Electronic Journal of Statistics
Vol. 3 (2009) 1322–1359
ISSN: 1935-7524
DOI: 10.1214/08-EJS204

Jump estimation in inverse regression

Leif Boysen, Sophie Bruns and Axel Munk

Institute for Mathematical Stochastics
Georg–August–Universität Göttingen
Goldschmidtstr. 7
37077 Göttingen
Germany
e-mail: boysen@math.uni-goettingen.de; sbruns@math.uni-goettingen.de
munk@math.uni-goettingen.de

Abstract: We consider estimation of a step function f from noisy observations of a deconvolution $\phi * f$, where ϕ is some bounded L_1-function. We use a penalized least squares estimator to reconstruct the signal f from the observations, with penalty equal to the number of jumps of the reconstruction. Asymptotically, it is possible to correctly estimate the number of jumps with probability one. Given that the number of jumps is correctly estimated, we show that for a bounded kernel ϕ the corresponding estimates of the jump locations and jump heights are $n^{-1/2}$ consistent and converge to a joint normal distribution with covariance structure depending on ϕ. As special case we obtain the asymptotic distribution of the least squares estimator in multiphase regression and generalizations thereof. Finally, singular integral kernels are briefly discussed and it is shown that the $n^{-1/2}$-rate can be improved.

AMS 2000 subject classifications: Primary 62G05, 62G20; secondary 42A82, 46E22.
Keywords and phrases: Change-point estimation, deconvolution, jump estimation, asymptotic normality, positive definite functions, native Hilbert spaces, entropy bounds, reproducing kernel Hilbert spaces, singular kernels, total positivity.

Received March 2008.

1. Introduction

Assume we have observations from a regression model given by

$$Y = \Big((\Phi f)\big(x_i\big) + \varepsilon_i \Big)_{i=1}^{n}, \tag{1}$$

where $\Phi f = \phi * f$ denotes convolution of some L_1-functions ϕ and f, $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. mean zero random variables with finite second moment, and $x_i = x_{i,n}$, $i = 1, \ldots, n$ is a triangular scheme of design points in $[0, 1]$, where we suppress throughout the dependence on n and simply write x_i. In the following we denote model (1) as an inverse (deconvolution) regression model and we assume throughout that ϕ is known. Suppose the objective function $f : \mathbb{R} \to \mathbb{R}$ is locally constant, i.e. a piecewise constant function with at most k jumps given

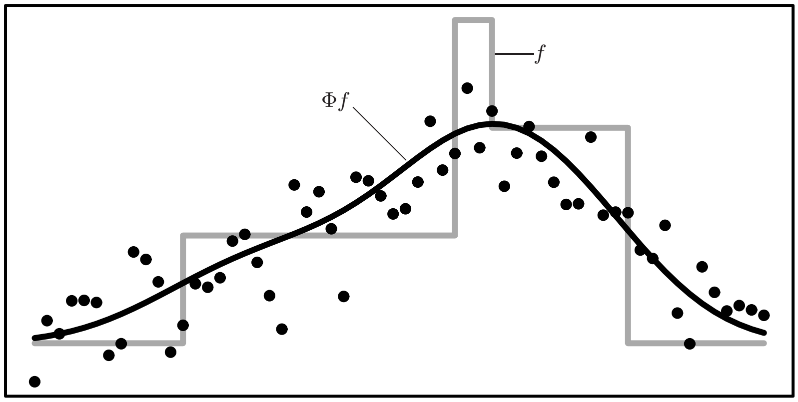

Fig 1. *Noisy observations of a blurred step function. The dots represent the observations and the black line the blurred function Φf, where Φ represents convolution with the gauss kernel. The gray line shows the original step function f, which is to be estimated.*

by

$$f(x) = \sum_{i=1}^{k+1} b_i 1_{[\tau_{i-1},\tau_i)}(x)\,, \tag{2}$$

$0 < \tau_1 < \cdots < \tau_k < 1$ and $k \in \mathbb{N}$ possibly unknown (see Figure 1). Throughout this paper we will set $\tau_0 := 0$ and $\tau_{k+1} := 1$. Hence, we will assume that all jumps occur in $[0,1]$, though the functions are defined on the reals for technical reasons. To keep the presentation simple, we further assume that f vanishes outside of $[0,1]$. We mention, however, that our results remain valid as long as $f \in L_1 \cap L_2$, where f has a continuous extension at the boundaries of $[0,1]$. From Figure 1 the difficulty of estimating jumps in inverse regression is apparent: Due to the smoothing by ϕ, jumps only appear as small smooth changes in Φf.
In this paper we show that the joint least squares estimator $\hat{\theta}_n$ of jumps and heights

$$\theta = (b_1, \tau_1, b_2, \tau_2, \ldots, b_k, \tau_k, b_{k+1}) \tag{3}$$

is $n^{-1/2}$ consistent and follows a multivariate normal limit law. Roughly speaking, convolution leads asymptotically to an average of an infinite number of r.v.'s around the jump location and hence a central limit theorem applies. This means that for a bounded kernel ϕ (under proper identifiability conditions) jump estimation for locally constant functions in inverse regression is a regular parametric estimation problem (see e.g. van der Vaart (1998) Theorems 5.21,31 for an elegant formulation of assumptions that do not require differentiability of the score function, which indeed is not valid in our case). This is in strong contrast to the case of direct regression, i.e. where Φ in (1) is the identity (see e.g. Carlstein and Müller (1994); Korostelev and Tsybakov (1993); van de Geer (1988); Müller (1992); Müller and Stadtmüller (1999); Yakir et al. (1999); Birgé and Massart (2007)). In this case it is known that the LSE converges at the (minimax) n^{-1} rate (see Korostelëv (1987)) and its distribution (after centering and scaling with n) is given as the minimizer of a certain random walk process (Yao and Au (1989)). Moreover, in direct regression problems the estimators of the jump heights and the locations of the jumps are mutually asymptotically independent. In inverse regression the situation is completely different. In general, all components of $n^{1/2}(\hat{\theta}_n - \theta)$ will be dependent asymptotically (depending on the kernel ϕ).

In fact, a main motivation to consider the space of locally constant functions as in (2) stems from the observation that, in general, deconvolution is a difficult problem, which is reflected by minimax rates of convergence that can be arbitrarily slow, e.g. $(\log n)^{-\beta}$ rates as for supersmooth (e.g. Gaussian) deconvolution when f is a function of Hölder smoothness $\beta > 0$ (cf. Fan (1991); Cavalier and Tsybakov (2002); Butucea and Tsybakov (2008a,b) among many others). However, we stress that in many practical situations, Gaussian deconvolution is still applied, leading to satisfactory results (see e.g. Bissantz et al. (2007) for an example in astrophysics). At first glance this seems to be contradictory. The reason is that often a minimax approach reflects a rather pessimistic point of view, in particular in large function classes such as Sobolev or Besov spaces. These spaces contain functions which are difficult to recover in deconvolution, although in many practical situations they can be excluded apriori. In fact, often more restrictive modeling is possible (and required) which may result in reasonably good rates of convergence. In this paper we will show that the assumption of locally constant functions will allow for an $n^{-1/2}$-rate of convergence under rather general conditions on the convolution kernel. These conditions are borrowed from the theory of radial basis functions in native Hilbert spaces and from total positivity. They cover super-smooth functions such as the Gauss-kernel, polynomial kernels $\phi(x) = x^p \, 1_{[0,1)}(x)$ with $p = 0, 1, \ldots$ and continuous symmetric functions ϕ which have a Fourier transform $\hat{\phi}$ not decaying faster than a rational function, i.e. $C(1 + |x|^{n_0})^{-1} \leq |\hat{\phi}(x)|$ for some $n_0 \in \mathbb{N}$, $C > 0$ and all $x \in \mathbb{R}$.

So far, we have assumed that the number of jumps is known in advance. If the number of jumps is unknown, we furthermore show that, under the additional assumption of subgaussian tails of the error distribution, the number of jumps can be asymptotically estimated correctly with probability one. This property has been denoted by Fan and Li (2004) as the “oracle property” which guarantees that the asymptotic distribution of the least squares estimator remains the same as for an unknown number of jumps. We are aware, of course, of possible pitfalls of the “oracle property” which transfer to our case as well. We refer to Leeb and Pötscher (2006, 2008) for a careful discussion.

We mention that our results can also be extended to more general Fredholm integral operators of the type $\Phi f = \int K(x, y) f(y) dy$ with continuous kernel $K : [0, 1] \times \mathbb{R} \to \mathbb{R}$ (see Boysen, 2006). For reasons of simplicity and ease of notation we do not treat this case here.

A classical model which fits into our framework was given by Quandt (1958). He introduced a linear regression model which obeys two separate regimes and where the change-point is not known. This model is called two-phase regression and inference in this setting was studied by Quandt (1960), Sprent (1961), Hinkley (1969) and more recently by van de Geer (1988), Yakir et al. (1999) and Koul et al. (2003), among others. If the objective function f is assumed to be continuous, two-phase regression can be modeled by an inverse regression model with a polynomial kernel with $p = 0$, i.e. $\phi(x) = 1_{[0,1)}(x)$. In this setting the $n^{-1/2}$ rate and the asymptotic distribution were derived by Hinkley (1969) and – for more general segmented regression models – by Feder (1975). From the perspective of statistical inverse problems their results are quite natural to understand: multiphase regression corresponds to estimation of a jump function in a noisy Volterra equation where the location of jumps correspond to the kinks of the multiphase regression function.

Our results generalize known results on the estimation of the intersection in two phase regression to the case where the objective function has an arbitrary number of phases and is piecewise polynomial of order $p+1$, with p continuous derivatives and a $(p+1)$-th derivative, which is a step function. For piecewise linear regression ($p = 1$) in a deconvolution context this problem occurs in rheology where the relaxation time spectrum has to be estimated from measurements of the dynamic moduli of materials (cf. Roths et al., 2000). Other applications stem from biophysics, where the ion-channel activity of lipid membranes are measured by impedance spectroscopy and the jump locations indicate different opening states (cf. Schmitt et al., 2006; Römer et al., 2004). We obtain the result that the rate of estimating the change-point does not depend on p, whereas in general nonparametric regression settings, the convergence rates for estimating a jump in the p-th derivative become slower as p grows (see Goldenshluger et al., 2006) and the references given there.

The first one to investigate the change-point problem in the framework of a statistical inverse problem was Neumann (1997), who considered the estimation of a change-point in a density deconvolution model $Y = X + \xi$ with known error density f_ξ. He treated the case that the density of X is bounded, has one jump at τ and is Lipschitz continuous elsewhere. In this setting τ can be estimated at a rate of $\min(n^{-1/(2\beta+1)}, n^{-1/(\beta+3/2)})$, provided the tails of the Fourier transform $\widehat{f_\xi}(x)$ decrease at a rate of $|x|^{-\beta}$. Moreover, he proved that these rates are optimal in a minimax sense. This result was extended by Goldenshluger et al. (2006) (in a white noise model) to classes of functions f which can be written as a sum of a step function and a function with smooth m-th derivative. They showed that in this case the minimax rates are of order $\min(n^{-1/(2\beta+1)}, n^{-(m+1)/(2\beta+2m+1)})$. If the smooth part of the function of interest belongs to a Paley-Wiener class, they show that a rate of $\min(n^{-1/2}, n^{-1/(2\beta+1)})$ can be obtained up to a logarithmic factor. Hence, the specific choice of locally constant jump functions in (2) used in this work comes close to the super-smooth case for $\beta \geq 1/2$, without the additional logarithmic factor. This seems quite natural, because piecewise constant functions are even more regular than piecewise super-smooth functions. Moreover, we will indicate that similar rates hold in the case of $\beta < 1/2$ if the assumption on the boundedness of the kernel is dropped (see Remark 3). Recently, Goldenshluger et al. (2008a,b) generalize the above mentioned results to a unifying framework of sequence space models covering delay and amplitude estimation, estimation of change-points in derivatives and change point estimation in a convolution white noise model. Finally, we stress another difference to our approach: These and other optimal estimation methods for jump points are performed in two steps, a localization step and an optimization step. For our method the estimator is computed in one (nonlinear) optimization step. We mention, however, that this may lead to difficult global numerical optimization problems, because in general the underlying (penalized) least squares problem need not be convex, similar as in nonlinear regression or more generally, for M-estimators. This is in contrast to the direct case, where computation of the location of jumps can be performed in $O(n^2)$ steps (Friedrich et al., 2008). A careful discussion is beyond the scope of this paper and will be treated separately.

This work is structured as follows. Section 2 gives some basic notation and the main assumptions. The estimate and its asymptotic properties are given in section 3 and the proof of the main result can be found in section 4. Finally, in section 5 we derive the required results from the theory of radial basis functions which yields sufficient conditions on ϕ for the asymptotic normality of the LSE.

2. Model assumptions and notation

2.1. Notation

Let $\gamma_0 = 0$, $\gamma_{k+1} = 1$ and define

$$\Gamma_k := \{(\gamma_0, \gamma_1, \dots, \gamma_{k+1}) : 0 < \gamma_1 < \dots < \gamma_k < 1\}$$

as the set of possible jumps of f in (1). Further denote the corresponding function space of locally constant functions with at most k jumps by

$$T_k := \left\{ \sum_{i=1}^{k+1} b_i 1_{[\tau_{i-1}, \tau_i)}(x) : \tau \in \Gamma_k, \, b_i \in \mathbb{R} \right\}, \ k \in \mathbb{N}.$$

Write $T_\infty := \bigcup_{k=1}^\infty T_k$ for the set of all step functions on $\mathbb{R}$ with a finite but arbitrary number of jumps, where we exclude an isolated jump at the end points of the interval $[0,1]$. Recall, that outside of $[0,1]$ these functions vanish. Let $T_{k,R} = \{g \in T_k : \|g\|_\infty \le R\}$ as well as $T_{\infty,R} := \bigcup_{k=1}^\infty T_{k,R}$ the corresponding spaces of uniformly bounded functions for some $R > 0$. If not mentioned otherwise, the restriction of these spaces to $[0,1]$ are considered to be subspaces of $L_2([0,1])$.

As usual, $\|\cdot\|_2$ will denote the $L_2(\mathbb{R})$ norm, $\langle \cdot, \cdot \rangle_2$ the corresponding inner product and $\|\cdot\|_\infty$ the supremum norm. Additionally, define the empirical norm $\|\cdot\|_n$ and the empirical inner product $\langle \cdot, \cdot \rangle_n$ by

$$\|f\|_n^2 := \frac{1}{n} \sum_{i=1}^n f^2(x_i) \qquad \text{as well as} \qquad \langle f, g \rangle_n := \frac{1}{n} \sum_{i=1}^n f(x_i) g(x_i),$$

where $x_1, \ldots, x_n$ are the design points. From the context it will be clear always that the index $n = 2$ refers to the L_2-norm. Similarly set

$$\|y\|_n^2 := \frac{1}{n}\sum_{i=1}^n y_i^2 \qquad \text{as well as} \qquad \langle y, z\rangle_n := \frac{1}{n}\sum_{i=1}^n y_i z_i$$

for $y, z \in \mathbb{R}^n$. Write $g(t_+) := \lim_{x \searrow t} g(x)$ for the right limit of g in t and $g(t_-) := \lim_{x \nearrow t} g(x)$ for the corresponding left limit. For some proper function $g : \mathbb{R} \to \mathbb{R}$ define the set of jump points of g as

$$\mathcal{J}(g) := \{t \in [0,1] : g(t_-) \neq g(t_+)\} \tag{4}$$

and $J_\#(f) := \#\mathcal{J}(f) + 1$, where $\#\mathcal{J}(f)$ denotes the number of jumps. Define the distance of some point $a \in \mathbb{R}$ to the set $B \subset \mathbb{R}$ as

$$\mathrm{d}(a, B) = \inf_{b \in B} |a - b|$$

and, slightly abusing notation, the Hausdorff distance of two sets A, B as

$$\mathrm{d}(A, B) = \max\{\sup_{a \in A} \mathrm{d}(a, B)\,, \sup_{b \in B} \mathrm{d}(b, A)\}\,.$$

Finally, for ease of notation for any $a, b \in \mathbb{R}$, $[a, b]$ and (a, b) always denote the intervals $[\min(a, b), \max(a, b)]$ and $(\min(a, b), \max(a, b))$, respectively.

2.2. Assumptions

Assumptions on the error If the number of jumps is known the following basic assumption is sufficient to deduce the $n^{-1/2}$ rates of convergence for the least squares estimates.

Assumption A. The array $(\varepsilon_1, \ldots, \varepsilon_n)$ consists of independent identically distributed random variables with mean zero for every n. Additionally, assume

$$\mathrm{E}(\varepsilon_1^2) = \sigma^2 < \infty\,.$$

If the number of jumps of the objective function is unknown, we will additionally need that the error satisfies the following subgaussian condition.

(A1) There exists some $\alpha > 0$ such that $\mathrm{E}(\exp(\varepsilon_1^2/\alpha)) < \infty$.

Assumptions on the kernel We require the following independence assumption.

Assumption B. Let

$$\Delta_\phi(x, a, b) := \begin{cases} \displaystyle\int_a^b \phi(x - y)dy & b \neq a\,, \\ \phi(x - a) & b = a\,. \end{cases} \tag{5}$$

Assume that $\phi \in L_1(\mathbb{R}) \cap L_\infty(\mathbb{R})$ is piecewise Lipschitz continuous with finitely many jumps. Additionally, the functions

$$\Delta_\phi(x, \tau_0, \tau_1)\,,\ \dots\,,\ \Delta_\phi(x, \tau_k, \tau_{k+1})\,,\ \Delta_\phi(x, \tau_1, \tau_1)\,,\ \dots\,,\ \Delta_\phi(x, \tau_k, \tau_k)$$

are linearly independent for every choice of $k \in \mathbb{N}$ and

$$0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_k < \tau_{k+1} = 1, \tag{6}$$

Remark 1. We mention that Assumption B implies that the functions

$$(\Phi 1_{[\tau_0,\tau_1)})(\cdot), (\Phi 1_{[\tau_1,\tau_2)})(\cdot), \dots, (\Phi 1_{[\tau_k,\tau_{k+1})})(\cdot) \tag{7}$$

are linearly independent for any parameter vector $(\tau_0, \dots, \tau_{k+1})$ as in (6).

The following Theorem 2.2 gives some general conditions, which are sufficient for ϕ to satisfy Assumption B. To this end recall the definition of extended sign-regularity as given in Karlin and Studden (1966). See also Karlin (1968) for many examples.

Definition 2.1. Fix $k \in \mathbb{N}$ and let $\phi \in C^{k-1}(\mathbb{R})$, $t_1 \le \cdots \le t_k$, $t_i \in \mathbb{R}$. Let for $j = 1, \dots, k$

$$\phi_{j,t_1,\dots,t_k}(x) = \left\{ \begin{array}{ccl} \phi(x - t_j) & : & t_{j-1} < t_j \\ \phi^{(r)}(x - t_j) & : & t_{j-r-1} < t_{j-r} = \cdots = t_j\,, \end{array} \right.$$

where t_0 is set to $-\infty$ and $1 \le r < k$. Moreover, define

$$\phi^* \left(\begin{array}{c} s_1, \dots, s_k \\ t_1, \dots, t_k \end{array} \right) = \det \left(\phi_{j,t_1,\dots,t_k}(s_i) \right)_{i,j=1}^{k}.$$

The function ϕ will be called extended sign regular of order k (ESR_k) on $\mathbb{R}$, provided that for each $r = 1, \dots, k$ there exists $\varepsilon_r \in \{-1, 1\}$ such that

$$\varepsilon_r \phi^* \left(\begin{array}{c} s_1, \dots, s_r \\ t_1, \dots, t_r \end{array} \right) > 0\,,$$

for all choices of $s_1 < s_2 < \cdots < s_r$ and $t_1 \le t_2 \le \cdots \le t_r$ with $s_i, t_i \in \mathbb{R}$. It is called ESR_∞ if this holds for any $k \in \mathbb{N}$.

Theorem 2.2. *The function ϕ satisfies Assumption B if one of the following conditions is satisfied.*

(i) $\phi \in C(\mathbb{R}) \cap L_1(\mathbb{R})$ *is a symmetric real-valued function with Fourier transform* $\widehat{\phi}(x) \ge 0$*, such that there exists* $n_0 \in \{0, 1, 2, \dots\}$ *and* $C > 0$ *with*

$$C(1 + |x|^{n_0})^{-1} \le |\widehat{\phi}(x)| \qquad \text{for all}\ x \in \mathbb{R}\,. \tag{8}$$

(ii) ϕ *is extended sign regular of order* $2k+2$ *on* $\mathbb{R}$*, with* $0 < \int \phi(x)dx < \infty$.

(iii) The function ϕ *is given by*

$$\phi(x) = \begin{cases} x^p & x \in [0,1] \\ 0 & else \end{cases} \qquad p \in \{0, 1, 2, \dots\}\,.$$

Examples of kernels satisfying Assumption B The most prominent example is the Gauss kernel $\phi(x) = (2\pi\sigma^2)^{-1/2}\exp(-(x/\sigma)^2/2)$ for some fixed $\sigma > 0$. In fact this kernel is well known to be ESR_∞ (see Section 3, Example 5 in Karlin and Studden, 1966).

Examples of kernels which satisfy condition (i) of Theorem 2.2 are the Laplace kernel $\phi(x) = \exp(-|x|)/2$, the kernel $\phi(x) = \cos(x)\exp(-|x|)$ and kernels of the type $\phi(x) = (1-|x|)_+^p$ for $p = 2, 3, \ldots$ where x_+ denotes the positive part of x. Moreover, the convolution of any two kernels ϕ_1, ϕ_2 satisfying condition (i) clearly also satisfies this condition.

As mentioned in the introduction it is possible to extend the identifiability Assumption B to general integrable kernels $k(x, y)$. For example, the Γ-kernels restricted to the positive reals

$$k(x,y) = \frac{1}{\Gamma(\alpha)y^\alpha} x^{\alpha-1}\exp\{-xy\}, \qquad \alpha > 0$$

are easily seen to be ESR_∞ because $\exp\{xy\}$ has this property. However, in this paper we will restrict to convolution and we postpone the issue of general integral kernels to a subsequent paper.

Assumptions on the design points We make the following assumption on the design points.

Assumption C. There exists a function $h : [0,1] \to [c_l, c_u]$ with $0 < c_l < c_u < \infty$ and $\int_0^1 h(x)dx = 1$, such that

$$\frac{i}{n} = \int_0^{x_{(i)}} h(x)dx + \delta_i$$

for all $i = 1, \ldots, n$, with $\max_{i=1,\ldots,n} |\delta_i| = O_P(n^{-1/2})$. Here $x_{(i)}$ denotes the i-th order statistic of $x_1, \ldots, x_n$.

Moreover, the design points $x_1, \ldots, x_n$ are independent of the error terms $\varepsilon_1, \ldots, \varepsilon_n$.

Dümbgen and Johns (2004) use a similar assumption on the design points. Note that the above assumption covers random designs as well as fixed designs generated by a regular density in the sense of Sacks and Ylvisaker (1970). It is easy to see that Assumption C holds if $x_1, x_2, \ldots$ are i.i.d. random variables with density g satisfying $\mathrm{supp}(g) = [0,1]$ and $0 < c_l < g(x) < c_u < \infty$ for all $x \in [0,1]$ and some $c_l, c_u \in \mathbb{R}$. If the design points $x_1, \ldots, x_n$ are nonrandom, the $O_P(n^{-1/2})$ term above is to be understood as $O(n^{-1/2})$. For h constant 1, this includes the case of a uniform (deterministic) design $x_{(i)} = i/n$. In fact, for fixed design points all results can be obtained essentially in the same way. The only argument which has to be slightly modified, is the one based on the law of the iterated logarithm in the proof of Lemma 4.14. Note that the respective inequalities remain valid because the random variables $\epsilon_{i,n}$ defining the error terms in a triangular scheme can be replaced by a sequence of i.i.d. random variables without changing the distribution.

3. Estimate and asymptotic results

Estimate Define the restricted least squares estimate $\hat{f}_n$ as approximate minimizer of the empirical L_2 distance to the data in the space $T_{k,R}$. More precisely $\hat{f}_n \in T_{k,R}$ and

$$\|\Phi\hat{f}_n - Y\|_n^2 \leq \min_{g\in T_{k,R}} \left(\|\Phi g - Y\|_n^2\right) + o_p(n^{-1})\,. \tag{9}$$

The minimizer of the functional on the right hand side always exists (compare Lemma 4.7). Note that we do not assume that the minimum is attained, but only that the functional above can be minimized up to some term of order $o_p(n^{-1})$. It does not need to be unique. This assumption allows for numerical approximation of the minimizer and gives an intuition of the needed precision for the asymptotic results to be valid. The restriction to functions with $\|f\|_\infty \leq R$ is a technical assumption, i.e. that some upper bound of the objective function is known beforehand.

Note that any estimator $\hat{f}_n$ has a representation as

$$\hat{f}_n(x) = \sum_{i=1}^{k+1} \hat{b}_i 1_{[\hat{\tau}_{i-1},\hat{\tau}_i)}(x)\,, \tag{10}$$

with vectors $\hat{b} = (\hat{b}_1, \ldots, \hat{b}_{k+1})^t$ and $\hat{\tau} = (\hat{\tau}_0, \ldots, \hat{\tau}_{k+1})^t$, where we formally set $\hat{\tau}_0 := 0$, $\hat{\tau}_{k+1} := 1$. This is the approximate least squares estimates (in the sense of (9)) of the true parameter vectors b and τ given by equation (2). Set

$$\hat{\theta}_n = (\hat{b}_1, \hat{\tau}_1, \hat{b}_2, \hat{\tau}_2, \ldots, \hat{b}_k, \hat{\tau}_k, \hat{b}_{k+1}) \tag{11}$$

as the least squares estimator of the combined parameter vector θ given in (3).

If the number of jumps is unknown, a different estimate is needed. In this case, assume that the penalized least squares estimate $\hat{f}_{\lambda_n}$ satisfies $\hat{f}_{\lambda_n} \in T_{\infty,R}$ and is defined as any solution of

$$\|\Phi\hat{f}_{\lambda_n} - Y\|_n^2 + \lambda_n J_\#(\hat{f}_{\lambda_n}) \leq \min_{g\in T_{\infty,R}} \left(\|\Phi g - Y\|_n^2 + \lambda_n J_\#(g)\right) + o_p(n^{-1})\,, \tag{12}$$

where $\lambda_n > 0$ is some smoothing parameter, s.t. $\lambda_n \to 0$ as $n \to \infty$.

Asymptotic results Before we state the main result, we first define the map $\nu : [0,1] \mapsto \mathbb{R}^{2k+1}$ by

$$\nu(x) = \begin{pmatrix} \Delta_\phi(x, \tau_0, \tau_1) \\ (b_1 - b_2)\Delta_\phi(x, \tau_1, \tau_1) \\ \Delta_\phi(x, \tau_1, \tau_2) \\ \vdots \\ (b_k - b_{k+1})\Delta_\phi(x, \tau_k, \tau_k) \\ \Delta_\phi(x, \tau_k, \tau_{k+1})) \end{pmatrix}, \tag{13}$$

and the $(2k+1) \times (2k+1)$ matrix V by its entries

$$(V)_{ij} = \int_0^1 (\nu(x)\nu(x)^t)_{ij}\, h(x)dx\,. \tag{14}$$

Here h is the design density satisfying Assumption C. Now we are able to formulate the asymptotic result for the least squares estimator.

Theorem 3.1. *Suppose the Assumptions A, B and C are met. Let $\hat{f}_n$ and V be given by (10) and (14), respectively. Set θ as the parameter vector of f given in (3), and $\hat{\theta}_n$ as the corresponding vector of estimates defined by (11). Given (9) and model (1), then*

(i) V *is positive definite.*
(ii) $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma^2 V^{-1})$.

Moreover,

(iii) $\|\Phi f - \Phi \hat{f}_n\|_2 = O_P(n^{-1/2})$.
(iv) $\mathrm{d}(\mathcal{J}(f), \mathcal{J}(\hat{f}_n)) = O_P(n^{-1/2})$.
(v) $\|f - \hat{f}_n\|_2 = O_P(n^{-1/4})$.

Note that parts (iv),(v) of the last theorem highlight the 'degree of ill posedness of 1/2' induced by the convolution compared to ordinary jump regression. Both rates become slower by an exponent 1/2 whereas for the 'forward problem' (in the terminology of inverse problems) in part (iii) the rates remain the same.

The following theorem implies that the penalized estimator $\hat{f}_{\lambda_n}$ and the restricted least squares estimator $\hat{f}_n$ asymptotically coincide, i.e. the number of jumps in T_∞ is asymptotically correctly estimated with probability one. Hence, conditionally on the correct number of jumps, Theorem 3.1 remains valid. We mention, however, that the finite sample behavior of the LSE may be affected significantly by this 'model selection step'. For a discussion of this issue in the context of regression see Leeb and Pötscher (2006, 2008).

Theorem 3.2. *Suppose condition (A1), (12) and the assumptions of Theorem 3.1 are satisfied. If $\lambda_n \to 0$ and $\lambda_n n^{1/(1+\epsilon)} \to \infty$ for some $\epsilon > 0$ as $n \to \infty$, then*

$$\lim_{n\to\infty} P\big(\#\mathcal{J}(\hat{f}_{\lambda_n}) = \#\mathcal{J}(f)\big) = 1\,.$$

The proofs of Theorem 3.1 and 3.2 can be outlined as follows. For a known number of jumps an entropy argument yields consistency of the least squares estimator. It is possible to represent the estimator as the minimizer of a stochastic process, which allows for a local stochastic expansion. This can be used to derive asymptotic normality. If the number of jumps is unknown, techniques from empirical process theory can be used to reduce this asymptotically to the case when the number is known.

The details of proof are given in several steps in section 4.

Remarks and Extensions

Remark 2. (Noisy Fredholm equations). All results of this chapter can also be shown for more general integral operators of the type $\Phi f = \int K(x,y)f(y)dy$ with continuous kernel $K : [0,1] \times \mathbb{R} \to \mathbb{R}$ satisfying $\sup_{x\in[0,1]} \|K(x,\cdot)\|_{L_1} < \infty$. In this case in definition (5) $\phi(x-y)$ has to be replaced by $K(x,y)$. Assumption B can be formulated in the same way. Details will be postponed to a separate paper.

Remark 3. (Singular kernels). If the integral kernel is unbounded, faster rates than $O_P(n^{-1/2})$ for estimating the jump location can be achieved. Indeed if ϕ is an Abel type kernel $\phi_\alpha(x) = x^{-\alpha}1_{(0,\infty)}(x)$ for $\alpha \in (0,1)$ then a jump can be recovered at a rate of $O_P(n^{-1/\min(2,3-2\alpha)})$. For details we refer to a separate publication. Interestingly, the n^{-1} sampling rate is achieved as $\alpha \to 1$, which is well known to be the best possible rate in direct regression for the estimation of a jump. Further, for singular kernels with a spike stronger than $|x|^{-1}$ jumps can be estimated with the same rate as for the direct case, which is achieved as $\alpha \to \infty$.

Note that the "elbow" in the rates of convergence occurs at $\alpha = 1/2$, and that the $n^{-1/2}$ rate holds for the case where ϕ_α is square integrable on bounded intervals.

This corresponds to findings of Neumann (1997) and Goldenshluger et al. (2006), who also observe an elbow in the rate of convergence of recovering a change point in an inverse problem at $\beta = 1/2$, if the Fourier transform of $\widehat{\phi}(x)$ decreases at rate of $|x|^{-\beta}$. Goldenshluger et al. (2006) give a rate of $O_P(n^{-1/\min(2,2\beta+1)})$ up to a logarithmic term if the smooth part of the function of interest is in a Paley-Wiener class. From

$$\left|\widehat{\phi}_\alpha(x)\right| = |x|^{-1+\alpha}\Gamma(1-\alpha)$$

it follows that the "elbow" for $\beta = 1/2$ can be identified with the elbow for $\alpha = 1/2$.

Remark 4. (Misspecification of the model). We will briefly discuss the behavior of the estimator $\hat{f}_n$, when the true function f from model (1) is not an element of $T_{k,R}$. In fact, under certain conditions $\hat{f}_n$ converges to a step function $t \in T_{k,R}$, s.t. Φt is a best approximation of Φf.

Lemma 3.3. *Assume $f \in L_2([0,1])$ with $\|f\|_\infty \le R$ and there exists a unique $t \in T_{k,R}$, such that*

$$\|\Phi f - \Phi t\|_2 = \min_{g\in T_{k,R}} \|\Phi f - \Phi g\|_2. \tag{15}$$

Furthermore, Assumption C holds with $h \equiv 1$, i.e. $x_{(i)} = i/n + \delta_i$. For $\hat{f}_n$, the least squares estimator from (9), it holds

$$\|\Phi t - \Phi \hat{f}_n\|_2 = o_P(1)$$

and

$$\|t - \hat{f}_n\|_2 = o_P(1). \tag{16}$$

Condition (15) is well known in nonlinear approximation theory (e.g. see Braess (1986)). Note that, due to injectivity of the operator Φ, the assumption of a unique minimizer t is equivalent to the assumption of existence of a unique $\Phi t \in \Phi(T_{k,R})$, which minimizes the L_2-distance from Φf. It would be of great interest, to relax the assumption of an approximate equidistant design in Lemma 3.3, as well as improving (16), e.g. to an asymptotic distributional law for $\hat{f}_n$. This is postponed to separate work.

4. Proof of Theorem 3.1 and 3.2

We start with some technical results and then give some entropy bounds on the spaces of interest. This requires tools of empirical process theory to prove consistency of the estimates. Afterwards we give a local stochastic expansion of the minimized process and use this to derive asymptotic normality. Finally we again use techniques from empirical process theory to show that the penalized estimate asymptotically coincides with the restricted least squares estimate. Note that Assumption B is needed to assure identifiability for any number k of jumps as well as positive definiteness of the asymptotic covariance matrix V.

4.1. Some technical lemmata

In the following we derive some properties of the mapping Φ restricted to the space of step functions.

Definition 4.1. A family of (possibly random) functions h_n, $n \in \mathbb{N}$ on a domain $D \subset \mathbb{R}$ is said to be equi-Lipschitz continuous with constant C, if for any $\delta > 0$,

$$\sup_{n\in\mathbb{N}} \sup_{x\in D} |h_n(x+\delta) - h_n(x)| \leq C\,\delta \qquad \text{(a.e.)}.$$

Lemma 4.2. *Let ϕ be a bounded integral kernel, then the following holds true.*

(i) For all $\epsilon > 0$ there exists $0 < C_0 < \infty$ such that for all $f \in T_\infty$

$$\|\Phi f\|_n^2 \leq C_0 \|f\|^2_{L_2([0,1])}\,.$$

(ii) For all $\epsilon > 0$ the map $\Phi : (\,T_k, \|\cdot\|_{L_2([0,1])}\,) \to L_2([0,1])$ is continuous.

(iii) Given Assumption B, $\Phi : T_k \to L_2([0,1])$ is one-to-one.

(iv) The function (Φf) is Lipschitz continuous on $\mathbb{R}$ for all $f \in T_\infty$.

(v) For functions $f_n \in T_{k,R}$, $n \in \mathbb{N}$ the family $h_n = \Phi f_n$, $n \in \mathbb{N}$ is equi-Lipschitz continuous with Lipschitz-constant $C_\phi = 2kR||\phi||_{L_\infty}$.

Proof. By Assumption B we have that $\|\phi\|_{L_\infty} = C < \infty$. Recall that f is 0 on $(-\infty, 0)$ and $[1, \infty)$. This gives

$$\begin{aligned} \|\Phi f\|_n^2 &\leq \frac{1}{n}\sum_{i=1}^n \int_0^1 f^2(y)\phi^2(x_i - y)dy \\ &\leq C^2 \int_0^1 f^2(y)dy, \end{aligned}$$

which proves (i). Similarly we can show $\|\Phi f\|_2 \leq C\|f\|^2_{L_2([0,1])}$ for $f \in T_k$ which gives continuity and hence (ii). As argued in the part on the assumptions on the kernel in section 2.2, (iii) follows from the independence of $\Delta_\phi(\cdot, \tau_i, \tau_{i+1})$.

To prove (iv), note that

$$\begin{aligned} |(\Phi 1_{[a,b)})(x) - (\Phi 1_{[a,b)})(x+\delta)| &= \Big| \int_{x-b}^{x-a} \phi(y)dy - \int_{x+\delta-b}^{x+\delta-a} \phi(y)dy \Big| \\ &\leq 2|\delta| \|\phi\|_{L_\infty}, \end{aligned}$$

for any $x, \delta \in \mathbb{R}$ and $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$. For $f \in T_\infty$ with $\#\mathcal{J}(f) < \infty$, this gives

$$|(\Phi f)(x) - (\Phi f)(x+\delta)| \leq \delta\, 2\, \#\mathcal{J}(f)\, \|f\|_\infty \|\phi\|_{L_\infty}. \tag{17}$$

The assertion (v) follows directly from (17). □

The following lemma provides a link of the empirical and the L_2 norm.

Lemma 4.3. *(i) Suppose h, c_l, c_u and δ_i, for $i = 1, \ldots, n$ satisfy Assumption C and f is piecewise Lipschitz continuous on $[0,1]$, i.e. there exist a partition $I_1, \ldots, I_k, k < \infty$, with $\bigcup_{i=1}^k I_k = [0,1]$ and $I_j \cap I_r = \emptyset$ for $j \neq r$ such that $f|_{I_j}$ is Lipschitz for all $j = 1, \ldots, k$. Then*

$$\int_0^1 f(x)h(x)dx = \frac{1}{n}\sum_{i=1}^n f(x_i) + O_P(n^{-1/2})$$

(ii) Let $\mathcal{F}_{k,R}$ the set of piecewise equi-Lipschitz continuous functions, with at most k jumps, uniformly bounded by $R > 0$. For any sequence f_n in $\mathcal{F}_{k,R}$, it holds

$$O_P(n^{-1/2}) + c_l\|f_n\|_2^2 \leq \|f_n\|_n^2 \leq c_u\|f_n\|_2^2 + O_P(n^{-1/2}), \tag{18}$$

with c_u, c_l depending on the density h only.

Proof. Let $H(x) = \int_0^x h(x)dx$, where h is as in Assumption C. Note that H is strictly monotone and the inverse H^{-1} is well defined on $[0,1]$. For $0 \leq a \leq b \leq 1$ we have that

$$b - a = H(H^{-1}(b)) - H(H^{-1}(a)) = \int_{H^{-1}(a)}^{H^{-1}(b)} h(x)dx \geq c_l(H^{-1}(b) - H^{-1}(a)).$$

Hence H^{-1} is Lipschitz and so is $(f\circ H^{-1})|_{H(I_j)}$ for all $j=1,\dots,k$. By Assumption C we have $H^{-1}(i/n-\delta_i)=x_{(i)}$ with $\nu_n := \max_{i=1,\dots,n}|\delta_i| = O_P(n^{-1/2})$. Assume

$$H^{-1}\Big(\Big[\frac{i-1}{n}\ ,\frac{i}{n}\Big]\Big)\subset I_j \qquad \text{and} \qquad H^{-1}\Big([i/n\ ,i/n-\delta_i]\Big)\subset I_r \tag{19}$$

for some $j,r\in\{1,\dots,k\}$. Here $[i/n\ \ ,i/n-\delta_i]$ is the interval spanned by $i/n, i/n-\delta_i$ as defined in Section 2. Consequently,

$$\begin{aligned} & n\int_{(i-1)/n}^{i/n} f(H^{-1}(x))dx \\ &= f(x_{(i)}) + n\int_{(i-1)/n}^{i/n} f(H^{-1}(x)) - f(H^{-1}(i/n))dx \\ &\qquad + n\int_{(i-1)/n}^{i/n} f(H^{-1}(i/n)) - f(H^{-1}(i/n-\delta_i))dx \end{aligned}$$

and by Lipschitz continuity of $f\circ H^{-1}$ on $[(i-1)/n, i/n]$ and $[i/n\ ,i/n-\delta_i]$ we obtain

$$\left| n\int_{(i-1)/n}^{i/n} f(H^{-1}(x))dx - f(x_{(i)})\right| \le c_f/c_l(1/n+\nu_n) = O_P(n^{-1/2}).$$

Here c_f is the maximum of all k Lipschitz constants for the intervals I_j. For general i, we get

$$\begin{aligned} n\int_{(i-1)/n}^{i/n} f(H^{-1}(x))dx &= f(x_{(i)}) + n\int_{(i-1)/n}^{i/n} f(H^{-1}(x)) - f(H^{-1}(i/n+\delta_i))dx \\ &\ge f(x_{(i)}) - 2\|f\|_\infty\,. \end{aligned}$$

Since f is piecewise Lipschitz continuous, f is bounded in supremum norm on $[0,1]$. Denote the points of discontinuity of f by $\mathcal{J}(f)=\{\vartheta_1,\dots,\vartheta_k\}$. The number of i, which do not satisfy (19) is bounded from above by

$$\begin{aligned} & k + \#\{i : \vartheta_j \in H^{-1}([i/n-\nu_n, i/n+\nu_n]) \text{ for some } j=1,\dots,k\} \\ &\quad = k + \#\{i : H^{-1}(i/n-\nu_n) \le \vartheta_j \le H^{-1}(i/n+\nu_n) \text{ for some } j=1,\dots,k\} \\ &\quad = k + \#\{i : H(\vartheta_j)-\nu_n \le i/n \le H(\vartheta_j)+\nu_n \text{ for some } j=1,\dots,k\} \\ &\quad = k + O(n\nu_n)\,. \end{aligned}$$

By application of the transformation formula and $c_f/c_l, k, \|f\|_\infty < \infty$ we get

$$\begin{aligned} \left|\frac{1}{n}\sum_{i=1}^n f(x_i) - \int_0^1 f(x)h(x)dx\right| &= \left|\frac{1}{n}\sum_{i=1}^n f(x_i) - \int_0^1 f(H^{-1}(x))dx\right| \\ &\le c_f/c_l(n^{-1}+\nu_n) + (k+n\nu_n)n^{-1}\|f\|_\infty \\ &\le (c_f/c_l + \|f\|_\infty)((k+1)n^{-1}+\nu_n), \end{aligned}$$

which proves i). Furthermore, the right hand side of the last equation is uniformly bounded for all $f \in \mathcal{F}_{k,R}$, which implies

$$\sup_{f\in\mathcal{F}_k}\left|\frac{1}{n}\sum_{i=1}^{n} f(x_i) - \int_0^1 f(x)h(x)dx\right| = O(n^{-1}+\nu_n) = O_P(n^{-1/2}).$$

Note, that the functions f^2, s.t. $f \in \mathcal{F}_{k,R}$, are bounded and equi-Lipschitz, too. Finally $c_l \le h(x) \le c_u$ for all $x \in [0,1]$ (Assumption C), this yields ii). □

4.2. Entropy results

To show consistency of the estimates, we wish to apply results from empirical process theory. To this end, let us first introduce some additional notation (cf. van de Geer, 2000).

Given a measure Q, a set of Q-measurable functions $\mathcal{G}$ and a real number $\delta > 0$, define the δ-covering number $N(\delta, \mathcal{G}, Q)$ as the smallest value of N for which there exist functions $g_1, \dots, g_N$ such that for every $g \in \mathcal{G}$ there is a $j \in 1, \dots N$ with

$$\Big(\int (g-g_j)^2 dQ\Big)^{1/2} \le \delta\,.$$

Moreover, define the δ-entropy H of $\mathcal{G}$ as

$$H(\delta,\mathcal{G},Q) = \log N(\delta,\mathcal{G},Q)\,.$$

If Q is the Lebesgue measure we will write $H(\delta,\mathcal{G})$ and $N(\delta,\mathcal{G})$ instead of $H(\delta,\mathcal{G},Q)$ and $N(\delta,\mathcal{G},Q)$. Given design points $x_1,\dots,x_n \in \mathbb{R}$, the empirical measure will be denoted by $Q_n = n^{-1}\sum_{i=1}^n \delta_{x_i}$. Note that $\|\cdot\|_n$ is the norm corresponding to the space $L_2(\mathbb{R}, Q_n)$.

Finally, define the entropy integral

$$J(\delta,\mathcal{G},Q) := \max\Big(\delta\,,\int_0^\delta H^{1/2}(u,\mathcal{G},Q)du\Big)\,.$$

Note that for our purposes, the relevant quantity is the entropy of the space $\mathcal{G}_{k,R} = \{\Phi f : f \in T_{k,R}\}$. However, it is convenient to first calculate the entropy of $(T_{k,R}, \|\cdot\|_{L_2([a,b])})$ and then use Lemma 4.2 to infer on the space $\mathcal{G}_{k,R}$.

Lemma 4.4. *For $-\infty < a < b < \infty$ there exists a constant $C > 0$ independent of δ,k and n, such that*

$$H(\delta,(T_{k,R},\|\cdot\|_{L_2([a,b])})) \le C(k+1)\Big(1+\log\Big(\frac{R(k+1)}{\delta}\Big)\Big)\,.$$

Proof. Define the sets

$$\Delta_\phi(\delta) = \Big\{-R + mc_2\delta : m = 0,\dots,\lceil 2R(c_2\delta)^{-1}\rceil\Big\}$$

and

$$\Gamma(\delta) = \left\{ a + mc_1\delta^2 : m = 1, \dots, \lfloor (b-a)(c_1\delta^2)^{-1} \rfloor \right\},$$

where c_1, c_2 will be defined later. Define the function class $\mathcal{H}(\delta)$ as

$$\begin{aligned}\mathcal{H}(\delta) \quad = \quad & \Bigg\{ g : g(x) = \sum_{i=1}^{k+1} b_i 1_{[\gamma_i - 1, \gamma_i)}(x) : b_i \in \Delta_\phi(\delta), i = 1, \dots, k+1, \\ & \qquad \gamma_0 = a, \gamma_{k+1} = b, \gamma_i \in \Gamma(\delta), \gamma_i < \gamma_{i+1}, i = 1, \dots, k \Bigg\}.\end{aligned}$$

Now for $g_0 \in T_{k,R}$ we can choose $g \in \mathcal{H}(\delta)$ such that $\mathrm{d}(\mathcal{J}(g), \mathcal{J}(g_0)) \le c_1\delta^2/2$, and that for any $x \in [a, b]$ with $\mathrm{d}(x, \mathcal{J}(g)) > c_1\delta^2/2$ we have $(g_0(x) - g(x))^2 \le c_2^2\delta^2/4$. Since g_0 has k jumps between a and b we get

$$\|g_0 - g\|^2_{L_2([a,b])} \le (b-a)c_2^2 \frac{\delta^2}{4} + k(2R)^2 c_1 \frac{\delta^2}{2}.$$

Choosing $c_1 = (4kR^2)^{-1}$ and $c_2 = (b-a)^{-1/2}$ gives $\|g_0 - g\|_2 \le \delta$. Hence $\mathcal{H}(\delta)$ is an δ-covering of $(T_{k,R}, \|\cdot\|_{L_2([a,b])})$. Since

$$\#\mathcal{H}(\delta) = \left\lceil \frac{2R\sqrt{b-a}}{\delta} \right\rceil^{k+1} \left\lceil \frac{(b-a)4kR^2}{\delta^2} \right\rceil^k = O\left(\left(\frac{R(k+1)}{\delta} \right)^{3k+1} \right)$$

the claim is proved. □

Lemma 4.4 directly gives that $(T_{k,R}, \|\cdot\|_{L_2([a,b])})$ is totally bounded for $-\infty < a < b < \infty$. Note that $(T_{k,R}, \|\cdot\|_{L_2([a,b])})$ also contains functions with less than k jumps and hence is closed. Consequently, it is compact.

Corollary 4.5. *The space $(T_{k,R}, \|\cdot\|_{L_2([a,b])})$ is compact for all a, b satisfying $-\infty < a < b < \infty$.*

We will now use the assumptions on the operator Φ or, to be more precise, Lemma 4.2, to deduce bounds on the entropy of the space

$$\mathcal{G}_{k,R}(\Phi) := \{\Phi g : g \in T_{k,R}\}.$$

Corollary 4.6. *Assume Φ satisfies Assumption B. There exists a constant C_2 independent of n,k and R such that*

$$H(\delta, \mathcal{G}_{k,R}(\Phi), Q_n) \le C_2(k+1)\left(1 + \log\left(\frac{R(k+1)}{\delta}\right)\right).$$

Proof. By Lemma 4.2, (i) there exist $-\infty < a < b < \infty$ and $0 < C_0 < \infty$ such that

$$\|\Phi f - \Phi g\|_n \le C_0 \|f - g\|_{L_2([a,b])}$$

for $f, g \in T_k$. Assume $\mathcal{H}(\delta)$ is a δ-covering of $(T_{k,R}, \|\cdot\|_{L_2([a,b])})$ for every $\delta > 0$. Then $\mathcal{H}(\delta/C_0)$ is a δ-covering of $\mathcal{G}_K(R)$. Consequently, the claim follows from Lemma 4.4. □

Again, this implies that the space $\mathcal{G}_{k,R}(\Phi)$ equipped with the empirical norm $\|\cdot\|_n$ is compact. Consequently the functional $\|\cdot - Y\|_n$ has a minimizer in $\mathcal{G}_{k,R}(\Phi)$ for every k. As $\lambda_n J_\#(\cdot)$ is strictly increasing in the number of jumps for every $\lambda_n > 0$ this implies the following lemma.

Lemma 4.7. *For each $\lambda_n > 0$ the functional $\|\cdot - Y\|_n + \lambda_n J_\#(\cdot)$ has a minimizer in $\mathcal{G}_{\infty,R}(\Phi)$.*

4.3. Consistency

To deduce consistency of the jump estimates from the L_2 consistency of the function estimator, a result on the dependency of $\mathrm{d}(\mathcal{J}(f), \mathcal{J}(g))$ on the L_2 distance of f and g is needed. This is given by the following lemma.

Lemma 4.8. *Assume $f, g \in T_\infty$. Then*

$$\mathrm{d}(\mathcal{J}(f), \mathcal{J}(g)) \leq \frac{4\|f-g\|_2^2}{(\min\{|f(t_+) - f(t_-)| : t \in \mathcal{J}(f)\})^2}\,.$$

Proof. Let $\tau \in \mathcal{J}(f)$ and $\gamma \in \mathcal{J}(g)$, such that $|\tau - \gamma| = \mathrm{d}(\mathcal{J}(f), \mathcal{J}(g))$. Then

$$\|f-g\|_2^2 \geq |\tau - \gamma| \Big(\frac{\min\{|f(t_+) - f(t_-)| : t \in \mathcal{J}(f)\}}{2} \Big)^2,$$

which proves the assertion. □

In order to show consistency of $\hat{f}_n$, we first prove the consistency of $\Phi\hat{f}_n$. To this end we require the following result which follows directly from the proof of Theorem 4.8, page 56 in van de Geer (2000).

Lemma 4.9. *Assume $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. with mean zero and $\mathrm{E}(\varepsilon_1^2) = \sigma^2 < \infty$. Set $\mathcal{G}_n(R) = \{g \in \mathcal{G} : \|g\|_n \leq R\}$ and suppose that*

$$\frac{1}{n}\, H(\delta, \mathcal{G}_n(R), Q_n) \to 0 \quad \textit{for all} \quad \delta > 0, R > 0.$$

Then

$$\sup_{g \in \mathcal{G}_n(R)} \big|\langle \varepsilon, g \rangle_n\big| = \sup_{g \in \mathcal{G}_n(R)} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i g(x_i) \right| = o_P(1)$$

for every $R > 0$.

Now we are able to prove consistency of $\hat{f}_n$.

Lemma 4.10. *Suppose the Assumptions A, B and C are met. Then Φ^{-1} is continuous as mapping from $\{\Phi f : f \in T_{k,R}\} \subset L_2([0,1])$ to the space $(T_{k,R}, \|\cdot\|_{L_2([0,1])})$ for any $k \in \mathbb{N}, R > 0$. Moreover $\|\Phi f - \Phi\hat{f}_n\|_2 = o_P(1)$ and consequently*

$$\|f - \hat{f}_n\|_2 = o_P(1)\,. \tag{20}$$

Proof. Use (9) and $Y = \Phi f + \varepsilon$ to obtain

$$\begin{aligned}\|\Phi\hat{f}_n - \Phi f\|_n &\leq 2\langle\Phi(\hat{f}_n - f), \varepsilon_n\rangle_n + o(n^{-1}) \\ &\leq 2 \sup_{g\in\mathcal{G}_{2k,2R}(\Phi)} \big|\langle g, \varepsilon_n\rangle_n\big| + o(n^{-1}),\end{aligned}$$

since $f - \hat{f}_n \in T_{2k,2R}$. By Corollary 4.6

$$n^{-1}H(\delta, \mathcal{G}_{2k,2R}(\Phi), Q_n) \to 0 \quad \text{for all} \quad \delta > 0 .$$

Hence Lemma 4.9 gives

$$\sup_{g\in\mathcal{G}_{2k,2R}(\Phi)} |\langle g, \varepsilon_n\rangle_n| = o_P(1) .$$

This proves $\|\Phi f - \Phi\hat{f}_n\|_n = o_P(1)$. Lemma 4.2 (v), together with Lemma 4.3, yields

$$\|\Phi f - \Phi\hat{f}_n\|_2 = o_P(1) . \tag{21}$$

By Corollary 4.5 the space $(T_{2k,2R}, \|\cdot\|_{L_2([0,1])})$ is compact. Lemma 4.2, (iii) and (ii) yield that the map

$$\Phi : (\ T_{2k,2R}, \|\cdot\|_{L_2([0,1])}\ ) \to L_2([0, 1])$$

is continuous and one-to-one. The inverse of a continuous injective mapping f restricted to the image $f(\Omega)$ is continuous if Ω is compact. This gives continuity of Φ^{-1} as mapping from $\{\Phi f : f \in T_{2k,2R}\} \subset L_2([0, 1])$ to $(T_{2k,2R}, \|\cdot\|_{L_2([0,1])})$. Hence, $\|\Phi f\|_2 \to 0$ implies $\|f\|_2 = \|f\|_{L_2([0,1])} = \|\Phi^{-1}\Phi f\|_{L_2([0,1])} \to 0$ for $f \in T_{2k,2R}$. Consequently (21) implies

$$\|f - \hat{f}\|_2 = o_P(1). \qquad \square$$

This allows us to infer the consistency of the parameter estimates. The following corollary is a direct consequence of Lemma 4.8 and 4.10.

Corollary 4.11. *Suppose the prerequisites of Lemma 4.10 are met. In this case*

$$\mathrm{d}(\mathcal{J}(f), \mathcal{J}(\hat{f}_n)) = o_P(1) ,$$

as well as $\#\mathcal{J}(f) = \#\mathcal{J}(\hat{f}_n)$. Moreover, if f is given by (2) and $\hat{f}_n$ by (10), we have for the estimates $\hat{b}_i$ of the levels b_i that

$$\max_{i=1,\ldots,k+1} |\hat{b}_i - b_i| = o_p(1) .$$

Similar to Lemma 4.10 we show Lemma 3.3.

Proof of Lemma 3.3. By definition of $\hat{f}_n$ we obtain

$$\|\Phi f + \varepsilon - \Phi\hat{f}_n\|_n^2 \leq \|\Phi f + \varepsilon - \Phi t\|_n^2 + o_P(n^{-1})$$

which yields

$$\|\Phi f - \Phi \hat{f}_n\|_n^2 \leq \|\Phi f - \Phi t\|_n^2 + 2\langle \Phi \hat{f}_n - \Phi t, \varepsilon\rangle_n + o_P(n^{-1}).$$

Since $\hat{f}_n - t \in T_{2k,2R}$, application of Lemma 4.9 gives an upper bound for the empirical process $|\langle \Phi \hat{f}_n - \Phi t, \varepsilon\rangle_n| \leq |\sup_{g \in T_{2k,2R}} \langle \Phi g - \Phi t, \varepsilon\rangle_n| = o_P(1)$. So we have

$$\|\Phi f - \Phi \hat{f}_n\|_n^2 \leq \|\Phi f - \Phi t\|_n^2 + o_p(1). \tag{22}$$

For all $x, y \in [0,1]$ it holds that $|\Phi f(y) - \Phi f(x)| \leq \|f\|_\infty |\Phi 1_{[0,1]}(y) - \Phi 1_{[0,1]}(x)| \leq \|f\|_\infty C_\phi |x-y|$, where C_ϕ is chosen as in Lemma 4.2 part (v). Hence the set of functions $\Phi f - \Phi g$, s.t. $g \in T_{k,R}$ is equi-Lipschitz and bounded by $2R$ and we can apply Lemma 4.3 part (ii) (with $c_l = c_u = 1$). This together with (22) leads to

$$\|\Phi f - \Phi t\|_2^2 \leq \|\Phi f - \Phi \hat{f}_n\|_2^2 \leq \|\Phi f - \Phi t\|_2^2 + o_P(1),$$

where the first inequality follows from the minimization property of t. Consequently we find

$$\left| \|\Phi f - \Phi t\|_2^2 - \|\Phi f - \Phi \hat{f}_n\|_2^2 \right| = o_P(1). \tag{23}$$

Now assume $\Phi \hat{f}_n$ does not converge to Φt in probability. Then we can choose a subsequence $(\Phi \hat{f}_{k_n})_{n\in\mathbb{N}}$ and $c, \delta_1 > 0$, such that $P(\|\Phi t - \Phi \hat{f}_{k_n}\|_2 \geq \delta_1) > c$ for all $n \in \mathbb{N}$. Since t is the unique minimum, we have $P(\|\Phi f - \Phi \hat{f}_{k_n}\|_2 - \|\Phi f - \Phi t\|_2 > \delta_2) > c$ for some $\delta_2 > 0$. This is a contradiction to (23), which proves the first claim, i.e.

$$\|\Phi t - \Phi \hat{f}_n\|_2 = o_P(1).$$

According to the proof of Lemma 4.10, $\Phi : T_{2k,2R} \longrightarrow \Phi(T_{2k,2R})$ is a homeomorphism. Therefore, it has a continuous inverse, which yields convergence of $\hat{f}_n$ to t in probability. This proves the second claim. ☐

4.4. Asymptotic normality

To show asymptotic normality for M-estimators, it is common to assume existence of the derivative of the function which is minimized. However, as ϕ is allowed to have discontinuities, a less restrictive result is needed.

As discussed in Chapter 5.3 of van der Vaart (1998) it is sufficient to assume existence of a second order Taylor-type expansion. Following this idea, the next theorem gives the asymptotic normality of the minimizer of a process $Z_n(\theta)$, provided it allows for a certain expansion. It is similar to Theorem 5.23 of van der Vaart (1998), but also covers the case of non i.i.d. random variables, which is required for the fixed design.

Theorem 4.12. *Assume $\Theta \subset \mathbb{R}^d$ is open and $\theta_0 \in \Theta$. Let $(Z_n(\theta))_{\theta\in\Theta}$ be a stochastic process. Assume there exists a sequence of random variables $(W_n)_{n\in\mathbb{N}} \subset \mathbb{R}^d$ and a positive definite matrix $V \in \mathbb{R}^{d\times d}$ such that*

$$Z_n(\theta_0 + \Delta) = Z_n(\theta_0) - 2n^{-1/2} W_n^t \Delta + \Delta^t V \Delta + R_n(\Delta) \tag{24}$$

with

$$\sup_{\|\Delta\|\le\delta} \frac{R_n(\Delta)}{\|\Delta\|^2+n^{-1}} \xrightarrow{p} 0 \qquad as \quad n\to\infty\,,\delta\to 0\,, \tag{25}$$

as well as

$$W_n \xrightarrow{\mathcal{L}} N(0,\Gamma)\,.$$

If $\hat\theta_n$ is a consistent estimator of θ_0 and $\hat\theta_n$ is an approximate minimizer of Z_n, i.e.

$$\|\hat\theta_n-\theta_0\| = o_P(1) \qquad and \qquad Z_n(\hat\theta_n) \le \inf_{\theta\in\Theta}(Z_n(\theta)) + o_P(n^{-1})\,,$$

then

$$\sqrt{n}(\hat\theta_n-\theta_0) = V^{-1}W_n + o_P(1)\,.$$

Proof. First, we show the $\sqrt{n}$ consistency of $\hat\theta_n$. Set $\Delta_n = (\hat\theta_n-\theta_0)$. Since $\hat\theta_n$ is an approximate minimizer of Z_n,

$$\begin{aligned} Z_n(\theta_0)+o_P(n^{-1}) &\ge Z_n(\hat\theta_n) = Z_n(\theta_0+(\hat\theta_n-\theta_0)) \\ &= Z_n(\theta_0) - 2n^{-1/2}W_n^t\Delta_n + \Delta_n^t V\Delta_n + R_n(\Delta_n)\,. \end{aligned}$$

Denote by λ_V the smallest eigenvalue of V. The expansion above implies

$$o_P(n^{-1}) \ge \frac{-\|\Delta_n\|}{\sqrt{n}}\,\frac{2W_n^t\Delta_n}{\|\Delta_n\|} + \lambda_V\|\Delta_n\|^2 + R_n(\Delta_n)\,.$$

Observe that the asymptotic normality of W_n implies $\|\Delta_n\|^{-1}W_n^t\Delta_n = O_P(1)$. Now divide by $\|\Delta_n\|^2+n^{-1}$ and use condition (25) and the consistency of $\hat\theta_n$. This gives

$$o_P\big((n\|\Delta_n\|^2+1)^{-1}\big) \ge \frac{O_P(1)}{\sqrt{n}\|\Delta_n\| + (\sqrt{n}\|\Delta_n\|)^{-1}} + \frac{\lambda_V}{1+(\sqrt{n}\|\Delta_n\|)^{-2}} + o_P(1)\,.$$

Now assume $\sqrt{n}\|\Delta_n\| \xrightarrow{p} \infty$. This leads to

$$o_P(1) \ge o_P(1) + \lambda_V\,,$$

which is a contradiction since $\lambda_V > 0$. This shows

$$\sqrt{n}\|\Delta_n\| = O_P(1)\,.$$

Now we derive the convergence of $\sqrt{n}\Delta_n$ to $V^{-1}W_n$. Observe that $V^{-1}W_n = O_P(1)$. By (25)

$$nR_n(n^{-1/2}V^{-1}W_n) = o_P(1) \qquad \text{as well as} \qquad nR_n(\Delta_n) = o_P(1)\,.$$

Together with (24) and the minimizing property of $\hat\theta_n$ this leads to

$$\begin{aligned} o_P(1) &\ge n\big(Z_n(\theta_0+\Delta_n) - Z_n(\theta_0+n^{-1/2}V^{-1}W_n)\big) \\ &= 2(V^{-1}W_n - \sqrt{n}\Delta_n)^t W_n + (\sqrt{n}\Delta_n)^t V(\sqrt{n}\Delta_n) \\ &\quad - (V^{-1}W_n)^t V(V^{-1}W_n) + o_P(1) \\ &= -2(\sqrt{n}\Delta_n)^t V(V^{-1}W_n) + (\sqrt{n}\Delta_n)^t V(\sqrt{n}\Delta_n) \\ &\quad + (V^{-1}W_n)^t V(V^{-1}W_n) + o_P(1) \\ &= (\sqrt{n}\Delta_n - V^{-1}W_n)^t V(\sqrt{n}\Delta_n - V^{-1}W_n) + o_P(1)\,. \end{aligned}$$

Since V is positive definite, it follows that

$$\|\sqrt{n}\Delta_n - V^{-1}W_n\|^2 = o_P(1)\,,$$

which proves the claim. □

A second order expansion for the minimized process To derive an expansion of type (24) for the problem in (9), let us first introduce some notation. For $b, \tilde{b} \in \mathbb{R}^{k+1}$ and $\tau, \tilde{\tau} \in \Gamma_k$ set

$$g(x, b, \tau) = \sum_{j=1}^{k+1} b_j \Phi 1_{[\tau_{j-1}, \tau_j)}(x)\,.$$

and

$$Z_n(\tilde{b}, \tilde{\tau}) = \frac{1}{n}\sum_{i=1}^{n}\Big(g(x_i, b, \tau) + \varepsilon_i - g(x_i, \tilde{b}, \tilde{\tau})\Big)^2\,. \tag{26}$$

Assume that f and the estimate $\hat{f}_n$ in (9) are defined through

$$\Phi f(x) = \sum_{i=1}^{k+1} b_i \Phi 1_{[\tau_{i-1}, \tau_i)}(x) \qquad \text{and} \qquad \Phi \hat{f}_n(x) = \sum_{i=1}^{k+1} \hat{b}_i \Phi 1_{[\hat{\tau}_{i-1}, \hat{\tau}_i)}(x)\,,$$

respectively. By definition of $Z_n(\tilde{b}, \tilde{\tau})$ it is clear that

$$Z_n(\hat{b}, \hat{\tau}) \le \min_{(\tilde{b}, \tilde{\tau}) \in [-R,R]^{k+1} \times \Gamma_k} Z_n(\tilde{b}, \tilde{\tau}) + o(n^{-1})\,. \tag{27}$$

To obtain an expansion for $Z_n(\tilde{b}, \tilde{\tau})$, first examine the difference $g(x, b, \tau) - g(x, \tilde{b}, \tilde{\tau})$.

Lemma 4.13. *Suppose Assumption B is satisfied and $\nu(x)$ is given by (13). Define Δ by*

$$\Delta = \left(\tilde{b}_1 - b_1, \tilde{\tau}_1 - \tau_1, \tilde{b}_2 - b_2, \tilde{\tau}_2 - \tau_2, \ldots, \tilde{\tau}_k - \tau_k, \tilde{b}_{k+1} - b_{k+1}\right)^t\,. \tag{28}$$

Then

$$\begin{aligned} & g(x, b, \tau) - g(x, \tilde{b}, \tilde{\tau}) \\ &= \sum_{j=1}^{k+1} b_j \Phi 1_{[\tau_{j-1}, \tau_j]}(x) - \tilde{b}_j \Phi 1_{[\tilde{\tau}_{j-1}, \tilde{\tau}_j]}(x) \\ &= -\Delta^t \nu(x) + O(\|\Delta\|^2) + \sum_{i=1}^{k} O(\|\tau - \tilde{\tau}\|) 1_{[x-\tau_i, x-\tilde{\tau}_i] \cap \mathcal{J}(\phi) \neq \emptyset}\,. \end{aligned}$$

Note that $[x - \tau_i, x - \tilde{\tau}_i] \cap \mathcal{J}(\phi) \neq \emptyset$ means that ϕ has a discontinuity in the interval with endpoints $x - \tau_i$ and $x - \tilde{\tau}_i$.

Proof of Lemma 4.13. Remember $\#\mathcal{J}(\phi) < \infty$ and $\|\phi\|_\infty < \infty$.

First assume that $\tilde{\tau}_j \geq \tau_j$ and ϕ is Lipschitz continuous on $[x - \tilde{\tau}_j, x - \tau_j]$, i.e. $\mathcal{J}(\phi) \cap [x - \tilde{\tau}_j, x - \tau_j] = \emptyset$. Then for all $y \in [x - \tilde{\tau}_j, x - \tau_j]$ we have $\phi(x - y) - \phi(x - \tau_j) = O(|y - \tau_j|)$. This leads to

$$
\begin{aligned}
\Phi 1_{[\tau_{j-1},\tau_j)}(x) - \Phi 1_{[\tau_{j-1},\tilde{\tau}_j)}(x) &= -\int_{\tau_j}^{\tilde{\tau}_j} \phi(x-y)dy \\
&= -(\tilde{\tau}_j - \tau_j)\phi(x-\tau_j) - \int_{\tau_j}^{\tilde{\tau}_j} (\phi(x-y) - \phi(x-\tau_j))dy \\
&= (\tau_j - \tilde{\tau}_j)\phi(x-\tau_j) - O(1)\int_{\tau_j}^{\tilde{\tau}_j} |y-\tau_j|dy \\
&= (\tau_j - \tilde{\tau}_j)\phi(x-\tau_j) + O((\tau_j - \tilde{\tau}_j)^2)\,.
\end{aligned}
$$

If ϕ has a discontinuity in $[x - \tilde{\tau}_j, x - \tau_j]$, then

$$
\begin{aligned}
\Phi 1_{[\tau_{j-1},\tau_j)}(x) - \Phi 1_{[\tau_{j-1},\tilde{\tau}_j)}(x) &= (\tau_j - \tilde{\tau}_j)\phi(x-\tau_j) + \int_{\tau_j}^{\tilde{\tau}_j} O(\|\phi\|_\infty)dy \\
&= (\tau_j - \tilde{\tau}_j)\phi(x-\tau_j) + O(|\tau_j - \tilde{\tau}_j|)\,.
\end{aligned}
$$

The same holds for $\tilde{\tau}_j < \tau_j$. Note that $1_{[x-\tau_j, x-\tilde{\tau}_j]\cap\mathcal{J}(\phi)\neq\emptyset}$ is one if and only if ϕ has a discontinuity in $[x - \tilde{\tau}_j, x - \tau_j]$. Consequently,

$$
\begin{aligned}
\Phi 1_{[\tau_{j-1},\tau_j)}(x) - \Phi 1_{[\tau_{j-1},\tilde{\tau}_j)}(x) &= (\tau_j - \tilde{\tau}_j)\phi(x-\tau_j) + O((\tau_j - \tilde{\tau}_j)^2) \\
&\quad + O(|\tau_j - \tilde{\tau}_j|)1_{[x-\tau_j, x-\tilde{\tau}_j]\cap\mathcal{J}(\phi)\neq\emptyset}\,.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\Phi 1_{[\tau_{j-1},\tau_j)}(x) - \Phi 1_{[\tilde{\tau}_{j-1},\tau_j)}(x) = (\tilde{\tau}_{j-1} - \tau_{j-1})\phi(x-\tau_{j-1}) \\
&\qquad + O((\tau_{j-1} - \tilde{\tau}_{j-1})^2) + O(|\tau_{j-1} - \tilde{\tau}_{j-1}|)1_{[x-\tau_{j-1}, x-\tilde{\tau}_{j-1}]\cap\mathcal{J}(\phi)\neq\emptyset}\,.
\end{aligned}
$$

Remember $\tau_0 = \tilde{\tau}_0$ and $\tau_{k+1} = \tilde{\tau}_{k+1}$, combine the preceding results to obtain

$$
\begin{aligned}
&\sum_{j=1}^{k+1} \Big(b_j \Phi 1_{[\tau_{j-1},\tau_j]}(x) - \tilde{b}_j \Phi 1_{[\tilde{\tau}_{j-1},\tilde{\tau}_j]}(x) \Big) \\
&= \sum_{j=1}^{k+1} \Big((b_j - \tilde{b}_j)\Phi 1_{[\tau_{j-1},\tau_j]}(x) + \tilde{b}_j \big(\Phi 1_{[\tau_{j-1},\tau_j]}(x) - \Phi 1_{[\tau_{j-1},\tilde{\tau}_j]}(x)\big) \\
&\quad + \tilde{b}_j \big(\Phi 1_{[\tau_{j-1},\tilde{\tau}_j]}(x) - \Phi 1_{[\tilde{\tau}_{j-1},\tilde{\tau}_j]}(x)\big)\Big) \\
&= \sum_{j=1}^{k+1} \Big((b_j - \tilde{b}_j)\Phi 1_{[\tau_{j-1},\tau_j]}(x) + \tilde{b}_j(\tau_j - \tilde{\tau}_j)\phi(x-\tau_j) + O((\tau_j - \tilde{\tau}_j)^2) \\
&\quad + O(|\tau_j - \tilde{\tau}_j|)1_{[x-\tau_j, x-\tilde{\tau}_j]\cap\mathcal{J}(\phi)\neq\emptyset} + \tilde{b}_j(\tilde{\tau}_{j-1} - \tau_{j-1})\phi(x-\tau_{j-1}) \\
&\quad + O((\tau_{j-1} - \tilde{\tau}_{j-1})^2) + O(|\tau_{j-1} - \tilde{\tau}_{j-1}|)1_{[x-\tau_{j-1}, x-\tilde{\tau}_{j-1}]\cap\mathcal{J}(\phi)\neq\emptyset}\Big)\,.
\end{aligned}
$$

By $\tilde b_j(\tau_j - \tilde\tau_j) = b_j(\tau_j - \tilde\tau_j) + O(\|b-\tilde b\|\,\|\tau-\tilde\tau\|)$, this gives

$$\begin{aligned} g(x,b,\tau) &- g(x,\tilde b,\tilde\tau) \\ &= \sum_{j=1}^{k+1}(b_j-\tilde b_j)\Phi 1_{[\tau_{j-1},\tau_j]}(x) + \sum_{j=1}^{k}(\tau_j-\tilde\tau_j)(b_j-b_{j+1})\phi(x-\tau_j) \\ &\quad + O(\|\tau-\tilde\tau\|^2) + O(\|b-\tilde b\|\,\|\tau-\tilde\tau\|) + \sum_{j=1}^{k} O(\|\Delta\|)1_{[x-\tau_i,x-\tilde\tau_i]\cap\mathcal{J}(\phi)\neq\emptyset}\,. \end{aligned}$$

Since $O(\|b-\tilde b\|\,\|\tau-\tilde\tau\|) = O(\|\Delta\|^2)$ this proves the claim. $\square$

Lemma 4.14. *Suppose the Assumptions A, B and C are met. Then the process $Z_n(\tilde b,\tilde\tau)$ allows an expansion of type (24), namely*

$$Z_n(\tilde b,\tilde\tau) = Z_n(b,\tau) + 2n^{-1/2}W_n^t\Delta + \Delta^t V\Delta + R_n(\Delta)\,,$$

where R_n satisfies condition (25), Δ is given by (28) and V is the $(2k+1)\times(2k+1)$ matrix defined by (14). Moreover

$$W_n \xrightarrow{\mathcal{L}} N(0,\mathrm{E}(\varepsilon_1^2)V)\,.$$

Before we give the proof, we need the following result on the number of design points contained in a sequence of intervals.

Lemma 4.15. *If the design points $x_1,\dots,x_n$ satisfy Assumption C, then for any two sequences $a_n, b_n, n\in\mathbb{N}$ with $0\le a_n<b_n\le 1$ we have*

$$n^{-1}\big(\#\{i : x_i\in[a_n,b_n]\}\big) = O_P(|b_n-a_n| + n^{-1/2}).$$

Proof. The proof is straightforward using that $H(x)=\int_0^x h(y)dy$ is strictly monotone, and that by Assumption C it holds $H^{-1}(i/n-\delta_i) = x_{(i)}$ with $\max_{i=1,\dots,n}|\,\delta_i| = O_P(n^{-1/2})$. $\square$

Proof of Lemma 4.14. Expand (26) to obtain

$$\begin{aligned} Z_n(\tilde b,\tilde\tau) = \frac{2}{n}\sum_{i=1}^{n}\varepsilon_i\Big(g(x_i,b,\tau) - g(x_i,\tilde b,\tilde\tau)\Big) \\ + \frac{1}{n}\sum_{i=1}^{n}\Big(g(x_i,b,\tau) - g(x_i,\tilde b,\tilde\tau)\Big)^2 + \|\varepsilon\|_n^2\,. \end{aligned} \tag{29}$$

Note that the last term equals $Z_n(b,\tau)$. We will first estimate the second term of (29). Denote the points of discontinuity of ϕ by $\mathcal{J}(\phi) = \{\vartheta_1,\dots,\vartheta_{\#\mathcal{J}(\phi)}\}$ with $\vartheta_1<\vartheta_2<\cdots<\vartheta_{\#\mathcal{J}(\phi)}$. This means

$$[x-\tau_i,x-\tilde\tau_i]\cap\mathcal{J}(\phi)\neq\emptyset \quad\Leftrightarrow\quad \exists s : x\in[\vartheta_s+\tau_i,\vartheta_s+\tilde\tau_i]\,.$$

By Lemma 4.15,

$$\#\{i : x_i \in [\vartheta_s + \tau_j, \vartheta_s + \tilde{\tau}_j]\} = O_P(n|\tau_j - \tilde{\tau}_j| + n^{1/2})\,.$$

This gives

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{s=1}^{\#\mathcal{J}(\phi)} 1_{[\vartheta_s+\tau_j,\vartheta_s+\tilde{\tau}_j]}(x_i) &= \frac{\#\mathcal{J}(\phi)}{n}\sum_{j=1}^{k} O_P(n|\tau_j - \tilde{\tau}_j| + n^{1/2}) \\ &= O_P(\|\Delta\| + n^{-1/2})\,.\end{aligned}$$

The functions $\nu_j(x)$ are piecewise equi-Lipschitz continuous by part (v) of Lemma 4.2. With the help of Lemma 4.3 this gives

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(\Delta^t\nu(x_i))^2 &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j,r}^{2k+1}\Delta_j\Delta_r\nu_j(x_i)\nu_r(x_j) \\ &= \sum_{j,r}^{2k+1}\Delta_j\Delta_r\left[\int_0^1 \nu_j(x)\nu_r(x)h(x)dx + o_P(1)\right] \\ &= \Delta^t V\Delta + o_P(\|\Delta\|^2)\,.\end{aligned}$$

Use Lemma 4.13 and the results above to obtain

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}(g(x_i,b,\tau) - g(x_i,\tilde{b},\tilde{\tau}))^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\Delta^t\nu(x_i) + O(\|\Delta\|^2) + O(\|\Delta\|)\sum_{j=1}^{k}\sum_{s=1}^{\#\mathcal{J}(\phi)} 1_{[\vartheta_s+\tau_j,\vartheta_s+\tilde{\tau}_j]}(x_i)\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\Delta^t\nu(x_i) + O(\|\Delta\|^2)\right)^2 + O(\|\Delta\|^2)O_P(\|\Delta\| + n^{-1/2}) \\ &= \Delta^t V\Delta + O_P(\|\Delta\|^3) + o_P(\|\Delta\|^2)\,,\end{aligned}$$

where V is given by (14). The remainder terms clearly satisfy condition (25).

Next, examine the first term of (29). Set

$$W_n = n^{-1/2}\sum_{i=1}^{n}\varepsilon_i\nu(x_i)$$

to derive

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(g(x_i,b,\tau) - g(x_i,\tilde{b},\tilde{\tau})) \\ &= -\sum_{i=1}^{n}\frac{\varepsilon_i}{n}\left(\Delta^t\nu(x_i) + O(\|\Delta\|^2) + O(\|\Delta\|)\sum_{j=1}^{k}\sum_{s=1}^{\#\mathcal{J}(\phi)} 1_{[\vartheta_s+\tau_j,\vartheta_s+\tilde{\tau}_j]}(x_i)\right) \\ &= -\frac{\Delta^t W_n}{\sqrt{n}} + \frac{O(\|\Delta\|^2)}{n}\sum_{i=1}^{n}\varepsilon_i + \frac{O(\|\Delta\|)}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{s=1}^{\#\mathcal{J}(\phi)}\varepsilon_i 1_{[\vartheta_s+\tau_j,\vartheta_s+\tilde{\tau}_j]}(x_i)\,.\end{aligned}$$

The second term is clearly $o_P(\|\Delta\|^2)$.

To obtain an upper bound for the third term suppose $\vartheta_s - \tau_j < \vartheta_s - \tilde\tau_j$. Set

$$i_l(s,j) = \min\{i : x_{(i)} \geq \theta_s - \tau_j\} \quad \text{and} \quad i_u(s,j) = \max\{i : x_{(i)} < \theta_s - \tilde\tau_j\}\,.$$

Consequently,

$$\left|\sum_{i=1}^{n} \varepsilon_i 1_{[\vartheta_s+\tau_j,\vartheta_s+\tilde\tau_j]}(x_i)\right| = \left|\sum_{i=i_l(s,j)}^{i_u(s,j)} \varepsilon_i\right|.$$

By the law of the iterated logarithm for $\varepsilon_1, \varepsilon_2, \ldots$ i.i.d. with $\mathrm{E}(\varepsilon_1) = 0$ and $\mathrm{E}(\varepsilon_1^2) < \infty$ we have for any sequence $(k_n)_{n\in\mathbb{N}} \subset \mathbb{N}$ with $\limsup_{n\to\infty} k_n = \infty$ that

$$\lim_{n\to\infty} \max_{j\in\{1,\ldots,k_n\}} (\mathrm{E}(\varepsilon_1^2) k_n \log\log k_n)^{-1/2} \left|\sum_{i=1}^{j} \varepsilon_i\right| = 1$$

almost surely. This implies for $\delta_n = i_u(s,j) - i_l(s,j)$ that

$$\max_{j=1,\ldots,\delta_n} \left|\sum_{i=i_l(s,j)}^{i_u(s,j)} \varepsilon_i\right| = O((\delta_n \log\log \delta_n)^{1/2})$$

holds almost surely. By Lemma 4.15,

$$\delta_n = \#\{i : \vartheta_s - \tau_j \leq x_{(i)} < \vartheta_s - \tilde\tau_j\} = O_P(n|\tau_j - \tilde\tau_j| + \sqrt{n}) = O_P(n\|\Delta\| + \sqrt{n})\,.$$

Consequently,

$$\sum_{i=1}^{n} \left|\varepsilon_i 1_{[\vartheta_s+\tau_j,\vartheta_s+\tilde\tau_j]}(x_i)\right| = O_P\Big(\sqrt{(n\|\Delta\| + n^{1/2}) \log\log(n\|\Delta\| + n^{1/2})}\Big)\,.$$

The same can be shown for $\vartheta_j - \tau_j \geq \vartheta_j - \tilde\tau_j$. Since $\mathcal{J}(\phi)$ is a finite set and $k < \infty$, it follows that

$$\begin{aligned} &\frac{O(\|\Delta\|)}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} \sum_{s=1}^{\#\mathcal{J}(\phi)} \varepsilon_i 1_{[\vartheta_s+\tau_j,\vartheta_s+\tilde\tau_j]}(x_i) \\ &\qquad = O(n^{-1}\|\Delta\|) O_P\Big(\sqrt{(n\|\Delta\| + n^{1/2}) \log\log(n\|\Delta\| + n^{1/2})}\Big)\,. \end{aligned} \tag{30}$$

To verify condition (25) for this term, note that for $\|\Delta\| < n^{-1/2}$,

$$(30) = O_P\big(n^{-5/4}\sqrt{\log\log(n^{1/2})}\big) = o_P(n^{-1})\,,$$

and for $\|\Delta\| \geq n^{-1/2}$,

$$(30) = O_P\big(\|\Delta\|^{3/2} n^{-1/2}\sqrt{\log\log(n)}\big) = o_P(\|\Delta\|^2)\,.$$

This gives

$$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(g(x_i,b,\tau)-g(x_i,\tilde{b},\tilde{\tau})) = -n^{-1/2}\Delta^t W_n + o_P(\|\Delta\|^2) + o_P(n^{-1})\,.$$

Next, take a closer look at W_n. For any $a \in \mathbb{R}^{2k+1}$,

$$a^t W_n = \sum_{i=1}^{n}\varepsilon_i\left(n^{-1/2}\sum_{j=1}^{2k+1}a_j\nu_j(x_i)\right)$$

and by similar calculations as in (30)

$$\sum_{i=1}^{n}\left(n^{-1/2}\sum_{j=1}^{2k+1}a_j\nu_j(x_i)\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(a^t\nu(x_i))^2 = a^tVa + o_P(1)\,.$$

By the central limit theorem and the Cramer-Wold device,

$$W_n \xrightarrow{\mathcal{L}} N(0,\sigma^2 V)\,,$$

where $\sigma^2 = \mathrm{E}(\varepsilon_1^2)$ and V is given by (14). □

Lemma 4.16. *Given the Assumptions C and B, the matrix V defined by (14) is positive definite.*

Proof. For any $\beta \in \mathbb{R}^{2k+1}$

$$\beta^t V\beta = \int\left(\sum_{i=1}^{2k+1}\beta_i\nu_i(x)\right)^2 h(x)dx \geq c_l\int_0^1\left(\sum_{i=1}^{2k+1}\beta_i\nu_i(x)\right)^2 dx\,.$$

Observe that by Assumption B, the functions $\nu_1,\ldots,\nu_{2k+1}$ are linearly independent as functions in $L_2([0,1])$, since $b_i - b_{i+1} \neq 0$ for all $i = 1,\ldots,k$. Consequently, for $\beta \neq 0$ we have that

$$\int_0^1\left(\sum_{i=1}^{2k+1}\beta_i\nu_i(x)\right)^2 dx > 0$$

and thus $\beta^t V\beta > 0$. □

4.5. Proof of Theorem 3.1

The proof of the main theorem is now a direct consequence of the results given above. Part (i) follows directly from the proof of Lemma 4.14.

Proof of part (ii) Corollary 4.11 implies $\|\theta - \hat{\theta}_n\| = o_P(1)$. By relation (27) and Lemma 4.14 the assumptions of Theorem 4.12 are satisfied. The claim follows by application of this theorem.

Proof of part (iii) By Lemma 4.13

$$
\begin{aligned}
&\int_0^1 \left(\sum_{i=1}^{k+1} b_i \Phi 1_{[\tau_{i-1},\tau_i)}(x) - \hat{b}_i \Phi 1_{[\hat{\tau}_{i-1},\hat{\tau}_i)}(x) \right)^2 dx? \\
&= \int_0^1 \big((\theta - \hat{\theta}_n)^t \nu(x) \big)^2 dx + O_P(\|\theta - \hat{\theta}\|^2) = O_P(n^{-1}),
\end{aligned}
$$

since ν is bounded. This proves the claim.

Proof of part (v) and part (iv) Note that

$$
\begin{aligned}
\|f - \hat{f}_n\|_2^2 &= \sum_{i=1}^{k+1} (b_i - \hat{b}_i)^2 \Big(\min(\tau_i, \hat{\tau}_i) - \max(\tau_{i-1}, \hat{\tau}_{i-1}) \Big) \\
&\quad + \sum_{i=1}^{k} \Big(1_{\tau_i \geq \hat{\tau}_i} (b_i - \hat{b}_{i+1})^2 + 1_{\tau_i < \hat{\tau}_i} (b_{i+1} - \hat{b}_i)^2 \Big) \big| \tau_i - \hat{\tau}_i \big| \\
&= O_P(n^{-1}) O_P(1) + O_P(1) O_P(n^{-1/2}) = O_P(n^{-1/2}).
\end{aligned}
$$

This proves part (v). Part (iv) follows by application of Lemma 4.8. □

4.6. Proof of Theorem 3.2

In this section we analyze the case where the number of jumps is unknown.

In order to reconstruct the number of jumps correctly, it is helpful to use a penalty function which is strictly increasing in the number of jumps. Any penalty term, which depends on the number of jumps only, is not a pseudo-norm on $T_{\infty,R}$, since $\#\mathcal{J}(\lambda f) = \#\mathcal{J}(f)$ for $\lambda \neq 0$. Hence, the standard results from empirical process theory do not apply. However, it is possible to use similar techniques in the proofs.

The fact that $\hat{f}_{\lambda_n}$ (approximately) minimizes the penalized L_2 functional, implies that for any $f \in T_{\infty,R}$, we get that

$$
\|\Phi \hat{f}_{\lambda_n} - Y\|_n^2 + \lambda_n J_\#(\hat{f}_{\lambda_n}) \leq \|\Phi f - Y\|_n^2 + \lambda_n J_\#(f) + o(n^{-1}).
$$

This gives

$$
\begin{aligned}
&\|\Phi \hat{f}_{\lambda_n} - \Phi f\|_n^2 + 2\langle \Phi \hat{f}_{\lambda_n} - \Phi f, -\varepsilon \rangle_n + \|\varepsilon\|_n + \lambda_n J_\#(\hat{f}_{\lambda_n}) \\
&\qquad\qquad \leq \|\varepsilon\|_n + \lambda_n J_\#(f) + o(n^{-1}),
\end{aligned}
$$

which yields the basic inequality

$$
\|\Phi \hat{f}_{\lambda_n} - \Phi f\|_n^2 + \lambda_n J_\#(\hat{f}_{\lambda_n}) \leq 2\langle \Phi \hat{f}_{\lambda_n} - \Phi f, \varepsilon \rangle_n + \lambda_n J_\#(f) + o(n^{-1}). \tag{31}
$$

Hence, a bound for the term $|\langle \Phi \hat{f}_{\lambda_n} - \Phi f, \varepsilon \rangle_n|$, would allow immediate conclusions on $\|\Phi \hat{f}_{\lambda_n} - \Phi f\|_n^2$ as well as $\lambda_n J_\#(\hat{f}_{\lambda_n})$.

Theorem 4.17. *Suppose Assumption A is met and the error satisfies (A1). Assume* $\sup_{g\in\mathcal{G}}\|g\|_n \le R$. *There exists a constant C depending only on Assumption (A1), such that for all $\delta > 0$ satisfying*

$$\sqrt{n}\delta \ge C\Big(\int_0^R H^{1/2}(u,\mathcal{G},Q_n)du \vee R\Big) \tag{32}$$

we have that

$$\mathrm{P}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{i=1}^n \varepsilon_i g(x_i)\right| \ge \delta\right) \le C\exp\Big(-\frac{n\delta^2}{C^2R^2}\Big). \tag{33}$$

Proof. See Lemma 3.2, page 29 in van de Geer (2000). □

A bound of this type can be obtained from the following exponential inequality.

Lemma 4.18. *Suppose Assumptions A and B are met and the error additionally satisfies (A1).*

There exist constants $c_1, c_2 > 0$, such that for all $t \ge c_1 n^{-1/2}$ we have

$$\mathrm{P}\Big(\sup_{f\in T_{\infty,R}}\frac{|\langle\varepsilon,\Phi f\rangle_n|}{\|\Phi f\|_n J_{\#}^{1/2}(f)\big(1+\log(J_{\#}(f)/\|\Phi f\|_n)_+\big)} \ge t\Big) \le c_2\exp\Big(-\frac{nt^2}{c_2^2}\Big).$$

Proof. Set $\mathcal{G}_{k,R}(\Phi) = \{\Phi g : g \in T_{k,R}\}$. By Corollary 4.6 there exists a constant $C > 0$ independent of u,k,R and n such that

$$H\big(u,\mathcal{G}_{k-1,R}(\Phi),Q_n\big) \le Ck\Big(1+\log\Big(\frac{Rk}{u}\Big)\Big).$$

Compute

$$\begin{aligned}
&\int_0^\delta H^{1/2}\big(u,\mathcal{G}_{k-1,R}(\Phi),Q_n\big)du\\
&\le \sqrt{Ck}\int_0^\delta\sqrt{\log\Big(\frac{\exp(1)Rk}{u}\Big)}du\\
&= eRk\sqrt{Ck}\int_0^{\frac{\delta}{eRk}}\sqrt{-\log(u)}du \le eRk\sqrt{Ck}\int_0^{\frac{\delta}{eRk}}(-\log(u))du\\
&= eRk\sqrt{Ck}\Big(\frac{\delta}{eRk}\Big(1-\log\Big(\frac{\delta}{eRk}\Big)\Big)\Big) = \delta\sqrt{Ck}(2+\log(R)+\log(k\delta^{-1}))\\
&\le C_1\delta\sqrt{k}\Big(1+\log\Big(\frac{k}{\delta}\vee 1\Big)\Big) = C_1\delta\sqrt{k}\Big(1+\log\Big(\frac{k}{\delta}\Big)_+\Big),
\end{aligned}$$

where C_1 is some finite constant independent of k and δ. By Theorem 4.17 there exists some constant C_2 only depending on α in the subgaussian error condition (A1), such that

$$\sqrt{n}\rho \ge C_2\Big(\int_0^\delta H^{1/2}\big(u,\mathcal{G}_{k-1,R}(\Phi),Q_n\big)du \vee \delta\Big)$$

implies

$$\mathrm{P}\Big(\sup_{g\in\mathcal{G}^{(n)}_{k-1,R}(\Phi,\delta)} |\langle g,\varepsilon\rangle_n| \geq \rho\Big) \leq C_2 \exp\Big(-\frac{n\rho^2}{C_2^2\delta^2}\Big),$$

where $\mathcal{G}^{(n)}_{k-1,R}(\Phi,\delta) = \{g \in \mathcal{G}_{k-1,R}(\Phi) : \|g\|_n \leq \delta\}$. Consequently, for all $t \geq C_2C_1n^{-1/2}$ we have that

$$\mathrm{P}\Big(\sup_{g\in\mathcal{G}^{(n)}_{k-1,R}(\Phi,\delta)} |\langle g,\varepsilon\rangle_n| \geq t\delta\sqrt{k}\Big(1+\log\Big(\frac{k}{\delta}\Big)_+\Big)\Big)$$
$$\leq C_2\exp\Big(-\frac{nt^2k\big(1+\log\big(\frac{k}{\delta}\big)_+\big)^2}{C_2^2}\Big).$$

We arrive at

$$\begin{aligned}
&\mathrm{P}\Big(\sup_{g\in\mathcal{G}_{k-1,R}(\Phi)} \frac{|\langle\varepsilon,g\rangle_n|}{\|g\|_n\sqrt{k}\big(1+\log(k/\|g\|_n)_+\big)} \geq t\Big)\\
&\leq \sum_{s=1}^{\infty}\mathrm{P}\Big(\sup_{g\in\mathcal{G}_{k-1,R}(\Phi,2^{1-s}R)} |\langle\varepsilon,g\rangle_n| \geq t(2^{-s}R)\sqrt{k}\Big(1+\Big(\log\Big(\frac{k}{2^{-s}R}\Big)\Big)_+\Big)\Big)\\
&\leq \sum_{s=1}^{\infty} C_2\exp\Big(\frac{-t^2nk(1+(\log(k/R)+s\log(2))_+)}{C_2^2}\Big)\\
&\leq \sum_{s=1}^{\infty} C_2\exp\Big(\frac{-t^2n(1+(s\log(2)-\log(R))_+)}{C_2^2}\Big).
\end{aligned}$$

Splitting this sum at $s_R := \lceil (1+\log(R))/\log(2)\rceil$ gives

$$\begin{aligned}
&\mathrm{P}\Big(\sup_{g\in\mathcal{G}_{k-1,R}(\Phi)} \frac{|\langle\varepsilon,g\rangle_n|}{\|g\|_n\sqrt{k}\big(1+\log(k/\|g\|_n)_+\big)} \geq t\Big)\\
&\leq C_2\Big\lceil\frac{1+\log(R)}{\log(2)}\Big\rceil\exp\Big(\frac{-t^2n}{C_2^2}\Big) + \sum_{s=s_R}^{\infty} C_2\exp\Big(\frac{-t^2nC_3(1+s\log(2))}{C_2^2}\Big)\\
&\leq C_5\exp\Big(\frac{-t^2n}{C_2^2}\Big) + \sum_{s=1}^{\infty} C_2\exp\Big(\frac{-t^2nC_4(1+s)}{C_2^2}\Big)\\
&\leq C_5\exp\Big(\frac{-t^2n}{C_2^2}\Big) + \exp\Big(\frac{-t^2nC_4}{C_2^2}\Big)\int_{s=0}^{\infty} C_2\exp\Big(\frac{-t^2nC_4s}{C_2^2}\Big)\\
&\leq C_5\exp\Big(\frac{-t^2n}{C_2^2}\Big) + \frac{C_2^3}{C_4t^2n}\exp\Big(\frac{-t^2nC_4}{C_2^2}\Big) \leq C_6\exp\Big(-\frac{t^2n}{C_4^2}\Big).
\end{aligned}$$

Here C_3, C_4, C_5, C_6 are constants depending on C_1, C_2 and R only. The last inequality holds by $t^2n \geq C_1^2C_2^2$.

Since the constant C_6 does not depend on k, the exponential inequality also holds if we additionally take the supremum over all k. This proves the claim. □

The above lemma yields upper bounds for the rate of $|\langle \Phi f, \varepsilon\rangle_n|$, which are stated in the subsequent corollary.

Corollary 4.19. *Suppose the prerequisites of Lemma 4.18 are met. Then*

$$\sup_{f\in T_{\infty,R}} \frac{|\langle \Phi f, \varepsilon\rangle_n|}{\|\Phi f\|_n \sqrt{J_\#(f)}\big(1+\log(J_\#(f)/\|\Phi f\|_n)_+\big)} = O_P(n^{-1/2})\,.$$

Moreover, for each $\epsilon > 0$ we have

$$\sup_{f\in T_{\infty,R}} \frac{|\langle \Phi f, \varepsilon\rangle_n|}{\|\Phi f\|_n^{1-\epsilon}(J_\#(f))^{(1+2\epsilon)/2}} = O_P(n^{-1/2})\,.$$

Proof. The first equation follows directly from Lemma 4.18. To show the second equation, observe that $J_\#(f) \geq 1$ and that $\sqrt{x}(1+\log(x)) \leq cx^{1/2+\epsilon}$ for $x \geq 1$, $\epsilon > 0$ and $c \geq (\epsilon^{-1} \vee 1)$ Moreover, if c is large enough and $x \geq 0$ then $x(1+\log(x^{-1})) \leq cx^{1-\epsilon}$. Combine these observations to derive the second equation from the first. □

Now we are in the position to prove that with probability one the penalized estimator $\hat f_{\lambda_n}$ correctly estimates the number of jumps as n tends to infinity (given a proper choice of the penalty term).

Proof of Theorem 3.2. Application of Corollary 4.19 to (31) gives

$$\begin{aligned}\|\Phi \hat f_{\lambda_n} - \Phi f\|_n^2 \leq \|\Phi \hat f_{\lambda_n} - \Phi f\|_n^{1-\epsilon} J_\#(\hat f_{\lambda_n} - f)^{1/2+\epsilon} O_P(n^{-1/2})\\ + \lambda_n(J_\#(f) - J_\#(\hat f_{\lambda_n})) + o(n^{-1})\,,\end{aligned} \tag{34}$$

where ϵ is given by the condition $\lambda_n n^{1/(1+\epsilon)} \to \infty$.

First, assume $J_\#(\hat f_{\lambda_n}) \leq J_\#(f)$. Then $J_\#(\hat f_{\lambda_n} - f)$ is bounded and (34) implies that either

$$\|\Phi \hat f_{\lambda_n} - \Phi f\|_n^2 = O(\lambda_n) + o(n^{-1}) \qquad \text{or} \qquad \|\Phi \hat f_{\lambda_n} - \Phi f\|_n^{1+\epsilon} = O_p(n^{-1/2})\,.$$

Thus, $\|\Phi \hat f_{\lambda_n} - \Phi f\|_n = o_P(1)$. Recall, that $J_\#(\hat f_{\lambda_n}) \leq J_\#(f)$, which allowes for application of Lemma 4.3, to deduce $\|\Phi \hat f_{\lambda_n} - \Phi f\|_2 = o_P(1)$. With the help of Lemma 4.10 and Lemma 4.8, it follows $\mathrm{d}(\mathcal{J}(\hat f_{\lambda_n}), \mathcal{J}(f)) = o_P(1)$, which in turn implies $J_\#(\hat f_{\lambda_n}) \geq J_\#(f)$ eventually.

Now assume $J_\#(\hat f_{\lambda_n}) \geq J_\#(f)$. Then (34) yields

$$\|\Phi \hat f_{\lambda_n} - \Phi f\|_n^2 \leq \|\Phi \hat f_{\lambda_n} - \Phi f\|_n^{1-\epsilon} J_\#(\hat f_{\lambda_n} - f)^{1/2+\epsilon} O_P(n^{-1/2}) + o(n^{-1})\,.$$

Assume n_k is a subsequence such that $\|\Phi \hat f_{\lambda_{n_k}} - \Phi f\|_{n_k}^{1-\epsilon} \geq c n_k^{-1/2}$ for some $c > 0$. Dividing the last equation by $\|\Phi \hat f_{\lambda_{n_k}} - \Phi f\|_{n_k}^{1-\epsilon}$ gives

$$\begin{aligned}\|\Phi \hat f_{\lambda_{n_k}} - \Phi f\|_{n_k}^{1+\epsilon} &\leq J_\#(\hat f_{\lambda_{n_k}} - f)^{1/2+\epsilon} O_P(n_k^{-1/2}) + o(n_k^{-1/2})\\ &= J_\#(\hat f_{\lambda_{n_k}} - f)^{1/2+\epsilon} O_P(n_k^{-1/2})\,.\end{aligned}$$

This yields

$$\|\Phi\hat f_{\lambda_{n_k}} - \Phi f\|_{n_k}^{1-\epsilon} \le J_\#(\hat f_{\lambda_{n_k}} - f)^{(1+\epsilon-2\epsilon^2)/(2+2\epsilon)} O_P(n_k^{-(1-\epsilon)/(2+2\epsilon)})\,.$$

Moreover, by (34)

$$\begin{aligned}\lambda_{n_k}&(J_\#(\hat f_{\lambda_{n_k}}) - J_\#(f)) \\ &\le O_P(n_k^{-1/2})\|\Phi\hat f_{\lambda_{n_k}} - \Phi f\|_{n_k}^{1-\epsilon} J_\#(\hat f_{\lambda_{n_k}} - f)^{1/2+\epsilon} + o(n_k^{-1})\,.\end{aligned}$$

Combine the last two equations to obtain

$$\lambda_{n_k}(J_\#(\hat f_{\lambda_{n_k}}) - J_\#(f)) \le O_P(n_k^{-1/(1+\epsilon)}) J_\#(\hat f_{\lambda_{n_k}} - f)^{(1+\epsilon-\epsilon^2)/(1+\epsilon)}\,. \tag{35}$$

Now assume n_k is a subsequence such that $\|\Phi\hat f_{\lambda_{n_k}} - \Phi f\|_{n_k}^{1-\epsilon} < c n_k^{-1/2}$ for some $c > 0$. Application of Corollary 4.19 to (31) and the observation that $J_\#(g) \ge 1$ for all g gives

$$\begin{aligned}&\lambda_{n_k}(J_\#(\hat f_{\lambda_{n_k}}) - J_\#(f)) \\ &\le O_P(n_k^{-1/2})\|\Phi\hat f_{\lambda_{n_k}} - \Phi f\|_{n_k}^{1-\epsilon} J_\#(\hat f_{\lambda_{n_k}} - f)^{1/2+\epsilon} + o(n_k^{-1}) \\ &\le O_P(n_k^{-1}) J_\#(\hat f_{\lambda_{n_k}} - f)^{1/2+\epsilon} \le O_P(n_k^{-1/(1+\epsilon)}) J_\#(\hat f_{\lambda_{n_k}} - f)^{(1+\epsilon-\epsilon^2)/(1+\epsilon)}\,.\end{aligned}$$

As each sequence can be decomposed into a subsequence containing only elements smaller than $cn^{-1/2}$ and a subsequence containing only elements greater or equal to $cn^{-1/2}$ for some $c > 0$, we have shown that $J_\#(\hat f_{\lambda_n}) \ge J_\#(f)$ implies (35).

Now we show that $J_\#(\hat f_{\lambda_{n_k}}) - J_\#(f) \to 0$ in probability. To this end, assume there exists some subsequence n_k such that

$$J_\#(\hat f_{\lambda_{n_k}}) - J_\#(f) \ge c > 0\,. \tag{36}$$

This implies $J_\#(f) \le J_\#(f) c^{-1}(J_\#(\hat f_{\lambda_{n_k}}) - J_\#(f))$ and

$$\begin{aligned}J_\#(\hat f_{\lambda_{n_k}} - f) &\le 2(J_\#(\hat f_{\lambda_{n_k}}) - J_\#(f)) + 2J_\#(f) \\ &\le (2 + 2J_\#(f)c^{-1})(J_\#(\hat f_{\lambda_{n_k}}) - J_\#(f)) \\ &= O(1)(J_\#(\hat f_{\lambda_{n_k}}) - J_\#(f))\,.\end{aligned}$$

Hence

$$\begin{aligned}O_P(n_k^{-1/(1+\epsilon)})&J_\#(\hat f_{\lambda_{n_k}} - f)^{(1+\epsilon-\epsilon^2)/(1+\epsilon)} \\ &= O_P(n_k^{-1/(1+\epsilon)})\big(J_\#(\hat f_{\lambda_{n_k}}) - J_\#(f)\big)^{(1+\epsilon-\epsilon^2)/(1+\epsilon)}\,.\end{aligned}$$

Together with (35), the assumption $\lambda_{n_k} n_k^{1/(1+\epsilon)} \to \infty$ and (36), this gives

$$0 < c^{\epsilon^2/(1+\epsilon)} \le \big(J_\#(\hat f_{\lambda_{n_k}}) - J_\#(f)\big)^{\epsilon^2/(1+\epsilon)} = O_P(\lambda_{n_k}^{-1} n_k^{-1/(1+\epsilon)}) = o_P(1)\,,$$

which is a contradiction and implies $J_\#(\hat{f}_n) - J_\#(f) \to 0$ in probability. Since $J_\#(f)$ and $J_\#(\hat{f}_n)$ are integers, this yields

$$\mathrm{P}\left(J_\#(\hat{f}_n) = J_\#(f)\right) \to 1\,,$$

for $n \to \infty$. This proves the claim. □

5. Proof of Theorem 2.2

Proof of part (i) To give the proof of Theorem 2.2, part (i) we will define the native Hilbert space $\mathcal{N}_\phi$ of a positive definite function ϕ and show that the elements of its dual space $\delta_x(f) = f(x)$ and $\rho_{x,y}(f) = \int_x^y f(t)dt$ are linearly independent, if ϕ has certain properties. Then we will deduce that the functions $\Delta_\phi(\cdot, \tau_0, \tau_1), \ldots, \Delta_\phi(x, \tau_k, \tau_{k+1})$ are linearly independent.

The assumptions $\widehat{\phi}(x) \geq 0$ and (8) imply that the Fourier transform $\widehat{\phi}$ is strictly positive. This means that ϕ is positive definite. (For a definition and characterization of real-valued positive definite functions, compare Chapter 6 in Wendland (2005).)

For a positive definite function ϕ and $\Omega \subset \mathbb{R}$ let $\mathcal{N}_\phi(\Omega)$ denote the unique Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ of functions $f : \Omega \to \mathbb{R}$ satisfying $f(x) = \langle f, \phi(x - \cdot)\rangle_{\mathcal{H}}$. $\mathcal{N}_\phi(\Omega)$ is called native space for ϕ and given by the closure of the span of the function set $\{\phi(x - \cdot) : x \in \Omega\}$ under the inner product induced by $\langle \phi(x - \cdot), \phi(y - \cdot)\rangle = \phi(x - y)$. A short introduction to native spaces along with some basic results of the theory can be found in Schaback (1999).

Denote by

$$\mathcal{S}(\mathbb{R}) = \left\{ f \in C^\infty(\mathbb{R}, \mathbb{C}) : \lim_{|x|\to\infty} |x^n f^{(m)}(x)| = 0 \text{ for all } n, m = 0, 1, 2, \ldots \right\}$$

the Schwartz space, where $C^\infty(\mathbb{R}, \mathbb{C})$ is the set of smooth functions from $\mathbb{R}$ to $\mathbb{C}$. The first result is, that the native space $\mathcal{N}_\phi(\Omega)$ contains all Schwartz functions which are compactly supported in Ω.

Lemma 5.1. *Assume $\Omega \subset \mathbb{R}$ and ϕ satisfies the conditions given by Theorem 2.2, part (i). Then all real Schwartz functions with support contained in Ω are elements of the native space $N_\phi(\Omega)$, this means that*

$$\Big\{ f \in \mathcal{S}(\mathbb{R}) : \mathrm{supp}(f) \subset \Omega \Big\} \subset \mathcal{N}_\phi(\Omega)\,.$$

Proof. We first proof the claim for $\Omega = \mathbb{R}$. Assume $f \in \mathcal{S}(\mathbb{R})$. Since Fourier transformation is a bijection from $\mathcal{S}(\mathbb{R})$ to $\mathcal{S}(\mathbb{R})$ $\widehat{f}$ and $\widehat{f}^{\,2}$ are also Schwartz functions. Hence for any $n_0 \in \mathbb{N}$, we can find a constant $c_1 > 0$ such that $|\widehat{f}(x)|^2 \leq c_1(1 + |x|^{n_0+2})^{-1}$. By (8) there exist $c_2 > 0$ and $n_0 \in \mathbb{N}$ such that $(\widehat{\phi}(x))^{-1} \leq c_2(1 + |x|^{n_0})$. We arrive at

$$\int_{\mathbb{R}} \frac{|\widehat{f}(x)|^2}{\widehat{\phi}(x)} dx \leq c_1 c_2 \int_{\mathbb{R}} \frac{1 + |x|^{n_0}}{1 + |x|^{n_0+2}} dx < \infty\,.$$

By Theorem 10.12 of Wendland (2005) the function f is in $\mathcal{N}_\phi(\mathbb{R})$ if and only if

$$\int_{\mathbb{R}} |\widehat{f}(x)|^2/\widehat{\phi}(x)dx < \infty\,.$$

This proves the claim for $\Omega = \mathbb{R}$.

Now assume $\Omega \subset \mathbb{R}$ is arbitrary and $f \in \mathcal{S}(\mathbb{R})$ with $\operatorname{supp} f \subset \Omega$. We have shown $f \in \mathcal{N}_\phi(\mathbb{R})$. By Theorem 10.47 in Wendland (2005) for $\Omega \subset \mathbb{R}$, $f \in \mathcal{N}_\phi(\mathbb{R})$ implies $f|_\Omega \in \mathcal{N}_\phi(\Omega)$. This proves the claim. □

Note that Lemma 5.1 implies that for any interval $(a,b) \subset \Omega$ there exists some test function $\psi \in \mathcal{N}_\phi(\Omega)$ satisfying $\operatorname{supp}(\psi) = [a,b]$. One example is

$$\psi(x) = 1_{(a,b)}(x)\,\exp((x-a)^{-1} + (b-x)^{-1})\,.$$

This observation can be used to show that point evaluation and integral mean are linearly independent as elements of the dual space of $\mathcal{N}_\phi(\Omega)$.

Definition 5.2. For $\gamma \in \mathbb{R}$ and $\gamma_1, \gamma_2 \in \mathbb{R} \cup \{-\infty, \infty\}$ with $\gamma_1 \le \gamma_2$ define the point evaluation functional $\delta_\gamma(f) = f(\gamma)$ and the functional $\rho_{\gamma_1,\gamma_2} : \mathcal{N}_\phi(\Omega) \to \mathbb{R}$ by

$$\rho_{\gamma_1,\gamma_2}(f) := \begin{cases} \displaystyle\int_{\gamma_1}^{\gamma_2} f(x)dx & \gamma_1 \ne \gamma_2\,, \\ f(\gamma_1) & \gamma_1 = \gamma_2\,. \end{cases}$$

Lemma 5.3. *Suppose ϕ satisfies the conditions given by Theorem 2.2, part (i). Assume $\tau_0 < \cdots < \tau_{k+1}$, $\gamma_1 < \cdots < \gamma_r$ and there exist an $\epsilon > 0$ such that $(\tau_1 - \epsilon, \tau_k + \epsilon) \subset \Omega$ as well as $(\gamma_1 - \epsilon, \gamma_r + \epsilon) \subset \Omega$. Then the functionals $\rho_{\tau_0,\tau_1}, \rho_{\tau_1,\tau_2}, \ldots, \rho_{\tau_k,\tau_{k+1}}, \delta_{\gamma_1}, \ldots, \delta_{\gamma_r}$ are linearly independent as elements of the dual space $\mathcal{N}_\phi(\Omega)'$.*

Proof. Assume

$$\sum_{i=1}^{k+1} \alpha_i \rho_{\tau_{i-1},\tau_i}(f) + \sum_{j=1}^{r} \beta_j \delta_{\gamma_j}(f) = 0$$

for all $f \in \mathcal{N}_\phi(\Omega)$. For each $i = 1, \ldots, k+1$ we can find an interval $J_i \subset [\tau_{i-1}, \tau_i] \cap \Omega$ such that $J_i \cap \gamma_j = \emptyset$ for all $j = 1, \ldots, r$. By Lemma 5.1 we can find a test function $f_i \in \mathcal{N}_\phi(\Omega)$ with $\operatorname{supp}(f_i) \subset J_i$ and $\int_{\mathbb{R}} f_i(x)dx = 1$ for all $i = 1, \ldots, k+1$. We then have that $\rho_{\tau_{l-1},\tau_l}(f_i) = 1_{i=l}$ and $\delta_{\gamma_j}(f_i) = 0$ for all $i = 1, \ldots, k+1$ and $j = 1, \ldots, r$. This leads to

$$0 = \sum_{l=1}^{k+1} \alpha_l \rho_{\tau_{l-1},\tau_l}(f_i) + \sum_{j=1}^{r} \beta_j \delta_{\gamma_j}(f_i) = \alpha_i$$

for all $i = 1, \ldots, k+1$. Similarly we can find test functions $f_j \in \mathcal{N}_\phi(\Omega)$ with $\delta_{\gamma_j}(f_i) = 1_{i=j}$ and deduce that $\beta_j = 0$ for all $j = 1, \ldots, r$. This proves the claim. □

Finally, we can prove Theorem 2.2, part (i).

Proof of Theorem 2.2, part (i). Assume

$$\Big\| \sum_{i=1}^{k+1} \alpha_i \Delta_\phi(\cdot, \tau_{i-1}, \tau_i) + \sum_{j=1}^{k} \beta_j \Delta_\phi(\cdot, \tau_j, \tau_j) \Big\|_2 = 0\,. \tag{37}$$

By continuity of ϕ, $\sum_{i=1}^{k+1} \alpha_i \Delta_\phi(x, \tau_{i-1}, \tau_i)$ is a continuous functions of x. Consequently, (37) implies

$$0 = \sum_{i=1}^{k+1} \alpha_i \Delta_\phi(x, \tau_{i-1}, \tau_i) + \sum_{j=1}^{k} \beta_j \Delta_\phi(x, \tau_j, \tau_j)\,,$$

for all $x \in [0,1]$. By definition of Δ_ϕ (see (5)), that means

$$0 = \sum_{i=1}^{k+1} \alpha_i \rho_{\tau_{i-1},\tau_i}\big(\phi(x-\cdot)\big) + \sum_{j=1}^{k} \beta_j \rho_{\tau_j,\tau_j}\big(\phi(x-\cdot)\big)\,,$$

for all $x \in [0,1]$. Set $\Omega = [0,1]$. By Theorem 8 in Schaback (1999), the native space $\mathcal{N}_\phi(\Omega)$ is the closure of the span of the set of functions $\{\phi(x-\cdot) : x \in \Omega\}$. It follows that

$$0 = \sum_{i=1}^{k+1} \alpha_i \rho_{\tau_{i-1},\tau_i}(f) + \sum_{j=1}^{k} \beta_j \rho_{\tau_j,\tau_j}(f)$$

for all $f \in \mathcal{N}_\phi(\Omega)$. By Lemma 5.3 we know that $\rho_{\tau_0,\tau_1}, \rho_{\tau_1,\tau_1}, \dots, \rho_{\tau_k,\tau_k}, \rho_{\tau_k,\tau_{k+1}}$ are linearly independent as elements of the dual space $\mathcal{N}_\phi(\Omega)'$. Consequently, $\alpha_i = 0 = \beta_j$ for all $i = 1, \dots, k+1, j = 1, \dots, k$, which proves the claim. □

Proof of part (ii) Again we have to show linear independence of the functions

$$\Big\{\Delta_\phi(x, \tau_i, \tau_{i+1}) : i = 0, \dots, k\Big\} \cup \Big\{\Delta_\phi(x, \tau_j, \tau_j) : j = 1, \dots, k\Big\}$$

in $L_2([0,1])$. Assume

$$\Big\| \sum_{i=0}^{k} \alpha_i \Delta_\phi(\cdot, \tau_i, \tau_{i+1}) + \sum_{j=1}^{k} \beta_j \Delta_\phi(\cdot, \tau_j, \tau_j) \Big\|_{L_2([0,1])} = 0\,.$$

Denote by $\phi_0(x) = \int_{-\infty}^{x} \phi(y)dy$ the primitive of ϕ. Since the functions $\Delta_\phi(\cdot, \tau_i, \tau_{i+1})$ and $\Delta_\phi(\cdot, \tau_i, \tau_i)$ are continuous we have for all $x \in [0,1]$ that

$$\begin{aligned} 0 &= \sum_{i=0}^{k} \alpha_i \Delta_\phi(x, \tau_i, \tau_{i+1}) + \sum_{j=1}^{k} \beta_j \Delta_\phi(x, \tau_j, \tau_j) \\ &= \sum_{i=0}^{k} \alpha_i \big(\phi_0(x-\tau_i) - \phi_0(x-\tau_{i+1})\big) + \sum_{j=1}^{k} \beta_j \phi(x-\tau_j) \end{aligned}$$

$$
\begin{aligned}
&= \alpha_0\phi_0(x-\tau_0) + \sum_{i=1}^{k}(\alpha_i-\alpha_{i-1})\phi_0(x-\tau_i) - \alpha_k\phi_0(x-\tau_{k+1}) \\
&\quad + \sum_{j=1}^{k}\beta_j\phi(x-\tau_j)\,.
\end{aligned}
$$

Consequently, this must also be true for the derivative, and the equation still holds if we replace ϕ_0 by ϕ and ϕ by ϕ'. Since the equality holds for all $x \in [0,1]$, it holds for a choice of $2k+2$ distinct points $x_0,\dots,x_{2k+1} \in [0,1]$. By the extended sign regularity of ϕ we know that the vectors

$$
\begin{pmatrix} \phi(x_0-\tau_i) \\ \vdots \\ \phi(x_{2k+1}-\tau_i) \end{pmatrix}_{i=0,\dots,k+1} \qquad \text{and} \qquad \begin{pmatrix} \phi'(x_0-\tau_j) \\ \vdots \\ \phi'(x_{2k+1}-\tau_j) \end{pmatrix}_{j=1,\dots,k}
$$

are linearly independent for $\tau_0 = 0 < \tau_i,\dots,\tau_k < 1 = \tau_{k+1}$. Hence, we immediately get that $\beta_j = 0$ for all $j = 1,\dots,k$. Moreover, $\alpha_0 = \alpha_{k+1} = 0$ and $(\alpha_{i-1}-\alpha_i) = 0$ for all $i = 1,\dots,k$. This leads to $\alpha_i = 0$ for all $i = 0,\dots,k+1$. □

Proof of part (iii) Assume

$$
\Big\|\sum_{i=1}^{k+1}\alpha_i\Delta_\phi(\cdot,\tau_{i-1},\tau_i) + \sum_{j=1}^{k}\beta_j\Delta_\phi(\cdot,\tau_j,\tau_j)\Big\|_{L_2([0,1])} = 0\,. \tag{38}
$$

Compute that

$$
\begin{aligned}
(\Phi 1_{[a,b)})(x) &= (p+1)\int_a^b (x-y)_+^p\,dy = (p+1)\int_{x-b}^{x-a} y_+^p\,dy \\
&= (x-a)_+^{p+1} - (x-b)_+^{p+1}\,.
\end{aligned}
$$

So $\sum_{i=1}^{k+1}\alpha_i\Delta_\phi(\cdot,\tau_{i-1},\tau_i)$ is a polynomial of degree $p+1$, whereas $\sum_{j=1}^{k}\beta_j\Delta_\phi(\cdot,\tau_j,\tau_j)$ has degree p. Since polynomials of different degrees are linearly independent (compare Achieser, 1992), this means Equation (38) holds if and only if

$$
\sum_{i=1}^{k+1}\alpha_i\Delta_\phi(\cdot,\tau_{i-1},\tau_i) = 0 = \sum_{j=1}^{k}\beta_j\Delta_\phi(\cdot,\tau_j,\tau_j).
$$

We show that $\alpha_i = 0$, for all $i = 0,\dots,k+1$ by induction. Therefore consider

$$
0 = \sum_{i=1}^{k+1}\alpha_i\Delta_\phi(x,\tau_{i-1},\tau_i)\Big|_{[0,\tau_1]} = \alpha_1(x-0)^{p+1}\Big|_{[0,\tau_1]}\,.
$$

This yields $\alpha_1 = 0$. Now assume $\alpha_i = 0$ for all $i \le j$. Then,

$$
\sum_{i=1}^{k+1}\alpha_i\Delta_\phi(x,\tau_{i-1},\tau_i)\Big|_{[\tau_j,\tau_{j+1}]} = \alpha_{j+1}(x-\tau_j)^{p+1}\,.
$$

this gives $\alpha_{j+1} = 0$.

In the same way it follows that $\beta_j = 0$ for all $j = 1,\dots,k$. □

Acknowledgments The authors wish to thank L. Dümbgen, T. Hohage, R. Schaback, U. Schneider and A. Tsybakov for interesting comments and bibliographic information. Further, the authors are indebted to an Associate Editor and two referees for various helpful remarks and suggestions which led to an improved version of this paper. L. Boysen gratefully acknowledges support by Georg Lichtenberg program "Applied Statistics & Empirical Methods", L. Boysen and S. Bruns DFG 1023 "Identification in Mathematical Models"; A. Munk was supported by DFG grant "Statistical Inverse Problems under Qualitative Shape Constraints" and DFG grant FOR916.

References

Achieser, N. I. (1992). *Theory of approximation*. Dover Publications Inc., New York. Translated from the Russian and with a preface by Charles J. Hyman, Reprint of the 1956 English translation. MR1217081

Birgé, L. and Massart, P. (2007). Minimal penalties for gaussian model selection. *Probab. Theory Relat. Fields* **138** 33–73. MR2288064

Bissantz, N., Dümbgen, L., Holzmann, H. and Munk, A. (2007). Nonparametric confidence bands in deconvolution density estimation. *J. Royal Statist. Society Ser. B.* **69** 483–506. MR2323764

Boysen, L. (2006). *Jump estimation for noisy blurred step function*. Ph.D. thesis, Georg-August-Universiät Göttingen.
URL `http://webdoc.sub.gwdg.de/diss/2006/boysen/boysen.pdf`

Braess, D. (1986). *Nonlinear approximation theory*, vol. 7 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin. MR0866667

Butucea, C. and Tsybakov, A. (2008a). Sharp optimality for density deconvolution with dominating bias. i. *Theory Prob. and Its Appl.* **52** 111–128. MR2354572

Butucea, C. and Tsybakov, A. (2008b). Sharp optimality for density deconvolution with dominating bias. ii. *Theory Prob. and Its Appl.* **52** 237–249. MR2354572

Carlstein, E. and Müller, H.-G. (eds.) (1994). *Change-point problems*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 23, Institute of Mathematical Statistics, Hayward, CA. Papers from the AMS-IMS-SIAM Summer Research Conference held at Mt. Holyoke College, South Hadley, MA, July 11–16, 1992. MR1477909

Cavalier, L. and Tsybakov, A. (2002). Sharp adaption for inverse problems with random noise. *Prob. Theory Rel. Fields* **123** 323–354. MR1918537

Dümbgen, L. and Johns, R. B. (2004). Confidence bands for isotonic median curves using sign tests. *J. Comput. Graph. Statist.* **13** 519–533. MR2063998

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journ. Amer. Statist. Assoc.* **99** 710–723. MR2090905

Feder, P. I. (1975). On asymptotic distribution theory in segmented regression problems–identified case. *Ann. Statist.* **3** 49–83. MR0378267

Friedrich, F., Kempe, A., Liebscher, V. and Winkler, G. (2008). Complexity penalized M-estimation: fast computation. *J. Comput. Graph. Statist.* **17** 201–224. MR2424802

Goldenshluger, A., Juditsky, A., Tsybakov, A. and Zeevi, A. (2008a). Change-point estimation from indirect observations. 1. minimax complexity. *Ann. Inst. H. Poincare Probab. Statist.* **44** 787–818. MR2453845

Goldenshluger, A., Juditsky, A., Tsybakov, A. and Zeevi, A. (2008b). Change-point estimation from indirect observations. 2. adaptiation. *Ann. Inst. H. Poincare Probab. Statist.* **44** 819–836. MR2453846

Goldenshluger, A., Tsybakov, A. and Zeevi, A. (2006). Optimal change-point estimation from indirect observations. *Ann. Statist.* **34** 350–372. MR2275245

Hinkley, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika* **56** 495–504. MR0254946

Karlin, S. (1968). *Total positivity. Vol. I.* Stanford University Press, Stanford, Calif. MR0230102

Karlin, S. and Studden, W. J. (1966). *Tchebycheff systems: With applications in analysis and statistics.* Pure and Applied Mathematics, Vol. XV, Interscience Publishers John Wiley & Sons, New York-London-Sydney. MR0204922

Korostelev, A. and Tsybakov, A. (1993). *Minimax theory of image reconstruction.* Lecture Notes in Statistics (Springer). 82. New York: Springer-Verlag. xi, 258 p. . MR1226450

Korostelëv, A. P. (1987). Minimax estimation of a discontinuous signal. *Teor. Veroyatnost. i Primenen.* **32** 796–799. MR0927265

Koul, H. L., Qian, L. and Surgailis, D. (2003). Asymptotics of M-estimators in two-phase linear regression models. *Stochastic Process. Appl.* **103** 123–154. MR1947962

Leeb, H. and Pötscher, B. M. (2006). Can one estimate the conditional distribution of post-model-selection estimators? *Ann. Stat.* **34** 2554–2591. MR2291510

Leeb, H. and Pötscher, B. M. (2008). Sparse estimators and the oracle property, or the return of the Hodges' estimator. *Journ. Econometrics* **142** 201–211. MR2394290

Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* **20** 737–761. MR1165590

Müller, H.-G. and Stadtmüller, U. (1999). Discontinuous versus smooth regression. *Ann. Statist.* **27** 299–337. MR1701113

Neumann, M. H. (1997). Optimal change-point estimation in inverse problems. *Scand. J. Statist.* **24** 503–521. MR1615339

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *J. Amer. Statist. Assoc.* **53** 873–880. MR0100314

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *J. Amer. Statist. Assoc.* **55** 324–330. MR0114269

Römer, W., Lam, Y. H., Fischer, D., Watts, A., Fischer, W. B., Göring, P., Wehrspohn, R. B., Gösele, U. and Steinem, C. (2004). Channel activity of a viral transmembrane peptide in micro-blms. *J. Am. Chem. Soc.* **49** 16267–16274.

Roths, T., Maier, D., Friedrich, C., Marth, M. and Honerkamp, J. (2000). Determination of the relaxation time spectrum from dynamic moduli using an edge preserving regularization method. *Rheol. Acta* **39** 163–173.

Sacks, J. and Ylvisaker, D. (1970). Designs for regression problems with correlated errors. III. *Ann. Math. Statist.* **41** 2057–2074. MR0270530

Schaback, R. (1999). Native Hilbert spaces for radial basis functions. I. In *New developments in approximation theory (Dortmund, 1998)*, vol. 132 of *Internat. Ser. Numer. Math.* Birkhäuser, Basel, 255–282. MR1724923

Schmitt, E. K., Vrouenraets, M. and Steinem, C. (2006). Channel activity of ompf monitored in nano-blms. *Biophys. J.* **91** 2163–2171.

Sprent, P. (1961). Some hypotheses concerning two phase regression lines. *Biometrics* **17** 634–645.

van de Geer, S. A. (1988). *Regression analysis and empirical processes*, vol. 45 of *CWI Tract*. Stichting Mathematisch Centrum Centrum voor Wiskunde en Informatica, Amsterdam. MR0937256

van de Geer, S. A. (2000). *Applications of empirical process theory*, vol. 6 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. MR1739079

van der Vaart, A. W. (1998). *Asymptotic statistics*, vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. MR1652247

Wendland, H. (2005). *Scattered data approximation*, vol. 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge. MR2131724

Yakir, B., Krieger, A. M. and Pollak, M. (1999). Detecting a change in regression: first-order optimality. *Ann. Statist.* **27** 1896–1913. MR1765621

Yao, Y.-C. and Au, S. T. (1989). Least-squares estimation of a step function. *Sankhyā Ser. A* **51** 370–381. MR1175613