Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2008, Vol. 44, No. 5, 819–836
DOI: 10.1214/07-AIHP144

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Change-point estimation from indirect observations. 2. Adaptation

A. Goldenshluger[a,1], A. Juditsky[b], A. Tsybakov[c] and A. Zeevi[d,2]

[a]*Department of Statistics, University of Haifa, Haifa 31905, Israel. E-mail: goldensh@stat.haifa.ac.il*
[b]*LMC-IMAG 51 rue des Mathematiques, B. P. 53 38041 Grenoble Cedex 9, France. E-mail: anatoli.iouditski@imag.fr*
[c]*Laboratoire de Probabilités et Modéles Aléatoires, CNRS UMR 7599 Université Paris VI 4, Place Jussieu, B. P. 188 75252 Paris Cedex 05, France. E-mail: tsybakov@ccr.jussieu.fr*
[d]*Graduate School of Business, Columbia University, New York, NY 10027, USA. E-mail: assaf@gsb.columbia.edu*

Received 11 October 2006; revised 13 August 2007; accepted 10 September 2007

Abstract. We focus on the problem of adaptive estimation of signal singularities from indirect and noisy observations. A typical example of such a singularity is a discontinuity (change-point) of the signal or of its derivative. We develop a change-point estimator which adapts to the unknown smoothness of a nuisance deterministic component and to an unknown jump amplitude. We show that the proposed estimator attains optimal adaptive rates of convergence. A simulation study demonstrates reasonable practical behavior of the proposed adaptive estimates.

Résumé. Nous étudions ici le problème d'estimation adaptative de singularités d'un signal à partir des observations indirectes et bruitées. Par exemple, cette definition de singularité inclut des points de discontinuité (points de rupture) du signal ou de ses derivées. Nous proposons un estimateur du point de rupture qui s'adapte à une regularité inconnue du paramètre de nuisance et à l'amplitude inconnue du saut, et dont la vitesse de convergence est optimale. Nous illustrons les propriétés théoriques de cet estimateur par quelques résultats de simulation.

MSC: 62G05; 62G20

Keywords: Singularities; Change-point; Estimation; Detection; Minimax risk; Adaption; Optimal rates

1. Introduction

Consider the following model in the space of sequences

$$y_k = a\exp(2\pi \mathrm{i} k\theta) + g_k + \varepsilon\sigma_k\xi_k, \quad k \in \mathbf{N}, \tag{1}$$

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 a 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, $k \in \mathbf{N}$.

We consider the above problem under the assumption that the nuisance sequence (g_k) belongs to a Sobolev ellipsoid

$$G_s(L) = \left\{ g \in \mathbf{C}^{\mathbf{N}} \Big| \sum_{k=1}^{\infty} |g_k|^2 k^{2s} \le L^2 \right\}, \quad s > -\frac{1}{2}. \tag{2}$$

[1]Supported by the ISF Grant No. 300/04.
[2]Supported by NSF Grant DMI-04-47652.

In addition we assume that

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

$$\underline{\sigma} k^{\beta} \le |\sigma_k| \le \bar{\sigma} k^{\beta}, \quad \forall k \in \mathbf{N}. \tag{3}$$

The motivation for considering model (1) under assumptions (2) and (3) is provided by the fact that various problems of change-point estimation from indirect observations can be stated in the form (1). In particular, the frequency θ in (1) corresponds to the change-point, while the amplitude a relates to the jump amplitude. The following three change-point estimation problems illustrate this relationship.

Estimation of a change-point in derivatives

Consider the Gaussian white noise model

$$\mathrm{d}Y(t) = f(t)\,\mathrm{d}t + \varepsilon\,\mathrm{d}W(t), \quad t \in [0,1], \tag{4}$$

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 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 $m = \lfloor \alpha \rfloor$ (where $\lfloor \alpha \rfloor$ stands for the integer part of α). We then say that $f^{(m)}$ has a cusp of the order $\alpha - m$ at θ.

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

$$f^{(\alpha)}(t) = aV(t-\theta) + q(t), \quad t \in [0,1], \tag{5}$$

where $V(t) = 1/2 - t - \lfloor t \rfloor$ is the "saw-tooth" function, and $q \in G_s(L)$, $s > \frac{1}{2}$. We note that (5) is the standard way of representing discontinuous functions in the theory of Fourier series (see, e.g., [5], p. 9). Then the model (4) is equivalent to the sequence-space model (1) where

$$g \in G_{s-1}(2\pi L) \text{ and } \sigma_k^2 = (2\pi k)^{2\alpha+2}. \tag{6}$$

Indeed,

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

and, due to the periodicity of f, $g_0 = 0$, the Fourier coefficients of the function $f^{(\alpha)}$ in (5) are

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

On the other hand, the model (4) is clearly equivalent to

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

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

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

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

$$y_k = a\mathrm{e}^{2\pi \mathrm{i} k\theta} + g_k + \sigma_k \varepsilon \xi_k, \quad k \in \mathbf{N}^+$$

with (σ_k) and (g_k) which satisfy (6).

Change-point estimation in the convolution white noise model

The white noise convolution model is given by the equation

$$\mathrm{d}Y(t) = (Kf)(t)\,\mathrm{d}t + \varepsilon\,\mathrm{d}W(t), \quad t \in [0, 1], \tag{7}$$

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]:

$$(Kf)(t) = \int_0^1 K(t-s) f(s)\,\mathrm{d}s.$$

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.

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

$$ck^{\alpha} \le \left|(K_k)^{-1}\right| \le Ck^{\alpha}.$$

Using the same arguments as above, we conclude that the model (7) can be equivalently rewritten in the form (1) with

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

Observe that the relation (3) holds with $\beta = \alpha + 1$, $\underline{\sigma} = 2\pi c$ and $\bar{\sigma} = 2\pi C$.

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

$$\mathrm{d}Y(t) = \left[aS(t-\theta) + q(t)\right]\mathrm{d}t + \varepsilon\,\mathrm{d}W(t), \quad t \in [0, 1], \tag{8}$$

$a \in \mathbf{R}\backslash\{0\}$ is an 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.

Suppose that in the model (8) $g \in G_s(L)$ and for some $\frac{1}{2} < \alpha < s$ and $0 < c \le C < \infty$

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

Obviously, the model (8) is equivalent to (1) with

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

and with $ck^{\alpha} \le \sigma_k \le Ck^{\alpha}$.

In the companion paper [2] (referred to hereafter as Part I), we studied the problem of minimax estimation of θ and a in the model (1). Our estimators of the jump amplitude and of the change-point were based on the so-called contrast functions

$$\widehat{J}_N(t) = \left|\sum_{k=N+1}^{2N} y_k \mathrm{e}^{-2\pi \mathrm{i} kt}\right|^2, \quad t \in [0, 1] \tag{9}$$

$$\widehat{H}_N(t) = 2\pi \mathrm{i} \sum_{k=N+1}^{2N}\sum_{j=N+1}^{2N} (k-j) y_k \bar{y}_j \mathrm{e}^{-2\pi \mathrm{i}(k-j)t} = -\widehat{J}_N'(t), \quad t \in [0, 1], \tag{10}$$

where $N \geq 1$ is a window-size parameter to be chosen. The estimator a_N of the jump magnitude $|a|$ is given by

$$a_N = N^{-1} \sqrt{\max_{t \in [0,1]} \widehat{J}_N(t)}, \tag{11}$$

while the estimate θ_N of the change-point θ is defined as a root of the equation $\widehat{H}_N(\theta) = 0$ on the interval with endpoints $\hat{\theta}_+$ and $\hat{\theta}_-$, where

$$\hat{\theta}_+ \equiv \arg\max_{t \in [0,1]} \widehat{H}_N(t), \qquad \hat{\theta}_- \equiv \arg\min_{t \in [0,1]} \widehat{H}_N(t). \tag{12}$$

These estimators are based on a characterization of a and θ in terms of deterministic counterparts J_N and H_N of $\widehat{J}_N$ and $\widehat{H}_N$ that are obtained from (9) and (10) by substituting $a\exp(2\pi i k\theta)$ for y_k. In particular, θ is the unique global maximizer of J_N, and the corresponding maximal value equals a^2N^2. Furthermore, if θ_+ and θ_- are the unique global maximizer and minimizer of H_N, then θ is the unique zero on the segment with the endpoints θ_- and θ_+ (it is also the midpoint of this segment). In Part I it was shown that the functions $\widehat{J}_N$ and $\widehat{H}_N$ converge to J_N and H_N uniformly on [0, 1].

The choice of the window parameter N in (11) and (12) is crucial to achieve the optimal estimation accuracy. In Part I we have shown that if parameters s and L of the class $G_s(L)$ are known, then N can be chosen (depending on s and L) in such a way that a_N and θ_N possess optimal minimax properties. In particular, if $N = c(L/(\varepsilon\sqrt{\ln \varepsilon^{-1}}))^{1/(\beta+s)}$ with some constant $c = c(\beta, s)$, then the estimator a_N is rate-optimal. On the other hand, construction of rate-optimal estimators for the change-point θ and the optimal rates of convergence are different for two zones: $1/2 < \beta \leq 3/2$, and $\beta > 3/2$. When $1/2 < \beta \leq 3/2$ then the optimal choice of N depends on the amplitude $|a|$ of the jump but does not depend on the regularity parameters s and L of the nuisance deterministic sequence (g_k). If $\beta > 3/2$ then the optimal choice of N depends on s and L but does not depend on $|a|$. We believe that the case $\beta > 3/2$ is more interesting because here the estimation problem is "truly inverse." Furthermore, in the case $1/2 < \beta \leq 3/2$, in order to construct the adaptive estimator it suffices to substitute an estimator $\tilde{a}$ of the amplitude $|a|$ into the estimator of θ. For instance, one can use to this end the adaptive estimator $\hat{a}$ of $|a|$ described below. That is the main reason, as far as the change-point estimation problem is concerned, that we limit our study to the case $\beta > 3/2$.

In this paper we develop adaptive estimators of $|a|$ and θ which do not require prior knowledge of the parameters of the class $G_s(L)$. We show that these estimators are rate optimal in the sense of [3] and [4]. The main results of this paper are given in Section 2. Simulation results, presented in Section 3, show reasonable practical behavior of the proposed estimators. The proofs are relegated to the Appendix.

2. Main results

We start a description of the adaptive jump amplitude estimator.

2.1. *Adaptive jump amplitude estimator*

The construction is based on the general adaptation scheme by Lepski [3]. In order to implement this scheme in the context of the jump amplitude estimation we need to control the stochastic error of the estimator a_N defined in (11). It was shown in Part I, Section 4.1 that the stochastic terms are determined via the complex-valued stationary Gaussian process $w_N(t) = \sum_{k=N+1}^{2N} \sigma_k \xi_k e^{-2\pi i k t}$, $t \in [0, 1]$ with zero mean, and variance

$$\sigma_w^2(N) \equiv 2 \sum_{k=N+1}^{2N} \sigma_k^2. \tag{13}$$

In view of Assumption (A), there exist constants $c_\beta \leq C_\beta < \infty$ depending only on β such that

$$c_\beta \underline{\sigma} N^{\beta+1/2} \leq \sigma_w(N) \leq C_\beta \bar{\sigma} N^{\beta+1/2}, \quad \text{for any } N \in \mathbf{N}^+. \tag{14}$$

Put $\bar{N} \equiv \lfloor \varepsilon^{-2} \rfloor$. For $N = 2, \ldots, \bar{N}$ we set

$$\bar{\sigma}_w(N) = C_\beta \bar{\sigma} N^{\beta+1/2}. \tag{15}$$

Let $\lambda \geq 1$ be a parameter to be specified. Consider the following iterative procedure:

Algorithm 1.

1. *Compute the estimate $a_{\bar{N}}$ with the window parameter $\bar{N}$ and the values*

$$\alpha_0^- = a_{\bar{N}} - 10\varepsilon\lambda\bar{N}^{-1}\bar{\sigma}_w(\bar{N}), \qquad \alpha_0^+ = a_{\bar{N}} + 10\varepsilon\lambda\bar{N}^{-1}\overline{\sigma}_w(\bar{N}).$$

2. *For $N = \bar{N} - 1, \ldots, 2$ compute the estimate a_N with the window parameter N. If*

$$\alpha_{N+1}^- \leq a_N + 10\varepsilon\lambda N^{-1}\bar{\sigma}_w(N) \quad \textit{and} \quad a_N - 10\varepsilon\lambda N^{-1}\bar{\sigma}_w(N) \leq \alpha_{N+1}^+,$$

declare N admissible and compute the brackets

$$\alpha_N^- = \max\{\alpha_{N+1}^-, a_N - 10\varepsilon\lambda N^{-1}\bar{\sigma}_w(N)\},$$
$$\alpha_N^+ = \min\{\alpha_{N+1}^+, a_N + 10\varepsilon\lambda N^{-1}\bar{\sigma}_w(N)\}.$$

3. *Define the adaptive estimate $\hat{a}_\varepsilon$ as any point of the segment $[\alpha_{\widehat{N}}^-, \alpha_{\widehat{N}}^+]$ (e.g. $\hat{a}_\varepsilon = (\alpha_{\widehat{N}}^+ + \alpha_{\widehat{N}}^-)/2$), where $\widehat{N}$ is the smallest admissible N.*

Note that the adaptive estimator $\hat{a}_\varepsilon$ is well-defined as the set of admissible window parameters N is non-empty (*barN* is always contained in this set). We also note that our construction depends only on a design parameter λ; it will be chosen in the sequel.

For $\lambda \geq 1$ define N_*, where

$$N_* = \min\{N\colon\ 2\varepsilon\lambda N^{-1}\bar{\sigma}_w(N) \geq \sqrt{3}LN^{-s-1/2}\}. \tag{16}$$

Note that N_* depends on the parameters s, L of the class $G_s(L)$. If we knew the true values of L, s (and could choose $N = N_*$ with $\lambda = O(\sqrt{\ln \varepsilon^{-1}})$), that choice of N would lead to a rate-optimal estimator of $|a|$; see Part I, Section 3.2.

The next statement establishes an upper bound on the accuracy of $\hat{a}_\varepsilon$.

Theorem 1. *Assume that Assumption* (A) *holds with $\beta > 1/2$. Let $\lambda \geq 1$, and let $g \in G_s(L)$, with $s > -1/2$ and $L > 0$ such that*

$$\varepsilon\lambda \leq \min\{2^{-1/2}, [c_\beta\underline{\sigma}(2L)^{-1}]^{1/(2\beta+2s-1)}, L(C_\beta\bar{\sigma})^{-1}\}. \tag{17}$$

Then there is a set $\mathcal{A}_J(\lambda) \subseteq \Omega$ such that

$$P(\mathcal{A}_J(\lambda)) \geq 1 - c(\beta)\lambda\bar{N}^2 \mathrm{e}^{-2\lambda^2}, \tag{18}$$

and for any $\omega \in \mathcal{A}_J(\lambda)$,

$$\left|\hat{a}_\varepsilon - |a|\right| \leq 20\varepsilon\lambda N_*^{-1}\bar{\sigma}_w(N_*). \tag{19}$$

The result of Theorem 1 shows that under proper choice of the design parameter λ, the event $\mathcal{A}_J(\lambda)$ is of large probability, and on $\mathcal{A}_J(\lambda)$ the estimation accuracy is controlled by the expression on the RHS of (19). An immediate corollary of the above result is as follows:

Corollary 1. *Let $\hat{a}_\varepsilon$ denote the adaptive estimator given by Algorithm* 1 *and associated with $\lambda = c(\beta)\sqrt{\ln \varepsilon^{-1}}$. Let $g \in G_s(L)$ with parameters $s > -1/2$, $0 < L < \infty$ such that* (17) *holds. Assume that $\beta > 1/2$. Then there exists a constant $C = C(\beta, s)$ such that*

$$\left(E\left|\hat{a}_\varepsilon - |a|\right|^2\right)^{1/2} \le C L^{(2\beta-1)/(2s+2\beta)}\left(\bar{\sigma}\varepsilon\sqrt{\ln \varepsilon^{-1}}\right)^{(2s+1)/(2s+2\beta)}, \quad \text{for any } a \in \mathbf{R}. \tag{20}$$

Comparing (20) to the results of Theorems 1 and 2 of Part 1, we conclude that the adaptive estimator $\hat{a}_\varepsilon$ is rate-optimal.

2.2. *Adaptive change-point estimator*

Let us start the presentation of the adaptive change-point estimator with some informal discussion. As we have emphasized in the introductory section, the optimal choice of window parameter N depends on the regularity parameters s and L and does not depend on the jump amplitude $|a|$ when $\beta > 3/2$. However, recall that when the jump amplitude is small, consistent estimation of the change-point is impossible. This is the case, in particular, when (cf. Theorem 5 of Part 1)

$$|a| \le c L^{(2\beta-1)/(2s+2\beta)}\left(\bar{\sigma}\varepsilon\sqrt{\ln \varepsilon^{-1}}\right)^{(2s+1)/(2s+2\beta)}$$

for some constant $c > 0$. If we recall now the definition of N_* in (16) with $\lambda \sim \sqrt{\ln \varepsilon^{-1}}$, we notice that the the "critical" amplitude value, i.e. the minimal jump amplitude for which consistent change-point estimation is conceivable, satisfies

$$|a| \ge c\lambda\varepsilon N_*^{-1}\bar{\sigma}_w(N_*).$$

In other words, the properties of the minimax change-point estimator $\hat{\theta}_*$ are quite different in two zones of the $(N, |a|)$ "plane" (see Fig. 1). In the *zone of detection*, which lies under the plot $|a| = c\lambda\varepsilon N^{-1}\bar{\sigma}_w(N)$, the estimator is not consistent. In the *zone of estimation* above the graph, the estimation of the change-point is feasible. Clearly, an adaptive change-point estimator will exhibit an analogous behavior. We expect its zone of estimation to be "comparable" to that of the minimax estimator, and its rate in this zone to be (up to a log factor in ε) the same as the minimax rate.

The following construction of the adaptive change-point estimator $\tilde{\theta}$ depends on two design parameters, λ and $\varkappa$; they will be specified in what follows.

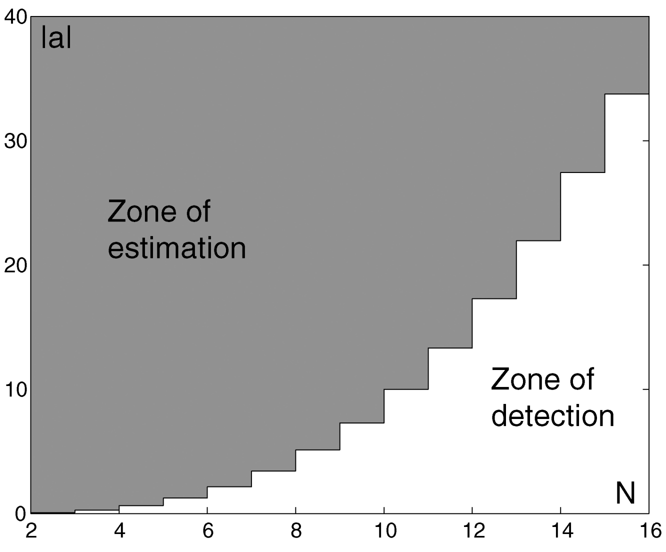

Fig. 1. "Typical" zone of estimation and zone of detection.

Let $\bar{N}$ and N_* be defined as above, and let $\hat{a}_\varepsilon$ be the adaptive estimate of $|a|$ defined in the previous section. For $N \in \{2, \dots, \bar{N}\}$ we define

$$\psi(N) \equiv 20\varepsilon\lambda N^{-1}\bar{\sigma}_w(N), \qquad \psi_* \equiv \psi(N_*). \tag{21}$$

Now, fix $\varkappa \geq 1$, and define $\mathcal{T}$ as a subset of $\{2, \dots, \bar{N}\}$,

$$\mathcal{T} = \{N \in \{2, \dots, \bar{N}\}\colon \hat{a}_\varepsilon \geq 2\varkappa\psi(N)\}. \tag{22}$$

If $\mathcal{T}$ is non-empty we denote $N_0 = \max_{N\in\mathcal{T}} N$, $N_m = \min_{N\in\mathcal{T}} N$ and put $\mathcal{T} = \{N_0, N_1, \dots, N_m\}$ where $N_0 > N_1 > \cdots > N_m$ (here $m + 1 \leq \bar{N} - 1$ is the cardinality of $\mathcal{T}$).

If the set $\mathcal{T}$ is empty we set the adaptive estimator $\hat{\theta}_\varepsilon = 0$, otherwise we perform the following iterative procedure:

Algorithm 2.

1. *Compute the estimate θ_{N_0} with the window parameter N_0 and the values*

$$\tau^-_{N_0} = \theta_{N_0} - 77\pi\varepsilon\lambda\hat{a}_\varepsilon^{-1}N_0^{-2}\bar{\sigma}_w(N_0), \qquad \tau^+_{N_0} = \theta_{N_0} + 77\pi\varepsilon\lambda\hat{a}_\varepsilon^{-1}N_0^{-2}\bar{\sigma}_w(N_0).$$

2. *For $N_i \in \mathcal{T}$, $i = 1, \dots, m$ compute the estimate θ_{N_i} with the window parameter N_i. If $\tau^-_{N_{i-1}} \leq \theta_{N_i} + 77\pi\varepsilon\lambda\hat{a}_\varepsilon^{-1}N_i^{-2}\bar{\sigma}_w(N_i)$ and $\theta_{N_i} - 77\pi\varepsilon\lambda\hat{a}_\varepsilon^{-1}N_i^{-2}\overline{\sigma}_w(N_i) < \tau^+_{N_{i-1}}$, declare N_i admissible and compute the brackets*

$$\tau^-_{N_i} = \max\{\tau^-_{N_{i-1}}, \theta_{N_i} - 77\pi\varepsilon\lambda\hat{a}_\varepsilon^{-1}N_i^{-2}\bar{\sigma}_w(N_i)\},$$
$$\tau^+_{N_i} = \min\{\tau^+_{N_{i-1}}, \theta_{N_i} + 77\pi\varepsilon\lambda\hat{a}_\varepsilon^{-1}N_i^{-2}\bar{\sigma}_w(N_i)\}.$$

3. *Define the adaptive estimate $\hat{\theta}_\varepsilon = (\tau^+_{\widehat{N}} + \tau^-_{\widehat{N}})/2)$ as the center of the interval $[\tau^-_{\widehat{N}}, \tau^+_{\widehat{N}}]$, where $\widehat{N}$ is the smallest admissible $N \in \mathcal{T}$.*

Observe first that the estimate $\hat{\theta}_\varepsilon$ is well-defined: if $\mathcal{T}$ is non-empty, the set of admissible window parameters always includes N_0, else, we have $\hat{\theta}_\varepsilon = 0$.

A reader familiar with Lepski's adaptive estimation procedure will notice an interesting characteristic of the proposed method. In the original Lepski procedure, in order to choose the adaptive estimator $\hat{\theta}$ from an ordered family of estimators $\theta_i, i \in I$, each estimate θ_i is to be compared with *all* subordinated estimators $\theta_k, 0 \leq k < i$. However, in Algorithm 2 above only the estimates θ_{N_i} with $N_i \in \mathcal{T}$ are compared. This modification of the Lepski method can be briefly justified as follows: suppose for a moment that the exact value $|a|$ of the jump amplitude is known. For each $2 \leq N \leq \bar{N}$ we have two possibilities: either $|a| \geq \varkappa\psi(N) = c\lambda\varepsilon N^{-1}\bar{\sigma}_w(N)$ or $|a| < \varkappa\psi(N)$. Let us consider in more detail the second possibility. Suppose that the window parameter N is the "optimal" one. Then $|a| < \varkappa\psi(N)$ implies that the corresponding minimax estimator is in the zone of detection where consistent estimation of θ is impossible. In this case any estimate $\hat{\theta}_\varepsilon \in [0, 1]$ is rate-minimax. Now, if the window parameter is not the "optimal" one, excluding the estimator θ_N from the set of tested estimators would not alter the properties of $\hat{\theta}_\varepsilon$. Hence, in both cases, one can safely exclude such θ_N from the set of candidate estimators. Now it suffices to substitute the pilot estimate $\hat{a}_\varepsilon$ for the true value $|a|$ to obtain the proposed adaptive algorithm.

The next statement establishes an upper bound on the accuracy of the estimate $\hat{\theta}_\varepsilon$.

Theorem 2. *Suppose that Assumption* (A) *holds with $\beta > 3/2$. Let $\lambda \geq 1$, and let $g \in G_s(L)$, with $s > -1/2$ and $L > 0$ such that* (17) *is valid*, *and*

$$\varepsilon\lambda \leq 6^{-(\beta+s)}L(C_\beta\underline{\sigma})^{-1}. \tag{23}$$

Let $\varkappa \geq 80\pi$, and assume that

$$|a| \geq 3\varkappa\psi_*. \tag{24}$$

Then there exists a set $\mathcal{A}_H(\lambda) \subseteq \Omega$ of probability at least $1 - c(\beta)\lambda\bar{N}^2 e^{-2\lambda^2}$, such that on $\mathcal{A}_H(\lambda)$ one has

$$|\hat{\theta}_\varepsilon - \theta| \le 155\pi\varepsilon\lambda|a|^{-1}N_*^{-2}\bar{\sigma}_w(N_*).$$

Corollary 2. *Let Assumption* (A) *hold with $\beta > 3/2$. Let $\hat{\theta}_\varepsilon$ denote the adaptive estimator given by Algorithm* 1 *and associated with $\lambda = 2\sqrt{\ln\varepsilon^{-1}}$ and $\varkappa \ge 80\pi$. Assume that $g \in G_s(L)$ with $s > -1/2$, $0 < L < \infty$ such that* (17) *and* (23) *are valid and let*

$$\varphi_\varepsilon(s, L, a) = |a|^{-1}L^{(2\beta-3)/(2\beta+2s)}\big(\bar{\sigma}\varepsilon\sqrt{\ln\varepsilon^{-1}}\big)^{(2s+3)/(2s+2\beta)}. \tag{25}$$

Assume that for some constant $c_1 = c_1(s, \beta)$

$$|a| \ge c_1 L^{(2\beta-1)/(2s+2\beta)}\big(\bar{\sigma}\varepsilon\sqrt{\ln\varepsilon^{-1}}\big)^{(2s+1)/(2s+2\beta)}. \tag{26}$$

Then there exists a constant $c_2 = c_2(s, \beta)$ such that

$$\big(E|\hat{\theta}_\varepsilon - \theta|^2\big)^{1/2} \le c_2\varphi_\varepsilon(s, L, a). \tag{27}$$

Corollary 2 is an immediate consequence of Theorem 2; its proof is therefore omitted. Let us compare the bound of Corollary 2 with that of Theorem 3 of Part I. One observes that the two bounds coincide (up to a different choice of constants) if one substitutes ε in the bound of Theorem 3 of Part I with $\varepsilon\sqrt{\ln\varepsilon^{-1}}$. Following [4] we refer to the factor $\sqrt{\ln\varepsilon^{-1}}$ as *the price of adaptation*. We now prove that in the change-point estimation problem, this price cannot be avoided even in the simple case when the class G contains at least two nuisance sequences with different regularity parameters.

Theorem 3. *Let Assumption* (A) *hold. If $\beta > 3/2$ then for any $s_0 > -1/2$, $s_1 > -1/2$, $s_0 \ne s_1$, any $a \ne 0$ and $L > 0$, there are two signals*

$$f_k^{(i)} = a\exp\big(2\pi i k\theta^{(i)}\big) + g_k^{(i)}, \quad k \in \mathbf{N}, i = 0, 1,$$

such that $g^{(i)} \in G_{s_i}(L)$ with the following property: for any estimator $\hat{\theta}$ of $\theta \in \{\theta^{(0)}, \theta^{(1)}\}$ from observation (y_k) as in (1), *one has*

$$\max_{f\in\{f^{(0)}, f^{(1)}\}} \varphi_\varepsilon^{-1}(s, L, a)\big(E|\hat{\theta} - \theta|^2\big)^{1/2} \ge c.$$

Here $\varphi_\varepsilon(\cdot)$ is defined as in (25) *and c is a positive constant depending only on β and s_0 and s_1.*

3. Simulation results

We have conducted a simulation study in order to evaluate practical performance of the adaptive change-point estimation procedures. The algorithms described in Section 2 depend on constants that guarantee adaptive optimality. These constants are derived from upper bounds on the stochastic terms that characterize the contrast functions $\widehat{J}_N$ and $\widehat{H}_N$. Note that $\widehat{J}_N$ and $\widehat{H}_N$ are quadratic functions of observations. In Part I we remarked that the change-point and jump amplitude estimation in model (1) can be based on the following contrast function that is linear in the observations (y_k)

$$\widehat{M}_N(t) \equiv \Re\Big(\sum_{k=N+1}^{2N} y_k e^{-2\pi i k t}\Big).$$

This contrast is an empirical counterpart of the function

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

Although theoretical analysis of the bias of estimators based on $\widehat{M}_N(\cdot)$ is much more involved, tight bounds on the stochastic error terms are easily derived. This is especially important for adaptive estimation because adaptive procedures use bounds on the stochastic error terms. Therefore in our experiments we implemented the estimators based on the function $\widehat{M}_N(\cdot)$.

The observations y_k, $k = 1, \dots, N_0 = \lfloor \varepsilon^{-2} \rfloor$ are generated in the frequency domain according to the model (1) for three different noise levels $\varepsilon = 0.1, 0.5, 0.25$. Via the inverse Fourier transform, this scheme is equivalent to a non-parametric regression model with regular design of step size $1/N_0$, and Gaussian zero mean errors with unit variance. The sequence σ_k is chosen to be $\sigma_k = (2\pi k)^\beta$. For instance, the value $\beta = 2$, as explained before, corresponds to the problem of estimating a change-point in the first derivative. The nuisance sequence (g_k) is chosen so that it belongs to the ellipsoid $G_s(L)$ [see (2)]. Below we present results of an experiment with $\beta = 2$ and $s = 3$, $L = 60$. The nuisance sequence $g_k \in G_s(L)$ is chosen to mask in the best way the change-point. The detailed description of such a sequence is given in the proof of the lower bound of Theorem 6 of Part 1.

Figure 2 displays a time domain observation corresponding the setup described above for $\varepsilon = 0.1$. The pure jump signal is the integral of the saw–tooth function; it models the jump in the first derivative. The combined signal represents the sum of the pure jump signal and the worst-case nuisance component from $G_3(60)$.

In Figs 3 and 4 we present the results of an experiment with $M = 100$ randomly generated observation samples (the values of θ are drawn from the uniform distribution on [0, 1]). We present in Fig. 3 the the mean square error of the estimator $\hat{a}_\varepsilon$ of $|a|$, and that of the estimator $\hat{\theta}_\varepsilon$ of θ, as a function of ε. In Fig. 4 the corresponding boxplots are provided.

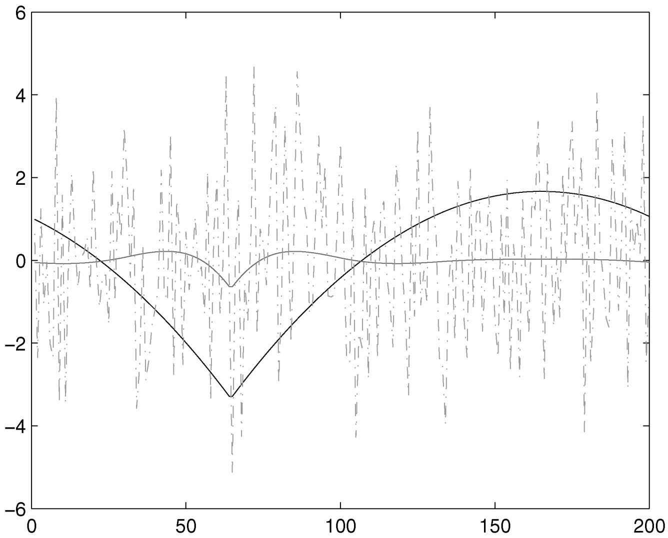

Fig. 2. The pure jump signal, combined signal and noisy observations.

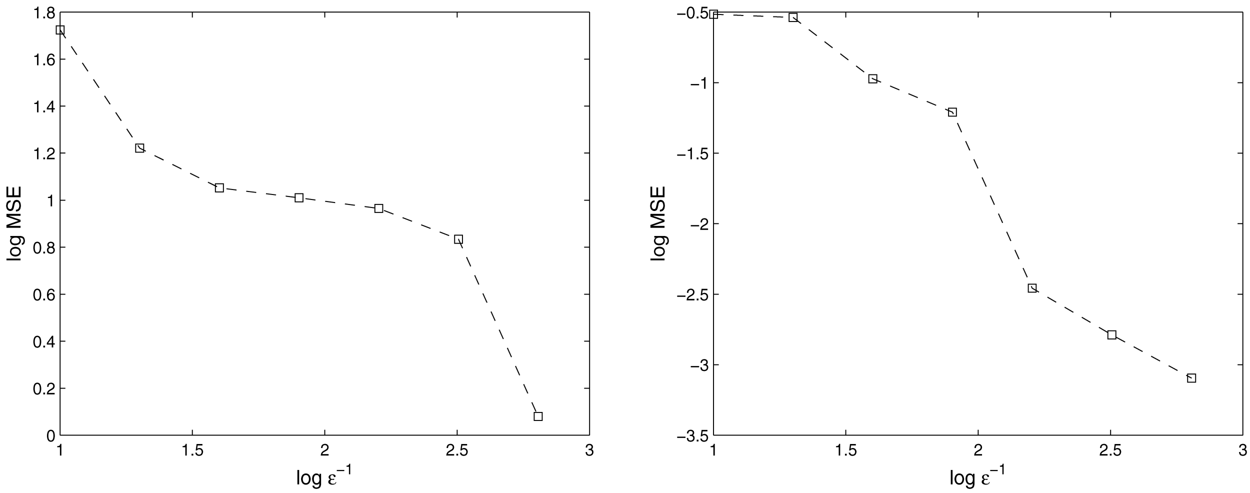

Fig. 3. The log of the MSE of the estimator $\hat{a}_\varepsilon$ (left) and that of the estimator $\hat{\theta}_\varepsilon$ (right) as a function of $\log \varepsilon^{-1}$.

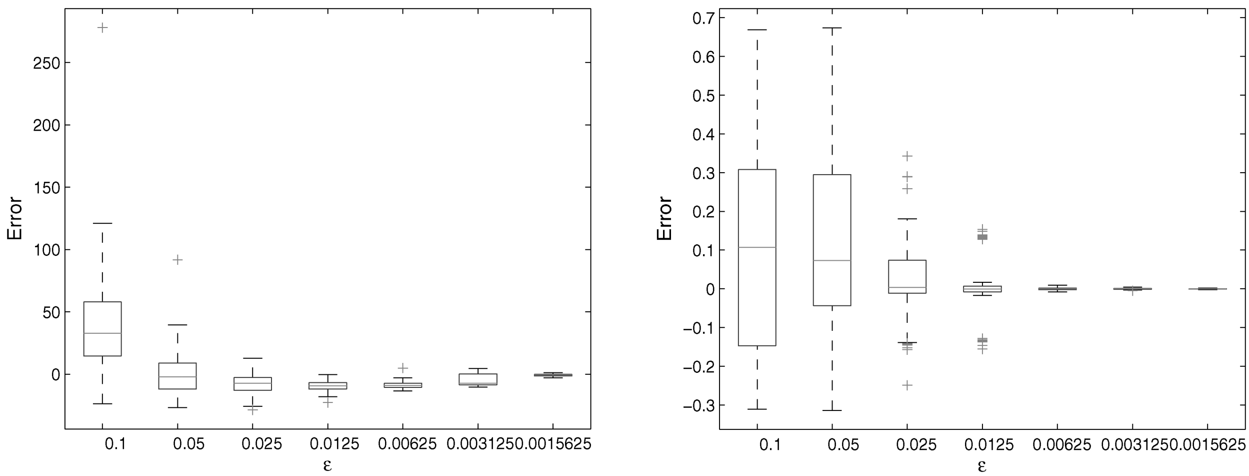

Fig. 4. Boxplot of the error $\hat{a}_\varepsilon - |a|$ (left) and that of $\hat{\theta}_\varepsilon - \theta$ (right) as a function of ε.

Appendix

A.1. *Proof of Theorem* 1

Let us first recall some notations used in Part I. The complex-valued Gaussian process $\{w_N(t), t \in [0,1]\}$ is defined as $w_N(t) = \sum_{k=N+1}^{2N} \sigma_k \xi_k \mathrm{e}^{-2\pi \mathrm{i} kt}$ and

$$\mathcal{A}_J(\lambda; N) = \Big\{\omega \in \Omega\colon \sup_{t\in[0,1]} \big|w_N(t)\big| \le 2\lambda\sigma_w(N)\Big\}.$$

Let

$$\rho_J(N) \equiv 2\left(\sum_{k=N+1}^{2N} |g_k|\right)^2 + 2|a|N \sum_{k=N+1}^{2N} |g_k|.$$

We also put

$$\mathcal{A}_J(\lambda) = \bigcap_{N=2}^{\bar{N}} \mathcal{A}_J(\lambda; N).$$

The bound (18) for the probability of $\mathcal{A}_J(\lambda)$ is readily given by Lemma 6 of Part I.

Following Part I, we denote by $J_N(t)$ the "ideal" version of the contrast function

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

It was shown in Part I that if the event $\mathcal{A}_J(\lambda; N)$ occurs, and if $g \in G_s(L)$ then

$$\begin{aligned} \sup_{t\in[0,1]} \bigl|\widehat{J}_N(t) - J_N(t)\bigr| &\le \Delta_J(\lambda; N) \equiv \rho_J(N) + 8\varepsilon^2\lambda^2\sigma_w^2(N) + 4\varepsilon\lambda|a|N\sigma_w(N) \\ &\le 6L^2N^{-2s+1} + 2\sqrt{3}|a|LN^{-s+3/2} + 8\varepsilon^2\lambda^2\sigma_w^2(N) + 4\varepsilon\lambda|a|N\sigma_w(N). \end{aligned} \tag{28}$$

For the sake of completeness, we also reproduce the following result from Part I.

Lemma 1. *Let a_N be the estimate of $|a|$ associated with the window parameter N. Then for any $\lambda \ge 1$ on the set $\mathcal{A}_J(\lambda; N)$*

$$\bigl|a_N - |a|\bigr| \le \min\left\{ \frac{\Delta_J(\lambda; N)}{N^2(|a| \vee a_N)}, \frac{\Delta_J^{1/2}(\lambda; N)}{N} \right\}, \quad \textit{for any } N \ge 1.$$

Note first that under the premises of Theorem 1, $2 \le N_* \le \bar{N}$. Indeed, by definition (16), $N_* \le [\sqrt{3}(2c_\beta\sigma)^{-1} \times L(\varepsilon\lambda)^{-1}]^{1/(\beta+s)} + 1$. Then the assumption (17) implies that this upper bound is $\le (\varepsilon\lambda)^{-2}$.

Step 1. Observe that by definition of N_*, we have from (28) that for any $N \ge N_*$ and $\omega \in \mathcal{A}_J(\lambda)$:

$$\begin{aligned} \frac{\Delta_J(\lambda; N)}{N^2} &\le \frac{\rho_J(N)}{N^2} + 8\varepsilon^2\lambda^2\frac{\sigma_w^2(N)}{N^2} + 4\varepsilon\lambda|a|\frac{\sigma_w(N)}{N} \\ &\le 16\varepsilon^2\lambda^2\frac{\bar{\sigma}_w^2(N)}{N^2} + 8\varepsilon\lambda|a|\frac{\bar{\sigma}_w(N)}{N}. \end{aligned}$$

Suppose now that for a given N the "true amplitude" $|a|$ satisfies

$$|a| \ge 8\varepsilon\lambda N^{-1}\bar{\sigma}_w(N).$$

Then

$$\frac{\Delta_J(\lambda; N)}{N^2} \le 2|a|\varepsilon\lambda N^{-1}\bar{\sigma}_w(N) + 8|a|\varepsilon\lambda N^{-1}\bar{\sigma}_w(N) \le 10\varepsilon\lambda|a|N^{-1}\bar{\sigma}_w(N).$$

Then by Lemma 1,

$$\bigl|a_N - |a|\bigr| \le \frac{\Delta_J(\lambda; N)}{N^2|a|} \le 10\varepsilon\lambda N^{-1}\bar{\sigma}_w(N).$$

On the other hand, if $|a| < 8\varepsilon\lambda N^{-1}\bar{\sigma}_w(N)$ then

$$\Delta_J(\lambda; N)N^{-2} \le 16\varepsilon^2\lambda^2N^{-2}\bar{\sigma}_w^2(N) + 64\varepsilon^2\lambda^2N^{-2}\bar{\sigma}_w^2(N) = 80\varepsilon^2\lambda^2N^{-2}\bar{\sigma}_w^2(N).$$

Again, by Lemma 1,

$$\big|a_N - |a|\big| \le \frac{\Delta_J^{1/2}(\lambda; N)}{N} \le 9\varepsilon\lambda N^{-1}\bar{\sigma}_w(N).$$

We conclude that on $\mathcal{A}_J(\lambda)$, for $N_* \le N \le \bar{N}$ the following holds:

$$\big|a_N - |a|\big| \le 10\varepsilon\lambda N^{-1}\bar{\sigma}_w(N). \tag{29}$$

Step 2. Let us suppose for an instant that N_*, which is defined in (16), is admissible. Then $\widehat{N} \le N_*$ and, by definition, $\hat{a}_\varepsilon$ belongs to the intersection of segments

$$S_N = \big[a_N - 10\varepsilon\lambda N^{-1}\bar{\sigma}_w(N), a_N - 10\varepsilon\lambda N^{-1}\bar{\sigma}_w(N)\big]$$

for all $N_* \le N \le \bar{N}$. In particular, $\hat{a}_\varepsilon \in S_{N_*}$, and, due to (29),

$$\begin{aligned}\big|\hat{a}_\varepsilon - |a|\big| &\le |\hat{a}_\varepsilon - a_{N_*}| + \big|a_{N_*} - |a|\big| \le 10\varepsilon\lambda N_*^{-1}\bar{\sigma}_w(N_*) + \big|a_{N_*} - |a|\big| \\ &\le 10\varepsilon\lambda N_*^{-1}\bar{\sigma}_w(N_*) + 10\varepsilon\lambda N_*^{-1}\bar{\sigma}_w(N_*) \le 20\varepsilon\lambda N_*^{-1}\bar{\sigma}_w(N_*).\end{aligned}$$

It remains to show that N_* is admissible for any $\omega \in \mathcal{A}_J(\lambda)$, or, what is exactly the same, that the sets S_N for $N_* \le N \le \bar{N}$ have a common point. But (29) means precisely that $|a|$ belongs to the intersection of all such sets S_N.

A.2. *Proof of Corollary* 1

Let $\lambda = c(\beta)\sqrt{\ln \varepsilon^{-1}}$. We use the decomposition

$$E\big(\hat{a}_\varepsilon - |a|\big)^2 = E\big(\hat{a}_\varepsilon - |a|\big)^2 1\big\{\mathcal{A}_J(\lambda)\big\} + E\big(\hat{a}_\varepsilon - |a|\big)^2 1\big\{\mathcal{A}_J^c(\lambda)\big\}. \tag{30}$$

For the first term of (30) we use the bound of Theorem 1:

$$E\big(\hat{a}_\varepsilon - |a|\big)^2 1\big\{\mathcal{A}_J(\lambda)\big\} \le 20\varepsilon\lambda N_*^{-1}\bar{\sigma}_w(N_*) \le C(\beta, s)L^{(2\beta-1)/(2s+2\beta)}\big(\bar{\sigma}\varepsilon\sqrt{\ln \varepsilon^{-1}}\big)^{(2s+1)/(2s+2\beta)}.$$

Let now $2 \le N \le \bar{N}$. By the definition of the estimator a_N (cf. Lemma 1),

$$\big|a_N - |a|\big| \le \min\left\{\frac{\Delta_N(\widehat{J}, J)}{N^2|a|}, \frac{\Delta_N^{1/2}(\widehat{J}, J)}{N}\right\}, \quad \text{for any } N \ge 1,$$

where

$$\Delta_N(\widehat{J}, J) \equiv \Big|\max_{t\in[0,1]}\big|\widehat{J}_N(t)\big| - \max_{t\in[0,1]}\big|J_N(t)\big|\Big|.$$

We now use the first of the above inequalities to bound the error in the case of "large" a, namely $|a| \ge 1$. The corresponding bound for the case of $|a| < 1$ can be obtained in the same way using the second inequality above. We write:

$$|a_N - a| = |a|^{-1}N^{-2}\Delta_N(\widehat{J}, J) \le |a|^{-1}N^{-2}\max_{t\in[0,1]}\big|\widehat{J}_N(t) - J_N(t)\big|.$$

Thus,

$$\begin{aligned}|a_N - a| &\le |a|^{-1}N^{-2}\Big[\rho_J(N) + \varepsilon^2\sup_{t\in[0,1]}\big|w_N(t)\big|^2 + \varepsilon|a|N\sup_{t\in[0,1]}\big|w_N(t)\big|\Big] \\ &\le N^{-2}\big[6L^2N^2 + 2\sqrt{3}LN^2 + \varepsilon^2\zeta_N^2 + \varepsilon\zeta_N\big],\end{aligned}$$

where we denoted $\zeta_N = \sup_{t\in[0,1]} |w_N(t)|$ and used the bound of Part I for $\rho_J(N)$ (cf. (28)). Now, assume that $\mathcal{A}^c_J(\lambda)$ holds. We have:

$$\begin{aligned} E|\hat{a}_\varepsilon - a|^2 1\{\mathcal{A}^c_J(\lambda)\} &\le E\Big(\max_{2\le N\le \bar{N}} |a_N - a|^2 1(\mathcal{A}^c_J(\lambda))\Big) \\ &\le 2\big(6L^2 + 2\sqrt{3}L\big)^2 P\big(\mathcal{A}^c_J(\lambda)\big) \\ &\quad + 2\varepsilon^4 E\Big(\max_{2\le N\le \bar{N}} N^{-4}\zeta_N^4 1\{\mathcal{A}^c_J(\lambda)\}\Big) \\ &\quad + 2\varepsilon^2 E\Big(\max_{2\le N\le \bar{N}} N^{-4}\zeta_N^2 1\{\mathcal{A}^c_J(\lambda)\}\Big) \\ &= I_1 + I_2 + I_3. \end{aligned}$$

Our goal is to bound I_i, $i = 1, \dots, 3$. By (18) we have

$$I_1 \le 2\big(6L^2 + 2\sqrt{3}L\big)^2 P\{\mathcal{A}^c_J(\lambda)\} \le c_1\big(L^4 + 1\big)\lambda \bar{N}^2 \exp\{-2\lambda^2\}. \tag{31}$$

Furthermore,

$$I_2 \le 2\varepsilon^4 \sum_{N=1}^{\bar{N}} N^{-4} E\big(\zeta_N^4 1\{\mathcal{A}^c_J(\lambda)\}\big) \le 2\varepsilon^4 \sum_{N=1}^{\bar{N}} N^{-4} E\big(\zeta_N^4 1\{\mathcal{A}^c_J(\lambda; N)\}\big).$$

On the other hand, by Lemma 6 of Part 1,

$$\begin{aligned} E\big(\zeta_N^4 1\{\mathcal{A}^c_J(\lambda; N)\}\big) &= E\big(\zeta_N^4 1\{\zeta > 2\lambda\sigma_w(N)\}\big) = \int_{(2\lambda\sigma_w(N))^4}^{\infty} \mathbf{P}\big(\zeta_N^4 > t\big)\,\mathrm{d}t \\ &= 4\big(2\lambda\sigma_w(N)\big)^4 \int_1^{\infty} t^3 P\big(\zeta_N > 2\lambda\sigma_w(N)t\big)\,\mathrm{d}t \\ &\le c_2\sigma_w^4(N)N\lambda \int_\lambda^{\infty} t^4 \mathrm{e}^{-2t^2}\,\mathrm{d}t \le c_3\sigma_w^4(N)N\lambda^3\mathrm{e}^{-2\lambda^2}. \end{aligned}$$

Because $\sigma_w(N) \le \bar{\sigma}_w(N)$, and $\bar{\sigma}_w(N)$ is monotone in N

$$I_2 \le c_3\varepsilon^4\bar{\sigma}_w^4(\bar{N})\lambda^3\mathrm{e}^{-2\lambda^2}. \tag{32}$$

In the same way we obtain the bound for I_3:

$$I_3 \le c_4\varepsilon^2 a^2\bar{\sigma}_w^2(\bar{N})\lambda\mathrm{e}^{-2\lambda^2}.$$

Along with (31) and (32) it implies that

$$E|a_\varepsilon - a|^2 1\{\mathcal{A}^c_J(\lambda)\} \le c_5\big[\big(L^4 + 1\big)\lambda\bar{N}^2 + \varepsilon^4\bar{\sigma}_w^4(\bar{N})\lambda^2 + \varepsilon^2\bar{\sigma}_w^2(\bar{N})\big]\mathrm{e}^{-2\lambda^2}.$$

When taking into account definitions of $\bar{N}$, $\bar{\sigma}_w$, and inequality $L \le \frac{1}{2}C_\beta\bar{\sigma}(\varepsilon\lambda)^{-1}$ [see (17)] we finally obtain

$$E|a_\varepsilon - a|^2 1\{\mathcal{A}^c_J(\lambda)\} \le c_6\big[\bar{\sigma}^4\varepsilon^{-8}\lambda + \bar{\sigma}^4\varepsilon^{-8\beta+4}\lambda^2 + a^2\bar{\sigma}^2\varepsilon^{-4\beta+2}\lambda\big]\mathrm{e}^{-2\lambda^2}.$$

Now one can easily exhibit a constant $c(\beta)$ such that for $\lambda = c(\beta)\sqrt{\ln\varepsilon^{-1}}$ the right-hand side above is bounded with ε^2.

A.3. Proof of Theorem 2

Again, we start with notations and results of Part I. The set $\mathcal{A}_H(\lambda; N)$ is defined as follows:

$$\mathcal{A}_H(\lambda; N) = \mathcal{A}_J(\lambda; N) \cap \Big\{\omega\colon \sup_{t\in[0,1]} \big|v_N(t)\big| \le 2\lambda\sigma_v\Big\},$$

where $v_N(t) = \sum_{k=N+1}^{2N} k\sigma_k\xi_k \mathrm{e}^{-2\pi \mathrm{i} kt}$ and $\sigma_v^2 = 2\sum_{k=N+1}^{2N} k^2\sigma_k^2$. Also

$$\sigma_u^2(N) = N^4\sigma_w^2(N) + N^2\sigma_v^2(N).$$

The following bounds on $\sigma_v(N)$ and $\sigma_u(N)$ in terms of $\sigma_w(N)$ can be easily derived:

$$N\sigma_w \le \sigma_v \le 2N\sigma_w, \qquad 2N^2\sigma_w \le \sigma_u \le 3N^2\sigma_w, \quad \forall N. \tag{33}$$

Define also $\mathcal{A}_H(\lambda) = \bigcap_{N=2}^{\bar N} \mathcal{A}_H(\lambda; N)$, and put

$$\rho_H(N) \equiv 8\pi \sum_{k=N+1}^{2N} k|g_k| \sum_{j=N+1}^{2N} |g_j| + 16\pi|a|N^2 \sum_{j=N+1}^{2N} |g_j|.$$

It was shown in Proposition 2 of Part I that if $g \in G_s(L)$ then on the set $\mathcal{A}_H(\lambda; N)$

$$\begin{aligned} &\sup_{t\in[0,1]} \big|\widehat{H}_N(t) - H_N(t)\big| \le \Delta_H(\lambda; N) \\ &\quad \equiv \rho_H(N) + 32\pi\varepsilon^2\lambda^2\sigma_w(N)\sigma_v(N) + 16\pi|a|\varepsilon\lambda\sigma_u(N) \\ &\quad \le 16\pi\big(2L^2N^{-2s+2} + 2L|a|N^{-s+5/2} + 2\varepsilon^2\lambda^2\sigma_w(N)\sigma_v(N) + |a|\varepsilon\lambda\sigma_u(N)\big). \end{aligned} \tag{34}$$

The following result has been proved in Part I.

Lemma 2. *Let $\hat\theta_N$ be the estimate of the change-point associated with the window parameter N. Let $\lambda \ge 1$ and $N \ge 6$ be such that*

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

Then

$$|\theta_N - \theta| \le \left(\frac{5}{4}a^2N^4\right)^{-1} \Delta_H(\lambda; N), \quad \forall\omega \in \mathcal{A}_H(\lambda; N).$$

Step 1. We start with the following lemma.

Lemma 3. *Suppose that Assumption* (A) *holds with $\beta > 3/2$. Let $g \in G_s(L)$ with $s > -1/2$ and $L > 0$ such that* (17) *holds. Let $\varkappa \ge 1$ and* (24) *is valid. Then for any $\omega \in \mathcal{A}_J(\lambda)$*

1. $N_* \in \mathcal{T}$;
2. $N \notin \mathcal{T}$ *if* $N_* < N \le \bar N$ *and* $\psi(N)(2\varkappa - 1) > |a|$.

Proof. We first note that (17) ensures $1 \le N_* \le \bar N$. Now we check that (22) holds for $N = N_*$ when $\mathcal{A}_J(\lambda)$ occurs. It follows from Theorem 1 that $|\hat a_\varepsilon - |a|| \le \psi_* \le \psi(N)$, for any $\omega \in \mathcal{A}_J(\lambda)$ and $N_* \le N \le \bar N$. This along with (24) implies that $\hat a_\varepsilon \ge |a| - \psi_* \ge (3\varkappa - 1)\psi_* \ge 2\varkappa\psi_*$. Hence $N_* \in \mathcal{T}$ as claimed.

On the other hand, on $\mathcal{A}_J(\lambda)$, when $N_* < N \le \bar N$ and $|a| < \psi(N)(2\varkappa - 1)$,

$$\hat a_\varepsilon \le a + \psi(N_*) < \psi(N)(2\varkappa - 1) + \psi(N) < 2\varkappa\psi(N),$$

and such $N \notin \mathcal{T}$. □

Step 2. Assume that $\mathcal{A}_H(\lambda)$ holds and recall that $\mathcal{A}_J(\lambda) \subseteq \mathcal{A}_H(\lambda)$. By Lemma 3, $N_* \in \mathcal{T}$. We will show that N_* is admissible. By definition of $\Delta_H(\lambda; N)$ and by (33) we have

$$\begin{aligned}\frac{1}{16\pi}\Delta_H(\lambda; N) &\le 2L^2 N^{-2s+2} + 2L|a|N^{-s+5/2} + 2\varepsilon^2\lambda^2\sigma_w(N)\sigma_v(N) + |a|\varepsilon\lambda\sigma_u(N) \\ &\le 2L^2 N^{-2s+2} + 2L|a|N^{-s+5/2} + 4\varepsilon^2\lambda^2 N\overline{\sigma}_w^2(N) + 3|a|\varepsilon\lambda N^2\bar{\sigma}_w(N).\end{aligned}$$

It follows from the definition of N_* that for all $N \ge N_*$

$$\begin{aligned}2L^2 N^{-2s+2} &\le \frac{8}{3}\varepsilon^2\lambda^2 N^{-1}\overline{\sigma}_w^2(N), \\ 2L|a|N^{-s+5/2} &\le \frac{4}{\sqrt{3}}\varepsilon\lambda|a|N\overline{\sigma}_w(N).\end{aligned}$$

Thus, for such N,

$$\frac{1}{16\pi}\Delta_H(\lambda; N) \le 7\varepsilon^2\lambda^2 N\bar{\sigma}_w^2(N) + 6\varepsilon\lambda|a|N^2\bar{\sigma}_w(N). \tag{35}$$

Recall that $\psi(N) = 20\varepsilon\lambda N^{-1}\bar{\sigma}_w(N)$, hence for all $N \ge N_*$, $N \in \mathcal{T}$

$$\varepsilon\lambda N^{-1}\bar{\sigma}_w(N) \le \frac{|a|}{20(2\varkappa - 1)}.$$

Now we have from (35):

$$\begin{aligned}\Delta_H(\lambda; N) &\le 16\pi\left(\frac{7N^3a^2}{400(2\varkappa-1)^2} + \frac{3N^3a^2}{10(2\varkappa-1)}\right), \\ (\text{as } \varkappa \ge 1) &\le \frac{127\cdot 16\pi}{400(2\varkappa-1)}a^2N^3 \le \frac{21\pi}{(2\varkappa-1)}a^2N^3 \le \frac{a^2N^3}{4}\end{aligned} \tag{36}$$

for $\varkappa \ge 41\pi + 1/2$. On the other hand, we can estimate $\Delta_H(\lambda; N)$ as follows:

$$\Delta_H(\lambda; N) \le 16\pi\left(\frac{7\varepsilon\lambda N^2|a|\bar{\sigma}_w(N)}{20(2\varkappa-1)} + 6\varepsilon\lambda|a|N^2\bar{\sigma}_w(N)\right) \le (96\pi+1)\varepsilon\lambda|a|N^2\bar{\sigma}_w(N). \tag{37}$$

We conclude from (36) and (37) that the conditions of Lemma 2 are satisfied for $N \ge N_*$ and

$$\begin{aligned}|\theta_N - \theta| &\le \left(\frac{5}{4}a^2N^4\right)^{-1}\Delta_H(\lambda; N) \le \frac{4(96\pi+1)}{5|a|}\varepsilon\lambda N^{-2}\bar{\sigma}_w(N) \\ &\le 77\pi\varepsilon\lambda N^{-2}\bar{\sigma}_w(N)\hat{a}_\varepsilon^{-1}\end{aligned} \tag{38}$$

(recall that due to (24) $\hat{a}_\varepsilon \le \frac{3\varkappa+1}{3\varkappa}|a|$ on $\mathcal{A}_J(\lambda)$). Now define

$$S_N \equiv \left[\theta_N - 77\pi\varepsilon\lambda\hat{a}_\varepsilon^{-1}N^{-2}\bar{\sigma}_w(N), \theta_N + 77\pi\varepsilon\lambda\hat{a}_\varepsilon^{-1}N^{-2}\bar{\sigma}_w(N)\right], \quad N \in \mathcal{T}.$$

Then (38) implies that on $\mathcal{A}_H(\lambda)$ the intersection of the segments S_N with $N \ge N_*$, $N \in \mathcal{T}$ contains (at least one point) θ and N_* is admissible.

Step 3. Thus we can write (cf. (38))

$$\begin{aligned}|\hat{\theta}_\varepsilon - \theta| &\le |\hat{\theta}_\varepsilon - \theta_{N_*}| + |\theta_{N_*} - \theta| \\ &\le 77\pi\varepsilon\lambda\hat{a}_\varepsilon^{-1}N_*^{-2}\bar{\sigma}_w(N_*) + 77\pi\varepsilon\lambda|a|^{-1}N_*^{-2}\bar{\sigma}_w(N_*)\end{aligned}$$

$$\left(\text{as } \hat{a}_\varepsilon \geq \frac{3\varkappa - 1}{3\varkappa}|a|\right) \leq 77\pi \frac{3\varkappa\varepsilon\lambda N_*^{-2}\bar{\sigma}_w(N_*)}{(3\varkappa - 1)|a|} + 77\pi\varepsilon\lambda|a|^{-1}N_*^{-2}\bar{\sigma}_w(N_*)$$
$$\leq 155\pi\varepsilon\lambda N_*^{-2}\bar{\sigma}_w(N_*).$$

This completes the proof.

A.4. *Proof of Theorem* 3

We start with the study of the minimax risk R_ε of a 2-point estimation problem:

$$R_\varepsilon = \sup_{i=0,1} \phi_i(\varepsilon)^{-1} E_i(\theta_i - \hat{\theta})^2.$$

Here $\theta_0 = 0$, $\theta_1 = \theta$, P_0 and P_1 are the corresponding probability distributions, $\phi_1(\varepsilon) = \theta^2$ and $\phi_0(\varepsilon)\phi_1(\varepsilon)^{-1} = \delta$. The following result is fairly known (see, e.g., [1]), we present it here for the sake of completeness.

Lemma 4. *Let $Z_1 = \frac{\mathrm{d}P_1}{\mathrm{d}P_0}$ be the likelihood ratio and $K(P_1, P_0) = -\int \ln Z_1 \,\mathrm{d}P_0$ the Kullback–Leibler distance between P_1 and P_0. Then*

$$R_\varepsilon \geq \max\{\mathrm{e}^{-K(P_0,P_1)-\delta}, 1 - \delta E_0 Z_1^2\}.$$

Proof. Clearly, R_ε is minorated with the Bayesian risk r_ε, which corresponds to the prior distribution $P(i = 0) = P(i = 1) = 1/2$:

$$r_\varepsilon = \inf_{\hat{\theta}} \left[\frac{\phi_0(\varepsilon)^{-1}}{2} E_0\hat{\theta}^2 + \frac{\phi_1(\varepsilon)^{-1}}{2} E_1(\theta - \hat{\theta})^2\right] = \frac{\phi_1(\varepsilon)^{-1}}{2} \inf_{\widehat{\theta}} E_0\left[\delta^{-1}\hat{\theta}^2 + (\theta - \hat{\theta})^2 Z_1\right], \tag{39}$$

Observe that $\theta^* = \frac{\delta\theta Z_1}{1+\delta Z_1}$ is the minimizer of (39), so that

$$r_\varepsilon = \frac{\phi_1(\varepsilon)^{-1}}{2} E_0\left[\theta^2 Z_1 - \frac{\delta\theta^2 Z_1^2}{1 + \delta Z - 1}\right] = \frac{\phi_1(\varepsilon)^{-1}\theta^2}{2} E_0 \frac{Z_1}{1 + \delta Z_1} = \frac{1}{2} E_0 \frac{Z_1}{1 + \delta Z_1}.$$

Now by the Jensen inequality,

$$\ln E_0 \frac{Z_1}{1 + \delta Z_1} \geq E_0(\ln Z_1) - E_0\big(\ln(1 + \delta Z_1)\big) \geq E_0(\ln Z_1) - \delta E_0 Z_1 = -K(P_0, P_1) - \delta.$$

On the other hand,

$$E_0 \frac{Z_1}{1 + \delta Z_1} \geq E_0 Z_1 - \delta E_0 Z_1^2 = 1 - \delta E_0 Z_1^2.$$

□

Now consider the following 2-point problem (cf. proof of Theorem 6 of Part I): given observations $y_k = f_k^{(i)} + \varepsilon\sigma_k\xi_k$, $i = 0, 1$ we are to estimate the parameter (change-point) $\theta \in \{\theta^{(0)}, \theta^{(1)}\}$, where

$$f_k^{(0)} = a, \quad \forall k \in \mathbf{N}^+ \quad \text{and} \quad f_k^{(1)} = a\mathrm{e}^{2\pi i k h} + g_k^{(1)}, \quad \forall k \in \mathbf{N}^+,$$

where $h > 0$, and

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

for some integer n to be chosen in the sequel. The hypotheses correspond to the model (1) 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.

Let us put for the sake of definiteness $s_0 > s_1$ and $L = 1$. Clearly, $g^{(0)} \in G_{s_0}(1)$. We will select n in such a way that $(g_k^{(1)})$ belongs to $G_{s_1}(1)$. We have

$$\sum_{k=1}^{\infty} \bigl|g_k^{(1)}\bigr|^2 k^{2s_1} \le a^2 \sum_{k=1}^{n} \bigl|1 - \mathrm{e}^{2\pi \mathrm{i} kh}\bigr|^2 k^{2s_1} \le c_1 a^2 \min\bigl\{h^2 n^{2s+3}, n^{2s_1+1}\bigr\},$$

where c_1 depends on s_1 only. Choosing

$$n = n_* \equiv c_2\bigl(|a|^{-1} h^{-1}\bigr)^{2/(2s_1+3)}$$

we obtain that $(g_k^{(1)}) \in G_{s_1}(1)$, provided that $n_* \le h^{-1}$.

Let P_0 and P_1 denote the probability measures associated with observations (y_k) in model (1) with $(f_k) = (f_k^{(0)})$ and $(f_k) = (f_k^{(1)})$, respectively. Note that the likelihood ratio $Z_1 = \frac{\mathrm{d}P_1}{\mathrm{d}P_0}$ satisfies

$$\begin{aligned} Z_1 &= \exp\Biggl(-\sum_{k=n_*+1}^{\infty} \frac{|y_k - a\mathrm{e}^{2\pi \mathrm{i} kh}|^2 - |y_k - a|^2}{2\varepsilon^2 \sigma_k^2}\Biggr) \\ &= \exp\Biggl(-\sum_{k=n_*+1}^{\infty} \frac{a(1-\mathrm{e}^{2\pi \mathrm{i} kh})\overline{\xi}_k + a(1-\mathrm{e}^{-2\pi kh})\xi_k}{2\varepsilon\sigma_k} + \frac{a^2(1-\cos 2\pi kh)}{\varepsilon^2\sigma_k^2}\Biggr) \\ &= \exp\Biggl(\sum_{k=n_*+1}^{\infty} \frac{a\eta_k(1-\cos 2\pi kh) + a\zeta_k \sin 2\pi kh}{\varepsilon\sigma_k} - \frac{a^2(1-\cos 2\pi kh)}{\varepsilon^2\sigma_k^2}\Biggr), \end{aligned}$$

where η_k and ζ_k are i.i.d. standard Gaussian random variables. Thus,

$$EZ_1^2 = \exp\Biggl(\frac{2a^2}{\varepsilon^2} \sum_{k=n_*+1}^{\infty} \frac{1-\cos 2\pi kh}{\sigma_k^2}\Biggr) = \exp\Biggl(\frac{4a^2}{\varepsilon^2} \sum_{k=n_*+1}^{\infty} \frac{\sin^2 \pi kh}{\sigma_k^2}\Biggr).$$

We can estimate the exponent as follows:

$$\begin{aligned} I &\equiv \frac{4a^2}{\varepsilon^2} \sum_{k=n_*+1}^{\infty} \frac{\sin^2 \pi kh}{\sigma_k^2} \le \frac{4a^2}{\underline{\sigma}^2\varepsilon^2} \sum_{k=n_*+1}^{\infty} k^{-2\beta} \sin^2 \pi kh \\ &\le \frac{4a^2}{\underline{\sigma}^2\varepsilon^2}\Biggl(\pi^2 h^2 \sum_{k=n_*+1}^{\lfloor(1/\pi h)\rfloor} k^{-2\beta+2} + \sum_{k=\lfloor(1/\pi h)\rfloor+1}^{\infty} k^{-2\beta}\Biggr) \\ &\le C\frac{a^2}{\underline{\sigma}^2\varepsilon^2}\bigl(h^2 n_*^{-2\beta+3} + h^{2\beta-1}\bigr). \end{aligned} \tag{40}$$

Let us choose for some small $c_3 > 0$,

$$h = c_3 a^{-2} n_*^{2\beta-3} \varepsilon^2 \ln \varepsilon^{-1} \quad \Longleftrightarrow \quad h = c_4 |a|^{-1} \bigl(\varepsilon\sqrt{\ln \varepsilon^{-1}}\bigr)^{(2s+3)/(2s+2\beta)}.$$

In view of (40), for ε small enough, $I \le c_5 \ln \varepsilon^{-1}$ with some small constant c_5. We conclude that for such choice of h and n_*

$$E_0 Z_1^2 \le \varepsilon^{-c_5}.$$

Let us now use Lemma 4. We set

$$\phi_0(\varepsilon) = c_6 |a|^{-1} \bigl(\varepsilon\sqrt{\ln \varepsilon^{-1}}\bigr)^{(2s_0+3)/(2s_0+2\beta)}, \qquad \phi_1(\varepsilon) = c_6 |a|^{-1} \bigl(\varepsilon\sqrt{\ln \varepsilon^{-1}}\bigr)^{(2s_1+3)/(2s_1+2\beta)},$$

so that

$$\delta = \frac{\phi_0(\varepsilon)}{\phi_1(\varepsilon)} \le \left(\varepsilon\sqrt{\ln\varepsilon^{-1}}\right)^{(4\beta-6)(s_0-s_1)/((2s_1+2\beta)(2s_0+2\beta))} \le c_7\varepsilon^{c_8(s_0-s_1)}$$

(recall that $\beta > 3/2$). We can now choose the constants in a way to obtain $\delta E_0 Z_1 < 1$. When applying Lemma 4 we get the required statement.

References

[1] L. Brown and M. Low. A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.* **24** (1996) 2524–2535. MR1425965

[2] A. Goldenshluger, A. Juditsky, A. Tsybakov and A. Zeevi. Change-point estimation from indirect observations. 1. Minimax complexity. *Ann. Inst. Henri Poincaré* **44** (2008) 787–818.

[3] O. V. Lepskii. A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** (1990) 454–466. MR1091202

[4] O. V. Lepskii and V. Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** (1997) 2512–2546. MR1604408

[5] A. Zygmund. *Trigonometric Series*, Vol. I, 2nd edition. Cambridge Univ. Press, 1959. MR0107776