Adaptive estimation of the spectrum of a stationary Gaussian sequence

FABIENNE COMTE

Laboratoire de Probabilités et Modèles Aléatoires, Boîte 188, Université Paris VI, 4 place Jussieu, 75252 Paris Cedex 05, France. E-mail: comte@ccr.jussieu.fr

In this paper, we study the problem of nonparametric adaptive estimation of the spectral density f of a stationary Gaussian sequence. For this purpose, we consider a collection of finite-dimensional linear spaces (e.g. linear spaces spanned by wavelets or piecewise polynomials on possibly irregular grids or spaces of trigonometric polynomials). We estimate the spectral density by a projection estimator based on the periodogram and constructed on a data-driven choice of linear space from the collection. This data-driven choice is made via the minimization of a penalized projection contrast. The penalty function depends on $\|f\|_\infty$, but we give results including the estimation of this bound. Moreover, we give extensions to the case of unbounded spectral densities (long-memory processes). In all cases, we state non-asymptotic risk bounds in L_2-norm for our estimator, and we show that it is adaptive in the minimax sense over a large class of Besov balls.

Keywords: adaptive estimation; long-memory process; penalty function; projection estimator; stationary sequence

1. Introduction

In this paper, we consider the problem of estimating the spectral density f of a stationary Gaussian sequence. We develop an adaptive nonparametric method that automatically selects an estimator $\hat{f}_{\hat{m}}$ from a collection of estimators $\hat{f}_m$ for varying index m. These estimators are constructed as projection estimators based on integrals of the periodogram. The index m denotes the linear functional subspace of $L_2([-\pi, \pi])$ on which the estimator is computed. Our method is very flexible and allows us to look for piecewise estimators associated with possibly irregular partitions of the interval $[-\pi, \pi]$. This is of particular interest for the recognition of peaks, which is often an important goal of spectral analysis. This explains why an appropriate method needs to be adaptive with respect to a possibly varying degree of smoothness of the function to be estimated.

We measure the performance of our estimator via the L_2-integrated risk and we give non-asymptotic risk bounds. We show that, without any prior knowledge of f, our estimate automatically reaches the minimax optimal rate on Besov balls.

We use a method of model selection inspired by Birgé and Massart (1997; 1998) and Barron *et al*. (1999), who studied several types of contrast and estimate in various contexts, but under assumptions of linearity or independence which are not valid here.

The problem of estimating the spectral density of a stationary (Gaussian) sequence has a

very rich history, of which it is beyond the scope of this paper to give an exhaustive presentation. Let us just mention that it was first studied mainly in a parametric framework: see, for example, the seminal paper by Whittle (1953), or the book by Hannan (1970), as well as the methods developped by Davies (1973). Minimum contrast estimation for spectral densities has also been explored in a parametric setting by Taniguchi in several papers (see, for example, Taniguchi 1987). A new wave of studies were published more recently (Fox and Taqqu 1986; Dahlhaus 1989) due to the development of fractional (long-memory) models in which the standard assumptions of boundedness from above and below of the spectral density are not fulfilled.

In the nonparametric framework, Bentkus and Rudzkis (1976) proved large-deviation results for a projection spectral estimate based on a tapered periodogram. Bentkus (1985) computed optimal rates of convergence of spectral estimates in some spaces of differentiable functions. In both cases, the variables are Gaussian but the rates of convergence are asymptotic and the methods are not adaptive: this means that the definition of their estimator requires a priori knowledge of the smoothness of the function f. Golubev (1993) builds a tapered estimate $\hat{f}_n(\omega) = \sum_{|k| \leqslant n} h_k \gamma_k e^{i\omega k}$ and proposes an adaptive procedure with respect to the choice of the weights h_k. His estimate is asymptotically minimax on some families of neighbourhoods of the true spectral density, under assumptions on f which are a little stronger than ours and for well-chosen families of weights among which the best is selected. Finally, Neumann (1996) considers a tapered wavelet projection estimator in a non-Gaussian framework. His conditions on f are given mainly in terms of the cumulants of the sequence. He uses first a fixed (unknown) and then a random thresholding method inspired by Donoho and Johnstone (1994; 1998). He reaches the optimal rates of convergence on standard Besov spaces with the fixed threshold. But the minimax rate is obtained up to a power of $\ln(n)$ with the random threshold. The adaptive method we develop here does not include this $\ln(n)$ factor. Again, in contrast to Neumann (1996), we give non-asymptotic results.

Note that Birgé and Massart (1997) have shown that wavelet thresholding can be obtained as a particular case of a model selection method similar to the one described here. It is also worth mentioning that we recover the adaptive result found by Efromovich (1998), who presents an adaptive method in the particular case of linear processes estimated with trigonometric bases. Most recently, an adaptive method selecting the optimal degree of a Fourier expansion has been developed by Soulier (1999): his method is based on log-periodogram regression in a semiparametric framework allowing long-memory cases to be dealt with.

The paper is organized as follows. Section 2 presents the model, the estimator, the collections of models and the assumptions under which we work. Section 3 gives the results obtained for our penalized minimum constrast estimator: we provide a uniform and non-asymptotic bound for the L_2-risk of our estimator. We work under the assumption of boundedness from above of the spectral density f, and we consider very general collections of models. But we find a $\|f\|_\infty^2$ multiplicative factor in the penalty function which must be used to implement the method. Therefore, we give an extension of our first theorem showing that we can preserve our result by replacing the $\|f\|_\infty^2$ factor in the penalty by an estimate. Moreover, we show how the bounds provided in our first two theorems lead to

asymptotic minimax rates of convergence of our estimate on Besov spaces: the adaptivity lies in the fact that no prior knowledge on the smoothness of f is required to reach the right rate. In Section 4, we present a robustness result proving that our fundamental inequality also holds when the spectral density is not bounded (long-memory case) in the case of trigonometric bases. Note that in any case, and contrary to a great deal of other work on the subject, we do not require the spectral density to be bounded away from 0. Finally, in Section 5 we give some simulation results that allow our method to be compared with Neumann's (1996) wavelets and kernel methods. Section 6 and the Appendix gather together the proofs of all our results.

2. The framework

2.1. The model

Throughout the following, we consider X_is satisfying the following assumption:

Assumption 1. *The sequence $(X_1, \ldots, X_n)$ is an n-sample drawn from a stationary sequence of Gaussian random variables*.

Let f be the spectral density of the process:

$$f(\omega) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \gamma_k e^{i\omega k}, \qquad \omega \in [-\pi, \pi],$$

where $\gamma_k = \mathrm{cov}(X_t, X_{t+k})$, $t \in \mathbb{Z}$. We need the following standard assumption on f:

Assumption 2. *The autocovariance function is such that $\sum_{k \in \mathbb{Z}} |\gamma_k| = M < +\infty$ and $\sum_{k \in \mathbb{Z}} |k\gamma_k^2| = M_1 < +\infty$*.

This condition is similar to the one given by Golubev (1993) for his set Σ except that he requires f to be separated from 0.

Assumption 2 implies in particular that the spectral density f is bounded by the constant M. As a consequence, it is also square-integrable. Nevertheless the boundedness of f is not necessary for square integrability, as illustrated by the fractional examples below. Moreover, Assumption 2 implies the following bound on the bias, which is proved in the Appendix:

Proposition 1. *Let X be a stationary sequence satisfying Assumption 2. Then*

$$\int_{-\pi}^{\pi} (f(\omega) - \mathrm{E}(I_n(\omega)))^2 \, d\omega \leqslant \frac{M_1 + 39M^2}{2\pi n} := \frac{M_2}{n}. \tag{1}$$

Remark 2.1. Assumption 2 does not require the spectral density to be bounded away from zero. Thus fractional models with intermediate memory are allowed. For instance, consider the process X_t defined by $(1 - B)^\alpha P(B)X_t = Q(B)\varepsilon_t$, where ε_t are independent and

identically distributed (i.i.d.) variables with mean 0 and variance σ^2, α is a real number in $]-\frac{1}{2}, \frac{1}{2}[$, B is the backward operator ($BX_t = X_{t-1}$) and P and Q are polynomials with roots outside the unit circle. Then X admits the following spectral density f:

$$f(\omega) = \frac{\sigma^2}{2\pi}\left(\sin\left(\frac{\omega}{2}\right)\right)^{-2\alpha}\frac{|Q(e^{i\omega})|^2}{|P(e^{i\omega})|^2} \underset{\omega \to 0}{\sim} C\omega^{-2\alpha},$$

where C is a constant (see, for instance, Brockwell and Davis 1991, Section 13.2, p. 520). Consequently, for negative α, the spectral density is zero when $\omega = 0$. Moreover, the autocovariance function γ_k is equivalent when $k \to +\infty$ to $C'k^{2\alpha-1}$, where C' is a constant. This implies that Assumption 2 is satisfied for $\alpha < 0$.

2.2. The collections of models

We begin the presentation of the collections of models with some examples. We describe five specific families of models $(S_m)_{m \in \mathcal{M}_n}$ which are suitable for estimating a spectral density f in $L_2([-\pi, \pi], dx)$. These fall into three categories: trigonometric polynomials, wavelets and piecewise polynomials.

(T) *Trigonometric polynomials*. We consider spaces S_m generated by the functions $\varphi_j(x) = \pi^{-1/2}\cos(jx)$, for $j = 0, \ldots, d$, where $D_m = d + 1$ is thus the dimension of S_m. In other words, any function t in S_m can be written $t(x) = \sum_{j=0}^{D_m-1} a_j\cos(jx)$ for some real numbers a_j, $j = 0, \ldots, D_m - 1$. Such a space S_m is thus entirely defined by its dimension D_m. The family of models $\mathcal{M}_n$ is in that case the set of all possible dimensions: $\mathcal{M}_n = \{1, \ldots, n\}$.

(P) *Regular piecewise polynomials*. We consider the dyadic partitions of $[0, \pi]$ given by $\mathcal{T}_m = \{[\pi j/2^m, \pi(j+1)/2^m), j = 0, \ldots, 2^m - 1\}$. Given some positive integer r, we define S_m as the space of piecewise polynomials with degree bounded by $r - 1$ on the partition $\mathcal{T}_m$ and defined on $[-\pi, 0]$ by $t(-x) = t(x)$. Here $D_m = r2^m$. This family is regular in the sense that the partition $\mathcal{T}_m$ has equispaced knots. Another regular family is obtained by considering general regular partitions $\mathcal{T}_m = \{[\pi j/m, \pi(j+1)/m), j = 0, \ldots, m - 1\}$, $D_m = rm$.

(GP) *General piecewice polynomials (of degree less than $r - 1$)*. The spaces are generated by polynomials of degree no greater than $r - 1$ on each subinterval of the grid whose knots are now all possible subsets of cardinality $m - 1$ of $\{\pi j/n, j = 1, \ldots, n - 1\}$, $D_m = rm \leqslant n$.

(W) *Regular compactly supported periodic wavelets*. An orthonormal basis of the L_2 space of 2π periodic functions on $[-\pi, \pi]$ is given (see Daubechies 1992, Section 9.3) by $\{\varphi_\lambda\}_{\{\lambda=(j,k),k\in\Lambda(j),j\in\mathbb{N}\}} = \{\phi_{l,k}\}_{k\in\Lambda(l)} \cup \{\psi_{j,k}\}_{j\geqslant l,k\in\Lambda(j)}$, where $\Lambda(j) = \{1, \ldots, 2^j\}$ and

$$\phi_{l,k}(x) = \frac{1}{\sqrt{2\pi}}\sum_{n\in\mathbb{Z}}\tilde{\phi}_{l,k}\left(\frac{x}{2\pi} + n\right), \qquad \psi_{l,k}(x) = \frac{1}{\sqrt{2\pi}}\sum_{n\in\mathbb{Z}}\tilde{\psi}_{l,k}\left(\frac{x}{2\pi} + n\right),$$

where $\tilde{\phi}_{l,k}(x) = 2^{l/2}\phi_{l,k}(2^l x - k)$ and $\tilde{\psi}_{j,k}(x) = 2^{j/2}\phi_{j,k}(2^j x - k)$ are such that ϕ

and ψ are compactly supported and $\{\tilde{\phi}_{l,k}\}_{k\in\Lambda(l)} \cup \{\tilde{\psi}_{j,k}\}_{j\geqslant l,k\in\Lambda(j)}$ forms an orthonormal basis of $L_2(\mathbb{R})$, denoted by $(\varphi_\lambda)_\lambda$. Then S_m is the space generated by the φ_λ for $\lambda \in \{(j, k), 0 \leqslant j \leqslant J, k \in \Lambda(j)\}$ which has dimension $D_m = \sum_{j=0}^J 2^k = 2^{J+1} - 1$ for $J = 0, \ldots, J_n$, where $2^{J_n+1} - 1 \leqslant n$, i.e. $J_n = [\ln(n)/\ln(2)] + 1$ ([.] denotes the integer part). The basis is used on $[0, \pi]$ and extended by parity to $[-\pi, 0]$.

(GW) *General compactly supported periodic wavelets.* For a description of these, see Barron *et al.* (1999, Section 3.2.1).

The definition of f implies some particular interest in the family (T) which is the most commonly used. Regular piecewise polynomials are used to estimate the bound $\|f\|_\infty$ and in the simulation study. General piecewise polynomials are often very flexible for the estimation of f. Wavelets are often considered, for instance by Neumann (1996) in a not necessarily Gaussian context for a similar estimator.

Remark 2.2. The first reason for the interest in general collections of models (and not just the regular ones) lies in the interest in peak detection for spectral densities. The second reason can be found in approximation theory. Let f be in the unit ball of a Sobolev space $W_p^{(\beta)}(\mathbb{R})$, β integer (i.e. $\|f^{(\beta)}\|_p \leqslant 1$). Assume that we are looking for some $\tilde{m} \in \mathscr{M}_n$ and for an optimal approximation $\tilde{f} \in S_{\tilde{m}}$ of f. If $p = 2$, it is well known that what we described as regular collections of models give an adequate approximation with an adequate rate (namely $D_m^{-\beta}$). But this is not true if $p < 2$. For instance, let f belong to $W_1^1(\mathbb{R})$. There is no hope in general of approximating f in L_2 at the optimal rate D_m^{-1} using piecewise constant functions on a partition with D_m equal pieces (see Pinkus 1985, Theorem 1.1, p. 232; Birgé and Massart 2000). But Birman and Solomjak (1967) reached the rate again by considering nonlinear procedures and piecewise polynomials based on some special irregular partition of $[0, 1]$. For details, see the discussion in Birgé and Massart (2000). This is illustrated by Corollary 4.

2.3. Assumptions on the collections of models

The general assumptions required for the collections of models are now described. The general set-up is as follows:

Assumption 3. *Each S_m is a linear finite-dimensional subspace of $L_2([-\pi, \pi])$ with dimension* $\dim(S_m) = D_m$. *Moreover,* $D_n := \max_{m\in\mathscr{M}_n} D_m \leqslant n$.

General collections of models are required to satisfy the following technical condition:

Assumption 4. *Let* $(\varphi_\lambda)_{\lambda\in\Lambda(m,m')}$ *be an orthonormal basis of* $S_m + S_{m'}$ *and* $\beta = (\beta_\lambda)_{\lambda\in\Lambda(m,m')} \in \mathbb{R}^{\dim(S_m+S_{m'})}$, *with* $|\beta|_\infty = \sup_{\lambda\in\Lambda(m,m')} |\beta_\lambda|$. *Then, for all* $m, m' \in \mathscr{M}_n$,

$$\bar{r}_{(m,m')} = \frac{1}{\sqrt{\dim(S_m + S_{m'})}} \sup_{\beta\neq 0} \frac{\|\sum_{\lambda\in\Lambda(m,m')}\beta_\lambda\varphi_\lambda\|_\infty}{|\beta|_\infty} \leqslant C_{\bar{r}}\sqrt{\frac{n}{\dim(S_m + S'_m)}}. \tag{2}$$

Remark 2.3. We know from Barron *et al.* (1999, pp. 322–323) that for the family (W), $\bar{r}_{(m,m')} \leqslant C$, with $C = 2 + \sqrt{2}$, and the same holds for (P), but with $C = \{(r+1)(2r+1)\}^{1/2}$, where r is the degree of the polynomials. In other words, $\bar{r}_{(m,m')}$ can in both these cases be bounded independently of (m, m').

Remark 2.4. The family (GP) satisfies inequality (2) with $C_{\bar{r}} = 2r\sqrt{r}$ (see inequality (3.8) in Barron *et al.* 1999). The family (GW) satisfies (2) with an unspecified $C_{\bar{r}}$ depending on the wavelets (Barron *et al.* 1999, p. 321). The family (T) satisfies (2) with $C_{\bar{r}} = \pi^{-1/2}$, provided that $D_n \leqslant \sqrt{n}$.

Finally, we need the following control on the number models in each collection:

Assumption 5. $\sum_{m \in \mathscr{M}_n} \mathrm{e}^{-L_m D_m} \leqslant \Sigma < +\infty$, for some positive weights L_m.

Remark 2.5. For the families (W) and (P), L_m is of order 1. For the family (P), for example, $L_m = 1/r$ can be used. Since there is at most one model per dimension, $\sum_{m \in \mathscr{M}_n} \mathrm{e}^{-L_m D_m}$ is less than $\sum_{d=1}^n \mathrm{e}^{-d} \leqslant 1/(1 - \mathrm{e}^{-1})$.

Remark 2.6. For the families (GP) and (GW), L_m is of order $\ln(n)$. Consider, for example, piecewise polynomials of degree $r - 1$. There are C_n^d possible subdivisions for a given dimension d. Thus, if we choose $L_m = L_n/r$, then

$$\sum_{m \in \mathscr{M}_n} \mathrm{e}^{-L_m D_m} = \sum_{d=1}^n C_n^d \mathrm{e}^{-L_n d} = (1 + n^{-1})^n - 1 \leqslant \mathrm{e} - 1.$$

Therefore $\Sigma = \mathrm{e} - 1$ can be used when $L_m = \ln(n)/r$. For details, see Barron *et al.* (1999, p. 328).

2.4. The procedure

Let $\{S_m, m \in \mathscr{M}_n\}$ be some finite collection of models as described above. We associate with each S_m the projection estimator $\hat{f}_m$ of f on S_m that minimizes over all t in S_m the contrast function

$$\gamma_n(t) = \int_{-\pi}^{\pi} t^2(\omega)\mathrm{d}\omega - 2\int_{-\pi}^{\pi} t(\omega)I_n(\omega)\mathrm{d}\omega,$$

where

$$I_n(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^n (X_t - \overline{X}_n)\mathrm{e}^{\mathrm{i}\omega t} \right|^2, \qquad \overline{X}_n = \frac{1}{n}\sum_{t=1}^n X_t. \tag{3}$$

First, we look for $\hat{f}_m$ such that

$$\gamma_n(\hat{f}_m) = \min_{t \in S_m} \gamma_n(t). \tag{4}$$

This is the same as looking for $\hat{f}_m$ so as to minimize $\int_{-\pi}^{\pi}(I_n(\omega) - t(\omega))^2 \, d\omega$. Note that $\hat{f}_m = \sum_{\lambda \in \Lambda_m} \hat{a}_\lambda \varphi_\lambda$, with $\hat{a}_\lambda = \int_{-\pi}^{\pi} I_n(\omega)\varphi_\lambda(\omega)d\omega$, if (φ_λ) is an L_2-orthonormal basis of S_m.

Second, we choose $\hat{m}$ among the family $\mathcal{M}_n$ such that

$$\hat{m} = \underset{m \in \mathcal{M}_n}{\arg\min} \, [\gamma_n(\hat{f}_m) + \mathrm{pen}(m)],$$

where $\mathrm{pen}(m)$ is a penalty function to be specified later. Then the estimator of f that we wish to study is $\tilde{f} = \hat{f}_{\hat{m}}$. Denoting by $\|g\|$ the usual L_2-norm $\|g\|^2 = \int_{-\pi}^{\pi} g^2(\omega)d\omega$, we wish to bound the risk $\mathrm{E}(\|f - \tilde{f}\|^2)$.

2.5. The aim

Let us briefly explain what kind of bound for the L_2-risk we are looking for. Under suitable assumptions on f, it is easy to prove (see Appendix) that the risk of $\hat{f}_m$ is bounded by

$$\mathrm{E}\|f - \hat{f}_m\|^2 \leqslant \|f - f_m\|^2 + s\frac{\dim(S_m)}{n}, \tag{5}$$

where f_m is the orthogonal L_2 projection of f on S_m and s is a constant depending on f. The first term of the right-hand side of (5) is known as the bias term, and the second as the variance term of the risk.

We aim to propose some suitable data-driven selection procedure to select some $\hat{m}$ from $\mathcal{M}_n$ in such a way that the projection estimator $\hat{f}_{\hat{m}}$ performs almost as well as the best $\hat{f}_m$ over the collection, the criterion for comparing the estimators being the L_2-risk. This implies that for a good choice of the penalty function, we aim to find inequalities

$$\mathrm{E}\|f - \hat{f}_{\hat{m}}\|^2 \leqslant C \inf_{m \in \mathcal{M}_n} \left[\|f - f_m\|^2 + L_m \frac{\dim(S_m)}{n} \right], \tag{6}$$

where C is a multiplicative constant and the L_ms are weights. When $L_m = 1$, our estimate turns out to minimize the risks among the collection, up to the multiplicative constant C.

3. General theorems

3.1. The bound on *f* is known

With L_2-L_∞ chaining methods (Barron *et al.*, 1999) we can prove our main theorem, which holds for general collections of models, whether regular or not.

Theorem 1. *Suppose that Assumptions* 1 *and* 2 *hold and consider a family of models satisfying Assumptions* 3–5. *Then the penalized projection estimator* (PPE) *defined by* (4) *and*

$$\tilde{f} = \hat{f}_{\hat{m}}, \quad where \; \hat{m} = \underset{m \in \mathcal{M}_n}{\arg\min} \left(\min_{t \in S_m} \int_{-\pi}^{\pi} t(\omega)(t(\omega) - 2I_n(\omega))d\omega + \mathrm{pen}(m) \right), \tag{7}$$

with

$$\text{pen}(m) \geq \kappa \|f\|_\infty^2 (1 + C_{\tilde{r}}^2)(1 + L_m^2) D_m/n, \tag{8}$$

in which κ *is a numerical* (*universal*) *constant, satisfies*

$$\mathrm{E}(\|f - \tilde{f}\|^2) \leq C \inf_{m \in \mathcal{M}_n} [\|f - f_m\|^2 + \text{pen}(m)] + 8\frac{M_2}{n} + C'\frac{(1 + C_{\tilde{r}}^2)\Sigma\|f\|_\infty^2}{n}, \tag{9}$$

where C *and* C' *are universal constants.*

From Remarks 2.5 and 2.6, it follows that the penalty is of order D_m/n for regular collections of models ((T), (P), (W)) and of order $\ln^2(n)D_m/n$ for general collections ((GP), (GW)).

3.2. Estimating the bound on *f*

Obviously the main drawback of Theorem 1 is that the (deterministic) penalty depends on an unknown quantity, namely $\|f\|_\infty$. In this subsection, we give an extension that leads to consideration of a random penalty function which no longer depends on the bound on f. The estimator of $\|f\|_\infty$ is constructed as proposed by Birgé and Massart (1997) in the following way: we take the infinite norm of $\hat{f}_n^P$, the (empirical) orthogonal projection of the periodogram I_n on a space of regular dyadic piecewise polynomials (a space of the family (P) based on dyadic partitions) of dimension depending on n and denoted by N_n. We denote by f_n^P the L_2-orthogonal projection of f on the same space.

Theorem 2. *Suppose that Assumptions* 1 *and* 2 *hold and consider a family of models satisfying Assumptions* 3–5. *Then if* $\|f - f_n^P\|_\infty \leq \frac{1}{4}\|f\|_\infty$ *and*

$$N_n \leq \frac{\kappa'}{(r+1)^2}\frac{n}{\ln(n)},$$

the PPE $\tilde{f}$ *defined by* (4) *and* (7) *with*

$$\text{pen}(m) \geq \kappa \|\hat{f}_n^P\|_\infty^2 (1 + C_{\tilde{r}}^2)(1 + L_m^2) D_m/n \tag{10}$$

where κ *and* κ' *are numerical constants, satisfies*

$$\mathrm{E}(\|f - \tilde{f}\|^2) \leq C \inf_{m \in \mathcal{M}_n} \left[\|f - f_m\|^2 + \|f\|_\infty^2 (1 + C_{\tilde{r}}^2)\frac{L_m^2 D_m}{n} \right]$$

$$+ \frac{C(M_2, \Sigma, \|f\|_\infty, C_{\tilde{r}})}{n}, \tag{11}$$

where C *is universal constant and* $C(M_2, \Sigma, \|f\|_\infty, C_{\tilde{r}})$ *a constant depending on* M_2, Σ, $\|f\|_\infty$ *and* $C_{\tilde{r}}$.

It remains to examine the condition $\|f - f_n^P\|_\infty \leq \frac{1}{4}\|f\|_\infty$. If some regularity conditions

are set on f, then it is generally satisfied, as is known from approximation theory. Indeed for $f \in \mathscr{B}_{\beta,p,\infty}$, $\beta > 1/p$, we know from DeVore and Lorentz (1993) that

$$\|f - f_n^P\|_\infty \leq C(\beta)|f|_{\beta,p} N_n^{-(\beta - 1/p)},$$

where the definitions are quickly recalled below. Therefore $\|f - f_n^P\|_\infty \leq \frac{1}{4}\|f\|_\infty$ if $N_n \geq (4C(\beta)|f|_{\beta,p}/\|f\|_\infty)^{1/(\beta - 1/p)} := C(f, \beta, p)$, where $C(f, \beta, p)$ is a constant depending on f, β and p.

Note that if C in (9) and (11) is allowed to depend on M_2, $C_{\bar{r}}$, Σ and $\|f\|_\infty$, then both inequalities can be written

$$\mathrm{E}(\|f - \tilde{f}\|^2) \leq C \inf_{m \in \mathscr{M}_n} \left[\|f - f_m\|^2 + \frac{L_m^2 D_m}{n} \right].$$

3.3. Adaptation to unknown smoothness

It is easy to derive adaptation results from inequalities such as (9) or (11). We give here an adaptation result with respect to the unknown smoothness of f. Moreover, we consider $f(2\pi\cdot)$ on $[0, 1]$. We recall quickly that a function f belongs to the Besov space $\mathscr{B}_{\beta,p,\infty}([0, 1])$ if it satisfies

$$|f|_{\beta,p} = \sup_{y > 0} y^{-\beta} w_d(f, y)_p < +\infty, \quad d = [\beta] + 1, \quad |t|_\infty = \sup_{x,y \in [0,1]} |t(x) - t(y)|,$$

where $w_d(f, y)_p$ denotes the modulus of smoothness. For a precise definition of these notions, see DeVore and Lorentz (1993, Section 2.7), where it also proved that $\mathscr{B}_{\beta,p,\infty}([0, 1]) \subset \mathscr{B}_{\beta,2,\infty}([0, 1])$ for $p \geq 2$. This justifies our now restricting our attention to $\mathscr{B}_{\beta,2,\infty}([0, 1])$.

Let us consider the three families (T), (P) and (W) previously described. It is well known from classical approximation theory that if $r \geq \beta$ and $f \in \mathscr{B}_{\beta,2,\infty}([0, 1])$, then

$$\|f - f_m\| \leq C(\beta)|f|_{\beta,2} D_m^{-\beta}, \tag{12}$$

where r is the regularity. See, for instance, Barron *et al*. (1999, Lemma 12, p. 404). It immediately follows from (9) and (11) that the following result holds:

Corollary 3. *Suppose that Assumptions* 1 *and* 2 *hold and that* $f(2\pi\cdot)$ *belongs to some* $\mathscr{B}_{\beta,2,\infty}([0, 1]) \cap L_\infty([0, 1])$. *Moreover, assume that* $\{S_m\}_{m \in \mathscr{M}_n}$ *is one of the families* (T)*,* (P) *or* (W) *with* $\beta > \frac{1}{2}$. *Let* $\tilde{f}$ *be the estimator defined by* (7) *with penalty* (8) *or* (10). *Then*

$$\mathrm{E}[\|f - \tilde{f}\|^2] \leq C|f|_{\beta,2}^{2/(1+2\beta)} n^{-2\beta/(2\beta+1)},$$

where the constant C *only depends on* β*,* M_2*,* $\|f\|_\infty$ *and* $C_{\bar{r}}$.

In other words

$$\sup_{f \in B_{\beta,2}(R)} \mathrm{E}[\|f - \tilde{f}\|^2] \leq C n^{-2\beta/(2\beta+1)},$$

where $B_{\beta,2}(R) = \{f \in \mathcal{B}_{\beta,2,\infty}, |f|_{\beta,2} \leqslant R\}$ and where the constant C now depends on β, R, M_2, $\|f\|_{\infty}$ and $\bar{r}$.

Remark 3.1. It is known from Bentkus (1985) that the minimax risk over some Besov ball $\{f/|f|_{\beta,2} \leqslant R\}$ cannot be smaller than $C'R^{2/(1+2\beta)}n^{-2\beta/(2\beta+1)}$, where C' is some positive constant, and therefore our estimator is minimax (up to constants) over all such balls simultaneously.

Remark 3.2. For the family (T), the bias–variance trade-off leads to the choice of $D_m = n^{1/(1+2\beta)}$, which is less than $\sqrt{n}$ when $\beta \geqslant 1/2$.

Remark 3.3. The rate obtained by Neumann (1996) in the non-Gaussian case with random thresholding and wavelet estimates is $(n/\ln(n))^{-2\beta/(2\beta+1)}$. We have improved this rate in the Gaussian case.

Remark 3.4. We recover the rate $\ln(n)/n$ found by Efromovich (1998) in the particular case of a spectral density with $\gamma_k \leqslant C\,e^{-\rho k}$. Indeed then, using the family (T), $\|f - f_m\|$ is of order $e^{-\rho D_m}$; thus the optimal choice on the collection is $D_m = 2\ln(n)/\rho$, which gives the global bound $\ln(n)/n$.

For general collections of models, if $L_m = \ln(n)$ we find the same rate up to a $\ln(n)$ factor, that is to say, we have, for (GP) or (GW), $\sup_{f \in B_{\beta,2}(R)} \mathrm{E}[\|f - \tilde{f}\|^2] \leqslant C(n/\ln^2(n))^{-2\beta/(2\beta+1)}$. But general collections of models can also be used, as per Remark 2.2, to handle the case of Besov bodies of the form $B_{\beta,p,\infty}(R_1, R_2) = \{t \in \mathcal{B}_{\beta,p,\infty}([0, 1])/|t|_{\beta,p} \leqslant R_1, |t|_{\infty} \leqslant R_2\}$ with $1 \leqslant p \leqslant 2$. The method is thus the same as in Baraud *et al.* (1999) and implies the same result.

More precisely, a modification of the general spaces allows us to use results of the compression algorithm field given in Birgé and Massart (2000). Let us define

$$K_j = [\mathcal{L}(2^{J-j})2^J], \qquad \text{with } \mathcal{L}(x) = \left(1 - \frac{\ln x}{\ln 2}\right)^{-s}, \quad s > 2, \, x \in (0, 1),$$

and

$$L(s) = 1 + \sum_{j=0}^{\infty} \frac{1 + (s + \ln(2))j}{(1 + j)^s}. \tag{13}$$

Then we define, using the notation of Section 2.2, the following restriction of (W):

(W') *Wavelets of regularity r.* For $J \in \{0, 1, \ldots, J_n\}$ with $\mathcal{P}(M)$ denoting the set of all the subsets of M, let

$$\mathcal{M}_n^J = \left\{\bigcup_{j=0}^{J_n} m_j / m_j \in \begin{cases} \mathcal{P}(\Lambda(j)) & \text{if } 0 \leqslant j < J, \\ \{A \in \mathcal{P}(\Lambda(j))/|A| = K_j\} & \text{if } j \geqslant J \end{cases}\right\}.$$

Let $\mathscr{M}_n = \bigcup_{j=0}^{J_n} \mathscr{M}_n^J$ and define S_m as the linear span of the φ_λ for $\lambda \in m$, $m \in \mathscr{M}_n$, and by $\mathbb{1}: x \mapsto 1$, $\forall x \in [0, 1]$.

An analogous restriction of (GP) could be given for piecewise polynomials. Then the following result holds:

Corollary 4. *Suppose that Assumptions* 1 *and* 2 *hold. Consider the collection of models* (W′) *with* $2^{J_n} \geqslant n/\Gamma$ *for some* $\Gamma \geqslant 1$, *and* $2^{J_n} \leqslant n$. *Let* $p \in [1, +\infty[$ *and*

$$\left(\frac{1}{2} - \frac{1}{p}\right)_+ \leqslant \beta_p := \begin{cases} \dfrac{1}{2}\left(\dfrac{1}{p} - \dfrac{1}{2}\right)\left[1 + \sqrt{\dfrac{2 + 3p}{2 - p}}\right] & \text{if } p < 2 \\[2mm] 0 & \text{otherwise.} \end{cases}$$

If $\beta_p < \beta \leqslant r$, *then* $\forall R_1, R_2 \in \mathbb{R}^+ \times \mathbb{R}^+$, *the PPE* $\tilde{f}$ *defined by* (7) *with penalty defined by* (8) *or* (10) *and by* $L_m = L(s)$ *given by* (13) $\forall m \in \mathscr{M}_n$, *satisfies*

$$\sup_{f \in B_{\beta, p, \infty}(R_1, R_2)} \mathrm{E}[\|f - \tilde{f}\|^2] \leqslant C_1 n^{-2\beta/(2\beta+1)},$$

where the constant C_1 *depends on* β, R_1, R_2 *and* Γ.

For the proof of this result, see the proof of Proposition 4.1 in Baraud *et al.* (1999).

4. The long-memory case

In this section, we study the case of unbounded f. This case, often known as the 'long-memory' case, has been extensively studied in recent years. It occurs for fractional models when the spectral density satisfies:

$$f(\lambda) \sim K\lambda^{-2\alpha} \qquad \text{when } \lambda \to 0, \, 0 < \alpha < \tfrac{1}{2} \tag{14}$$

which implies that $\gamma_k \sim C_\gamma k^{2\alpha-1}$ when $k \to +\infty$ (see Adenstedt 1974). Therefore, Assumption 2 is not satisfied for $0 < \alpha < \tfrac{1}{2}$. An example of such models is given by the fractional ARMA models described in Remark 2.1 above when $0 < \alpha < \tfrac{1}{2}$.

In this framework, we can prove the next theorem, which must be considered as a robustness result only, for two reasons. First, there is a constraint on α which cannot easily be checked. Second, the penalty depends not on the bound on f, but on the bound on $f(\omega)\omega^{2\alpha}$ which, like α, is unknown and not easy to estimate. Note that it also depends on the L_2-norm $\|f\|$ which may be estimated in the light of work by Laurent and Massart (1998).

Theorem 5. *Suppose that Assumption* 1 *holds and that f satisfies*

$$f(\lambda) \leqslant K\lambda^{-2\alpha}, \qquad \forall \lambda \in [-\pi, \pi], \quad \text{with } 0 < \alpha < \tfrac{1}{4}. \tag{15}$$

Consider the family of models (T) *with* $D_n \leqslant n$. *Then the PPE* $\tilde{f}$ *defined by* (7) *with*

$$\text{pen}(m) \geqslant \kappa(K^2 + \|f\|^2 + C_\gamma^2)\ln^2(D_m)D_m/n,$$

where κ is a numerical constant, satisfies

$$\mathrm{E}(\|f - \tilde{f}\|^2) \leqslant C \inf_{m \in \mathscr{M}_n}[\|f - f_m\|^2 + \text{pen}(m)] + C' \frac{(K^2 + \|f\|^2)}{n}, \qquad (16)$$

where C and C' are numerical constants and C_γ is such that $\gamma_k \leqslant C_\gamma k^{2\alpha-1}$ for all $k \geqslant 1$.

The restriction on α comes from the non-square integrability of f when α is greater than $\frac{1}{4}$.

Remark 4.1. This result shows that we keep (up to a factor $\ln^2(D_m)$) the same order for the penalty and for the variance term of the risk as in Theorems 1 and 2, even for non-bounded spectral densities (or non-summable covariance functions). The loss here comes from the order of $\|f - f_m\|$, which is different from that found for classical functions in Besov spaces.

For instance, since $f(\lambda) = (1/2\pi)(\gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j \cos(\lambda j))$ then $\|f - f_m\|^2$ has the same order as $\sum_{j \geqslant D_m}\gamma_j^2$; and since, in the fractional case, $\gamma_j \sim C_\gamma j^{2\alpha-1}$, $\|f - f_m\|^2$ is of order $D_m^{4\alpha-1}$. Optimizing both terms on the right-hand side of (16) gives a rate of order $(n/\ln^2(n))^{-(1-4\alpha)/(1+(1-4\alpha))}$. This implies a rate of convergence tending to 0 when α tends to $\frac{1}{4}$. This is what is often obtained in such cases: for example, Hall and Hart (1990) also find a rate of convergence which has the same feature, in a regression problem with long-range dependent errors.

Remark 4.2. Since Assumption 2 is not fulfilled, the result of Proposition 1 no longer holds. Equation (37) in the Appendix implies that it can be replaced by

$$\|f - \mathrm{E}(I_n)\|^2 = O(n^{4\alpha-1}) \qquad \text{if } |\gamma_k| \leqslant C_\gamma k^{2\alpha-1}, \forall k \geqslant 1. \qquad (17)$$

This is the reason why Theorem 5 holds for trigonometric bases only: improving the bound in (17) requires a precise order for the scalar product $\langle f - \mathrm{E}(I_n), s \rangle$ for $s \in S_m$. Since $f - \mathrm{E}(I_n)$ admits a trigonometric development, the order is better if S_m is a trigonometric space.

Remark 4.3. The proof of Theorem 5 does not rely on a chaining method but on some simpler arguments. The method used to prove Theorem 1 and 2 would give a result of the same kind but with a different penalty function. The loss in the penalty implies a loss in the global rate (which is therefore less accurate than that found in Remark 4.1). More precisely, in the case of models satisfying Assumptions 4 and 5 and under the same assumptions on f and γ_k as in Theorem 5, we would find, for $0 < \alpha < \frac{1}{4}$,

$$\text{pen}(m) \geqslant \kappa K^2(1 + C_{\tilde{r}}^2)(1 + L_m)^2 D_m/n^{1-4\alpha},$$

where κ is a numerical constant, and a penalized least-squares estimator $\tilde{f}$ satisfying

$$\mathrm{E}(\|f - \tilde{f}\|^2) \leqslant C \inf_{m \in \mathscr{M}_n}[\|f - f_m\|^2 + \text{pen}(m)] + C' \frac{(1 + C_{\tilde{r}}^2)K^2\Sigma}{n^{1-4\alpha}} + C'' \frac{C_\gamma^2}{n^{1-4\alpha}}, \qquad (18)$$

where C, C' and C'' are numerical constants and C_γ is such that $\gamma_k \leqslant C_\gamma k^{2\alpha-1}$ for all $k \geqslant 1$. See the note on the proof of this result at the end of Section 6.

5. Simulations

The programs for our simulations were implemented in collaboration with Y. Rozenholc using the Matlab programming environment. Attempts to test the performance of the method described in this paper have already been made by Birgé and Rozenholc (2000) and Comte and Rozenholc (2000). They are very convenient to use, but have high initial cost; indeed, the constants in the penalty functions are unknown. Therefore, many investigations must be carried out over large families of theoretical functions f to find the right formula. On the other hand, once the right calibration is found, the method is very flexible and all choices are data-driven. For instance, one can try both regular and irregular piecewise polynomials with arbitrary degree (say) 0, 1, 2, 3 to be automatically chosen on each considered subinterval of any subdivision. Various degrees have been tested for regression and volatility function estimation by Comte and Rozenholc (2000). General histograms have been investigated for density estimation of i.i.d. random variables by Rozenholc, but this implies a great number of models to examine: the algorithms must be highly optimized for the computational time to remain reasonable.

This is the reason why we have concerned ourselves here with simple regular histograms. This collection of models quickly gives a very good idea of the potential of the method. We simulated the same model as Neumann (1996), namely

$$X_t = Y_t + \sigma Z_t,$$

where Y_t is an ARMA(2,2) process,

$$Y_t + a_1 Y_{t-1} + a_2 Y_{t-2} = b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + b_2 \varepsilon_{t-2},$$

and $\{\varepsilon_t\}$, $\{Z_t\}$ are independent Gaussian white noise processes with unit variance. The constants were chosen as $a_1 = 0.2$, $a_2 = 0.9$, $b_0 = 1$, $b_1 = 0$, $b_2 = 1$ and $\sigma = 0.5$. We simulated $T = 1000$ samples with length $n = 1024$, whereas Neumann (1996) used $T = 100$ samples of the same length: this has no influence on the mean value of the L_2-risk nor on its standard deviation $\hat{v}$, but may slightly reduce the $\pm 95\%$ intervals computed as in Neumann (1996) as $1.96\,\hat{v}/\sqrt{T}$.

A given dimension $D_m = d$ is associated with a histogram basis

$$\varphi_j^{(d)} = \sqrt{\frac{d}{\pi}} \mathbb{1}_{[\pi j/d,\pi(j+1)/d[}, \qquad \text{for } j = 0, 1, \ldots, d-1,$$

and with the estimated coefficients

$$\hat{a}_j^{(d)} = \sqrt{\frac{d}{\pi}} \left\{ \frac{c_0}{2d} + \frac{1}{\pi} \sum_{r=1}^{n-1} \frac{c_r}{r} \left[\sin\left(\frac{\pi(j+1)r}{d}\right) - \sin\left(\frac{\pi jr}{d}\right) \right] \right\}$$

for $j = 0, 1 \ldots, d-1$, where $c_r = c_{r,n}$, $r = 0, 1, \ldots, n-1$, are the empirical covariances

defined by (32) in the Appendix. The penalized contrast values leading to the selection of the optimal dimension among $d = 1, 2, \ldots, 100$ are

$$-\sum_{j=0}^{d-1}\left(\hat{a}_j^{(d)}\right)^2 + \kappa C_\infty^2(d_{\text{ref}})\frac{d}{n}, \tag{19}$$

where $C_\infty(d_{\text{ref}}) = \max_{0 \leqslant j \leqslant d_{\text{ref}}-1} |\hat{a}_j^{(d_{\text{ref}})}|$ is the estimate of $\|f\|_\infty$ and κ is a constant.

We experimented with several values of the constant κ and of the dimension d_{ref}. We found $d_{\text{ref}} = [n/36\ln(n)] = 4$ to be a good choice: this is the maximal value authorized by Theorem 2. Since the constant $1/36$ in the bound seemed somewhat arbitrary, we expected a much higher dimension to be chosen. We also found $\kappa = 4$ to be a good choice. The results of these experiments are given in Tables 1 and 2, in terms of the L_2-risk and the confidence intervals (CI). Table 3 recalls Neumann's results as a benchmark.

We can see that with this rough version of our method, we already achieve results as good as those of Neumann (1996), who uses a somewhat more complicated wavelet thresholding. Figure 1 illustrates also the type of estimated spectral density found with our method. The oscillations of our histograms are clearly analogous to those of Neumann's wavelets.

Table 1. Adaptive histogram method: different choices of κ

κ	3	3.5	4	4.5	5
L_2-risk	0.053	0.049	0.048	0.047	0.048
$\pm 95\%$ CI	0.0026	0.0023	0.0021	0.0018	0.0018

$T = 1000$ samples of size $n = 1024$. The penalized contrast is computed as in (19) with $d_{\text{ref}} = 4$.

**Table 2.** Adaptive histogram method: different choices of d_{ref}

d_{ref}	2	3	4	5	6	10	25
L_2-risk	0.048	0.050	0.048	0.054	0.047	0.049	0.052
$\pm 95\%$ CI	0.0018	0.0024	0.0021	0.0026	0.0022	0.0023	0.0025

$T = 1000$ samples of size $n = 1024$. The penalized contrast is computed as in (19) with $\kappa = 4$.

Table 3. Neumann's results

Estimator	Kernel	$w(1, \text{u})$	$w(2, \text{u})$	$w(1, \text{s})$	$w(2, \text{s})$
L_2-risk	0.042	0.066	0.054	0.044	0.040
$\pm 95\%$ C.I.	0.0032	0.0042	0.0053	0.0038	0.0043

$T = 100$ samples of size $n = 1024$ $w(i, \text{u})$ and $w(i, \text{s})$ refer to unshifted and shifted versions of the thresholding with threshold $\hat{\lambda}_{j,k}^{(i)}$ developed by Neumann. The kernel method is based on the optimal bandwith (unknown in practice).
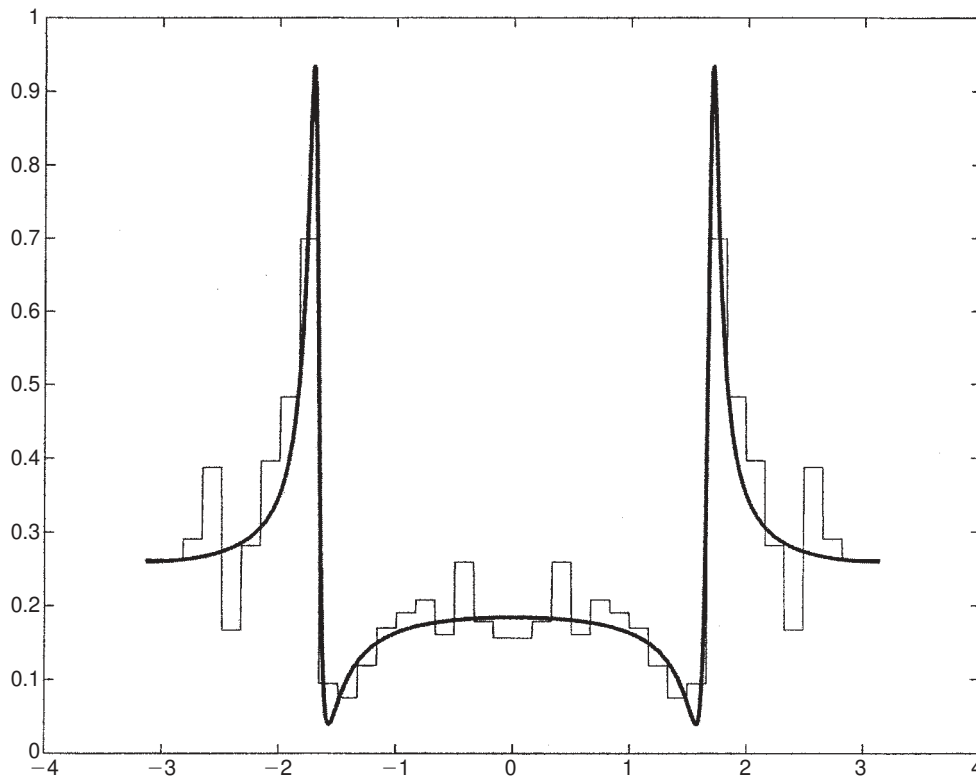
Figure 1. True spectral density and regular histogram estimation.

Of course, true piecewise polynomials may be smoother and therefore nicer, but it is not certain that the values of the L_2-risk would be significantly better. What appears in the example considered here is that irregular partitions would lead to very relevant choices. The dimensions selected by the algorithm with regular partitions are quite high in order to capture the peak. Irregular partitions would lead to much smaller dimensions and more adapted choices. This is a further programming challenge.

6. Proofs

6.1. Proof of Theorem 1

6.1.1. *Decomposition of the contrast*

We start with a decomposition of the contrast which is common to the proofs of all theorems. Let $\langle u, v \rangle = \int_{-\pi}^{\pi} u(\omega)v(\omega)\mathrm{d}\omega$, and $\|u\|^2 = \langle u, u \rangle$. Then we can write:

$$\gamma_n(t) = \|t\|^2 - 2\langle I_n, t \rangle = \|t - f\|^2 - 2\langle I_n - f, t \rangle - \|f\|^2.$$

By definition $\tilde{f} = \hat{f}_{\hat{m}}$ satisfies, $\forall m \in \mathcal{M}_n$, the inequality $\gamma_n(\tilde{f}) + \mathrm{pen}(\hat{m}) \leq \gamma_n(f_m) + \mathrm{pen}(m)$, where f_m denotes the orthogonal projection of f on S_m. This implies that, $\forall m \in \mathcal{M}_n$,

$$\|f - \tilde{f}\|^2 - 2\langle I_n - f, \tilde{f}\rangle + \mathrm{pen}(\hat{m}) \leq \|f - f_m\|^2 - 2\langle I_n - f, f_m\rangle + \mathrm{pen}(m),$$

that is,

$$\|f - \tilde{f}\|^2 \leq \|f - f_m\|^2 + 2\langle I_n - f, \tilde{f} - f_m\rangle + \mathrm{pen}(m) - \mathrm{pen}(\hat{m})$$

$$\leq \|f - f_m\|^2 + 2\langle f - \mathrm{E}(I_n), f_m - \tilde{f}\rangle + 2\langle I_n - \mathrm{E}(I_n), \tilde{f} - f_m\rangle \qquad (20)$$

$$+ \mathrm{pen}(m) - \mathrm{pen}(\hat{m}).$$

Then there are two different strategies for the proofs of Theorems 1 and 5. But they both require a fundamental Bernstein-type inequality. For the latter the Gaussian feature is crucial.

6.1.2. A Bernstein-type inequality

Let us write $X = (X_1, \ldots, X_n)'$ and $\overline{X} = (1/n)\sum_{i=1}^n X_i$. For $u \in S_{m'}$, let

$$Z_n(u) = \langle I_n - \mathrm{E}(I_n), u\rangle. \qquad (21)$$

Writing $e = (1, \ldots, 1)'$, which is of size $n \times 1$, Z_n can also be written as

$$Z_n(u) = \frac{1}{n}[(X - \overline{X}e)' T_n(u)(X - \overline{X}e) - \mathrm{E}((X - \overline{X}e)' T_n(u)(X - \overline{X}e))],$$

where $T_n(u)$ is the Toeplitz matrix associated with the function u, that is,

$$[T_n(u)]_{j,k} = \int_{-\pi}^{\pi} u(\omega)\mathrm{e}^{-i\omega(j-k)}\,\mathrm{d}\omega, \qquad 1 \leq j, k \leq n.$$

Since we aim to estimate a positive function, we can restrict our attention to positive functions of $S_{m'}$, without loss of generality. We prove in this section the following result:

Proposition 2. *Assume that the stationary sequence $X_1, \ldots, X_n$ satisfies Assumption* 1 *and let f denote its spectral density. Then for any even bounded positive function u,*

$$\mathrm{P}(Z_n(u) \geq 2\|f\|_\infty(\|u\|\sqrt{x} + \|u\|_\infty x)) \leq \mathrm{e}^{-nx} \qquad (22)$$

and

$$\mathrm{P}(Z_n(u) \geq 2\|u\|_\infty(\|f\|\sqrt{x} + \rho(T_n(f))x)) \leq \mathrm{e}^{-nx}. \qquad (23)$$

$\|f\|_\infty$ *in* (22) *can be replaced by $\rho(T_n(f))$, where $\rho(M)$ denotes the spectral radius of a symmetric matrix M.*

Proof. Since u is even positive on $[-\pi, \pi]$, $T_n(u)$ is a symmetric positive definite matrix in $\mathbb{M}_n(\mathbb{R})$ (see Dacunha-Castelle and Duflo 1993, p. 41). We denote by $\sqrt{T_n(u)}$ the usual symmetric square root of a symmetric positive matrix. Then

$$Z_n(u) = \frac{1}{n}[\| \sqrt{T_n(u)}(X - \overline{X}e)\|_n^2 - \mathrm{E}(\| \sqrt{T_n(u)}(X - \overline{X}e)\|_n^2)],$$

with $\|x\|_n$ denoting the Euclidian norm in $\mathbb{R}^n$. We use the following inequality, valid in the Gaussian case only, which is a result proved by Laurent and Massart (1998, Lemma 1, p. 24):

Proposition 3. *Let* $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_p)'$ *be a vector of n i.i.d. centred Gaussian random variables with common variance* σ^2, *and let A be some matrix in* $\mathbb{M}_n(\mathbb{R})$. *Let* $\lambda_1, \ldots, \lambda_p$ *be the eigenvalues of the positive symmetric matrix* AA', $v^2 = \sum_{i=1}^p \lambda_i^2$ *and* $b = \max_{i=1,\ldots,p} \lambda_i$. *Then* $\chi^2(A) = \|A\varepsilon\|_p^2$ *satisfies, for each* $x > 0$,

$$\mathrm{P}(\chi^2(A) \geqslant \mathrm{tr}(AA')\sigma^2 + 2\sigma^2\sqrt{v^2 x} + 2b\sigma^2 x) \leqslant \mathrm{e}^{-x}, \tag{24}$$

where tr *denotes the trace of the matrix.*

Let $\chi^2 = n\langle I_n, u \rangle = (X - \overline{X}e)'T_n(u)(X - \overline{X}e)$. If $\mu = \mathrm{E}(X_1)$, since $X - \overline{X}e = (X - \mu e) - (\overline{X} - \mu)e$, we can assume without loss of generality that $\mu = 0$ (otherwise we work with $\tilde{X} = X - \mu e$). Let H be the hyperplane $\mathrm{Vect}(e)^\perp$ (i.e. the orthogonal in the Euclidian space $\mathbb{R}^n$ of the linear space generated by the vector e). Since $\overline{X}e$ is the orthogonal projection of X on $\mathrm{Vect}(e)$, $X - \overline{X}e = P_H X$, where P_H is the matrix of the orthogonal projection on H. Therefore, $X - \overline{X}e$ is a Gaussian vector with degenerated variance, $\mathrm{var}(X - \overline{X}e) = P_H \Sigma_X P_H'$, where the rank of P_H is $n - 1$ and Σ_X denotes the variance matrix of X. Since Σ_X is symmetric positive semi-definite (it would be positive definite and thus invertible if we had assumed f to be bounded away from zero), Σ_X can be written $\Sigma_X = P'DP$ where $P'P = Id_n$ (the $n \times n$ identity matrix) and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_p, 0, \ldots, 0)$ with $\lambda_i > 0$ for $i = 1, \ldots, p$, with $p \leqslant n$ denoting the rank of Σ_X. Let $Y = \Delta PX$, where $\Delta = \mathrm{diag}(1/\sqrt{\lambda_1}, \ldots, 1/\sqrt{\lambda_p}, 1 \ldots, 1)$. We have $\mathrm{var}(Y) = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$, where there are p ones and $n - p$ zeros on the diagonal. Since $\mathrm{E}(Y) = 0$, this implies that $Y_{p+1}, \ldots, Y_n$ are almost surely null random variables. Moreover,

$$\chi^2 = \| \sqrt{T_n(u)}P_H X\|_n^2 = \| \sqrt{T_n(u)}P_H(\Delta P)^{-1}Y\|_n^2 = Y'(\Delta P)'^{-1}P_H'T_n(u)P_H(\Delta P)^{-1}Y$$

$$= (Y')^{(p)}[(\Delta P)'^{-1}P_H'T_n(u)P_H(\Delta P)^{-1}]_p Y^{(p)},$$

where $Y^{(p)} = (Y_1, \ldots, Y_p)$, and $M_p = (m_{i,j})_{1\leqslant i,j\leqslant p}$ is the $p \times p$ truncated matrix deduced from $M = (m_{i,j})_{1\leqslant i,j\leqslant n}$. We can apply inequality (24) to χ^2 with

$$AA' = [(\Delta P)'^{-1}P_H'T_n(u)P_H(\Delta P)^{-1}]_p = [BB']_p, \qquad B = (\Delta P)'^{-1}P_H'\sqrt{T_n(u)}.$$

First, it is clear that $\mathrm{E}(\chi^2) = \mathrm{tr}(AA')$ and $\chi^2 - \mathrm{E}(\chi^2) = nZ_n(u)$. Moreover, $b = \rho(AA') \leqslant \rho(BB')$ since any eigenvalue of AA' is clearly an eigenvalue of BB' (complete the eigenvector with zeros) and

$$v^2 = \mathrm{tr}((AA')^2) \leqslant \mathrm{tr}((BB')^2) = \mathrm{tr}(\Sigma_X P_H'T_n(u)P_H \Sigma_X P_H'T_n(u)P_H).$$

Then we use the inequalities $\text{tr}((MN)^2) \leq \rho(M)^2\,\text{tr}(N^2)$ and $\text{tr}((MN)^2) \leq \rho(N)^2\,\text{tr}(M^2)$ for any pair (M, N) of symmetric matrices. Since $\rho(P_H) = 1$ and $\rho(\Sigma_X) = \rho(T_n(f)) \leq \|f\|_\infty$, we have $\rho(P_H\Sigma_X P'_H) \leq \rho(\Sigma_X)$ and therefore we find that $\text{tr}((P_H\Sigma_X P'_H T_n(u))^2) \leq \|f\|_\infty^2\,\text{tr}(T_n(u)^2)$. Moreover, we know from Davies (1973, Lemma 3.1(iii)) that, for any function $u \in L_2([-\pi, \pi])$,

$$\lim_{n \to +\infty} \uparrow \frac{1}{n}\sum_{i=1}^{n}\lambda_i^2(T_n(u)) = \int_{-\pi}^{\pi} u^2(x)\mathrm{d}x$$

for any Toeplitz matrix $T_n(u)$ and with $\lambda_i(T_n(u))$, $i = 1, \ldots, n$ denoting the eigenvalues of $T_n(u)$. Then $\text{tr}(T_n(u)^2) = \sum_{i=1}^{n}\lambda_i^2$ is of order $n\|u\|^2$ when n grows to infinity, and in particular $\text{tr}(T_n(u)^2) \leq n\|u\|^2$. Finally, since $\rho(BB') = \sqrt{\rho((BB')^2)}$, we also find that $b \leq \rho(\Sigma_X)\rho(T_n(u)) \leq \|f\|_\infty\|u\|_\infty$. Then inequality (24) gives (22).

Inequality (23) is derived in the same way but with the inequalities $b \leq \rho(T_n(f))\|u\|_\infty$ and $v^2 \leq \rho(T_n(u))^2\,\text{tr}(T_n(f)^2) \leq \|u\|_\infty^2\text{tr}(T_n(f)^2)$. Then we note that $\text{tr}(T_n(f)^2)$ is less than $n\int_{-\pi}^{\pi} f^2(x)\mathrm{d}x$. $\square$

6.1.3. Proof of Theorem 1

Since m is fixed and m' is varying, let $S(m') = S_m + S_{m'}$, $D(m') = \dim(S_m + S_{m'}) \leq D_m + D_{m'}$ and $\Lambda(m')$ a set such that $(\varphi_\lambda)_{\lambda \in \Lambda(m')}$ is an orthonormal basis of $S(m')$. Then from (20) we obtain

$$\|f - \tilde{f}\|^2 \leq \|f - f_m\|^2 + 2\langle f - \mathrm{E}(I_n), f_m - \tilde{f}\rangle + 2\|\tilde{f} - f_m\|G(\hat{m}) + \text{pen}(m) - \text{pen}(\hat{m}),$$

with Z_n defined by (21) and

$$G(m') = \sup_{u \in B_{m,m'}(0,1)} \langle I_n - \mathrm{E}(I_n), u\rangle = \sup_{u \in B_{m,m'}(0,1)} Z_n(u), \tag{25}$$

where $B_{m,m'}(0, 1)$ is the unit L_2 ball of $S_m + S_{m'}$. This implies that

$$\|f - \tilde{f}\|^2 \leq \|f - f_m\|^2 + 8\|f - \mathrm{E}(I_n)\|^2 + \tfrac{1}{4}\|\tilde{f} - f_m\|^2 + 8G^2(\hat{m}) (26) + \text{pen}(m) - \text{pen}(\hat{m}).$$

Therefore we study the expectation of $G^2(\hat{m})$.

Proposition 4. *Under the assumptions of Theorem* 1,

$$\mathrm{E}[(G^2(m') - p_1(m) - p_1(m'))_+] \leq C\,\mathrm{e}^{-L_{m'}D_{m'}}\frac{(1 + C_{\tilde{F}}^2)\|f\|_\infty^2}{n},$$

where $p_1(m) = \tilde{\kappa}(1 + C_{\tilde{F}}^2)\|f\|_\infty^2(1 + L_m)^2 D_m/n$, *with* C *and* $\tilde{\kappa}$ *being numerical constants.*

From Proposition 4 and Assumption 5, it follows that, if $p(m') = p_1(m') + p_1^{(m)}$,

$$\mathrm{E}[G^2(\hat{m})] \leqslant \mathrm{E}[(G^2(\hat{m}) - p(\hat{m}))_+] + p(\hat{m}) \leqslant \sum_{m' \in \mathscr{M}_n} \mathrm{E}[(G^2(m') - p(m'))_+] + p(\hat{m})$$

$$\leqslant C \frac{(1 + C_{\bar{r}}^2)\|f\|_\infty^2}{n} \sum_{m' \in \mathscr{M}_n} \mathrm{e}^{-L_{m'} D_{m'}} + p_1(\hat{m}) + p_1(m)$$

$$\leqslant C \frac{(1 + C_{\bar{r}}^2)\|f\|_\infty^2 \Sigma}{n} + p_1(\hat{m}) + p_1(m). \tag{27}$$

Using the fact that $\|\tilde{f} - f_m\|^2 \leqslant 2\|f - \tilde{f}\|^2 + 2\|f - f_m\|^2$, (26) can be rewritten

$$\|f - \tilde{f}\|^2 \leqslant 3\|f - f_m\|^2 + 8\|f - \mathrm{E}(I_n)\|^2 + 8G^2(\hat{m}) + 8(p_1(\hat{m})$$

$$+ p_1(m)) - 2\mathrm{pen}(\hat{m}) + 2\mathrm{pen}(m), \tag{28}$$

which, combined with (1) under Assumption 2 and (27), entails that

$$\mathrm{E}(\|f - \tilde{f}\|^2) \leqslant 3\|f - f_m\|^2 + 8\frac{M_2}{n} + C\frac{(1 + C_{\bar{r}}^2)\|f\|_\infty^2 \Sigma}{n} + 3\mathrm{pen}(m),$$

where C is a numerical constant and since $\mathrm{pen}(m) \geqslant 8p_1(m)$ (with $\kappa = 8\tilde{\kappa}$). Finally, the infimum over $\mathscr{M}_n$ can be taken since the result holds for any m.

Proof of Proposition 4. The proof uses arguments similar to those developed in the proof of Theorem 5 in Birgé and Massart (1998) and in the proofs of Proposition 7 and Theorems 8 and 9 in Barron *et al.* (1999). It is based on a chaining argument from which we know (see Barron *et al.* 1999, Lemma 9, p. 400) that, in a linear subspace $S \subset L_2$ of dimension D, we can find a finite δ-net, $T_\delta \subset \mathscr{B}$, where $\mathscr{B}$ denotes the ball of S centred in 0 and with radius σ, and a mapping π from S to T_δ, such that:

- for each $0 < \delta \leqslant \sigma/5$, $|T_\delta| \leqslant (5\sigma/\delta)^D$;
- $\|u - \pi(u)\| \leqslant \delta$, $\forall u \in S$ and $\sup_{u \in \pi^{-1}(t)} \|u - t\|_\infty \leqslant \bar{r}\delta$, $\forall t \in T$, with $\bar{r}$ defined as in (2).

We apply the above result to the linear space $S_m + S_{m'}$ of dimension $D(m')$ and $\bar{r} = \bar{r}_{(m,m')}$ defined by Assumption 4. We consider δ_k-nets, $T_k = T_{\delta_k}$, with $\delta_k = \delta_0 2^{-k}$, where $\delta_0 \leqslant \frac{1}{5}$ is to be chosen later. We set $H_k = \ln(|T_k|)$. Given some point $u \in B_{m,m'}(0, 1)$, we can find a sequence $\{u_k\}_{k \geqslant 0}$ with $u_k \in T_k$ such that $\|u - u_k\|^2 \leqslant \delta_k^2$ and $\|u - u_k\|_\infty \leqslant \bar{r}_{(m,m')}\delta_k$. Therefore we have the following decomposition, which holds for any $u \in B_{m,m'}(0, 1)$:

$$u = u_0 + \sum_{k=1}^\infty (u_k - u_{k-1}).$$

Clearly $\|u_0\| \leqslant \delta_0$, $\|u_0\|_\infty \leqslant \bar{r}_{(m,m')}\delta_0$ and, for all $k \geqslant 1$, $\|u_k - u_{k-1}\|^2 \leqslant 2(\delta_k^2 + \delta_{k-1}^2) = 5\delta_{k-1}^2/2$ and $\|u_k - u_{k-1}\|_\infty \leqslant 3\bar{r}_{(m,m')}\delta_{k-1}/2$.

Let $(\eta_k)_{k \geqslant 0}$ be a sequence of positive numbers that will be chosen later on. We have

$$P\left[\sup_{u\in B_{m,m'}(0,1)} Z_n(u) > \eta\right] = P\left[\exists (u_k)_{k\in\mathbb{N}} \in \prod_{k\in\mathbb{N}} T_k / Z_n(u_0) + \sum_{k=1}^{+\infty} Z_n(u_k - u_{k-1}) > \eta_0 + \sum_{k\geqslant 1}\eta_k\right]$$

$$\leqslant P_1 + P_2,$$

where

$$P_1 = \sum_{u_0\in T_0} P(Z_n(u_0) > \eta_0), \qquad P_2 = \sum_{k=1}^{\infty} \sum_{\substack{u_{k-1}\in T_{k-1}\\ u_k\in T_k}} P(Z_n(u_k - u_{k-1}) > \eta_k),$$

with $\eta_0 + \sum_{k\geqslant 1}\eta_k \leqslant \eta$. Using inequality (22), we straightforwardly infer that $P_1 \leqslant \exp(H_0 - nx_0)$ and $P_2 \leqslant \sum_{k\geqslant 1}\exp(H_{k-1} + H_k - nx_k)$ provided that we choose

$$\begin{cases} \eta_0 = 2\|f\|_\infty \delta_0(\sqrt{x_0} + \bar{r}_{(m,m')}x_0), \\ \eta_k = 2\|f\|_\infty \delta_{k-1}(\sqrt{\frac{5}{2}x_k} + \frac{3}{2}\bar{r}_{(m,m')}x_k). \end{cases}$$

We choose x_0 such that $nx_0 = H_0 + L_{m'}D_{m'} + \tau$ and, for $k \geqslant 1$,

$$nx_k = H_{k-1} + H_k + kD_{m'} + L_{m'}D_{m'} + \tau.$$

If $D_{m'} \geqslant 1$, we infer that

$$P\left(\sup_{t\in B_{m'}} Z_n(t) > \eta_0 + \sum_{k\geqslant 1}\eta_k\right) \leqslant e^{-L_{m'}D_{m'}-\tau}\left(1 + \sum_{k=1}^{\infty} e^{-kD_{m'}}\right) \leqslant 1.6\, e^{-L_{m'}D_{m'}-\tau}.$$

All that remains is to bound $\sum_{k\geqslant 0}\eta_k$. Since $\sum_{k=0}^{\infty}\delta_k = 2\delta_0$ and $\sum_{k=0}^{\infty} k\delta_k = 2\delta_0$, we find that

$$\delta_0 x_0 + \sum_{k=1}^{\infty}\delta_{k-1}x_k \leqslant 3\delta_0\frac{D(m')}{n}(a(\delta_0) + L_{m'}) + 3\delta_0\frac{\tau}{n},$$

where $a(\delta_0) = (5\ln(5/\delta_0) + 6\ln(2) + 4)/3$. With repeated use of the inequality $(x + y)^2 \leqslant (1 + \theta)x^2 + (1 + 1/\theta)y^2$ for positive x, y and θ, we obtain

$$\left(\sum_{k=0}^{\infty}\eta_k\right)^2 \leqslant 4\|f\|_\infty^2\left[\left(\delta_0\sqrt{x_0} + \sqrt{\frac{5}{2}}\sum_{k=1}^{\infty}\delta_{k-1}\sqrt{x_k}\right) + 1.5\bar{r}_{(m,m')}\left(\delta_0 x_0 + \sum_{k=1}^{\infty}\delta_{k-1}x_k\right)\right]^2$$

$$\leqslant 8\|f\|_{\infty}^2\left[\left(\delta_0+\frac{5}{2}\sum_{k=1}^{\infty}\delta_{k-1}\right)\left(\delta_0 x_0+\sum_{k=1}^{\infty}\delta_{k-1}x_k\right)+(1.5)^2\bar{r}_{(m,m')}^2\left(\delta_0 x_0+\sum_{k=1}^{\infty}\delta_{k-1}x_k\right)^2\right]$$

$$\leqslant 8\|f\|_{\infty}^2\delta_0^2\left[180\frac{D(m')}{n}(a(\delta_0)+L_{m'})+9(1.5)^2\left(\frac{101}{100}\right)\left(\frac{11}{10}\right)\bar{r}_{(m,m')}^2\left(\frac{D(m')}{n}(a(\delta_0)+L_{m'})\right)^2\right.$$

$$\left.+180\frac{\tau}{n}+9(1.5)^2100\bar{r}_{(m,m')}^2\frac{\tau^2}{n^2}\right]$$

$$\leqslant(8\times 0.385)\|f\|_{\infty}^2\left[\frac{D(m')}{n}(1+L_{m'})+4\bar{r}_{(m,m')}^2\left(\frac{D(m')}{n}\right)^2(1+L_{m'})^2+\frac{\tau}{n}+\bar{r}_{(m,m')}^2\frac{\tau^2}{n^2}\right]$$

$$\leqslant 3.1\|f\|_{\infty}^2\left((1+C_{\bar{r}}^2)\frac{D(m')(1+L_{m'})^2}{n}+\frac{\tau}{n}+\bar{r}_{(m,m')}^2\frac{\tau^2}{n^2}\right),$$

provided that we choose $\delta_0=1/100\leqslant 1/5$, $a(\delta_0)\sim 13.08$. The last inequality holds due to Assumption 4. This implies

$$\mathrm{P}\left[\sup_{u\in B_{m,m'}(0,1)}Z_n^2(u)>\kappa\|f\|_{\infty}^2\left(\frac{(1+C_{\bar{r}}^2)D(m')(1+L_{m'})^2}{n}+2\left(\frac{\tau}{n}\vee\bar{r}_{(m,m')}^2\frac{\tau^2}{n^2}\right)\right)\right]$$

$$\leqslant\mathrm{P}\left[\sup_{u\in B_{m,m'}(0,1)}Z_n^2(u)>\eta^2\right]\leqslant 2\mathrm{P}\left[\sup_{u\in B_{m,m'}(0,1)}Z_n(u)>\eta\right]\leqslant 3.2\mathrm{e}^{-L_{m'}D_{m'}-\tau}.$$

Therefore, for $\kappa=3.1$,

$$\mathrm{E}\left(G^2(m')-\kappa\|f\|_{\infty}^2(1+C_{\bar{r}}^2)\frac{D(m')(1+L_{m'})^2}{n}\right)_+$$

$$\leqslant\int_0^{\infty}\mathrm{P}\left(G^2(m')>\kappa(1+C_{\bar{r}}^2)\|f\|_{\infty}^2\frac{D(m')(1+L_{m'})^2}{n}+u\right)\mathrm{d}u$$

$$\leqslant\mathrm{e}^{-L_{m'}D_{m'}}\left(\int_{2\kappa\|f\|_{\infty}^2/\bar{r}_{(m,m')}^2}^{\infty}\mathrm{e}^{-nu/(2\kappa\|f\|_{\infty}^2)}\mathrm{d}u+\int_0^{2\kappa\|f\|_{\infty}^2/\bar{r}_{(m,m')}^2}\mathrm{e}^{-n\sqrt{u}/(2\sqrt{\kappa}\bar{r}_{(m,m')}\|f\|_{\infty})}\mathrm{d}u\right)$$

$$\leqslant\mathrm{e}^{-L_{m'}D_{m'}}\frac{2\kappa\|f\|_{\infty}^2}{n}\left(\int_0^{\infty}\mathrm{e}^{-v}\mathrm{d}v+\frac{2\bar{r}_{(m,m')}^2}{n}\int_0^{\infty}\mathrm{e}^{-\sqrt{v}}\mathrm{d}v\right)\leqslant\kappa'\,\mathrm{e}^{-L_{m'}D_{m'}}\frac{(1+4C_{\bar{r}}^2)\|f\|_{\infty}^2}{n}$$

using Assumption 4 again. This completes the proof. $\qquad\square$

6.2. Proof of Theorem 2

Recall that f_n^P is the L_2 orthogonal projection of f on the space $\mathscr{P}_n$ of piecewise polynomials of degree r on a dyadic partition with step 2^{-J_n}. The dimension of $\mathscr{P}_n$ is $N_n=(r+1)2^{J_n}$. Let $\hat{f}_n^P$ similarly be the orthogonal projection of I_n on $\mathscr{P}_n$ and

$$\Omega_{n,b} = \left\{ \left| \frac{\|\hat{f}_n^P\|_\infty}{\|f\|_\infty} - 1 \right| < b \right\}, \qquad 0 < b < 1.$$

On $\Omega_{n,b}$, the following inequalities hold:

$$\|f\|_\infty \le \frac{1}{1-b} \|\hat{f}_n^P\|_\infty, \qquad \|\hat{f}_n^P\|_\infty \le (1+b)\|f\|_\infty.$$

We start from (28), with $p_1(m)$ as defined in Proposition 4. On $\Omega_{n,b}$ we have

$$p_1(m') \le \frac{\tilde{\kappa}}{(1-b)^2} \|\hat{f}_n^P\|_\infty^2 (1 + C_{\tilde{r}}^2) \frac{D_{m'}(1+L_{m'})^2}{n}.$$

Then the choice

$$\text{pen}(m) = \frac{4\tilde{\kappa}}{(1-b)^2} \|\hat{f}_n^P\|_\infty^2 (1 + C_{\tilde{r}}^2) \frac{D_m(1+L_m)^2}{n}$$

ensures that, on $\Omega_{n,b}$, $8p_1(\hat{m}) - 2\text{pen}(\hat{m}) \le 0$ and

$$8p_1(m) + 2\text{pen}(m) \le 8\left[1 + \frac{(1+b)^2}{(1-b)^2} \right] \tilde{\kappa}\|f\|_\infty^2 (1 + C_{\tilde{r}}^2) \frac{D_m(1+L_m)^2}{n}.$$

Therefore, (28) becomes, on $\Omega_{n,b}$,

$$\|f - \tilde{f}\|^2 \le 3\|f - f_m\|^2 + 8\frac{M_2}{n} + 8G^2(\hat{m}) + C(b)\|f\|_\infty^2 (1 + C_{\tilde{r}}^2) \frac{D_m(1+L_m)^2}{n}.$$

Taking the expectation on $\Omega_{n,b}$ gives

$$\mathrm{E}(\|f - \tilde{f}\|^2 \mathbb{1}_{\Omega_{n,b}})$$

$$\le 3\|f - f_m\|^2 + 8\frac{M_2}{n} + C\frac{(1 + C_{\tilde{r}}^2)\|f\|_\infty^2 \Sigma}{n}$$

$$+ C(b)\|f\|_\infty^2 (1 + C_{\tilde{r}}^2) \frac{D_m(1+L_m)^2}{n}.$$

Next we must study $\mathrm{E}(\|f - \tilde{f}\|^2 \mathbb{1}_{\Omega_{n,b}^c})$. We note that

$$\|f - \tilde{f}\|^2 = \|f - f_{\hat{m}}\|^2 + \|f_{\hat{m}} - \hat{f}_{\hat{m}}\|^2 \le 2\|f\|^2 + \|f - I_n\|^2$$

$$\le 2\|f\|_\infty^2 + 2\|f - \mathrm{E}I_n\|^2 + 2\|I_n - \mathrm{E}I_n\|^2$$

using Pythagoras and the fact that $f_{\hat{m}} - \hat{f}_{\hat{m}}$ is the L_2-orthogonal projection of $f - I_n$ on $S_{\hat{m}}$. This implies that

$$\mathrm{E}(\|f - \tilde{f}\|^2 \mathbb{1}_{\Omega_{n,b}^c}) \le \left(2\|f\|_\infty^2 + \frac{M_2}{n} \right) \mathrm{P}(\Omega_{n,b}^c) + 2(\mathrm{E}\|I_n - \mathrm{E}I_n\|^4 \mathrm{P}(\Omega_{n,b}^c))^{1/2}.$$

For $\mathrm{E}\|I_n - \mathrm{E}I_n\|^4$, we have the inequality

$$\mathrm{E}\|I_n - \mathrm{E}(I_n)\|^4 = \mathrm{E}\left[\int_{-\pi}^{\pi}(I_n(\lambda) - \mathrm{E}I_n(\lambda)^2\,\mathrm{d}\lambda\right]^2 = 4\pi^2\mathrm{E}\left(\sum_{|k|<n}\left(c_{k,n} - \left(1 - \frac{|k|}{n}\right)\gamma_k\right)^2\right)^2$$

$$\leq \left(2\pi\sum_{|k|<n}\sqrt{\mathrm{E}\left(c_{k,n} - \left(1 - \frac{|k|}{n}\right)\gamma_k\right)^4}\right)^2,$$

where $c_{k,n}$ is the empirical covariance defined by (32) below. With rough bounds it can be proved that $\mathrm{E}(c_{k,n} - (1 - |k|/n)\gamma_k)^4$ is uniformly bounded; the computations are somewhat tedious. Therefore $\mathrm{E}\|I_n - \mathrm{E}(I_n)\|^4$ is of order less than n^2, and we need to prove that $\mathrm{P}(\Omega_{n,b}^c)$ is of order $1/n^4$ for some choice of N_n.

We note that $|\|\hat{f}_n^P\|_\infty - \|f\|_\infty| \leq \|\hat{f}_n^P - f_n^P\|_\infty + \|f_n^P - f\|_\infty$. Since $\|f_n^P - f\|_\infty \leq \frac{1}{4}\|f\|_\infty$, if we take $b \geq \frac{3}{4}$, this implies $\|f_n^P - f\|_\infty \leq (b/3)\|f\|_\infty$. We follow the line of Section 5.3 in Birgé and Massart (1997) and use their relation (2.8). Denoting by I_j the intervals $[\pi(j-1)/2^{J_n}, \pi j/2^{J_n})$ of the partition involved in $\mathscr{P}_n$, we have

$$\|\hat{f}_n^P - f_n^P\|_\infty \leq \sup_{1 \leq j \leq 2^{J_n}}\|(\hat{f}_n^P - f_n^P)\mathbb{1}_{I_j}\|_\infty \leq \sup_{1 \leq j \leq 2^{J_n}}\frac{\sqrt{r+1}2^{J_n/2}}{\sqrt{\pi}}\|(\hat{f}_n^P - f_n^P)\mathbb{1}_{I_j}\|$$

$$\leq \frac{(r+1)2^{J_n/2}}{\sqrt{\pi}}\sup_{\lambda\in\Lambda_n}|\langle I_n - f,\,\varphi_\lambda\rangle|$$

$$\leq (r+1)2^{J_n/2}\pi^{-1/2}\left(\sup_{\lambda\in\Lambda_n}|\langle I_n - \mathrm{E}(I_n),\,\varphi_\lambda\rangle| + \sup_{\lambda\in\Lambda_n}|\langle\mathrm{E}(I_n) - f,\,\varphi_\lambda\rangle|\right)$$

$$\leq ((r+1)\sqrt{N_n/\pi})\left(\sup_{\lambda\in\Lambda_n}|Z_n(\varphi_\lambda)| + \sqrt{\frac{M_2}{n}}\right),$$

where $(\varphi_\lambda)_{\lambda\in\Lambda_n}$ is the orthonormal basis of $\mathscr{P}_n$, and using Proposition 1 and the definition (21) of Z_n. This implies

$$\mathrm{P}(\Omega_{n,b}^c) = \mathrm{P}(|\|\hat{f}_n^P\|_\infty - \|f\|_\infty| > b\|f\|_\infty)$$

$$\leq \mathrm{P}\left((r+1)\sqrt{\frac{N_n}{\pi}}\left(\sup_{\lambda\in\Lambda_n}|Z_n(\varphi_\lambda)| + \sqrt{\frac{M_2}{n}}\right) + \frac{b}{3}\|f\|_\infty > b\|f\|_\infty\right)$$

$$\leq \mathrm{P}\left(\sup_{\lambda\in\Lambda_n}|Z_n(\varphi_\lambda)| > \frac{2b/\sqrt{\pi}}{3(r+1)\sqrt{N_n}}\|f\|_\infty - \sqrt{\frac{M_2}{n}}\right)$$

$$\leq 2\sum_{\lambda\in\Lambda_n}\mathrm{P}\left(Z_n(\varphi_\lambda) > \frac{1}{\sqrt{N_n}}\left(\frac{2\sqrt{\pi}b\|f\|_\infty}{3(r+1)} - \sqrt{\frac{M_2N_n}{n}}\right)\right).$$

Let

$$\xi = \frac{1}{\sqrt{N_n}}\left(\frac{2\sqrt{\pi}b\|f\|_\infty}{3(r+1)} - \sqrt{\frac{M_2 N_n}{n}}\right).$$

If we set the constraint $(M_2 N_n/n)^{1/2} \leqslant (b\pi^{1/2}\|f\|_\infty)/(6(r+1))$, i.e. the constraint on N_n given by

$$N_n \leqslant \frac{n}{M_2}\left(\frac{b\sqrt{\pi}\|f\|_\infty}{6(r+1)}\right)^2, \tag{29}$$

then $\xi \geqslant b\pi^{1/2}\|f\|_\infty/(2(r+1)N_n^{1/2})$. It follows from (22) that, since $\|\varphi_\lambda\| = 1$ and $\|\varphi_\lambda\|_\infty \leqslant \{(r+1)N_n/\pi\}^{1/2}$,

$$P(Z_n(\varphi_\lambda) > 2\|f\|_\infty(\sqrt{x} + \sqrt{(r+1)N_n/\pi}x)) \leqslant e^{-nx}.$$

This entails that

$$P(Z_n(\varphi_\lambda) > \xi) \leqslant \exp\left(-n\left(\frac{\xi^2}{2v^2} \wedge \frac{\xi}{2c}\right)\right),$$

where $v = 2^{1/2}\|f\|_\infty$ and $c = 2\|f\|_\infty\{(r+1)N_n/\pi\}^{1/2}$. Since

$$\frac{n\xi^2}{2v^2} = \frac{n\xi^2}{4\|f\|_\infty^2} \geqslant \frac{\pi b^2}{16(r+1)^2}\frac{n}{N_n} \quad\text{and}\quad \frac{n\xi}{2c} = \frac{n\xi\sqrt{\pi}}{4\|f\|_\infty\sqrt{(r+1)N_n/\pi}} \geqslant \frac{\pi b}{8(r+1)^{3/2}}\frac{n}{N_n},$$

we find that

$$P(Z_n(\varphi_\lambda) \geqslant \xi) \leqslant \exp\left(-\frac{\pi b^2}{16(r+1)^2}\frac{n}{N_n}\right),$$

and consequently

$$P(\Omega_{n,b}^c) \leqslant N_n \exp\left(-\frac{\pi b^2}{16(r+1)^2}\frac{n}{N_n}\right).$$

Then choosing

$$b = \frac{4}{6}\sqrt{\frac{5}{\pi}} \sim 0.841 \geqslant \frac{3}{4} \quad\text{and}\quad N_n \leqslant \frac{1}{36(r+1)^2}\frac{n}{\ln(n)}$$

implies that

$$P(\Omega_{n,b}^c) \leqslant \frac{1}{36(r+1)^2}\frac{1}{n^4\ln(n)}$$

provided that $n \geqslant \exp(\pi b^2 M_2/M^2)) = \exp(9M_2/20M^2)$. For small values of n, the result is trivially fulfilled as soon as the universal constant is chosen large enough.

6.3. Proof of Theorem 5

We use the same notation as in the proof of Theorem 1. We start from equation (20). The first problem is to find the maximum value of the term $2\langle f - \mathrm{E}(I_n), f_m - \tilde{f}\rangle$ because the result of Proposition 1 no longer holds. Consequently a more precise study must be carried out. We can prove the following lemma (see Appendix):

Lemma 6.1. *For any $m \in \mathscr{M}_n$ and $s \in S_m$ where S_m is a space of family* (T), $\|s\| = 1$,

$$|\langle f - \mathrm{E}(I_n), s\rangle| \leq 2\sqrt{2}\sqrt{\frac{1}{n^2}\sum_{j=0}^{D_m-1} j^2\gamma_j^2 + 36\frac{D_m}{n^2}\left(\sum_{j=0}^{n}|\gamma_j|\right)^2}.$$

This implies that

$$
\begin{aligned}
2|\langle f - \mathrm{E}(I_n), f_m - \tilde{f}\rangle| &\leq \frac{1}{8}\|f_m - \tilde{f}\|^2 + 8\langle f - \mathrm{E}(I_n), \frac{f_m - \tilde{f}}{\|f_m - \tilde{f}\|}\rangle^2 \\
&\leq \frac{1}{8}\|f_m - \tilde{f}\|^2 + 64\left[\frac{1}{n^2}\sum_{j=0}^{D(\hat{m})-1} j^2\gamma_j^2 + 36\frac{D(\hat{m})}{n^2}\left(\sum_{j=0}^{n}|\gamma_j|\right)^2\right] \\
&\leq \frac{1}{8}\|f_m - \tilde{f}\|^2 + 64\left[C_\gamma^2\left(\frac{D_m^{4\alpha+1} + D_{\hat{m}}^{4\alpha+1}}{n^2}\right) + 36C_\gamma^2\left(\frac{D_m}{n} + \frac{D_{\hat{m}}}{n}\right)\right] \\
&\leq \frac{1}{8}\|f_m - \tilde{f}\|^2 + \kappa C_\gamma^2\left(\frac{D_m}{n} + \frac{D_{\hat{m}}}{n}\right), \tag{30}
\end{aligned}
$$

using the inequalities $\gamma_j \leq C_\gamma j^{2\alpha-1}$ and $4\alpha - 1 \leq 0$, and where κ is a numerical constant.

For the bound of $\mathrm{E}(G^2(m'))$, with G given by (25), we do not use Proposition 4 and proceed more simply. We know that

$$G^2(m') = \left[\sup_{\sum a_\lambda^2 \leq 1}\sum_{\lambda \in \Lambda(m')} a_\lambda Z_n(\varphi_\lambda)\right]^2 = \sum_{\lambda \in \Lambda(m')} Z_n^2(\varphi_\lambda).$$

Under assumption (15) on f, we straightforwardly derive from Dahlhaus (1989, p. 1781) that $\rho(T_n(f)) \leq Kn^{2\alpha}$, and this holds for $\alpha \in (0, \frac{1}{2})$. Here, we use inequality (23) instead of (22), since for $0 < \alpha < \frac{1}{4}$, f is still square-integrable:

$$\mathrm{P}(Z_n(\varphi_\lambda) \geq 2\|\varphi_\lambda\|_\infty(\|f\|\sqrt{u} + Kn^{2\alpha}u)) \leq \mathrm{e}^{-nu}. \tag{31}$$

Here $\varphi_\lambda(x) = B\cos(\lambda x)$ and $B = \|\varphi_\lambda\|_\infty = \pi^{-1/2}$. This leads to the following bound: for every $\gamma \geq 0$,

$$P(G^2(m') > \gamma^2) = P\left(\sum_{\lambda \in \Lambda(m')} Z_n^2(\varphi_\lambda) > \gamma^2\right) \leqslant \sum_{\lambda \in \Lambda(m')} P(Z_n^2(\varphi_\lambda) > \gamma_\lambda^2)$$

$$= \sum_{\lambda \in \Lambda(m')} (P(Z_n(\varphi_\lambda) > \gamma_\lambda) + P(Z_n(-\varphi_\lambda) > \gamma_\lambda)) \leqslant 2 \sum_{\lambda \in \Lambda(m')} e^{-nx_\lambda},$$

with $\gamma^2 = \sum_{\lambda \in \Lambda(m')} \gamma_\lambda^2$ and $\gamma_\lambda = 2B(\|f\| x_\lambda^{1/2} + Kn^{2\alpha} x_\lambda)$. Choosing

$$x_\lambda = \frac{5\ln(D(m'))}{n} + \frac{\eta}{n}, \qquad \eta > 0,$$

implies that

$$P(G^2(m') > \gamma^2) \leqslant \frac{2e^{-\eta}}{D^4(m')}.$$

Let C be a numerical constant that may change from line to line. We find

$$\gamma^2 = \sum_\lambda \gamma_\lambda^2 \leqslant 8B^2 \sum_\lambda [\|f\|^2 x_\lambda + K^2 n^{4\alpha} x_\lambda^2]$$

$$\leqslant C\left[\|f\|^2 \frac{D(m)\ln(D(m'))}{n} + K^2 \frac{D(m')\ln^2(D(m'))}{n^{2-4\alpha}} + \eta\|f\|^2 \frac{D(m')}{n} + \eta^2 K^2 \frac{D(m')^2}{n^{2-4\alpha}}\right]$$

$$\leqslant C\left[(K^2 + \|f\|^2)\frac{D(m')\ln^2(D(m'))}{n}\right] + K_1 D(m')\left[\frac{\eta}{n} \vee \frac{D(m')\eta^2}{n}\right]$$

using the fact that $1 - 4\alpha > 0$ and setting $K_1 = 2C \max(\|f\|^2, K^2)$.

This gives the following choice for $p(m')$:

$$p(m') = C(K^2 + \|f\|^2)\frac{D(m')\ln^2(D(m'))}{n},$$

and leads to

$$P\left(G^2(m') > p(m') + 2K_1 D(m')\left[\frac{\eta}{n} \vee \frac{D(m')\eta^2}{n}\right]\right) \leqslant 2e^{-\eta} D(m')^{-4}.$$

Now we set

$$\eta = \frac{n\tau}{2K_1 D(m')} \qquad \text{if } \tau < 2K_1/n,$$

and

$$\eta = \sqrt{\frac{n\tau}{2K_1 D^2(m')}} \qquad \text{otherwise,}$$

whence

$$\int_0^\infty P(G^2(m') > p(m') + \tau)d\tau$$

$$= \int_0^{2K_1/n} P(G^2(m') > p(m') + \tau)d\tau + \int_{2K_1/n}^\infty P(G^2(m') > p(m') + \tau)d\tau$$

$$\leqslant 2D(m')^{-4}\left\{\int_0^{2K_1/n} \exp\left\{-\frac{n\tau}{2K_1}\right\} d\tau + \int_{2K_1/n}^\infty \exp\left\{-\sqrt{\frac{n\tau}{2K_1 D^2(m')}}\right\}\right\}$$

$$\leqslant 4K_1\frac{D(m')^{-3}}{n} + 8K_1\frac{D(m')^{-2}}{n} \leqslant 12K_1\frac{D_{m'}^{-2}}{n},$$

since $D(m') \geqslant D_{m'}$ and $\int_0^\infty e^{-u} du = 1$, $\int_0^\infty e^{-\sqrt{u}} du = 2$. This gives the bound

$$E(G^2(m') - p(m'))_+ \leqslant \frac{12K_1}{n} D_{m'}^{-2}.$$

Since in families (T), $D(m') = D_m \vee D_{m'}$, we have $p(\hat{m}) \leqslant p_1(m) + p_1(\hat{m})$, where $p_1(m) = C(K^2 + \|f\|^2)(D_m \ln^2(D_m))/n$. The end of the proof is the same as the end of the proof of Theorem 1 using

$$\sum_{m' \in \mathcal{M}_n} D_{m'}^{-2} = \sum_{k=1}^n k^{-2} \leqslant \frac{\pi^2}{6} = \Sigma.$$

instead of Assumption 5. Gathering these bounds with (30) gives the result claimed in Theorem 5, provided that pen is chosen such that $\text{pen}(m) \geqslant p_1(m) + \kappa C_\gamma^2 D_m/n$, $\forall m$.

Comments on the proof of Remark 4.3. Under assumption (15) on f, we use the inequality $\rho(T_n(f)) \leqslant Kn^{2\alpha}$ in the proof of Proposition 4 and Theorem 1. We apply inequality (22) again with $\|f\|_\infty$ replaced by $\rho(T_n(f))$. The order of $E\|f - I_n\|^2$ is no longer given by Proposition 1 but is computed from equation (37) in the Appendix, and is of order $n^{4\alpha-1}$.

Appendix

Proof of (5). Trivially, $\|f - \hat{f}_m\|^2 = \|f - f_m\|^2 + \|f_m - \hat{f}_m\|^2$, and if $(\varphi_\lambda)_{\{\lambda \in \Lambda_m\}}$ is an L_2-orthonormal basis of the linear space S_m with card $\Lambda_m = D_m$, then $f_m = \sum_{\lambda \in \Lambda_m}\langle f, \varphi_\lambda\rangle\varphi_\lambda$ and $\hat{f}_m = \sum_{\lambda \in \Lambda_m}\langle I_n, \varphi_\lambda\rangle\varphi_\lambda$. Therefore,

$$\|f_m - \hat{f}_m\|^2 = \sum_{\lambda \in \Lambda_m}\langle f - I_n, \varphi_\lambda\rangle^2 \leqslant 2\sum_{\lambda \in \Lambda_m}(\langle f - E(I_n), \varphi_\lambda\rangle^2 + \langle E(I_n) - I_n, \varphi_\lambda\rangle^2)$$

$$\leqslant 2D_m\|f - I_n\|^2 + 2\sum_{\lambda \in \Lambda_m}\langle E(I_n) - I_n, \varphi_\lambda\rangle^2.$$

First, Proposition 1 implies

$$D_m \mathrm{E}(\|f - I_n\|^2) \leqslant M_2 \frac{D_m}{n}$$

under Assumption 2. Second, Theorem 10.3.2 (ii) in Brockwell and Davis (1991) implies that under standard assumptions on linear processes, $\mathrm{cov}(I_n(\omega), I_n(\omega')) = O(1/n)$, uniformly in ω, ω' for $\omega \neq \omega'$. Thus

$$\mathrm{E} \sum_{\lambda \in \Lambda_m} \langle \mathrm{E}(I_n) - I_n, \varphi_\lambda \rangle^2 = \sum_{\lambda \in \Lambda_m} \int \varphi_\lambda(\omega) \varphi_\lambda(\omega') \mathrm{cov}(I_n(\omega), I_n(\omega')) \mathrm{d}\omega \, \mathrm{d}\omega'$$

$$\leqslant \frac{c}{n} \sum_{\lambda \in \Lambda_m} \left(\int \varphi_\lambda(\omega) \mathrm{d}\omega \right)^2 \leqslant c \frac{D_m}{n}.$$

Gathering all terms gives the result.   □

Proof of Proposition 1. First we rewrite $I_n^0(\omega)$ and $I_n(\omega)$ as follows:

$$I_n^0(\omega) = \frac{1}{2\pi} \left(c_{0,n}^0 + 2 \sum_{r=1}^{n-1} c_{r,n}^0 \cos r\omega \right), \qquad I_n(\omega) = \frac{1}{2\pi} \left(c_{0,n} + 2 \sum_{r=1}^{n-1} c_{r,n} \cos r\omega \right),$$

where, for $0 \leqslant r \leqslant n - 1$,

$$c_{r,n}^0 = \frac{1}{n} \sum_{k=1}^{n-r} (X_k - \mu)(X_{k+r} - \mu), \qquad c_{r,n} = \frac{1}{n} \sum_{k=1}^{n-r} (X_k - \overline{X}_n)(X_{k+r} - \overline{X}_n). \qquad (32)$$

Then we have

$$(2\pi)^2 (f(\omega) - \mathrm{E}I_n(\omega))^2 = (2\pi)^2 (f(\omega) - \mathrm{E}I_n^0(\omega) + \mathrm{E}I_n^0(\omega) - \mathrm{E}I_n(\omega))^2$$

$$= (2\pi)^2 \left(\left\{ 2 \sum_{k \geqslant n} \gamma_k \cos k\omega - 2 \sum_{k=1}^{n-1} \frac{k}{n} \gamma_k \cos k\omega \right\} \right.$$

$$\left. + \left\{ \mathrm{E}c_{0,n}^0 - \mathrm{E}c_{0,n} + 2 \sum_{r=1}^{n-1} \cos r\omega (\mathrm{E}c_{r,n}^0 - \mathrm{E}c_{r,n}) \right\} \right)^2$$

$$:= (2\pi)^2 (A_n^2(\omega) + 2A_n(\omega)B_n(\omega) + B_n^2(\omega)).$$

We obtain

$$\int_{-\pi}^{\pi} A_n^2(\omega) \mathrm{d}\omega = 4\pi \left(\sum_{k \geqslant n} \gamma_k^2 + \frac{1}{n^2} \sum_{k=1}^{n-1} k^2 \gamma_k^2 \right) \leqslant 4\pi \left(\frac{1}{n} \sum_{k \geqslant n} k \gamma_k^2 + \frac{n}{n^2} \sum_{k=1}^{n-1} k \gamma_k^2 \right) = \frac{2\pi M_1}{n}. \quad (33)$$

Next

$$\int_{-\pi}^{\pi} B_n^2(\omega) \mathrm{d}\omega = 2\pi (\mathrm{E}c_{0,n}^0 - \mathrm{E}c_{0,n})^2 + 4\pi \sum_{r=1}^{n-1} (\mathrm{E}c_{r,n}^0 - \mathrm{E}c_{r,n})^2.$$

It is easily seen from Theorem 8.3.2 in Anderson (1971) that if $\sum_{k=-\infty}^{\infty}|\gamma_k| < \infty$, then

$$|n(\mathrm{E}c_{r,n}^0 - \mathrm{E}c_{r,n})| \leqslant 3 \sum_{|k|<n} |\gamma_k| \leqslant 6 \sum_{k=0}^{\infty} |\gamma_k| \tag{34}$$

for all n and r. Therefore, we derive that

$$\int_{-\pi}^{\pi} B_n^2(\omega)\mathrm{d}\omega \leqslant 18\pi \frac{M^2}{n}\left(2 + \frac{1}{n}\right). \tag{35}$$

Finally, the same arguments yield

$$2\left|\int_{-\pi}^{\pi} A_n(\omega)B_n(\omega)\mathrm{d}\omega\right| \leqslant 24\pi M \sum_{k=1}^{n-1} \frac{k}{n^2}|\gamma_k| \leqslant 24\pi \frac{M^2}{n}. \tag{36}$$

Then combining (33)–(36), we derive that

$$\int_{-\pi}^{\pi} (f(\omega) - \mathrm{E}I_n(\omega))^2 \, \mathrm{d}\omega \leqslant \frac{M_1}{2\pi n} + 15\frac{M^2}{\pi n} + 9\frac{M^2}{2\pi n^2},$$

which concludes the proof. □

Note that, using (33) and (34), we obtain more generally the bound

$$\int_{-\pi}^{\pi} (f(\omega) - \mathrm{E}(I_n(\omega)))^2 \, \mathrm{d}\omega \leqslant 8\pi \left[\sum_{k\geqslant n} \gamma_k^2 + \frac{1}{n^2}\sum_{k=1}^{n-1} k^2\gamma_k^2 + \frac{6}{n}\left(\sum_{k=1}^{n} |\gamma_k|\right)^2\right] \tag{37}$$

which allows us to compute an order even if $\sum|\gamma_k|$ is not finite. In particular, if $|\gamma_k| \sim k^{2\alpha-1}$, then all terms are of order $n^{4\alpha-1}$.

Proof of Lemma 6.1. Let $s \in S_m$ with $\|s\| = 1$. Then $s(x) = B\sum_{j=0}^{D_m-1} a_j \cos(jx)$ and

$$|\langle f - \mathrm{E}(I_n), s\rangle| = B\left|\sum_{j=0}^{D_m-1} a_j\langle f - \mathrm{E}(I_n), \cos(j\cdot)\rangle\right|$$

$$\leqslant B\left(\sum_{j=0}^{D_m-1} a_j^2\right)^{1/2}\left(\sum_{j=0}^{D_m-1} \langle f - \mathrm{E}(I_n), \cos(j\cdot)\rangle^2\right)^{1/2}$$

$$\leqslant B\left(\sum_{j=0}^{D_m-1} \langle f - \mathrm{E}(I_n), \cos(j\cdot)\rangle^2\right)^{1/2}$$

With the same notation as previously,

$$B^2 \langle f - \mathrm{E}(I_n), \cos(j\cdot) \rangle = \int_{-\pi}^{\pi} (f(\omega) - \mathrm{E}I_n(\omega))B^2 \cos(j\omega)\mathrm{d}\omega$$

$$= \int_{-\pi}^{\pi} (A_n(\omega) + B_n(\omega))B^2 \cos(j\omega)\mathrm{d}\omega$$

$$:= I_1(j) + I_2(j).$$

Here,

$$I_1(j) = \int_{-\pi}^{\pi} \left(2 \sum_{k \geqslant n} \gamma_k \cos(k\omega) - 2 \sum_{k=1}^{n} \frac{k}{n} \gamma_k \cos(k\omega) \right) B^2 \cos(j\omega)\mathrm{d}\omega$$

$$= -\frac{2}{n} \sum_{k=1}^{n} k\gamma_k \int_{-\pi}^{\pi} B^2 \cos(k\omega)\cos(j\omega)\mathrm{d}\omega = -\frac{2j\gamma_j}{n},$$

all terms $\int_{-\pi}^{\pi} B^2 \cos(k\omega)\cos(j\omega)\mathrm{d}\omega$ being zero in the first sum because $j < D_m \leqslant n \leqslant k$ so that $j = k$ never holds. For $I_2(j)$, we use (34):

$$I_2(j) = \int_{-\pi}^{\pi} (\mathrm{E}c_{0,n}^0 - \mathrm{E}c_{0,n})B^2 \cos(j\omega)\mathrm{d}\omega + 2 \sum_{r=1}^{n-1} (\mathrm{E}c_{r,n}^0 - \mathrm{E}c_{r,n}) \int_{-\pi}^{\pi} B^2 \cos(r\omega)\cos(j\omega)\mathrm{d}\omega$$

$$= (\mathrm{E}c_{0,n}^0 - \mathrm{E}c_{0,n})\delta_{j,0} + 2(\mathrm{E}c_{r,n}^0 - \mathrm{E}c_{r,n})\delta_{r,j}\mathbb{1}_{r \geqslant 1} \leqslant \frac{6}{n} \sum_{k=0}^{n} |\gamma_k|(\delta_{j,0} + 2\delta_{r,j}\mathbb{1}_{r \geqslant 1}).$$

Gathering the terms gives:

$$|\langle f - \mathrm{E}(I_n), s \rangle| \leqslant \left(\sum_{j=0}^{D_m-1} (I_1(j) + I_2(j))^2 \right)^{1/2} \leqslant \left(\sum_{j=0}^{D_m-1} 2(I_1^2(j) + I_2^2(j)) \right)^{1/2}$$

$$\leqslant \sqrt{2} \left(\sum_{j=0}^{D_m-1} \frac{4j^2\gamma_j^2}{n^2} + \frac{144D_m}{n^2} \left(\sum_{k=0}^{n} |\gamma_k| \right)^2 \right)^{1/2},$$

which ends the proof of Lemma 6.1.   □

Acknowledgements

References

Adenstedt, R.K. (1974) On large sample estimation for the mean of a stationary sequence. *Ann. Statist.*, **2**, 1095–1107.

Anderson, T.W. (1971) *The Statistical Analysis of Time Series*. New York: Wiley.

Baraud, Y., Comte, F. and Viennet, G. (1999) Model selection for (auto-)regression with dependent data. Working paper no. 9929, Centre de Recherche en Économie et Statistique.

Barron, A., Birgé, L. and Massart, P. (1999) Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, **113**, 301–413.

Bentkus, R. (1985) Rate of uniform convergence of statistical estimators of spectral density in spaces of differentiable functions. *Lithuanian Math. J.*, **25**, 209–219.

Bentkus, R. and Rudzkis, R. (1976) Large deviations for the spectral estimate of a stationary Gaussian sequence. *Lithuanian Math. J.*, **16**, 519–529.

Birgé, L. and Massart, P. (1997) From model selection to adaptive estimation. In D. Pollard, E. Torgersen and G. L. Yang (eds), *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, pp. 55–87. New York: Springer-Verlag.

Birgé, L. and Massart, P. (1998) Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, **4**, 329–375.

Birgé, L. and Massart, P. (2000) An adaptive compression algorithm in Besov spaces. *Constr. Approx.*, **16**, 1–36.

Birgé, L. and Rozenholc, Y. (2000) How many bins should be put in a regular histogram? Unpublished manuscript.

Birman, M.S. and Solomjak, M.Z. (1967) Piecewise polynomials approximation of functions of the classes W_p. *Mat. Sb.*, **73**, 295–317.

Brockwell, J.B. and Davis, R.A. (1991), *Time Series: Theory and Methods*, 2nd edition. Berlin: Springer-Verlag.

Comte, F. and Rozenholc, Y. (2000) Adaptive estimation of mean and volatility functions in (auto-)regressive models. In revision for Stochastic Process. Appl.

Dacunha-Castelle, D. and Duflo, M. (1993) *Probabilités et Statistiques, 2. Problèmes à Temps Mobile*, 2nd edition. Paris: Masson.

Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes. *Ann. Statist.*, **17**, 1749–1766.

Daubechies, I. (1992) *Ten Lectures on Wavelets*. Philadelphia: Society of Industrial and Applied Mathematics.

Davies, R.B. (1973) Asymptotic inference in stationary Gaussian time-series. *Adv. Appl. Probab.*, **5**, 469–497.

DeVore, R.A. and Lorentz, G.G. (1993) *Constructive Approximation*. Berlin: Springer-Verlag.

Donoho, D.L. and Johnstone, I.M. (1994) Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81**, 425–455.

Donoho, D.L. and Johnstone, I.M. (1998) Minimax estimation via wavelet shrinkage. *Ann. Statist.*, **26**, 879–921.

Efromovich, S. (1998) Data-driven estimation of the spectral density. *J. Amer. Statist. Assoc.*, **92**, 762–769.

Fox, R. and Taqqu, M.S. (1986) Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, **14**, 517–532.

Gobulev, G.K. (1993) Nonparametric estimation of smooth spectral densities of Gaussian stationary sequences, *Theory Probab. Appl.*, **38**, 630–639.

Hall, P. and Hart, J.D. (1990) Nonparametric regression with long-range dependence, *Stochastic Process. Appl.*, **36**, 339–351.

Hannan, E.J. (1970) *Multiple Time Series*. New York: Wiley.

Laurent, B. and Massart, P. (1998) Adaptive estimation of a quadratic functional by model selection. Working paper 98.81, Université Paris-Sud.

Neumann, M.H. (1996) Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series, *J. Time Ser. Anal.*, **17**, 601–633.

Pinkus, A. (1985). *n-Widths in Approximation Theory*. Berlin: Springer-Verlag.

Soulier, P. (1999). Estimation adaptative de la densité spectrale d'un processus faiblement ou fortement dépendant. Preprint, Université d'Evry.

Taniguchi, M. (1987) Minimum contrast estimation for spectral densities of stationary processes. *J. Roy. Statist. Soc. Ser. B*, **49**, 315–325.

Whittle, P. (1953) The analysis of multiple stationary time series. *J. Roy. Statist. Soc. Ser. B*, **15**, 125–139.