James-Stein Shrinkage and Second Order Efficiency in Semiparametrics

Arnak S. Dalalyan

James-Stein Shrinkage and Second Order Efficiency in Semiparametrics

Arnak Dalalyan

Laboratoire de Probabilités
Université Paris 6, Boîte courrier 188
75252 Paris Cedex 05, France
e-mail: dalalyan@ccr.jussieu.fr

Université Paris VI

Abstract: The problem of estimating the centre of symmetry of an unknown periodic function observed in Gaussian white noise is considered. Using the penalized blockwise James-Stein method, a smoothing filter allowing to define the penalized profile likelihood is proposed. The estimator of the centre of symmetry is then the maximizer of this penalized profile likelihood. This estimator is shown to be semiparametrically adaptive and efficient. Moreover, the second order term of its risk expansion is proved to behave at least as well as the second order term for the best possible estimator using monotone smoothing filter. Under mild assumptions, this estimator is shown to be second order minimax sharp adaptive over the whole scale of Sobolev balls with smoothness $\beta > 1$. Thus, these results improve on Dalalyan, Golubev and Tsybakov (2003), where $\beta \geq 2$ is required.

AMS 2000 subject classifications: 62G05, 62G20.

Keywords and phrases: semiparametric estimation, second order efficiency, penalized profile likelihood, exact minimax asymptotics, James-Stein shrinkage, adaptive smoothing.

1. Introduction

Extensive development of semiparametric models in recent years is in a great part explained by the compromise they offer between the flexibility of nonparametric modeling and the relative simplicity of theoretical treatment of parametric models. In a regular semiparametric model, though the ignorance of an infinite dimensional nuisance parameter, the finite-dimensional parameter of interest can be estimated as well as if the nuisance parameter were known. More precisely, there are estimators not exploiting the value of the nuisance parameter that have asymptotically the same quadratic risk as the

September 10, 2005

efficient estimators in the associated parametric model where the nuisance parameter is known.

One of the most popular general methods allowing to construct efficient estimators in a semiparametric model is perhaps the profile likelihood maximization. The asymptotic properties of this method are studied by Severini and Wong (1992) and, in a more general fashion, by Murphy and van der Vaart (2000). Profile likelihood techniques are effectively applied in a number of contexts such as laser vibrometry signals (Lavielle and Lévy-Leduc, 2004), varying coefficient partially linear models (Fan and Huang, 2005), and so forth.

If the parameter of interest is partitioned as (ϑ, f), with ϑ being a low-dimensional parameter of interest and f a higher dimensional nuisance parameter, and $l_n(\vartheta, f)$ is the log-likelihood of the model, then the profile likelihood for ϑ is defined as $pl_n(\vartheta) = \sup_{f\in\mathcal{F}} l_n(\vartheta, f)$ and the Profile Likelihood Estimator (PLE) is $\vartheta_{\rm PLE} = \arg\max_\vartheta pl_n(\vartheta)$. Thus, the nuisance parameter f is eliminated by taking the sup over all possible values of f in some a priori chosen class $\mathcal{F}$ (typically a set of two times differentiable functions with somehow bounded second order derivative).

A natural question arises: what is the best way of choosing the class $\mathcal{F}$ and what is the impact of this choice on the accuracy of the PLE. The theory fails to answer this question as long as only the first order term of the risk is considered. It appears that the most appealing way to study the dependence on $\mathcal{F}$ of the accuracy of the PLE is to consider the second order term of the quadratic risk. This approach is developed in Golubev and Härdle (2000) for partial linear models and in Dalalyan *et al.* (2003) for a nonlinear model with a shift parameter. It is shown there that the penalized PLE is second order asymptotically efficient among all possible estimators of ϑ, whenever f ranges over a ball in a Sobolev space. Analogous results for scaling parameter estimation are obtained by Castillo (2005).

These results grant an increasing importance to the second order terms in that they show that in a semiparametric estimation problem the second order term is not dramatically smaller than the first order term, especially when the nuisance parameter is not very smooth (or not very sparse). Thus, to find a procedure minimizing the second order term is not only a challenging theoretical problem, but also is of high practical interest.

Note that the study of second order term of the risk is important for other methods (one-step procedure, invariance principle, etc.) of semiparametric estimation as well. See Härdle and Tsybakov (1993), Mammen and Park

(1997), Kang *et al.* (2000) for details. Interestingly, the investigation of second order properties of semiparametric estimators contributes to the emergence of the nonparametric techniques in semiparametrics.

In Golubev and Härdle (2000), Dalalyan *et al.* (2003) an estimator of ϑ minimizing asymptotically the second order term of the risk among all possible estimators of ϑ is proposed. The parameter f is then assumed to belong to a Sobolev ball with *known* smoothness and radius. In practice, the Statistician does not know the smoothness and the radius of the nuisance parameter exactly. Therefore, he should base his strategy on the worst possible smoothness of f. The obvious drawback of this method is that the accuracy of estimating ϑ will not be better for very smooth functions f than for functions of low smoothness. This indicates that the adaptation to the unknown smoothness widely investigated in nonparametrics can be of substantial use in semiparametrics.

The main goal of the present paper is to define an estimator of ϑ which adapts automatically to the smoothness of f and is simultaneously second order efficient over a large variety of Sobolev balls. For linear models, such a procedure has been proposed by Golubev and Härdle (2002). They use the well known method of unbiased risk minimization in order to determine the data driven filter. However, their procedure is provably second order efficient only when the data driven filter is based on a relatively small part of the sample. This sample-splitting technique is frequently used in theory (cf. Bickel (1982), Pfanzagl (1990), p. 19), but it is rather unattractive from the practical point of view.

In this paper, we use the celebrated blockwise shrinkage introduced by James and Stein (1961) for estimating the mean of a multivariate normal distribution. The efficiency of this method in nonparametric estimation has been proven by Donoho and Johnstone (1995), Cai (1999), Cai and Low (2005) for wavelets, by Cavalier and Tsybakov (2001,2002) for inverse problems and by Rigollet (2004) for density estimation. The construction we use is closer to the one of Cavalier and Tsybakov (2002) in that a penalized version of the James-Stein estimator with weakly geometrically increasing blocks is considered. We believe that, unlike in nonparametric inference, in semiparametric inference the penalization of the filter is crucial for the second order efficiency.

To avoid technical difficulties, we focus here on a simple semiparametric model previously considered in Dalalyan *et al.* (2003), namely the "signal in Gaussian white noise" model with symmetric periodic signal. The shape of the signal is unknown but only the centre of symmetry is to be estimated. Our results improve on Dalalyan *et al.* (2003) in that the second order efficiency is proven for smoothness $\beta > 1$ instead of $\beta \geq 2$. A further key point is that, unlike in Golubev and Härdle (2002), our procedure is not based on the sample-splitting technique.

The paper is organized as follows. Section 2 describes the model and the Penalized Maximum Likelihood Estimator (PMLE) based on a filtering sequence h. In Section 3, the local concavity of the penalized profile likelihood is proven and the PMLE based on a data dependent choice of h is introduced. Oracle inequalities for adaptive PMLE and its second order efficiency over Sobolev balls are stated and proven in Section 4. Some numerical results are presented in Section 5. Finally, Section 6 contains the definition of a preliminary estimator and the technical details of the proofs.

2. A simple semiparametric model

Consider the "signal in Gaussian white noise model", that is the observations

$$dx^{\varepsilon}(t) = f_{\vartheta}(t)\,dt + \varepsilon dW(t), \qquad t \in [-1/2, 1/2], \tag{1}$$

are available, where $W(t)$ is a Brownian motion. Assume that the signal has the form $f_{\vartheta}(t) = f(t-\vartheta)$ where $f : \mathbb{R} \to \mathbb{R}$ is a symmetric periodic function with period 1, that is $f \in \mathcal{F}_0$ with

$$\mathcal{F}_0 = \Big\{ f \in L^2([-1/2, 1/2]) \,:\, f(x) = f(-x),\ f(x) = f(x+1),\ \forall\, x \in \mathbb{R} \Big\}.$$

The goal is to estimate the parameter $\vartheta \in \Theta \subset]-T, T]$ with $T < 1/4$. The unknown function f is considered as an infinite dimensional nuisance parameter.

For any integer $k \geq 1$, let us denote

$$x_k = \sqrt{2}\int_{-1/2}^{1/2} \cos(2\pi k t)\,dx^{\varepsilon}(t), \qquad x_k^* = \sqrt{2}\int_{-1/2}^{1/2} \sin(2\pi k t)\,dx^{\varepsilon}(t). \tag{2}$$

Clearly,

$$\begin{cases} x_k &= f_k \cos(2\pi k\vartheta) + \varepsilon\xi_k, \\ x_k^* &= f_k \sin(2\pi k\vartheta) + \varepsilon\xi_k^*, \end{cases} \tag{3}$$

where $f_k = \sqrt{2}\int_{-1/2}^{1/2} \cos(2\pi kt)\, f(t)\, dt$ and $(\xi_k, \xi_k^*, k = 1, 2, \ldots)$ are independent standard Gaussian random variables. The first order asymptotic properties of estimators in closely related models have been studied by Golubev (1990), Härdle and Marron (1990), Schick (1998,1999). This model is an idealized version of the symmetric location model (Stone (1975), Mammen and Park (1997)) and the shifted curves model (Gamboa *et al.* (2005)).

Note that, the laws of likelihood processes (indexed by $(\vartheta, f) \in \mathbb{R} \times \mathcal{F}_0$) of the models (1) and (3) coincide implying thus the equivalence of these models. The Fisher information when estimating ϑ in these models is $I^\varepsilon(f) = \varepsilon^{-2}\int_{-1/2}^{1/2} f'(x)^2\, dx = \varepsilon^{-2}\sum_{k\in\mathbb{N}}(2\pi k)^2 f_k^2$.

In this paper, we estimate the parameter ϑ by a version of the well known method of profile likelihood maximization (cf. van der Vaart (2001), p. 106), namely the penalized maximum likelihood estimator such as it is defined in Dalalyan, Golubev and Tsybakov (2003) ([DGT] hereafter). We recall briefly its definition. Write $\mathbf{P}_{\vartheta,f}$ (resp. $\mathbf{E}_{\vartheta,f}$) for probability measure (resp. expectation) induced by x^ε on the canonical space $C([-1/2, 1/2])$ equipped with the Wiener measure. As no confusion is possible, we use the same notation in the "sequence model" given by (3). The Radon density $L_\varepsilon(\tau, f, \cdot)$ of $\mathbf{P}_{\tau,f}$ with respect to $\mathbf{P}_{0,0}$ is given by

$$\begin{aligned} L_\varepsilon(\tau, f, x^\varepsilon) &= \exp\left(\varepsilon^{-2}\int_{-1/2}^{1/2} f(t-\tau)\, dx^\varepsilon(t) - \frac{\varepsilon^{-2}}{2}\int_{-1/2}^{1/2} f^2(t)\, dt\right) \\ &= \exp\left(\frac{\sqrt{2}}{\varepsilon^2}\sum_{k=1}^{\infty} f_k \int_{-1/2}^{1/2} \cos[2\pi k(t-\tau)]\, dx^\varepsilon(t) - \frac{1}{2\varepsilon^2}\sum_{k=1}^{\infty} f_k^2\right). \end{aligned}$$

Easy algebra yields

$$\max_{(f_k)_k\in\mathbb{R}^{\mathbb{N}}} L_\varepsilon(\tau, f, x^\varepsilon) = \exp\left\{\varepsilon^{-2}\sum_{k=1}^{\infty}\left(\int_{-1/2}^{1/2} \cos[2\pi k(t-\tau)]\, dx^\varepsilon(t)\right)^2\right\}. \quad (4)$$

For any $\tau \in \mathbb{R}$, this expression is equal to infinity for almost all paths x^ε. Thus, some restrictions on the nuisance parameter f are necessary. However, the profile likelihood over suitable subsets of $\mathbb{R}^{\mathbb{N}}$ cannot be written explicitly. To overcome this difficulty, we apply the profiling technique to the penalized likelihood, where the penalization corresponds to ellipsoids in ℓ^2. More details on this method can be found in [DGT], we here content ourselves with giving the final definition.

We call *filtering sequence* or *filter* any sequence $h = (h_k)_{k\in\mathbb{N}} \in [0, 1]^{\mathbb{N}}$ such that only a finite number of h_k are non-zero. Define the functional

$$\Phi_\varepsilon(\tau, h) \;=\; \sum_{k=1}^{\infty} h_k \Big(\int_{-1/2}^{1/2} \cos[2\pi k(t-\tau)]\, dx^\varepsilon(t) \Big)^2. \tag{5}$$

The PMLE of ϑ is then $\hat\vartheta_{AD} = \arg\max_\tau \Phi_\varepsilon(\tau, h)$. The role of the sequence h is thus to filter out the irrelevant terms in the right side of (4), that is to assign a value h_k close to zero to the terms corresponding to a small signal-to-noise ratio $|f_k|/\varepsilon$.

For deterministic filters h, the asymptotic behavior of the estimator $\hat\vartheta_{AD}$ has been studied in [DGT]. Under some smoothness assumptions on f, for a broad choice of filters h, the first order asymptotic efficiency of $\hat\vartheta_{AD}$ is proven. Moreover, it is shown that the second order term of its risk expansion is $\varepsilon^2 R^\varepsilon[f,h]/\|f'\|^4$, where

$$R^\varepsilon[f,h] \;=\; \sum_{k=1}^{\infty} (2\pi k)^2 \big[(1-h_k)^2 f_k^2 + \varepsilon^2 h_k^2\big].$$

This result suggests to use the filter $h_{opt} = \arg\min_h R^\varepsilon[f,h]$ for defining the PMLE of ϑ. However, this minimizer is inapplicable since it depends on f. To get rid of this dependence, the minimax approach recommends the utilization of the filter $h_{\mathcal F} = \arg\inf_h \sup_{f\in\mathcal F} R^\varepsilon[f,h]$. If $\mathcal F$ is a ball in a Sobolev space, a solution of this minimization problem is given by the Pinsker filter (cf. Pinsker (1980)). Although this filter leads to a second order minimax estimator of ϑ ([DGT], Thm. 2 and 3), it suffers from the well known drawbacks of the minimax theory: the obtained estimator is rather pessimistic and requires a quite precise information on the smoothness of the unknown function.

The aim of the present paper is to propose a data-driven filter $\hat h$ so that the resulting PMLE of ϑ circumvents these drawbacks, more precisely, it mimics well the behavior of the oracles related to some interesting and possibly vast classes of estimators (for a fixed function f) and is second order sharp adaptive over a broad scale of Sobolev balls.

3. PMLE based on a data-driven filter

3.1. Local properties of $\Phi_\varepsilon(\tau, h)$

Let us introduce some auxiliary notation:

$$y_k(\tau) = x_k(\tau) + i x_k^*(\tau) = \sqrt{2}\int_{-1/2}^{1/2} e^{2i\pi k(t-\tau)} dx^\varepsilon(t),$$
$$z_k(\tau) = \xi_k(\tau) + i \xi_k^*(\tau) = \sqrt{2}\int_{-1/2}^{1/2} e^{2i\pi k(t-\tau)} dW(t).$$

We will write $x_k = x_k(0)$, $x_k^* = x_k^*(0)$ and similarly for ξ_k and ξ_k^*. The symmetry of f implies that $x_k(\vartheta) = f_k + \varepsilon\xi_k(\vartheta)$ and $x_k^*(\vartheta) = \varepsilon\xi_k^*(\vartheta)$. Moreover, for all $\tau \in \mathbb{R}$, the random variables $\{\xi_i(\tau), \xi_i^*(\tau);\, i \in \mathbb{N}\}$ are i.i.d. standard Gaussian. Using these notation, the functional Φ_ε can be rewritten as follows

$$\Phi_\varepsilon(\tau, h) = \frac{1}{2}\sum_{k=1}^{\infty} h_k\Big(x_k(\vartheta)\cos[2\pi k(\tau-\vartheta)] + x_k^*(\vartheta)\sin[2\pi k(\tau-\vartheta)]\Big)^2.$$

Our aim is to show that under some assumptions on h, the function $\Phi_\varepsilon(\cdot, h)$ has, with a probability close to one, a local maximum in a neighborhood of ϑ. Note that the derivative of the function $\tau \mapsto \Phi_\varepsilon(\tau, h)$ is given by

$$\begin{aligned}\Phi_\varepsilon'(\tau, h) = &-\sum_{k=1}^{\infty} h_k\pi k(x_k(\vartheta)^2 - x_k^*(\vartheta)^2)\sin[4\pi k(\tau-\vartheta)]\\ &+\sum_{k=1}^{\infty} 2h_k\pi k x_k(\vartheta)x_k^*(\vartheta)\cos[4\pi k(\tau-\vartheta)].\end{aligned} \tag{6}$$

Proposition 1. *Let* $\|f^{(\beta_*)}\|^2 = \sum_k (2\pi k)^{2\beta_*} f_k^2 < \infty$ *for some* $\beta_* > 1.5$ *and set* $\hat\beta_* = \beta_* \wedge 1$, $N_\varepsilon = [(\varepsilon^2\log\varepsilon^{-5})^{-\frac{1}{2\beta_*+1}}]$. *Let* $h \in [0,1]^N$ *be a random vector depending on* (x_k, x_k^*) *only via* $x_k^2 + x_k^{*2}$. *For any* $\varepsilon < 1$, *there exists an event* $\mathcal{A}_\varepsilon$ *such that* $\mathbf{P}_{\vartheta,f}(\mathcal{A}_\varepsilon^c) \le 2\varepsilon^4$ *and on* $\mathcal{A}_\varepsilon \cap \{\sum_{k=1}^\infty (1-h_k)(2\pi k)^2 f_k^2 \le \|f'\|^2/4\}$, *for all* ε *verifying*

$$N_\varepsilon^{1-\hat\beta_*} \le \frac{\|f'\|}{2\|f^{(\beta_*)}\| + 6\pi}, \tag{7}$$

the function $\tau \mapsto \Phi_\varepsilon(\tau, h)$ *is strictly concave and admits a unique maximum* $\hat\vartheta_\varepsilon$ *in the interval* $[\vartheta - (4\pi N_\varepsilon)^{-1}, \vartheta + (4\pi N_\varepsilon)^{-1}]$, *satisfying*

$$|\hat\vartheta_\varepsilon - \vartheta| \le 8\varepsilon\sqrt{\log\varepsilon^{-5}}\;\|f'\|^{-1}. \tag{8}$$

Proof. Set $\Theta_\varepsilon = [\vartheta - (4\pi N_\varepsilon)^{-1}, \vartheta + (4\pi N_\varepsilon)^{-1}]$. Assume that (7) is fulfilled and $\sum_{k=1}^{\infty}(1-h_k)(2\pi k)^2 f_k^2 \le \|f'\|^2/4$. On the one hand, the first inequality of Proposition 4 (see Section 6.2 below) implies that, on $\mathcal{A}_\varepsilon$,

$$\Phi''_\varepsilon(\tau,h) \le -\|f'\|^2 + \sum_{k=1}^{\infty}(1-h_k)(2\pi k)^2 f_k^2 + \|f'\|^2/4 \le -\|f'\|^2/2,$$

for all $\tau \in \Theta_\varepsilon$ and for ε small enough. Therefore, $\Phi_\varepsilon(\cdot,h)$ is strictly concave. On the other hand, the second inequality of Proposition 4 implies that

$$\begin{aligned}\frac{\Phi'_\varepsilon(\tau,h)}{\tau-\vartheta} &\le -\frac{\|f'\|^2}{2} + \frac{2\varepsilon(\|f'\|\sqrt{\log\varepsilon^{-5}} + 2\pi N_\varepsilon^{1-\hat\beta_*})}{|\tau-\vartheta|}\\ &\le -\frac{\|f'\|^2}{2} + \frac{4\varepsilon\sqrt{\log\varepsilon^{-5}}\,\|f'\|}{|\tau-\vartheta|}\end{aligned} \tag{9}$$

for sufficiently small values of ε. Therefore, $\pm\Phi_\varepsilon(\vartheta\pm(4\pi N_\varepsilon)^{-1},h) < 0$, which guarantees that the maximum $\hat\vartheta_\varepsilon$ of $\Phi_\varepsilon(\cdot,h)$ is attained in the interior of Θ_ε and $\Phi'_\varepsilon(\hat\vartheta_\varepsilon,h) = 0$. Applying (9) to $\tau = \hat\vartheta_\varepsilon$ we get (8). □

Remark 1. The choice $N_\varepsilon = [(\varepsilon^2\log\varepsilon^{-5})^{-\frac{1}{2\beta_*+1}}]$ has a simple interpretation. If $f_k^2 \le \varepsilon^2$ for some $k \in \mathbb{N}$, then the kth observation in (3) is not relevant for estimating the parameter ϑ. Let $\mathcal{K} = \{k \in \mathbb{N} : f_k^2 > \varepsilon^2\}$ and $K = \#\mathcal{K}$. Then

$$\sum_{k\in\mathcal{K}}(2\pi k)^{2\beta_*} f_k^2 \ge \varepsilon^2 \sum_{k\in\mathcal{K}}(2\pi k)^{2\beta_*} \ge \varepsilon^2\sum_{k=1}^{K}(2\pi k)^{2\beta_*} \ge \frac{\varepsilon^2(2\pi)^{2\beta_*}K^{2\beta_*+1}}{2\beta_*+1}.$$

Thus, the number K of Fourier coefficients f_k larger than ε is at most $O(\varepsilon^{-\frac{2}{2\beta_*+1}})$. Thus, for ε small enough, all observations relevant for estimating ϑ lie in $\{y_1,\ldots,y_N\}$.

Remark 2. In [DGT], the estimator $\hat\vartheta_{AD}$ is defined as the maximizer of $\Phi_\varepsilon(\cdot,h)$ over the whole interval Θ. Instead, we define it as the local maximizer in the neighborhood of a preliminary estimator. This modification is explained by the fact that the function $\tau \mapsto \Phi_\varepsilon(\tau,h)$ is only locally concave when $\beta_* < 2$. Furthermore, the computation of the local minimum is faster than the computation of the global minimum.

Remark 3. Elaborating on the arguments of Proposition 1, it can be shown that Thm. 1 from [DGT] remains true for $\beta > 1$, provided that ϑ_{AD} is defined as the local maximizer of $\Phi_\varepsilon(\cdot,h)$ and condition B2 is replaced by $h_k = 0$ for any $k > N_\varepsilon$.

3.2. Blockwise constant James-Stein filter with penalization

Let J be a positive integer and $\kappa_1, \ldots, \kappa_{J+1} \in \{1, \ldots, N_\varepsilon\}$ be a strictly increasing sequence such that $\kappa_1 = 1$. Set $B_j = \{k \in \mathbb{N} : \kappa_j \leq k < \kappa_{j+1}\}$. Let $\mathcal{H}^*(B)$ be the set of all filters $h \in [0,1]^N$ that are constant on the blocks $B = \{B_j\}_{j=1}^J$:

$$h \in \mathcal{H}^*(B) \qquad \Longleftrightarrow \qquad h_k = h_{k'}, \quad \forall\, k, k' \in B_j.$$

If the function f is known, we can choose the best possible filter $h^* \in \mathcal{H}^*(B)$ by minimizing $R^\varepsilon[f,h]$ over $\mathcal{H}^*(B)$. We call h^* "oracle". Simple computations show that

$$h_k^* = \frac{\|f'\|_{(j)}^2}{\|f'\|_{(j)}^2 + \varepsilon^2\sigma_j^2}, \qquad k \in B_j, \quad j = 1, \ldots, J,$$

where $\|f'\|_{(j)}^2 = \sum_{k \in B_j} (2\pi k)^2 f_k^2$ and $\sigma_j^2 = \sum_{k \in B_j} (2\pi k)^2$.

Since the oracle h^* is unusable, we replace it by a suitable estimator. Denote $R_j^\varepsilon[f,a] = (1-a)^2\|f'\|_{(j)}^2 + \varepsilon^2 a^2 \sigma_j^2$ so that $R^\varepsilon[f,h] = \sum_j R_j^\varepsilon[f, h_{\kappa_j}] + \sum_{k > N_\varepsilon} (2\pi k)^2 f_k^2$. Then $h_{\kappa_j}^*$ is the minimizer of $R_j^\varepsilon[f,a]$ over $a \in \mathbb{R}$. For small values of $h_{\kappa_j}^*$, the minimizer of

$$\hat{R}_j^\varepsilon[a] := (1-a)^2(\|y'\|_{(j)}^2 - 2\varepsilon^2\sigma_j^2)_+ + \varepsilon^2 a^2 \sigma_j^2,$$

which is an estimator of $R_j^\varepsilon[f,a]$, can be large with respect to h_k^*. To avoid such a configuration, we penalize the large values of a and define the estimator $\hat{h}^{JS}$ of h^* as the minimum over $[0,1]$ of the function:

$$a \mapsto \hat{R}_j^\varepsilon[a] + 2\varphi_j \varepsilon^2 \sigma_j^2 a,$$

where $\varphi_j > 0$ is a factor of penalization tending to zero. This leads us to the penalized James-Stein filter

$$\hat{h}_k^{JS} = \left(1 - \frac{\varepsilon^2\sigma_j^2(1+\varphi_j)}{(\|y'\|_{(j)}^2 - 2\varepsilon^2\sigma_j^2)_+ + \varepsilon^2\sigma_j^2}\right)_+, \quad \forall k \in B_j, \tag{10}$$

with $\|y'\|_{(j)}^2 = \sum_{k \in B_j} (2\pi k)^2 |y_k|^2$ and $y_k = x_k + i x_k^*$.

3.3. Weakly geometrically increasing blocks

The aim of this section is to propose a concrete scheme for defining the blockwise constant data-driven filter. We use the weakly geometrically increasing blocks introduced by Cavalier and Tsybakov (2001,2002). These blocks have the advantage of being simple (the construction is driven by only one parameter) and of having good approximation properties with respect to the class of monotone filters (cf. Cavalier and Tsybakov (2002), Lemma 1).

Let $\nu = \nu_\varepsilon$ be a positive integer that increases as ε decreases. Set $\rho_\varepsilon = \nu_\varepsilon^{-1/3}$ and define

$$\kappa_j = \begin{cases} (1+\nu_\varepsilon)^{j-1}, & j = 1, 2, \\ \kappa_{j-1} + \lfloor \nu_\varepsilon \rho_\varepsilon (1+\rho_\varepsilon)^{j-2} \rfloor, & j = 3, 4, \ldots, \end{cases} \tag{11}$$

where $\lfloor x \rfloor$ stands for the largest integer strictly smaller than x. Let J be the smallest integer j such that $\kappa_j \geq N_\varepsilon + 1$. We redefine $\kappa_{J+1} = N_\varepsilon + 1$ and set $B_j = \{\kappa_j, \ldots, \kappa_{j+1} - 1\}$ for all $j = 1, \ldots, J$.

3.4. Brief description of the procedure

The outlined scheme can be implemented as follows.

1. Choose a real number $\beta_* > 1$ and set $N_\varepsilon = 5 \vee [(\varepsilon^2 \log \varepsilon^{-5})^{-\frac{1}{2\beta_*+1}}]$, $\nu_\varepsilon = e^{\sqrt{\log N_\varepsilon}}$ and $\rho_\varepsilon = \nu_\varepsilon^{-1/3}$.
2. Define the sequence $(\kappa_j)_j$ by (11).
3. Set $\varphi_j = \sqrt{24 \log \varepsilon^{-5}}/(\kappa_{j+1} - \kappa_j)$, $\sigma_j^2 = \sum_{\kappa_j \leq k < \kappa_{j+1}} (2\pi k)^2$ and define the data-dependent filter $\hat{h}^{JS}$ by (10).
4. Compute the preliminary estimator $\bar{\vartheta}_\varepsilon$ and set $\bar{\Theta}_\varepsilon = [\bar{\vartheta}_\varepsilon - \delta, \bar{\vartheta}_\varepsilon + \delta]$ with $\delta_\varepsilon = \varepsilon \log(\varepsilon^{-2})$.
5. Define $\hat{\vartheta}_\varepsilon$ as the minimum in $\bar{\Theta}_\varepsilon$ of $\Phi_\varepsilon(\cdot, \hat{h}^{JS})$ (see (5)).

Note that the only "free" parameter in this procedure is β_*. In practice, if no information on the regularity of f is available, it appears plausible to assume that f has Sobolev smoothness $\beta_* = 2$. This is the value we use in simulations.

Motivated by Remark 1, we recommend to replace N_ε in practice by $N_\varepsilon \wedge [3\varepsilon^{-\frac{2}{3}} \widehat{\|f'\|}^{\frac{2}{3}}/(4\pi^2)]$, where $\widehat{\|f'\|^2}$ is a rate optimal estimator of $\|f'\|^2$ (see, for example, Fan (1991), Efromovich and Samarov (2000)). Additionally, it may be useful to multiply δ_ε by the estimated standard deviation of $\bar\vartheta_\varepsilon$. These changes do not affect the rates of N_ε and δ_ε, therefore our theoretical results remain true after these modifications.

4. Main theoretical results

4.1. Comparison with the blockwise constant oracle

In this section, $\hat h^{JS}$ denotes the blockwise constant filter defined by (10), φ_j is the penalization we use on the block B_j and $\varphi_\varepsilon = \max_j \varphi_j$. We emphasize that in this section no condition on the blocks B_j is required. Let T_j be the length of the block B_j and $T_\varepsilon = \inf_j T_j$. The oracle choice of h in the class $\mathcal{H}^*(B)$ of all filters constant on the blocks $B = \{B_j\}_j$ is denoted by h^*. Define

$$\hat\vartheta_\varepsilon^{JS} = \arg\max_{\tau\in\bar\Theta_\varepsilon} \Phi_\varepsilon(\tau, \hat h^{JS}), \tag{12}$$

where $\bar\Theta_\varepsilon = [\bar\vartheta_\varepsilon - \delta_\varepsilon, \bar\vartheta_\varepsilon + \delta_\varepsilon]$ and $\bar\vartheta_\varepsilon$ is a rate optimal initial estimator of ϑ (cf. Section 6.1). Introduce the functional class

$$\mathcal{F}(\beta_*, L_*, \rho) = \Big\{ f \in \mathcal{F}_0 \to \mathbb{R} :\ f(x) = f(-x),\ \|f^{(\beta_*)}\| \le L_*,\ |f_1| \ge \rho \Big\},$$

where $\beta_* > 1$, $\rho > 0$, $L_* > 0$ are some constants.

Theorem 1. *Let $\hat\vartheta_\varepsilon^{JS}$ be defined by (12) with blocks B_j verifying* $\log\varepsilon^{-1} = o(T_\varepsilon)$ *as $\varepsilon \to 0$. If the penalty φ_j is equal to $\sqrt{24T_j^{-1}\log\varepsilon^{-5}}$, then*

$$\varepsilon^{-2}\|f'\|^2 \mathbf{E}_{\vartheta,f}[(\hat\vartheta_\varepsilon^{JS} - \vartheta)^2] \le 1 + (1+\alpha_\varepsilon)\frac{R^\varepsilon[f,h^*]}{\|f'\|^2},$$

where $\alpha_\varepsilon \to 0$ as $\varepsilon \to 0$ uniformly in $f \in \mathcal{F}(\beta_*, L_*, \rho)$.

Remark 4. If the block B_j is large, then more observations (x_k, x_k^*) are used for estimating the value of the oracle $h^*_{\kappa_j}$. Hence, it is natural to expect that α_ε decreases as T_ε increases. A thorough inspection of the proof allows to describe this feature with the help of the order relation $\alpha_\varepsilon^2 \asymp T_\varepsilon^{-1}\log\varepsilon^{-1}$.

Proof. Let us denote $\mathcal{E} = \{\hat h_k^{JS} \in [h_k^* - \sqrt{8}\varphi_{j(k)}(1-h_k^*), h_k^*],\ \forall k = 1,\dots,N_\varepsilon\}$ and $\tilde h = \hat h^{JS}\mathbb{1}_{\mathcal{E}}$, where $j(k)$ is the number of the block containing k. Lemma 5 implies that $\mathbf{P}_{\vartheta,f}(\mathcal{E}^c) \le 2\varepsilon^4$. For ε small enough, we have $(1-\tilde h_k) \le 2(1-h_k^*)$ and the inequality $\sum_k (1-\hat h_k)(2\pi k)^2 f_k^2 \le \|f'\|^2/4$ is fulfilled on $\mathcal{E}$.

By virtue of Proposition 3, for sufficiently small values of ε, the event $\mathcal{A}_0 = \Big\{|\bar\vartheta_\varepsilon - \vartheta| \le (\frac{1}{4\pi N_\varepsilon} - \delta_\varepsilon) \wedge (\delta_\varepsilon - \frac{8\varepsilon\sqrt{\log\varepsilon^{-5}}}{\|f'\|})\Big\}$ verifies $\mathbf{P}_{\vartheta,f}(\mathcal{A}_0^c) = O(\varepsilon^4)$. It can be checked that on $\mathcal{A}_0$, $\Big[\vartheta \pm \frac{8\varepsilon\sqrt{\log\varepsilon^{-5}}}{\|f'\|}\Big] \subset [\bar\vartheta_\varepsilon \pm \delta_\varepsilon] \subset [\vartheta \pm \frac{1}{4\pi N_\varepsilon}]$, where we have used the notation $[\vartheta \pm \delta] := [\vartheta - \delta, \vartheta + \delta]$.

According to Proposition 1, on the event $\mathcal{A}_1 = \mathcal{A}_0 \cap \mathcal{A}_\varepsilon \cap \mathcal{E}$, the function $\Phi_\varepsilon(\cdot, \hat h)$ is strictly concave and has a unique maximum in $[\bar\vartheta_\varepsilon \pm \delta_\varepsilon]$. Hence the estimator $\hat\vartheta_\varepsilon^{JS}$ verifies $\Phi'_\varepsilon(\hat\vartheta_\varepsilon^{JS}, \tilde h) = 0$ on the event $\mathcal{A}_1$. By Taylor's formula, there exits a point $\tilde\vartheta \in [\vartheta, \hat\vartheta_\varepsilon^{JS}]$ such that

$$0 = \Phi'_\varepsilon(\hat\vartheta_\varepsilon^{JS}, \tilde h) = \Phi'_\varepsilon(\vartheta, \tilde h) + (\hat\vartheta_\varepsilon^{JS} - \vartheta)\Phi''_\varepsilon(\vartheta, \tilde h) + \frac{(\hat\vartheta_\varepsilon^{JS} - \vartheta)^2}{2}\Phi'''_\varepsilon(\tilde\vartheta, \tilde h).$$

Therefore, on $\mathcal{A}_1$,

$$\hat\vartheta_\varepsilon^{JS} - \vartheta = -\frac{\Phi'_\varepsilon(\vartheta, \tilde h)}{\Phi''_\varepsilon(\vartheta, \tilde h) + \frac{1}{2}(\hat\vartheta_\varepsilon^{JS} - \vartheta)\Phi'''_\varepsilon(\tilde\vartheta, \tilde h)}\ .$$

Using (6), one checks that

$$\Phi''_\varepsilon(\vartheta, \tilde h) = -\sum_{k=1}^{\infty} h_k(2\pi k)^2[(f_k + \varepsilon\xi_k(\vartheta))^2 - \varepsilon^2\xi_k^*(\vartheta)^2].$$

In the sequel, we write ξ_k, ξ_k^* instead of $\xi_k(\vartheta), \xi_k^*(\vartheta)$. On the one hand, Lemmas 6, 7 (with $x^2 = \log\varepsilon^{-5}$) and 8 combined with (22) imply that, on an event $\mathcal{A}_2$ of probability higher than $1 - 4\varepsilon^4$, we have

$$\begin{aligned}
\Phi''_\varepsilon(\vartheta, \tilde h) &= -\|f'\|^2 + \sum_{k=1}^{\infty}(1 - h_k^*)(2\pi k)^2 f_k^2 - 2\varepsilon\sum_{k=1}^{N} h_k^*(2\pi k)^2 f_k\xi_k + o(R^\varepsilon)\\
&= -\|f'\|^2(1 - \|f'\|^{-2}R^\varepsilon[f, h^*] - \zeta - o(R^\varepsilon[f, h^*])),
\end{aligned}$$

where $\zeta = 2\varepsilon\|f'\|^{-2}\sum_k h_k^*(2\pi k)^2 f_k\xi_k$ is a zero mean Gaussian random variable. By virtue of (16), its variance verifies

$$4\varepsilon^2\sum_{k=1}^{N} {h_k^*}^2(2\pi k)^4 f_k^2 \le \frac{12\varepsilon^2}{\min_j T_j}\sum_{j=1}^{J} {h_{\kappa_j}^*}^2\sigma_j^2\|f'\|_{(j)}^2 \le \frac{12\|f'\|^2 R^\varepsilon[f, h^*]}{T_\varepsilon}. \quad (13)$$

Therefore, $\mathbf{E}_{\vartheta,f}[\zeta^{2p}]^{1/p} = o(R^\varepsilon[f, h^*])$ for any $p > 0$. On the other hand, in view of Lemma 9, there is an event $\mathcal{A}_3$ such that $\mathbf{P}_{\vartheta,f}(\mathcal{A}_3^c) = O(\varepsilon^4)$ and

$$(\hat\vartheta_\varepsilon^{JS} - \vartheta)\Phi'''_\varepsilon(\tilde\vartheta, \tilde h) = o(R^\varepsilon)$$

on this event. Using the inequality $(1-x)^{-2} \leq 1 + 2x + 16x^2$ for all $x \in [-1/2, 1/2]$, we get

$$\|f'\|^4(\hat{\vartheta}_\varepsilon^{JS} - \vartheta)^2 = \frac{\Phi'_\varepsilon(\vartheta, \tilde{h})^2}{\big(1 - \|f'\|^{-2}R^\varepsilon[f,h^*] - \zeta - o(R^\varepsilon)\big)^2}$$
$$\leq \Phi'_\varepsilon(\vartheta, \tilde{h})^2 + 2\Phi'_\varepsilon(\vartheta, \tilde{h})^2(\|f'\|^{-2}R^\varepsilon[f,h^*] + \zeta) + o(\varepsilon^2 R^\varepsilon),$$

on the event $\mathcal{A}_4 = \mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3$. Using Lemma 3, we infer that

$$\mathbf{E}_{\vartheta,f}[\Phi'_\varepsilon(\vartheta, \tilde{h})^2] \leq \varepsilon^2\|f'\|^2 + \varepsilon^2 \sum_{k=1}^{\infty} (2\pi k)^2[(h_k^{*2} - 1)f_k^2 + h_k^{*2}\varepsilon^2]$$
$$= \varepsilon^2(\|f'\|^2 - R^\varepsilon[f,h^*]).$$

Combining these relations with Lemmas 10 and 11, we get an event $\mathcal{A}$ such that $\mathbf{P}_{\vartheta,f}(\mathcal{A}^c) = O(\varepsilon^4)$ and

$$\|f'\|^4\mathbf{E}_{\vartheta,f}[(\hat{\vartheta}_\varepsilon^{JS} - \vartheta)^2\mathbb{1}_{\mathcal{A}}] \leq \mathbf{E}_{\vartheta,f}[\Phi'_\varepsilon(\vartheta,\tilde{h})^2]\left(1 + \frac{2R^\varepsilon[f,h^*]}{\|f'\|^2}\right) + o(\varepsilon^2 R^\varepsilon[f,h^*])$$
$$\leq \varepsilon^2\|f'\|^2 + \varepsilon^2 R^\varepsilon[f,h^*] + o(\varepsilon^2 R^\varepsilon[f,h^*]).$$

Since $|\hat{\vartheta}_\varepsilon^{JS} - \vartheta| \leq 1$, we have $\mathbf{E}_{\vartheta,f}[(\hat{\vartheta}_\varepsilon^{JS} - \vartheta)^2\mathbb{1}_{\mathcal{A}^c}] \leq \mathbf{P}_{\vartheta,f}(\mathcal{A}^c) = O(\varepsilon^4)$. In view of (23), $\varepsilon^2 = o(R^\varepsilon[f,h^*])$. Therefore $\mathbf{E}_{\vartheta,f}[(\hat{\vartheta}_\varepsilon^{JS} - \vartheta)^2\mathbb{1}_{\mathcal{A}^c}] = O(\varepsilon^4) = o(\varepsilon^2 R^\varepsilon[f,h^*])$ and the assertion of the theorem follows. $\square$

4.2. Comparison with the monotone oracle

Now we consider the class $\mathcal{H}_{\text{mon}}$ of filters having decreasing components, that is

$$\mathcal{H}_{\text{mon}} = \Big\{h \in [0,1]^N \ :\ h_k \geq h_{k+1},\ 1 \leq k \leq N_\varepsilon - 1\Big\}.$$

The class $\mathcal{H}_{\text{mon}}$ is of high interest in statistics because it contains the most common filters such as the projection filter, the Pinsker filter, the Tikhonov or smoothing spline filter and so forth. See Efromovich (1997), Ch. 7 for a comprehensive overview on adaptive filtering.

Proposition 2. *Set* $\gamma_\varepsilon = \max_{1\leq j\leq J-1}(\sigma_{j+1}^2/\sigma_j^2)$. *Then*

$$R^\varepsilon[f,h^*] \leq \gamma_\varepsilon \inf_{h\in\mathcal{H}_{\text{mon}}} R^\varepsilon[f,h] + \varepsilon^2\sigma_1^2.$$

A more general version of this result is Lemma 1 in Cavalier and Tsybakov (2002). Since the proof in our setting is simple, we give it below.

Proof. Let h be a filter from $\mathcal{H}_{\rm mon}$. Define $\bar h$ by $\bar h_k = h_{\kappa_j}$ if $k \in B_j$ for some j and $\bar h_k = 0$ if $k > N_\varepsilon$. Since the components of h are decreasing, we have $1 - \bar h_k \le 1 - h_k$ and therefore

$$R^\varepsilon[f, \bar h] \le \sum_{k=1}^{N} (1-h_k)^2 (2\pi k)^2 f_k^2 + \varepsilon^2 \sum_{j=1}^{J} h_{\kappa_j}^2 \sigma_j^2. \tag{14}$$

Again by monotonicity of h, we have $h_{\kappa_j} \le h_k$ for all $k \in B_{j-1}$. Hence,

$$\sum_{j=2}^{J} h_{\kappa_j}^2 \sigma_j^2 \le \gamma_\varepsilon \sum_{j=2}^{J} h_{\kappa_j}^2 \sigma_{j-1}^2 \le \gamma_\varepsilon \sum_{k=1}^{N} h_k^2 (2\pi k)^2.$$

Combining this inequality with (14) and bounding h_1 by 1, we get $R^\varepsilon[f, \bar h] \le \gamma_\varepsilon R^\varepsilon[f,h] + \sigma_1^2$. Since $\bar h \in \mathcal{H}^*$ and h^* minimizes $R^\varepsilon[f,h]$ over all $h \in \mathcal{H}^*$, we have $R^\varepsilon[f,h^*] \le \gamma_\varepsilon R^\varepsilon[f,h] + \varepsilon^2\sigma_1^2$. This inequality holds for every $h \in \mathcal{H}_{\rm mon}$, therefore the assertion of the proposition follows. □

Combining this proposition with Theorem 1 we get the following result.

Corollary 1. Assume that the conditions of Theorem 1 are fulfilled, then

$$\varepsilon^{-2}\|f'\|^2 \mathbf{E}_{\vartheta,f}[(\hat\vartheta_\varepsilon^{JS} - \vartheta)^2] \le 1 + \gamma_\varepsilon (1+\alpha_\varepsilon) \frac{\min_{h\in\mathcal{H}_{\rm mon}} R^\varepsilon[f,h]}{\|f'\|^2},$$

where $\alpha_\varepsilon \to 0$ as $\varepsilon \to 0$ uniformly in $f \in \mathcal{F}(\beta_*, L_*, \rho)$.

Remark 5. For the blocks defined by (11), we have $T_\varepsilon = \nu_\varepsilon\rho_\varepsilon(1+\rho_\varepsilon)$, $\sigma_1^2 \le 4\pi^2\nu_\varepsilon^3$ and $-\nu_\varepsilon\rho_\varepsilon + \nu_\varepsilon(1+\rho_\varepsilon)^j \le \kappa_{j+1} \le 1 + \nu_\varepsilon(1+\rho_\varepsilon)^j$. One also checks that $\gamma_\varepsilon = \max_j \sigma_{j+1}^2/\sigma_j^2$ is asymptotically equivalent to $(1+\rho_\varepsilon)^3 \sim 1 + 3\rho_\varepsilon$ as $\varepsilon \to 0$. Therefore the factor in the oracle inequality of Corollary 1 is of order $(1 + 3\rho_\varepsilon + \alpha_\varepsilon)$. We have already mentioned that $\alpha_\varepsilon^2 = O(T_\varepsilon^{-1}\log\varepsilon^{-1})$. The trade-off between α_ε and ρ_ε leads us to $\rho_\varepsilon \asymp \nu_\varepsilon^{-1/3}$. This clarifies our choice of ρ_ε slightly differing from the one of Cavalier and Tsybakov (2002).

4.3. Second order minimax sharp adaptation

To complete the theoretical analysis, we show below that the estimator $\hat\vartheta_\varepsilon^{JS}$ corresponding to the blocks (11) enjoys minimax properties over a large scale of Sobolev balls. Assume that $\bar f \in \mathcal{F}(\beta^*, L^*, \rho)$ and define

$$\mathcal{F}_{\delta,\beta,L}(\bar f) = \Big\{ f = \bar f + v \,:\, \|v\| \le \delta,\ \|v^{(\beta)}\| \le L \Big\}.$$

Theorem 2. *Assume that ν_ε verifies $\varepsilon^{\frac{2}{2\beta+1}}\nu_\varepsilon \to 0$ as $\varepsilon \to 0$ and the conditions of Theorem 1 are fulfilled. If δ is small enough and $\bar f \in \mathcal{F}(\beta^*, L^*, \rho)$ with $\beta^* > \beta \ge \beta_*$, then the estimator $\hat\vartheta_\varepsilon^{JS}$ defined in Section 3.4 satisfies*

$$\sup_{\vartheta\in\Theta, f\in F_{\delta,\beta,L}(\bar f)} \varepsilon^{-2}\|f'\|^2 \mathbf{E}_{\vartheta,f}[(\hat\vartheta_\varepsilon^{JS} - \vartheta)^2] \le 1 + (1+o(1))\frac{C(\beta,L)\varepsilon^{\frac{4\beta-4}{2\beta+1}}}{\|\bar f'\|^2},$$

when $\varepsilon \to 0$, with $C(\beta,L) = \frac{1}{3}\big(\frac{\beta-1}{2\pi(\beta+2)}\big)^{\frac{2\beta-2}{2\beta+1}}(L(2\beta+1))^{\frac{3}{2\beta+1}}$. Moreover, the following lower bound holds:

$$\inf_{\tilde\vartheta_\varepsilon} \sup_{\vartheta\in\Theta, f\in F_{\delta,\beta,L}(\bar f)} \varepsilon^{-2}\|f'\|^2 \mathbf{E}_{\vartheta,f}[(\tilde\vartheta_\varepsilon - \vartheta)^2] \ge 1 + (1+o(1))\frac{C(\beta,L)\varepsilon^{\frac{4\beta-4}{2\beta+1}}}{\|\bar f'\|^2},$$

where the inf *is taken over all possible estimators $\tilde\vartheta_\varepsilon$.*

Proof. According to Lemma 12, there exists a filter $\lambda^* \in \mathcal{H}_{\text{mon}}$ such that $\sup_{f\in F_{\delta,\beta,L}(\bar f)} R^\varepsilon[f,\lambda^*]/\|f'\|^2 \le (1+o(1))C(\beta,L)\varepsilon^{\frac{4\beta-4}{2\beta+1}}\|\bar f'\|^{-2}$. Combining this result with Corollary 1, we get the first inequality. The second inequality is Theorem 2 of [DGT]. Although the latter is stated for $\beta \ge 2$, a thorough inspection of its proof shows that the same claim is true for any $\beta > 1$. □

5. Numerical results

We now assess the finite-sample performance of the estimator $\hat\vartheta_\varepsilon^{JS}$ via Monte Carlo simulations. The main aim is to illustrate that (a) the second order efficiency is important even for moderate sample sizes, (b) the penalization used in the James-Stein filter improves the quality of estimation.

Example 1. The first example deals with a very smooth function. Set

$$g(t) = \sin[1 + \cos(2\pi t) - \cos(4\pi t) - \cos(6\pi t)] + \sin^2(2\pi t),$$

$f(t) = 10g(t(1-t))$ and $\vartheta = 0$. The curve of this function is given in Figure 1. We draw 50000 random samples of size $n = \varepsilon^{-2} = 2^k \cdot 50$ for $k = 1, \ldots, 14$. Table 1 contains the Means of the Squared Errors (MSE) of

Table 1: MSE with 50000 replications for Example 1.

n	$\bar{\vartheta}$	ϑ^{Pr}	$\hat{\vartheta}^{JS}$	$\tilde{\vartheta}^{JS}$
100	7.6686	1.0161	1.0200	1.0195
200	7.6586	1.0310	1.0242	1.0243
400	7.6892	1.0217	1.0070	1.0096
800	7.6911	1.0276	1.0083	1.0111
1600	7.7292	1.0348	1.0120	1.0146
3200	7.6115	1.0251	1.0036	1.0057
6400	7.7707	1.0307	1.0097	1.0112
12800	7.7878	1.0276	1.0068	1.0083

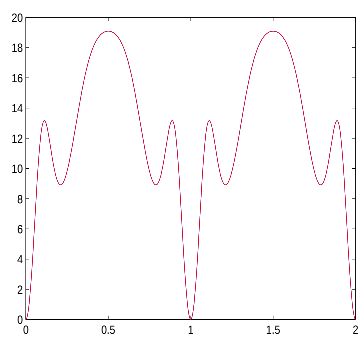

Figure 1. The function of Ex. 1.

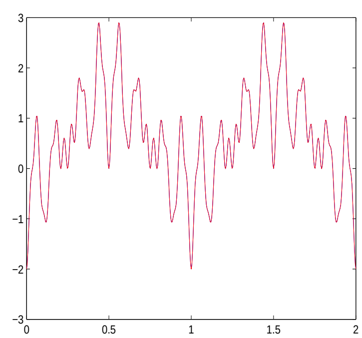

Figure 2. The function of Ex. 2.

different estimators normalized by the Fisher information $\varepsilon^{-2}\|f'\|^2$. Here $\bar{\vartheta}$ is the preliminary estimator of Section 6.1, $\hat{\vartheta}^{JS}$ is the estimator we recommend (Section 3.4), $\tilde{\vartheta}^{JS}$ is the same estimator without penalization ($\varphi_j = 0$) and ϑ^{Pr} is the PMLE based on the projection filter $h_k = \mathbb{1}(k \leq 2\nu)$. Note that ϑ^{Pr} is asymptotically efficient according to [DGT], Thm. 1.

Example 2. The second example deals with a less regular function (cf. Figure 2). It is given by $f(t) = 2\sin^2(\pi t) - \cos^3(16\pi t) - \cos^4(10\pi t)$. The true value of the parameter ϑ is 0. The number of replications is 50000. We see from Table 2 that in this example, as in the previous one, the estimator $\hat{\vartheta}^{JS}$ almost always outperforms the other procedures.

Table 2: MSE with 50000 replications for Example 2.

n	$\bar{\vartheta}$	ϑ^{Pr}	$\hat{\vartheta}^{JS}$	$\tilde{\vartheta}^{JS}$
200	105.2159	31.4007	31.4208	31.4696
400	104.5065	3.8246	3.8245	3.8221
800	105.0829	1.6813	1.6581	1.6597
1600	104.7602	1.0514	1.0443	1.0436
3200	103.6639	1.0280	1.0324	1.0295
6400	103.3527	1.0282	1.0106	1.0126
12800	103.9512	1.0204	1.0011	1.0028
25600	104.8528	1.0268	1.0009	1.0029
51200	104.2793	1.0339	1.0071	1.0092
102400	104.9999	1.0268	1.0027	1.0040

The MatLab code of this procedure is available on request. It is also downloadable from `http://proba.jussieu.fr/pageperso/dalalyan/Eng/research.html`.

6. Preliminary estimator and technical lemmas

6.1. Preliminary estimator

Having the observation $(x^\varepsilon(t), |t| \le 1/2)$, we can compute x_1, x_1^* by (2). Then we have $x_1 = f_1 \cos(2\pi\vartheta) + \varepsilon\xi_1$ and $x_1^* = f_1 \sin(2\pi\vartheta) + \varepsilon\xi_1^*$, where ξ_1, ξ_1^* are independent standard Gaussian random variables. We define

$$\bar{\vartheta}_\varepsilon = \frac{1}{2\pi}\ \arctan\left(\frac{x_1^*}{x_1}\right),$$

if $x_1 \ne 0$ and $\bar{\vartheta}_\varepsilon = 1/4$ if $x_1 = 0$. One easily checks that $\bar{\vartheta}_\varepsilon$ is the maximum likelihood estimator in the model induced by observations (x_1, x_1^*). The following result describes its asymptotic behavior.

Proposition 3. *If ε is sufficiently small, then*

$$\sup_{|\vartheta|\le T} \mathbf{P}_{\vartheta,f}\big(|\bar{\vartheta}_\varepsilon - \vartheta| \ge x\big) \le \exp\left(-\frac{x^2 f_1^2 \cos^2(2\pi T)}{2\pi^2\varepsilon^2}\right),$$

for all $x \in [0, 1/2]$ and for all $T < 1/4$.

Proof. Let us introduce $X=\sqrt{\xi_1^2+\xi_1^{*2}}$. One checks that $\sin[2\pi(\bar\vartheta_\varepsilon-\vartheta)]=\varepsilon(\xi_1\sin(2\pi\bar\vartheta_\varepsilon)-\xi_1^*\cos(2\pi\bar\vartheta_\varepsilon))f_1^{-1}$. The Cauchy-Schwarz inequality yields $|\sin[2\pi(\bar\vartheta_\varepsilon-\vartheta)]|\le\varepsilon X|f_1^{-1}|$. Since $\vartheta\in[-1/4,1/4]$ and $\bar\vartheta_\varepsilon\in[-T,T]$, we have $\frac{\sin[2\pi(\bar\vartheta_\varepsilon-\vartheta)]}{2\pi(\bar\vartheta_\varepsilon-\vartheta)}\ge\frac{\sin(2\pi(T+1/4))}{4\pi T}\ge\frac{\cos(2\pi T)}{\pi}$. Therefore,

$$\begin{aligned}\mathbf{P}_{\vartheta,f}\big(|\bar\vartheta_\varepsilon-\vartheta|\ge x\big)&\le\mathbf{P}_{\vartheta,f}\bigg(\frac{\pi|\sin[2\pi(\bar\vartheta_\varepsilon-\vartheta)]|}{\cos(2\pi T)}\ge x\bigg)\\&\le\mathbf{P}_{\vartheta,f}\big(X\ge x(\pi\varepsilon)^{-1}|f_1|\cos(2\pi T)\big),\end{aligned}$$

and the fact that $X^2/2$ follows the law $\mathscr{E}(1)$ completes the proof. ☐

6.2. Proofs of Lemmas used in Proposition 1

Let us start with some basic facts that will be often used in the proofs. For any $n,m,p\in\mathbb{N}$, we have

$$n^p(n-m)\ge\sum_{k=m+1}^{n}k^p\ge\frac{n^p(n-m)}{p+1}.\tag{15}$$

Applying this inequality to $p=2$, we get

$$\max_{k\in B_j}(2\pi k)^2\le 3\sigma_j^2/T_j.\tag{16}$$

Assume now that ξ is a random variable of law $\mathcal{N}(0,1)$. For any $\sigma^2\le1/4$, we have $(1-2\sigma^2)^{-1}\le2$ and $(1-2\sigma^2)^{-1/2}\le e^{2\sigma^2}$, therefore

$$\mathbf{E}[e^{(\mu+\sigma\xi)^2}]=\frac{e^{\frac{\mu^2}{(1-2\sigma^2)}}}{\sqrt{1-2\sigma^2}}\le e^{\frac{\mu^2}{(1-2\sigma^2)}+\sigma^2}\le\exp\big(2\mu^2+2\sigma^2\big).\tag{17}$$

Using the more precise inequalities $(1-2\sigma^2)^{-1}\le1+4\sigma^2$ and $\log(1-2\sigma^2)^{-1}\le2\sigma^2+4\sigma^4$, we get $\mathbf{E}[e^{(\mu+\sigma\xi)^2}]\le\exp(\mu^2+\sigma^2+2\sigma^2(2\mu^2+\sigma^2))$ or equivalently,

$$\mathbf{E}[e^{2\mu\sigma\xi+\sigma^2(\xi^2-1)}]\le e^{2\sigma^2(2\mu^2+\sigma^2)}.\tag{18}$$

Throughout this section, we assume that $\|f^{(\beta_*)}\|^2=\sum_k(2\pi k)^{2\beta_*}f_k^2<\infty$ for some $\beta_*>1$, $N_\varepsilon=[(\varepsilon^2\log\varepsilon^{-5})^{-\frac{1}{2\beta_*+1}}]$ and $h\in[0,1]^{N_\varepsilon}$ is a random vector depending on (x_k,x_k^*) only via $x_k^2+x_k^{*2}$. Without loss of generality, we give the proofs in the case $\vartheta=0$.

Lemma 1. *Set* $\hat\beta_* = \beta_* \wedge 1.5$. *For all* τ *such that* $4\pi|\tau-\vartheta| \le N_\varepsilon^{-1}$,

$$\Big|\Phi''_\varepsilon(\tau,h) + \sum_{k=1}^N (2\pi k)^2 h_k f_k^2\Big| \le N_\varepsilon^{2-2\hat\beta_*}\Big(\|f^{(\beta_*)}\| + 2\pi\hat X/\sqrt{\log\varepsilon^{-5}}\,\Big)^2,$$

where $\hat X = \max_{1\le k\le N_\varepsilon}\sqrt{\xi_k^2+\xi_k^{*2}}$.

Proof. One easily checks that

$$\begin{aligned}\Phi''_\varepsilon(\vartheta+\tau,h) = &-\sum_{k=1}^N (2\pi k)^2 h_k f_k^2 \cos[4\pi k\tau]\\ &-2\varepsilon\sum_{k=1}^N (2\pi k)^2 h_k f_k(\xi_k\cos[4\pi k\tau] + \xi_k^*\sin[4\pi k\tau])\\ &-\varepsilon^2\sum_{k=1}^N (2\pi k)^2 h_k\big[(\xi_k^2-\xi_k^{*2})\cos[4\pi k\tau] + 2\xi_k^*\xi_k\sin[4\pi k\tau]\big].\end{aligned}$$

On the one hand, thanks to inequality $|1-\cos x| \le |x|$,

$$\begin{aligned}\Big|\sum_{k=1}^N (2\pi k)^2 h_k f_k^2(1-\cos[4\pi k\tau])\Big| &\le 2\tau\sum_{k=1}^N (2\pi k)^3 f_k^2\\ &\le 2\tau(2\pi N_\varepsilon)^{3-2\hat\beta_*}\|f^{(\beta_*)}\|^2\\ &\le N_\varepsilon^{2-2\hat\beta_*}\|f^{(\beta_*)}\|^2.\end{aligned}$$

On the other hand, the Cauchy-Schwarz inequality yields $\xi_k\cos[4\pi k\tau] + \xi_k^*\sin[4\pi k\tau] \le \hat X$ and $(\xi_k^2-\xi_k^{*2})\cos[4\pi k\tau]+2\xi_k^*\xi_k\sin[4\pi k\tau] \le \hat X^2$. Therefore, it holds

$$\begin{aligned}\Big|\sum_{k=1}^N (2\pi k)^2 h_k f_k(\xi_k\cos[4\pi k\tau] + \xi_k^*\sin[4\pi k\tau])\Big| &\le \hat X\sum_{k=1}^N (2\pi k)^2|f_k|\\ &\le \hat X\|f^{(\beta_*)}\|\sqrt{\sum_{k=1}^N (2\pi k)^{4-2\beta_*}} \le 2\pi\hat X N_\varepsilon^{\frac52-\hat\beta_*}\|f^{(\beta_*)}\|,\end{aligned}$$

and

$$\Big|\sum_{k=1}^N (2\pi k)^2 h_k\big[(\xi_k^2-\xi_k^{*2})\cos[4\pi k\tau] + 2\xi_k^*\xi_k\sin[4\pi k\tau]\big]\Big| \le 4\pi^2 N_\varepsilon^3\hat X^2.$$

Taking into account the identity $\varepsilon^2 N_\varepsilon^3 = N_\varepsilon^{2-2\beta_*}/\log(\varepsilon^{-5})$, for all τ verifying $|\tau| \leq (4\pi N_\varepsilon)^{-1}$, we get

$$\left|\Phi_\varepsilon''(\vartheta+\tau,h) + \sum_{k=1}^{N} h_k(2\pi k)^2 f_k^2\right| \leq N_\varepsilon^{2-2\hat\beta_*}(\|f^{(\beta_*)}\|^2 + 2\pi\hat X/\sqrt{\log\varepsilon^{-5}})^2$$

and the assertion of the lemma follows. □

Lemma 2. *Let $h \in [0,1]^N$ be a random vector depending on (x_k, x_k^*) only via $x_k^2 + x_k^{*2}$. For any $x \in [0, \sqrt{N_\varepsilon/6}]$, it holds*

$$\mathbf{P}_{\vartheta,f}\Big(|\Phi_\varepsilon'(\vartheta,h)| > 2x\varepsilon(\|f'\| + 2\pi\varepsilon N_\varepsilon^{3/2})\Big) \leq 2e^{-x^2}$$

Proof. The random variables $X_k = (2\pi k)(f_k + \varepsilon\xi_k)\xi_k^*$, $k = 1, \ldots, N_\varepsilon$ and $h_1, \ldots, h_N$ fulfill the conditions of Lemma 14 with $\varrho_k^* = 1$, $T_k = 1/(2\sqrt{2}\pi k\varepsilon)$, $g_k^2 = (2\pi k)^2(f_k^2 + \varepsilon^2)$, since due to (17),

$$\mathbf{E}[e^{tX_k}] = \mathbf{E}[e^{(2\pi k t)^2(f_k+\varepsilon\xi_k)^2/2}] \leq e^{(2\pi k t)^2(f_k^2+\varepsilon^2)}. \tag{19}$$

By definition, $\Phi_\varepsilon'(\vartheta,h) = \varepsilon\sum_{k=1}^{N_\varepsilon} h_k X_k$, and therefore,

$$\mathbf{P}_{\vartheta,f}\Big(|\Phi_\varepsilon'(\vartheta,h)| \geq 2x\varepsilon\Big(\sum_{k=1}^{N}(2\pi k)^2(f_k^2+\varepsilon^2)\Big)^{1/2}\Big) \leq 2e^{-x^2}$$

for all $x \in [0, (\sum_k (2\pi k)^2(f_k^2+\varepsilon^2))^{1/2}/(2\sqrt{2}\pi N_\varepsilon\varepsilon)]$. To complete the proof, it suffices to remark that

$$\frac{\sum_{k=1}^{N}(2\pi k)^2(f_k^2+\varepsilon^2)}{8\pi^2 N_\varepsilon^2\varepsilon^2} \geq \frac{\sum_{k=1}^{N} k^2}{2N_\varepsilon^2} \geq \frac{N_\varepsilon}{6}$$

and $\sum_{k=1}^{N}(2\pi k)^2(f_k^2+\varepsilon^2) \leq (\|f'\| + \varepsilon(2\pi)N_\varepsilon^{3/2})^2$. □

Proposition 4. *Assume that $\|f^{(\beta_*}\| < \infty$ for some $\beta_* > 1$ and set $\hat\beta_* = \beta_* \wedge 1.5$. There exists an event $\mathcal{A}_\varepsilon$ such that for every $\varepsilon < 1/2$, $\mathbf{P}_{\vartheta,f}(\mathcal{A}_\varepsilon) \geq 1 - 2\varepsilon^4$ and on $\mathcal{A}_\varepsilon$ it holds:*

$$\Phi_\varepsilon''(\tau,h) \leq -\sum_{k=1}^{N} h_k(2\pi k)^2 f_k^2 + N_\varepsilon^{2-2\hat\beta_*}(\|f^{(\beta_*)}\| + 3\pi)^2 \tag{20}$$

$$\begin{aligned}\frac{\Phi_\varepsilon'(\tau,h)}{\tau-\vartheta} \leq &-\sum_{k=1}^{N} h_k(2\pi k)^2 f_k^2 + N_\varepsilon^{2-2\hat\beta_*}(\|f^{(\beta_*)}\| + 3\pi)^2\\ &+ \frac{2\varepsilon(\|f'\|\sqrt{\log\varepsilon^{-5}} + 2\pi N_\varepsilon^{1-\beta_*})}{|\tau-\vartheta|}.\end{aligned} \tag{21}$$

for all $\tau \in [\vartheta - (4\pi N_\varepsilon)^{-1}, \vartheta + (4\pi N_\varepsilon)^{-1}]$.

Proof. According to Lemma 1, we have

$$\Phi''_\varepsilon(\tau,h) \le -\sum_{k=1}^{N} h_k(2\pi k)^2 f_k^2 + N_\varepsilon^{2-2\hat\beta_*}\big(\|f^{(\beta_*)}\| + 2\pi\hat X/\sqrt{\log\varepsilon^{-5}}\,\big)^2.$$

Since for every k, $X_k^2/2 = (\xi_k^2 + \xi_k^{*2})/2$ follows the exponential law with mean 1, we have

$$\mathbf{P}\big(4\hat X^2 > 9\log\varepsilon^{-5}\big) \le N_\varepsilon \mathbf{P}\big(X_1^2/2 > \log\varepsilon^{-5}\big) \le N_\varepsilon\varepsilon^5 \le \varepsilon^4.$$

This inequality completes the proof of (20).

To prove (21), note that for some $\tilde\tau \in [\vartheta,\tau]$, we have $\Phi'_\varepsilon(\tau,h) = \Phi'_\varepsilon(\vartheta,h) + (\tau-\vartheta)\Phi''_\varepsilon(\tilde\tau,h)$. Lemma 2 and (20) yield (21). □

6.3. Lemmas used in Theorem 1

Let us start with some simple algebra allowing to obtain a rough evaluation of $R^\varepsilon[f,h^*]$, where h^* is the ideal filter an oracle would choose in the class of blockwise constant filters. For this filter h^*, it holds

$$R^\varepsilon[f,h^*] = \sum_{j=1}^{J} \frac{\varepsilon^2\sigma_j^2\|f'\|_{(j)}^2}{\|f'\|_{(j)}^2 + \varepsilon^2\sigma_j^2} + \sum_{k>N_\varepsilon}(2\pi k)^2 f_k^2.$$

Using the explicit form of h_k^*, we get

$$R^\varepsilon[f,h^*] = \sum_{k=1}^{\infty}(1-h_k^*)(2\pi k)^2 f_k^2 \ge \varepsilon^2\sum_{k=1}^{N} h_k^*(2\pi k)^2. \tag{22}$$

Since $\varepsilon^2 N_\varepsilon^3 \to 0$ as $\varepsilon\to 0$, we have $R^\varepsilon[f,h^*] \le (2\pi)^2\varepsilon^2 N_\varepsilon^3 \to 0$ as $\varepsilon \to 0$. In view of $\|f'\|_{(1)}^2 \ge 4\pi^2\rho^2$ and $\varepsilon^2\sigma_1^2 \le 4\pi^2\varepsilon^2 N_\varepsilon^3 \to 0$ as $\varepsilon\to 0$, for ε small enough the inequality $\varepsilon^2\sigma_1^2 \le \|f'\|_{(1)}^2$ holds. Therefore,

$$R^\varepsilon[f,h^*] \ge \frac{\varepsilon^2\sigma_1^2\|f'\|_{(1)}^2}{\|f'\|_{(1)}^2 + \varepsilon^2\sigma_1^2} \ge \frac{\varepsilon^2\sigma_1^2}{2}. \tag{23}$$

Hence, for every function f, the quantity $R^\varepsilon[f,h^*]$ tends to zero as $\varepsilon \to 0$ slower than ε^2 and faster than $\varepsilon^2 N_\varepsilon^3$.

Lemma 3. *It holds* $\mathbf{E}[\Phi'_\varepsilon(\vartheta,\tilde h)^2] \le \varepsilon^2\sum_{k=1}^{N} h_k^{*2}(2\pi k)^2(f_k^2+\varepsilon^2)$.

Proof. Using (6), one checks that

$$\Phi'_\varepsilon(\vartheta,\tilde h)=\varepsilon\sum_{k=1}^{N}\tilde h_k(2\pi k)(f_k+\varepsilon\xi_k(\vartheta))\xi_k^*(\vartheta). \tag{24}$$

In the sequel, we write ξ_k,ξ_k^* instead of $\xi_k(\vartheta),\xi_k^*(\vartheta)$. For any $k'\neq k$, the random variable $\tilde h_k(f_k+\varepsilon\xi_k)\xi_k^*\tilde h_{k'}(f_{k'}+\varepsilon\xi_{k'})\xi_{k'}^*$ is symmetric. Therefore it has zero mean and

$$\begin{aligned}\mathbf{E}[\Phi'_\varepsilon(\vartheta,\tilde h)^2]&=\varepsilon^2\sum_{k=1}^{N}(2\pi k)^2\mathbf{E}\big[\tilde h_k^2(f_k+\varepsilon\xi_k)^2\xi_k^{*2}\big]\\&\le\varepsilon^2\sum_{k=1}^{N}(2\pi k)^2h_k^{*2}\mathbf{E}\big[(f_k+\varepsilon\xi_k)^2\xi_k^{*2}\big],\end{aligned}$$

and the assertion of the lemma follows. □

Lemma 4. *Let us denote*

$$\eta_j=\frac{2\varepsilon\sum_{k\in B_j}(2\pi k)^2f_k\xi_k+\varepsilon^2\sum_{k\in B_j}(2\pi k)^2(\xi_k^2+\xi_k^{*2}-2)}{\|f'\|_{(j)}^2+\varepsilon^2\sigma_j^2}.$$

For any positive x such that $x^2\le T_j/10$, it holds

$$\mathbf{P}\Big(|\eta_j|>x\sqrt{24(1-h^*_{\kappa_j})T_j^{-1}}\Big)\le 2e^{-x^2}.$$

Proof. Set $Y_k=2\varepsilon(2\pi k)^2f_k\xi_k+\varepsilon^2(2\pi k)^2(\xi_k^2+\xi_k^{*2}-2)$, $\sigma=\varepsilon(2\pi k)\sqrt{t}$ and $\mu=(2\pi k)f_k\sqrt{t}$. Using (18) we infer that

$$\begin{aligned}\mathbf{E}[e^{tY_k}]&=\mathbf{E}[e^{2\mu\sigma\xi_k+\sigma^2(\xi_k^2-1)}]\,\mathbf{E}[e^{\sigma^2(\xi_k^{*2}-1)}]\le e^{4\sigma^2(\mu^2+\sigma^2)}\\&=e^{4\varepsilon^2(2\pi k)^4t^2(f_k^2+\varepsilon^2)}\end{aligned}$$

as soon as $\sigma=\varepsilon(2\pi k)\sqrt{t}\le 1/2$, or equivalently $t\le 1/4\varepsilon^2(2\pi k)^2$. By [27, Thm. 2.7], for any $x>0$, we get

$$\mathbf{P}\Big(\Big|\sum_{k\in B_j}Y_j\Big|\ge x\Big(2\sum_{k\in B_j}4\varepsilon^2(2\pi k)^4(f_k^2+\varepsilon^2)\Big)^{1/2}\Big)\le 2e^{-x^2(1\wedge Q_\varepsilon x^{-1})}$$

where

$$Q_\varepsilon=\frac{(8\varepsilon^2\sum_{k\in B_j}(2\pi k)^4(f_k^2+\varepsilon^2))^{1/2}}{4(2\varepsilon\pi\kappa_{j+1})^2}.$$

It is clear that

$$Q_\varepsilon \geq \frac{(8\sum_{k\in B_j}(2\pi k)^4)^{1/2}}{4(2\pi\kappa_{j+1})^2} \geq \frac{(8(2\pi)^4\kappa_{j+1}^4 T_j/5)^{1/2}}{4(2\pi\kappa_{j+1})^2} \geq \sqrt{T_j/10}$$

$$\sum_{k\in B_j}(2\pi k)^4(f_k^2+\varepsilon^2) \leq (2\pi\kappa_{j+1})^2(\|f'\|_{(j)}^2+\varepsilon^2\sigma_j^2) \leq 3T_j^{-1}\sigma_j^2(\|f'\|_{(j)}^2+\varepsilon^2\sigma_j^2)$$
$$\leq \frac{3(1-h_{\kappa_j}^*)(\|f'\|_{(j)}^2+\varepsilon^2\sigma_j^2)^2}{T_j\varepsilon^2}.$$

These inequalities combined with the identity $\eta_j = \sum_{k\in B_j} Y_j/(\|f'\|_{(j)}^2+\varepsilon^2\sigma_j^2)$ yield the desired result. □

Lemma 5. *Assume that* $\varphi_j^2 T_j \geq 24\log\varepsilon^{-5}$. *Then*

$$\mathbf{P}\Big(\hat h_k^{JS} \in \Big[h_k^* - \frac{2(1-h_k^*)\varphi_j}{(1-\varphi_j)}\,,h_k^*\Big]\Big) \geq 1-2\varepsilon^5.$$

Proof. Note that

$$\hat h_k^{JS} = \Big(1-\frac{(1-h_k^*)(1+\varphi_j)}{1+\eta_j}\Big)\mathbb{1}_{\{\eta_j > -h_k^*+\varphi_j(1-h_k^*)\}}. \tag{25}$$

One checks that $\hat h_k^{JS} > h_k^*$ if and only if $\eta_j > \varphi_j$. Therefore, $\mathbf{P}(\hat h_k^{JS} > h_k^*) = \mathbf{P}(\eta_j > \varphi_j) \leq \varepsilon^4$. Similarly,

$$\mathbf{P}\Big(\hat h_k^{JS} < h_k^* - \frac{2(1-h_k^*)\varphi_j}{(1-\varphi_j)}\Big) = \mathbf{P}\Big(\varphi_j-(1+\varphi_j)h_k^* \leq \eta_j \leq -\varphi_j\Big)$$
$$+\mathbf{P}(\eta_j < \varphi_j-(1+\varphi_j)h_k^*)\mathbb{1}_{\{h_k^* > \frac{2(1-h_k^*)\varphi_j}{(1-\varphi_j)}\}}$$
$$\leq \mathbf{P}(\eta_j \leq -\varphi_j),$$

since $h_k^* > \frac{2(1-h_k^*)\varphi_j}{(1-\varphi_j)}$ if and only if $\varphi_j-(1+\varphi_j)h_k^* < -\varphi_j$. Therefore

$$\mathbf{P}\Big(\hat h_k^{JS} \in \Big[h_k^* - \frac{2(1-h_k^*)\varphi_j}{(1-\varphi_j)}\,,h_k^*\Big]\Big) \geq \mathbf{P}\big(|\eta_j| \leq \varphi_j\big) \geq 1-2\varepsilon^5$$

and the assertion of the lemma follows. □

Lemma 6. *For any positive* x *verifying* $x^2 \leq T_\varepsilon/5$ *it holds*

$$\mathbf{P}\Big(\Big|\sum_{k=1}^N \tilde h_k(2\pi k)^2(\xi_k^2-\xi_k^{*2})\Big| \geq \frac{12\sqrt{2}x}{\sqrt{T_\varepsilon}}\ R^\varepsilon[f,h^*]\Big) \leq 2Je^{-x^2}.$$

Proof. Using (16), we get $\left(\sum_{k\in B_j}(2\pi k)^4\right)^{1/2} \le 3\sigma_j^2/\sqrt{T_j}$. This inequality combined with the fact that $h_k \le h_k^*$ on $\mathcal{E}$, allows us to bound the probability of the event of interest by

$$\mathbf{P}\Big(\Big|\sum_{k=1}^{N}\tilde h_k(2\pi k)^2(\xi_k^2-\xi_k^{*2})\Big| \ge \sqrt{32}\,x\sum_{j=1}^{J}h^*_{\kappa_j}\Big(\sum_{k\in B_j}(2\pi k)^4\Big)^{\frac12}\Big)$$
$$\le \sum_{j=1}^{J}\mathbf{P}\Big(\tilde h_{\kappa_j}\Big|\sum_{k\in B_j}(2\pi k)^2(\xi_k^2-\xi_k^{*2})\Big| \ge \sqrt{32}\,x h^*_{\kappa_j}\Big(\sum_{k\in B_j}(2\pi k)^4\Big)^{1/2}\Big)$$
$$\le \sum_{j=1}^{J}\mathbf{P}\Big(\Big|\sum_{k\in B_j}(2\pi k)^2(\xi_k^2-\xi_k^{*2})\Big| \ge \sqrt{32}\,x\Big(\sum_{k\in B_j}(2\pi k)^4\Big)^{1/2}\Big).$$

The desired result follows now from Lemma 13 and (15). □

Lemma 7. *For any $x>0$,*

$$\mathbf{P}\bigg(\varepsilon\Big|\sum_{k=1}^{N}(\tilde h_k - h_k^*)(2\pi k)^2\xi_k f_k\Big| \ge 5x\sqrt{\frac{\varphi_\varepsilon}{T_\varepsilon}}\;R^\varepsilon[f,h^*]\bigg) \le 2Je^{-x^2}.$$

Proof. Remark first that

$$(\tilde h_{\kappa_j} - h^*_{\kappa_j})^2 \le h^{*\,2}_{\kappa_j}\wedge 8(1-h^*_{\kappa_j})^2\varphi_j^2 \le \sqrt{8}\varphi_j h^*_{\kappa_j}(1-h^*_{\kappa_j})$$

for all $j=1,\dots,J$. Set $Y_j = \varepsilon\sum_{k\in B_j}(2\pi k)^2\xi_k f_k$. The random variables $Y_1,\dots,Y_J$ are independent zero mean Gaussian with variance

$$\mathbf{E}[Y_j^2] = \varepsilon^2\sum_{k\in B_j}(2\pi k)^4 f_k^2 \le 3\varepsilon^2\sigma_j^2 T_j^{-1}\sum_{k\in B_j}(2\pi k)^2 f_k^2.$$

Therefore, $\mathbf{P}(|Y_j| \ge \sqrt{6/T_j}\;x\varepsilon\sigma_j\|f'\|_{(j)}) \le 2e^{-x^2}$ and consequently,

$$\mathbf{P}\Big(\sum_{j=1}^{n}|(\tilde h_{\kappa_j} - h^*_{\kappa_j})Y_j| \ge \sqrt{\frac{6}{T_\varepsilon}}\;x\varepsilon\sum_{j=1}^{J}\sqrt{4\varphi_j h^*_{\kappa_j}(1-h^*_{\kappa_j})}\sigma_j\|f'\|_{(j)}\Big) \le 2Je^{-x^2}.$$

To complete the proof, note that by the Cauchy-Schwarz inequality,

$$\varepsilon\sum_{j=1}^{J}\sqrt{h^*_{\kappa_j}(1-h^*_{\kappa_j})}\;\sigma_j\|f'\|_{(j)} \le \frac12\sum_{j=1}^{J}(1-h^*_{\kappa_j})\|f'\|^2_{(j)} + \frac12\sum_{j=1}^{J}h^*_{\kappa_j}(\varepsilon\sigma_j)^2$$

and the right side is bounded by $R^\varepsilon[f,h^*]$. □

Lemma 8. *We have* $\sum_{k=1}^{N}(\tilde{h}_k - h_k^*)(2\pi k)^2 f_k^2 \le 4\varphi_\varepsilon R^\varepsilon[f, h^*]$.

Proof. Since $0 \le h_k^* - \tilde{h}_k \le 4\varphi_\varepsilon(1-h_k^*)$, $\forall k \in B_j$, we have

$$\sum_{k=1}^{\infty}|\tilde{h}_k - h_k^*|(2\pi k)^2 f_k^2 \le 4\varphi_\varepsilon \sum_{k=1}^{N}(1-h_k^*)(2\pi k)^2 f_k^2 \le 4\varphi_\varepsilon R^\varepsilon[f, h^*]. \qquad \square$$

Lemma 9. *Set* $\hat{X} = \max_{1\le k\le N_\varepsilon}\sqrt{\xi_k^2(\vartheta) + \xi_k^{*2}(\vartheta)}$. *There exists an event of probability at least* $1 - 2\varepsilon^4$ *such that on this event, for all* $\tau \in \mathbb{R}$, *we have*

$$\frac{|\Phi_\varepsilon'''(\tau, \tilde{h})|}{R^\varepsilon[f, h^*]} \le \frac{12|\tau-\vartheta| \cdot \|f'\|^2}{\varepsilon^2 T_\varepsilon} + \frac{12\|f'\|\sqrt{\log\varepsilon^{-8}}}{\varepsilon T_\varepsilon} + \frac{16\sqrt{3}\,\pi N_\varepsilon|\tau-\vartheta|\hat{X}\|f'\|}{\varepsilon\sqrt{T_\varepsilon}} + 4\pi N_\varepsilon \hat{X}^2.$$

Proof. Using (6), one checks that

$$\begin{aligned}\Phi_\varepsilon'''(\tau, h) &= 2\sum_{k=1}^{\infty} h_k(2\pi k)^3[(f_k + \varepsilon\xi_k(\vartheta))^2 - \varepsilon^2\xi_k^*(\vartheta)^2]\sin[4\pi k(\tau-\vartheta)] \\ &\quad - 4\varepsilon\sum_{k=1}^{\infty} h_k(2\pi k)^3(f_k + \varepsilon\xi_k(\vartheta))\xi_k^*(\vartheta)\cos[4\pi k(\tau-\vartheta)].\end{aligned}$$

Without loss of generality, we assume in the sequel that $\vartheta = 0$. Then

$$\begin{aligned}|\Phi_\varepsilon'''(\tau, \tilde{h})| &\le 2\sum_{k=1}^{N}\tilde{h}_k(2\pi k)^3 f_k^2|\sin(4\pi k\tau)| + 4\varepsilon\sum_{k=1}^{N}\tilde{h}_k(2\pi k)^3 f_k\xi_k^* \\ &\quad + 4\varepsilon\sum_{k=1}^{N}\tilde{h}_k(2\pi k)^3 f_k\big[\xi_k\sin(4\pi k\tau) + \xi_k^*(\cos(4\pi k\tau) - 1)\big] \\ &\quad + 2\varepsilon^2\sum_{k=1}^{N}\tilde{h}_k(2\pi k)^3\big[(\xi_k^2 - \xi_k^{*2})\sin(4\pi k\tau) - 2\xi_k\xi_k^*\cos(4\pi k\tau)\big].\end{aligned}$$

Using the inequalities $|\sin(4\pi k\tau)| \le 4\pi k|\tau|$ and

$$\sum_{k=1}^{N}\tilde{h}_k(2\pi k)^4 f_k^2 \le \sum_{j=1}^{J} h_{\kappa_j}^*\|f'\|_{(j)}^2(2\pi\kappa_{j+1})^2 \le \frac{3\|f'\|^2}{T_\varepsilon}\sum_{j=1}^{J} h_{\kappa_j}^*\sigma_j^2,$$

as well as the inequality $\varepsilon^2\sum_{j=1}^{J} h_{\kappa_j}^*\sigma_j^2 \le R^\varepsilon[f, h^*]$, we get the desired bound for the first sum. The bound on the second term is obtained using Lemma 14, the well known bound on the Laplace transform of a Gaussian distribution and the inequality

$$\sum_{k=1}^{N} h_k^{*2}(2\pi k)^6 f_k^2 \le \sum_{j=1}^{J} h_{\kappa_j}^{*2}(2\pi\kappa_{j+1})^4\|f'\|_{(j)}^2 \le \frac{9}{T_\varepsilon^2}\sum_{j=1}^{J} h_{\kappa_j}^{*2}\sigma_j^4\|f'\|_{(j)}^2$$
$$\le \frac{9\|f'\|^2}{T_\varepsilon^2}\max_j h_{\kappa_j}^{*2}\sigma_j^4 \le \frac{9\|f'\|^2 R^\varepsilon[f,h^*]^2}{\varepsilon^4 T_\varepsilon^2}.$$

The bounds on the two remaining sums are obtained by combining the inequalities

$$|\xi_k \sin(4\pi k\tau) + \xi_k^*(\cos(4\pi k\tau) - 1)| \le |4\pi k\tau| \cdot |\xi_k \cos(4\pi k\tilde\tau) - \xi_k^* \sin(4\pi k\tilde\tau)|$$
$$\le 4\pi k|\tau|\hat X,$$
$$|(\xi_k^2 - \xi_k^{*2})\sin(4\pi k\tau) - 2\xi_k\xi_k^*\cos(4\pi k\tau)| \le \hat X^2$$

with arguments similar to those used to bound the first two sums. □

Lemma 10. *For any $x > 0$, it holds*

$$\mathbf{P}\Big(\Big|\Phi_\varepsilon'(\vartheta,\tilde h) - \Phi_\varepsilon'(\vartheta,h^*)\Big|^2 \ge 12\varepsilon^2 x^2\varphi_\varepsilon R^\varepsilon[f,h^*]\Big) \le 2e^{-x^2(1\wedge\sqrt{T_\varepsilon/8x^2})}.$$

Proof. Let us denote $X_k = (2\pi k)(f_k + \varepsilon\xi_k)\xi_k^*$. According to (24),

$$\Phi_\varepsilon'(\vartheta,\tilde h) - \Phi_\varepsilon'(\vartheta,h^*) = \varepsilon\sum_{j=1}^{k}(\tilde h_k - h_k^*)X_k.$$

According to (19), for all $t \le 1/(2\sqrt{2}\pi k\varepsilon)$, we have $\mathbf{E}[e^{tX_k}] \le e^{t^2(2\pi k)^2(f_k^2+\varepsilon^2)}$. Thus the conditions of Lemma 14 are fulfilled with $\varrho_k = \tilde h_k - h_k^*$, $\varrho_k^* = 3\sqrt{h_k^*\varphi_j(1-h_k^*)}$, $T_k = 1/(\sqrt{8}\,\pi k\varepsilon)$ and $g_k^2 = (2\pi k)^2(f_k^2+\varepsilon^2)$. Therefore,

$$\mathbf{P}\Big(\Big|\Phi_\varepsilon'(\vartheta,\tilde h) - \Phi_\varepsilon'(\vartheta,h^*)\Big|^2 \ge 12\varepsilon^2x^2\varphi_\varepsilon\sum_{k=1}^{N}(2\pi k)^2 h_k^*(1-h_k^*)(f_k^2+\varepsilon^2)\Big) \le 2e^{-x^2}$$

for any $x > 0$ verifying

$$x^2 \le \frac{3\sum_{k=1}^{N}\varphi_{j(k)}(2\pi k)^2 h_k^*(1-h_k^*)(f_k^2+\varepsilon^2)}{8\varepsilon^2\max_j\varphi_j h_{\kappa_j}^*(1-h_{\kappa_j}^*)(2\pi\kappa_{j+1})^2}.$$

To complete the proof, it suffices to remark that

$$\sum_{k=1}^{N}(2\pi k)^2 h_k^*(1-h_k^*)(f_k^2+\varepsilon^2) = \sum_{j=1}^{J} h_{\kappa_j}^*(1-h_{\kappa_j}^*)[\|f'\|_{(j)}^2 + \varepsilon^2\sigma_j^2] \le R^\varepsilon[f,h^*]$$

and

$$\frac{3\sum_{k=1}^{N}\varphi_{j(k)}(2\pi k)^2 h_k^*(1-h_k^*)(f_k^2+\varepsilon^2)}{8\varepsilon^2\max_j\varphi_j h_{\kappa_j}^*(1-h_{\kappa_j}^*)(2\pi\kappa_{j+1})^2} \geq \frac{3\max_j\varphi_j h_{\kappa_j}^*(1-h_{\kappa_j}^*)\sigma_j^2}{8\max_j\varphi_j h_{\kappa_j}^*(1-h_{\kappa_j}^*)(2\pi\kappa_{j+1})^2}.$$

Since $\sigma_j^2 \geq T_j(2\pi\kappa_{j+1})^2/3$, the assertion of the lemma follows. $\square$

Lemma 11. *For any event $\mathcal{A}$ verifying $\mathbf{P}(\mathcal{A}^c)=O(\varepsilon^4)$, we have*

$$\mathbf{E}[\Phi_\varepsilon'(\vartheta,h^*)^2\zeta\mathbb{1}_{\mathcal{A}}]=o(\varepsilon^2R^\varepsilon[f,h^*]).$$

Proof. We have

$$\begin{aligned}\mathbf{E}[\Phi_\varepsilon'(\vartheta,h^*)^2\zeta] &= \varepsilon^2\mathbf{E}\Big[\sum_{k=1}^{N}(2\pi k)^2 h_k^{*2}(f_k+\varepsilon\xi_k(\vartheta))^2\xi_k^*(\vartheta)^2\zeta\Big]\\ &= 2\varepsilon^3\mathbf{E}\Big[\sum_{k=1}^{N}(2\pi k)^2h_k^{*2}f_k\xi_k\zeta\Big]=4\varepsilon^4\sum_{k=1}^{N}(2\pi k)^4h_k^{*3}f_k^2\\ &\leq 12\|f'\|^2T_\varepsilon^{-1}\varepsilon^4\sum_{j=1}^{J}{h_{\kappa_j}^*}^3\sigma_j^2 \leq 12\|f'\|^2T_\varepsilon^{-1}\varepsilon^2R^\varepsilon[f,h^*].\end{aligned}$$

Using the Rosenthal inequality, one checks that $\mathbf{E}[\Phi_\varepsilon'(\vartheta,h^*)^{2p}]=O(\varepsilon^{2p})$ and $\mathbf{E}[\zeta^{2p}]=o(R^\varepsilon[f,h^*]^p)$ for any integer $p>0$. Therefore, the Cauchy-Schwarz inequality yields,

$$\big|\mathbf{E}[\Phi_\varepsilon'(\vartheta,h^*)^2\zeta\mathbb{1}_{\mathcal{A}^c}]\big| \leq o(\varepsilon^2\sqrt{R^\varepsilon[f,h^*]}\,)\sqrt{\mathbf{P}(\mathcal{A}^c)}=o(\varepsilon^4\sqrt{R^\varepsilon[f,h^*]}\,)$$

and the assertion of the lemma follows. $\square$

6.4. Lemma used in Theorem 2

We assume that $f\in\mathcal{F}_{\delta,\beta,L}(\bar f)$ with $\bar f\in\mathcal{F}(\beta^*,L^*,\rho)$ and $\beta^*>\beta\geq\beta_*$. For seek of completeness we give below a suitable version of the Pinsker theorem (cf. Pinsker (1980)).

Lemma 12. *Set $\gamma_\varepsilon=1/\log\varepsilon^{-2}$, $W_\varepsilon=\big(\frac{L}{\varepsilon^2}\frac{(\beta+2)(2\beta+1)}{(2\pi)^{2\beta}(\beta-1)}\big)^{1/(2\beta+1)}$ and define*

$$\lambda_k^*=\begin{cases}1, & k\leq\gamma_\varepsilon W_\varepsilon,\\ \Big[1-\Big(\dfrac{k}{W_\varepsilon}\Big)^{\beta-1}\Big]_+, & k>\gamma_\varepsilon W_\varepsilon.\end{cases}$$

The filter λ_k^* *satisfies*

$$\sup_{f\in\mathcal{F}_{\delta,\beta,L}} R^\varepsilon[f,\lambda^*] \le (1+o(1))\, C(\beta,L)\, \varepsilon^{\frac{4\beta-4}{2\beta+1}},$$

where the inf *is taken over all possible sequences* $h=(h_k)_{k\in\mathbb{N}}$.

Proof. Set $v=f-\bar f$. Using the inequality $(\bar f_k+v_k)^2\le 2z^{-1}\bar f_k^2+(1+z)v_k^2$, $\forall z\in[0,1]$, we obtain

$$\begin{aligned} R^\varepsilon[f,\lambda^*] &= \sum_{k>\gamma_\varepsilon W_\varepsilon}(2\pi k)^2(1-\lambda_k^*)^2(\bar f_k+v_k)^2+\varepsilon^2\sum_{k=1}^\infty (2\pi k)^2\lambda_k^{*2}\\ &\le 2z^{-1}\sum_{k>\gamma_\varepsilon W_\varepsilon}(2\pi k)^2(1-\lambda_k^*)^2\bar f_k^2+(1+z)R^\varepsilon[v,\lambda^*]. \end{aligned}$$

Since $\bar f\in\mathcal{F}(\beta^*,L^*,\rho)$, we have

$$\sum_{k>\gamma_\varepsilon W_\varepsilon}(2\pi k)^2(1-\lambda_k^*)^2\bar f_k^2\le L^*(\gamma_\varepsilon W_\varepsilon)^{2-2\beta^*}=o(W_\varepsilon^{2-2\beta}). \tag{26}$$

On the other hand, setting $\tilde\lambda_k^*=(1-(k/W_\varepsilon)^{\beta-1})_+$,

$$R^\varepsilon[v,\lambda^*]-R^\varepsilon[v,\tilde\lambda^*]\le\varepsilon^2\sum_{k\le\gamma_\varepsilon W_\varepsilon}(2\pi k)^2\le 4\pi^2\varepsilon^2(\gamma_\varepsilon W_\varepsilon)^3=o(\varepsilon^2W_\varepsilon^3).$$

Using the relation $W_\varepsilon^{2-2\beta}=O(\varepsilon^2W_\varepsilon^3)$ and choosing $z=z_\varepsilon$ appropriately, we get

$$\sup_{f\in\mathcal{F}_{\delta,\beta,L}(\bar f)} R^\varepsilon[f,\lambda^*]\le(1+o(1))\sup_{v\in\mathcal{W}(\beta,L)}R^\varepsilon[v,\tilde\lambda^*]+o(\varepsilon^2W_\varepsilon^3),$$

where $\mathcal{W}(\beta,L)$ is the Sobolev ball $\{v:\sum_{k\ge1}(2\pi k)^{2\beta}f_k^2\le L\}$. It then follows from Belitser and Levit (1995), Thm. 1 and Example 1, p. 265 (with $\alpha=\beta-1$ and $\delta=3$) that $\sup_{v\in\mathcal{W}(\beta,L)}R^\varepsilon[v,\tilde\lambda^*]=C(\beta,L)\,\varepsilon^{\frac{4\beta-4}{2\beta+1}}(1+o(1))$. To conclude, it suffices to remark that $o(\varepsilon^2W_\varepsilon^3)=o(\varepsilon^{\frac{4\beta-4}{2\beta+1}})$. □

6.5. General results

Lemma 13. *Assume that* $a_1,a_2,\ldots,a_n\in\mathbb{R}$ *and* $\varsigma=\sum_{k=1}^n a_k(\xi_k^2-\xi_k^{*2})$, *where* $(\xi_1,\ldots,\xi_n,\xi_1^*,\ldots,\xi_n^*)$ *is a zero mean Gaussian vector with identity covariance matrix. For any* $y\in[0,\|a\|/\max_k a_k]$, *it holds*

$$\mathbf{P}\Big(\varsigma^2\ge 32y^2\sum_{k=1}^n a_k^2\Big)\le 2e^{-y^2}.$$

Proof. Using the formula of the Laplace transform of a chi-squared distribution, for any $t \in [-(\sqrt{8}a_k)^{-1}, (\sqrt{8}a_k)^{-1}]$, we get

$$\mathbf{E}[e^{a_k t(\xi_k^2 - \xi_k^{*2})}] = \frac{1}{1 - 4a_k^2 t^2} \le e^{8a_k^2 t^2}.$$

Applying [27, Thm. 2.7] with $g_k = 16a_k^2$ and $x = \sqrt{32}y\|a\|$, we get the desired result. □

Lemma 14. *Let $X_1, \ldots, X_n$ be independent symmetric random variables. Let $\boldsymbol{\varrho} = (\varrho_1, \ldots, \varrho_n)$ be a random vector satisfying $|\varrho_j| \le \varrho_j^*$, $\forall j = 1, \ldots, n$ with some deterministic sequence $(\varrho_j^*)_{j=1}^n$ and $\mathscr{L}(\boldsymbol{\varrho}|X_j = x) = \mathscr{L}(\boldsymbol{\varrho}|X_j = -x)$ for all $j \in \{1, \ldots, n\}$. If*

$$\mathbf{E}[e^{tX_j}] \le e^{t^2 g_j^2}$$

for some sequence $(g_j)_{j=1,\ldots,n}$ and for $|t| \le T_j$, then

$$\mathbf{P}\Big(\Big|\sum_{j=1}^n \varrho_j X_j\Big|^2 \ge 4x^2 \sum_{j=1}^n \varrho_j^{*2} g_j^2\Big) \le 2e^{-x^2(1\wedge Q_n x^{-1})}, \ \forall\, x > 0$$

where $Q_n = (\sum_{j=1}^n \varrho_j^{*2} g_j^2)^{1/2} \min_j (T_j/\varrho_j^*)$.

Proof. Set $Y_j = \varrho_j X_j$ and $\bar{Y}_j = \varrho_j^* X_j$. For any $p_1, \ldots, p_n \in \mathbb{N}^n$, the expectation $\mathbf{E}[Y_1^{p_1} \cdot \ldots \cdot Y_n^{p_n}]$ vanishes if at least one p_j is odd. Therefore, $\mathbf{E}[(\sum_j Y_j)^k] = 0$ if k is odd and $\mathbf{E}[(\sum_j Y_j)^k] \le \mathbf{E}[(\sum_j \bar{Y}_j)^k]$ if k is even. Hence

$$\begin{aligned}\mathbf{E}\Big[\exp\Big(t\sum_j Y_j\Big)\Big] &= \sum_{k=0}^\infty \frac{t^k \mathbf{E}[(\sum_j Y_j)^k]}{k!} = \sum_{k=0}^\infty \frac{t^{2k}\mathbf{E}[(\sum_j \bar{Y}_j)^{2k}]}{(2k)!} \\ &= \mathbf{E}[e^{t\sum_j \varrho_j^* X_j}] \le e^{t^2 \sum_j \varrho_j^{*2} g_j^2}, \quad \forall\, |t| \le \min_j (T_j/\varrho_j^*).\end{aligned}$$

According to the Markov inequality,

$$\mathbf{P}\Big(\Big|\sum_{j=1}^n \varrho_j X_j\Big| \ge y\Big) \le 2e^{-ty}\mathbf{E}[e^{t\sum_j Y_j}] \le 2\exp\Big(-ty + t^2 \sum_{j=1}^n \varrho_j^{*2} g_j^2\Big).$$

Setting $t = \frac{x\wedge Q_n}{\sqrt{\sum_j \varrho_j^{*2} g_j^2}}$ and $y^2 = 4x^2 \sum_j \varrho_j^{*2} g_j^2$ we get the desired result. □

References

[1] Belitser, E. and Levit, B. (1995). On minimax filtering on ellipsoids. *Math. Methods Statist.* **4** 259-273.

[2] Bickel, P. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.

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

[4] Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist.*, **27**, 898–924.

[5] Cai, T. and Low, M. (2005). Nonparametric estimation over shrinking neighborhoods: superefficiency and adaptation. *Ann. Statist.*, **33** (1), 184–213.

[6] Castillo, I. (2005) Semiparametric second order efficient estimation of the period of a signal and applications, preprint.

[7] Cavalier, L. and Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. *Proba. Theory and Related Fields* 123, 323–354.

[8] Cavalier, L. and Tsybakov, A. (2001). Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation. *Math. Methods of Stat.* 10, 247–282.

[9] Dalalyan, A., Golubev, G. and Tsybakov, A. (2006). Penalized Maximum Likelihood and Semiparametric Second Order Efficiency. *Ann. Statist.*, **34** (1), to appear.

[10] Donoho, D. and Johnstone, I. (1995), Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, **90** (432), 1200–1224.

[11] Efromovich, S. Yu. (1999). *Nonparametric curve estimation. Methods, theory, and applications.* Springer Series in Statistics. Springer-Verlag, New York.

[12] Efromovich, S. and Samarov, A. (2000). Adaptive estimation of the integral of squared regression derivatives. *Scand. J. Statist.* **27**(2), 335–351.

[13] Fan, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* **19** (3), 1273–1294.

[14] Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Manuscript downloadable from `http://www.orfe.princeton.edu/∼jqfan/fan/manuscripts.html`

[15] Gamboa, F., Loubes, J.-M. and Maza, E. (2005). Shifts estimation for high dimensional data. Manuscript downloadable from `http://mahery.math.u-psud.fr/∼loubes/`

[16] Golubev, G. (1990). On estimation of time delay of a signal under nuisance parameters. *Probl. Inf. Transm.* **25** 3-12.

[17] Golubev, G. and Härdle W. (2000). On the second order minimax estimation in partial linear models. *Math. Methods of Stat.* 2000, 2, 160–175.

[18] Golubev, G. and Härdle, W. (2002) On adaptive smoothing in partial linear models. *Math. Methods of Stat.* 1, 98–117.

[19] Härdle, W. and Marron, J. (1990). Semiparametric comparison of regression curves. *Ann. Statist.* **18** 63-89.

[20] Härdle, W. and Tsybakov, A. (1993). How sensitive are average derivatives? *J. Econometrics* **58** 31-48.

[21] Hallin, M. and Werker, B. (2003). Semi-parametric efficiency, distribution-freeness and invariance. *Bernoulli* **9** 137-165.

[22] James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Forth Berkeley Symp. Math. Statist. Prob.* **1** 361-379. Univ. of California Press.

[23] Kang, K., Kim, W. and Park, B. (2000). Higher order kernels in adaptive location estimation. *J. Nonparametr. Statist.* 12(6), 879–888.

[24] Lavielle, M. and Lévy-Leduc, C. (2004). Semiparametric estimation of the frequency of unknown periodic functions and its application to laser vibrometry signals. *To appear in IEEE Transactions on Signal Processing.*

[25] Mammen, E. and Park, B. (1997). Optimal smoothing in adaptive location estimation. *J. Statist. Plann. Inference* 58(2), 333–348.

[26] Murphy, S. and van der Vaart, A. (2000), On Profile Likelihood. *J. Amer. Statist. Assoc.*, **95**, 449-485.

[27] Petrov, V. *Limit theorems of probability theory. Sequences of independent random variables.* (1995) Oxford Studies in Probability, 4. Oxford University Press, New York.

[28] Pfanzagl, J. (1990). *Estimation in Semiparametric Models,* Lecture Notes in Statist, **63**, Springer, New York.

[29] Pinsker, M. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Probl. Inf. Transm.* **16** 120-133.

[30] Rigollet, Ph. (2004). Adaptive density estimation using Stein's blockwise method. Preprint PMA-913.

[31] Schick, A. (1998). Estimating a shift in nonparametric regression via U-statistics. *J. Statist. Plann. Inference.* **67** (2), 259–271.

[32] Schick, A. (1999). Efficient estimation of a shift in nonparametric regression. *Statist. Probab. Lett.*, **41** (3), 287–301.

[33] Severini, T. and Wong, W. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist.* **20** 1768-1802.

[34] Stone, C. (1975). Adaptive maximum likelihood estimation of a location parameter. *Ann. Statist.* **3** 267-284.

[35] van der Vaart, A. (2001). *Semiparametric Statistics,* manuscript downloadable from `http://www.math.vu.nl/sto/publications.php`