C.M. Hurvich and W.W. Chen. An efficient taper for potentially overdifferenced long-memory time series. *Journal of Time Series Analysis*, 21(2):155–180, 2000.
- [17] C.M. Hurvich, E. Moulines, and Ph. Soulier. The FEXP estimator for non Gaussian, potentially no stationary processes. Prépublication 133 de l’Université d’Evry Val d’Essonne, 2000.

- [18] A. Iouditsky, E. Moulines, and Ph. Soulier. Adaptive estimation of the fractional differencing coefficient. Prépublication de l'université d'Evry val d'Essonne, to appear in *Bernoulli*, 1999.
- [19] N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions 2*. New York, Wiley, 1995.
- [20] E. Moulines and Ph. Soulier. Broadband log-periodogram regression of time series with long-range dependence. *Annals of Statistics*, 27(4):1415–1439, 1999.
- [21] E. Moulines and Ph. Soulier. Data driven order selection for projection estimator of the spectral density of time series with long range dependence. *Journal of Time Series Analysis*, 21(2):193–218, 2000.
- [22] B.S. Polyak and A.B. Tsybakov and. Asymptotic optimality of the C_p test for the orthogonal series estimation of regression. *Theory Probab. Appl.*, 35(2):293–306, 1989.
- [23] P.M. Robinson. Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, 23(3):1043–1072, 1995a.
- [24] Ph. Soulier. Some new bounds and a central limit theorem for functions of Gaussian vectors. Prépublication 71 de l'Université d'Evry Val d'Essonne, to appear in *Statistics and Probability Letters*, 1998.
- [25] M.S. Taqqu. Law of the iterated logarithm for sums of nonlinear functions of Gaussian variables that exhibit long range dependence. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 40:203–238, 1977.
- [26] N. Terrin and C.M. Hurvich. An asymptotic Wiener-Ito representation for the low frequency ordinates of the periodogram of a long memory time series. *Stochastic Processes and their Applications*, 54(2):297–307, 1994.
- [27] J.W. Tukey. *An introduction to the calculation of numerical spectrum analysis*. Advanced seminar on spectral analysis of time series (ed. B. Harris), Wiley, 1967.
- [28] S.A. Van de Geer. *Applications of empirical process theory*. Cambridge Series on Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2000.
- [29] C. Velasco. Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, 20(1):87–127, 1999a.
- [30] C. Velasco. Non-Stationary log-periodogram regression. *Journal of Econometrics*, 91(2):325–371, 1999b.