

Change-point estimation from indirect observations

1. Minimax complexity

A. Goldenshluger *
University of Haifa

A. Juditsky†
Université Grenoble I

A.B. Tsybakov ‡
Université Paris VI

A. Zeevi
Columbia University §

January 18, 2007

Abstract

We consider the problem of nonparametric estimation of signal singularities from indirect and noisy observations. Here by singularity we mean a discontinuity (change-point) of the signal or of its derivative. The model of indirect observations we consider is that of a linear transform of the signal, observed in white noise. The estimation problem is analyzed in a minimax framework. We provide lower bounds for minimax risks and propose rate-optimal estimation procedures.

Résumé

Cet article a pour but d'étudier le problème d'estimation non-paramétrique de singularités d'un signal à partir des observations indirectes et bruitées. Les singularités que nous considérons ici sont des points de discontinuité (points de rupture) du signal ou de ses dérivées. Nous étudions le modèle où l'on dispose d'observations indirectes d'une transformée linéaire du signal dans le bruit blanc gaussien. Le problème de l'estimation est analysé dans un cadre minimax. Nous obtenons des minorations du risque minimax et nous proposons des estimateurs qui sont optimaux en vitesse de convergence.

1 Introduction

Let us start with three estimation problems which motivate our study.

Problem 1. [*Estimation of a change-point in derivatives.*] Consider the Gaussian white noise model

$$dY(t) = f(t)dt + \varepsilon dW(t), \quad t \in [0, 1], \quad (1)$$

where f is an unknown periodic function on $[0, 1]$, $\varepsilon > 0$, and W is the standard Wiener process. Assume that f is α times differentiable, and $f^{(\alpha)}$ is smooth apart from a single discontinuity of the first kind at the point $\theta \in [0, 1]$. We are interested in estimating the

*Department of Statistics, University of Haifa, Haifa 31905. Supported by the ISF grant No. 300/04. e-mail: goldensh@stat.haifa.ac.il

†LMC-IMAG 51 rue des Mathématiques B. P. 53 38041 Grenoble Cedex 9 France. e-mail: anatoli.iouditski@imag.fr

‡Laboratoire de Probabilités et Modèles Aléatoires, CNRS UMR 7599, Université Paris VI, 4, Place Jussieu, Boîte courrier 188, 75252 Paris Cedex 05 France. e-mail: tsybakov@ccr.jussieu.fr

§Graduate School of Business, Columbia University, New York, NY 10027, USA. e-mail: as-saf@gsb.columbia.edu

change-point θ , and the amplitude a of the jump. When α is not integer then $f^{(\alpha)}$ is understood as the Weyl fractional derivative of f . Let $m = \lfloor \alpha \rfloor$ (here $\lfloor \alpha \rfloor$ stands for the integer part of α). We say that $f^{(m)}$ has a cusp of the order $\alpha - m$ at θ .

Problem 2. [*Change-point estimation in the convolution white model.*] The white noise convolution model is given by the equation

$$dY(t) = (Kf)(t)dt + \varepsilon dW(t), \quad t \in [0, 1], \quad (2)$$

where f is a periodic function on $[0, 1]$, $\varepsilon > 0$, W is the standard Wiener process, and the operator K is that of the periodic convolution on $[0, 1]$:

$$(Kf)(t) = \int_0^1 K(t-s)f(s)ds.$$

The function f is assumed to be smooth apart from a single discontinuity at $\theta \in [0, 1]$. The goal here is to estimate the change-point θ and the jump amplitude a .

Note that the derivative change-point estimation is a special case of Problem 2. Indeed, assume without loss of generality that f has zero mean value. Then for any $\alpha > 0$ the operation of Weyl's fractional integration is expressed in terms of convolution:

$$f^{(-\alpha)}(t) = (Kf)(t) = \int_0^1 K_\alpha(t-s)f(s)ds, \quad K_\alpha(t) \equiv \sum_{k=-\infty, k \neq 0}^{\infty} \frac{e^{-2\pi ikt}}{(-2\pi ik)^\alpha}.$$

In this case estimating a change-point in f from observations (2) amounts to estimating a change-point in $f^{(\alpha)}$ from observations (1). For more details on fractional derivatives and integrals of periodic functions we refer to Samko, Kilbas and Marichev (1993, §19).

Problem 3. [*Delay and amplitude estimation.*] Let S be a known periodic signal. Assume that we observe the trajectory $Y = (Y(t))$, $t \in [0, 1]$ where

$$dY(t) = [aS(t-\theta) + q(t)]dt + \varepsilon dW(t), \quad t \in [0, 1], \quad (3)$$

$a \in \mathbf{R} \setminus \{0\}$ is unknown nuisance parameter, $\theta \in [0, 1]$, q is an unknown smooth periodic nuisance function, $\varepsilon > 0$, and W is the standard Wiener process. We are interested in estimation of the delay parameter θ and the signal amplitude a . To ensure that θ is identifiable in this setup, some additional conditions on q should be imposed. For instance, one can require that function q is smoother, in a certain sense, than S .

Change-point estimation and detection is one of most important tasks of statistics and as such it retained much attention of statistical and signal processing community (cf. the books Ibragimov and Has'minski (1981), Basseville and Nikiforov (1993), and the references therein). This problem is also well represented in the literature on nonparametric regression estimation (cf. the works of Korostelev (1987), Yin (1988), Müller (1992), Wang (1995), Raimondo (1998), Gijbels, Hall and Kneip (1999), Antoniadis and Gijbels (2002), among many others). Certainly, among the estimation problems, presented above, it is Problem 1 that was treated most extensively. For instance, Korostelev (1987) considered minimax estimation of the change-point in f from direct observations ($\alpha = 0$). A remarkable result in Korostelev (1987) states that the minimax risk over the class of functions having a single change-point and satisfying the Lipschitz condition away from the change-point is ε^2 , while the minimax rate for the sequential (Markov) estimator is $\varepsilon^2 \ln \varepsilon^{-1}$. In the problems of sequential estimation of a change-point in the signal ($\alpha = 0$), or in its derivative ($\alpha = 1$), precise asymptotic expressions for the minimax risk have been obtained in Freidlin and Korostelev (1995). It has been proved there that minimax rate of sequential estimation of the change-point in the derivative ($\alpha = 1$) from observations (1) is $(\varepsilon^2 \ln \varepsilon^{-1})^{1/3}$. The

problem of minimax estimation of cusps and change-points in derivatives ($\alpha > 0$) is much less studied. For instance, Raimondo (1998) has developed a suboptimal wavelet estimator, some suboptimal change-point estimators were proposed recently by Huh and Carriere (2002) and Park and Kim (2004).

A problem of change-point estimation closely related to Problem 2 has been studied by Neumann (1997). In particular, the minimax rates of convergence for change-point estimation in the density convolution model are derived in that paper. Namely, for the class of probability densities which are Lipschitz-continuous away from the change-point, Neumann (1997) shows that the minimax rate of estimation of the change-point is $\min\{n^{-2/(2\beta+3)}, n^{-1/(2\beta+1)}\}$ where n is the size of the observation sample, and β is the ill-posedness index of the convolution. With usual calibration $n^{-1/2} = \varepsilon$ this result can be translated to the white noise convolution model. Recently Goldenshluger, Tsybakov and Zeevi (2006) have studied the problem of change-point estimation in the convolution model (2). They have extended the results of Neumann (1997) to the classes functions which are smooth, except at the point of discontinuity and provided a number of minimax results for this problem.

In the statistical literature Problem 3 is mainly studied under parametric assumptions, i.e. when $q(t) = 0$ (see, e.g. Ibragimov and Has'minski (1981, § 7.2), Kutoyants (2004)). For some related models we refer also to Golubev (1989) and Dalalyan, Golubev and Tsybakov (2006).

In this paper we propose a unified framework to study Problems 1–3. We show that estimation of the change-point θ and the jump amplitude a in Problems 1–3 can be reduced to the problem of recovering the frequency and the amplitude of a complex harmonic oscillation in the presence of random noise and a deterministic nuisance. To be more precise, consider the following sequence space model

$$y_k = a \exp(2\pi i k \theta) + g_k + \varepsilon \sigma_k \xi_k, \quad k \in \mathbf{N},$$

where $g = (g_k) \in \mathbf{C}^{\mathbf{N}}$ is an unknown nuisance sequence, $\sigma = (\sigma_k) \in \mathbf{C}^{\mathbf{N}}$ is a given sequence, and $\xi = (\xi_k) \in \mathbf{C}^{\mathbf{N}}$ is a sequence of independent standard complex-valued normal random variables. We demonstrate below that this model includes the aforementioned Problems 1–3 as special cases. Then we concentrate on the study of theoretical accuracy limitations in estimating θ and a , and develop corresponding rate-optimal procedures. Our frequency domain estimation technique is closely related to spectral analysis of time series and frequency estimation. There is vast literature on estimation of complex harmonic signals from noisy observations (see, e.g., Hannan (1973), Rice and Rosenblatt (1988), and the recent book Quinn and Hannan (2001) for further references). In the forthcoming second part of this paper (Goldenshluger, Juditsky, Tsybakov and Zeevi (2006)) we develop adaptive estimators of θ and a that do not require prior knowledge of regularity of the nuisance sequence (g_k) .

The rest of the paper is organized as follows. In Section 2 we formulate the estimation problem in the sequence space and establish its relationship to Problems 1 – 3. We define the estimator of the jump amplitude and study its properties in Section 3.2. Finally, Section 3.3 is devoted to the change-point estimation. Proofs of main results are given in Section 4, auxiliary results are relegated to Appendix.

2 Sequence space problem formulation

Consider the following model in the space of sequences

$$y_k = a \exp(2\pi i k \theta) + g_k + \varepsilon \sigma_k \xi_k, \quad k \in \mathbf{N}, \quad (4)$$

where $a \in \mathbf{R}$, $\theta \in [0, 1]$ are unknown constants, $g = (g_k) \in \mathbf{C}^{\mathbf{N}}$ is an unknown nuisance sequence, $\sigma = (\sigma_k) \in \mathbf{C}^{\mathbf{N}}$ is a given sequence, and $\xi = (\xi_k) \in \mathbf{C}^{\mathbf{N}}$ is the sequence of

independent standard complex-valued Gaussian random variables: $(\Re\xi_k, \Im\xi_k) \sim \mathcal{N}(0, I)$. The goal is to estimate θ and a using observations $y_k, k \in \mathbf{N}$.

For a real-valued function $f \in L_2[0, 1]$ we denote (f_k) the sequence of Fourier coefficients of f :

$$f_k \equiv \int_0^1 f(t)e^{2\pi ikt} dt, \quad k \in \mathbf{Z}.$$

Note that, as $f(t)$ is real, $f_{-k} = \bar{f}_k$ (here \bar{x} is the complex conjugate of x). This is why, in what follows, we only consider the part of the sequence (f_k) for $k \geq 0$.

We will assume that the nuisance sequence (g_k) belongs to the Sobolev ellipsoid $G_s(L)$, $s > -\frac{1}{2}$, $0 < L < \infty$,

$$G_s(L) = \left\{ g \in \mathbf{C}^{\mathbf{N}} \mid \sum_{k=1}^{\infty} |g_k|^2 k^{2s} \leq L^2 \right\}.$$

We will refer to the function f to belong to the ellipsoid $G_s(L)$ if the sequence of its Fourier coefficients is in $G_s(L)$. The sequence (σ_k) is assumed to satisfy the following assumption:

Assumption (A). For some $\beta > 1/2$ and $0 < \underline{\sigma} \leq \bar{\sigma}$

$$\underline{\sigma} k^\beta \leq |\sigma_k| \leq \bar{\sigma} k^\beta, \quad \forall k \in \mathbf{N}. \quad (5)$$

The general scheme that we develop in this paper can be applied to a variety of cases, including the setup of infinitely smooth nuisance functions and/or severely ill-posed problems when $|\sigma_k|$ grows exponentially with k . These extensions do not introduce major conceptual difficulties and require only a technical care. They are left beyond the scope of our work.

Let us show that Problems 1 – 3, formulated in Section 1, can be expressed in terms of the observation model (4).

Problem 1. If $f^{(\alpha)}$ has a single discontinuity of size a at $\theta \in [0, 1]$, then it can be uniquely represented as

$$f^{(\alpha)}(t) = aV(t - \theta) + q(t), \quad t \in [0, 1], \quad (6)$$

where $V(t) = 1/2 - t - [t]$ is the ‘‘saw-tooth’’ function, and $q \in G_s(L)$, $s > \frac{1}{2}$. We note that (6) is the standard way of representation of discontinuous functions in the theory of Fourier series [see, e.g., Zygmund (1959, p. 9)]. Then the model (1) is equivalent to the sequence-space model (4) where

$$g \in G_{s-1}(2\pi L), \quad \sigma_k^2 = (2\pi k)^{2\alpha+2} \quad (7)$$

Indeed, as

$$V_k = \begin{cases} (2\pi ik)^{-1}, & k = 1, 2, \dots \\ 0, & k = 0, \end{cases}$$

and, due the periodicity of f , $g_0 = 0$, the Fourier coefficients of the function $f^{(\alpha)}$ in (6) are

$$f_k^{(\alpha)} = a(2\pi ik)^{-1} e^{2\pi ik\theta} + q_k, \quad k \in \mathbf{N}^+, \quad \text{and} \quad f_0^{(\alpha)} = 0.$$

On the other hand, the model (1) is clearly equivalent to

$$z_k = f_k + \varepsilon\eta_k, \quad k = 0, 1, 2, \dots$$

where $z_k = \int_0^1 e^{2\pi ikt} dY(t)$, and η_k are i.i.d. standard complex-valued Gaussian random variables. Note that

$$f_k^{(\alpha)} = (-2\pi ik)^\alpha f_k, \quad k \in \mathbf{N}^+.$$

Thus we obtain for $y_k = (-1)^\alpha (2\pi ik)^{\alpha+1} z_k$, $g_k = (2\pi ik)q_k$ and $\xi_k = (-1)^\alpha i^{\alpha+1} \eta_k$,

$$y_k = ae^{2\pi ik\theta} + g_k + \sigma_k \varepsilon \xi_k, \quad k \in \mathbf{N}^+$$

with (σ_k) and (g_k) which satisfy (7).

Note that the problem of estimating the change-point in f from observations (1) [as in Korostelev (1987)] corresponds to the model (4) with $\sigma_k = 2\pi k$, i.e. $\beta = 1$.

Problem 2. Suppose, as above that the decomposition $f(t) = aV(t - \theta) + q(t)$ hold, and $q \in G_s(L)$. Assume that the Fourier coefficients (K_k) of the kernel K do not vanish, moreover, let for some $\alpha > 1/2$ and $0 < c \leq C < \infty$ the kernel K satisfy:

$$ck^\alpha \leq |(K_k)^{-1}| \leq Ck^\alpha.$$

When using the same arguments as above we conclude that the model (2) can be equivalently rewritten in the form (4) with

$$\sigma_k = 2\pi k |(K_k)^{-1}|, \quad k \in \mathbf{N}^+; \quad g \in G_{s-1}(2\pi L).$$

Observe that the relation (5) holds with $\beta = \alpha + 1$ and $\underline{\sigma} = 2\pi c$ and $\bar{\sigma} = 2\pi C$.

Problem 3. Suppose that in the model (3) $g \in G_s(L)$ and for some $\frac{1}{2} < \alpha < s$ and $0 < c \leq C < \infty$

$$ck^\alpha \leq |S_k^{-1}| \leq Ck^\alpha, \quad k \in \mathbf{N}^+.$$

Obviously, the model (3) is equivalent to (4) with

$$\sigma_k = |S_k^{-1}|, \quad k \in \mathbf{N}^+, \quad g \in G_{s-\alpha}(CL),$$

with $ck^\alpha \leq \sigma_k \leq Ck^\alpha$.

3 Main results

Our objective in this section is to bound the minimax complexity of the problem of estimating parameters a and θ in the model (4). With some terminology abuse we will refer to θ and $|a|$ as change-point and jump amplitude. Let us first introduce some notation.

3.1 Preliminaries

Complexity measures. Let $\hat{\theta}$ be an estimator of θ based on observations (y_k) , as in (4). We measure the accuracy of $\hat{\theta}$ with the maximal risk

$$R_\theta[\hat{\theta}; G_s(L)] \equiv \sup_{g \in G_s(L), \theta \in [0,1]} \left(E(\hat{\theta} - \theta)^2 \right)^{1/2},$$

(here $E(\cdot) = E_{\theta,g}(\cdot)$ stands for the expectation with respect to the distribution of (ξ_k)). The minimax complexity $R_\theta^*[G_s(L)]$ of the estimation problem (the minimax risk) is defined by

$$R_\theta^*[G_s(L)] \equiv \inf_{\hat{\theta}} R_\theta[\hat{\theta}; G_s(L)],$$

where the infimum is taken over all estimators $\hat{\theta}$ (i.e., measurable functions of observations). Note that the minimax risks of θ -estimation depend on $|a|$, but we do not indicate this explicitly to alleviate the notation. Similarly, in the problem of estimating the jump amplitude $|a|$, we define

$$R_a[\hat{a}; G_s(L)] \equiv \sup_{g \in G_s(L), \theta \in [0,1], a \in \mathbf{R}} \left(E(\hat{a} - |a|)^2 \right)^{1/2}, \quad R_a^*[G_s(L)] \equiv \inf_{\hat{a}} R_a[\hat{a}; G_s(L)].$$

The following construction underlies the estimator design:

Contrast functions. Let N be a positive integer number, a design parameter to be chosen. Consider the following random functions:

$$\widehat{J}_N(t) \equiv \left| \sum_{k=N+1}^{2N} y_k e^{-2\pi i k t} \right|^2, \quad t \in [0, 1] \quad (8)$$

$$\widehat{H}_N(t) \equiv 2\pi i \sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} (k-j) y_k \bar{y}_j e^{-2\pi i (k-j)t}, \quad t \in [0, 1]. \quad (9)$$

In fact, \widehat{J}_N is nothing but the rescaled periodogram of the data y_k confined to the spectral window $\{N+1, \dots, 2N\}$. Observe also that $\widehat{H}_N(t) = -\widehat{J}'_N(t)$. We denote

$$J_N(t) \equiv a^2 \left| \sum_{k=N+1}^{2N} \exp(-2\pi i k(t-\theta)) \right|^2$$

$$H_N(t) \equiv 2\pi i a^2 \sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} (k-j) \exp(-2\pi i (k-j)(t-\theta)).$$

Note that J_N and H_N correspond to the functions, defined in (8) and (9), when $y_k = a \exp(2\pi i k \theta)$. Let

$$F_N(t) \equiv \left| \sum_{k=N+1}^{2N} e^{-2\pi i k t} \right|^2 = \left(\frac{\sin \pi t N}{\sin \pi t} \right)^2, \quad U_N(t) \equiv -F'_N(t); \quad (10)$$

$F_N(\cdot)$ is the standard Fejer summability kernel up to a normalization (Zygmund 1959). It is easily seen that

$$J_N(t) = a^2 F_N(t-\theta), \quad \text{and} \quad H_N(t) = a^2 U_N(t-\theta).$$

The parameters θ and a admit useful characterization in terms of functions J_N and H_N . In particular, θ is the unique global maximizer of J_N , and the corresponding maximal value is $a^2 N^2$. Furthermore, let θ_- and θ_+ be, respectively, the (unique) global maximizer and the unique global minimizer of H_N . Then θ is the unique zero of H_N on the segment with the endpoints at θ_- and θ_+ (it is also the midpoint of this segment). Further, as we show in Section 4.1, the functions \widehat{J}_N and \widehat{H}_N converge to J_N and H_N uniformly on $[0, 1]$.

Construction of our estimates for a and θ is based on statistics $\widehat{J}_N(t)$ and $\widehat{H}_N(t)$ and exploits the aforementioned properties of their ‘‘ideal counterparts’’, $J_N(t)$ and $H_N(t)$.

3.2 Jump amplitude estimation

Now we turn to the problem of the jump amplitude estimation. We set

$$\widehat{a}_N \equiv N^{-1} \max_{t \in [0, 1]} \sqrt{\widehat{J}_N(t)}.$$

Theorem 1 *Suppose that Assumption (A) holds true. Let \widehat{a}_* be the estimate, associated with*

$$N = N_a \equiv \min\{N : \varepsilon \sqrt{\ln \varepsilon^{-1}} \sigma_w(N) \geq (\sqrt{3}/2) L N^{-s+1/2}\}, \quad (11)$$

where $\sigma_w(N) \equiv 2 \sum_{k=N+1}^{2N} \sigma_k^2$. Denote

$$\varphi_\varepsilon \equiv L^{\frac{2\beta-1}{2\beta+2s}} \left(\bar{\sigma} \varepsilon \sqrt{\ln \varepsilon^{-1}} \right)^{\frac{2s+1}{2\beta+2s}}. \quad (12)$$

Then there exists a constant $C = C(\beta, s)$ such that for all $0 < \varepsilon < 1$

$$R_a[\widehat{a}_*; G_s(L)] \leq C \varphi_\varepsilon, \quad \text{for any } a \in \mathbf{R}. \quad (13)$$

Our next result is the lower bound on the minimax risk.

Theorem 2 *Suppose that Assumption (A) holds. Denote*

$$\phi_\varepsilon = L^{\frac{2\beta-1}{2\beta+2s}} \left(\sigma \varepsilon \sqrt{\ln \varepsilon^{-1}} \right)^{\frac{2s+1}{2\beta+2s}}. \quad (14)$$

Then there is $c = c(\beta, s) > 0$ such that for any ε small enough the minimax risk satisfy

$$R_a^*[G_s(L)] \geq c\phi_\varepsilon.$$

Remarks:

1. When comparing the bounds of Theorems 1 and 2 we conclude that the estimator \hat{a}_* of $|a|$ is rate minimax. This result comes without much surprise. Consider, for instance, the problem of recovering the amplitude of the discontinuity in the derivative $f^{(\alpha)}$. In this case the rates φ_ε and ϕ_ε , coincide, up to a constant, with the minimax rate of estimating the function $f^{(\alpha)}$ in the uniform norm from observations (1) or (4).
2. The derived minimax rate of convergence follows from the fact that in our definition of the minimax risk the supremum is taken over all $a \in \mathbf{R}$. In particular, the lower bound of Theorem 2 is achieved on worst-case signals with jump amplitude a tending to zero as $\varepsilon \rightarrow 0$. If in the definition of the minimax risk the supremum is taken over all $|a|$ separated away from zero by a given constant, then the minimax rate of convergence is faster by the logarithmic factor.
3. In the case when an estimate of the sign of a is also required the corresponding estimator can be easily constructed. To this end, let us consider the random function

$$\widehat{M}_{N_a}(t) \equiv \Re \left(\sum_{k=N_a+1}^{2N_a} y_k e^{-2\pi i k t} \right),$$

and set the estimator \hat{s} of sign a as follows: $\hat{s} = \text{sign}(\widehat{M}_{N_a}(\hat{\tau}))$, where $\hat{\tau}$ is a global maximizer of $\widehat{J}_{N_a}(t)$. Note that \widehat{M}_{N_a} is a perturbed version of the “ideal” function

$$M_{N_a}(t) = a \Re \left(\sum_{k=N_a+1}^{2N_a} y_k \exp(-2\pi i k(t - \theta)) \right) = a \frac{\sin \pi N_a(t - \theta)}{\sin \pi(t - \theta)} \cos[(3N_a + 1)\pi(t - \theta)].$$

One can show that $P(\hat{s} \neq \text{sign } a)$ converges to zero exponentially as ε goes to zero.

The aggregate \widehat{M}_{N_a} can be of interest on its own. It can be proved that the estimator

$$\hat{a}' = N_a^{-1} \max_{t \in [0,1]} |\widehat{M}_{N_a}(t)|$$

obeys the bound of Theorem 1. This study is, however, beyond the scope of the present paper.

3.3 Change-point estimation

Estimator construction. The proposed estimation procedure is based on the representation of θ as the unique zero-crossing of the function H_N on the segment delimited by the global minimizer and the global maximizer of H_N . Therefore, the estimation algorithm is two-staged: at the first stage a localization for the “active segment” of H_N is computed

using the global extrema of the empirical function \widehat{H}_N ; then $\widehat{\theta}$ is taken as a zero of \widehat{H}_N on the localizer.

Let

$$\widehat{\theta}_- = \arg \min_{t \in [0,1]} \widehat{H}_N(t), \quad \widehat{\theta}_+ = \arg \max_{t \in [0,1]} \widehat{H}_N(t). \quad (15)$$

Consider the localizer $\mathcal{I}_\varepsilon = [\min(\widehat{\theta}_-, \widehat{\theta}_+), \max(\widehat{\theta}_-, \widehat{\theta}_+)]$. The estimator $\widehat{\theta}_N$ is defined as a root of the equation $\widehat{H}_N(\widehat{\theta}_N) = 0$ in the interval \mathcal{I}_ε :

$$\widehat{\theta}_N \in \mathcal{I}_\varepsilon : \widehat{H}_N(\widehat{\theta}_N) = 0.$$

If $\widehat{H}_N(\cdot)$ has several roots on \mathcal{I}_ε we pick any one of them, if there are none, we set $\widehat{\theta} = \widehat{\theta}_+$.

Risk bounds for $\widehat{\theta}_N$. We present here the upper bounds for the maximal risk of $\widehat{\theta}_N$. We split the admissible range $\beta > 1/2$ of Assumption (A) into two zones: $1/2 \leq \beta \leq 3/2$ and $\beta > 3/2$. According to the argument in Section 2, the first zone corresponds to estimation of a change-point or a cusp of the order $\leq 1/2$ from direct observations, while the second one covers the problems of estimating change-points in derivatives. The choice of the window parameter (the value of N) and the rate of convergence of $\widehat{\theta}_N$ are quite different in those zones. The next two theorems state the upper bounds on the maximal risks separately for these two cases. Let us denote

$$\sigma_u^2 = \sigma_u^2(N) = N^4 \sum_{k=N+1}^{2N} \sigma_k^2 + N^2 \sum_{k=N+1}^{2N} k^2 \sigma_k^2. \quad (16)$$

Due to Assumption (A), one can easily verify that for some $C_u < \infty$, $\sigma_u \leq C_u \bar{\sigma} N^{\beta+5/2}$.

Theorem 3 *Let Assumption (A) hold with $\beta > 3/2$ and let $\widehat{\theta}_*$ denote the change-point estimate $\widehat{\theta}_N$ associated with*

$$N = N_\theta \equiv \min\{N : \varepsilon \sigma_u(N) \geq 2LN^{-s+5/2}\}. \quad (17)$$

where σ_u is given in (16). Let $\varepsilon \leq 6^{-(\beta+s)} L(C_u \bar{\sigma})^{-1}$ and

$$|a| \geq cL^{\frac{2\beta-1}{2s+2\beta}} (\bar{\sigma}\varepsilon)^{\frac{2s+1}{2s+2\beta}} \sqrt{\ln \varepsilon^{-1}} \quad (18)$$

for some constant c which depends only on β, s . Then there exists a positive constant $C = C(\beta, s)$ such that

$$R_\theta[\widehat{\theta}_*; G_s(L)] \leq C|a|^{-1} L^{\frac{2\beta-3}{2\beta+2s}} (\bar{\sigma}\varepsilon)^{\frac{2s+3}{2\beta+2s}}. \quad (19)$$

Theorem 4 *Let Assumption (A) hold with $\beta \in (1/2, 3/2]$. Assume that*

$$|a| \geq cL^{\frac{2\beta-1}{2s+2\beta}} \left(\bar{\sigma}\varepsilon \sqrt{\ln \varepsilon^{-1}} \right)^{\frac{2s+1}{2s+2\beta}} \quad (20)$$

and

$$|a| \geq c' \bar{\sigma} \varepsilon \sqrt{\ln \varepsilon^{-1}}$$

for some constants $c = c(\beta)$ and $c' = c'(\beta)$. Let $C_N = C_N(\beta)$ be a certain positive constant (it is specified explicitly in the proof of the theorem), and let \widehat{a}_* be the estimator of $|a|$ defined in Section 3.2. Consider the estimator $\widehat{\theta}_*$, associated with

$$\widehat{N}_* = \left\lfloor C_N \left(\frac{\widehat{a}_*}{\bar{\sigma}\varepsilon \sqrt{\ln \varepsilon^{-1}}} \right)^{2/(2\beta-1)} \right\rfloor.$$

Then

$$R_\theta[\widehat{\theta}_*; G_s(L)] \leq C \left(\frac{\bar{\sigma}\varepsilon}{|a|} \right)^{\frac{2}{2\beta-1}} \left(\ln \varepsilon^{-1} \right)^{\frac{3-2\beta}{2(2\beta-1)}}, \quad (21)$$

where C depends on β only.

Lower bounds. We present here two minorations of the minimax risk. The first one states that if the amplitude of the jump is less than a specific threshold, namely, the minimax rate of estimation of $|a|$, then consistent estimation of θ is impossible.

Theorem 5 *Let*

$$\phi_\varepsilon = L^{\frac{2\beta-1}{2s+2\beta}} \left(\underline{\sigma}\varepsilon\sqrt{\ln\varepsilon^{-1}} \right)^{\frac{2s+1}{2s+2\beta}}$$

(cf. also (14)). Then there is $c = c(\beta, s) > 0$ such that for any $0 < |a| \leq c\phi_\varepsilon$ one holds :

$$R_\theta^*[G_s(L)] \geq c.$$

The following bounds remain valid for “large” values of $|a|$:

Theorem 6 *Let Assumption (A) hold. If $\beta > 3/2$ then for some $c = c(\beta, s)$,*

$$R_\theta^*[G_s(L)] \geq c|a|^{-1} L^{\frac{2\beta-3}{2\beta+2s}} (\underline{\sigma}\varepsilon)^{\frac{2s+3}{2s+2\beta}}.$$

If $1/2 < \beta \leq 3/2$ then for some $c' = c'(\beta)$,

$$R_\theta^*[G_s(L)] \geq c'|a|^{-1} \begin{cases} \underline{\sigma}\varepsilon(\ln\varepsilon^{-1})^{-1/2}, & \beta = 3/2 \\ (\underline{\sigma}\varepsilon)^{2/(2\beta-1)}, & 1/2 < \beta < 3/2. \end{cases}$$

Remarks:

1. Let us consider the result of Theorem 4 along with the lower bounds in Theorems 5 and 6. We observe that in the zone $1/2 < \beta \leq 3/2$ the risk of the estimator $\widehat{\theta}_*$ differs from the corresponding lower bound by a logarithmic factor. We are almost certain that there is no logarithmic factor in minimax rate (cf. the result of Korostelev (1987) for the case $\beta = 1$). In fact one can show the rate with the iterated logarithm. However, we do not know if the estimator $\widehat{\theta}_N$ can be modified to attain the minimax rate.

It is worth to mention that in the case $1/2 < \beta \leq 3/2$ the preliminary estimates $\widehat{\theta}_-$ and $\widehat{\theta}_+$ attain the rate of convergence $O([\varepsilon\sqrt{\ln\varepsilon^{-1}}]^{2/(2\beta-1)})$ which differs only by the factor $\sqrt{\ln\varepsilon^{-1}}$ from the rate, given in (21) of Theorem 4.

2. As far as the zone $\beta > 3/2$ is concerned, we observe that the upper and the lower bounds (Theorems 3 and 6) are the same up to a constant, at least in the case when $|a| \geq c\phi_\varepsilon(\ln\varepsilon^{-1})^{\frac{2\beta-1}{2(2\beta+2s)}}$ (see also Goldenshluger, Tsybakov and Zeevi (2006)). On the other hand, we have seen (cf. Theorem 5) that in the case $|a| \leq c'\phi_\varepsilon$ the consistent estimation of θ is impossible. It can be shown (and it can be readily seen from the proof of Theorem 3) that for “intermediate” values of $|a|$, i.e., for $c\phi_\varepsilon < |a| \leq c\phi_\varepsilon(\ln\varepsilon^{-1})^{\frac{2\beta-1}{2(2\beta+2s)}}$, the consistent estimation is possible, though the maximal risk in this case is larger than that in (19) by a logarithmic factor.
3. Let $\widehat{\tau} \equiv \arg \max_{t \in [0,1]} \widehat{J}_{N_\theta}(t)$ [see the proof of Theorem 1]. One can choose $\widehat{\tau}$ as an estimator of the change-point θ . It can be shown, however, that this estimator suboptimal. In particular, the maximal risk of this estimator converges to zero at the rate $\varepsilon^{(2s+5)/(4\beta+4s)}$, which is worse than the minimax rate $\varepsilon^{(2s+3)/(2\beta+2s)}$. On the other hand, the estimator $\widehat{\tau}$ can be used on the preliminary stage of $\widehat{\theta}_*$ to provide a “good” localizer $[\widehat{\tau} - \frac{1}{N_\theta}, \widehat{\tau} + \frac{1}{N_\theta}]$ for the active segment of $H_{N_\theta}(\cdot)$.

3.4 Application to derivative change–point estimation

Let us see how the main results in the previous section apply to Problem 1 from the introduction.

Suppose that the derivative $f^{(\alpha)}$ admits the representation

$$f^{(\alpha)}(t) = aV(t - \theta) + q(t), \quad t \in [0, 1],$$

where V is the saw-tooth function, and $q \in G_s(L)$ with some $s > 1/2$. The argument in Section 2 implies that the model (1) transcribes into the sequence space model (4) with the parameters (cf. (7))

$$g \in G_{s-1}(2\pi L), \quad \sigma_k^2 = (2\pi k)^{2\alpha+2}.$$

Thus, the results above apply with the substitutions $s \mapsto s-1$, $\beta = \alpha + 1$, $\underline{\sigma} = \bar{\sigma} = (2\pi)^{\alpha+1}$.

Assume, for instance, that our objective is to estimate the change-point in the first derivative of f (i.e., $\alpha = 1$), and that the nuisance regular component g belongs to $G_s(L)$. The results of the previous section (for $\beta = 2$) imply that the minimax risk of the amplitude estimation is of the order $[\varepsilon\sqrt{\ln \varepsilon^{-1}}]^{(2s-1)/(2s+2)}$, the same as the “critical” amplitude (which bounds the zone where the consistent estimation of the change-point is impossible). The minimax risk of estimating θ is $\varepsilon^{(2s+1)/(2s+2)}$ as soon as $|a| \geq O(\varepsilon^{\frac{2s-1}{2s+2}}\sqrt{\ln \varepsilon^{-1}})$. In particular, for $s = 1$ we have the rates $(\varepsilon\sqrt{\ln \varepsilon^{-1}})^{1/4}$ and $\varepsilon^{3/4}$ respectively.

If we wish to estimate the cusp of the order, say $\alpha = 1/2$, then $\beta = \alpha + 1 = 3/2$ and the minimax rate of estimating θ is ε (up to the factor $\sqrt{\ln \varepsilon^{-1}}$) independently of the smoothness of the deterministic nuisance component q . At the same time, the jump amplitude can be estimated with the accuracy $[\varepsilon\sqrt{\ln \varepsilon^{-1}}]^{\frac{2s-1}{2s+1}}$ that depends on smoothness of q . The case of direct observations (cf. Korostelev (1987)) corresponds to $\alpha = 0$, $\beta = 1$. The accuracy of the change–point estimator, proposed here, is $\varepsilon^2\sqrt{\ln \varepsilon^{-1}}$, while the jump amplitude is estimated with the rate $(\varepsilon\sqrt{\ln \varepsilon^{-1}})^{\frac{2s-1}{2s}}$.

4 Proofs of main results

In what follows C_i and c_i stand for positive constants, which can depend only on β and s , and which values are of no importance.

We start with some technical statements which will be used in the sequel.

4.1 Preliminary results

Now we are to establish uniform bounds on deviation of $\widehat{J}_N(t)$ and $\widehat{H}_N(t)$ from $J_N(t)$ and $H_N(t)$, which are in the basis of all further developments. For this purpose we introduce the following notation. For $N \geq 1$ let $\{w_N(t), t \in [0, 1]\}$ and $\{v_N(t), t \in [0, 1]\}$ be given by

$$w_N(t) = \sum_{k=N+1}^{2N} \sigma_k \xi_k e^{-2\pi ikt}, \quad v_N(t) = \sum_{k=N+1}^{2N} k \sigma_k \xi_k e^{-2\pi ikt}. \quad (22)$$

Clearly, $\{w_N(t)\}$ and $\{v_N(t)\}$ are zero mean complex–valued stationary Gaussian processes with variances

$$\sigma_w^2 \equiv \sigma_w^2(N) = 2 \sum_{k=N+1}^{2N} \sigma_k^2, \quad \sigma_v^2 \equiv \sigma_v^2(N) = 2 \sum_{k=N+1}^{2N} k^2 \sigma_k^2. \quad (23)$$

Let $\lambda \geq 1$, B be a subinterval of $[0, 1]$, $B \subseteq [0, 1]$, we define

$$\mathcal{A}_J(\lambda; N, B) \equiv \{\omega \in \Omega : \sup_{t \in B} |w_N(t)| \leq 2\lambda\sigma_w\}, \quad (24)$$

$$\mathcal{A}_H(\lambda; N, B) \equiv \mathcal{A}_J(\lambda; N, B) \cap \{\omega \in \Omega : \sup_{t \in B} |v_N(t)| \leq 2\lambda\sigma_v\}. \quad (25)$$

In all what follows we write $\mathcal{A}_J(\lambda; N)$ and $\mathcal{A}_H(\lambda; N)$ for $\mathcal{A}_J(\lambda; N, [0, 1])$ and $\mathcal{A}_H(\lambda; N, [0, 1])$ respectively. The probability of events $\mathcal{A}_J(\lambda; N, B)$ and $\mathcal{A}_H(\lambda; N, B)$ are easily controlled; see Lemma 6 in Appendix.

The next two statements establish upper bounds on estimation accuracy of $\widehat{J}_N(t)$ and $\widehat{H}_N(t)$ when these events occur.

Proposition 1 *Assume that $g \in G_s(L)$. If $\omega \in \mathcal{A}_J(\lambda; N, B)$ then*

$$\sup_{t \in B} |\widehat{J}_N(t) - J_N(t)| \leq \rho_J(N) + 8\varepsilon^2 \lambda^2 \sigma_w^2 + 4\varepsilon \lambda |a| N \sigma_w,$$

where

$$\begin{aligned} \rho_J(N) &\equiv 2 \left(\sum_{k=N+1}^{2N} |g_k| \right)^2 + 2|a|N \sum_{k=N+1}^{2N} |g_k| \\ &\leq 6L^2 N^{-2s+1} + 2\sqrt{3}|a|LN^{-s+3/2}. \end{aligned} \quad (26)$$

Proof: By definition of $\widehat{J}_N(t)$ and in view of (4) we have

$$\begin{aligned} \widehat{J}_N(t) &= a^2 \sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} \exp\{-2\pi i(k-j)(t-\theta)\} \\ &\quad + \sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} g_k \bar{g}_j \exp\{-2\pi i(k-j)t\} \\ &\quad + \varepsilon^2 \sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} \sigma_k \xi_k \bar{\sigma}_j \bar{\xi}_j \exp\{-2\pi i(k-j)t\} \\ &\quad + 2\Re \left(\sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} a e^{2\pi i k \theta} \bar{g}_j \exp\{-2\pi i(k-j)t\} \right) \\ &\quad + 2\Re \left(\sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} a e^{2\pi i k \theta} \varepsilon \bar{\sigma}_j \bar{\xi}_j \exp\{-2\pi i(k-j)t\} \right) \\ &\quad + 2\Re \left(\sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} \varepsilon g_k \bar{\sigma}_j \bar{\xi}_j \exp\{-2\pi i(k-j)t\} \right) \\ &= J_N(t) + \sum_{l=1}^5 I_J^{(l)}(t), \end{aligned}$$

and now we derive upper bounds on the terms $I_J^{(l)}(t)$, $l = 1, \dots, 5$.

It is immediately seen that

$$|I_J^{(1)}(t)| \leq \left(\sum_{k=N+1}^{2N} |g_k| \right)^2, \quad |I_J^{(3)}(t)| \leq 2|a|N \sum_{k=N+1}^{2N} |g_k|.$$

On the set $\mathcal{A}_J(\lambda; N, B)$

$$I_J^{(2)}(t) = \varepsilon^2 |w_N(t)|^2 \leq 4\varepsilon^2 \lambda^2 \sigma_w^2, \quad |I_J^{(4)}(t)| \leq 2\varepsilon |a|N |w_N(t)| \leq 4|a|\varepsilon \lambda \sigma_w.$$

Furthermore, noticing that $|I_J^{(5)}(t)| \leq |I_J^{(1)}(t)| + |I_J^{(2)}(t)|$ we finally obtain that

$$\begin{aligned} |\widehat{J}_N(t) - J_N(t)| &\leq 2 \left(\sum_{k=N+1}^{2N} |g_k| \right)^2 + 8\varepsilon^2 \lambda^2 \sigma_w^2 + 2|a|N \sum_{k=N+1}^{2N} |g_k| + 4|a|\varepsilon \lambda \sigma_w \\ &= \rho_J(N) + 8\varepsilon^2 \lambda^2 \sigma_w^2 + 4|a|\varepsilon \lambda \sigma_w, \end{aligned}$$

as claimed. Inequality (26) is a straightforward consequence of Lemma 7 of Appendix. \blacksquare

Proposition 2 *Assume that $g \in G_s(L)$. If $\omega \in \mathcal{A}_H(\lambda; N, B)$ then*

$$\sup_{t \in B} |\widehat{H}_N(t) - H_N(t)| \leq \rho_H(N) + 32\pi \varepsilon^2 \lambda^2 \sigma_w \sigma_v + 16\pi |a| \varepsilon \lambda \sigma_u, \quad (27)$$

where $\sigma_u \equiv \sigma_u(N) = N^2\sigma_w + N\sigma_v$, and

$$\begin{aligned}\rho_H(N) &\equiv 8\pi \sum_{k=N+1}^{2N} k|g_k| \sum_{j=N+1}^{2N} |g_j| + 16\pi|a|N^2 \sum_{j=N+1}^{2N} |g_j|. \\ &\leq 32\pi L^2 N^{-2s+2} + 32\pi L|a|N^{-s+5/2}.\end{aligned}\quad (28)$$

Proof: By definition of $\widehat{H}_N(t)$ we have

$$\begin{aligned}\widehat{H}_N(t) &= 2\pi ia^2 \sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} (k-j) \exp(-2\pi i(k-j)(t-\theta)) \\ &\quad + 2\pi i \sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} (k-j) g_k \bar{g}_j \exp(-2\pi i(k-j)t) \\ &\quad + 2\pi i \varepsilon^2 \sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} (k-j) \sigma_k \bar{\sigma}_j \xi_k \bar{\xi}_j \exp(-2\pi i(k-j)t) \\ &\quad + 4\pi \Im \left(\sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} (k-j) a \exp(2\pi i k \theta) \bar{g}_j \exp(-2\pi i(k-j)t) \right) \\ &\quad + 4\pi \varepsilon \Im \left(\sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} (k-j) a \exp(2\pi i k \theta) \bar{\sigma}_j \bar{\xi}_j \exp(-2\pi i(k-j)t) \right) \\ &\quad + 4\pi \varepsilon \Im \left(\sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} (k-j) g_k \bar{\sigma}_j \bar{\xi}_j \exp(-2\pi i(k-j)t) \right) \\ &\equiv H_N(t) + \sum_{l=1}^5 I_H^{(l)}(t),\end{aligned}$$

and we bound $I_H^{(l)}(t)$, $l = 1, \dots, 5$ from above.

First we note that

$$\begin{aligned}|I_H^{(1)}(t)| &\leq 4\pi \sum_{k=N+1}^{2N} k|g_k| \sum_{j=N+1}^{2N} |g_j| \\ |I_H^{(3)}(t)| &\leq 4\pi|a| \sum_{k=N+1}^{2N} k \sum_{j=N+1}^{2N} |g_j| + 4\pi|a|N \sum_{j=N+1}^{2N} j|g_j| \\ &\leq 2\pi|a|N \left[(3N+1) \sum_{j=N+1}^{2N} |g_j| + 2 \sum_{j=N+1}^{2N} j|g_j| \right] \leq 16\pi|a|N^2 \sum_{j=N+1}^{2N} |g_j|.\end{aligned}$$

Furthermore, $|I_H^{(5)}(t)| \leq |I_H^{(1)}(t)| + |I_H^{(2)}(t)|$ and on the set $\mathcal{A}_H(\lambda; N, B)$

$$\begin{aligned}|I_H^{(2)}(t)| &\leq 4\pi \varepsilon^2 |w_N(t)| |v_N(t)| \leq 16\pi \varepsilon^2 \lambda^2 \sigma_v^2 \sigma_w^2 \\ |I_H^{(4)}(t)| &\leq 4\pi \varepsilon |a| \left(|w_N(t)| \sum_{k=N+1}^{2N} k + N |v_N(t)| \right) \\ &\leq 16\pi \varepsilon \lambda |a| (N^2 \sigma_w + N \sigma_v) = 16\pi \varepsilon \lambda |a| \sigma_u.\end{aligned}$$

Combining all these inequalities we obtain (27). Inequality (28) is an immediate consequence of Lemma 7. \blacksquare

Under Assumption (A), $\sigma_w^2(N)$, $\sigma_v^2(N)$ and $\sigma_u^2(N)$ admit the following bounds in terms of N

$$c_w^2 \underline{\sigma}^2 N^{2\beta+1} \leq \sigma_w^2(N) \leq C_w^2 \bar{\sigma}^2 N^{2\beta+1}, \quad (29)$$

$$c_v^2 \underline{\sigma}^2 N^{2\beta+3} \leq \sigma_v^2(N) \leq C_v^2 \bar{\sigma}^2 N^{2\beta+3} \quad (30)$$

$$c_u^2 \underline{\sigma}^2 N^{2\beta+5} \leq \sigma_u^2(N) \leq C_u^2 \bar{\sigma}^2 N^{2\beta+5}. \quad (31)$$

Here the constants $c_w, C_w, c_v, C_v, c_u, C_u$ depend on β only and can be easily computed explicitly. In what follows all these constants are regarded as known because the sequence (σ_k) is given.

The next proposition establishes a bound on the accuracy of preliminary estimators $\widehat{\theta}_+$ and $\widehat{\theta}_-$ (see (15)).

Proposition 3 *Let*

$$\Delta_H(\lambda; N) \equiv \rho_H(N) + 16\pi|a|\varepsilon\lambda\sigma_u + 32\pi\varepsilon^2\lambda^2\sigma_w\sigma_v,$$

where $\rho_H(N)$ is given in (28). Let $N \geq 6$ and $\lambda \geq 1$ be such that

$$\Delta_H(\lambda; N) < \frac{a^2 N^3}{4}. \quad (32)$$

Then for all $\omega \in \mathcal{A}_H(\lambda; N)$

$$\theta \leq \hat{\theta}_+ \leq \theta + \frac{4}{5N}, \quad \text{and} \quad \theta - \frac{4}{5N} \leq \hat{\theta}_- \leq \theta.$$

Proof: Assume that $\mathcal{A}_H(\lambda; N)$ holds; then by Proposition 2

$$\sup_{t \in [0,1]} |\hat{H}_N(t) - H_N(t)| \leq \Delta_H(\lambda; N). \quad (33)$$

We prove the statement of the proposition by contradiction. Define

$$t_+ = \arg \max_{t \in [0,1]} H_N(t), \quad t_- = \arg \min_{t \in [0,1]} H_N(t).$$

By Lemma 8 in Appendix, $t_+ \in [\theta, \theta + \frac{1}{N}]$ and $t_- \in [\theta - \frac{1}{N}, \theta]$. It is sufficient to prove the statement of the proposition for $\hat{\theta}_+$; the proof for $\hat{\theta}_-$ is identical in every detail.

First, assume that $\theta + \frac{1}{N} < \hat{\theta}_+ < \theta + 1 - \frac{1}{N}$. We have

$$\begin{aligned} \hat{H}_N(t_+) &\leq \hat{H}_N(\hat{\theta}_+) \leq H_N(\hat{\theta}_+) + \Delta_H(\lambda; N) \\ &= H_N(t_+) + \Delta_H(\lambda; N) + [H_N(\hat{\theta}_+) - H_N(t_+)], \end{aligned}$$

where the first inequality follows from definition of $\hat{\theta}_+$, and the second one is a consequence of (33). Now we observe that by Lemma 8(iv) and by the origin of $U_N(t)$

$$H_N(t_+) - \max_{1/N < t < 1 - (1/N)} H_N(t) = a^2 \left[U_N(t_+) - \max_{1/N < t < 1 - (1/N)} U_N(t) \right] \geq \frac{1}{2} a^2 N^3$$

so that

$$\hat{H}_N(t_+) \leq H_N(t_+) + \Delta_H(\lambda; N) - \frac{1}{2} a^2 N^3 < H_N(t_+) - \frac{1}{4} a^2 N^3, \quad (34)$$

where the last inequality is in view of (32). Thus we have that

$$|\hat{H}_N(t_+) - H_N(t_+)| > \frac{1}{4} a^2 N^3, \quad \forall \omega \in \mathcal{A}_H(\lambda; N),$$

which contradicts (33) and (32). Hence $\hat{\theta}_+ \notin (\theta + 1/N, \theta + 1 - 1/N)$ whenever $\mathcal{A}_H(\lambda; N)$ holds.

Now assume that $\hat{\theta}_+ \in [\theta + (1 - \eta)/N, \theta + 1/N]$. In this case, by the same reasoning and by Lemma 8(iv) with choice $\eta = 1/5$ we obtain that

$$\begin{aligned} H_N(t_+) - \max_{t \in [\theta + (1-\eta)N^{-1}, \theta + N^{-1}]} H_N(t) &\geq a^2 \left[U_N(t_+) - \max_{t \in [4/(5N), 1/N]} U_N(t) \right] \\ &\geq a^2 \left(\frac{3}{2} N^3 - N^3 \right) = \frac{1}{2} a^2 N^3; \end{aligned}$$

recall that $U_N(\cdot)$ is defined in (10). Therefore the same inequality (34) leads to the contradiction. It is shown similarly that on the set $\mathcal{A}_H(\lambda; N)$, $\widehat{\theta}_+$ cannot lie in the interval $(\theta - \frac{1}{N}, \theta)$; here we use the fact that

$$H_N(t_+) - \max_{t \in [\theta - 1/N, \theta]} H_N(t) = H_N(t_+) \geq 3a^2 N^3$$

because $\max_{t \in [\theta - \frac{1}{N}, \theta]} H_N(t) = 0$. ■

The bound of Proposition 3 indicates that on the set $\mathcal{A}_H(\lambda; N)$ the preliminary estimates $\widehat{\theta}_+$ and $\widehat{\theta}_-$ are within distance $4/(5N)$ from the target value θ . In other words, the preliminary estimates localize properly the target value when $\mathcal{A}_H(\lambda; N)$ occurs. Choosing λ in an appropriate way, we will ensure that the event $\mathcal{A}_H(\lambda; N)$ will be of "large" probability. This will allow to control the probability of "proper localization":

$$P\left\{|\widehat{\theta}_- - \theta| \vee |\widehat{\theta}_+ - \theta| > \frac{4}{5N}\right\} \leq P\left\{\mathcal{A}_H^c(\lambda; N)\right\}, \quad \forall 1 \leq \lambda \leq \lambda_{\max}(N),$$

where

$$\lambda_{\max}(N) \equiv \max\{\lambda \geq 1 \mid \Delta_H(\lambda; N) \leq \frac{1}{4}a^2 N^3\}.$$

4.2 Proof of Theorem 1

Let $\widehat{\tau} \equiv \arg \max_{t \in [0, 1]} \widehat{J}_{N_a}(t)$; then

$$|\widehat{a}_*^2 - a^2| = N_a^{-2} |\widehat{J}_{N_a}(\widehat{\tau}) - J_{N_a}(\theta)|. \quad (35)$$

The following bounds on N_a follows from (11) and (29)

$$\left(\sqrt{3}(2C_w\overline{\sigma})^{-1} \frac{L}{\varepsilon\sqrt{\ln \varepsilon^{-1}}}\right)^{1/(\beta+s)} \leq N_a \leq \left(\sqrt{3}(2c_w\underline{\sigma})^{-1} \frac{L}{\varepsilon\sqrt{\ln \varepsilon^{-1}}}\right)^{1/(\beta+s)} + 1. \quad (36)$$

For any $\lambda \geq 1$ we have on the set $\mathcal{A}_J(\lambda; N_a)$

$$\widehat{J}_{N_a}(\widehat{\tau}) \geq \widehat{J}_{N_a}(\theta) \geq J_{N_a}(\theta) - \Delta_J(\lambda; N_a) \quad (37)$$

$$J_{N_a}(\theta) \geq J_{N_a}(\widehat{\tau}) \geq \widehat{J}_{N_a}(\widehat{\tau}) - \Delta_J(\lambda; N_a), \quad (38)$$

where

$$\Delta_J(\lambda; N_a) \equiv \rho_J(N_a) + 8\varepsilon^2 \lambda^2 \sigma_w^2(N_a) + 4\varepsilon \lambda |a| N_a \sigma_w(N_a). \quad (39)$$

These inequalities follow from definition of $\widehat{\tau}$, Proposition 1, and properties of function J_{N_a} [see Lemma 8 (ii)]. We obtain from (37) and (38) that for all $\lambda \geq 1$

$$|\widehat{J}_{N_a}(\widehat{\tau}) - J_{N_a}(\theta)| \leq \Delta_J(\lambda; N_a), \quad \forall \omega \in \mathcal{A}_J(\lambda; N_a). \quad (40)$$

Therefore

$$\begin{aligned} E|\widehat{J}_{N_a}(\widehat{\tau}) - J_{N_a}(\theta)|^2 &= E|\widehat{J}_{N_a}(\widehat{\tau}) - J_{N_a}(\theta)|^2 \mathbf{1}\{\mathcal{A}_J(\lambda; N_a)\} \\ &\quad + E|\widehat{J}_{N_a}(\widehat{\tau}) - J_{N_a}(\theta)|^2 \mathbf{1}\{\mathcal{A}_J^c(\lambda; N_a)\} \\ &\equiv P_1 + P_2, \end{aligned}$$

where as shown above, $P_1 \leq \Delta_J^2(\lambda; N_a)$.

Our current goal is to bound P_2 . Using the fact that θ is the point of global maximum of function $J_{N_a}(\cdot)$ we have

$$\widehat{J}_{N_a}(\widehat{\tau}) - J_{N_a}(\theta) = \widehat{J}_{N_a}(\widehat{\tau}) - J_{N_a}(\widehat{\tau}) + J_{N_a}(\widehat{\tau}) - J_{N_a}(\theta) \leq \widehat{J}_{N_a}(\widehat{\tau}) - J_{N_a}(\widehat{\tau}).$$

On the other hand, by definition of $\hat{\tau}$ we have

$$\hat{J}_{N_a}(\hat{\tau}) - J_{N_a}(\theta) = \hat{J}_{N_a}(\hat{\tau}) - \hat{J}_{N_a}(\theta) + \hat{J}_{N_a}(\theta) - J_{N_a}(\theta) \geq \hat{J}_{N_a}(\theta) - J_{N_a}(\theta).$$

Combining two last inequalities we obtain

$$\begin{aligned} |\hat{J}_{N_a}(\hat{\tau}) - J_{N_a}(\theta)| &\leq |\hat{J}_{N_a}(\hat{\tau}) - J_{N_a}(\hat{\tau})| \vee |\hat{J}_{N_a}(\theta) - J_{N_a}(\theta)| \\ &\leq \sup_{t \in [0,1]} |\hat{J}_{N_a}(t) - J_{N_a}(t)|. \end{aligned}$$

Therefore

$$P_2 \leq 4E \sup_{t \in [0,1]} |\hat{J}_{N_a}(t) - J_{N_a}(t)|^2 \mathbf{1}\{\mathcal{A}_J^c(\lambda; N_a)\}.$$

Using the same reasoning as in the proof of Proposition 1 we can write

$$\sup_{t \in [0,1]} |\hat{J}_{N_a}(t) - J_{N_a}(t)| \leq 2\rho_J(N_a) + 2\varepsilon^2 \sup_{t \in [0,1]} |w_{N_a}(t)|^2 + 2\varepsilon|a|N_a \sup_{t \in [0,1]} |w_{N_a}(t)|.$$

Denote $\zeta = \sup_{t \in [0,1]} |w_{N_a}(t)|$; then

$$\begin{aligned} E \sup_{t \in [0,1]} |\hat{J}_{N_a}(t) - J_{N_a}(t)|^2 \mathbf{1}\{\mathcal{A}_J^c(\lambda; N_a)\} \\ \leq 16\rho_J^2(N_a)P\{\mathcal{A}_J^c(\lambda; N_a)\} + 16\varepsilon^4 E[\zeta^4 \mathbf{1}\{\mathcal{A}_J^c(\lambda; N_a)\}] + 8\varepsilon^2|a|^2 N_a^2 E[\zeta^2 \mathbf{1}\{\mathcal{A}_J^c(\lambda; N_a)\}] \\ \equiv P_3 + P_4 + P_5. \end{aligned}$$

By Lemma 6 in Appendix, and by definition of N_a we obtain

$$P_3 \leq 16\Delta_J^2(\lambda; N_a)P\{\mathcal{A}_J^c(\lambda; N_a)\} \leq c_1\Delta_J^2(\lambda; N_a)\lambda N_a \exp\{-2\lambda^2\}.$$

Furthermore,

$$\begin{aligned} P_4 &= 16\varepsilon^4 E[\zeta^4 \mathbf{1}\{\zeta > 2\lambda\sigma_w\}] = 8\varepsilon^4 \int_{(2\lambda\sigma_w)^4}^{\infty} P\{\zeta^4 > t\} dt \\ &= 64\varepsilon^4 (2\lambda\sigma_w)^4 \int_1^{\infty} t^3 P\{\zeta > 2\lambda\sigma_w t\} dt \leq c_2\varepsilon^4 (\lambda\sigma_w)^4 \lambda N_a \int_1^{\infty} t^4 \exp\{-2\lambda^2 t^2\} dt \\ &= c_3\varepsilon^4 \sigma_w^4 N_a \int_{\lambda}^{\infty} t^4 e^{-2t^2} dt \leq c_4\varepsilon^4 \sigma_w^4 N_a \lambda^3 \exp\{-2\lambda^2\}, \end{aligned}$$

and

$$\begin{aligned} P_5 &\leq 8\varepsilon^2|a|^2 N_a^2 \int_{(2\lambda\sigma_w)^2}^{\infty} P\{\zeta^2 > t\} dt \\ &= 16\varepsilon^2|a|^2 N_a^2 (2\lambda\sigma_w)^2 \int_1^{\infty} t P\{\zeta > 2\lambda\sigma_w t\} dt \leq c_4\varepsilon^2|a|^2 N_a^3 (\lambda\sigma_w)^2 \lambda \int_1^{\infty} t^2 \exp\{-2\lambda^2 t^2\} dt \\ &\leq c_5\varepsilon^2|a|^2 N_a^3 \sigma_w^2 \int_{\lambda}^{\infty} t^2 e^{-2t^2} dt \leq c_6\varepsilon^2|a|^2 N_a^3 \sigma_w^2 \lambda \exp\{-2\lambda^2\}. \end{aligned}$$

Thus we get

$$\begin{aligned} P_2 &\leq \sum_{i=3}^5 P_i \\ &\leq c_7\lambda \exp(-2\lambda^2) \left(N_a \Delta_J^2(\lambda; N_a) + \varepsilon^4 N_a \sigma_w^4 \lambda^2 + \varepsilon^2 |a|^2 N_a^3 \sigma_w^2 \right) \\ &\leq c_8\lambda \exp\{-2\lambda^2\} \left(N_a \Delta_J^2(\lambda; N_a) + \varepsilon^4 N_a^{4\beta+3} \lambda^2 + \varepsilon^2 |a|^2 N_a^{2\beta+4} \right), \end{aligned}$$

where the last inequality is a consequence of (29). Choosing $\lambda = \lambda_* \equiv \sqrt{c_9 \ln \varepsilon^{-1}}$ with sufficiently large constant c_9 depending on β , s , and combining the above inequalities we obtain

$$\begin{aligned} \left\{ E|\widehat{a}_*^2 - a^2|^2 \right\}^{1/2} &\leq N_a^{-2} \left\{ E|\widehat{J}_{N_a}(\widehat{\tau}) - J_{N_a}(\theta)|^2 \right\}^{1/2} \\ &\leq N_a^{-2} [\Delta_J^2(\lambda_*; N_a) + P_2]^{1/2} \leq c_{10} (|a|\varphi_\varepsilon + \varphi_\varepsilon^2), \end{aligned}$$

where φ_ε is defined in (12). The last inequality follows from the fact that by choice of c_9 , $\Delta_J^2(\lambda_*; N_a)$ dominates P_2 , and in view of (39), (26), and definitions of N_a and φ_ε

$$\Delta_J(\lambda_*; N_a) \leq c_{11} (\varepsilon^2 \lambda_*^2 \sigma_w^2(N_a) + \varepsilon \lambda_* |a| N_a \sigma_w(N_a)) = c_{12} (\varphi_\varepsilon^2 + |a|\varphi_\varepsilon).$$

To get (13) we use the elementary inequality $\sqrt{|x-y|} \geq |\sqrt{|x|} - \sqrt{|y|}|$ and the fact that $|\widehat{a}_*^2 - a^2| \geq |\widehat{a}_* - |a|| \cdot |a|$, thus coming to

$$R_a[\widehat{a}_*; G_s(L)] \leq c_{13} \left[\varphi_\varepsilon + (|a|^{-1} \varphi_\varepsilon^2) \wedge (|a|\varphi_\varepsilon)^{1/2} \right] \leq c_{14} \varphi_\varepsilon. \quad \blacksquare$$

4.3 Proof of Theorem 2.

Let for $0 < \varkappa \leq 1$ $\psi = \varkappa\phi$ and

$$N = \max \{ N \in \mathbf{N} : \phi^2 N^{2s+1} \leq L^2 \}, \quad (41)$$

so that

$$\left(\frac{L^2}{\underline{\sigma}^2 \varepsilon^2 \ln \varepsilon^{-1}} \right)^{\frac{1}{2\beta+2s}} - 1 < N \leq \left(\frac{L^2}{\underline{\sigma}^2 \varepsilon^2 \ln \varepsilon^{-1}} \right)^{\frac{1}{2\beta+2s}}.$$

Consider the observations

$$y_k = s_k^{(j)} + \varepsilon \sigma_k \xi_k, \quad (42)$$

where (ξ_k) , $k = 0, 1, \dots$ are standard complex valued i.i.d. random variables (i.e., $(\Re \xi_k, \Im \xi_k)^T \sim N(0, I)$). Here $N+1$ signals $s^{(j)}$, $j = 0, \dots, N$ are defined as follows:

$$s_k^{(j)} = \psi e^{2\pi i k \theta_j} \mathbf{1}(k \geq N+1) = \psi e^{2\pi i k \theta_j} - g_k^{(j)}, \quad k = 0, 1, \dots, \quad \text{for } j = 1, \dots, N$$

where $\theta_j = j/N$, and $s^{(0)} \equiv 0$.

Note first that the sequences $(g_k^{(j)})$,

$$g_k^{(j)} = \psi e^{2\pi i k \theta_j} \mathbf{1}(0 \leq k \leq N), \quad j = 1, \dots, N,$$

all belongs to $G_s(L)$. Indeed,

$$\sum_{k=0}^{\infty} |g_k^{(j)}|^2 k^{2s} = \psi^2 \sum_{k=0}^N k^{2s} \leq \psi^2 N^{2s+1} \leq L^2$$

by (41).

Let now P_j , $j = 1, \dots, N$ stand for the distributions of observations (42) which correspond to signals from the family $S = \{s^{(j)}, j = 1, \dots, N\}$, and let π be the uniform prior probability on S : $\pi_j = \frac{1}{N}$. Let now P_π denote the Bayes measure for the prior π : $P_\pi = \frac{1}{N} \sum_{j=1}^N P_j$. We also denote $Z_j = \frac{dP_j}{dP_0}$, $j = 1, \dots, N$ and $Z_\pi = \frac{dP_\pi}{dP_0}$ the corresponding likelihood ratios. Now consider the minimax risk R_ε of a $N+1$ -point estimation problem:

$$R_\varepsilon = \sup_{j=0, \dots, N} E_j(\psi_j - \widehat{\psi})^2,$$

where $\psi_j = \psi$, $j = 1, \dots, N$ and $\psi_0 = 0$.

Lemma 1 $R_\varepsilon \geq \frac{\psi^2}{4} \left(1 - \frac{E(Z_\pi - 1)^2}{2}\right)$.

The proof of the lemma is presented in the appendix. Now the result of Theorem 2 is an immediate consequence of the following statement:

Proposition 4 *Let $\varkappa > 0$ be a small enough absolute constant. Then*

$$E_0(Z_\pi - 1)^2 \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0.$$

Proof: Note that the likelihood ratio

$$\begin{aligned} Z_j &= \exp\left(-\sum_{k=N+1}^{\infty} \frac{|y_k - s_k^{(j)}|^2 - |y_k|^2}{2\varepsilon^2\sigma_k^2}\right) \\ &= \exp\left(\sum_{k=N+1}^{\infty} \frac{s_k^{(j)}\overline{\xi_k} + \overline{s_k^{(j)}}\xi_k}{2\varepsilon\sigma_k} - \frac{|s^{(j)}|^2}{2\varepsilon^2\sigma_k^2}\right) \\ &= \exp\left(\sum_{k=N+1}^{\infty} \frac{\psi(\eta_k \cos 2\pi k\theta_j + \zeta_k \sin 2\pi k\theta_j)}{\varepsilon\sigma_k} - \frac{\psi^2}{2\varepsilon^2\sigma_k^2}\right) \end{aligned}$$

where (η_k) and (ζ_k) are the sequences of i.i.d. standard gaussian random variables, $(\eta_k, \zeta_k)^T \sim N(0, I)$. Then

$$Z_\pi = \frac{1}{N} \sum_{j=1}^N Z_j.$$

For obvious reason $E_0 Z_\pi = 1$. Let us compute $E_0(Z_\pi^2)$. We have

$$\begin{aligned} E_0(Z_\pi^2) &= \frac{1}{N^2} \sum_{j=1}^N \sum_{\ell=1}^N E \exp\left(\sum_{k=N+1}^{\infty} \frac{\psi}{\varepsilon\sigma_k} [\eta_k(\cos 2\pi k\theta_j + \cos 2\pi k\theta_\ell) \right. \\ &\quad \left. + \zeta_k(\sin 2\pi k\theta_j + \sin 2\pi k\theta_\ell)] - \frac{\psi^2}{\varepsilon^2\sigma_k^2}\right) \\ &= \frac{1}{N^2} \sum_{j=1}^N \sum_{\ell=1}^N \exp\left(\sum_{k=N+1}^{\infty} \frac{\psi^2}{\varepsilon^2\sigma_k^2} \cos 2\pi k(\theta_j - \theta_\ell)\right) \\ &= \frac{1}{N^2} \sum_{j=1}^N \sum_{\ell=1}^N \exp\left(\sum_{k=N+1}^{\infty} \frac{\psi^2}{\varepsilon^2\sigma_k^2} \cos \frac{2\pi k(j - \ell)}{N}\right) \\ &= \frac{1}{N} \sum_{j=0}^{N-1} \exp\left(\sum_{k=N+1}^{\infty} \frac{\psi^2}{\varepsilon^2\sigma_k^2} \cos \frac{2\pi kj}{N}\right), \end{aligned}$$

by definition of θ_j .

Lemma 2 *Let $0 \leq x \leq n/9$. Then $e^x \leq \sum_{j=0}^{n-1} \frac{x^j}{j!} + \frac{e^{-n}}{\sqrt{2\pi n}}$.*

The proof is postponed till the appendix.

Using the result of Lemma 2 we obtain for $n \geq 9\frac{\psi^2}{\varepsilon^2} \sum_{k=N+1}^{\infty} \sigma_k^{-2}$ the bound:

$$\begin{aligned} E(Z_\pi)^2 &\leq 1 + \frac{1}{N} \sum_{j=0}^{N-1} \sum_{\ell=1}^{n-1} \frac{1}{\ell!} \left[\sum_{k=N+1}^{\infty} \frac{\psi^2}{\varepsilon^2\sigma_k^2} \cos \frac{2\pi kj}{N} \right]^\ell + \frac{e^{-n}}{\sqrt{2\pi n}} \\ &\equiv 1 + \frac{e^{-n}}{\sqrt{2\pi n}} + \sum_{\ell=1}^{n-1} \frac{I_\varepsilon^{(\ell)}}{\ell!}. \end{aligned} \tag{43}$$

Note that

$$I_\varepsilon^{(1)} = \frac{\psi^2}{N\varepsilon^2} \sum_{k=N+1}^{\infty} \sigma_k^{-2} \sum_{j=0}^{N-1} \cos \frac{2\pi k j}{N}.$$

However, for $k \neq rN$, $r \in \mathbf{Z}$, $\sum_{j=0}^{N-1} \cos \frac{2\pi k j}{N} = 0$. Indeed,

$$2 \sin \frac{\pi k}{N} \sum_{j=0}^{N-1} \cos \frac{2\pi k j}{N} = \sin \frac{2\pi k(N-1/2)}{N} + \sin \frac{\pi k}{N} = 0.$$

Thus

$$\begin{aligned} I_\varepsilon^{(1)} &\leq \frac{\psi^2}{\varepsilon^2} \sum_{k \geq N+1, k=rN, r \in \mathbf{Z}} \sigma_k^{-2} \leq \frac{\psi^2}{\underline{\sigma}^2 \varepsilon^2} \sum_{k=1}^{\infty} (N+kN)^{-2\beta} \\ &\leq \frac{\psi^2}{\underline{\sigma}^2 \varepsilon^2 N^{2\beta}} \sum_{k=1}^{\infty} (1+k)^{-2\beta} \leq \frac{\psi^2}{\underline{\sigma}^2 \varepsilon^2 N^{2\beta}} (2\beta-1)^{-1}. \end{aligned}$$

The treatment of the generic term $I_\varepsilon^{(\ell)}$ is a bit more involved. We have

$$I_\varepsilon^{(\ell)} = \frac{1}{N} \sum_{j=0}^{N-1} \sum_{\mathbf{k} \geq N+1} \left(\frac{\psi^2}{\varepsilon^2} \right)^\ell \frac{\cos \frac{2\pi k_1 j}{N}}{\sigma_{k_1}^2} \cdots \frac{\cos \frac{2\pi k_\ell j}{N}}{\sigma_{k_\ell}^2},$$

where $\mathbf{k} = (k_1, \dots, k_\ell)$ is the multi-index. Observe that

$$\prod_{l=1}^{\ell} \cos \frac{2\pi k_l j}{N} = 2^{-\ell} \sum_{e \in \{-1,1\}^\ell} \cos \frac{2\pi e^T \mathbf{k} j}{N}.$$

Now we have for any \mathbf{k} such that $e^T \mathbf{k} \neq rN$, $r \in \mathbf{Z}$:

$$\sum_{j=0}^{N-1} \prod_{l=1}^{\ell} \cos \frac{2\pi k_l j}{N} = 2^{-\ell} \sum_{e \in \{-1,1\}^\ell} \sum_{j=0}^{N-1} \cos \frac{2\pi e^T \mathbf{k} j}{N} = 0.$$

In other words, the non-vanishing terms correspond to such \mathbf{k} that

$$\sum_{l=1}^{\ell} e_l k_l + rN = 0, \quad r \in \mathbf{Z}.$$

Therefore, the sum for $I_\varepsilon^{(\ell)}$ can be rewritten as follows:

$$\begin{aligned} I_\varepsilon^{(\ell)} &= \frac{1}{N} \sum_{j=0}^{N-1} \sum_{\mathbf{k} \geq N+1} \left(\frac{\psi^2}{\varepsilon^2} \right)^\ell \prod_{l=1}^{\ell} \sigma_{k_l}^{-2} 2^{-\ell} \sum_{e \in \{-1,1\}^\ell} \cos \frac{2\pi e^T \mathbf{k} j}{N} \\ &\leq 2^{-\ell} \left(\frac{\psi^2}{\varepsilon^2} \right)^\ell \sum_{e \in \{-1,1\}^\ell} \sum_{\substack{\mathbf{k} \geq N+1, r \in \mathbf{Z} \\ e^T \mathbf{k} = rN}} \prod_{l=1}^{\ell} \sigma_{k_l}^{-2} \\ &\leq 2^{-\ell} \left(\frac{\psi^2}{\varepsilon^2 \underline{\sigma}^2} \right)^\ell \sum_{e \in \{-1,1\}^\ell} \sum_{\mathbf{k} \geq N+1, r \in \mathbf{Z}, e^T \mathbf{k} = rN} \prod_{l=1}^{\ell} k_l^{-2\beta} \end{aligned} \tag{44}$$

We approximate the last sum in the RHS of (44) with an integral: note that

$$\begin{aligned}
& \sum_{k_1=N+1}^{\infty} \dots \sum_{k_\ell=N+1}^{\infty} \prod_{l=1}^{\ell} k_l^{-2\beta} \mathbf{1}\{e^T \mathbf{k} = rN\} \\
&= N^{-2\beta\ell} \sum_{k_1=1}^{\infty} \dots \sum_{k_\ell=1}^{\infty} \prod_{l=1}^{\ell} \left(1 + \frac{k_l}{N}\right)^{-2\beta} \mathbf{1}\left\{\frac{e^T \mathbf{k}}{N} = r - \ell\right\} \\
&\leq C \frac{N^{\ell-1}}{N^{2\beta\ell} \sqrt{\ell}} \int_0^\infty \dots \int_0^\infty \frac{dx_1}{(1+x_1)^{2\beta}} \dots \frac{dx_\ell}{(1+x_\ell)^{2\beta}} \delta(e^T x - r + \ell),
\end{aligned}$$

where $x = (x_1, \dots, x_\ell)^T$ (we have used the fact that the unit volume of the hyperplane $e^T x = a$ contains $O(\frac{N}{\sqrt{\ell}})$ points). We conclude that

$$I_\varepsilon^{(\ell)} \leq \frac{C}{N} \left(\frac{\psi^2}{\varepsilon^2 N^{2\beta-1} \underline{\sigma}^2} \right)^\ell 2^{-\ell} \sum_{e \in \{-1,1\}^\ell} J_\ell(e), \quad (45)$$

where

$$J_\ell(e) = \frac{1}{\sqrt{\ell}} \sum_{m=-\infty}^{\infty} \int_0^\infty \dots \int_0^\infty \frac{dx_1}{(1+x_1)^{2\beta}} \dots \frac{dx_\ell}{(1+x_\ell)^{2\beta}} \delta(e^T x - m),$$

Lemma 3 *There is $C(\beta) < \infty$ such that $J_\ell(e) \leq C^\ell(\beta)$.*

The proof is given in the appendix.

When substituting the result of Lemma 3 into (45) we obtain for $I_\varepsilon^{(\ell)}$ the bound

$$I_\varepsilon^{(\ell)} \leq \frac{1}{N} \left(\frac{C(\beta)\psi^2}{\varepsilon^2 N^{2\beta-1} \underline{\sigma}^2} \right)^\ell \quad (46)$$

Putting (46) into (43) results in

$$\begin{aligned}
E(Z_\pi^2) &\leq 1 + \frac{e^{-n}}{\sqrt{2\pi n}} + \sum_{\ell=1}^{n-1} \frac{1}{N\ell!} \left(\frac{C(\beta)\psi^2}{\underline{\sigma}^2 \varepsilon^2 N^{2\beta-1}} \right)^\ell \\
&\leq 1 + \frac{e^{-n}}{\sqrt{2\pi n}} + N^{-1} \exp\left(\frac{C(\beta)\psi^2}{\underline{\sigma}^2 \varepsilon^2 N^{2\beta-1}} \right).
\end{aligned} \quad (47)$$

By the choice of $\phi = \phi_\varepsilon$ in (14) and (41),

$$\frac{\psi^2}{\underline{\sigma}^2 \varepsilon^2 N^{2\beta-1}} = \varkappa^2 \frac{\psi^2}{\underline{\sigma}^2 \varepsilon^2 N^{2\beta-1}} \leq \varkappa^2 \ln \varepsilon^{-1} \leq C(s, \beta) \varkappa^2 \ln N$$

for ε small enough. Now the bound (47) implies that

$$\begin{aligned}
E(Z_\pi - EZ_\pi)^2 &= EZ_\pi^2 - 1 \leq \frac{e^{-n}}{\sqrt{2\pi n}} + N^{-1} \left[\exp\left(\frac{C(\beta)\psi^2}{\underline{\sigma}^2 \varepsilon^2 N^{2\beta-1}} \right) - 1 \right] \\
&\leq \frac{e^{-n}}{\sqrt{2\pi n}} + \frac{\exp(C'(s, \beta) \varkappa^2 \ln N) - 1}{N} \rightarrow 0
\end{aligned}$$

as $\varepsilon \rightarrow 0$ if $\varkappa^2 C'(s, \beta) < 1$, what implies the proposition. ■

4.4 Proof of Theorem 3

Step 1. The following bounds on N_θ defined in (17) are easily derived using (31):

$$\left[2(C_u\bar{\sigma})^{-1}\frac{L}{\varepsilon}\right]^{1/(\beta+s)} \leq N_\theta \leq \left[2(c_u\underline{\sigma})^{-1}\frac{L}{\varepsilon}\right]^{1/(\beta+s)} + 1. \quad (48)$$

Let

$$\lambda_* = c_1|a|\varepsilon^{-1}N_\theta^{-\beta+1/2}, \quad (49)$$

where $c_1 = c_1(\beta, s, \underline{\sigma}, \bar{\sigma})$ is a fixed constant. We note that $\lambda_* \geq 1$ due to (18) with appropriate constant c . We have

$$E|\widehat{\theta}_* - \theta|^2 = E|\widehat{\theta}_* - \theta|^2 \mathbf{1}\{\mathcal{A}_H(\lambda_*; N_\theta)\} + E|\widehat{\theta}_* - \theta|^2 \mathbf{1}\{\mathcal{A}_H^c(\lambda_*; N_\theta)\}, \quad (50)$$

and our current goal is to bound the two terms on the RHS. From (48), (49) and Lemma 6 we obtain

$$\begin{aligned} E|\widehat{\theta}_* - \theta|^2 \mathbf{1}\{\mathcal{A}_H^c(\lambda_*; N_\theta)\} &\leq P\{\mathcal{A}_H^c(\lambda_*; N_\theta)\} \\ &\leq c_2\lambda_*N_\theta \exp\{-2\lambda_*^2\} \\ &\leq c_3|a|L^{-\frac{2\beta-3}{2\beta+2s}}\varepsilon^{-\frac{2s+3}{2\beta+2s}} \exp\{-c_4a^2\varepsilon^{-\frac{2s+1}{\beta+s}}L^{-\frac{2\beta-1}{\beta+s}}\}. \end{aligned} \quad (51)$$

Step 2. Now our goal is to bound $E|\widehat{\theta}_* - \theta|^2 \mathbf{1}\{\mathcal{A}_H(\lambda_*; N_\theta)\}$ from above. By definition of N_θ , and due to (48), (18), and (49) one can choose constants c in (18) and c_1 in (49) depending on $s, \beta, \underline{\sigma}$, and $\bar{\sigma}$ only so that

$$\begin{aligned} 16\pi|a|\varepsilon\sigma_u(N_\theta) &\geq 32\pi L|a|N_\theta^{-s+5/2}, \\ 32\pi\varepsilon^2\lambda_*^2\sigma_w(N_\theta)\sigma_v(N_\theta) &\geq 32\pi L^2N_\theta^{-2s+2}, \\ 16\pi|a|\varepsilon\lambda_*\sigma_u(N_\theta) &\geq 32\pi\varepsilon^2\lambda_*^2\sigma_w(N_\theta)\sigma_v(N_\theta). \end{aligned}$$

Therefore, Proposition 2 implies that on the set $\mathcal{A}_H(\lambda_*; N_\theta)$ we have

$$\sup_{t \in [0,1]} |\widehat{H}_{N_\theta}(t) - H_{N_\theta}(t)| \leq \Delta_H(\lambda_*; N_\theta) \leq 64\pi|a|\varepsilon\lambda_*\sigma_u(N_\theta). \quad (52)$$

Furthermore,

$$64\pi|a|\varepsilon\lambda_*\sigma_u(N_\theta) \leq 64\pi C_u\bar{\sigma}c_1a^2N_\theta^3 < \frac{1}{4}a^2N_\theta^3 \quad (53)$$

where the last inequality is ensured by choice of the constant c_1 . The condition $\varepsilon \leq 6^{-(\beta+s)}L(C_u\bar{\sigma})^{-1}$ guarantees that $N_\theta \geq 6$. Therefore Proposition 3 can be applied. Using (52) and (53) we obtain by Proposition 3 that $|\widehat{\theta}_* - \theta| \leq 4/(5N_\theta)$ whenever $\mathcal{A}_H(\lambda_*; N_\theta)$ holds.

Now we argue that on the set $\mathcal{A}_H(\lambda_*; N_\theta)$ function $\widehat{H}_{N_\theta}(\cdot)$ necessarily has a zero in \mathcal{I}_ε . Indeed,

$$\mathcal{A}_H(\lambda_*; N_\theta) \subseteq \left\{ \omega : \sup_{t \in [0,1]} |\widehat{H}_{N_\theta}(t) - H_{N_\theta}(t)| \leq 64\pi|a|\varepsilon\lambda_*\sigma_u(N_\theta) \right\}. \quad (54)$$

On the other hand, if D denotes the event that $\widehat{H}_{N_\theta}(\cdot)$ has no zeros in \mathcal{I}_ε , then

$$D \subseteq \left\{ \omega : \sup_{t \in [0,1]} |\widehat{H}_{N_\theta}(t) - H_{N_\theta}(t)| \geq \max_{t \in [0,1]} |H_{N_\theta}(t)| \right\}. \quad (55)$$

By Lemma 8 (iv), $\max_{t \in [0,1]} |H_{N_\theta}(t)| \geq \frac{3}{2}a^2N_\theta^3$, while $64\pi|a|\varepsilon\lambda_*\sigma_u(N_\theta) < a^2N_\theta^3/4$. This shows that the sets on the RHS of (54) and (55) are disjoint, so that $\widehat{H}_{N_\theta}(\cdot)$ must vanish in \mathcal{I}_ε if $\mathcal{A}_H(\lambda_*; N_\theta)$ occurs. Thus on the set $\mathcal{A}_H(\lambda_*; N_\theta)$ we have

$$|H_{N_\theta}(\widehat{\theta}_*)| \leq 64\pi|a|\varepsilon\lambda_*\sigma_u(N_\theta) \quad (56)$$

because $\widehat{H}_{N_\theta}(\widehat{\theta}_*) = 0$, and in view of (52). Furthermore, (31) and (49) imply that

$$64\pi|a|\varepsilon\lambda_*\sigma_u(N_\theta)\left(\frac{5}{4}a^2N_\theta^4\right)^{-1} \leq \frac{1}{5N_\theta}.$$

This along with (56) and Lemma 8(v) leads to the inequality $|\widehat{\theta}_* - \theta| \leq 1/(5N_\theta)$.

Let $\delta \in (0, 1/(5N_\theta))$; then

$$E|\widehat{\theta}_* - \theta|^2 1\{\mathcal{A}_H(\lambda_*; N_\theta)\} \leq \delta^2 + \sum_{l=1}^{l_0} \delta^2 2^{2l} P\{(\delta 2^{l-1} \leq |\widehat{\theta}_* - \theta| \leq \delta 2^l) \cap \mathcal{A}_H(\lambda_*; N_\theta)\}, \quad (57)$$

where $l_0 = \min\{l : \delta 2^l > 1/(5N_\theta)\}$. Let $B_l \equiv \{t : \delta 2^{l-1} \leq |t - \theta| \leq \delta 2^l\}$, $l = 1, \dots, l_0$; then

$$\begin{aligned} & P\{(\delta 2^{l-1} \leq |\widehat{\theta}_* - \theta| \leq \delta 2^l) \cap \mathcal{A}_H(\lambda_*; N_\theta)\} \\ & \leq P\{(|H_{N_\theta}(\widehat{\theta}_*)| \geq \frac{5}{4}a^2N_\theta^4\delta 2^{l-1}) \cap \mathcal{A}_H(\lambda_*; N_\theta)\} \\ & \leq P\{(\sup_{t \in B_l} |H_{N_\theta}(t) - \widehat{H}_{N_\theta}(t)| \geq \frac{5}{4}a^2N_\theta^4\delta 2^{l-1}) \cap \mathcal{A}_H(\lambda_*; N_\theta)\} \\ & = P\{(\sup_{t \in B_l} |H_{N_\theta}(t) - \widehat{H}_{N_\theta}(t)| \geq 64\pi|a|\varepsilon\sigma_u(N_\theta)\lambda_l) \cap \mathcal{A}_H(\lambda_*; N_\theta)\} \end{aligned}$$

where the first inequality follows from the properties of H_{N_θ} and because $\widehat{H}_{N_\theta}(\widehat{\theta}) = 0$, and

$$\lambda_l \equiv \frac{5}{4}a^2N_\theta^4\delta 2^{l-1}[64\pi|a|\varepsilon\sigma_u(N_\theta)]^{-1}, \quad l = 1, 2, \dots, l_0. \quad (58)$$

Choosing

$$\delta = \delta_* \equiv \frac{64\pi|a|\varepsilon\sigma_u(N_\theta)}{\frac{5}{4}a^2N_\theta^4} = c_6|a|^{-1}\varepsilon N_\theta^{\beta-3/2} = c_7|a|^{-1}L^{\frac{2\beta-3}{2\beta+2s}}\varepsilon^{\frac{2s+3}{2\beta+2s}}$$

and taking into account that $2^{l_0-1} \leq (5N_\theta\delta_*)^{-1}$, we get that $\lambda_l \leq \lambda_*$, $\forall l = 1, \dots, l_0$. This along with Lemma 6 implies that

$$\begin{aligned} P\{\sup_{B_l} |H_{N_\theta}(t) - \widehat{H}_{N_\theta}(t)| \geq 64\pi|a|\varepsilon\sigma_u(N_\theta)\lambda_l\} & \leq P\{\mathcal{A}_H^c(\lambda_l; N_\theta, B_l)\} \\ & \leq c_8\lambda_l|B_l|N_\theta \exp\{-2\lambda_l^2\}. \end{aligned}$$

Hence

$$\begin{aligned} & \sum_{l=1}^{l_0} \delta^2 2^{2l} P\{(\delta 2^{l-1} \leq |\widehat{\theta}_* - \theta| \leq \delta 2^l) \cap \mathcal{A}_H(\lambda_*; N_\theta)\} \\ & \leq \delta^2 \sum_{l=1}^{l_0} 2^{2l} \lambda_l |B_l| N_\theta \exp\{-2\lambda_l^2\} \\ & \leq \delta^2 (\delta N_\theta) \sum_{l=1}^{l_0} 2^{3l-1} \exp\{-2^{2l-1}\} \leq c_9 \delta^2. \end{aligned}$$

Combining the last inequality with (57) we finally obtain

$$E|\widehat{\theta}_* - \theta|^2 1\{\mathcal{A}_H(\lambda_*; N_\theta)\} \leq c_{10}\delta^2. \quad (59)$$

We complete the proof using (59), (58), (51), (50) and taking into account (18). \blacksquare

4.5 Proof of Theorem 4

We start with the following statement:

Step 1. Let \widehat{a}_* be the estimate of $|a|$ defined in Section 3.2. Under conditions of Theorem 1 for any $\lambda \geq 1$ one has

$$P\left\{\left|\widehat{a}_* - |a|\right| \geq \frac{\Delta_J(\lambda; N_a)}{|a|N_a^2}\right\} \leq c_1\lambda N_a \exp\{-2\lambda^2\}, \quad (60)$$

where $\Delta_J(\lambda; N)$ is defined in (39), and c_1 is a constant that may depend on β , $\bar{\sigma}$ and $\underline{\sigma}$ only. Indeed, the inequality (60) follows immediately from the proof of Theorem 1 [cf. (35) and (40)], and Lemma 6.

Step 2. Let $\lambda_* = \sqrt{\eta \ln \varepsilon^{-1}}$ where $\eta \geq 4/(2\beta - 1)$. Define

$$K \equiv \min \left\{ (32\pi C_w C_v \bar{\sigma}^2)^{-1/(2\beta-1)}, (16\pi C_u \bar{\sigma})^{-2/(2\beta-1)} \right\},$$

and let

$$N_* = \left\lfloor K \left(\frac{|a|}{64\varepsilon\lambda_*} \right)^{2/(2\beta-1)} \right\rfloor.$$

First we show that $\Delta_H(\lambda_*; N_*) < a^2 N_*^3/16$, provided that (20) is valid with appropriate constant c . Indeed, by definition of N_* we have

$$32\pi(\varepsilon\lambda_*)^2 \sigma_w(N_*) \sigma_v(N_*) \leq 32\pi(\varepsilon\lambda_*)^2 C_w C_v \bar{\sigma}^2 N_*^{2\beta+2} \leq a^2 N_*^3/64,$$

where the last inequality follows by choice of K . Similarly we verify that

$$16\pi|a|\varepsilon\lambda_* \sigma_u(N_*) \leq 16\pi|a|\varepsilon\lambda_* C_u \bar{\sigma} N_*^{\beta+5/2} \leq a^2 N_*^3/64$$

again in view of our choice of K . The inequalities

$$32\pi L^2 N_*^{-2s+2} \leq a^2 N_*^3/64, \quad 32\pi L|a| N_*^{-s+5/2} \leq a^2 N_*^3/64$$

are implied by the condition (20) with appropriate constant c . Combining these bounds we obtain that $\Delta_H(\lambda_*; N_*) < a^2 N_*^3/16$ as claimed. This means that if N_* could be taken as the window parameter of the change-point estimate then all conditions of Proposition 3 would be fulfilled. However, this choice of the window parameter is not feasible because $|a|$ is unknown.

Step 3. Let \hat{a}_* be the estimate of $|a|$ defined in Section 3.2. Recall that the estimate \hat{a}_* is associated with the window parameter N_a given in (11). We set

$$\hat{N}_* = \left\lfloor K \left(\frac{\hat{a}_*}{64\varepsilon\lambda_*} \right)^{2/(2\beta-1)} \right\rfloor.$$

Now assume that $\omega \in \mathcal{A}_J(\lambda_*; N_a)$. In view of (60),

$$\left| \frac{|\hat{a}_*|}{|a|} - 1 \right| \leq \gamma \equiv \frac{\Delta_J(\lambda_*; N_a)}{|a|^2 N_a^2}, \quad \forall \omega \in \mathcal{A}_J(\lambda_*; N_a).$$

Using (39), (36) and (20) we obtain that

$$\gamma \leq c_2 \frac{\varepsilon\lambda_* |a| N_a^{\beta+3/2}}{a^2 N_a^2} = c_3(\varepsilon\lambda_*) |a|^{-1} N_a^{\beta-1/2} \leq c_4 |a|^{-1} L^{\frac{2\beta-1}{2\beta+2s}} (\varepsilon\lambda_*)^{\frac{2s+1}{2s+2\beta}} \leq 1/2$$

where the last inequality follows from (20) with appropriate constant c . Further, because

$$\frac{1}{64} \hat{a}_* (\varepsilon\lambda_*)^{-1} = \frac{1}{64} |a| (\varepsilon\lambda_*)^{-1} \left[1 + \left(\frac{\hat{a}_*}{|a|} - 1 \right) \right]$$

we have that $N_*(1 - \gamma)^{2/(2\beta-1)} \leq \hat{N}_* \leq N_*(1 + \gamma)^{2/(2\beta-1)}$ on the set $\mathcal{A}_J(\lambda_*; N_a)$, which, in turn, implies that

$$\left(\frac{1}{2} \right)^{2/(2\beta-1)} N_* \leq \hat{N}_* \leq \left(\frac{3}{2} \right)^{2/(2\beta-1)} N_*, \quad \forall \omega \in \mathcal{A}_J(\lambda_*; N_a).$$

Therefore, by the same computation as for $\Delta_H(\lambda_*, N_*)$ we obtain that

$$\Delta_H(\lambda_*; \hat{N}_*) < \frac{a^2 \hat{N}_*^3}{8}, \quad \forall \omega \in \mathcal{A}_J(\lambda_*; N_a).$$

Let $N_0 \equiv (1/2)^{2/(2\beta-1)} N_*$, $N_1 \equiv (3/2)^{2/(2\beta-1)} N_*$, and define

$$\mathcal{B} \equiv \mathcal{B}_H(\lambda_*) \cap \mathcal{A}_J(\lambda_*; N_a), \quad \mathcal{B}_H(\lambda_*) \equiv \bigcap_{n=N_0}^{N_1} \mathcal{A}(\lambda_*; N).$$

We have

$$\begin{aligned} P\{\mathcal{B}^c\} &\leq P\{\mathcal{B}_H^c(\lambda_*)\} + P\{\mathcal{A}_J^c(\lambda; N_a)\} \\ &\leq \sum_{N=N_0}^{N_1} P\{\mathcal{A}_H^c(\lambda_*; N)\} + P\{\mathcal{A}_J^c(\lambda; N_a)\} \leq c_5 \lambda_* (N_*^2 + N_a) \exp\{-2\lambda^2\}, \end{aligned}$$

where the last inequality follows from definition of N_0 , N_1 and from Lemma 6.

The further proof goes along the same lines as the proof of Theorem 3 with the event \mathcal{B} playing the role of $\mathcal{A}_H(\lambda_*; N_\theta)$. Any $\eta \geq 4/(2\beta - 1)$ guarantees that the contribution of the error on the set \mathcal{B} to the risk will be small. \blacksquare

4.6 Proof of Theorem 5

We provide the proof of the theorem for the particular case $\sigma_k^2 = (2\pi k)^{2\beta}$, $\beta \in \mathbf{N}^+$. The proof for the general case is much more technical and is based on the same ideas as that of Theorem 2.

Consider the operation of multiple integration, i.e. for $\beta \in \mathbf{N}^+$

$$f^{(-\beta)}(t) = \int_0^t f^{(-\beta+1)}(s) ds = \frac{1}{\beta!} \int_0^t (t-s)^{\beta-1} f(s) ds$$

Note that $f_k^{(-\beta)} = (-2\pi i k)^{-\beta} f_k$ for any periodic function $f \in L_2[0, 1]$.

Now, consider the following construction: let $\ell = \lfloor s \rfloor + 1$ (here $\lfloor a \rfloor$ stands for the largest integer $\leq a$). Consider a function $\phi : [0, 1] \rightarrow \mathbf{R}$ which is ℓ times continuously differentiable and $\|\phi^{(\ell)}\|_\infty \leq C$. Furthermore, we require

$$\int \phi(t) dt = 1, \quad \int t^k \phi(t) dt = 0, \quad k = 1, 2, \dots, \beta.$$

Now, consider for some $\psi > 0$, $N \in \mathbf{N}$, $N > 1$ the function $g_N(t) = \psi[N\phi(Nt) - 1]$.

Let

$$s_N(t) = \psi[\delta(t) - 1] - g_N(t) = \psi[\delta(t) - N\phi(Nt)].$$

We start with the technical result:

Lemma 4 *The signal s_N possesses the following properties:*

- 1) there is $c_0 > 0$ such that for any $N \geq 1$, $g_N \in G(s, L)$ if $\psi \leq c_0 L N^{-s+1/2}$;
- 2) s_N has $\beta + 1$ vanishing moments:

$$\int_0^{\frac{1}{N}} t^k s_N(t) dt = 0, \quad k = 0, \dots, \beta.$$

We leave to the reader the proof of this simple statement.

Let S be the family of translations of signal s_N :

$$s^{(j)}(t) = s_N\left(t - \frac{j}{N}\right), \quad j = 1, \dots, N-1.$$

With some abuse of notations we denote $s_k^{(j)}$, $k \geq 0$ the Fourier coefficients of $s^{(j)}$. Note that

$$s_k^{(j)} = \psi e^{2\pi i k \theta_j} - g_k^{(j)}, \quad k = 1, 2, \dots, \quad \text{for } j = 1, \dots, N$$

where $\theta_j = j/N$.

We consider the problem of estimation of the shift parameter θ_j of the signal $s^{(j)}$ from the noisy observations (y_k) , $k = 0, 1, 2, \dots$ of its Fourier coefficients as in (42). Recall that

$$y_k = s_k^{(j)} + \varepsilon \sigma_k \xi_k,$$

where (ξ_k) , $k = 0, 1, \dots$ are standard complex valued i.i.d. random variables (i.e., $(\Re \xi_k, \Im \xi_k)^T \sim N(0, I)$), $\sigma_k^2 = (2\pi k)^{2\beta}$, $\beta \in \mathbf{N}^+$.

Let, as above, P_j , $j = 0, \dots, N-1$ stand for the distributions of observations (42) which correspond to signals from the family $S = \{s^{(j)}, j = 1, \dots, N\}$, and let π be the uniform prior probability on S . We denote P_π the Bayes measure for the prior π , $Z_j = \frac{dP_j}{dP_0}$, $j = 1, \dots, N$ and $Z_\pi = \frac{dP_\pi}{dP_0}$.

Now consider the minimax risk R_ε of a N-point estimation problem:

$$R_\varepsilon = \sup_{j=1, \dots, N} E_j(\theta_j - \hat{\theta})^2.$$

Lemma 5 *For any $0 < \delta < 1$,*

$$R_\varepsilon \geq \frac{1}{3} P_0(Z_\pi \geq 1 - \delta) - \delta - \frac{E_0 \left(\sum_{j=1}^N j Z_j \right)^2}{N^4(1 - \delta)}.$$

The proof of this result is put in the appendix.

Now the statement of the theorem follows from the following result:

Proposition 5 *Let*

$$N = \left(\frac{L^2}{\varepsilon^2 \ln \varepsilon^{-1}} \right)^{\frac{1}{2\beta+2s}},$$

and let $\psi = c\phi_\varepsilon$, $c > 0$ being a small enough absolute constant, where ϕ_ε is defined in (14). Then

$$E_0(Z_\pi - 1)^2 \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0; \tag{61}$$

$$N^{-4} E_0 \left(\sum_{j=1}^N j Z_j \right)^2 - \frac{1}{4} \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0. \tag{62}$$

Proof: Observe that the observation model (42) is equivalent to the following white-noise model:

$$dY_t = f^{(-\beta)}(t)dt + \varepsilon dW_t, \quad t \in [0, 1],$$

where W_t is a complex-valued Wiener process. Recall that by 2) of Lemma 4, all moments of s_N (and those of $s^{(j)}$) are vanishing up to the order β . If we denote $u^{(j)}(t) = K s^{(j)}(t)$, then $\text{supp } u^{(j)} \subseteq [\frac{j}{N}, \frac{j+1}{N}]$, and $u^{(j)}$ and $u^{(\ell)}$ have disjoint supports. Furthermore, by construction of $s^{(j)}$, $u^{(j)}(t) \leq C\psi N^{-\beta+1}$, and

$$\|u^{(j)}\|_2^2 \leq C^2 \psi^2 N^{-2\beta+1}. \tag{63}$$

The likelihood ratio

$$Z_j = \exp \left(\varepsilon^{-1} \int_0^1 u^{(j)}(t) dW_t - \frac{\int_0^1 (u^{(j)})^2(t) dt}{2\varepsilon^2} \right) = \exp \left(\varepsilon^{-1} \mu \eta_j - \frac{\mu^2}{2\varepsilon^2} \right),$$

where η_j , $j = 1, \dots, N$ are i.i.d. normal random variables, $E\eta_1 = 0$, $E\eta_1^2 = 1$, and $\mu = \|u^{(1)}\|_2$. Note that the likelihood ratios Z_j and $Z_{j'}$ are independent when $j \neq j'$.

Now

$$E(Z_\pi - 1)^2 = \frac{1}{N^2} \left(\sum_{j=1}^N Z_j - EZ_j \right)^2 = \frac{1}{N} E(Z_1^2 - E(Z_1)^2) = \frac{1}{N} e^{\mu^2 \varepsilon^{-2}} \rightarrow 0$$

when, for instance, $\mu^2 \varepsilon^{-2} < \frac{\ln N}{2}$. Further,

$$\begin{aligned} E \left(\sum_{j=1}^N j Z_j \right)^2 &= \left(\sum_{j=1}^N j \right)^2 + E \left(\sum_{j=1}^N j (Z_j - 1) \right)^2 \\ &= \frac{N^2(N+1)^2}{4} + \sum_{j=1}^N j^2 e^{\mu^2 \varepsilon^{-2}} = \frac{N^4}{4} (1 + o(1)) \end{aligned}$$

if $\mu^2 \varepsilon^{-2} < \frac{\ln N}{2}$. When taking into account the bound (63), we conclude that the relations (61) and (62) of Proposition 5 hold true when

$$\psi^2 = c^2 N^{2\beta-1} \varepsilon^2 \ln N,$$

for some absolute constant $c > 0$. ■

4.7 Proof of Theorem 6

Consider the following 2-point testing problem: given observations $y_k = f_k + \varepsilon \sigma_k \xi_k$ we would like to discriminate between two hypotheses

$$\begin{aligned} H_0 : f_k &= f_k^{(0)} \equiv a, \quad \forall k \in \mathbf{N}^+, \quad \text{and} \\ H_1 : f_k &= f_k^{(1)} \equiv a e^{2\pi i k h} + g_k^{(1)}, \quad \forall k \in \mathbf{N}^+ \end{aligned}$$

where $h > 0$, and

$$g_k^{(1)} \equiv \begin{cases} a(1 - e^{2\pi i k h}), & 0 < k \leq n \\ 0, & k > n \end{cases}$$

for some integer n to be chosen in the sequel. The hypotheses correspond to the model (4) with $a^{(0)} = a^{(1)} = a$, $\theta^{(0)} = 0$, $\theta^{(1)} = h$, $g_k^{(0)} = 0, \forall k \in \mathbf{N}^+$, and $g_k^{(1)}$ as defined above.

We will select n in such a way that $(g_k^{(1)})$ belongs to $G_s(L)$. We have

$$\sum_{k=1}^{\infty} |g_k^{(1)}|^2 k^{2s} \leq a^2 \sum_{k=1}^n |1 - e^{2\pi i k h}|^2 k^{2s} \leq c_1 a^2 \min\{h^2 n^{2s+3}, n^{2s+1}\},$$

where c_1 depends on s only. Choosing

$$n = n_* \equiv c_2 (L|a|^{-1} h^{-1})^{2/(2s+3)} \tag{64}$$

we obtain that $(g_k^{(1)}) \in G_s(L)$, provided that $n_* \leq h^{-1}$.

Let P_0 and P_1 denote the probability measures associated with observations (y_k) in model (4) with $(f_k) = (f_k^{(0)})$ and $(f_k) = (f_k^{(1)})$ respectively. The Kullback–Leibler divergence between these measures is

$$\begin{aligned}\mathcal{K}(P_0, P_1) &= \sum_{k=1}^{\infty} \frac{1}{2\varepsilon^2\sigma_k^2} |f_k^{(0)} - f_k^{(1)}|^2 \\ &\leq \frac{a^2}{\varepsilon^2\sigma_{\min}^2} \sum_{k>n_*}^{\infty} k^{-2\beta} |1 - e^{2\pi i k h}|^2 \\ &\leq \frac{8a^2}{\underline{\sigma}^2\varepsilon^2} \left(\pi^2 h^2 \sum_{k=n_*+1}^{\lfloor (1/\pi h) \rfloor} k^{-2\beta+2} + \sum_{k=\lfloor (1/\pi h) \rfloor+1}^{\infty} k^{-2\beta} \right).\end{aligned}\quad (65)$$

First assume that $\beta > 3/2$. Choosing

$$h = c_3 \varepsilon^2 a^{-2} n_*^{2\beta-3} \Leftrightarrow h = c_4 |a|^{-1} L^{\frac{2\beta-3}{2s+2\beta}} \varepsilon^{\frac{2s+3}{2s+2\beta}}$$

we see that $n_* \leq h^{-1}$ in view of (65), and

$$\mathcal{K}(P_0, P_1) \leq c_5 \frac{a^2}{\varepsilon^2} \left[h^2 n_*^{-2\beta+3} + h^{2\beta-1} \right] \leq c_6 < \infty$$

for ε small enough. On the other hand, $|\theta^{(0)} - \theta^{(1)}| = h$; hence by the standard argument (see, e.g., Tsybakov (2004, Theorem 2.2)) $\sup_{g \in G_s(L)} E|\hat{\theta} - \theta|^2 \geq c_7 h$. This completes the proof for the case $\beta > 3/2$.

If $\beta = 3/2$ then we have from (65) that $\mathcal{K}(P_0, P_1) \leq c_8 a^2 \varepsilon^{-2} h^2 \ln h^{-1}$; hence the choice $h = c_9 |a|^{-1} \varepsilon (\ln \varepsilon^{-1})^{-1/2}$ implies that $n_* \leq h^{-1}$ and guarantees the boundedness of $\mathcal{K}(P_0, P_1)$. This leads to the announced result. If $1/2 < \beta < 3/2$ then instead of (64) we can choose $n_* = c_{10} (L/|a|)^{\frac{2}{2s+1}}$ and the second term on the RHS of (65) is dominant. Thus, $\mathcal{K}(P_0, P_1) \leq c_{11} a^2 \varepsilon^{-2} h^{2\beta-1}$ and choosing $h = c_{11} |a|^{-1} \varepsilon^{2/(2\beta-1)}$, we complete the proof of the theorem. \blacksquare

Appendix

Lemma 6 *Let $B \subseteq [0, 1]$ be a subinterval of $[0, 1]$, and $|B|$ its Lebesgue measure. Then there exists constants C_1 and C_2 depending on β , $\underline{\sigma}$ and $\bar{\sigma}$ only such that*

$$P\{\sup_{t \in B} |v_N(t)| \geq \lambda\} \leq C_1 |B| N^{\beta+5/2} \lambda \sigma_v^{-2} \exp\{-\lambda^2/(2\sigma_v^2)\}, \quad \forall \lambda \geq 2\sigma_v \quad (66)$$

$$P\{\sup_{t \in B} |w_N(t)| \geq \lambda\} \leq C_2 |B| N^{\beta+3/2} \lambda \sigma_w^{-2} \exp\{-\lambda^2/(2\sigma_w^2)\}, \quad \forall \lambda \geq 2\sigma_w. \quad (67)$$

Here $\{w_N(t), t \in [0, 1]\}$, $\{v_N(t), t \in [0, 1]\}$, σ_w and σ_v are defined in (22) and (23) respectively. Moreover, for all $N \geq 1$ and $\lambda \geq 1$ one has

$$\begin{aligned}P\{\mathcal{A}_H^c(\lambda; N, B)\} &\leq C_3 \lambda |B| N \exp\{-2\lambda^2\}, \\ P\{\mathcal{A}_J^c(\lambda; N, B)\} &\leq C_4 \lambda |B| N \exp\{-2\lambda^2\},\end{aligned}$$

where events $\mathcal{A}_J(\lambda; N, B)$ and $\mathcal{A}_H(\lambda; N, B)$ are defined in (24) and (25).

Proof: In the proof c_1, c_2, \dots stand for positive constants that may depend on β , $\underline{\sigma}$ and $\bar{\sigma}$ only. We use the general exponential inequality of Talagrand; see, e.g., Wellner and

van der Vaart (1996, Proposition A.2.2). Clearly, $E|v_N(t)|^2 \leq \sigma_v^2$, and $E|w_N(t)|^2 \leq \sigma_w^2$. Further, for $t, s \in [0, 1]$

$$\begin{aligned} r^2(s, t) &\equiv E|v_N(t) - v_N(s)|^2 = E \left| \sum_{k=N+1}^{2N} k \sigma_k \xi_k (e^{-2\pi i k t} - e^{-2\pi i k s}) \right|^2 \\ &\leq 4\pi^2 \sum_{k=N+1}^{2N} k^4 |\sigma_k|^2 |t - s|^2 \leq c_1 |t - s|^2 N^{2\beta+5}. \end{aligned}$$

Then the minimal number of balls of radius ν in the seminorm $r(\cdot, \cdot)$ covering the index set $B \subseteq [0, 1]$ does not exceed $c_2 |B| N^{\beta+5/2} \nu^{-1}$. Therefore applying Proposition A.2.2 from van der Vaart and Wellner (1996) [in their notation we put $K \sim |B| N^{\beta+5/2}$, $\varepsilon_0 = \sigma_v \sim N^{\beta+3/2}$] we obtain for any $\lambda \geq 2\sigma_v$ that

$$P\{\sup_{t \in B} |v_N(t)| \geq \lambda\} \leq c_3 \lambda |B| N^{\beta+5/2} \sigma_v^{-2} \exp\{-\lambda^2 / (2\sigma_v^2)\}.$$

Similarly, for $\{w_N(t); t \in [0, 1]\}$ we have for $\lambda \geq 2\sigma_w$

$$P\{\sup_{t \in B} |w_N(t)| \geq \lambda\} \leq c_4 \lambda |B| N^{\beta+3/2} \sigma_w^{-2} \exp\{-\lambda^2 / (2\sigma_w^2)\}.$$

It follows from (66), (67) and (29), (30) that for any $\lambda \geq 1$

$$\begin{aligned} P\{\sup_{t \in B} |v_N(t)| \geq 2\lambda \sigma_v\} &\leq c_5 \lambda |B| N \exp(-2\lambda^2) \\ P\{\sup_{t \in B} |w_N(t)| \geq 2\lambda \sigma_w\} &\leq c_6 \lambda |B| N \exp(-2\lambda^2). \end{aligned}$$

■

Lemma 7 *Let $N \geq 1$. If $g \in G_s(L)$, $s > -1/2$ then*

$$\sum_{k=N+1}^{2N} |g_k| \leq \sqrt{3} L N^{-s+1/2}, \quad \sum_{k=N+1}^{2N} k |g_k| \leq 2 L N^{-s+3/2}. \quad (68)$$

Proof: By the Cauchy–Schwarz inequality

$$\sum_{k=N+1}^{2N} |g_k| \leq \left[\sum_{k=N+1}^{\infty} |g_k|^2 k^{2s} \right]^{1/2} \left[\sum_{k=N+1}^{2N} k^{-2s} \right]^{1/2} \leq L \left[\sum_{k=N+1}^{2N} k^{-2s} \right]^{1/2}.$$

We obtain

$$\sum_{k=N+1}^{2N} k^{-2s} = \sum_{k=1}^{2N} k^{-2s} - \sum_{k=1}^N k^{-2s} \leq 2N^{1-2s} \frac{2^{1-2s} - 1}{1 - 2s} \leq 3N^{1-2s}$$

for all $s \in (-1/2, 1/2)$, $\sum_{k=N+1}^{2N} k^{-1} \leq 2 \ln 2$, and

$$\begin{aligned} \sum_{k=N+1}^{2N} k^{-2s} &= \sum_{k=N+1}^{\infty} k^{-2s} - \sum_{k=2N+1}^{\infty} k^{-2s} \\ &\leq 2(N+1)^{-2s+1} \frac{1}{2s-1} \left[1 - \left(\frac{1}{2}\right)^{2s-1} \right] \leq 2 \ln 2 (N+1)^{-2s+1}. \end{aligned}$$

for all $s > 1/2$. Thus, $\sum_{k=N+1}^{2N} |g_k| \leq \sqrt{3}LN^{-s+1/2}$, $\forall s > -1/2$ as claimed in the first inequality in (68). Similarly,

$$\sum_{k=N+1}^{2N} k|g_k| \leq \left[\sum_{k=N+1}^{\infty} |g_k|^2 k^{2s} \right]^{1/2} \left[\sum_{k=N+1}^{2N} k^{-2s+2} \right]^{1/2} \leq L \left[\sum_{k=N+1}^{2N} k^{-2s+2} \right]^{1/2},$$

and $\sum_{k=N+1}^{2N} k^{-2s+2} \leq 4N^{-2s+3}$ for all $s > -1/2$ so that the second inequality in (68) follows. \blacksquare

We establish some useful properties of functions J_N and H_N .

Lemma 8 (i) Function $F_N(\cdot)$ is periodic on $[-1/2, 1/2]$, $F_N(t) = F_N(-t)$.

(ii) $F_N(0) = N^2$ is the global maximum of F_N . Moreover, F_N is decreasing when $0 < t \leq 1/N$, and for $N \geq 4$

$$\max_{|t| \geq \frac{1}{5N}} F_N(t) \leq \frac{N^2}{3}. \quad (69)$$

(iii) Function $U_N(\cdot)$ is periodic on $[-1/2, 1/2]$, $U_N(t) = -U_N(-t)$, $U_N(0) = U_N(1/N) = 0$, and for all $N \geq 2$, $U_N(t) > 0$, for $0 < t < \frac{1}{N}$.

(iv) For all $N \geq 6$

$$\max_{t \in [0, \frac{1}{N}]} U_N(t) \geq \frac{3}{2}N^3, \quad \text{and} \quad \max_{t \in [\frac{1}{N}, \frac{1}{2}]} |U_N(t)| \leq N^3.$$

(v) $U'_N(0) > 0$, $U'_N(1/N) < 0$, $U'_N(\cdot)$ has a unique zero in $(0, \frac{1}{N})$, and for all $N \geq 6$ we have

$$\begin{aligned} U_N(t) &\geq \frac{5}{4}N^4 t, \quad t \in [0, \frac{1}{2N}] \\ U_N(t) &\geq \frac{1}{3}N^4 \left(\frac{1}{N} - t \right), \quad t \in [\frac{1}{2N}, \frac{1}{N}]. \end{aligned} \quad (70)$$

(vi) Let $0 < \eta < 1/2$; then for all $N \geq 6$

$$\max_{t \in [\frac{1-\eta}{N}, \frac{1}{N}]} U_N(t) \leq 5\eta N^3.$$

Proof: (ii). We prove (69). For $0 \leq t \leq 1/6$, $\sin \pi t \geq 3t$; hence

$$\max_{\frac{1}{5N} \leq |t| \leq 1/6} F_N(t) \leq \max_{\frac{1}{5N} \leq |t| \leq 1/6} \frac{1}{9\pi^2 t^2} \leq \frac{25N^2}{9\pi^2} \leq \frac{N^2}{3}.$$

For $\frac{1}{6} \leq t \leq 1/2$, $\sin \pi t \geq 1/2$ hence

$$\max_{\frac{1}{6} \leq t \leq \frac{1}{2}} F_N(t) \leq 4 \leq \frac{N^2}{3}.$$

(iii). The proof of this statement is immediate.

(iv). Differentiating we have the following explicit formula for $U_N(\cdot)$:

$$U_N(t) = -2\pi \frac{\sin \pi t N}{\sin \pi t} \left[N \frac{\cos \pi t N}{\sin \pi t} - \frac{\sin \pi t N}{\sin^2 \pi t} \cos \pi t \right]. \quad (71)$$

$$= 2\pi \frac{\sin \pi t N}{\sin^3 \pi t} \left[\sin \pi t N \cos \pi t - N \cos \pi t N \sin \pi t \right]. \quad (72)$$

Another useful representation of $U_N(\cdot)$ is easily obtained from the definition:

$$\begin{aligned}
U_N(t) &= 2\pi i \sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} (k-j) \exp[-2\pi i(k-j)t] \\
&= 2\pi \sum_{k=N+1}^{2N} \sum_{j=N+1}^{2N} (k-j) \sin[2\pi(k-j)t] \\
&= 4\pi \sum_{k=1}^{N-1} k(N-k) \sin(2\pi kt).
\end{aligned} \tag{73}$$

We have from (72)

$$\max_{t \in [0, \frac{1}{N}]} U_N(t) \geq U_N\left(\frac{1}{2N}\right) = 2\pi \frac{1}{\sin^3 \frac{\pi}{2N}} \cos \frac{\pi}{2N} \geq \frac{16}{\pi^2} N^3 \cos \frac{\pi}{12} \geq \frac{3}{2} N^3.$$

For $0 \leq t \leq \frac{1}{6}$, $\sin \pi t \geq 3t$, so that

$$\begin{aligned}
\max_{\frac{1}{N} \leq t \leq \frac{1}{6}} |U_N(t)| &\leq 2\pi \max_{\frac{1}{N} \leq t \leq \frac{1}{6}} \frac{1}{\sin^2 \pi t} \left[\frac{1}{\sin \pi t} + N \right] \\
&\leq \frac{2\pi}{9} N^2 (N + N/3) = \frac{8\pi}{27} N^3 \leq N^3, \quad \forall \frac{1}{N} \leq t \leq \frac{1}{6}.
\end{aligned}$$

If $\frac{1}{6} \leq t \leq \frac{1}{2}$ then $\sin \pi t \geq 1/2$; hence we obtain from (72)

$$\max_{t \in [\frac{1}{6}, \frac{1}{2}]} |U_N(t)| \leq 8\pi(N+2) \leq N^3.$$

Combining these bounds we complete the proof of (iv).

(v). The mentioned properties of $U'_N(\cdot)$ follow from standard properties of the Fejer kernel. To prove (70) we use (73). Note that $\sin(2\pi kt) > 0$ for all $k = 1, \dots, N-1$ whenever $t \in [0, \frac{1}{2N}]$. Assume for simplicity that N is even. Then

$$\begin{aligned}
U_N(t) &\geq 4\pi \sum_{k=1}^{N/2} k(N-k) \sin(2\pi kt) \geq 16\pi t \sum_{k=1}^{N/2} k^2(N-k) \\
&= 16\pi t \left[\frac{N^2}{24} (N+2)(N+1) - \frac{N^2}{64} (N+2)^2 \right] \geq \frac{5}{12} \pi N^4 t,
\end{aligned}$$

as claimed.

Define $R_N(t) \equiv \sin \pi t N \cos \pi t - N \cos \pi t N \sin \pi t$; then by (71)

$$U_N(t) = 2\pi \frac{\sin \pi t N}{\sin^3 \pi t} R_N(t).$$

Using the fact that $\sin \alpha \geq -(2/\pi)\alpha + 2$ for $\alpha \in [\pi/2, \pi]$ we obtain for $t \in (1/(2N), 1/N)$

$$\frac{\sin \pi t N}{\sin \pi t} \geq \frac{-\frac{2}{\pi} \pi t N + 2}{\sin \pi t} \geq \frac{1}{\sin \frac{\pi}{N}} 2N \left(\frac{1}{N} - t \right) \geq \frac{2}{\pi} N^2 \left(\frac{1}{N} - t \right).$$

Note that $R_N(t) > 0$, $\forall t \in (\frac{1}{2N}, \frac{1}{N})$; in addition,

$$R'_N(t) = \pi(N^2 - 1) \sin \pi t N \sin \pi t > 0, \quad \forall \frac{1}{2N} < t < \frac{1}{N}.$$

Thus $R_N(t)$ is monotone increasing in $(1/(2N), 1/N)$ so that

$$\begin{aligned}
U_N(t) &\geq 4N^2 \left(\frac{1}{N} - t \right) \frac{R_N(1/2N)}{\sin^2 \frac{\pi}{N}} \\
&\geq \frac{4}{\pi^2} N^4 \left(\frac{1}{N} - t \right) \cos \frac{\pi}{12} \geq \frac{1}{3} N^4 \left(\frac{1}{N} - t \right),
\end{aligned}$$

as claimed.

(vi). We have from (73) for all $t \in [-\frac{1}{2}, \frac{1}{2}]$

$$\begin{aligned} |U'_N(t)| &= 8\pi^2 \sum_{k=1}^{N-1} k^2(N-k) \cos(2\pi kt) \\ &\leq 8\pi^2 \left[N \sum_{k=1}^{N-1} k^2 - \sum_{k=1}^{N-1} k^3 \right] = \frac{2}{3}\pi^2 N^2(N-1)^2. \end{aligned}$$

Then because $U_N(1/N) = 0$ and $N \geq 6$

$$\max_{\frac{1-\eta}{N} \leq t \leq \frac{1}{N}} U_N(t) \leq \frac{2}{3}\pi^2 N^2(N-1)^2 \frac{\eta}{N} \leq \frac{2}{3}\pi^2 \left(\frac{5}{6}\right)^2 \eta N^3 \leq 5\eta N^3.$$

This completes the proof of the lemma. ■

Proof of Lemma 1: Clearly, R_ε is minorated with the Bayesian risk r_ε , which corresponds to the prior distribution $P(j=0) = P(1 \leq j \leq N) = 1/2$:

$$r_\varepsilon = \inf_{\hat{\psi}} \left[\frac{1}{2} E_0 \hat{\psi}^2 + \frac{1}{2} E_\pi (\psi - \hat{\psi})^2 \right] = \inf_{\hat{\psi}} \frac{1}{2} E_0 \left[\hat{\psi}^2 + Z_\pi (\psi - \hat{\psi})^2 \right] \quad (74)$$

Observe that $\bar{\psi} = \frac{\psi Z_\pi}{1+Z_\pi}$ is the minimizer of (74), so that

$$r_\varepsilon = \frac{\psi^2}{2} E_0 \left[\frac{Z_\pi}{1+Z_\pi} \right] = \frac{\psi^2}{2} E_0 \left[\frac{Z_\pi}{2+(Z_\pi-1)} \right] \geq \frac{\psi^2}{4} \left(1 - E_0 \frac{(Z_\pi-1)^2}{2} \right),$$

as $\frac{1}{1+x} \geq 1-x$ for $x > -1$. ■

Proof of Lemma 2: Indeed,

$$e^x = \sum_{j=0}^{\infty} \frac{x^j}{j!} \leq \sum_{j=0}^{n-1} \frac{x^j}{j!} + g_n(x),$$

where $g_n(x) = \frac{x^n}{n!} e^x$. By the Stirling formula,

$$g_n(x) \leq \frac{1}{\sqrt{2\pi n}} \left(\frac{x}{n}\right)^n e^{n+x} \leq \frac{e^{-n}}{\sqrt{2\pi n}}$$

for $0 \leq x \leq n/9$. ■

Proof of Lemma 3: Let, for the sake of definiteness, $e = (1, \dots, 1)^T$. We have

$$\begin{aligned}
J_\ell(e) &= \frac{1}{\sqrt{\ell}} \sum_{m=-\infty}^{\infty} \int_0^\infty \cdots \int_0^\infty \frac{dx_1}{(1+x_1)^{2\beta}} \cdots \frac{dx_\ell}{(1+x_\ell)^{2\beta}} \delta(e^T x - m) \\
&= \frac{1}{\sqrt{\ell}} \sum_{m=0}^{\infty} \sum_{j=0}^{m-1} \underbrace{\int_0^\infty \cdots \int_0^\infty}_{\ell-1} \frac{dx_1}{(1+x_1)^{2\beta}} \cdots \frac{dx_{\ell-1}}{(1+x_{\ell-1})^{2\beta}} \int_j^{j+1} \frac{dx_\ell}{(1+x_\ell)^{2\beta}} \delta(e^T x - m) \\
&\leq \sum_{m=0}^{\infty} \sum_{j=0}^{m-1} \frac{1}{(1+j)^{2\beta}} \int_0^\infty \frac{dx_1}{(1+x_1)^{2\beta}} \cdots \int_0^\infty \frac{dx_{\ell-1}}{(1+x_{\ell-1})^{2\beta}} 1(m-j-1 \leq x_1 + \dots + x_{\ell-1} \leq m-j) \\
&\leq \sum_{m=0}^{\infty} \sum_{j=0}^{m-1} \frac{1}{(1+j)^{2\beta}} \int_0^\infty \frac{dx_1}{(1+x_1)^{2\beta}} \cdots \int_0^\infty \frac{dx_{\ell-1}}{(1+x_{\ell-1})^{2\beta}} 1(m \leq x_1 + \dots + x_{\ell-1} \leq m+1) \\
&\leq (1 + (2\beta - 1)^{-1}) \int_0^\infty \frac{dx_1}{(1+x_1)^{2\beta}} \cdots \int_0^\infty \frac{dx_{\ell-1}}{(1+x_{\ell-1})^{2\beta}} \leq \frac{2\beta}{(2\beta - 1)^\ell}.
\end{aligned}$$

■

Proof of Lemma 5: Again, R_ε is minorated with the Bayesian risk r_ε , which corresponds to the prior distribution $P(0 \leq j \leq N-1) = 1/N$:

$$r_\varepsilon = N^{-1} \inf_{\hat{\theta}} \left[\sum_{j=1}^N E(\theta_j - \hat{\theta})^2 \right] = N^{-1} \inf_{\hat{\theta}} E_0 \left[\sum_{j=1}^N Z_j (\theta_j - \hat{\theta})^2 \right] \quad (75)$$

Observe that $\bar{\theta} = \frac{\sum_{j=1}^N \theta_j Z_j}{\sum_{j=1}^N Z_j}$ is the minimizer of (75), and

$$r_\varepsilon = N^{-1} E_0 \left[\sum_{j=1}^N \theta_j^2 Z_j - \frac{\left(\sum_{j=1}^N \theta_j Z_j \right)^2}{\sum_{j=1}^N Z_j} \right] = N^{-3} E_0 \left[\sum_{j=1}^N j^2 Z_j - \frac{\left(\sum_{j=1}^N j Z_j \right)^2}{\sum_{j=1}^N Z_j} \right].$$

Let now $A = \{\omega \in \Omega \mid Z_\pi \geq 1 - \delta\}$. By the Cauchy inequality,

$$r_\varepsilon \geq N^{-3} E_0 \left[\left(\sum_{j=1}^N j^2 Z_j - \frac{\left(\sum_{j=1}^N j Z_j \right)^2}{\sum_{j=1}^N Z_j} \right) 1(A) \right].$$

Note that

$$\begin{aligned}
E_0 \sum_{j=1}^N j^2 Z_j 1(A) &= \sum_{j=1}^N j^2 P(A) - E_0 \sum_{j=1}^N j^2 (1 - Z_j) 1(A) \\
&\geq \frac{N^3}{3} P(A) - N^2 E_0 \sum_{j=1}^N (1 - Z_j) 1(A) \geq \frac{N^3}{3} P(A) - N^3 \delta.
\end{aligned}$$

On the other hand,

$$E_0 \frac{\left(\sum_{j=1}^N j Z_j \right)^2}{\sum_{j=1}^N Z_j} 1(A) \leq \frac{E_0 \left(\sum_{j=1}^N j Z_j \right)^2}{N(1 - \delta)}$$

■

References

- ANTONIADIS, A., GIJBELS, I. (2002). Detecting abrupt changes by wavelet methods. *J. Nonparametr. Statist.* **14**, 7-29.
- BASSEVILLE, M. and NIKIFOROV, I.V. (1993) *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, Inc., Upper Saddle River, NJ.
- DALALYAN, A.S., GOLUBEV, G.K. and TSYBAKOV, A. (2006). Penalized maximum likelihood and semiparametric second order efficiency. *Annals of Statistics* **34**, 169-201.
- FREIDLIN, M. I. and KOROSTELEV, A. P. (1995). Image processing for plane domains: change-point problems for the domain's area. *Problems of Information Transmission* **31**, 27-45.
- GIJBELS, I., HALL, P. and KNEIP, A. (1999). On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.* **51**, 231-251.
- GOLDENSHLUGER, A., TSYBAKOV, A. and ZEEVI, A. (2006). Optimal change-point estimation from indirect observations. *Annals of Statistics* **34**, 350-372.
- GOLDENSHLUGER, A., TSYBAKOV, A., JUDITSKY, A. and ZEEVI, A. (2006). Change-point estimation from indirect observations 2. Adaptation. Submitted to *Annales de l'IHP*.
- GOLUBEV, G. K. (1989). Estimation of the time delay of a signal under nuisance parameters. *Problems Inform. Transmission* **25**, 173-180.
- HANNAN, E. J. (1973). The estimation of frequency. *J. Appl. Probab.* **10**, 510-519.
- HUH, J. and CARRIERE, K. C. (2002). Estimation of regression functions with a discontinuity in a derivative with local polynomial fits. *Statist. Probab. Lett.* **56**, 329-343.
- IBRAGIMOV, I. A. and HAS'MINSKI, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer-Verlag, New York.
- KOROSTELEV, A. P. (1987). Minimax estimation of a discontinuous signal. *Theory Probab. Applic.* **32**, 727-730.
- KUTOYANTS, Y. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer, London.
- MÜLLER, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20**, 737-761.
- NEUMANN, M. H. (1997). Optimal change-point estimation in inverse problems. *Scand. J. Statist.* **24**, 503-521.
- PARK, C.-W. and KIM, W.-C. (2004). Estimation of a regression function with a sharp change point using boundary wavelets. *Statist. Probab. Lett.* **66**, 435-448.
- QUINN, B. G. and HANNAN, E. J. (2001). *The Estimation and Tracking of Frequency*. Cambridge University Press.
- RAIMONDO, M. (1998). Minimax estimation of sharp change points. *Ann. Statist.* **26**, 1379-1397.
- RICE, J. A. and ROSENBLATT, M. (1988). On frequency estimation. *Biometrika* **74**, 477-484.

- SAMKO, S. G., KILBAS, A. A. and MARICHEV, O. I. (1993). *Fractional integrals and derivatives. Theory and applications*. Translated from the 1987 Russian original. Gordon and Breach Science Publishers.
- TSYBAKOV, A. B. (2004). *Introduction à l'estimation non-paramétrique*. Springer, Berlin.
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
- WANG, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82**, 385-397.
- YIN, Y. Q. (1988). Detection of the number, locations and amplitudes of jumps. *Comm. Statist. Stochastic Models* **4**, 445-455.
- ZYGMUND, A. (1959). *Trigonometric Series*. vol. I, 2nd edition, Cambridge University Press.