Electronic Journal of Statistics
Vol. 5 (2011) 102–126
ISSN: 1935-7524
DOI: 10.1214/11-EJS599

Semiparametric shift estimation based on the cumulated periodogram for non-regular functions

Ismaël Castillo*

LPMA, CNRS and Universités Paris 6 et 7
175, rue du Chevaleret, 75013 Paris, France
e-mail: ismael.castillo@upmc.fr

and

Eric Cator

Department of Mathematics, Delft University of Technology,
Mekelweg 4, 2628CD Delft, Nederland
e-mail: e.a.cator@tudelft.nl

Abstract: The problem of estimating the center of symmetry of a symmetric signal in Gaussian white noise is considered. The underlying nuisance function f is not assumed to be differentiable, which makes a new point of view to the problem necessary.

We investigate the well-known sieve maximum likelihood estimators based on the cumulated periodogram, and study minimax rates over classes of irregular functions. It is shown that if the class appropriately controls the growth to infinity of the Fisher information over the sieve, semiparametric fast rates of convergence are obtained. We prove a lower bound result which implies that these semiparametric rates are really slower than the parametric ones, contrary to the regular case. Our results also suggest that there may be room to improve on the popular cumulated periodogram estimator.

AMS 2000 subject classifications: Primary 62G05, 62G20.
Keywords and phrases: Semi-parametric models, non-regular case, Gaussian white noise models, estimation of the center of symmetry.

Received February 2010.

1. Introduction and main results

Let $L^2[0,1]$ be the Hilbert space of all real-valued square integrable functions on the interval $[0,1]$. The model we consider, which will be referred to as the *translation model* in the sequel, consists of observing a path Y, which is a solution of the diffusion equation

$$dY(t) = f(t-\theta)dt + \varepsilon dW(t)\ , \quad t \in [-1/2, 1/2], \quad (\varepsilon > 0) \tag{1}$$

*Research partly supported by a Postdoctoral fellowship from the Vrije Universiteit Amsterdam, the Netherlands.

where the unknown function f is *symmetric* (that is $f(-u) = f(u)$ for all u), 1-periodic and when restricted to $[0, 1]$, belongs to $L^2[0, 1]$. The unknown parameter θ, that is the center of symmetry of the *signal* $f(\cdot - \theta)$, is supposed to belong to a compact subset of the real numbers, and W is here standard Brownian motion on $[-1/2, 1/2]$. We shall work in the asymptotic framework $\varepsilon \to 0$. For basic properties of this model and related ones, such as estimation of the period of a periodic function, estimation of the amplitude, etc., see [6].

The problem of estimating features (e.g. location of symmetry, period, amplitude ...) of an unknown signal corrupted by noise is a central issue in signal processing and its applications (telecommunications, laser vibrometry, to name a few), see [7] and the references therein. Model (1) can be seen as an idealized version in continuous time of a finite sample model. One of the most important methods of estimation involves the maximization of the so-called *cumulated periodogram*. In the context of model (1), this is the same as maximizing in τ the quantity

$$\ell_K(\tau) = \sum_{k=1}^{K} \left| \sqrt{2} \int_{-1/2}^{1/2} \cos(2\pi k(t-\tau)) dY(t) \right|^2 .$$

One still needs to choose the integer K, which corresponds to the dimension of the finite-dimensional space (or *sieve*) approximating the functional space where f lives. Some issues related to this (delicate) choice are discussed in the sequel. In this paper we will study the cumulated periodogram estimator when the underlying function f is irregular, find uniform rates of convergence over specific "irregularity" classes, and show that there may be room to improve on the periodogram estimator.

1.1. Framework

In the sequel we shall expand functions in $L^2[0, 1]$ over the trigonometric basis. Since in model (1) f is even, we can expand it over the elements $\varepsilon_k(\cdot) = \sqrt{2}\cos(2\pi k\cdot)$, for $k \geq 0$. For f in $L^2[0, 1]$, we denote by $\|\cdot\|_2$ its L^2-norm and by $\{f_k\}_{k\geq 0}$ the sequence of its Fourier coefficients, so that

$$\|f\|_2^2 = \int_0^1 f(t)^2 dt, \quad f_k = \sqrt{2} \int_0^1 f(t) \cos(2\pi kt) dt \quad (k \geq 1).$$

The only further assumption that we make a priori on the function f is that there exist positive constants ρ and L with $L^2 > 2\rho^2$ such that f belongs to the set $\mathcal{F}$ defined as

$$\textbf{(F)} \qquad \mathcal{F} \triangleq \mathcal{F}(\rho, L) = \{f \in L^2[0, 1], \quad |f_1| \geq \rho, \quad \|f\|_2^2 \leq L^2\}.$$

We also extend the functions in $\mathcal{F}$ by 1-periodicity so that they can be seen as functions over the real line. Note that we do not assume a form of differentiability of f, as opposed to works studying the models at stake in the regular framework (for instance, in [3, 5, 2] and [4], existence of strictly more than one derivative in the L^2-sense is required). The assumption on f_1 is a way to ensure that the quantity $\sum_{k\geq 0} k^2 f_k^2$ (up to a factor ε^{-2} it equals the Fisher information), which might be infinite in our context, is far away from zero. This corresponds to the fact that one wants to exclude functions which could be arbitrarily close in $L^2[0,1]$ to a constant function, for which estimation of θ in (1) is impossible. It also ensures that the function f has 1 as the smallest period, which is important for identifiability reasons.

We further assume that the parameter θ belongs to a compact interval Θ^S

$$\Theta^S = [-\tau_0, \tau_0] \subset]-1/4, 1/4[.$$

One easily checks that to ensure identifiability, the parameter set should have a diameter smaller than 1/2, which explains the choice for Θ^S.

The aim of the statistical problem at hand is to understand how and how fast one can estimate the unknown θ, the function f being unknown as well. An estimator will be as usual a measurable function of the observations, which means here a measurable function with respect to the σ-field generated by the process Y, solution of (1), on the Banach space $\mathcal{B}^S$ of continuous functions on $[-1/2, 1/2]$. We denote by $\mathbf{P}^S_{\theta,f}$ the probability distribution generated by Y on this space.

The natural statistical object to work with is the *likelihood*, which, in the context of model (1), is defined as a likelihood ratio. Let $\mathbf{P}^S_0$ be the probability distribution generated by ε times Brownian motion on $\mathcal{B}^S$. Then the likelihood $L^S(\theta, f)$ is defined as a Radon-Nikodym derivative, for which an explicit expression is given by Girsanov's formula: for any $\mathbf{Y} \in \mathcal{B}^{\mathbf{S}}$,

$$L^S(\theta, f)(\mathbf{Y}) \triangleq \frac{d\mathbf{P}^S_{\theta,f}}{d\mathbf{P}^S_0}(\mathbf{Y}) = \exp\left(\varepsilon^{-2}\int_{-1/2}^{1/2} f(t-\theta)d\mathbf{Y}(t) - \frac{\varepsilon^{-2}}{2}\int_{-1/2}^{1/2} f(t-\theta)^2 dt\right).$$

We denote by $\mathbf{E}^S_{\theta,f}$ the expectation under the probability distribution $\mathbf{P}^S_{\theta,f}$. For simplicity we often drop the index S.

1.2. Irregular signals

In the sequel, an irregular function will be an element f of $L^2[0,1]$ such that the sum $\sum_{k\geq 1} k^2 f_k^2$ is infinite. Informally, the Fisher information becomes infinite in this case. We start by some basic examples of such functions.

1. *Step function.* Define the element of $L^2[-1/2, 1/2]$

$$g_1(u) = \mathbf{1}_{|u|<1/4}$$

and extend it by periodicity to $\mathbb{R}$. The Fourier coefficients $g_{1,k}$ of this function behave like k^{-1} as k increases.

2. *Infinite jump.* More generally, for any $0 < \eta < 1/2$,

$$g_\eta(u) = |u|^{-\eta} \mathbf{1}_{0<|u|<1/4}$$

extended by periodicity to $\mathbb{R}$ has Fourier coefficients $g_{\eta,k}$ that behave like $k^{-1+\eta}$.

It is important to note that no continuity or differentiability assumption is made on f. In particular, here we do not impose any constraint on quantities such as "the number of jumps of f" or the regularity of f between its jumps points. This is in contrast to the problems considered in [6] for irregular functions, and more generally, in so-called "change-points" problems; see for instance [8]. In fact, the semiparametric rates we obtain below will be different.

Classes of irregular functions based on sieve-Fisher information. Since the goal of efficient inference cannot be the "Fisher information" $\sum_{k\geq 1} k^2 f_k^2$, which is now infinite, a natural approach is to define classes of functions quantifying by how much partial sums of this quantity grow. Define, for an integer $K_0 > 0$, reals $M > 0$ and $\alpha \in]0,1[$, the class

$$S_\alpha(K_0, M) = \left\{ f \in \mathcal{F} : \quad \forall K \geq K_0, \quad \sum_{k=1}^{K} k^2 f_k^2 \geq M K^{2-2\alpha} \right\}. \tag{2}$$

For instance, the functions g_1 and g_η considered above belong respectively to $S_{1/2}(K, M)$ and $S_{1/2-\eta}(K, M)$, for appropriate choices of constants K, M.

Classes of irregular functions based on Kullback-Leibler divergence. We will also consider the following class, which defines the irregularity using the intrinsic distance on the statistical problem, that is here the L^2-distance between the signals. Let us define, for positive η, M_1 and $0 < \alpha < 1$,

$$\begin{aligned} \mathcal{F}_\alpha(\eta, M_1) \quad = \quad & \Big\{ f \in \mathcal{F}, \quad \forall \tau, \theta \in \Theta^S : |\tau - \theta| < \eta, \\ & \int_{-1/2}^{1/2} \{ f(t-\theta) - f(t-\tau) \}^2 \, dt \geq M_1 (\tau - \theta)^{2\alpha} \Big\}. \end{aligned}$$

For instance, we have $g_1 \in \mathcal{F}^S_{1/2}(\delta, M)$ and $g_\eta \in \mathcal{F}^S_{1/2-\eta}(\delta, M)$ for appropriate δ, M. In fact, it can be checked that for any $0 < \alpha < 1$, we have the inclusion $S_\alpha(K_0, M) \subset \mathcal{F}_\alpha(\delta, M_1)$, for small enough K_0 and large enough M. This can be seen by expanding the L^2-distance over the Fourier basis and using the bound $|\sin(u)| \geq 2|u|/\pi$ for $0 \leq |u| \leq \pi/2$.

1.3. Semiparametric rates for sieve maximum likelihood estimators

A natural way to obtain estimators in model (1) is to use the profile likelihood method (see [10], Chap. 25) which consists in maximizing the likelihood in two steps: first one maximizes it with respect to the nuisance parameter, obtaining a function -the *profile likelihood*- independent of f, and then one chooses the maximizer of this quantity as estimator. In fact, the first step cannot involve a maximum over the whole space $\mathcal{F}$, which is too large, thus one restricts the maximization to a sieve, which here will be the linear space generated by the first K elements of the trigonometric basis. The final estimator is obtained by choosing K and is called the *sieve maximum likelihood*, or *sieve-MLE*.

This method is particularly useful here since the profile likelihood can be written explicitly and in fact leads to a criterion function known as *cumulated periodogram*. The calculations leading to the criterion together with refinements with weights have been studied in detail in [3] and [2], which is why we directly give the expression of the obtained profile likelihood,

$$\ell_K(\tau) = \sum_{k=1}^{K} \left| \sqrt{2} \int_{-1/2}^{1/2} \cos(2\pi k(t-\tau)) dY(t) \right|^2, \tag{3}$$

where τ belongs to Θ^S, $Y(t)$ is assumed to satisfy (1), and K is an integer to be chosen. For any $K > 0$, we obtain a sieve-MLE $\hat{\theta}(K)$ by setting

$$\hat{\theta}(K) = \underset{\tau \in \Theta^S}{\mathrm{Argmax}}\ \ell_K(\tau). \tag{4}$$

The choice of K will be made precise below.

Considered class of nuisance functions. A first candidate for the class is certainly the whole class $\mathcal{F}$ defined above. However, this class contains (almost) all smooth functions, for which we know that estimation is in fact more difficult in the considered model, and the obtained rate turns out to be ε (the usual parametric $n^{-1/2}$); see Theorem 1.2 and Section 1.4 for a corresponding (and in fact stronger) lower bound result.

A smaller class allows us to obtain uniform fast rates for sieve-MLEs: the class $S_\alpha(K_0, M)$ defined by (2), which enables us to quantify the "irregularity" of a function through the parameter α using a control on the first K coefficients of f in the (Fourier) basis. Note that the definition of $S_\alpha(K_0, M)$ is quite natural in the sense that it *precisely* gives control over the Fisher information for estimating θ in the sequence (depending on K) of sieve models where (θ, f) is of the form $\{(\theta, f_1, \ldots, f_K)\}$. Indeed, one easily checks that the Fisher information for the latter sieve model used in (1) equals $\varepsilon^{-2} \sum_{k=1}^{K} k^2 f_k^2$. In that sense the considered class is perhaps the most natural one to consider in the non-regular case. We shall see that considering an essentially larger class leads again to the usual regular rates, see Theorem 1.4.

Let us now make the following choice for the cut-off parameter K

$$K_S^* = K_{S,\alpha}^* = \begin{cases} \lfloor \varepsilon^{-\frac{2}{1+2\alpha}} \rfloor & \text{if} \quad 1/4 \le \alpha \le 1, \\ \lfloor \varepsilon^{-\frac{4}{1+4\alpha}} \rfloor & \text{if} \quad 0 < \alpha < 1/4 \end{cases}. \tag{5}$$

The associated sieve-MLE is defined by $\hat{\theta}_S = \hat{\theta}(K_S^*)$. Let us define the rate

$$r_\varepsilon = \begin{cases} \varepsilon^{\frac{4}{1+4\alpha}} & \text{if} \quad 0 < \alpha < 1/4, \\ \varepsilon^{\frac{3}{1+2\alpha}} & \text{if} \quad 1/4 \le \alpha \le 1. \end{cases} \tag{6}$$

Our first result is the following upper bound:

Theorem 1.1. *Let $\hat{\theta}_S = \hat{\theta}(K_S^*)$ and let r_ε be defined by (6).Then with $l_\varepsilon = -\log\varepsilon$, for any $0 < \alpha < 1$, $K_0 > 0$ and $M > 0$,*

$$\lim_{\varepsilon\to 0} \sup_{\theta\in\Theta^S,\ f\in S_\alpha(K_0,M)} \mathbf{P}_{\theta,f}\left(\left|\hat{\theta}_S - \theta\right| > l_\varepsilon^{1/\alpha} r_\varepsilon\right) \quad = 0.$$

For $1/4 \le \alpha < 1$, we also have that

$$\lim_{\varepsilon\to 0} \sup_{\theta\in\Theta^S,\ f\in S_\alpha(K_0,M)} r_\varepsilon^{-2}\mathbf{E}_{\theta,f}\left(\hat{\theta}_S - \theta\right)^2 < +\infty.$$

The rate of convergence r_ε is a *fast rate* in that it is faster than the usual parametric rate $r_{0,\varepsilon}^S = \varepsilon$ (that is $1/\sqrt{n}$ if we take $\varepsilon = 1/\sqrt{n}$). Note that the smaller α (that is the more irregular the function in the sense of the classes $S_\alpha(K_0, M)$), the faster we can estimate θ, as we can see immediately from the expression (6). A close examination of the proof of Theorem 1.1 reveals that the rate of $\hat{\theta}(K)$ is quite sensitive to the choice of the cut-off K. Note also that there is a change of slope in the power of the rate (6) at $\alpha = 1/4$, which is reminiscent of the nonparametric effect observed in the problem of estimating the L^2-norm, where $\alpha = 1/4$ is also a transition point. Finally note that the choice of K depends on the parameter α, which is often unknown in practice. A natural follow-up of the previous result is then to build a fully adaptive estimate, see Section 1.6 for a further discussion.

When we consider the larger class $\mathcal{F}$, we can use our methods to extend the semiparametric results known in the regular case where f is differentiable and the Fisher information is finite, saying that semiparametric estimation is possible at rate ε, whatever the regularity of f in $\mathcal{F}$:

Theorem 1.2. *Let $\tilde{K} = \lfloor \varepsilon^{-1/2} \rfloor$ and let us define $\tilde{\theta}_S = \hat{\theta}(\tilde{K})$.Then*

$$\lim_{\varepsilon\to 0} \sup_{\theta\in\Theta^S,\ f\in\mathcal{F}} \varepsilon^{-2}\mathbf{E}_{\theta,f}(\tilde{\theta}_S - \theta)^2 \quad < +\infty.$$

In the next section, it will become clear that consider the class $\mathcal{F}_\alpha$ will give the same upper bound as Theorem 1.2.

Possible other choices of basis functions. One could also consider the cumulated periodogram with respect to other basis functions, for instance wavelets. But here we shall consider only the Fourier basis, which leads to the interpretation of the criterion as sum of estimated harmonics, and which also has good properties with respect to a shift of the parameter, which makes the mathematical analysis simpler.

1.4. Lower bounds

As we have seen in Section 1.3, the rate of sieve-MLE estimators is, uniformly over the class $S_\alpha(K_0, M)$, faster than the smooth rate. It is then natural to investigate the best possible rate of convergence of semiparametric estimators in the minimax sense over such classes. In particular, we would like to know whether it is possible or not to do as well as in the parametric case, which is the case in the regular framework, at least asymptotically. Let us define

$$\underline{r_\varepsilon} = \varepsilon^{\frac{5}{1+4\alpha}}. \tag{7}$$

Theorem 1.3. *For any $0 < \alpha < 1$, positive K_0, M, we can choose L, defining the class $\mathcal{F}$, large enough such that there exist C such that, using (7),*

$$\liminf_{\varepsilon\to 0} \inf_{\hat\theta} \sup_{\theta\in\Theta^S,\ f\in S_\alpha(K_0,M)} \mathbf{E}_{\theta,f}((\hat\theta-\theta)^2 \underline{r_\varepsilon}^{-2}) \geq C > 0,$$

where the infima are taken over all estimators of θ in the translation model.

We will see in the next section that the semiparametric rate found in this Theorem is slower than the rates obtained in the parametric case, which is the model where $f \in S_\alpha(K_0, M)$ would be known. This means that a significant *loss of speed* occurs due to the nuisance parameter f. Note also that this lower bound is actually strictly faster than the rates we found for the sieve-MLE. This suggest that maybe we can improve on the cumulative periodogram estimator; see also the discussion in Section 1.6.

The fact that in Theorem 1.3 we might need to choose L very big is related to the fact that the upper bound on the L^2-norms in $\mathcal{F}$ might conflict with the lower bound on the L^2-distance of the signals.

Another interesting question is the following: what happens if we seek for uniform results over classes larger than S_α, for instance the classes $\mathcal{F}_\alpha$ defined by (3)? The next lower bound result shows that the latter are, somehow, too large, in that one cannot do better than the rate in the smooth case.

Theorem 1.4. *For any $0 < \alpha < 1$ and positive η, M_1, we can choose L, defining the class $\mathcal{F}(\rho, L)$, big enough such that*

$$\liminf_{\varepsilon\to 0} \inf_{\hat\theta} \sup_{\theta\in\Theta^S,\ f\in\mathcal{F}_\alpha(\eta,M_1)} \mathbf{E}_{\theta,f}((\hat\theta-\theta)^2 \varepsilon^{-2}) > 0,$$

where the infima are taken over all estimators of θ in the translation model.

In the semiparametric context, it is thus not possible to obtain a better minimax rate of convergence over the classes $\mathcal{F}_\alpha$ than in the case where f is smooth.

1.5. Parametric rates

Let us consider model (1) in the parametric case, that is for known f. Suppose that f is continuously differentiable on $[0, 1]$, then it is not hard to check that the parametric models $\{P_{\theta,f},\ \theta \in \Theta^S\}$ form a *regular* statistical experiment as defined for instance in [6], page 65. In particular, the Fisher information $\mathcal{I}_\varepsilon(f)$ is finite and equals

$$\mathcal{I}_\varepsilon(f) = \varepsilon^{-2}\|f'\|_2^2 = \varepsilon^{-2}\sum_{k\geq 1}(2\pi k)^2 f_k^2. \tag{8}$$

The maximum likelihood estimator

$$\overline{\theta} = \operatorname*{Argmax}_{\tau\in\Theta^S}\int_{-1/2}^{1/2} f(t-\tau)d\mathbf{Y}(t)$$

is locally asymptotically minimax and its mean square risk satisfies, as $\varepsilon \to 0$,

$$\sup_{\theta\in\Theta^S}\mathbf{E}_{\theta,f}((\overline{\theta}-\theta)^2\mathcal{I}_\varepsilon(f)) = 1 + o(1).$$

Interestingly, in the semiparametric (regular) framework where f is not known but assumed to be continuously differentiable, it can be shown that exactly the same rate of convergence with the same constant can be attained for properly chosen estimates, see, for instance, [3] and references therein. The situation is completely different for less regular functions, as the comparison of the rates obtained in Theorem 1.3 (lower bound) and in the Proposition below (upper bound for known f) reveals.

Rates of convergence for known f. Both for purposes of comparison with the smooth case and for comparison with the rates for unknown f obtained in Section 1.3, it is now important to mention what type of rates of convergence arise in the parametric case where f is known and irregular. The case of functions with a bounded number of jump points has been studied in [6]. For instance, to estimate θ when $f = g_1$, the rate in terms of the quadratic risk can be shown to be within a constant times ε^2. More generally, for an "α"-regular f, one expects a rate in $\varepsilon^{1/\alpha}$. Indeed, we show below that this is true (up to a log-factor) for any f in a class S_α.

Since here f is known, so is also $f^{[K]}(\cdot) = \sum_{k=1}^{K} f_k\varepsilon_k(\cdot)$ and we define, for any sequence of integers $K = K(\varepsilon)$ such that $K(\varepsilon) \to +\infty$ as $\varepsilon \to 0$,

$$\overline{\theta}(K) \quad \triangleq \quad \operatorname*{Argmax}_{\tau\in\Theta^S}\int_{-1/2}^{1/2} f^{[K]}(t-\tau)d\mathbf{Y}(t).$$

Let $\lfloor x \rfloor$ be the integer part of the real x. The proof of the following proposition is sketched in Section 4.

Proposition 1. *Fix positive K_0, M and $\alpha \in]0,1[$. Let us define $l_\varepsilon = \log(\varepsilon^{-1})$ and $\overline{K}_\alpha = \lfloor (l_\varepsilon\varepsilon)^{-1/\alpha}\rfloor$. Then for any f in $S_\alpha(K_0, M)$,*

$$\lim_{\varepsilon\to 0}\sup_{\theta\in\Theta^S}\mathbf{P}_{\theta,f}\left(|\overline{\theta}(\overline{K}_\alpha)-\theta| \geq l_\varepsilon^{1/\alpha}\varepsilon^{1/\alpha}\right) = 0.$$

Notice that arbitrarily high rates can be achieved as the known signal f gets more and more irregular. We have seen that in the semiparametric framework, this will no longer be possible, which is very different from the regular case.

In the proposition an extra logarithmic factor arises. This might be due to the fairly broad class we consider (with no special continuity assumptions on f) or to the fact that our -fairly simple, see Section 4- method of proof is suboptimal.

1.6. Discussion

The obtained rates of convergence are summarized in Figure 1. Uniform fast rates of convergence are obtained for sieve-MLE estimators (the 'square'-curve) over appropriate spaces and the best possible rates in the minimax sense over these spaces are strictly below the parametric ones ('plus' and 'circle'-curves). This loss of speed is even worse if one considers larger spaces ('diamond'-curve). Hence, not assuming differentiability of f enables us to discover new properties of the models at stake, in particular the breakdown in the semiparametric rate for low regularities.

Furthermore, from a practical point of view, it is appealing to do as few regularity assumptions on f as possible to have a broader framework of study available. The fine tuning of K that is needed here, see (5), confirms the sensitivity observed in practice with respect to the choice of K of estimation algorithms like the ones considered in [7] or [1].

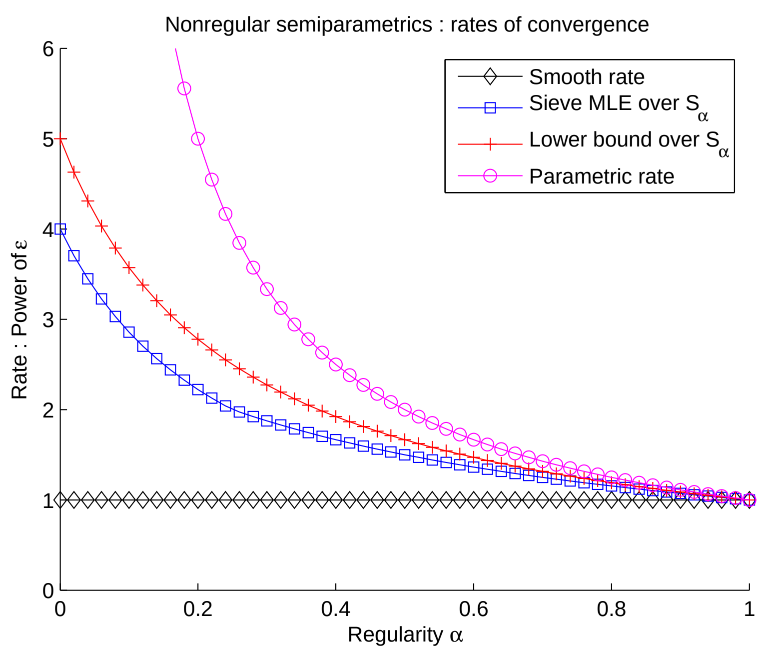

Fig 1. *Rates of convergence.*

The results presented here obviously only constitute a first step in the understanding of rates of convergence for the cumulated periodogram in the nonregular case, and many questions remain open. For instance, the "gap" between the 'square' and 'plus'-curves leads to two questions. The first one is: can we do better with sieve-MLEs (4) than the rate we obtain for the choice (5) of K^*? In other words, is the choice $K = K^*$ optimal? Although we shall not prove it here, we think it is, see the discussion at the end of the proof of Theorem 1.1. The second question is, if the sieve-MLEs cannot do better than the 'square'-curve, is the 'plus'-curve not optimal or can we improve estimation? We conjecture that there are estimators outperforming the sieve-MLEs, which would then be based on a different criterion than the cumulative periodogram. Since the most efficient practical algorithms we know (for instance the one in [7]) are based on model selection ideas with penalization using as a criterion the cumulated periodogram, it would also be highly desirable to answer the latter question to possibly obtain a different criterion leading to a faster algorithm.

A first improvement on the cumulated periodogram estimator would be to chose K in an adaptive (possibly data-driven) way and obtain a theorem which would yield the above rates for the resulting estimator in both regular and nonregular cases. A further direction would be to allow for (possibly data-driven) *weights* in front of each square in (3). For instance, one could consider a type of thresholding estimator which would put to zero the elements of the sum of (3) which are below the expected noise level. Understanding the behavior of such non-linear estimators in the present semiparametric context would be very desirable too, from both practical and theoretical perspective.

2. Upper bound proofs

Let us denote, for a process Y indexed by a subset Θ of the real numbers,

$$\| Y \| \triangleq \sup_{t\in\Theta} Y(t).$$

The notation $\lesssim$ (resp. $\gtrsim$) is used for "smaller (resp. larger) than or equal to a universal constant times". Sometimes we make universal constants explicit and denote them by C. Furthermore, $\mathcal{O}$ and o are the usual Landau symbols. For any integers m, n, we denote by $m \wedge n$ their minimum.

The criterion ℓ_K defined in (3) can be written

$$\ell_K(\tau) \quad = \quad \sum_{k=1}^{K} f_k^2 \cos^2(2\pi k(\tau-\theta)) + \varepsilon\eta_1(\tau) + \varepsilon^2\eta_2(\tau),$$

where

$$\eta_1(\tau) \quad = \quad 2\sum_{k=1}^{K} f_k \left\{\cos(2\pi k\tau)\xi_k + \sin(2\pi k\tau)\xi_k^*\right\} \cos(2\pi k(\tau-\theta)) \qquad (9)$$

$$\eta_2(\tau) \quad = \quad \sum_{k=1}^{K} \left\{\cos(2\pi k\tau)\xi_k + \sin(2\pi k\tau)\xi_k^*\right\}^2, \qquad (10)$$

and

$$\xi_k = \sqrt{2}\int_{-1/2}^{1/2} \cos(2\pi k t)dW(t), \qquad \xi_k^* = \sqrt{2}\int_{-1/2}^{1/2} \sin(2\pi k t)dW(t).$$

Note that $\{\xi_k\}_{k\geq 1}$ and $\{\xi_k^*\}_{k\geq 1}$ form two independent sequences of independent $\mathcal{N}(0,1)$ variables. Note also that η_2 is a pure noise term (that is, it does not depend on f). Let us write

$$\ell_K(\tau) - \ell_K(\theta) = -\sum_{k=1}^{K} f_k^2 \sin^2(2\pi k(\tau-\theta)) + \varepsilon(\eta_1(\tau)-\eta_1(\theta)) + \varepsilon^2(\eta_2(\tau)-\eta_2(\theta)).$$

Using the properties of the Fourier basis and setting $\delta = \tau - \theta$ one checks that

$$\mathbf{E}(\{\eta_1(\tau)-\eta_1(\theta)\}^2) = 4\sum_{k=1}^{K} f_k^2 \sin^2(2\pi k\delta) \triangleq 4v(\delta, K), \tag{11}$$

$$\mathbf{E}(\{\eta_2(\tau)-\eta_2(\theta)\}^2) = 4\sum_{k=1}^{K} \sin^2(2\pi k\delta). \tag{12}$$

2.1. Control of the stochastic parts of the criterion

To prove our upper-bound results, it is useful to be able to control the processes $\eta_1(\tau)-\eta_1(\theta)$ and $\eta_2(\tau)-\eta_2(\theta)$. We establish that their supremum over the sets of interest is controlled up to a logarithmic factor by the supremum of the standard deviations. For Gaussian processes (thus a priori only for $\eta_1(\tau)-\eta_1(\theta)$), this property is known to be true if one is able to control the entropy of the indexing set of the process with respect to the natural semimetric given by its covariance structure (see for example [9], Theorem 2.4 or [11], Appendix A.2.2). Here we shall instead use the fact that the two processes at stake are differentiable a.s. (which comes from their explicit dependence in τ, see Equations (9) and (10)) and we will apply the useful Lemma 2.1, that we know from [5].

Lemma 2.1. *Let $Z(t)$ be a stochastic process differentiable a.s., μ and x positive real numbers and I an interval of $\mathbb{R}$. Then it holds*

$$\mathbf{P}\left(\sup_{\tau\in I} Z(\tau) > x\right) \leq e^{-\mu x}\sup_{\tau\in I}\left(\mathbf{E}e^{2\mu Z(\tau)}\right)^{1/2}\left(1+\mu\int_{\tau\in I}(\mathbf{E}|Z'(\tau)|^2)^{1/2}d\tau\right).$$

Lemma 2.2. *There exist positive constants C_1, C_2, C_3, C_4 such that for K large enough, for any $y > 0$,*

$$\mathbf{P}\left(\|\eta_2 - K\| > y\sqrt{K}\right) \leq C_1 K\exp(-C_2 y\{y\wedge K^{1/2}\})$$

$$\mathbf{P}\left(\|\eta_2'\| > yK^{3/2}\right) \leq C_3 K^2\exp(-C_4 y\{y\wedge K^{3/2}\}).$$

Lemma 2.3. *Let γ_ε be a sequence tending to zero as $\varepsilon \to 0$ and $v(\cdot, K)$ be defined by (11). There exist positive constants C_5, C_6 such that for K large enough, for any positive real y,*

$$\sup_{\theta \in \Theta^S, f \in \mathcal{F}} \mathbf{P}_{\theta,f}\left(\sup_{\gamma_\varepsilon \leq |\tau-\theta| < 2\tau_0} \frac{|\eta_1(\tau) - \eta_1(\theta)|}{v(\tau-\theta, K)^{1/2}} \geq y\right) \leq C_5 \gamma_\varepsilon^{-4} K^4 \exp(-C_6 y^2).$$

The proof of Lemmas 2.2 and 2.3 can be found in Section 4.3. Note that Lemma 2.2 implies that if K is polynomial in ε^{-1}, then $\mathbf{P}(\|\eta_2 - K\| > l_\varepsilon \sqrt{K})$ and $\mathbf{P}(\|\eta_2'\| > l_\varepsilon K^{3/2})$ decrease to zero faster than any polynomial in ε, the same also holding for $\mathbf{P}(\|\eta_2(\cdot) - \eta_2(\theta)\| > l_\varepsilon \sqrt{K})$.

2.2. Proof of Theorems 1.1 and 1.2

Let us first briefly comment on the strategy of proof of Theorem 1.1. The main part of the proof is the local study of the criterion $\ell_K(\tau)$ around the true value $\tau = \theta$. Proving that it takes its global maximum close to $\tau = \theta$ gives us already a first "rough" rate of convergence. Then, we refine this first rate in successive steps by analyzing the dependence of $\ell_K(\tau) - \ell_K(\theta)$ in $\tau - \theta$.

The usual approach when f is sufficiently smooth is to expand the criterion ℓ_K around $\tau = \theta$ using Taylor's formula and to control the rest terms by bounds on the derivatives of ℓ_K, we refer to [3, 4] (and [5, 2] in the case of the period model). This approach cannot be used here, at least not until a sufficiently fast rate has been reached. The reason is that as long as $\delta > K^{-1}$, the functional $v(\delta, \varepsilon)$ does not behave like δ^2 but like $\delta^{2\alpha}$. A similar phenomenon (but this time independently of f) occurs with the pure noise part $\eta_2(\tau) - \eta_2(\theta)$: we shall use the bound $|\sin(2\pi k\delta)| \leq 1 \wedge 2\pi k|\delta|$, which yields different estimates depending on how close τ is to θ. In the "smooth" case, this discussion was unnecessary since K could be chosen small enough, corresponding to the fact that the smoother the function, the smaller the number of significative Fourier coefficients.

Proof of Theorem 1.1. For simplicity of notation, the cut-off K_S^* defined by (5) will be denoted by K. Let $\gamma_\varepsilon \geq \gamma_\varepsilon'$ be two sequences of positive reals and $v(\delta, K)$ defined by (11). Since by definition $\hat{\theta}_S$ is a point of maximum of the criterion, we have that

$$\begin{aligned}
&\mathbf{P}(\gamma_\varepsilon' < |\hat{\theta}_S - \theta| \leq \gamma_\varepsilon) \\
&\quad \leq \mathbf{P}\left(\sup_{\tau:\ \gamma_\varepsilon' < |\tau-\theta| \leq \gamma_\varepsilon} \ell_K(\tau) \geq \ell_K(\theta)\right) \\
&\quad \leq \mathbf{P}\left(\sup_{\delta:\ \gamma_\varepsilon' < |\delta| \leq \gamma_\varepsilon} -v(\delta, K) + 2\varepsilon(\eta_1(\tau) - \eta_1(\theta)) + \varepsilon^2(\eta_2(\tau) - \eta_2(\theta)) \geq 0\right).
\end{aligned}$$

This quantity can be further bounded above by

$$\begin{aligned}
&\mathbf{P}(\gamma'_\varepsilon < |\hat\theta_S - \theta| \le \gamma_\varepsilon) \\
&\qquad \le \mathbf{P}\left(4\varepsilon \sup_{\gamma'_\varepsilon < |\delta| \le \gamma_\varepsilon} \frac{\eta_1(\tau) - \eta_1(\theta)}{v(\delta, K)^{1/2}} \ge \inf_{\gamma'_\varepsilon < |\delta| \le \gamma_\varepsilon} v(\delta, K)^{1/2}\right) && (13) \\
&\qquad + \mathbf{P}\left(2\varepsilon^2 \sup_{\gamma'_\varepsilon < |\delta| \le \gamma_\varepsilon} \frac{\eta_2(\tau) - \eta_2(\theta)}{v(\delta, K)^{1/2}} \ge \inf_{\gamma'_\varepsilon < |\delta| \le \gamma_\varepsilon} v(\delta, K)^{1/2}\right). && (14)
\end{aligned}$$

We first establish that $\hat\theta_S$ converges at least at rate $\gamma_{1,\varepsilon} = l_\varepsilon \varepsilon K^{1/4}$. For this we prove that the right-hand side of the last display tends to zero if $\gamma_\varepsilon = 2\tau_0$ and $\gamma'_\varepsilon = \gamma_{1,\varepsilon}$. Note that $\gamma_{1,\varepsilon}$ tends to zero due to the choice $K = K^*$.

Note that $v(\delta, K) \ge f_1^2 \sin^2(2\pi\delta)$. Since $|\delta| < 2\tau_0$ and $2\tau_0$ is strictly less than $1/2$, we have that, using **(F)**, $\inf_{\delta:\ |\delta| > \gamma_{1,\varepsilon}} v(\delta, K)^{1/2} \gtrsim |f_1|\gamma_{1,\varepsilon} \gtrsim \gamma_{1,\varepsilon}$. Thus using Lemma 2.3, we obtain that (13) tends to zero as $\varepsilon \to 0$. To see that the same holds for (14), note that for some small $C > 0$,

$$\begin{aligned}
(14) \quad &\le \quad \mathbf{P}\left(\varepsilon^2 \sup_{\delta:\ \gamma'_\varepsilon < |\delta| \le \gamma_\varepsilon} \eta_2(\tau) - \eta_2(\theta) \ge \inf_{\delta:\ |\delta| > \gamma_\varepsilon} v(\delta, K)\right) && (15) \\
&\le \quad \mathbf{P}\left(\sup_{\delta:\ \gamma'_\varepsilon < |\delta| \le \gamma_\varepsilon} \eta_2(\tau) - \eta_2(\theta) \ge C l_\varepsilon^2 K^{1/2}\right).
\end{aligned}$$

Finally we use Lemma 2.2. Now we shall use repeatedly (13)-(14), improving each time on the rate of convergence γ'_ε assuming that f belongs to $S_\alpha(K_0, M)$.

Let us check that the rate of convergence over $S_\alpha(K_0, M)$ is at least $\gamma_{2,\varepsilon} = D_2 x_\varepsilon K^{-1}$, where $x_\varepsilon = l_\varepsilon^{1/4\alpha}$ and D_2 is a large enough constant (we do this only in the case $\alpha \le 3/4$, since for $3/4 < \alpha \le 1$, $\gamma_{1,\varepsilon} \le \gamma_{2,\varepsilon}$). Thus in this paragraph we work on the set $\{\delta :\ \gamma_{2,\varepsilon} < |\delta| \le \gamma_{1,\varepsilon}\}$. Since δ now tends to zero, $\lfloor 1/4|\delta| \rfloor \ge 1/8|\delta|$. Using the fact that for any $x \in [-\pi/2, \pi/2]$, $|\sin(x)| \ge 2|x|/\pi$, we have that for δ positive,

$$v(\delta, K) \ge \sum_{k=1}^{K \wedge \lfloor 1/4\delta \rfloor} f_k^2 \sin^2(2\pi k\delta) \gtrsim \delta^2 \sum_{k=1}^{K \wedge 1/8\delta} k^2 f_k^2.$$

The definition of the class S_α now implies that $v(\delta, K) \gtrsim \delta^2 \{K \wedge 1/8\delta\}^{2-2\alpha} \gtrsim \delta^{2\alpha}$ and thus it holds $\inf_{\delta:\ \gamma_{2,\varepsilon} < |\delta| \le \gamma_{1,\varepsilon}} v(\delta, K) \gtrsim x_\varepsilon^{2\alpha} K^{-2\alpha}$. To show that (13) and (14) (via (15)) tend to zero, we make use of Lemmas 2.3, 2.2. This can be done if the two following conditions are satisfied:

$$\varepsilon^{-1} x_\varepsilon^\alpha K^{-\alpha} \gtrsim \sqrt{l_\varepsilon} \quad \text{and} \quad \varepsilon^{-2} x_\varepsilon^{2\alpha} K^{-2\alpha} \gtrsim \sqrt{l_\varepsilon K}.$$

This happens if K is less than $K_{max} = \varepsilon^{\frac{-4\alpha}{1+4\alpha}}$. Since this is the case for the choice $K = K^*$, we have proved that $\hat\theta_S$ achieves the rate $\gamma_{2,\varepsilon}$. Finally we check that $\hat{\theta}_S$ achieves the rate $\gamma_{3,\varepsilon} = x_\varepsilon^4 r_\varepsilon$. On the set $\{\delta: \ \gamma_{3,\varepsilon} < |\delta| \leq \gamma_{2,\varepsilon}\}$,

$$v(\delta, K) \geq \sum_{k=1}^{K/x_\varepsilon} f_k^2 \sin^2(2\pi k \delta) \gtrsim \delta^2 (K/x_\varepsilon)^{2-2\alpha},$$

and thus $\inf_{\delta:\ \gamma_{3,\varepsilon}<|\delta|\leq\gamma_{2,\varepsilon}} v(\delta,K)^{1/2}$ is bounded from below by $\gamma_{3,\varepsilon}(K/x_\varepsilon)^{1-\alpha}$, which in view of the definitions of K^* and r_ε together with Lemma 2.3 implies that (13) tends to zero. Now note that since η_2 is differentiable, there exists a real c_τ such that $\eta_2(\tau) - \eta_2(\theta) = (\tau - \theta)\eta_2'(c_\tau) = \delta\eta_2'(c_\tau)$. Thus

$$\sup_{\gamma_{3,\varepsilon}<|\delta|\leq\gamma_{2,\varepsilon}} \frac{|\eta_2(\tau) - \eta_2(\theta)|}{v(\delta, K)^{1/2}} \lesssim (K/x_\varepsilon)^{\alpha-1} \sup_{\tau\in\Theta} |\eta_2'(\tau)|,$$

where we have used the bound in the one but last display to bound the denominator from below. We finally obtain that

$$\begin{aligned}
(14) \quad &\leq \quad \mathbf{P}\left(\sup_{\tau\in\Theta} |\eta_2'(\tau)| \gtrsim \varepsilon^{-2}(K/x_\varepsilon)^{1-\alpha} \inf_{\delta:\ \gamma_{3,\varepsilon}<|\delta|\leq\gamma_{2,\varepsilon}} v(\delta,K)^{1/2}\right) \\
&\leq \quad \mathbf{P}\left(\sup_{\tau\in\Theta} |\eta_2'(\tau)| \gtrsim \varepsilon^{-2}(K/x_\varepsilon)^{2-2\alpha}\gamma_{3,\varepsilon}\right).
\end{aligned}$$

Using the definitions of K^* and r_ε, we see that $\varepsilon^{-2}(K/x_\varepsilon)^{2-2\alpha}\gamma_{3,\varepsilon} \gtrsim K^{3/2} l_\varepsilon$. Now Lemma 2.2 implies that (14) tends to zero faster than any power of ε, which concludes the proof of the first part of the Theorem.

To obtain the second part of the Theorem, we apply Lemma 4.1 with the rate $v_\varepsilon = l_\varepsilon^{1/\alpha} r_\varepsilon^S$. From what precedes we have that $\mathbf{P}_{\theta,f}(|\hat{\theta}_S - \theta| > v_\varepsilon)$ tends to zero faster than any power of ε. Since f belongs to the class $S_\alpha(K_0, M)$, we have $I_2(K) \geq K^{2-2\alpha}$ and (18) is bounded by $\varepsilon^2 K^{2\alpha-2} + \varepsilon^4 K^{4\alpha-1}$. By definition of $K = K_S^*$, this is within a constant of the target rate r_ε^S. Now note that $I_4(K) \leq K^2 I_2(K)$ and $I_3(K) \leq K I_2(K)$. For $1/4 \leq \alpha < 1$, using these bounds together with the expressions of $r(\varepsilon)$, K_S^* and v_ε, we obtain that $r(\varepsilon) = o(r_\varepsilon^S)$. Hence in this case the quadratic risk of $\hat{\theta}_S$ is within a constant of r_ε^S. □

Proof of Theorem 1.2. We reproduce the beginning of the proof of Theorem 1.1 and obtain a rate $\gamma_\varepsilon = l_\varepsilon \varepsilon \tilde{K}^{1/4}$ (note that this intermediate rate is faster than $\tilde{K}^{-1}$). Now we proceed as in the last step of the proof for $\hat{\theta}_S$. Using the estimates $v(\delta, \tilde{K}) \gtrsim \delta^2$ together with Lemmas 2.3 and 2.2, we obtain a deviation inequality with rate $l_\varepsilon \varepsilon$.

To obtain the result for the quadratic risk, we can now apply Lemma 4.1 with $v_\varepsilon = l_\varepsilon \varepsilon$. Since $I_2(K)$ is bounded from below by a positive constant, (18) is at most a constant times ε^2. It is now enough to check that $r(\varepsilon)$ is negligible. Applying the Cauchy-Schwarz inequality, $I_4(K) \leq (I_2(K) I_6(K))^{1/2} \leq I_2(K)^{1/2} K^3$. Further using that $I_3(K) \leq K^3$ and $K = \tilde{K}$, we obtain $r(\varepsilon) = o(\varepsilon^2)$. □

Remark. One can note that the rate in Theorem 1.1 is determined by the value of $\tau - \theta$ such that the pure-noise process $\varepsilon^2(\eta_2(\tau) - \eta_2(\theta))$ becomes dominant in the balancing of $-\sum_{k=1}^{K} f_k^2 \sin^2(2\pi k(\tau-\theta))+\varepsilon(\eta_1(\tau)-\eta_1(\theta))+\varepsilon^2(\eta_2(\tau)-\eta_2(\theta))$. Values of K essentially faster or slower than K^* eventually lead to rates for $\tau-\theta$ slower than r_ε^S when doing the preceding balancing. By contrast, η_1 determines the rate in the parametric case studied in Theorem 1.2.

3. Lower bound proofs

To prove a lower bound on the rate of convergence when estimating θ, we will follow a well-known approach, which is outlined beautifully in Pollard's so far unpublished book Asymptopia: choose $\mathbf{P}_{\theta,g_1},\dots,\mathbf{P}_{\theta,g_m}$ and $\mathbf{P}_{\tau,g_1},\dots,\mathbf{P}_{\tau,g_m}$ with τ far enough away from θ, such that

$$\mathbf{Q}_\theta \triangleq \frac{1}{m}\sum_{k=1}^{m}\mathbf{P}_{\theta,g_k} \quad \text{and} \quad \mathbf{Q}_\tau \triangleq \frac{1}{m}\sum_{k=1}^{m}\tilde{\mathbf{P}}_{\tau,g_k}$$

remain close in L^1. Then we can use the following inequality:

$$\inf_{\hat\theta}\sup_{(\theta,f)} \mathbf{E}_{\theta,f}\left(\hat\theta-\theta\right)^2 \geq \frac{1}{4}(\theta-\tau)^2\left(1-\frac{1}{2}\,\|\mathbf{Q}_\theta-\mathbf{Q}_\tau\|_1\right).$$

So our goal is to maximize $(\theta-\tau)^2$, keeping $\|\mathbf{Q}_\theta-\mathbf{Q}_\tau\|_1$ away from 2. In both the proofs of Theorem 1.4 and of Theorem 1.3, we do this by choosing a smooth function f_m (which may in fact not depend on m), and bounding respectively $\chi^2(\mathbf{P}_{\theta,f_m},\mathbf{Q}_\theta)$, $\chi^2(\mathbf{P}_{\tau,f_m},\mathbf{Q}_\tau)$ (these can be bounded using the same technique) and $\chi^2(\mathbf{P}_{\theta,f_m},\mathbf{P}_{\tau,f_m})$, which is relatively easy since f_m is smooth. Since for any two probability measures $\mathbf{P}$ and $\mathbf{Q}$ we have

$$\|\mathbf{P}-\mathbf{Q}\|_1 \leq \sqrt{\chi^2(\mathbf{P},\mathbf{Q})},$$

we can then use the triangle inequality for the L^1-norm to bound $\|\mathbf{Q}_\theta-\mathbf{Q}_\tau\|_1$. This scheme is depicted in Figure 2.

3.1. Proof of Theorem 1.4

To prove Theorem 1.4, we need the following lemma.

Lemma 3.1. *Suppose $f\in L^2[0,1]$ and the perturbations $\eta_1,\dots,\eta_m\in L^2[0,1]$. Define*

$$\mathbf{P}_1=\mathbf{P}_{\theta,f} \quad \textit{and} \quad \mathbf{P}_2=\frac{1}{m}\sum_{k=1}^{m}\mathbf{P}_{\theta,f+\eta_k}.$$

Then it holds

$$\chi^2(\mathbf{P}_1,\mathbf{P}_2)+1=\mathbf{E}_{\mathbf{P}_1}\left(\frac{d\mathbf{P}_2}{d\mathbf{P}_1}\right)^2=\frac{1}{m^2}\sum_{1\leq i,j\leq m}\exp\left(\frac{1}{\varepsilon^2}\int_{-1/2}^{1/2}\{\eta_i\eta_j\}(t-\theta)dt\right).$$

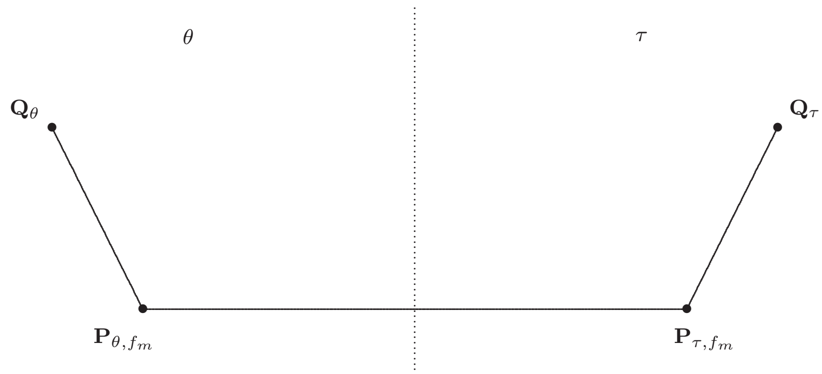

FIG 2. *Idea behind lower bound proofs.*

The proof of this Lemma follows from the Girsanov formula and some elementary calculations. Note that the result does not depend on f.

Now choose the smooth function

$$f(x) = \cos(2\pi x).$$

We will choose perturbations $\eta_1, \dots, \eta_m$ such that $f + \eta_k \in \mathcal{F}_\alpha(\eta, M_1)$, and such that $\mathbf{P}_{\theta,f}$ and $\frac{1}{m}\sum_{k=1}^m \mathbf{P}_{\theta,f+\eta_k}$ are arbitrarily close together as $m \to \infty$. This will show that you cannot estimate θ in the class $\mathcal{F}_\alpha$ faster than in the smooth case.

We start with an element $\phi \in L^2[-1/2, 1/2]$, with support in $(-1/4, 1/4)$, and such that $\phi \in \mathcal{F}_\alpha(1/2, R)$, for some constant R we will specify later. Recall that elements of $\mathcal{F}$ are extended by 1-periodicity. However, to define our perturbations, we shall just need the values of ϕ within $(-1/2, 1/2]$, so we set $\phi_0(x) = \phi(x)\mathbf{1}_{(-1/2,1/2]}(x)$ for all x, where $\mathbf{1}_B$ is the indicator of B. For fixed m and any x in $(-1/2, 1/2]$, we define for $k = 1, \dots, m$

$$\eta_k(x) = \sqrt{m}\,\phi_0\left(m\left(x + \frac{1}{4m} - \frac{k}{2m}\right)\right) + \sqrt{m}\,\phi_0\left(m\left(-x + \frac{1}{4m} - \frac{k}{2m}\right)\right).$$

Then we define $\eta_k(x)$ for any x in $\mathbb{R}$ by 1-periodicity. These rescaled, translated and symmetrized versions of ϕ are all orthogonal, since their supports are disjoint. We wish to check that $f + \eta_k \in \mathcal{F}_\alpha(\eta, M_1)$. First we remark that bounding the first Fourier coefficient of $f + \eta_k$ from below is not a problem, since

$$\left|\int_{-1/2}^{1/2} \eta_k(x) e^{-2\pi i x}\, dx\right| \le \int_{-1/2}^{1/2} |\eta_k(x)|\, dx \le \frac{2}{\sqrt{m}} \int_{-1/2}^{1/2} |\phi(x)|\, dx,$$

so this can be done by choosing m large enough. Also note that for every $m \ge 2$ and $k = 1, \dots, m$, we have

$$\|\eta_k\|_2^2 = 2\|\phi\|_2^2,$$

so the L^2-norm of the η_k's remains bounded. Finally, using the inequality $(a+b)^2 \geq \frac{1}{2}a^2 - b^2$, we see that

$$\begin{aligned}
&\int_{-1/2}^{1/2} \left[f(x-\theta) + \eta_k(x-\theta) - f(x-\tau) - \eta_k(x-\tau)\right]^2 \, dx \\
&\qquad \geq \frac{1}{2}\int_{-1/2}^{1/2} \left[\eta_k(x-\theta) - \eta_k(x-\tau)\right]^2 \, dx - \int_{-1/2}^{1/2} \left[f(x-\theta) - f(x-\tau)\right]^2 \, dx \\
&\qquad \geq \frac{1}{2}\int_{-1/2}^{1/2} \left[\eta_k(x-\theta) - \eta_k(x-\tau)\right]^2 \, dx - 2\pi^2(\theta-\tau)^2.
\end{aligned}$$

In order to control the first part of the right-hand side, we assume that m is even, since otherwise $\eta_{(m+1)/2}$ will have period $1/2$, which might cause problems. If m is even and $|\theta - \tau| > 1/2m$, but $|\theta - \tau| < \eta \leq 1/2$, then for every k, $\eta_k(x-\theta)$ will have at least one "bump" (either the left scaled copy of ϕ or the right one) which is disjoint from the support of $\eta_k(x-\tau)$, and vice versa. This means that

$$\int_{-1/2}^{1/2} \left[\eta_k(x-\theta) - \eta_k(x-\tau)\right]^2 \, dx \geq 2\|\phi\|_2^2 \qquad \left(\text{for } |\theta-\tau| > \frac{1}{2m}\right).$$

Furthermore, if $|\theta - \tau| \leq 1/2m$, then it is not hard to see that

$$\begin{aligned}
\int_{-1/2}^{1/2} \left[\eta_k(x-\theta) - \eta_k(x-\tau)\right]^2 \, dx \quad &\geq \quad \int_{-1/2}^{1/2} \left[\phi(y-m\theta) - \phi(y-m\tau)\right]^2 \, dy \\
&\geq \quad Rm^{2\alpha}(\theta-\tau)^{2\alpha}.
\end{aligned}$$

To guarantee that $f + \eta_k \in \mathcal{F}_\alpha(\eta, M_1)$, it is therefore enough to make sure that $\|\phi\|_2$ is larger than some lower bound depending on η and M_1 (but not on m!), in which case the relevant inequalities are satisfied for all m big enough. Increasing $\|\phi\|_2 = \|\eta_k\|_2$ might cause the function $f + \eta_k$ to leave the class $\mathcal{F}(\rho, L)$, unless we choose L big enough. Now define

$$\mathbf{Q}_\theta = \frac{1}{m}\sum_{k=1}^{m} \mathbf{P}_{\theta, f+\eta_k}.$$

Then Lemma 3.1 tells us that, using the fact that the η_k's are orthogonal,

$$\|\mathbf{P}_{\theta,f} - \mathbf{Q}_\theta\|_1 \leq \sqrt{\chi^2(\mathbf{P}_{\theta,f}, \mathbf{Q}_\theta)} \leq \frac{1}{\sqrt{m}}\sqrt{e^{2\|\phi\|_2^2/\varepsilon^2} - 1}.$$

This means that by choosing m large enough, we can put $\mathbf{P}_{0,\theta}$ arbitrarily close to $\mathbf{P}_{\theta,f}$. Furthermore, it is clear that

$$\|\mathbf{P}_{\theta,f} - \mathbf{Q}_\theta\|_1 = \|\mathbf{P}_{\tau,f} - \mathbf{Q}_\tau\|_1.$$

Since f is smooth, we can keep $\|\mathbf{P}_{\theta,f} - \mathbf{P}_{\tau,f}\| \leq 1$ when $|\theta - \tau|$ is of the order ε:

$$\begin{aligned}
\|\mathbf{P}_{\theta,f} - \mathbf{P}_{\tau,f}\| \quad &\leq \quad \sqrt{\chi^2(\mathbf{P}_{\theta,f}, \mathbf{P}_{\tau,f})} \\
&= \quad \sqrt{\exp\left(\varepsilon^{-2}\int_0^1 (f(t-\tau) - f(t-\theta))^2 \, dt\right) - 1} \leq 1,
\end{aligned}$$

where the last inequality holds if

$$\int_0^1 (f(t-\tau) - f(t-\theta))^2 \, dt = 2\sin^2(2\pi(\theta - \tau)) \le \frac{1}{2}\varepsilon^2.$$

This can be done by choosing $|\theta - \tau| = \varepsilon/4\pi$. By the arguments belonging to Figure 2 the minimax result of Theorem 1.4 then follows.

3.2. Proof of Theorem 1.3

Here we will need slightly more complicated perturbations than in the previous section. We therefore also need a new lemma:

Lemma 3.2. *Suppose $\phi_1, \ldots, \phi_m \in L^2[0,1]$ are orthogonal and $f \in L^2[0,1]$ and $\|\phi_i\|_2^2 \le M$. Extend these functions periodically. Suppose the interval I has integer length. Define $\mathbf{P}_f$ as the measure corresponding to the model*

$$dX(t) = f(t)dt + dW(t) \quad (t \in I).$$

For $m \ge 1$ define $\mathcal{W} = \{-1,1\}^m$ and for $w \in \mathcal{W}$, define

$$\phi_w = \sum_{i=1}^m w_i \phi_i.$$

Finally, let us define

$$\mathbf{Q}^{\mathcal{W}} = 2^{-m} \sum_{w \in \mathcal{W}} \mathbf{P}_{f+\phi_w}.$$

Then there exists a constant $C > 0$ depending only on M such that

$$\mathbf{E}_{\mathbf{P}_f} \left(\frac{d\mathbf{Q}^{\mathcal{W}}}{d\mathbf{P}_f} \right)^2 \le \exp\left(C \sum_{i=1}^m \left(\int_I \phi_i^2(t)\, dt \right)^2 \right).$$

Proof. We start by noting that if F is a primitive of f, we have

$$\begin{aligned} \frac{d\mathbf{P}_{f+\phi_w}}{d\mathbf{P}_f}(X) &= \exp\left(\int_I \phi_w(t)\, dX(t) - \frac{1}{2}\int_I \phi_w^2(t)\, dt - \int_I f(t)\phi_w(t)\, dt \right) \\ &= \exp\left(\int_I \phi_w(t)\, d(X-F)(t) - \frac{1}{2}\int_I \phi_w^2(t)\, dt \right). \end{aligned}$$

This means that, using the fact that the ϕ_i's are orthogonal,

$$\begin{aligned} \frac{d\mathbf{Q}^{\mathcal{W}}}{d\mathbf{P}_f}(X) &= 2^{-m} \sum_{w \in \mathcal{W}} \exp\left(\int_I \phi_w(t)\, d(X-F)(t) - \frac{1}{2}\int_I \phi_w^2(t)\, dt \right) \\ &= \exp\left(-\frac{1}{2} \sum_{i=1}^n \int_I \phi_i^2(t)\, dt \right) 2^{-m} \sum_{w \in \mathcal{W}} \exp\left(\int_I \phi_w(t)\, d(X-F)(t) \right) \end{aligned}$$

$$= \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\int_I \phi_i^2(t)\,dt\right) \times$$
$$\prod_{i=1}^{m}\frac{1}{2}\left[\exp\left(\int_I \phi_i(t)\,d(X-F)(t)\right) + \exp\left(-\int_I \phi_i(t)\,d(X-F)(t)\right)\right].$$

So we get, using the fact that the random variables $\int_I \phi_i(t)\,dW(t)$ are independent,

$$\begin{aligned}
\mathbf{E}_{\mathbf{P}_f}\left(\frac{d\mathbf{Q}^{\mathcal{W}}}{d\mathbf{P}_f}\right)^2 &= \exp\left(-\sum_{i=1}^{n}\int_I \phi_i^2(t)\,dt\right) \times \\
&\quad \mathbf{E}\left(\prod_{i=1}^{m}\frac{1}{4}\left[2 + e^{2\int_I \phi_i(t)\,dW(t)} + e^{-2\int_I \phi_i(t)\,dW(t)}\right]\right) \\
&= \prod_{i=1}^{m}\frac{1}{2}\left[\exp\left(-\int_I \phi_i^2(t)\,dt\right) + \exp\left(\int_I \phi_i^2(t)\,dt\right)\right].
\end{aligned}$$

Now choose $C > 0$ such that for all $|u| \le M$, $\cosh(u) \le 1 + Cu^2 \le e^{Cu^2}$. Then we get

$$\mathbf{E}_{\mathbf{P}_f}\left(\frac{d\mathbf{Q}^{\mathcal{W}}}{d\mathbf{P}_f}\right)^2 \le \exp\left(C\sum_{i=1}^{m}\left(\int_I \phi_i^2(t)\,dt\right)^2\right). \qquad \square$$

Remember the definition of our class:

$$S_\alpha(K_0, M) = \left\{ f \in \mathcal{F} : \quad \forall K \ge K_0 : \sum_{k=-K}^{K} k^2 |c_k|^2 \ge M K^{2-2\alpha} \right\}.$$

Choose $C > 1$ big enough such that $g \in S_\alpha(K_0, M)$, where g has Fourier coefficients

$$g_k = C k^{-\frac{1}{2}-\alpha} \ (k \ge 1) \text{ and } g_{-k} = g_k.$$

Here we need to choose L, the bound on the L^2-norm of functions in $\mathcal{F}$, big enough. Now fix $m \ge 2$ and denote $\hat{\psi}(k) = \int_0^1 \psi(x) e^{-2ik\pi x} dx$ the Fourier coefficients of a square-integrable function ψ. Define $\phi_1, \ldots, \phi_m \in L^2[0,1]$ by their Fourier coefficients.

$$\begin{aligned}
\hat{\phi}_1(m) &= C \cdot m^{-\frac{1}{2}-\alpha} \\
\hat{\phi}_i(k) &= C \cdot k^{-\frac{1}{2}-\alpha} \quad (2 \le i \le m \text{ and } m + n_{i-1} \le k < m + n_i).
\end{aligned}$$

For all other combinations of i and $k \ge 0$ we take $\hat{\phi}_i(k) = 0$, and we set $\hat{\phi}_i(-k) = \hat{\phi}_i(k)$. Here, $n_1 = 1$, $n_m = +\infty$, and the other n_i's are chosen such that for a fixed $R > 4$ (and all $m \ge 2$) we have

$$\|\phi_i\|_2^2 \le R C^2 m^{-2\alpha-1}.$$

The reason that this is possible, is because there exists $R > 4$ such that

$$\sum_{k=m}^{\infty} C^2 k^{-1-2\alpha} \leq \frac{1}{4} R C^2 m^{-2\alpha} \quad (\forall\ m \geq 2).$$

It is clear that we can make sure that for $1 \leq i \leq m-1$,

$$\frac{1}{m} \sum_{k=m}^{\infty} C^2 k^{-1-2\alpha} - C^2 \cdot m^{-2\alpha-1} \leq \frac{1}{2} \|\phi_i\|_2^2 \leq \frac{1}{4} R C^2 m^{-2\alpha-1}.$$

This means that

$$\|\phi_m\|_2^2 \leq \frac{2}{m} \sum_{k=m}^{\infty} C^2 k^{-1-2\alpha} + 2C^2 \cdot m^{-2\alpha} \leq R C^2 m^{-2\alpha-1}.$$

Now define $f_m \in L^2[0,1]$ through its Fourier coefficients:

$$\hat{f}_m(k) = \begin{cases} 0 & \text{for } k = 0, \\ C|k|^{-\frac{1}{2}-\alpha} & \text{for } 1 \leq k \leq m-1 \text{ and } -m+1 \leq k \leq -1, \\ 0 & \text{for } |k| \geq m. \end{cases}$$

Define for $w \in \mathcal{W} = \{-1,1\}^m$

$$\phi_w = \sum_{i=1}^{m} w_i \phi_i.$$

Clearly, $|\hat{f}_m(k) + \hat{\phi}_w(k)| = |\hat{g}(k)|$, so $f_m + \phi_w \in S_\alpha(K_0, M)$. Also, f_m and $\phi_1, \ldots, \phi_m$ are all orthogonal. Define

$$\mathbf{Q}_\theta^{\mathcal{W}} = 2^{-m} \sum_{w \in \mathcal{W}} \mathbf{P}_{\theta, f_m + \phi_w}.$$

According to (a rescaled version of) Lemma 3.2, if we wish to bound $\chi^2(\mathbf{Q}_\theta^{\mathcal{W}}, \mathbf{P}_{\theta, f_m})$, it is enough to bound

$$\sum_{i=1}^{m} \varepsilon^{-4} \|\phi_i\|_2^4 \leq R^2 C^4 m^{-1-4\alpha} \varepsilon^{-4}.$$

This remains bounded if we choose

$$m \approx \varepsilon^{-\frac{4}{1+4\alpha}}.$$

Now define

$$\mathbf{Q}_\tau^{\mathcal{W}} = 2^{-m} \sum_{w \in \mathcal{W}} \mathbf{P}_{\tau, f_m + \phi_w}.$$

Clearly, the above argument also shows that we have bounded $\chi^2(\mathbf{Q}_\tau^{\mathcal{W}}, \mathbf{P}_{\tau, f_m})$. Following the scheme depicted in Figure 2, we then need to bound $\chi^2(\mathbf{P}_{\theta, f_m}, \mathbf{P}_{\tau, f_m})$. This can be done by bounding

$$\varepsilon^{-2} \int_0^1 \left(f_m(x-\theta) - f_m(x-\tau)\right)^2 dx.$$

However,

$$\begin{aligned}\varepsilon^{-2}\int_0^1 (f_m(x-\theta)-f_m(x-\tau))^2\,dx &= 2\varepsilon^{-2}\sum_{k=1}^{m-1}\left|e^{i2\pi k\theta}-e^{i2\pi k\tau}\right|^2\hat{f}(k)^2\\ &\leq 8\pi^2(\theta-\tau)^2\varepsilon^{-2}\sum_{k=1}^{m-1}k^2\hat{f}(k)^2\\ &\lesssim (\theta-\tau)^2\varepsilon^{-2}m^{2-2\alpha}\lesssim(\theta-\tau)^2\varepsilon^{-\frac{10}{1+4\alpha}}.\end{aligned}$$

This means that if we choose

$$\theta-\tau\approx\varepsilon^{\frac{5}{1+4\alpha}},$$

we can bound $\chi^2(\mathbf{P}_{\theta,f_m},\mathbf{P}_{\tau,f_m})$, which proves the lower bound in Theorem 1.3, using the arguments belonging to Figure 2.

4. Appendix

4.1. Rates in the parametric case

We start by proving Proposition 1.

Proof of Propositions 1. The proof is very much in the spirit of the proof of Theorem 1.1 but much easier since there are just two terms in the criterion and no pure noise term η_2. Indeed, starting again from the definition of $\overline{\theta}(\overline{K}_\alpha)$ and using the noteworthy fact that the functions $f^{[K]}(\cdot-\tau)$ and $(f-f^{[K]})(\cdot-\theta)$ are orthogonal in $L^2[-1/2,1/2]$, one obtains

$$\begin{aligned}\overline{\theta}(K) &= \operatorname*{Argmax}_{\tau\in\Theta^S}\quad \int_{-1/2}^{1/2}\{f^{[K]}(t-\tau)-f^{[K]}(t-\theta)\}dW(t)\\ &\qquad -\frac{1}{2}\int_{-1/2}^{1/2}\{f^{[K]}(t-\tau)-f^{[K]}(t-\theta)\}^2dt.\end{aligned}$$

The remainder of the proof now closely follows the proof of Theorem 1.1. The deterministic part in the preceding display equals $-2\sum_{k=1}^{K}f_k^2\sin^2(\pi k\delta)$ whereas the process part plays the role of $\eta_1(\tau)-\eta_1(\theta)$. □

We also notice that a similar result can be obtained for f in some class $\mathcal{F}_\alpha$, any $0<\alpha<1$. In this case, one rather considers the (full-, instead of sieved-) maximum-likelihood estimator

$$\overline{\theta}\quad\triangleq\quad\operatorname*{Argmax}_{\tau\in\Theta^S}\int_{-1/2}^{1/2}f(t-\tau)d\mathbf{Y}(t).$$

Using the fact that the observation process $\mathbf{Y}$ follows (1), one verifies that

$$\overline{\theta}=\operatorname*{Argmax}_{\tau\in\Theta^S}\int_{-1/2}^{1/2}\{f(t-\tau)-f(t-\theta)\}dW(t)-\frac{1}{2}\int_{-1/2}^{1/2}\{f(t-\tau)-f(t-\theta)\}^2dt.$$

Proposition 2. *Suppose for $f \in \mathcal{F}$ there exist $0 < \beta < \alpha$, $\eta > 0$ and $M_1, M_2 > 0$ such that*

$$M_1(\tau-\theta)^{2\alpha} \leq \int_{-1/2}^{1/2} \{f(t-\theta) - f(t-\tau)\}^2 \, dt \leq M_2(\tau-\theta)^{2\beta}$$

for all $\theta, \tau \in \Theta^S$ with $|\theta - \tau| < \eta$. Then

$$\lim_{\varepsilon \to 0} \sup_{\theta \in \Theta^S} \mathbf{P}_{\theta,f}\left(|\overline{\theta} - \theta| \geq l_\varepsilon \varepsilon^{1/\alpha}\right) = 0.$$

The control from below defines the degree of irregularity of the functions in the class, whereas the control from above is to ensure that the supremum of the process $\int_{-1/2}^{1/2}\{f(t-\tau) - f(t-\theta)\}dW(t)$ is essentially controlled by the supremum of the standard deviations. Note that the control from below alone does not guarantee that this last process has continuous sample paths. The proof of Proposition 2 is similar to that of Proposition 1 and is omitted.

4.2. Expansion of the quadratic risk of the estimators

From probabilities of deviation from the true θ_0 it is sometimes possible to obtain results for the quadratic risk without logarithmic factors. This can be done by a "localization" argument following an idea used in [3].

The idea is to introduce a quantity $\hat{\tau}(K)$ defined by a well-chosen linear approximation of the Taylor expansion of the function $\tau \to \ell_K(\tau)$ around $\tau = \theta$. By definition of $\hat{\theta}(K)$, we have $\ell'_K(\hat{\theta}(K)) = 0$. Taylor's theorem applied to ℓ'_K around $\tau = \theta$ thus yields the existence of a (random) real ζ such that

$$0 = \ell'_K(\theta) + (\hat{\theta}(K) - \theta)\ell_K^{(2)}(\theta) + \frac{1}{2}(\hat{\theta}(K) - \theta)^2 \ell_K^{(3)}(\zeta). \tag{16}$$

Now let us define $\hat{\tau}(K)$ by the identity

$$0 = \ell'_K(\theta) + (\hat{\tau}(K) - \theta)\mathbf{E}_{\theta,f}\ell_K^{(2)}(\theta). \tag{17}$$

For any integer $p \geq 1$ let us denote $I_p(K) = \sum_{k=1}^K k^p f_k^2$.

Lemma 4.1. *Let $\hat{\tau}(K)$ be defined by (17). As $\varepsilon \to 0$, it holds,*

$$\mathbf{E}_{\theta,f}(\hat{\tau}(K) - \theta)^2 = \varepsilon^2 I_2(K)^{-1} + \mathcal{O}(\varepsilon^4 K^3 I_2(K)^{-2}). \tag{18}$$

Assume, as $\varepsilon \to 0$, that $\mathbf{P}_{\theta,f}(|\hat{\theta}(K) - \theta| > v_\varepsilon) = o(\varepsilon^\eta)$ for any $\eta > 0$, where K is bounded by some power of ε^{-1}. Then as $\varepsilon \to 0$, it holds

$$\mathbf{E}_{\theta,f}(\hat{\theta}(K) - \theta)^2 = \mathcal{O}\left(\mathbf{E}_{\theta,f}(\hat{\tau}(K) - \theta)^2 + r(\varepsilon)\right),$$

where $r(\varepsilon) = v_\varepsilon^2\left[\varepsilon^2(I_4(K) + \varepsilon^2 K^5)I_2(K)^{-2} + (I_3(K) + \varepsilon^2 K^4)v_\varepsilon^2 I_2(K)^{-2}\right]$.

Let us now briefly explain how this Lemma can sometimes be used to obtain rates without log-factors. Assume that one already has obtained some rate say $v_\varepsilon = l_\varepsilon^a r_\varepsilon$, for some $a > 0$ in the sense that $\mathbf{P}_{\theta,f}(|\hat{\theta}(K) - \theta| > v_\varepsilon)$ tends to zero faster than any power of ε^{-1}. This is the case for Theorems 1.1 and 1.2 for instance. If for this rate v_ε it holds that $r(\varepsilon) = o(\mathbf{E}_{\theta,f}(\hat{\tau}(K) - \theta)^2)$ then the quadratic risk of $\hat{\theta}(K)$ is in fact given by (18).

The lemma also provides a further justification of the choice $K = K_S^*$ in (5) and an interesting interpretation of the rate r_ε^S in (6) in the case $1/4 < \alpha < 1$. In that case it can be checked that the choice $K = K_S^*$ provides the best trade-off in K in (18) for the class of functions $S_\alpha(K_0, M)$, and that for this K the remainder term $r(\varepsilon)$ is negligible. Moreover, the obtained risk in (18) is then $(r_\varepsilon^S)^2$. The situation for $\alpha < 1/4$ is more complex. The best K from the point of view of (18) on $S_\alpha(K_0, M)$ would be $K = +\infty$, but then of course the remainder $r(\varepsilon)$ comes in. For $K = K_S^*$ the term $r(\varepsilon)$ is in fact larger than v_ε. This might be due to the fact that our bound $r(\varepsilon)$ is suboptimal for $\alpha < 1/4$ or that this method breaks down for low regularities.

4.3. Proof of technical lemmas

Proof of Lemma 2.2. By definition,

$$\begin{aligned}
\eta_2(\tau) - K &= \sum_{k=1}^{K} \{\cos(2\pi k\tau)\xi_k + \sin(2\pi k\tau)\xi_k^*\}^2 \\
\eta_2'(\tau) &= 2\sum_{k=1}^{K} (2\pi k)\{\cos(2\pi k\tau)\xi_k + \sin(2\pi k\tau)\xi_k^*\} \times \\
&\qquad \{-\sin(2\pi k\tau)\xi_k + \cos(2\pi k\tau)\xi_k^*\}
\end{aligned}$$

To control their Laplace transforms note that for any fixed τ,

$$\eta_2(\tau) \overset{\mathcal{D}}{=} \sum_{k=1}^{K} \alpha_k^2 \quad \text{and} \quad \eta_2'(\tau) \overset{\mathcal{D}}{=} 2\sum_{k=1}^{K}(2\pi k)\alpha_k\tilde{\alpha}_k \overset{\mathcal{D}}{=} 2\sum_{k=1}^{K}(2\pi k)(\beta_k^2 - \tilde{\beta}_k^2),$$

where $\overset{\mathcal{D}}{=}$ denotes equality in distribution and $\alpha_k, \tilde{\alpha}_k, \beta_k, \tilde{\beta}_k$ are independent standard normal. Thus for any $0 < \mu < 1/8$, due to the fact that $-\log(1-v) \leq v + v^2$ for $-1/2 < v < 1/2$,

$$\mathbf{E}(\exp\{2\mu(\eta_2 - K)\}) = \exp(-2\mu K)\exp\left\{-\sum_{k=1}^{K}\frac{1}{2}\log(1-4\mu)\right\} \leq \exp(8\mu^2 K).$$

Similarly, there exists $C > 0$ such that

$$\mathbf{E}(\exp\{2\mu\eta_2'\}) \leq \exp(C\mu^2 K^3).$$

By direct computation one checks $\mathbf{E}(\eta_2'(\tau)^2) \lesssim K^3$ and $\mathbf{E}(\eta_2''(\tau)^2) \lesssim K^5$. We apply Lemma 2.1 to the processes $\eta_2(\tau) - K$ and $\eta_2'(\tau)$ with the respective choices $x = K^{1/2}y$, $\mu = (K^{-1/2}y \wedge 1)/8$ and $x = K^{3/2}y$, $\mu = (K^{-3/2}y \wedge 1)/D$, with D a large enough constant. $\square$

Proof of Lemma 2.3. We shall again apply Lemma 2.1. Let us denote

$$\zeta_1(\tau) = \frac{\eta_1(\tau) - \eta_1(\theta)}{v(\tau - \theta, K)^{1/2}}.$$

Note that due to (11), $\mathbf{E}(\zeta_1(\tau)^2) = 1$ thus $\mathbf{E}(\exp\{2\mu\zeta_1(\tau)\}) = \exp(2\mu^2)$. On the other hand, by direct calculation and using the inequality $(a+b)^2 \le 2a^2 + 2b^2$,

$$\begin{aligned}\mathbf{E}(\zeta_1'(\tau)^2) &\lesssim \frac{\mathbf{E}(\eta_1'(\tau)^2)}{v(\tau - \theta, K)} + \mathbf{E}\{\eta_1(\tau) - \eta_1(\theta)\}^2 \left\{ \frac{\sum_{k=1}^K (2\pi k) f_k^2 \sin(4\pi k\{\tau - \theta\})}{2v(\tau - \theta, K)^{3/2}} \right\}^2 \\ &\lesssim K^3 v(\delta, K)^{-1} + \left\{ \sum_{k=1}^K (2\pi k) f_k^2 \right\}^2 v(\delta, K)^{-2} \lesssim K^3 \gamma_\varepsilon^{-2} + K^2 \|f\|_2^4 \gamma_\varepsilon^{-4},\end{aligned}$$

for any τ such that $\gamma_\varepsilon \le |\tau - \theta| \le 2\tau_0$, which concludes the proof. $\square$

Proof of Lemma 4.1. The identity (18) follows using the explicit expressions of the derivatives $\ell_K'(\tau)$ and $\ell_K^{(2)}(\tau)$ evaluated at $\tau = \theta$. To obtain the second part of the Lemma, one introduces the event $\mathcal{A} = \{|\hat\theta(K) - \theta| \le v_\varepsilon\}$ and use the bounds, for any $\eta > 0$

$$\begin{aligned}\mathbf{E}_{\theta,f}(\hat\theta(K) - \theta)^2 &= \mathbf{E}_{\theta,f}\{(\hat\theta(K) - \theta)^2 1_{\mathcal{A}}\} + \mathbf{E}_{\theta,f}\{(\hat\theta(K) - \theta)^2 1_{\mathcal{A}^c}\} \\ &\le 2\mathbf{E}_{\theta,f}(\hat\tau(K) - \theta)^2 1_{\mathcal{A}} + 2\mathbf{E}_{\theta,f}(\hat\theta(K) - \hat\tau(K))^2 1_{\mathcal{A}} + o(\varepsilon^\eta) \\ &\lesssim (18) + \mathbf{E}_{\theta,f}(\hat\theta(K) - \hat\tau(K))^2 1_{\mathcal{A}}.\end{aligned}$$

It suffices bound the last term by $\mathcal{O}(r(\varepsilon))$. To do so, substracting (17) and (16) we see that it is enough to control the moments of order 2 of $\{\ell_K^{(2)}(\theta) - \mathbf{E}_{\theta,f}\ell_K^{(2)}(\theta)\}$ and $\sup_{\zeta\in\Theta} |\ell_K^{(3)}(\zeta)|$. Simple computations similar to [3], Lemma 6 lead to

$$\begin{aligned}\mathbf{E}_{\theta,f}\{\ell_K^{(2)}(\theta) - \mathbf{E}_{\theta,f}\ell_K^{(2)}(\theta)\}^2 &= \mathcal{O}(\varepsilon^2 I_4(K) + \varepsilon^4 K^5) \\ \mathbf{E}_{\theta,f} \sup_{\zeta\in\Theta} |\ell_K^{(3)}(\zeta)|^2 &= \mathcal{O}(I_3(K) + \varepsilon^2 K^4),\end{aligned}$$

as $\varepsilon \to 0$, from which the result easily follows. $\square$

Acknowledgments

The authors would like to thank the Statistics department of the Université Paris-Sud for offering them the possibility to work together on this subject. The first author would like to thank Professor Aad van der Vaart for stimulating discussions.

References

[1] Castillo, I. Penalized profile likelihood methods and second order properties in semiparametrics. *PhD Thesis, Université Paris-Sud* (2006).

[2] Castillo, I. Semi-parametric second-order efficient estimation of the period of a signal. *Bernoulli 13*, 4 (2007), 910–932. MR2364219

[3] Dalalyan, A., Golubev, G., and Tsybakov, A. Penalized maximum likelihood and semiparametric second order efficiency. *Ann. Statist. 34*, 1 (2006), 169–201. MR2275239

[4] Dalalyan, A. S. Stein shrinkage and second-order efficiency for semiparametric estimation of the shift. *Math. Methods Statist. 16*, 1 (2007), 42–62. MR2319470

[5] Golubev, G. K. Estimating the period of a signal of unknown shape corrupted by white noise. *Problems of Information Transmission 24*, 4 (1988), 288–299. MR0979694

[6] Ibragimov, I. A., and Has'minskii, R. Z. *Statistical estimation: asymptotic theory*. Springer, New York e.a., 1981. MR0620321

[7] Lavielle, M., and Lévy-Leduc, C. Semiparametric estimation of the frequency of unknown periodic functions and its application to laser vibrometry signals. *IEEE Trans. Signal Process. 53*, 7 (2005), 2306–2314. MR2168952

[8] Raimondo, M. Minimax estimation of sharp change points. *Ann. Statist. 26*, 4 (1998), 1379–1397. MR1647673

[9] Talagrand, M. Sharper bounds for Gaussian and empirical processes. *Ann. Probab. 22*, 1 (1994), 28–76. MR1258865

[10] van der Vaart, A. W. *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. Cambridge, 1998. MR1652247

[11] van der Vaart, A. W., and Wellner, J. A. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics. MR1385671