Semiparametric second order efficient estimation of the period of a signal

I. Castillo*

* Department of Mathematics, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, the Netherlands. E-mail: i.castillo@few.vu.nl

Abstract

We consider the problem of estimating the period of an unknown periodic function in Gaussian white noise. We construct a class of estimators of the period by a penalized maximum likelihood method. We obtain a second order asymptotic expansion of the risk of such estimators and construct a second order efficient estimator when the unknown function belongs to some known class.

Keywords: Exact Minimax Asymptotics, Penalized Maximum Likelihood, Second Order Efficiency, Semiparametric Estimation, Unknown Period.
Subject Class. MSC-2000 : 62G05, 62G20.

1 Introduction

The framework. We consider the following model,

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

where f is an unknown real periodic function with period 1, θ is a period parameter which we want to estimate and $W(t)$ is the standard Brownian motion on $[-T/2, T/2]$. We assume that θ belongs to an interval $\Theta = [\alpha_T, \beta_T]$, where α_T, β_T are reals such that $0 < \alpha_T < \beta_T < +\infty$. We consider the asymptotic framework $T \to +\infty$.

Motivation. When f is known, the maximum likelihood estimator is asymptotically efficient, see Ibragimov & Has'minskii (1981, Theorem 5.1 p.203). Here, we are interested in the semiparametric problem: the parameter θ has to be estimated but f is an infinite-dimensional nuisance parameter. In a seminal paper, Golubev (1988) gives an asymptotically efficient estimator of the period in this framework. In a recent work, Gassiat & Lévy-Leduc (2006) obtain an asymptotically efficient estimator of the period in a discretized version of (1) and address the problem of estimation of multiple periods when the signal is a sum of different periodic functions.

From a practical point of view, the problem of period estimation arises in many different areas, such as communications, seismic signal processing or laser vibrometry (see Prenat (2001)). Model (1) can be seen as the signal observed by a receptor situated at some distance from a vibrating source. The frequency of the source vibration equals θ^{-1} and the noise $dW(t)$ stands for the degradation of the original signal due to the distance between the source and receptor. For other practical applications and further references on the subject, we refer to Gassiat & Lévy-Leduc (2006) and Lavielle & Lévy-Leduc (2005). In this last work, the authors propose a practical method to estimate the frequency in the discrete-time model cited above, when the number of observations at hand is fixed.

Semiparametric estimators and second order efficiency. We now return to the general semiparametric framework. As explained in van der Vaart (1998, Chap. 25) or in Bickel, Klaassen, Ritov & Wellner (1998), the aim is to estimate the parameter without knowing the nuisance function f and, if possible, to obtain the same optimal asymptotic variance for the estimator as in the parametric framework. If this is possible, which is the case in model (1), we say that there is no information loss.

However, for a given model, there often is a large choice of asymptotically efficient estimators. This motivates the study of the second order term for the estimators quadratic risk. This problem was studied for partial linear models by Golubev & Härdle (2000). In this paper, the authors construct second order efficient estimators when the nuisance function belongs to a known functional class. In Golubev & Härdle (2002), the authors give nonparametric adaptive versions of their estimators. For essentially nonlinear models, the first result on semiparametric second order efficiency was established by Dalalyan, Golubev & Tsybakov (2006). In this paper, the translation model $x(t) = f(t-\theta)+\varepsilon n(t)$, $t \in [-1/2, 1/2]$ is considered, and the authors study, as ε tends to zero, the asymptotic second order properties of a class of penalized maximum likelihood estimators. Second order adaptive versions of these results are obtained in Dalalyan (2005).

As noted in Dalalyan et al. (2006), the study of this problem is of particular interest since, *asymptotically*, the second order terms in a semiparametric setup are often not much smaller than the first order ones. Typically, as we shall see for model (1), the first order term being $T^{-3/2}$, the second order term can be of the order of $T^{-3/2} \times T^{-2/5}$.

Objective and results. The goal of this paper is to investigate the second order efficiency in the semiparametric problem of period estimation. One of the key tools in the paper of Dalalyan et al. (2006) is the fact that the model can be projected onto a basis and thus be described as a discrete collection of independent submodels. As we shall see in the sequel, this property does not hold in our framework. Thus we have to work globally on the whole model and hence introduce appropriate methodologies, as for instance the formulation of the criterion upon which the estimation is based as the action of a compact operator over some functional space.

In this paper, we obtain the risk second order term over a large family of period estimators. We also obtain a lower bound for the second order term, which is achieved by an estimator of the preceding family. This is the first result about second order estimation in the period model. It provides a theoretical basis for the choice of the smoothing parameters in the construction of the estimator. It also highlights the important role played by the nonparametric nuisance part of the model. For instance it originally motivated the work of Castillo, Lévy-Leduc & Matias (2006), where the authors study the problem of sharp adaptive estimation of f in model (1).

Structure of the paper. In Section 2, we construct a class of estimators using the penalized maximum likelihood method. Section 3 contains our main asymptotic results: we give the second order properties of the preceding estimators and study the lower bound for the risk over all estimators. Moreover, we construct an estimator achieving the optimal second order rate. Section 4 is devoted to technical proofs.

2 Penalized Maximum Likelihood Estimator

Let us first introduce some useful notation. As the function f in (1) is 1-periodic, we may assume that f can be written as a convergent Fourier series

$$f(x) = \sum_{k\in\mathbb{Z}} c_k e^{2i\pi kx} = a_1 + \sqrt{2}\sum_{k\geqslant 1} a_{2k}\cos(2\pi kx) + a_{2k+1}\sin(2\pi kx) = \sum_{k\geqslant 1} a_k\varepsilon_k(x) \ , \quad (2)$$

where $\varepsilon_1(x) = 1$ and for $k \geqslant 1$, $\varepsilon_{2k}(x) = \sqrt{2}\cos(2\pi kx)$ and $\varepsilon_{2k+1}(x) = \sqrt{2}\sin(2\pi kx)$.

The Fisher information for known f in model (1) is, as T tends to infinity,

$$\mathrm{I}_T(f,\theta) = \{1 + o(1)\}\frac{T^3}{12\theta^4}\sum_{k\in\mathbb{Z}}(2\pi k)^2|c_k|^2. \quad (3)$$

Note that this quantity is four times smaller than the one given in Ibragimov & Has'minskii (1981, p.209) since here the observation interval is $[-T/2, T/2]$ and not $[0, T]$. Note also that it depends on *both* f and θ. To simplify the notation, we denote it by I_T in the sequel when we do not want to emphasize this dependence.

We know that there are estimators whose asymptotic variance is the inverse of I_T, for instance, the estimator in Golubev (1988). However, among all asymptotically efficient estimators of θ, is it possible to choose one which would be the best one in a certain sense ? In particular, is it possible to find an estimator achieving an optimal second order rate ?

We now construct a family of estimators of the period with smoothing parameters which allows us to deal with efficiency at the second order. We use the method of penalized maximum likelihood introduced in Dalalyan et al. (2006). Note that here, model (1) cannot be partitioned in a discrete collection of projected 1-dimensional submodels since the space of projection would be $L^2([0,\theta])$ and θ is unknown. Thus the object of our study is the global likelihood function of the model.

The likelihood for estimating θ in model (1) depends on θ and f, which is given by its Fourier coefficients (a_k). To eliminate the nuisance parameters (a_k), we first assume that the a_ks are independent centered Gaussian random variables with variance ${\sigma_k}^2$ and independent of the noise, that is we put a prior distribution on (a_k). To estimate the a_ks, we maximize the posterior distribution of (a_k) conditionally to the observations $\{x(t)\}$. Note that this is equivalent to the maximization of the joint likelihood of $(\{x(t)\},(a_k))$ or also of its logarithm which, thanks to the Girsanov formula, see for example Ibragimov & Has'minskii (1981, Appendix 2), is given by the following function Φ

$$\Phi\left[\tau,\{x(t)\},(a_k)\right] = -\frac{1}{2}\int_{-T/2}^{T/2} f(t/\tau)^2 dt + \int_{-T/2}^{T/2} f(t/\tau)dx(t) - \sum_{k\geqslant 1}\frac{{a_k}^2}{2{\sigma_k}^2}. \quad (4)$$

Taking the partial derivatives of the last quantity with respect to the Fourier coefficients a_k and using the approximation $\int_{-T/2}^{T/2}\varepsilon_k(t/\tau)f(t/\tau)dt \approx a_k T$, we find that the maximum is approximately obtained for (a_k^*) such that

$$(T+\sigma_k^{-2})a_k^* = \int_{-T/2}^{T/2} \varepsilon_k(t/\tau)dx(t).$$

Let f^* be the function with Fourier coefficients (a_k^*), then $\int_{-T/2}^{T/2} f^*(t/\tau)dx(t) = \sum_{k\geqslant 1}(T+\sigma_k^{-2})a_k^{*2}$ and $\int_{-T/2}^{T/2} f^*(t/\tau)^2 dt \approx T\sum_{k\geqslant 1} a_k^{*2}$. Thus,

$$\Phi[\tau,\{x(t)\},(a_k^*)] \approx \sum_{k\geqslant 1} \frac{1}{T+\sigma_k^{-2}} \left(\int_{-T/2}^{T/2} \varepsilon_k(t/\tau)dx(t) \right)^2 .$$

For symmetry reasons (for the minimax problem, we will assume that f lies in a Sobolev ellipsoid, see Section 3.3), for $k \geqslant 1$, we impose $\sigma_{2k} = \sigma_{2k+1}$. This means that we put same weights on sine and cosine for a given frequency. Hence, writing, for $k \geqslant 0$, $\lambda_k = (T+\sigma_{2k+1}^{-2})^{-1}T$, which is in $[0,1]$, we obtain the following weighted criterion,

$$\mathrm{L}(\tau) = \sum_{k\geqslant 1} \frac{\lambda_k}{T} \left| \int_{-T/2}^{T/2} e^{2ik\pi t/\tau} dx(t) \right|^2 . \tag{5}$$

Note that we have dropped the first term in the sum since it does not depend on θ. Moreover there is no restriction using the Fourier basis, the preceding construction can be used with any orthonormal basis $\{\varepsilon_k\}$ of $\mathrm{L}^2[0,1]$.

Equation (5) can be seen as a weighted version of the estimator in Golubev (1988) . Note that the weight λ_k is outside the square of the integral. We are thus introducing a weight on the 'energy' and not directly on the data, contrary to data tapers methods, studied in the context of frequency estimation among others by Chen, Wu & Dahlhaus (2000). In fact, the results obtained by the two methods are of different natures. We also note from equation (4) that the estimation of the nuisance parameters a_k is done by an (approximate) *penalized* maximum likelihood method and thus the 'Bayesian approach' leads, in fact, to penalization (see for instance Kimeldorf & Wahba (1970)).

We are now able to construct our estimator. As in Golubev (1988) or Gassiat & Lévy-Leduc (2006), direct maximization of (5) would not allow us to distinguish the multiples of the unknown period. To avoid this problem, we aim to take the smallest approximate minimizer of $\mathrm{L}(\tau)$ by choosing

$$\mathcal{E}_T = \left\{ \tau\in\Theta\ ,\ \ \mathrm{L}(\tau) \geqslant (1-\log^{-1/4} T)\sup_{\tau\in\Theta} \mathrm{L}(\tau) \right\} \tag{6}$$

$$e_T = \inf \mathcal{E}_T. \tag{7}$$

Now let $B(x,R) = \{\tau\in\Theta,\ |\frac{x}{\tau}-1| < R\}$. We define our estimator as:

$$\theta^* = \operatorname*{Argmax}_{\tau\in B(e_T,1/4)} \mathrm{L}(\tau). \tag{8}$$

In order to understand the behavior of the criterion $\mathrm{L}(\tau)$, we introduce some useful notation. We define symmetrized weights over $\mathbb{Z}$ by letting $\lambda_{-k} = \lambda_k$ and we set $\lambda_0 = 0$. The symbols $\sum_k$, $\sum_{k\neq 0}$, $\sum_{k\geqslant 0}$ denote respectively sums over $\mathbb{Z}$, $\mathbb{Z}^*$ and $\mathbb{N}$. Then

$$\mathrm{L}(\tau) = [\Gamma(\tau)+\mathrm{X}(\tau)+\Psi(\tau)]/2, \qquad \text{where}$$

$$\begin{cases} \Gamma(\tau) &= \sum_k \lambda_k T^{-1} \left| \int_{-T/2}^{T/2} e^{2ik\pi t/\tau} f(t/\theta) dt \right|^2 \\ \mathrm{X}(\tau) &= 2 \sum_k \lambda_k T^{-1} \int_{-T/2}^{T/2} e^{2ik\pi t/\tau} f(t/\theta) dt \int_{-T/2}^{T/2} e^{-2ik\pi t/\tau} dW(t) \\ \Psi(\tau) &= \sum_k \lambda_k T^{-1} \left| \int_{-T/2}^{T/2} e^{2ik\pi t/\tau} dW(t) \right|^2 . \end{cases} \quad (9)$$

3 Second order asymptotics in the period model

Let us begin this section by some definitions. We say that a function is a $o(1)$ (respectively a $O(1)$) if it tends to zero (resp. is bounded) as T goes to infinity. If (z_k) is a sequence of complex numbers indexed by $\mathbb{Z}$,

$$\|z\|^2 = \sum_k |z_k|^2 \ , \quad \|z\|_1 = \sum_k |z_k| \quad \text{and for } m \geqslant 1 \ , \quad \|z^{(m)}\|^2 = \sum_k |z_k|^2 (2\pi k)^{2m}.$$

We denote simply $\|z'\|$ instead of $\|z^{(1)}\|$. If v is a square-integrable function, $\|v\|$ denotes its L^2-norm on $[0,1]$. In the sequel, C denotes a universal constant.

3.1 Assumptions on the model

Let us recall that $\Theta = [\alpha_T, \beta_T]$. We assume that, as T tends to $+\infty$,

$$(\mathbf{P1}) \quad \alpha_T^{-1} = O(T) \qquad\qquad (\mathbf{P2}) \quad \beta_T = O(\log T).$$

We assume that f belongs to some class $F = F(\rho, C_0)$ of smooth functions whose Fourier coefficients (c_k) satisfy the Fourier expansion (2) and

$$(\mathbf{F1}) \quad |c_1|^2 \geqslant \rho > 0 \qquad\qquad (\mathbf{F2}) \quad \sum_k (2\pi k)^4 |c_k|^2 \leqslant C_0 < +\infty \,.$$

We consider sequences of weights (λ_k) such that $\lambda_0 = 0$, $\lambda_1 = 1$ and for all integer k. $\lambda_{-k} = \lambda_k$ and $0 \leqslant \lambda_k \leqslant 1$. We also assume that there exists N_T going to $+\infty$ such that

$$(\mathbf{W0}) \quad \lambda_k = 0 \text{ for } k \geqslant N_T, \qquad \text{and } N_T^4 = o(T), \text{ as } T \text{ tends to } +\infty.$$

Moreover, we assume that there exist positive constants ρ_1 and C_1 such that:

$$(\mathbf{W1}) \quad \|\lambda'\| \geqslant \rho_1 \log^2 T \max_{k \geqslant 1} \lambda_k (2\pi k)$$

$$(\mathbf{W2}) \quad \sum_k \lambda_k (2\pi k)^4 \leqslant C_1 T.$$

Finally, we use the following technical assumption:

$$(\mathbf{T}) \quad \left[\sum_k (1-\lambda_k)(2\pi k)^2 |c_k|^2 \right]^2 = o \left[\frac{1}{\log T} \sum_k (1-\lambda_k)^2 (2\pi k)^2 |c_k|^2 \right].$$

The notation o and O in the previous assumptions are meant to be *uniform* with respect to f and λ. In the sequel, all o and O will be so. Assumptions (**F1**) and (**F2**) are regularity conditions on f. Assumptions (**W0**), (**W1**) and (**W2**) are satisfied for a quite large variety of weight sequences. For instance, they are fulfilled for projection weights $(\mathbf{1}_{|k| \leqslant N_T})$ provided that $N_T \geqslant C \log^4 T$ and $N_T^5 = O(T)$.

Let us briefly study two consequences of the preceding assumptions. First, (**F1**) and (**F2**) imply that there exists a constant $h > 0$ - for example $h = \rho/C_0$ -, such that for any integer $p \geqslant 2$,

$$\sum_{q \neq 0} |c_{pq}|^2 \leqslant (1-h) \sum_{q \neq 0} |c_q|^2. \tag{10}$$

Second, since the weights are between 0 and 1, we have $(1-\lambda_k)^2 \leqslant (1-\lambda_k)$ and thus

$$\sum_k (1-\lambda_k)(2\pi k)^2 |c_k|^2 \times \sum_k (1-\lambda_k)^2 (2\pi k)^2 |c_k|^2 \leqslant \left[\sum_k (1-\lambda_k)(2\pi k)^2 |c_k|^2\right]^2.$$

Using (**T**), we get

$$\sum_k (1-\lambda_k)(2\pi k)^2 |c_k|^2 = o(\log^{-1} T). \tag{11}$$

In particular note that for each fixed k, the weight λ_k tends to 1 as T tends to $+\infty$.

3.2 Second order asymptotics for the risk

Let us introduce the following functional, where the c_ks are the Fourier coefficients of f.

$$R_T(f, \lambda) = \sum_k (2\pi k)^2 \left((1-\lambda_k)^2 |c_k|^2 + \frac{1}{T} \lambda_k^2 \right). \tag{12}$$

This functional corresponds to a term of *nonparametric* estimation. In fact, suppose you have at hand a Gaussian sequence model defined by $y_k = c_k + T^{-1/2}\varepsilon_k$, where $\varepsilon_k = (1/\sqrt{2})(\varepsilon_{1,k} + i\varepsilon_{2,k})$ and $\{\varepsilon_{1,k}\}$, $\{\varepsilon_{2,k}\}$ are independent sequences of standard normal random variables. Denoting by f the function with Fourier coefficients c_k, suppose you want to estimate the derivative f' given by the Fourier coefficients $c'_k = 2ik\pi c_k$. Now consider the linear estimator $\widehat{f'}$ defined by the Fourier coefficients $(\widehat{c'_k}) = (\lambda_k(2ik\pi)y_k)$. Then (12) is nothing but the quadratic risk $\mathbf{E}(\|\widehat{f'} - f'\|^2)$.

Theorem 1. *Under assumptions* (**P**),(**F**),(**W**) *and* (**T**)*, uniformly in $f \in F$ and in $\theta \in \Theta$, the estimator θ^* defined by* (8) *satisfies, as T tends to $+\infty$,*

$$\mathbf{E}_{\theta,f}((\theta^* - \theta)^2 \mathrm{I}_T(f, \theta)) = 1 + \{1 + o(1)\} \frac{R_T(f, \lambda)}{\|f'\|^2}. \tag{13}$$

Remark 1. The preceding theorem gives an expansion at the order 2 for the risk. Thanks to (11), $\sum_k (2\pi k)^2 (1-\lambda_k)^2 |c_k|^2 = o(\log^{-1} T)$. Moreover, (**W0**) gives $T^{-1}\sum_k (2\pi k)^2 \lambda_k^2 \leqslant CT^{-1} N_T^3 = o(T^{-1/4})$. Hence $R_T(f, \lambda) = o(1)$ as T tends to $+\infty$.

Remark 2. The second order term $\|f'\|^{-2} R_T(f, \lambda)$ is the same as the one obtained by Dalalyan et al. (2006) and similar to the one obtained by Golubev & Härdle (2000). It seems to be a general feature of smooth semiparametric models that the term of order 2 reflects the estimation of the score function of the model.

Remark 3. In this paper we consider model (1) on the centered interval $[-T/2, T/2]$. It can be checked that for a different interval of length T, for example $[0, T]$, Equation (13) still holds. Note that the semiparametric optimal first order convergence rate is still I_T. In fact, one can prove that $\|f'\|^2(12\theta^4)^{-1}$ is the *efficient Fisher information* for the model $dX(t) = f(t/\theta)dt + dW(t),\ t \in \mathcal{I}$, where $\mathcal{I}$ is any interval of length T.

Proof. Here we give the main lines of the proof, the technical aspects being postponed to Section 4. A transversal object is the criterion L on which the definition of θ^* relies. We study its deterministic part Γ in Subsection 4.1 by Fourier techniques. In Subsection 4.2 we study the stochastic parts X and Ψ by introducing a well-chosen operator on $\mathrm{L}^2[-T/2, T/2]$. We need both global expansions, which means for all τ in Θ, and local expansions in a neighborhood of the true θ.

The first step is to obtain consistency results for θ^*. Let us define the event

$$\mathcal{A}_0 = \left\{ \sup_{\tau\in\Theta} |\,\mathrm{X}(\tau) + \Psi(\tau)| \leqslant T^{3/4} \right\}. \tag{14}$$

The complementary of $\mathcal{A}_0$ has probability $o(T^{-p})$ for any integer p, thanks to Lemma 6.

Thanks to Lemma 1, with $\gamma_T = T^{1/8}$, we have $\sup_{\tau\in\Theta} \Gamma(\tau) \leqslant T\{\sum_{k\neq 0} |c_k|^2 + o({\gamma_T}^{-1/2})\}$. Thus, on the event $\mathcal{A}_0$, it holds $\sup_{\tau\in\Theta} \mathrm{L}(\tau) \leqslant T\{\sum_{k\neq 0} |c_k|^2 + o({\gamma_T}^{-1/2}) + T^{-1/4}\}/2$.

On the other hand, using again Lemma 1 and the definition of $\mathcal{A}_0$, we have, on $\mathcal{A}_0$, $2L(\theta) \geqslant T\{\sum_{k\neq 0} |c_k|^2 + o(\log^{-1} T) - T^{-1/4}\}$. Thus $\theta \in \mathcal{E}_T$ on $\mathcal{A}_0$ for T large enough.

For any integers $p \geqslant 2$ and $j \geqslant 1$, upper-bounding the weights by 1 and using (10), we get $\sum_q \lambda_{qj} |c_{qp}|^2 T \leqslant \sum_{q\neq 0} |c_{qp}|^2 T \leqslant (1-h) \sum_{q\neq 0} |c_q|^2 T$. Hence with $\varepsilon_T = T^{-3/8}$, we have $\mathcal{E}_T \subset \cup_{j\geqslant 1} B(j\theta, \varepsilon_T)$ on $\mathcal{A}_0$. Therefore $e_T \in B(\theta, \varepsilon_T)$ on $\mathcal{A}_0$.

The definition of θ^* then implies that $\theta^* \notin B(2\theta, \varepsilon_T)$ on $\mathcal{A}_0$, thus:

$$\theta^* \in B(\theta, \varepsilon_T) \text{ on } \mathcal{A}_0.$$

In fact, we need to refine this rate. For any $D > 0$, let us define the event $\mathcal{A}_1$ as

$$\mathcal{A}_1 = \{|\theta^* - \theta| T^{3/2} \theta^{-2} \leqslant D \log^{1/2} T\}\ , \tag{15}$$

Lemma 11 establishes that for every integer p, for D large enough, $\mathbf{P}({\mathcal{A}_1}^c)$ is a $o(T^{-p})$.

The next step is to introduce the following variable $\widehat{\tau}$, which is a theoretical tool for our proof (of course, it is not an estimator since it requires the knowledge of θ)

$$\mathrm{L}'(\theta) + (\widehat{\tau} - \theta)\mathbf{E}(\mathrm{L}''(\theta)) = 0. \tag{16}$$

By Lemma 12, $\widehat{\tau}$ has the desired expansion at the order 2. It remains to show that θ^* and $\widehat{\tau}$ are close enough, so that θ^* has the same expansion as $\widehat{\tau}$ at the order 2.

$$\mathbf{E}((\theta^* - \theta)^2 \mathrm{I}_T) = \mathbf{E}((\theta^* - \theta)^2 \mathrm{I}_T \mathbf{1}_{\mathcal{A}_1}) + \mathbf{E}((\theta^* - \theta)^2 \mathrm{I}_T \mathbf{1}_{\mathcal{A}_1^c}).$$

Using (**P1**), (**P2**) and (43), reaches the conclusion that

$$\mathbf{E}((\theta^* - \theta)^2 \mathrm{I}_T \mathbf{1}_{\mathcal{A}_1^c}) \leqslant C\beta_T^2 \alpha_T^{-4} T^3 \mathbf{P}(\mathcal{A}_1^c) \ ,$$

which is negligible compared to a given power of T choosing D large enough in (15). Now

$$\mathbf{E}((\theta^* - \theta)^2 \mathrm{I}_T \mathbf{1}_{\mathcal{A}_1}) = \mathbf{E}((\widehat{\tau} - \theta)^2 \mathrm{I}_T) + \mathbf{E}([(\theta^* - \widehat{\tau})^2 + 2(\theta^* - \widehat{\tau})(\widehat{\tau} - \theta)]\mathrm{I}_T \, 1_{\mathcal{A}_1}) - \mathbf{E}((\widehat{\tau} - \theta)^2 \mathrm{I}_T \mathbf{1}_{\mathcal{A}_1^c}).$$

The first term in the above expression gives the appropriate expansion thanks to Lemma 12, while the last one is negligible thanks to (43), (**P1**) and (**P2**). The middle term must still be dealt with and this is carried out by Lemma 13. Hence the expansions of the risk for θ^* and $\widehat{\tau}$ are the same, which proves Theorem 1. □

3.3 Lower bound for the risk

In this section, we establish a lower bound in the minimax sense for the second order term over all possible estimators of θ. For any $\beta > 0$, $L > 0$ let us define

$$\mathcal{W}_{\beta,L} = \left\{ f = \{c_k\}_{k \in \mathbb{Z}} \ , \quad \sum_k |2\pi k|^{2\beta} |c_k|^2 \leqslant L \right\}.$$

The expression of the second order term in Theorem 1 suggests to minimize the functional $R_T(f, \lambda)$. The behavior of this functional is well understood, see Pinsker (1980) or Tsybakov (2004, Chap.3) for a complete overview on the subject. There exist a function s in $\mathcal{W}_{\beta,L}$ and a sequence q in l^2 such that (s, q) is a saddle point for $R_T(f, \lambda)$ over $\mathcal{W}_{\beta,L} \times l^2(\mathbb{Z})$. It holds

$$r_T = R_T(s, q) = \inf_{\lambda \in l^2} \sup_{f \in \mathcal{W}_{\beta,L}} R_T(f, \lambda) = \frac{1}{T} \sum_k (2\pi k)^2 q_k. \tag{17}$$

One has explicit expressions of s and q in terms of the solution W_T of the equation

$$\frac{1}{T} \sum_{k \in \mathbb{Z}^*} \left[\left(\frac{W_T}{|k|} \right)^{\beta - 1} - 1 \right]_+ (2\pi |k|)^{2\beta} = L. \tag{18}$$

Let g be a function in $F(\rho, C_0)$. Let us denote by g_k its Fourier coefficients. For any $\delta > 0$ and fixed $\beta > 0, L > 0$, we define a neighborhood of g as follows:

$$F_\delta(g) = \{ f = g + v, \ \|v\| \leqslant \delta, \ v \in \mathcal{W}_{\beta,L} \}.$$

Theorem 2. *Suppose that $\beta \geqslant 2$ and $L > 0$. For any $\delta_T \to 0$ such that there exists $\alpha > 0$ with $\delta_T^2 T / W_T^{1+\alpha} \to +\infty$, as T tends to $+\infty$,*

$$\inf_{\hat{\theta}} \sup_{\theta, \ f \in F_{\delta_T}(g)} \mathbf{E}_{\theta,f}((\hat{\theta} - \theta)^2 \mathrm{I}_T(\theta, f)) \geqslant 1 + \{1 + o(1)\} \frac{r_T}{\|g'\|^2} \ ,$$

where the infimum is taken over all estimators $\hat{\theta}$ based on the observations $\{x(t)\}$ and where r_T is defined by (17).

Proof. The first step is to change variables so that, contrary to (3), the Fisher information will not depend on the parameter of interest anymore. Let us define $\omega = \theta^{-1}$. It is not hard to check that the Fisher information $J_T(f)$ for estimating ω in model (1) equals

$$J_T(f) = \theta^4 I_T(\theta, f) = \{1 + o(1)\}T^3\|f'\|^2/12$$

and in particular, up to the $o(1)$ term, does not depend on θ. Since the mapping $\hat\theta \to \hat\theta/\theta^2$ between the set of the estimators of θ and the set of the estimators of ω is one to one, we obtain

$$\inf_{\hat\theta} \sup_{\theta\in\Theta,\ f\in F_{\delta_T}(g)} \mathbf{E}_{\theta,f}((\hat\theta - \theta)^2 \mathrm{I}_T(\theta, f)) = \inf_{\hat\omega} \sup_{\omega,\ f\in F_{\delta_T}(g)} \mathbf{E}_{\omega,f}((\hat\omega - \omega)^2 J_T(f)) ,$$

where the last infimum is taken over all estimators $\hat\omega$ based on the observations $\{x(t)\}$ and where the supremum is taken over $\omega \in [\beta_T^{-1}, \alpha_T^{-1}]$.

The second step is to lower bound the minimax risk by a well chosen Bayes risk. Let us define a prior on f by giving a prior to each Fourier coefficient of f, concentrating around the Fourier coefficient g_k of g in the following way

$$f_k = g_k + \sigma_k\{\xi_{k,1} + i\xi_{k,2}\} \text{ , where}$$

for $p = 1, 2$, $\xi_{k,p} \sim \mathcal{N}(0, 1/2)$ are independent and independent of $\{W(t)\}$ and

$$\sigma_k{}^2 = \begin{cases} 0, & |k| \leqslant \gamma_T W_T \\ (1-\gamma_T)|s_k|^2, & |k| > \gamma_T W_T \end{cases} , \tag{19}$$

where $\gamma_T = 1/\log T$ and the s_k's are the Fourier coefficients of the function defined in (17). Then we choose as prior for ω a distribution with density π on $[\alpha_T, \beta_T]$, where π vanishes at the endpoints of the interval and such that the Fisher information $\mathrm{J}_\pi = \int \pi'(\omega)^2\pi^{-1}(\omega)dx$ is finite.

We denote by $\Psi_\sigma(f)$ the distribution associated to the preceding prior on f. Then let

$$\overline{J_T} = \int J_T(f)d\Psi_\sigma(f) = \{1 + o(1)\}\frac{T^3}{12}\sum_k (2\pi k)^2(|g_k|^2 + 2\sigma_k^2).$$

In the sequel, we shall denote by $\mathbf{E}$ the expectation in the full Bayesian model.

For the preceding choice of σ_k^2, the random function f is close to g with high probability. More precisely, one can check as in Dalalyan et al. (2006) that $\mathbf{P}(F_{\delta_T}(g)^c)$ decreases at exponential rate. Moreover, it is also not difficult to check that with this choice of σ_k^2, it holds

$$\left|12\overline{J_T}T^{-3} - \|g'\|^2\right| = o(1). \tag{20}$$

$$T^{-1}\sum_k (2\pi k)^2\mu_k = \{1 + o(1)\}r_T , \quad \text{with} \quad \mu_k = \frac{\sigma_k{}^2}{T^{-1} + \sigma_k{}^2}. \tag{21}$$

Let us lower bound the minimax risk using the full Bayesian model,

$$\begin{aligned}\inf_{\hat\omega} \sup_{\omega,\ f\in F_{\delta_T}(g)} \mathbf{E}_{\omega,f}((\hat\omega - \omega)^2 J_T(f)) \geqslant\ & \inf_{\hat\omega} \mathbf{E}((\hat\omega - \omega)^2\overline{J_T}\mathbf{1}_{F_{\delta_T}(g)}(f)) & (I)\\ & - \sup_{\hat\omega} \mathbf{E}((\hat\omega - \omega)^2(\overline{J_T} - J_T(f))\mathbf{1}_{F_{\delta_T}(g)}) & (II).\end{aligned}$$

The outline of the proof of the theorem is then the following. We bound (I) from below:

$$(I) \quad \geqslant \quad \inf_{\hat{\omega}} \overline{J_T}\mathbf{E}((\hat{\omega}-\omega)^2) - C\beta_T^2 T^3 \mathbf{P}(F_{\delta_T}(g)^c).$$

The second term in this difference is negligible thanks to the exponential bound on $\mathbf{P}(F_{\delta_T}(g)^c)$. Let us denote by $\mathcal{J}_T(\omega)$ the Fisher information associated to the observations $\{x(t)\}$ in the Bayesian model with respect to f *with fixed* ω. The Van Trees inequality (see Gill & Levit (1995)) applied to the full Bayesian model gives,

$$\inf_{\hat{\omega}} \mathbf{E}((\hat{\omega}-\omega)^2) \geqslant \frac{1}{\int \mathcal{J}_T(\omega)\pi(d\omega) + J_\pi}.$$

The key point of the proof is then to obtain an expansion of $\mathcal{J}_T(\omega)$. Note that, as the model cannot be partitioned in 1-dimensional submodels, the Fisher information $\mathcal{J}_T(\omega)$ is not a sum of Fisher information over submodels and thus has to be handled globally. As T tend to $+\infty$, with the σ_k^2's defined by (19),

$$\mathcal{J}_T(\omega) = \overline{\mathrm{J}_T} - \{1+o(1)\}\frac{T^2}{12}\sum_k (2\pi k)^2 \mu_k. \tag{22}$$

The proof of (22) is not difficult but technical. For detailed calculations, we refer the reader to Castillo (2006).

Thanks to Lemma (22), we deduce

$$\inf_{\hat{\omega}} \overline{J_T}\mathbf{E}((\hat{\omega}-\omega)^2) \quad \geqslant \quad \frac{\overline{\mathrm{J}_T}}{\overline{\mathrm{J}_T} - \{1+o(1)\}\frac{T^2}{12}\sum_k (2\pi k)^2\mu_k + J_\pi}.$$

Now using (20) and (21), we get

$$\inf_{\hat{\omega}} \overline{J_T}\mathbf{E}((\hat{\omega}-\omega)^2) \geqslant \frac{1}{1 - \{1+o(1)\}\|g'\|^{-2}(r_T - 12T^{-3}J_\pi)}.$$

Using the fact that $T^{-3} = o(r_T)$ and the inequality $(1+x)^{-1} \geqslant 1-x$, valid for $x > -1$,

$$\inf_{\hat{\omega}} \overline{J_T}\mathbf{E}((\hat{\omega}-\omega)^2) \geqslant 1 + \{1+o(1)\}\|g'\|^{-2} r_T$$

Finally, (II) is negligible with respect to (I) as in Dalalyan et al. (2006). □

Remark 4. Changing variables works well in our framework since the Fisher information (3) has a quite simple separated form in f and θ. This might not be the case for other models. For instance, in the Cox model, the form of the efficient Fisher information as given in van der Vaart (1998, p.416) is more involved and a change of variables does not seem to make the parameter of interest vanish. We think that this difficulty could be overcome by also carefully choosing a prior on the θ-parameter.

3.4 Achieving the lower bound

Theorem 3. *Assume that the conditions of Theorem 2 are fulfilled and that there is some $p > \beta \geqslant 2$ such that the sum $\sum |2\pi k|^{2p}|g_k|^2$ is finite. Then there exists a sequence of weights λ^* such that the corresponding estimator $\widehat{\theta}(\lambda^*)$ defined by (8) achieves the bound of Theorem 2. That is, as T tends to $+\infty$,*

$$\sup_{\theta\in\Theta, f\in F_{\delta_T}(g)} \mathbf{E}_{\theta,f}\left((\widehat{\theta}(\lambda^*)-\theta)^2 I_T(\theta,f)\right) = 1 + \{1+o(1)\}\|g'\|^{-2} r_T.$$

Let us comment about the rate of convergence for the second order term obtained above. Pinsker's theory states that, up to a constant which can be computed in terms of β and L, r_T is of order $T^{(2-2\beta)/(2\beta+1)}$. For example, in the case $\beta = 2$, the optimal second order term is of the order $T^{-2/5}$. This semiparametric rate is fairly slow compared to the first order rate, which is, up to a constant, $T^{-3/2}$, which highlights the importance of second order terms in semiparametric estimation. Moreover, the less regular the function, the more significative the second order term.

Proof. The idea is to take for λ^* slightly modified Pinsker weights by letting

$$\lambda_k^* = \begin{cases} 1\ , & |k| \leqslant \gamma_T W_T \\ \left[1 - (|k|/W_T)^{\beta-1}\right]_+ , & |k| > \gamma_T W_T \end{cases}, \tag{23}$$

where W_T satisfies (18). It is not difficult to check that these weights satisfy assumptions (**W**) and (**T**). Thanks to Theorem 1 it suffices to check that the supremum of $R_T(f,\lambda^*)$ for f in the considered vanishing neighborhood is indeed equivalent to r_T. There is no difficulty with respect to Dalalyan et al. (2006) thus this argument is omitted. □

4 Proof of Theorem 1

Definition 1. We say that the probability of a measurable set $\mathcal{A}$ is *negligible* if, for all integer p, we have $\mathbf{P}(\mathcal{A}(K)) = o(T^{-p})$ as T tends to $+\infty$.

4.1 Behavior of the deterministic part Γ

Let us denote by $\hat{\phi}$ the Fourier transform of the indicator function of $[-1/2,1/2]$: $\hat{\phi}(x) = \int_{-1/2}^{1/2} e^{2i\pi xt}dt = \sin(\pi x)/(\pi x)$. In the sequel, we use the following bounds on $\hat{\phi}$. For any $p \in \mathbb{N}$, there exist constants $M_p > 0$ depending only on p such that

$$|\hat{\phi}^{(p)}(u)| \leqslant M_p \text{ for } u \in [-1,1]\ ,\quad |\hat{\phi}^{(p)}(u)| \leqslant M_p/|u| \text{ for } |u| > 1/4. \tag{24}$$

$$|\hat{\phi}'(u)| \leqslant C_1|u| \text{ for all } u \in \mathbb{R}\ ,\quad |\hat{\phi}^{(3)}(u)| \leqslant C_2|u| \text{ for all } u \in \mathbb{R}. \tag{25}$$

For any real number x, let us denote by $\Delta(x)$ its distance to $\mathbb{Z}$ and by $]x[$ the (smallest) integer realizing this distance. Let us also introduce the auxiliary notation

$$a_{k,l} = \frac{T}{\theta}\left(l - \frac{k\theta}{\tau}\right)\quad ,\qquad b_{k,l} = \frac{T}{\theta}(l-k).$$

Using (2) one obtains:

$$\Gamma(\tau) = \sum_k \lambda_k T \left| \sum_l c_l \hat{\phi}\left(a_{k,l}\right) \right|^2, \tag{26}$$

$$\Gamma(\tau) = \sum_k \lambda_k T \left[\left| c_{]k\theta/\tau[} \hat{\phi}\left(\frac{T}{\theta}\Delta(k\theta/\tau)\right) + \sum_{l \neq]k\theta/\tau[} c_l \hat{\phi}\left(a_{k,l}\right) \right|^2 \right]. \tag{27}$$

For any $p, j \in \mathbb{Z}$, we denote by $p \wedge j$ the greatest common divisor of p and j.

Lemma 1. *Let $\gamma_T = T^{1/8}$ and $\varepsilon_T = T^{-3/8}$. Then, as T tends to $+\infty$,*

$$\Gamma(\tau) = o({\gamma_T}^{-1/2}T) \quad \text{if } \tau \notin \cup_{j\geqslant 1,\ 0<p\leqslant\gamma_T} B(j\theta/p, \varepsilon_T) \tag{28}$$

$$\Gamma(\tau) \leqslant \{\textstyle\sum_q \lambda_{qj}|c_{qp}|^2 + o({\gamma_T}^{-1/2})\}T \quad \text{if } \tau \in B(j\theta/p, \varepsilon_T)\ (p \leqslant \gamma_T,\ p \wedge j = 1) \tag{29}$$

$$\Gamma(\theta) = \{\textstyle\sum_{k\neq 0} |c_k|^2 + o(\log^{-1} T)\}T. \tag{30}$$

Proof. First note that without restriction we can just study the sum over $|l| \leqslant \gamma_T$ in (26). Indeed, using the Cauchy-Schwarz inequality, $\|\lambda\|_1 \leqslant N_T = o(T^{1/4})$ and (**F2**),

$$\sum_k \lambda_k T \left| \sum_{|l|>\gamma_T} c_l \hat{\phi}(a_{k,l}) \right|^2 \leqslant \|\lambda\|_1 T \sum_{|l|>\gamma_T} \frac{1}{l^4} \sum_{|l|>\gamma_T} l^4 |c_l|^2 = o\left({\gamma_T}^{-1}T\right). \tag{31}$$

In the remainder of the proof, p and j are two relatively prime integers such that $p \leqslant \gamma_T$. Note also that by symmetry we can always assume that k is positive.

Assume that τ is not in a ball $B(j\theta/p, \varepsilon_T)$ with $p \leqslant \gamma_T$ and $j \geqslant 1$. For any l such that $|l| \leqslant \gamma_T$ and any integer k this implies that $|a_{k,l}| \geqslant T\beta_T^{-1}\varepsilon_T \geqslant 1$. Thanks to (24), we have that $|\hat{\phi}(a_{k,l})|^2 \leqslant \beta_T^2 T^{-2} {\varepsilon_T}^{-2}$. Using (26) and (31), this yields $\Gamma(\tau) = o({\gamma_T}^{-1/2}T)$.

Assume that τ is in a ball $B(j\theta/p, \varepsilon_T)$ with $p \leqslant \gamma_T$ and $p \wedge j = 1$. In this case we notice that the sum over $l \neq]k\theta/\tau[$ in (27) is negligible thanks to the triangle inequality, (**F2**) and (24). Then there are two cases for the integer k in (27).

- There exists an integer q such that $k = qj$. Then for T large enough, $]k\theta/\tau[= pq$ since
$$|k\theta/\tau - pq| < pq\varepsilon_T \leqslant \gamma_T N_T \varepsilon_T = o(1).$$
- The integer k is not a multiple of j. Then $k = qj + r$ with $r < j$. Note that
$$\left| \frac{k\theta}{\tau} - \left(pq + \frac{rp}{j}\right) \right| \leqslant pq\varepsilon_T + \frac{rp\varepsilon_T}{j} \leqslant Cpq\varepsilon_T \leqslant C\frac{\gamma_T N_T \varepsilon_T}{j}.$$
Since $p \wedge j = 1$, $k\theta/\tau$ is at a distance to the integers larger than $1/2j$. Since τ must lie in Θ, we have that $j\theta/p \leqslant 2\beta_T$. Thus $j^{-1} \geqslant \theta/(2p\beta_T)$. Therefore
$$\left| \frac{T}{\theta}\Delta(k\theta/\tau) \right| \geqslant \frac{T}{4\gamma_T\beta_T},$$
which allows us to prove that the corresponding term is negligible using (24).

Finally writing (27) at $\tau = \theta$, it is easy to check, by similar arguments, that

$$\Gamma(\theta) = \sum_k \lambda_k |c_k|^2 T + o\{(\beta_T T^{-1} + N_T \beta_T^2 T^{-2})T\}.$$

Moreover thanks to (11), $\sum_{k\neq 0}(1-\lambda_k)|c_k|^2 \leqslant \sum_k (1-\lambda_k)(2\pi k)^2 |c_k|^2 = o(\log^{-1} T)$, which establishes (30). $\square$

Let us define the following weighted Fisher information.

$$\mathrm{I}(\lambda) = \frac{T^3}{12\theta^4} \sum_k \lambda_k (2\pi k)^2 |c_k|^2 \ , \quad \mathrm{I}(\lambda^2) = \frac{T^3}{12\theta^4} \sum_k \lambda_k^2 (2\pi k)^2 |c_k|^2. \tag{32}$$

Lemma 2. *As T tends to $+\infty$,*

$$\Gamma'(\theta)^2 = T^3\theta^{-4} o\left(\|\lambda'\|^2 T^{-1}\right), \quad \Gamma''(\theta) = -2\,\mathrm{I}(\lambda) + T^3\theta^{-4} o\left(\|\lambda'\|^2 T^{-1}\right).$$

Proof. For any τ in Θ , taking the derivative with respect to τ in (9),

$$\Gamma'(\tau) = \frac{T^2}{\tau^2} \sum_k \lambda_k k \sum_{p,l} \overline{c_p} c_l \left(\hat{\phi}'(a_{k,l})\hat{\phi}(a_{k,p}) + \hat{\phi}'(a_{k,p})\hat{\phi}(a_{k,l})\right). \tag{33}$$

Now consider the case $\tau = \theta$ and write the sum over p, l in (33) as

$$\sum_{p,l} = \sum_{p=k,l=k} + \left\{ \sum_{p=k,l\neq k} + \sum_{p\neq k,l=k} \right\} + \sum_{p\neq k,l\neq k} = a + b + c. \tag{34}$$

Then note that $a = 0$, $|b| \leqslant |c_k| \sum_p |c_p| \theta T^{-1}$ and $|c| \leqslant (\sum_p |c_p|)^2 \theta^2 T^{-2}$. Hence $\Gamma'(\theta) = O(T/\theta)$, which, using (**P2**) and (**W1**) gives the first expansion. The expansion for $\Gamma''(\theta)$ is obtained similarly. $\square$

Let us define the set $\mathcal{V}_{\mathcal{A}_1}$ as

$$\mathcal{V}_{\mathcal{A}_1} = \{\tau \in \Theta,\ |\tau - \theta| \leqslant D T^{-3/2} \theta^2 \log^{1/2} T\}. \tag{35}$$

Lemma 3. *As T tends to $+\infty$, uniformly in $\tau \in \mathcal{V}_{\mathcal{A}_1}$, it holds*

$$\Gamma'(\tau) = O(\theta^{-2} T^{3/2} \log^{1/2} T)\ , \quad \Gamma''(\tau) = O(\theta^{-4} T^3)\ , \quad \Gamma^{(3)}(\tau) = o(\theta^{-6} T^4).$$

Proof. Using (25) and the fact that τ lies in $\mathcal{V}_{\mathcal{A}_1}$, $|\hat{\phi}'(a_{k,k})| \leqslant C_1 |a_{k,k}| \leqslant C k T^{-1/2} \log^{1/2} T$. We write again the sum (33) as in (34). Thanks to (25) and (**F2**), the term corresponding to $\sum_{p=k=l}$ is bounded above by $C T^{3/2} \theta^{-2} \log^{1/2} T \sum \lambda_k k^2 |c_k|^2 \leqslant C T^{3/2} \theta^{-2} \log^{1/2} T$. Similarly, one concludes that the term $\sum_{p=k,l\neq k} + \sum_{p\neq k,l=k}$ is bounded above by $C(T\theta^{-1} + T^{3/2}\theta^{-2}\log^{1/2} T)$ and that the term $\sum_{p\neq k,l\neq k}$ is bounded above by $\sum \lambda_k k \leqslant N_T^2$, which with (**P2**) and (**W0**) gives the result. The two other expansions are obtained similarly. $\square$

4.2 Behavior of the stochastic parts X and Ψ.

Let $\mathbb{K}_T^\tau$ be the operator on $\mathrm{L}^2([-T/2,T/2])$ such that for any g in $\mathrm{L}^2([-T/2,T/2])$,

$$\mathbb{K}_T^\tau g\ (t) = \int_{-T/2}^{T/2} \left\{ T^{-1} \sum_k \lambda_k e^{2i\pi k(t-u)/\tau} \right\} g(u)du = \int_{-T/2}^{T/2} K_T^\tau(t-u)g(u)du.$$

Then $\mathrm{X}(\tau)$ is nothing but $2\int \mathbb{K}_T^\tau f(\cdot/\theta)\ (t)\, dW(t)$. Thus X and its derivative are centered Gaussian random variables of variance

$$\mathbf{E}(\mathrm{X}^{(p)}(\tau)^2) = 4\int_{-T/2}^{T/2} \left| [\mathbb{K}_T^\tau f(\cdot/\theta)]^{(p)}\ (t) \right|^2 dt. \tag{36}$$

On the other hand, note that for every T and τ, the operator $\mathbb{K}_T^\tau$ acting on $\mathrm{L}^2[-T/2,T/2]$ is self-adjoint and compact. Namely, it is a convolution operator with bounded kernel K_T^τ. The spectral theorem (see for example Yoshida (1978)) then states that it can be characterized by eigenvalues $\{\beta_k\}_{k\in\mathbb{Z}}$ and an orthonormal basis of eigenvectors $\{v_k\}_{k\in\mathbb{Z}}$.

Using expansions over the basis $\{v_k\}$, one can check that there exists a sequence $\{\alpha_k\}_{k\in\mathbb{Z}}$ of independent $\mathcal{N}(0,1)$ random variables such that

$$\Psi(\tau) = \int_{-T/2}^{T/2}\int_{-T/2}^{T/2} K_T^\tau(t-u)dW(u)dW(t) = \sum_k \beta_k\, \alpha_k^2. \tag{37}$$

This also holds for the derivatives of Ψ, replacing the kernel K_T^τ by its derivatives. Let us denote by $\beta_k^{(p)}$ the eigenvalues of $\mathbb{K}_T^{\tau\ (p)}$. Then it follows from (37) that, for $p \geqslant 0$,

$$\begin{aligned} \mathbf{E}[\Psi^{(p)}(\tau)] &= \sum_k \beta_k^{(p)} = \int_{-T/2}^{T/2} K_T^{\tau\ (p)}(0)dt & (38)\\ \mathbf{Var}[\Psi^{(p)}(\tau)] &= 2\sum_k \beta_k^{(p)\ 2} = 2\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} \left| K_T^{\tau\ (p)}(t-u) \right|^2 dtdu, & (39)\end{aligned}$$

where **Var** denotes the variance. These formulas allow the deviations of the process $\Psi^{(p)}$ to be controlled through the study of its Laplace transform.

Lemma 4. *There exists $C > 0$ such that, for all $\tau \in \Theta$, $\mathbf{E}(\mathrm{X}(\tau)^2) \leqslant CT$.*

Proof. Let us denote $\gamma_k = \sum_l c_l \hat{\phi}(a_{k,l})$. Then

$$\begin{aligned} \mathbf{E}(\mathrm{X}(\tau)^2) &= 4T\int_{-1/2}^{1/2} \left| \sum_k \lambda_k \sum_l c_l \hat{\phi}(a_{k,l}) \exp\{2ik\pi tT/\tau\} \right|^2 dt \\ \mathbf{E}(\mathrm{X}(\tau)^2) &= 4T\sum_k {\lambda_k}^2 |\gamma_k|^2 + 4T\sum_{k\neq k'} \lambda_k\lambda_{k'}\gamma_k\overline{\gamma_{k'}}O(\beta_T T^{-1})\ , \end{aligned}$$

where we used the fact that $\int_{-1/2}^{1/2} \exp\{2i(k-k')\pi tT/\tau\}dt = \mathbf{1}_{k=k'} + \mathbf{1}_{k\neq k'}O(\beta_T T^{-1})$ for all integers k,k'. The first term in the last sum is similar to $\Gamma(\tau)$ (see (26) with ${\lambda_k}^2$ instead of λ_k) and thus it is a $O(T)$ uniformly in τ thanks to Lemma 1. The second term is a $O(\|\lambda\|_1^2\beta_T T^{-1})$, which proves the lemma. □

Lemma 5. *For all $\tau \in \Theta$, for all positive integer p,*

$$\mathbf{E}(\mathrm{X}^{(p)}(\tau)^2) \leqslant CT^{2p+1}\tau^{-4p}\sum_k {\lambda_k}^2(2\pi k)^{2p}\left(|c_{]k\theta/\tau[}|^2 + O\left(\beta_T^2 T^{-2}\right)\right) \ ,$$
$$\mathbf{E}(\Psi^{(p)}(\tau)^2) \leqslant CT^{2p}\tau^{-4p}\|\lambda^{(p)}\|^2.$$

Proof. Denoting by $\xi_k(\tau) = \{\sum_l c_l\hat{\phi}(a_{k,l}(\tau))\}^{(p)}$, the derivative being with respect to τ ,

$$\mathbf{E}(\mathrm{X}^{(p)}(\tau)^2) = 4T\int_{-1/2}^{1/2}\left|\sum_k \lambda_k e^{2i\pi kTt/\tau}\xi_k\right|^2 dt \leqslant 4T\left[\sum_k {\lambda_k}^2|\xi_k|^2 + O(\beta_T T^{-1})\sum_{k\neq m}\lambda_k\lambda_m\xi_k\overline{\xi_m}\right].$$

Since $|\sum_{k\neq m}\lambda_k\lambda_m\xi_k\overline{\xi_m}| \leqslant (\sum_k {\lambda_k}^2|\xi_k|)^2 \leqslant N_T\sum_k {\lambda_k}^2|\xi_k|^2$ by the Cauchy-Schwarz inequality,

$$\mathbf{E}(\mathrm{X}^{(p)}(\tau)^2) \leqslant CT\sum_k {\lambda_k}^2|\xi_k|^2.$$

Now one checks that the term of higher speed coming from the p-th derivative while evaluating ξ_k is $T^p\tau^{-2p}k^p\hat{\phi}^{(p)}(a_{k,l})$. Hence,

$$|\xi_k|^2 \leqslant CT^{2p}\theta^{-4p}k^{2p}\sum_l |c_l\hat{\phi}^{(p)}(a_{k,l})|^2 \leqslant CT^{2p}\tau^{-4p}(2\pi k)^{2p}\left[|c_{]k\theta/\tau[}|^2 + O([\sum_l |c_l|\beta_T T^{-1}]^2)\right] \ ,$$

which proves the result for $\mathrm{X}^{(p)}$.

To bound $\Psi^{(p)}$, we use (39). Again, one checks that when evaluating the p-th derivative, there is one dominating term which is :

$$\mathbf{E}(\Psi^{(p)}(\tau)^2) \leqslant CT^{-2}\tau^{-4p}\sum_{k,l}\lambda_k\lambda_l(2\pi)^{2p}(kl)^p\int_{-T/2}^{T/2}\int_{-T/2}^{T/2}(t-u)^{2p}\exp\{2i\pi(t-u)(k-l)/\tau\}dudt$$
$$\leqslant CT^{-2}\tau^{-4p}\sum_k {\lambda_k}^2(2\pi k)^{2p}O(T^{2p+2}). \quad \square$$

Remark 5. In the sequel, we use Lemma 5 for neighborhoods $\mathcal{V}$ of θ such that, for any $\tau \in \mathcal{V}$, for any integer k such that $|k| \leqslant N_T$, it holds $]k\theta/\tau[= k$.

Lemma 6. *For any integer p, as T tends to $+\infty$,*

$$\mathbf{P}\left(\sup_{\tau\in\Theta}|\mathrm{X}(\tau) + \Psi(\tau)| > \log^2 T\left[T^{1/2} + \sum_k \lambda_k\right]\right) = o(T^{-p}).$$

Proof. The proof is not difficult using Lemma 14 as in Golubev (1988) and is omitted. $\square$

Lemma 7. *Let $\mathrm{I}(\lambda^2)$ be defined by (32), then as T tends to $+\infty$,*

$$\mathbf{E}(\mathrm{X}'(\theta)^2) = 4\,\mathrm{I}(\lambda^2) + T^3\theta^{-4}o(\|\lambda'\|^2T^{-1}).$$

Proof. Let us denote $y_s = \sum_l c_l(2i\pi)\hat{\phi}(b_{l,s})$ and $z_s = -\sum_l c_l\hat{\phi}'(b_{l,s})$. Then

$$\begin{aligned} T^{-3}\theta^4\mathbf{E}(\mathrm{X}'(\theta)^2) &= 4\int_{-1/2}^{1/2}\left|\sum_k \lambda_k k(ty_k - z_k)\exp\{2i\pi ktT/\theta\}\right|^2 dt \\ &= 4\sum_k \lambda_k{}^2k^2(|y_k|^2/12+|z_k|^2) + O(\theta T^{-1})\sum_{k\neq p}\lambda_k\lambda_p kp(|y_k\overline{y_p}|+|y_k\overline{z_p}|+|\overline{y_p}z_k|+|z_k\overline{z_p}|). \end{aligned}$$

Now note that $y_s = 2i\pi c_s + O(\theta T^{-1})$ and $z_s = O(\theta T^{-1})$. This yields the result using for the remainder term the fact that, by the Cauchy-Schwarz inequality and (**F2**), we have $\sum_k \lambda_k{}^2k^2|c_k| \leqslant C\|\lambda\|$ and $\sum_k \lambda_k k|c_k| \leqslant C$. □

Lemma 8. *As T tends to $+\infty$,*

$$\mathbf{E}(\Psi'(\theta)^2) = T^2\theta^{-4}\{1+o(1)\}\sum_k(2\pi k)^2\lambda_k{}^2/3.$$

Proof. Using (36) , we get

$$\begin{aligned} \mathbf{E}(\Psi'(\theta)^2) &= 8\pi^2T^2\theta^{-4}\sum_{k,p}kp\lambda_k\lambda_p\int_{-T/2}^{T/2}\int_{-T/2}^{T/2}\exp(2i\pi(t-u)(k-p)/\theta)(t-u)^2dtdu \\ \mathbf{E}(\Psi'(\theta)^2) &= T^2\theta^{-4}\{(1/3)+O(\theta^2T^{-2})\}\sum_k(2\pi k)^2\lambda_k{}^2. \end{aligned}$$

□

Lemma 9. *Recall that $\mathcal{M}_T = \max_k \lambda_k(2\pi k)$, then the quantities*

$$\mathbf{P}(\sup_{\mathcal{V}_{\mathcal{A}_1}}|\mathrm{X}''(\tau)| > T^{5/2}\log T\theta^{-4}), \qquad \mathbf{P}(\sup_{\mathcal{V}_{\mathcal{A}_1}}|\Psi''(\tau)| > T^{5/2}\log T\theta^{-4})\ ,$$
$$\mathbf{P}(\sup_{\mathcal{V}_{\mathcal{A}_1}}|\mathrm{X}^{(3)}(\tau)| > T^{7/2}\log T\mathcal{M}_T\theta^{-6}), \qquad \mathbf{P}(\sup_{\mathcal{V}_{\mathcal{A}_1}}|\Psi^{(3)}(\tau)| > T^{3}\log T\|\lambda^{(3)}\|\theta^{-6})$$

are negligible in the sense of Definition 1. Moreover,

$$\mathbf{E}(\sup_{\mathcal{V}_{\mathcal{A}_1}}|\mathrm{X}^{(3)}(\tau)|^2) = O(T^8\theta^{-12}), \qquad \mathbf{E}(\sup_{\mathcal{V}_{\mathcal{A}_1}}|\Psi^{(3)}(\tau)|^2) = O(T^8\theta^{-12}\log^{1/2}T).$$

Proof. The proof is standard using Lemmas 14 and 5 and is omitted. □

4.3 End of the proof.

Lemma 10. *Let $\mu_T = T^{1/16}$ and $\mathcal{B} = B(\theta, \theta T^{-1}\mu_T^{-1})$. Then for any $\tau \in \mathcal{B}$ it holds*

$$\Gamma(\tau)-\Gamma(\theta) = -\,(\tau-\theta)^2\,\mathrm{I}(\lambda) + o\{(\tau-\theta)^2\,\mathrm{I}(\lambda) + (\tau-\theta)\,\mathrm{I}(\lambda)^{1/2}\} \tag{40}$$
$$\mathrm{X}(\tau)-\mathrm{X}(\theta) = 2(\tau-\theta)\{1+o(1)\}\,\mathrm{I}(\lambda^2)^{1/2}\,\mathcal{N} + (\tau-\theta)\mathcal{R}_1(\tau) \tag{41}$$
$$\Psi(\tau)-\Psi(\theta) = (\tau-\theta)\mathcal{R}_2(\tau)\ , \tag{42}$$

where $\mathcal{N}$ is a $\mathcal{N}(0,1)$ random variable and for $i = 1, 2$, the process $\mathcal{R}_i$ satisfies

$$\exists C_1, C_2 > 0, \forall x \in [0, \log T], \quad \mathbf{P}(\sup_{\mathcal{B}}|\mathcal{R}_i| > x\mathrm{I}_T^{1/2}) \leqslant (1 + C_1\mu_T x)\exp(-C_2x^2).$$

Proof. To prove (40) note that

$$\Gamma(\tau)-\Gamma(\theta) = \sum_k \lambda_k T\left[\left|c_k\hat\phi(a_{k,k})+\sum_{l\neq k} c_l\hat\phi(a_{k,l})\right|^2-\left|c_k+\sum_{l\neq k}c_l\hat\phi(b_{k,l})\right|^2\right]$$
$$= \sum_k \lambda_k T|c_k|^2\left\{\left|\hat\phi(a_{k,k})\right|^2-1\right\}+\sum_k\lambda_k T\left[\left|\sum_{l\neq k}c_l\hat\phi(a_{k,l})\right|^2-\left|\sum_{l\neq k}c_l\hat\phi(b_{k,l})\right|^2\right]+\gamma$$
$$= \alpha+\beta+\gamma,$$

where γ regroups the crossed terms coming from each of the squares. But for all real u, $\hat\phi(u)=1-\pi^2u^2/6+u^3\psi(u)$ with ψ bounded on $\mathbb{R}$,

$$\alpha = -(T^3\tau^{-2}\theta^{-2}/12)\sum_k \lambda_k(2\pi k)^2|c_k|^2(\tau-\theta)^2\{1+kT(\tau^{-1}-\theta^{-1})O(1)\}.$$

Now note that, thanks to (**F1**), the rate of $\mathrm{I}(\lambda)$ is a constant times $T^3\theta^{-4}$. Using the fact that $\sum k^3|c_k|^2$ is finite and the fact that τ lies in $\mathcal{B}$, we obtain $\alpha=-\mathrm{I}(\lambda)(\tau-\theta)^2\{1+o(1)\}$.

To bound $|\beta|$, we use the inequality $||a|^2-|b|^2|\leqslant|a-b|(|a|+|b|)$ and (24),

$$\left|\sum_{l\neq k}c_l\{\hat\phi(a_{k,l})-\hat\phi(b_{k,l})\}\right|\leqslant\sum_l|c_l|\left[\sup_{[a_{k,l},b_{k,l}]}|\hat\phi'|\right]|a_{k,k}|\leqslant C(2\theta/T)|k||1-\theta/\tau|T/\theta$$
$$\sum_{l\neq k}|c_l|\left[|\hat\phi(a_{k,l})|+|\hat\phi(b_{k,l})|\right]\leqslant C(2\theta/T).$$

Thus $|\beta|$ is upper-bounded by $C\sum_k\lambda_k|k||\tau-\theta|\leqslant CN_T^2|\tau-\theta|$ which is a $o((\tau-\theta)\,\mathrm{I}(\lambda)^{1/2})$. The same holds for γ, similarly.

Note that by Taylor's expansion, there exist random reals c and d such that :

$$\mathrm{X}(\tau)-\mathrm{X}(\theta)=\mathrm{X}'(\theta)(\tau-\theta)+(\mathrm{X}'(c)-\mathrm{X}'(\theta))(\tau-\theta)$$
$$\Psi(\tau)-\Psi(\theta)=\Psi'(d)(\tau-\theta),$$

where the distribution of $\mathrm{X}'(\theta)$ is $\mathcal{N}(0,\mathbf{E}(\mathrm{X}'(\theta)^2))$ and $\mathbf{E}(\mathrm{X}'(\theta)^2)$ is given by Lemma 7.

Let us now control the deviations of $\mathrm{X}'(\tau)-\mathrm{X}'(\theta)$. Writing X' with the help of the operator $\mathbb{K}_T^\tau$ and using Taylor's formula, we can find a *deterministic* $w\in\mathcal{B}=B(\theta,\theta T^{-1}\mu_T^{-1})$ such that $\mathrm{X}'(\tau)-\mathrm{X}'(\theta)=(\tau-\theta)\,\mathrm{X}''(w)$. Then for any $\gamma>0$,

$$\mathbf{E}(\exp\{2\gamma(\mathrm{X}'(\tau)-\mathrm{X}'(\theta))\})=\mathbf{E}(\exp\{2\gamma(\tau-\theta)\,\mathrm{X}''(w)\})=\exp\{2\gamma^2(\tau-\theta)^2\mathbf{E}(\mathrm{X}''(w)^2)\}.$$

Using Lemma 5 and the definition of $\mathcal{B}$,

$$\mathbf{P}(\sup_{\tau\in\mathcal{B}}|\mathcal{R}_1(\tau)|>x\mathrm{I}_T^{1/2}) \leqslant \exp(-\gamma x\mathrm{I}_T^{1/2})\exp(C\gamma^2T^5\theta^{-8}\sup_{\tau\in\mathcal{B}}(\tau-\theta)^2)(1+\gamma CT^{5/2}\int_{\mathcal{B}}\frac{d\tau}{\tau^4})$$
$$\leqslant \exp(-\gamma x\mathrm{I}_T^{1/2})\exp(C\gamma^2\mu_T^{-2}\mathrm{I}_T)(1+C\gamma\theta^{-2}T^{-1}\mu_T^{-1})$$

Letting $\gamma=\eta\mu_T^2\mathrm{I}_T^{-1/2}x$ with η small enough, we obtain the desired bound for $\mathcal{R}_1$.

Finally we control the deviations of Ψ' on $\mathcal{B}$. For any real γ such that $\gamma^{-1} > 8\sup_p \beta_p^{(1)}$, using (37) and the inequality $-\log(1-u) \leqslant u + u^2$ for $u < 1/2$,

$$\mathbf{E}(\exp\{2\gamma\Psi'(\tau)\}) = \exp\left\{-\sum_p \log(1-4\gamma\beta_p^{(1)})/2\right\} \leqslant \exp\left\{2\gamma\sum_p \beta_p^{(1)} + 8\gamma^2\sum_p \beta_p^{(1)\,2}\right\}.$$

Then using (38), $\sum_p \beta_p^{(1)} = 0$ and using (39) and Lemma 5, $\sum_p \beta_p^{(1)\ 2} \leqslant T^2\tau^{-4}\|\lambda^{(1)}\|^2$. In particular, $\sup_p \beta_p^{(1)} \leqslant CT\theta^{-2}\|\lambda^{(1)}\|$. To conclude we apply Lemma 14 with $\gamma = \eta \mathrm{I}_T^{-1/2}x$ and η small enough. □

Lemma 11. *As T tends to $+\infty$, for any integer p, for D large enough,*

$$\mathbf{P}(|\theta^* - \theta|T^{3/2}\theta^{-2} \geqslant D\log^{1/2} T) = o(T^{-p}). \tag{43}$$

Proof. We proceed in two steps. Recall that from Section 3, we have $\theta^* \in B(\theta, \varepsilon_T)$ on $\mathcal{A}_0$, where $\mathbf{P}(\mathcal{A}_0^c)$ is negligible. We first show that $\mathbf{P}(\theta^* \in B(\theta, \theta T^{-1}\mu_T^{-1})^c)$ is negligible. Note that if V is a neighborhood of θ such that $]k\theta/\tau[= k$ for all $|k| \leqslant N_T$ and τ in V (for example $V = B(\theta, \varepsilon_T)$), we have

$$\Gamma(\tau) - \Gamma(\theta) = \sum_k \lambda_k T|c_k|^2(|\hat{\phi}(a_{k,k})|^2 - 1) + O(TN_T\frac{\beta_T}{T}).$$

Let us define $\mathcal{C} = B(\theta, \varepsilon_T) \cap B(\theta, \frac{\theta}{T\mu_T})^c$. Next we prove that $\mathbf{P}(\theta^* \in \mathcal{C})$ is negligible. Note that for $C = \pi^2/2$, $D = \pi^2$ and any $u \in \mathbb{R}$, we have $|\hat{\phi}(u)|^2 - 1 \leqslant -Cu^2(D+u^2)^{-1}$, thus thus for any $\tau \in \mathcal{C}$,

$$\begin{aligned}
\Gamma(\tau) - \Gamma(\theta) &\leqslant -T\sum_k \lambda_k|c_k|^2\frac{Ck^2}{D\mu_T^2 + k^2} + O(N_T\beta_T) \\
&\leqslant -T\mu_T^{-2}C\sum_{|k|\leqslant\mu_T} \lambda_k k^2|c_k|^2 + O(N_T\beta_T) \leqslant -CT\mu_T^{-2}\ \text{, since } \mu_T^2\beta_T N_T = o(T).
\end{aligned}$$

Recall that θ^* is defined by (8) and let us denote $\eta = \mathrm{X} + \Psi$.

$$\begin{aligned}
\mathbf{P}(\theta^* \in \mathcal{C}) &\leqslant \mathbf{P}(\theta \notin B(e_T, 1/4)) + \mathbf{P}\left(\mathrm{L}(\theta) \leqslant \sup_{\mathcal{C}} \mathrm{L}(\tau)\right) \\
\mathbf{P}(\theta^* \in \mathcal{C}) - \mathbf{P}(\mathcal{A}_0^c) &\leqslant \mathbf{P}\left(\Gamma(\theta) \leqslant \sup_{\mathcal{C}}\Gamma(\tau) + 2\sup_{\Theta}|\eta(\tau)|\right) \leqslant \mathbf{P}\left(2\sup_{\Theta}|\eta(\tau)| \geqslant CT\mu_T^{-2}\right).
\end{aligned}$$

Using Lemma 6, we conclude that $\mathbf{P}(\theta^* \in B(\theta, \frac{\theta}{T\mu_T})^c)$ is negligible.

Denoting by $\mathcal{A}' = \{\theta^* \in B(\theta, \frac{\theta}{T\mu_T})\}$ and $E_x = \{\frac{\theta^2}{2T\mu_T} > |\tau - \theta| > x\mathrm{I}_T^{-1/2}\}$ for $x > 0$,

$$\mathbf{P}(|\theta^* - \theta|\mathrm{I}_T^{1/2} > x) \leqslant \mathbf{P}\left(\sup_{|\tau-\theta|>x\mathrm{I}_T^{-1/2}} \mathrm{L}(\tau) \geqslant \mathrm{L}(\theta)\right) \leqslant \mathbf{P}\left(\sup_{E_x} \mathrm{L}(\tau) - \mathrm{L}(\theta) \geqslant 0\right) + \mathbf{P}(\mathcal{A}'^c).$$

Now write $\mathrm{L}(\tau)-\mathrm{L}(\theta)=\Gamma(\tau)-\Gamma(\theta)+\mathrm{X}(\tau)-\mathrm{X}(\theta)+\Psi(\tau)-\Psi(\theta)$. Note that for T large enough, thanks to (40), we have $\Gamma(\tau)-\Gamma(\theta)\leqslant-(\tau-\theta)^2\,\mathrm{I}(\lambda)/2$ for any $\tau\in E_x$. Then using expansions (41) and (42), we obtain

$$\begin{aligned}
&\mathbf{P}(|\theta^*-\theta|\mathrm{I}_T^{1/2}>x)\\
&\leqslant\ \mathbf{P}(\sup_{E_x}(\tau-\theta)\left[-\frac{1}{2}(\tau-\theta)\,\mathrm{I}(\lambda)+2\,\mathrm{I}(\lambda^2)^{1/2}\mathcal{N}+(\mathcal{R}_1+\mathcal{R}_2)\right]\geqslant 0)+\mathbf{P}(\mathcal{A}'^c)\\
&\leqslant\ \mathbf{P}(x\leqslant C\mathcal{N})+\mathbf{P}(x\leqslant \mathrm{I}_T^{-1/2}\sup_{E_x}|\mathcal{R}_1+\mathcal{R}_2|)+\mathbf{P}(\mathcal{A}'^c)
\end{aligned}$$

To conclude, we use the standard bound for the tail of a Gaussian random variable, Lemma 10 and we set $x=D\log^{1/2}T$ with D large enough. □

Lemma 12. *As T tends to $+\infty$,*

$$\mathbf{E}((\widehat{\tau}-\theta)^2\mathrm{I}_T)=1+\{1+o(1)\}\frac{R_T(f,\lambda)}{\|f'\|^2}.$$

Proof. The definition of $\widehat{\tau}$ implies $\mathbf{E}((\widehat{\tau}-\theta)^2\mathrm{I}_T)\,[\mathbf{E}(\mathrm{L}''(\theta))]^2=\mathbf{E}(\mathrm{L}'(\theta)^2)\mathrm{I}_T$. To compute the expectations, note that (38) is zero for $p\geqslant 1$ and remark that $\mathrm{X}'(\theta)\Psi'(\theta)$ has zero mean since it is a product of an uneven number of stochastic integrals with respect to $dW(t)$. Thus $4\,[\mathbf{E}(\mathrm{L}''(\theta))]^2=\Gamma''(\theta)^2$ and $4\mathbf{E}(\mathrm{L}'(\theta)^2)=\Gamma'(\theta)^2+\mathbf{E}(\mathrm{X}'(\theta)^2)+\mathbf{E}(\Psi'(\theta)^2)$.

Using Lemmas 7, 8 and 2, we can write the expansion of the risk for $\widehat{\tau}$:

$$\begin{aligned}
\mathbf{E}((\widehat{\tau}-\theta)^2) &= \left[\mathrm{I}(\lambda^2)+\frac{T^3}{\theta^4}o(\frac{\|\lambda'\|^2}{T})+\frac{T^2}{\theta^4}\frac{\pi^2}{3}\sum_k{\lambda_k}^2k^2\{1+o(1)\}\right]\left[\mathrm{I}(\lambda)+\frac{T^3}{\theta^4}o(\frac{\|\lambda'\|^2}{T})\right]^{-2}\\
\mathbf{E}((\widehat{\tau}-\theta)^2\mathrm{I}_T) &= \left[1+\|f'\|^{-2}\left\{\sum_k({\lambda_k}^2-1)(2\pi k)^2|c_k|^2+\frac{1}{T}\sum_k{\lambda_k}^2(2\pi k)^2+o(\frac{\|\lambda'\|^2}{T})\right\}\right]\\
&\quad\times\left[1+\|f'\|^{-2}\sum_k(\lambda_k-1)(2\pi k)^2|c_k|^2+o(\frac{\|\lambda'\|^2}{T})\right]^{-2}.
\end{aligned}$$

Thanks to (11), one can expand the denominator. Then using the expansion of $(1-\varepsilon)^{-2}$ around $\varepsilon=0$ and the fact that, thanks to (**T**) , $\sum_k(1-\lambda_k)(2\pi k)^2|c_k|^2=o(R_T(f,\lambda)^{1/2})$, one obtains

$$\begin{aligned}
\mathbf{E}((\widehat{\tau}-\theta)^2\mathrm{I}_T) &= \left[1+\|f'\|^{-2}\left\{\sum_k({\lambda_k}^2-1)(2\pi k)^2|c_k|^2+\frac{1}{T}\sum_k{\lambda_k}^2(2\pi k)^2+o(\frac{\|\lambda'\|^2}{T})\right\}\right]\\
&\quad\times\left[1-2\|f'\|^{-2}\sum_k(\lambda_k-1)(2\pi k)^2|c_k|^2+o(R_T(f,\lambda))\right]\\
\mathbf{E}((\widehat{\tau}-\theta)^2\mathrm{I}_T) &= 1+\|f'\|^{-2}\sum_k(2\pi k)^2\left((1-\lambda_k)^2|c_k|^2+\frac{1}{T}{\lambda_k}^2\right)+o(R_T(f,\lambda)).
\end{aligned}$$

□

Lemma 13. *Let R_T be the functional defined by* (12)*, then, as T tends to $+\infty$,*

$$\mathbf{E}((\theta^* - \widehat{\tau})^2 \mathrm{I}_T \, 1_{\mathcal{A}_1}) = o\left(R_T(f,\lambda)T^{-1}\right) \tag{44}$$
$$\mathbf{E}((\theta^* - \widehat{\tau})(\widehat{\tau} - \theta) \mathrm{I}_T \, 1_{\mathcal{A}_1}) = o\left(R_T(f,\lambda)T^{-1}\right) \tag{45}$$

Proof. Note that thanks to Lemma 1 combined with Lemma 6 the criterion L admits on $\mathcal{A}_0$ a local maximum inside the ball $B(e_T, 1/4)$. Thus $\mathrm{L}'(\theta^*) = 0$ on this event. Using Taylor's expansion of $\mathrm{L}(\tau)$, there exists $\omega \in [\theta, \theta^*]$ such that:

$$\mathrm{L}'(\theta^*) = 0 = \mathrm{L}'(\theta) + (\theta^* - \theta)\mathrm{L}''(\theta) + \frac{1}{2}(\theta^* - \theta)^2 \mathrm{L}^{(3)}(\omega).$$

Using (16) we deduce:

$$(\theta^* - \widehat{\tau})\mathbf{E}(\mathrm{L}''(\theta)) = (\theta - \theta^*)\{\mathrm{L}''(\theta) - \mathbf{E}(\mathrm{L}''(\theta))\} - \frac{1}{2}(\theta - \theta^*)^2 \mathrm{L}^{(3)}(\omega). \tag{46}$$

Note that using Lemmas 3 and 9,

$$\mathbf{E}\left[\sup_{\mathcal{V}_{\mathcal{A}_1}} |\mathrm{L}^{(3)}(\omega)|^2\right] \leqslant \mathbf{E}\left[\sup_{\mathcal{V}_{\mathcal{A}_1}} |\Gamma^{(3)}|^2 + |\,\mathrm{X}^{(3)}\,|^2 + |\Psi^{(3)}|^2\right] \leqslant CT^8\theta^{-12}\log^{1/2} T.$$

Note that on $\mathcal{A}_1$, $(\theta^* - \theta)^2 \leqslant D^2\theta^4 T^{-3}\log T$. Moreover, thanks to Lemma 2, $\mathbf{E}(\mathrm{L}''(\theta))^{-2} = 4\mathbf{E}(\Gamma''(\theta))^{-2} \leqslant (CT^3\theta^{-4})^{-2}$. Thanks to Lemma 5,

$$\mathbf{E}\{\mathrm{L}''(\theta) - \mathbf{E}(\mathrm{L}''(\theta))\}^2 \leqslant \mathbf{E}(\mathrm{X}''(\theta)^2) + \mathbf{E}(\Psi''(\theta)^2) \leqslant CT^5\theta^{-8}.$$

From the preceding inequalities we conclude that

$$\mathbf{E}((\theta^* - \widehat{\tau})^2 \mathrm{I}_T \mathbf{1}_{\mathcal{A}_1}) \leqslant C(\log T)^{5/2}T^{-1} \leqslant C(\log T)^{-3/2}T^{-1}\|\lambda'\|^2 = o(R_T(f,\lambda)),$$

thanks to (**W1**), which proves (44). Finally, (45) is proved by similar arguments. □

Lemma 14. *Let (X_t) be a stochastic process differentiable a.s., μ and x positive real numbers and I an interval of $\mathbb{R}$. Then it holds*

$$\mathbf{P}(\sup_{\tau \in I} \mathrm{X}_\tau > x) \leqslant \exp(-\mu x) \sup_{\tau \in I} (\mathbf{E}\exp(2\mu\,\mathrm{X}_\tau))^{1/2}\left(1 + \mu \int_{\tau \in I} (\mathbf{E}|\,\mathrm{X}'_\tau\,|^2)^{1/2} d\tau\right).$$

Proof. The proof of this lemma can be found in Golubev (1988). □

References

Bickel, P., Klaassen, C., Ritov, Y. & Wellner (1998). *Efficient and Adaptive Estimation for Semiparametric Models*, Springer, New York.

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

Castillo, I., Lévy-Leduc, C. & Matias, C. (2006). Exact adaptive estimation of the shape of a periodic function with unknown period corrupted by white noise, *Mathematical Methods of Statist.* . To appear.

Chen, Wu & Dahlhaus (2000). Hidden frequency estimation with data tapers., *Journal of time series analysis* **21**(2): 113–142.

Dalalyan, A. (2005). Stein shrinkage with penalization and second order efficiency in semiparametrics. Preprint.

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

Gassiat, E. & Lévy-Leduc, C. (2006). Efficient semi-parametric estimation of the periods in a superposition of periodic functions with unknown shape, *Journal of Time Series Analysis* . To appear.

Gill, R. D. & Levit, B. Y. (1995). Applications of the Van Trees inequality : a Bayesian Cramér-Rao bound, *Bernoulli* **1**(1-2): 59–79.

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

Golubev, G. K. & Härdle, W. (2000). Second order minimax estimation in partial linear models, *Math. Methods Statist.* **9**(2): 160–175.

Golubev, G. K. & Härdle, W. (2002). On adaptive smoothing in partial linear models, *Math. Methods Statist.* **11**(1): 98–117.

Ibragimov, I. A. & Has'minskii, R. Z. (1981). *Statistical estimation: asymptotic theory*, Springer, New York e.a.

Kimeldorf, G. S. & Wahba, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, *Ann. Math. Statist.* **41**: 495–502.

Lavielle, M. & Lévy-Leduc, C. (2005). Semiparametric estimation of the frequency of unknown periodic functions and its application to laser vibrometry, *IEEE Transactions on Signal Processing* **53**(7): 2306–2315.

Pinsker, M. S. (1980). Optimal filtration of square-integrable signals in gaussian noise., *Problems Information Transmission* **16**: 120–133.

Prenat, M. (2001). Vibration modes and laser vibrometry performance with noise, *Proceedings of the Physics in Signal and Image Processing (PSIP2001) conference, Marseille, France* .

Tsybakov, A. B. (2004). *Introduction à l'estimation non-paramétrique. (Introduction to nonparametric estimation)*, Mathématiques & Applications 41. Paris: Springer.

van der Vaart, A. W. (1998). *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. Cambridge.

Yoshida, K. (1978). *Functional Analysis.*, Die Grundlehren der mathematischen Wissenschaften, 123. Springer.