Bernoulli **19**(5A), 2013, 1576–1611
DOI: 10.3150/12-BEJ422

Adaptive circular deconvolution by model selection under unknown error distribution

JAN JOHANNES* and MAIK SCHWARZ**

*Institut de statistique, biostatistique et sciences actuarielles, Voie du Roman Pays 20, Boîte L1.04.01, 1348 Louvain-la-Neuve, Belgium. E-mail: *jan.johannes@uclouvain.be; **maik.schwarz@uclouvain.be*

We consider a circular deconvolution problem, in which the density f of a circular random variable X must be estimated nonparametrically based on an i.i.d. sample from a noisy observation Y of X. The additive measurement error is supposed to be independent of X. The objective of this work was to construct a fully data-driven estimation procedure when the error density φ is unknown. We assume that in addition to the i.i.d. sample from Y, we have at our disposal an additional i.i.d. sample drawn independently from the error distribution. We first develop a minimax theory in terms of both sample sizes. We propose an orthogonal series estimator attaining the minimax rates but requiring optimal choice of a dimension parameter depending on certain characteristics of f and φ, which are not known in practice. The main issue addressed in this work is the adaptive choice of this dimension parameter using a model selection approach. In a first step, we develop a penalized minimum contrast estimator assuming that the error density is known. We show that this partially adaptive estimator can attain the lower risk bound up to a constant in both sample sizes n and m. Finally, by randomizing the penalty and the collection of models, we modify the estimator such that it no longer requires any previous knowledge of the error distribution. Even when dispensing with any hypotheses on φ, this fully data-driven estimator still preserves minimax optimality in almost the same cases as the partially adaptive estimator. We illustrate our results by computing minimal rates under classical smoothness assumptions.

Keywords: adaptive density estimation; circular deconvolution; minimax theory; model selection; orthogonal series estimation; spectral cut-off

1. Introduction

This work deals with the estimation of circular probability densities from noisy observations. "Circular" means that the observations are points on the circle. Such models arise in numerous and various fields of application. Data with temporal structure are most naturally represented in this way; for example, times of day when events of interest occur such as requests in a computer network, financial transactions, or gun crimes, can be represented as points on a clock face (Gill and Hangartner [22]), as illustrated in Figure 1. Replacing the clock face by a compass rose, directional data also can be treated in the circular setting. Curray [14] considered the analysis of directional data in the context of geological research. Cochran, Mouritsen and Wikelski [9] investigated migrating birds' navigation abilities using circular data.

The applications of circular data are not restricted to a spatiotemporal context. Gill and Hangartner [22] provided an overview of circular data in political science, where they can be used to, for example, model political preferences, which are not of a temporal or a spatial na-

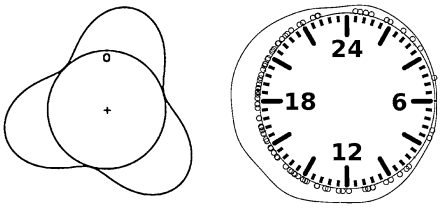

Figure 1. A trimodal circular density and a density estimator from periodic data.

ture. For a more detailed discussion of the specifics of circular data, see Mardia [32]. Numerous circular data sets and examples of their statistical analysis have been provided by Fisher [21].

Let X be the circular random variable whose density f we are interested in and let ε be an independent additive circular error with unknown density φ. Denote by Y the contaminated observation and by g its density. Throughout this work, we identify the circle with the unit interval $[0, 1)$ for notational convenience. Thus, X and ε take their values in $[0, 1)$. Let $\lfloor\cdot\rfloor$ be the floor function. Taking into account the circular nature of the data, the model can be written as $Y = X + \varepsilon - \lfloor X + \varepsilon \rfloor$ or, equivalently, $Y = X + \varepsilon \bmod [0, 1)$. We then have

$$g(y) = (f * \varphi)(y) := \int_{[0,1)} f\big((y - s) - \lfloor y - s \rfloor\big)\varphi(s)\,\mathrm{d}s, \qquad y \in [0, 1),$$

such that $*$ denotes circular convolution. Therefore, the estimation of f is called a circular deconvolution problem. Let $L^2 := L^2([0, 1))$ be the Hilbert space of square-integrable complex-valued functions defined on $[0, 1)$ endowed with the usual inner product $\langle f, g\rangle = \int_{[0,1)} f(x)\overline{g(x)}\,\mathrm{d}x$, where $\overline{g(x)}$ denotes the complex conjugate of $g(x)$. In this work, we suppose that f and φ, and hence also g, belong to the subset $\mathcal{D}$ of all densities in L^2. Consequently, they admit representations as discrete Fourier series with respect to the exponential basis, $\{e_j\}_{j\in\mathbb{Z}}$, of L^2, where $e_j(x) := \exp(-\mathrm{i}2\pi jx)$ for $x \in [0, 1)$ and $j \in \mathbb{Z}$. Given $p \in \mathcal{D}$ and $j \in \mathbb{Z}$, let $[p]_j := \langle p, e_j\rangle$ be the jth Fourier coefficient of p. In particular, $[p]_0 = 1$. The key to the analysis of the circular deconvolution problem is the convolution theorem, which states that $g = f * \varphi$ if and only if $[g]_j = [f]_j[\varphi]_j$ for all $j \in \mathbb{Z}$. Therefore, as long as $[\varphi]_j \neq 0$ for all $j \in \mathbb{Z}$, which we assume from here on, we have

$$f = 1 + \sum_{|j|>0} \frac{[g]_j}{[\varphi]_j} e_j \qquad \text{with } [g]_j = \mathbb{E}e_j(-Y) \text{ and } [\varphi]_j = \mathbb{E}e_j(-\varepsilon)\ \forall j \in \mathbb{Z}. \tag{1.1}$$

Note that an analogous representation holds in the case of deconvolution on the real line when the X-density is compactly supported but the error term ε, and hence Y, take their values in $\mathbb{R}$. In this situation, the deconvolution density still admits a discrete representation as in (1.1), but involving the characteristic functions of φ and g rather than their discrete Fourier coefficients. There is a vast literature on deconvolution on the real line, with or without compactly supported deconvolution density. In the case where the error density is fully known, a very popular approach based on kernel methods has been considered by, among many others, Carroll and Hall

[7], Devroye [15], Fan [18,19], Stefanski [45], Zhang [47], Goldenshluger [23,24], and Kim and Koo [29]. Mendelsohn and Rice [35] and Koo and Park [30], for example, studied spline-based methods, whereas Pensky and Vidakovic [42], Fan and Koo [20], and Bigot and Van Bellegem [2], used wavelet decomposition. Situations with only partial knowledge of the error density have been considered as well (e.g., Butucea and Matias [4], Meister [33], Schwarz and Van Bellegem [44]). Consistent deconvolution without previous knowledge of the error distribution is also possible in the case of panel data (e.g., Horowitz and Markatou [26], Hall and Yao [25], or Neumann [40]) or by assuming an additional sample from the error distribution (e.g., Diggle and Hall [16], Neumann [39], Johannes [27], or Comte and Lacour [11]). For a broader overview on deconvolution problems, see the monograph of Meister [34].

We now return to the circular case. In this paper, we assume that we do not know the density $g = f * \varphi$ of the contaminated observations or the error density φ, but we have at our disposal two independent samples of i.i.d. random variables

$$Y_k \sim g \qquad (k = 1, \dots, n) \quad \text{and} \quad \varepsilon_k \sim \varphi \qquad (k = 1, \dots, m) \tag{1.2}$$

of size $n \in \mathbb{N}$ and $m \in \mathbb{N}$, respectively. Our aim is to establish a fully data-driven estimation procedure for the deconvolution density f that attains optimal convergence rates in a minimax sense. More precisely, given classes $\mathcal{F}_\gamma^r$ and $\mathcal{E}_\lambda^d$ (defined below) of deconvolution and error densities, respectively, we measure the accuracy of an estimator $\widetilde{f}$ of f by the maximal weighted risk $\sup_{f\in\mathcal{F}_\gamma^r} \sup_{\varphi\in\mathcal{E}_\lambda^d} \mathbb{E}\|\widetilde{f} - f\|_\omega^2$ defined with respect to some weighted norm $\|\cdot\|_\omega^2 := \sum_{j\in\mathbb{Z}} \omega_j |[\cdot]_j|^2$, where $\omega := (\omega_j)_{j\in\mathbb{Z}}$ is a strictly positive sequence of weights. This allows us to quantify the estimation accuracy in terms of the mean integrated squared error (MISE) not only of f itself, but also of its derivatives, for example. It is well known that even in case of a known error density, the maximal risk in terms of the MISE in the circular deconvolution problem is essentially determined by the asymptotic behavior of the sequences of Fourier coefficients $([f])_{j\in\mathbb{Z}}$ and $([\varphi])_{j\in\mathbb{Z}}$ of the deconvolution density and the error density, respectively. For a fixed deconvolution density f, a faster decay of the ε-density's Fourier coefficients $([\varphi])_{j\in\mathbb{Z}}$ results in a slower optimal rate of convergence. For example, in the standard context of an ordinary smooth deconvolution density, when $([f])_{j\in\mathbb{Z}}$ decays polynomially, logarithmic rates of convergence appear when the error density is super smooth, that is, $([\varphi])_{j\in\mathbb{Z}}$ has exponential decay. Efromovich [17] treated this special case exclusively. However, this situation and many others are covered by the density classes

$$\mathcal{F}_\gamma^r := \left\{ p \in \mathcal{D} \colon \sum_{j\in\mathbb{Z}} \gamma_j \big|[p]_j\big|^2 =: \|p\|_\gamma^2 \le r \right\} \quad \text{and}$$

$$\mathcal{E}_\lambda^d := \left\{ p \in \mathcal{D} \colon 1/d \le \frac{|[p]_j|^2}{\lambda_j} \le d \ \forall j \in \mathbb{Z} \right\},$$

where $r, d \ge 1$ and the positive weight sequences $\gamma := (\gamma_j)_{j\in\mathbb{Z}}$ and $\lambda := (\lambda_j)_{j\in\mathbb{Z}}$ specify the asymptotic behavior of the respective sequence of Fourier coefficients. In Section 2, we present a lower bound of the maximal weighted risk that is determined essentially by the sequences γ, λ, and ω. This lower bound is composed of two main terms, each of which depends on the size of

one sample but not of the other sample. Let us define an orthogonal series estimator by replacing the unknown Fourier coefficients in (1.1) by empirical counterparts, that is,

$$\begin{aligned}\widehat{f_k} &:= 1 + \sum_{0<|j|\le k} \frac{\widehat{[g]}_j}{\widehat{[\varphi]}_j} \mathbb{1}\left\{\left|\widehat{[\varphi]}_j\right|^2 \ge 1/m\right\} e_j \\ &\text{with } \widehat{[g]}_j := \frac{1}{n}\sum_{i=1}^{n} e_j(-Y_i) \text{ and } \widehat{[\varphi]}_j := \frac{1}{m}\sum_{i=1}^{m} e_j(-\varepsilon_i).\end{aligned} \tag{1.3}$$

For each j, we introduce a threshold for the estimated coefficient $\widehat{[\varphi]}_j$ that corresponds, in accordance with Neumann [39], to the rate at which $[\varphi]_j$ can be estimated. Again, things work out analogously to deconvolution on the real line, where we need only replace the empirical Fourier coefficients with the corresponding values of the empirical characteristic functions. Similar estimators have been studied by, for example, Neumann [39] on the real line and by Efromovich [17] in the circular case.

We show below that the estimator $\widehat{f_k}$ attains the lower bound and thus is minimax optimal. By comparing the minimax rates in the cases of known and unknown error density, we can characterize the influence of the estimation of the error density on the quality of the estimation. In particular, depending on the Y sample size n, we can determine the minimal ε sample size m_n needed to attain the same upper risk bound as in the case of a known error density, up to a constant. Interestingly, the required sample size, m_n, is far smaller than n in a wide range of situations. For example, in the super smooth case, it is sufficient that the size of the ε sample be a polynomial in n, that is, $m_n = n^r$ for any $r > 0$.

Of course, minimax optimality can be achieved only if the dimension parameter k is chosen in an optimal way. In general, this optimal choice of k depends on, among other things, the sequences γ and λ. However, in the special case where the error density is known to be super smooth and the deconvolution density is ordinary smooth, the optimal dimension parameter depends only on λ and not on γ. Thus, the estimator is automatically adaptive with respect to γ under the optimal choice of k. In this situation, Efromovich [17] provided an estimator that is also adaptive with respect to the super smooth error density. In contrast, Cavalier and Hengartner [8], deriving oracle inequalities in an indirect regression problem based on a circular convolution contaminated by Gaussian white noise, treated only the ordinary smooth case. As in our setting, their observation scheme involves two independent samples. Of note, application of these estimators requires knowledge of whether the error density is ordinary or super smooth. In this work, we provide a unified estimation procedure that can attain minimax rates in both cases, being adaptive over a class including both ordinary and super smooth error densities. This fully adaptive method of choosing the parameter k depends only on the observations, not on characteristics of either f or φ. Our main result is that for this automatic choice $\widehat{k}$, the estimator $\widehat{f_{\widehat{k}}}$ attains the lower bound up to a constant, and thus is minimax-optimal, over a wide range of sequences γ and λ, covering in particular both ordinary and super smooth error densities. A similar result was recently derived in the context of a functional linear regression model by Comte and Johannes [10].

Regarding the two sample sizes, the assumption of Cavalier and Hengartner [8] on the respective noise levels can be translated to our model by stating that the ε sample size m is at least as

large as the Y sample size n. This assumption was also made by Efromovich [17]. Also note that in the functional linear regression model, only one sample size, n, occurs (Comte and Johannes [10]); however, as mentioned earlier, without changing the minimax rates, the ε-sample size can be reduced to m_n, which can be much smaller than n. This is a desirable property, given that the observation of the additional sample from ε may be expensive in practice. Nevertheless, the minimal choice of m depends on, among other things, the sequences γ and λ and thus is unknown in general. Despite the eventual deterioration of the minimax rate resulting from choosing the sample size m smaller than n, the proposed estimator still attains this rate in many cases; that is, no price, in terms of convergence rate, is paid for adaptivity.

The adaptive choice of k is motivated by the general model selection strategy developed by Barron, Birgé and Massart [1]. Concretely, following Comte and Taupin [13], who treated the case of a known error density only, $\widehat{k}$ is the minimizer[1] of a penalized contrast

$$\widehat{k} := \operatorname*{argmin}_{1\le k\le K}\big[-\|\widehat{f}_k\|_\omega^2 + \operatorname{pen}(k)\big].$$

Note that we can compute $\|\widehat{f}_k\|_\omega^2 = 1+\sum_{0<|j|\le k}\omega_j|[\widehat{g}]_j|^2|[\widehat{\varphi}]_j|^{-2}\mathbb{1}\{|[\widehat{\varphi}]_j|^2\ge 1/m\}$. As in case of a known error density, it turns out that both the penalty function $\operatorname{pen}(\cdot)$ and the upper bound K needed for the correct choice of k depend on a characteristic of the error density, which is now unknown. This quantity is often referred to as the degree of ill-posedness of the underlying inverse problem. Therefore, as an intermediate step, we allow the penalty function $\operatorname{pen}(\cdot)$ and the upper bound K to depend on the error density. We then show an upper risk bound for the resulting partially adaptive estimator. We prove that over a wide range of sequences γ, this choice of k yields the same upper risk bound as the optimal choice, up to a constant. Finally, we choose k fully adaptively by replacing $\operatorname{pen}(\cdot)$ and K by their empirical versions, which depend only on the data. As in the case of known degree of ill-posedness, we show an upper risk bound for the now fully adaptive estimator.

Let us return briefly to deconvolution on the real line with compactly supported X density. We note that in this situation, the adaptive choice of k can be performed in the same way. Moreover, the upper risk bounds remain valid, and the adaptive estimator is minimax optimal over a wide range of cases. In fact, the circular structure of the model is exploited only in the proof of the lower bound and to guarantee the existence of the discrete representation in (1.1), which still holds in case of a compactly supported deconvolution density.

This paper is organized as follows. In the next section, we develop the minimax theory for the circular deconvolution model with respect to the weighted norms introduced above and compute the rates which we can obtain in different configurations for the weight sequences. We devote the final section to constructing the adaptive estimator and show an upper risk bound. We illustrate our results with example configurations considered in Section 2. All proofs are deferred to the Appendix.

[1]For a sequence a_n attaining a minimum on $N\subseteq\mathbb{N}$, let $\operatorname{argmin}_{n\in N} a_n := \min\{n\in N | a_n \le a_k\ \forall k\in N\}$.

2. Minimax optimal estimation

In this section, we develop the minimax theory for estimating a circular deconvolution density under unknown error density when two independent samples from Y and ε, of size n and m, respectively, are available. We derive a lower bound depending on both sample sizes and show that the orthogonal series estimator $\widehat{f_k}$ defined in (1.3) attains this lower bound up to a constant if k is chosen in an appropriate way. All results in this paper are derived under the following minimal regularity conditions:

Assumption A1. *Let $(\gamma_j)_{j\in\mathbb{Z}}$, $(\omega_j)_{j\in\mathbb{Z}}$ and $(\lambda_j)_{j\in\mathbb{Z}}$ be strictly positive symmetric sequences of weights with $\gamma_0 = \omega_0 = \omega_1 = \lambda_0 = \lambda_1 = 1$ such that $(\omega_n/\gamma_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ are nonincreasing, respectively with $\Lambda := \sum_{j\in\mathbb{Z}} \lambda_j < \infty$.*

Here and subsequently, we refer to any sequence $(a_n)_{n\in\mathbb{Z}}$ as a whole by omitting its index as in, for example "the sequence a". We define arithmetic operations on sequences element-wise. Furthermore, we denote by C universal numerical constants and by $C(\cdot)$ constants depending only on the arguments. In both cases, the values of the constants may change from line to line. Moreover, we write $a_n \lesssim b_n$ when $a_n \le Cb_n$ for all sufficiently large $n \in \mathbb{N}$, and $a_n \sim b_n$ when $a_n \lesssim b_n$ and $b_n \lesssim a_n$ simultaneously.

Lower bounds

The next assertion provides a lower bound in the case of a known error density, which obviously will depend on the size of the Y sample only. Of course, this lower bound is still valid in the case of an unknown error density.

Theorem 2.1. *Assume an i.i.d. Y sample of size n. Consider sequences ω, γ, and λ satisfying Assumption A1 such that $\sum_{j\in\mathbb{Z}} \gamma_j^{-1} = \Gamma < \infty$ and $\varphi \in \mathcal{E}_\lambda^d$ for some $d \ge 1$. Define, for all $n \ge 1$,*

$$\begin{aligned} k_n^* &:= k_n^*(\gamma, \lambda, \omega) := \operatorname*{argmin}_{k\in\mathbb{N}} \left[\max\left(\frac{\omega_k}{\gamma_k}, \sum_{0<|j|\le k} \frac{\omega_j}{n\lambda_j} \right) \right] \quad \text{and} \\ \psi_n &:= \psi_n(\gamma, \lambda, \omega) := \max\left(\frac{\omega_{k_n^*}}{\gamma_{k_n^*}}, \sum_{0<|j|\le k_n^*} \frac{\omega_j}{n\lambda_j} \right). \end{aligned} \tag{2.1}$$

If, in addition, $\eta := \inf_{n\ge1}\{\psi_n^{-1} \min(\omega_{k_n^*}\gamma_{k_n^*}^{-1}, \sum_{0<|l|\le k_n^*} \omega_l (n\lambda_l)^{-1})\} > 0$, then, for all $n \ge 2$

$$\inf_{\widetilde{f}} \sup_{f\in\mathcal{F}_\gamma^r} \left\{ \mathbb{E}\|\widetilde{f} - f\|_\omega^2 \right\} \ge \frac{\eta \min(r-1, 1/(8d\Gamma))}{16} \psi_n,$$

where the infimum is taken over all possible estimators of f.

Remark 2.2. When φ is known, it is natural to consider the orthogonal series estimator $\widetilde{f}_k := 1 + \sum_{1<|j|\le k}(\widehat{[g]}_j/[\varphi]_j)e_j$. It is easily seen that for $|j| \le k$, we have $\mathbb{E}[[\widetilde{f}]_j] = [f]_j$ and $\mathbb{V}\mathrm{ar}([\widetilde{f}]_j) \le (n|[\varphi]_j|^2)^{-1}$, whereas $\mathbb{E}[[\widetilde{f}]_j] = 0$ and $\mathbb{V}\mathrm{ar}([\widetilde{f}]_j) = 0$ for $|j| > k$. Thus, for all $f \in \mathcal{F}_\gamma^r$ and $\varphi \in \mathcal{E}_\lambda^d$, we have

$$\mathbb{E}\big[\|\widetilde{f}_k - f\|_\omega^2\big] \le \sum_{|j|>k} \omega_j |[f]_j|^2 + \frac{1}{n}\sum_{0<|j|\le k} \frac{\omega_j}{|[\varphi]_j|^2} \le (r+d)\max\bigg(\frac{\omega_k}{\gamma_k}, \sum_{0<|j|\le k}\frac{\omega_j}{n\lambda_j}\bigg).$$

Thus, the choice k_n^* of k from (2.1) realizes the best variance–bias trade-off, ψ_n. This demonstrates that when φ is known, $\widetilde{f}_{k_n^*}$ attains the rate ψ_n, which thus is minimax optimal.

The proof of the last assertion is based on Assouad's cube technique (Korostelëv and Tsybakov [31]), which involves constructing $2^{2k_n^*}$ candidates of deconvolution densities that have the largest possible $\|\cdot\|_\omega$-distance but are still statistically indistinguishable. Of note, the additional assumption $\sum_{j\in\mathbb{Z}} \gamma_j^{-1} = \Gamma < \infty$ is used only to ensure that these candidates are densities. Also of note, in the case where $r = 1$, the lower bound is equal to 0, because in this situation the set $\mathcal{F}_\gamma^r$ reduces to a singleton containing only the uniform density. In the next theorem, we state a lower bound characterizing the additional complexity due to the unknown error density, which, surprisingly, depends only on the error sample size.

Theorem 2.3. *Assume* (1.2) *and let* ω, γ, *and* λ *be sequences satisfying Assumption* A1. *For all* $m \ge 2$, *let*

$$\kappa_m := \kappa_m(\gamma, \lambda, \omega) := \max_{j\in\mathbb{N}}\bigg\{\omega_j \gamma_j^{-1} \min\bigg(1, \frac{1}{m\lambda_j}\bigg)\bigg\}. \tag{2.2}$$

If in addition there exists a density in $\mathcal{E}_\lambda^{\sqrt{d}}$ *that is bounded from below by* $1/2$, *then, for all* $m \ge 2$,

$$\inf_{\widetilde{f}} \sup_{f\in\mathcal{F}_\gamma^r} \sup_{\varphi\in\mathcal{E}_\lambda^d} \big\{\mathbb{E}\|\widetilde{f} - f\|_\omega^2\big\} \ge \frac{\min(r-1,1)\min(1/(4d), (1-d^{-1/4})^2)}{4\sqrt{d}}\kappa_m,$$

where the infimum is taken over all possible estimators of f.

The proof of the last assertion takes its inspiration from a proof given by Neumann [39], who proved a similar lower bound for deconvolution on the real line when both densities f and φ are ordinary smooth, that is, γ and λ have polynomial decay. In contrast to the proof of Theorem 2.1, here we only need compare two candidates of error densities that are still statistically indistinguishable. However, to ensure that these candidates are densities, we impose the additional condition. It is easily seen that this condition is satisfied if $\ell := \sum_{j\in\mathbb{Z}} \lambda_j^{-1/2} < \infty$ and $\sqrt{d} \ge \max(4\ell^2, 1)$. Of note, in case where $d = 1$, the set $\mathcal{E}_\lambda^d$ of possible error densities reduces to a singleton, and thus the lower bound is equal to 0. Finally, by a combination of both lower bounds, we obtain the next corollary.

Corollary 2.4. *Under the assumptions of Theorem 2.1 and 2.3 for all* $n, m \geq 2$

$$\inf_{\widetilde{f}} \sup_{f \in \mathcal{F}_\gamma^r} \sup_{\varphi \in \mathcal{E}_\lambda^d} \left\{ \mathbb{E} \| \widetilde{f} - f \|_\omega^2 \right\} \geq C(\eta, r, d, \Gamma) \max(\psi_n, \kappa_m).$$

Upper bound

In the next theorem and all subsequent results, we assume observations according to (1.2). First, we summarize sufficient conditions to ensure the optimality of the orthogonal series estimator $\widehat{f_k}$ defined in (1.3), provided that the dimension parameter k is chosen appropriately. We use the value k_n^* defined in (2.1), which, although obviously involving the sequences ω, γ, and λ, surprisingly does not depend on the ε sample size m. With this choice, the estimator attains the lower bound given in Corollary 2.4 up to a constant and thus is minimax-optimal.

Theorem 2.5. *Under Assumption* A1, *we have, for all* $n, m \geq 1$,

$$\sup_{f \in \mathcal{F}_\gamma^r} \sup_{\varphi \in \mathcal{E}_\lambda^d} \left\{ \mathbb{E} \| \widehat{f}_{k_n^*} - f \|_\omega^2 \right\} \leq C \left\{ (d + r) \psi_n + dr \kappa_m \right\}.$$

Note that under slightly stronger conditions on the sequences ω, γ, and λ than those in Assumption A1, it can be shown that in the case of equally large samples from Y and ε, we always have the same rate as in the case of known error density. However, in special cases, the required ε sample size can be much smaller than the Y sample size, as we show below.

Illustration: Estimation of derivatives

Here we illustrate our results considering classical smoothness assumptions. Regarding the deconvolution density f, it is interesting to recall that the class $\mathcal{F}_\gamma^r$ is a subset of the Sobolev space of p-times differentiable periodic functions if $\gamma_j \sim |j|^{2p}$ (Neubauer [37,38]). We call this case *ordinary smooth*. Moreover, up to a constant, for any function $h \in \mathcal{F}_\gamma^r$, the weighted norm $\|h\|_\omega$ with $\omega_j \sim j^{2s}$ equals the L^2 norm of the sth weak derivative $h^{(s)}$ for each integer $0 \leq s \leq p$. By virtue of this relationship, the results in the previous section imply both a lower bound and an upper bound of the L^2 risk for estimation of the sth weak derivative of f. If, in contrast, $\gamma_j \sim \exp(|j|^{2p})$ with $p > 1$, then $\mathcal{F}_\gamma^r$ is a class of analytic functions (Kawata [28]). We refer to this situation as *super smooth*.

As for the error densities, we consider two special cases corresponding to a regular decay of their Fourier coefficients. The error density is called *ordinary smooth* if $\lambda_j \sim |j|^{-2a}$ for some $a > 1/2$ and *super smooth* if $\lambda_j \sim \exp(-|j|^{2a})$ for some $a > 0$.

We consider the following three situations: In the cases [o-o] and [s-o], the error density is ordinary smooth and the deconvolution density is either ordinary smooth or super smooth case, respectively. Case [o-s] is the opposite of case [s-o].

It is readily seen that in all of these cases, the minimal regularity conditions given in Assumption A1 and the additional conditions in Theorems 2.1 and 2.3 translate to simple restrictions on

p, a, and s, which are given in the following proposition. Roughly speaking, these restrictions imply that both the deconvolution density and the error density are at least continuous. The lower bounds presented in the following assertion follow directly from Corollary 2.4:

Proposition 2.6.

[o-o] *For $p > 1/2$, $a > 1$, and $0 \le s \le p$, we have for all $n, m \ge 1$*

$$\inf_{\widetilde{f}^{(s)}} \sup_{f \in \mathcal{F}_\gamma^r} \sup_{\varphi \in \mathcal{E}_\lambda^d} \{\mathbb{E}\| \widetilde{f}^{(s)} - f^{(s)} \|^2\} \gtrsim n^{-2(p-s)/(2p+2a+1)} + m^{-((p-s)\wedge a)/a}.$$

[s-o] *For $p > 0$, $a > 1$, and $s \ge 0$, we have for all $n, m \ge 1$*

$$\inf_{\widetilde{f}^{(s)}} \sup_{f \in \mathcal{F}_\gamma^r} \sup_{\varphi \in \mathcal{E}_\lambda^d} \{\mathbb{E}\| \widetilde{f}^{(s)} - f^{(s)} \|^2\} \gtrsim n^{-1} (\log n)^{(2a+2s+1)/(2p)} + m^{-1}.$$

[o-s] *For $p > 1/2$, $a > 0$, and $0 \le s \le p$, we have for all $n, m \ge 1$*

$$\inf_{\widetilde{f}^{(s)}} \sup_{f \in \mathcal{F}_\gamma^r} \sup_{\varphi \in \mathcal{E}_\lambda^d} \{\mathbb{E}\| \widetilde{f}^{(s)} - f^{(s)} \|^2\} \gtrsim (\log n)^{-(p-s)/a} + (\log m)^{-(p-s)/a}.$$

Remark 2.7. We do not treat the doubly exponential case [s-s] here, because doing so would require rather intricate computations and distinctions of cases. A detailed analysis of this case in the context of density deconvolution on the real line has been provided by Butucea and Tsybakov [5,6]. Note that the expressions in n in the foregoing result coincide with the lower bounds for the deconvolution problem on the real line, which can be found in the literature. For example, in cases where the error distribution is known, Fan [18] have addressed the cases [o-o] and [o-s] and Butucea [3] examined the case [s-o].[2] Those authors developed kernel-based estimation procedures which attain these lower bounds. In the case [o-o] (still on the real line), for cases where the error density is unknown, [39] also investigated the impact of estimating the error density and obtained the same lower bound as in the foregoing result.

As an estimator of $f^{(s)}$, we consider the sth weak derivative of the estimator $\widehat{f}_k$ defined in (1.3), with k as specified below. Given the exponential basis $\{e_j\}_{j \in \mathbb{Z}}$, we recall that for each integer $0 \le s \le p$, the sth derivative in a weak sense of the estimator $\widehat{f}_k$ is

$$\widehat{f}_k^{(s)} = \sum_{j \in \mathbb{Z}} (2i\pi j)^s [\widehat{f}_k]_j e_j. \tag{2.3}$$

As an immediate consequence of Theorem 2.5, the rates of the lower bound given by Proposition 2.6 are attained for $k = k_n^*$, as summarized in the next result. Thus, we have proven that these rates are optimal and that the proposed estimator $\widehat{f}_{k_n^*}^{(s)}$ is minimax optimal in both cases. Furthermore, it is of interest to characterize the minimal size m of the additional sample from ε

[2]When comparing the bounds, attention must be given to the slightly different parameterizations of the density classes in the cited articles.

needed to attain the same rate as in case of a known error density. Thus, we let the ε-sample size depend on the Y-sample size n as well.

Proposition 2.8. *Let $(m_n)_{n\geq 1}$ be a sequence of positive integers*:

[o-o] *For $p > 1/2$, $a > 1$, and $0 \leq s \leq p$ with $k_n^* \sim n^{1/(2p+2a+1)}$, we have for all $n, m \geq 1$*

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\left\{\mathbb{E}\left\|\widehat{f}_{k_n^*}^{(s)} - f^{(s)}\right\|^2\right\} \lesssim n^{-2(p-s)/(2p+2a+1)} + m^{-((p-s)\wedge a)/a}$$

and if $q_{\text{o-o}} := \lim_{n\to\infty} n^{2((p-s)\vee a)/(2p+2a+1)} m_n^{-1}$ exists,[3] *then it follows that as $n \to \infty$*

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\left\{\mathbb{E}\left\|\widehat{f}_{k_n^*}^{(s)} - f^{(s)}\right\|^2\right\} = \begin{cases} \mathrm{O}\big(n^{-2(p-s)/(2p+2a+1)}\big) & \text{if } q_{\text{o-o}} < \infty, \\ \mathrm{O}\big(m_n^{-((p-s)\wedge a)/a}\big) & \text{otherwise.} \end{cases}$$

[s-o] *For $p > 0$, $a > 1$, and $s \geq 0$ with $k_n^* \sim (\log n)^{1/(2p)}$, we have, for all $n, m \geq 1$,*

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\left\{\mathbb{E}\left\|\widehat{f}_{k_n^*}^{(s)} - f^{(s)}\right\|^2\right\} \lesssim n^{-1}(\log n)^{(2a+2s+1)/(2p)} + m^{-1}$$

and if $q_{\text{s-o}} := \lim_{n\to\infty} n(\log n)^{-(2a+2s+1)/(2p)} m_n^{-1}$ exists, it follows as $n \to \infty$

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\left\{\mathbb{E}\left\|\widehat{f}_{k_n^*}^{(s)} - f^{(s)}\right\|^2\right\} = \begin{cases} \mathrm{O}\big(n^{-1}(\log n)^{(2a+2s+1)/(2p)}\big) & \text{if } q_{\text{s-o}} < \infty, \\ \mathrm{O}\big(m_n^{-1}\big) & \text{otherwise.} \end{cases}$$

[o-s] *For $p > 1/2$, $a > 0$, and $0 \leq s \leq p$ with $k_n^* \sim (\log n)^{1/(2a)}$, we have, for all $n, m \geq 1$,*

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\left\{\mathbb{E}\left\|\widehat{f}_{k_n^*}^{(s)} - f^{(s)}\right\|^2\right\} \lesssim (\log n)^{-(p-s)/a} + (\log m)^{-(p-s)/a}$$

and if $q_{\text{o-s}} := \lim_{n\to\infty} (\log n)(\log m_n)^{-1}$ exists, then it follows, as $n \to \infty$,

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\left\{\mathbb{E}\left\|\widehat{f}_{k_n^*}^{(s)} - f^{(s)}\right\|^2\right\} = \begin{cases} \mathrm{O}\big((\log n)^{-(p-s)/a}\big) & \text{if } q_{\text{o-s}} < \infty, \\ \mathrm{O}\big((\log m_n)^{-(p-s)/a}\big) & \text{otherwise.} \end{cases}$$

The existence of the limits $q_{\text{o-o}}$, $q_{\text{o-s}}$, and $q_{\text{s-o}}$ is required only to exclude the case of oscillating sequences, which we are not interested in here. In this case, none of the two terms in the upper bound is asymptotically dominant, and the convergence rate is the alternating maximum of the two terms.

In the case [o-o], whenever $n^{2((p-s)\vee a)/(2p+2a+1)} = \mathrm{O}(m_n)$, which is much less than $m_n = n$, we obtain the rate of known error density. This is even more visible in the case [o-s], where the rate of known error density is attained even if $m_n = n^r$ for arbitrarily small $r > 0$. Moreover, we

[3] The limit "∞" is authorized, with $\lim_{n\to\infty} a_n = \infty :\Longleftrightarrow \forall K > 0\ \exists n_0 \in \mathbb{N}\ \forall n \geq n_0\colon\ a_n \geq K$.

emphasize the influence of the parameter a that characterizes the rate of decay of the Fourier coefficients of the error density φ. Because a smaller value of a leads to faster rates of convergence, this parameter is often called *degree of ill-posedness* (e.g., Natterer [36]).

3. Adaptive estimation

Our aim is to construct an adaptive estimator of the deconvolution density f. Adaptation means that despite an unknown error density in $\mathcal{E}_\lambda^d$, the estimator should attain the optimal rate of convergence $\max(\psi_n, \kappa_m)$ over the ellipsoid $\mathcal{F}_\gamma^r$ for a wide range of different weight sequences γ and λ.

In a first step, we suppose that φ is known, but γ and r are unknown. In what follows, we consider the orthogonal series estimator $\widehat{f_k}$ defined in (1.3) and construct a procedure to choose the dimension parameter k based on a model selection approach via penalization. This partially adaptive choice $\widetilde{k}$ will involve only the data and the error density φ.

In a second step, we replace φ with its empirical version and thus dispense with any knowledge about φ. Doing so, we obtain a fully adaptive choice $\widehat{k}$ of the dimension parameter.

Partially adaptive estimation knowing φ

We first introduce sequences that are used below.

Definition 3.1. *For all $n, m \geq 1$ and $k \geq 0$, define*

(i) $\Delta_k := \Delta_k(\varphi) := \max_{-k\leq j\leq k} \frac{\omega_j}{|[\varphi]_j|^2}$ *and* $\delta_k := \delta_k(\varphi) := 2k\Delta_k \frac{\log(\Delta_k \vee (k+2))}{\log(k+2)}$;

(ii) *given* $\omega_k^+ := \max_{0\leq j\leq k} \omega_j$ *and* $N_n^\circ := \max\{1 \leq N \leq n | \omega_N^+ \leq n\}$, *let*

$$N_n := N_n(\varphi) := \min\left\{1 \leq j \leq N_n^\circ \Big| \frac{|[\varphi]_j|^2}{j\omega_j^+} \leq \frac{\log(n+2)}{n}\right\} - 1,$$

defining further $b_m := (8\log(\log(m+20)))^{-1}$, *let*

$$M_m := M_m(\varphi) := \min\left\{1 \leq j \leq m \big| \big|[\varphi]_j\big|^2 \leq m^{-1+b_m}\right\} - 1;$$

with $N_n := N_n^\circ$ and $M_m := m$ when the respective set in the definition is empty.

These sequences are used for small sample sizes as well, which explains their rather complicated form. We can now define a partially adaptive choice of the dimension parameter k,

$$\widetilde{k} := \operatorname*{argmin}_{0\leq k\leq (N_n \wedge M_m)} \left[-\|\widehat{f_k}\|_\omega^2 + 60\frac{\delta_k}{n}\right], \tag{3.1}$$

which obviously depends only on the data and the error density φ. We obtain the fully adaptive estimator below by introducing the empirical versions of δ, N, and M given in Definition 3.1.

For a fixed φ, we could now derive an upper risk bound for the partially adaptive estimator $\widehat{f}_{\widetilde{k}}$, which would depend on δ, N, and M. But because we wish to obtain a uniform upper risk bound over the class $\mathcal{E}_{\lambda}^{d}$, instead we now redefine the foregoing objects referring only to the weight sequence λ and the constant d.

Definition 3.2. *Let ω^{+}, N°, and b as in Definition 3.1.*

(i) *For all $k \geq 0$, define $\Delta_k^{\lambda} := \max_{-k \leq j \leq k} \omega_j / \lambda_j$ and*

$$\delta_k^{\lambda} := 2k\Delta_k^{\lambda} \frac{\log(\Delta_k^{\lambda} \vee (k+2))}{\log(k+2)}.$$

(ii) *Define two sequences, N^{λ} and M^{λ}, as follows*:

$$N_n^{\lambda} := \min\left\{1 \leq j \leq N_n^{\circ} \middle| \frac{\lambda_j}{j\omega_j^{+}} < \frac{4d\log(n+2)}{n}\right\} - 1,$$

$$M_m^{\lambda} := \min\left\{1 \leq j \leq m | \lambda_j < 4dm^{-1+b_m}\right\} - 1.$$

If the set in the definition is empty, then we set $N_n^{\lambda} := 0$ or $M_m^{\lambda} := 0$, respectively.

(iii) *Define two sequences, N^{u} and M^{u}, as follows*:

$$N_n^{u} := N_n^{u}(\lambda) := \min\left\{1 \leq j \leq n \middle| \frac{\lambda_j}{j\omega_j^{+}} < \frac{\log(n+2)}{4dn}\right\} - 1,$$

$$M_m^{u} := M_m^{u}(\lambda) := \min\left\{1 \leq j \leq m \middle| \lambda_j < \frac{m^{-1+b_m}}{4d}\right\} - 1.$$

If the set in the definition is empty, we set $N_n^{u} := n$ or $M_m^{u} := m$.

(iv) *Let $\Sigma : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function such that, for all $C > 0$,*

$$\sum_{k \geq 1} C\Delta_k^{\lambda} \exp\left(-\frac{k\log(\Delta_k^{\lambda} \vee (k+2))}{3C\log(k+2)}\right) \leq \Sigma(C) < \infty.$$

It is easy to see that there exists always a function Σ satisfying the defining condition. Moreover, as we show in Lemma A1 in the Appendix, the sequences defined above satisfy $N_n^{\lambda} \leq N_n \leq N_n^{u}$ and $M_m^{\lambda} \leq M_m \leq M_m^{u}$ for all $n, m \in \mathbb{N}$. In the illustration below we compute these objects explicitly.

Theorem 3.3. *Let $\zeta_d := \log(3d)/\log(d)$. Under Assumption A1, for all $n, m \geq 1$,*

$$\sup_{f \in \mathcal{F}_{\gamma}^{r}} \sup_{\varphi \in \mathcal{E}_{\lambda}^{d}} \left\{\mathbb{E}\|\widehat{f}_{\widetilde{k}} - f\|_{\omega}^{2}\right\} \leq C\left\{(r + d\zeta_d) \min_{0 \leq k \leq (N_n^{\lambda} \wedge M_m^{\lambda})} \left[\max\left(\frac{\omega_k}{\gamma_k}, \frac{\delta_k^{\lambda}}{n}\right)\right] + rd\kappa_m\right\}$$

$$+ C(r, d, \Lambda, \Sigma)\left[\frac{1}{m} + \frac{1}{n}\right].$$

A comparison with the lower bound from Corollary 2.4 shows that this upper bound ensures minimax optimality of the estimator $\widehat{f}_{\widetilde{k}}$ only if

$$\psi_{n,m}^{\diamond} := \min_{1\le k\le (N_n^\lambda \wedge M_m^\lambda)} \left[\max\left(\frac{\omega_k}{\gamma_k}, \frac{\delta_k^\lambda}{n} \right) \right]$$

is in the same order as $\psi_n = \min_{k\in\mathbb{N}}\{\max(\frac{\omega_k}{\gamma_k}, \sum_{0<|j|\le k} \frac{\omega_j}{n\lambda_j})\}$. Note that, by construction, $\delta_k^\lambda \ge \sum_{0<|j|\le k} \omega_j \lambda_j^{-1}$ for all $k \ge 1$. In addition, δ^λ is direcly related to the penalty function. The next assertion is a immediate consequence of Theorem 3.3, and we omit its proof.

Corollary 3.4. *Under Assumption* A1, *and if*

$$\eta^{\diamond} := \sup_{n,m\ge 1} \left\{ \psi_{n,m}^{\diamond} / \max(\psi_n, \kappa_m) \right\} < \infty,$$

then we have, for all $n, m \ge 1$,

$$\sup_{f\in\mathcal{F}_\gamma^r} \sup_{\varphi\in\mathcal{E}_\lambda^d} \left\{ \mathbb{E}\| \widehat{f}_{\widetilde{k}} - f \|_\omega^2 \right\} \le C\big(\eta^{\diamond}, \Sigma, r, d, \Lambda\big) \max(\psi_n, \kappa_m).$$

In Theorem 2.5, we have shown the minimax optimality of the orthogonal series estimator under the optimal choice k_n^* of the dimension parameter. Comparing Corollary 3.4 with this theorem, it is noteworthy that the only additional assumption needed to ensure minimax optimality of the partially adaptive estimator is $\eta^{\diamond} < \infty$.

Remark 3.5. The partially adaptive choice $\widetilde{k}$ still depends on $\varphi \in \mathcal{E}_\lambda^d$. However, we can already define a procedure depending only on the sequence λ and the constant d, namely

$$\widetilde{k}^\lambda := \operatorname*{argmin}_{1\le k\le (N_n^\lambda \wedge M_m^\lambda)} \left[-\| \widehat{f}_k \|_\omega^2 + 60 \frac{d\delta_k^\lambda}{n} \right].$$

Roughly speaking, this choice requires knowledge of the degree of ill-posedness of the underlying inverse problem only. It is straightforward to derive an upper risk bound for $\widehat{f}_{\widetilde{k}^\lambda}$, which is, up to minor changes in the constants, the same as that in Theorem 3.3. Its proof follows the lines of the proof of Theorem 3.3, using the new penalty term $\mathrm{pen}(k) = 60 d\delta_k^\lambda$. The only change occurs when applying Lemma A4, which uses $\delta_k^* = d\delta_k^\lambda$ and $\Delta_k^* = d\Delta_k^\lambda$ rather than $\delta_k^* = \delta_k$ and $\Delta_k^* = \Delta_k$.

Fully adaptive estimation

We begin by defining empirical versions of the sequences given in Definition 3.1.

Definition 3.6. *For all $n, m \geq 1$ and $k \geq 0$, define*

(i) $\widehat{\Delta}_k := \max_{-k \leq j \leq k} \frac{\omega_j}{|[\widehat{\varphi}]_j|^2} \mathbb{1}\{|[\widehat{\varphi}]_j|^2 \geq 1/m\}$ *and* $\widehat{\delta}_k := k\widehat{\Delta}_k \frac{\log(\widehat{\Delta}_k \vee (k+2))}{\log(k+2)}$;

(ii) *given N_n°, ω^+, and b from Definition 3.1,*

$$\widehat{N}_n := \min\left\{1 \leq j \leq N_n^\circ \,\middle|\, \frac{\min(|[\widehat{\varphi}]_j|^2, |[\widehat{\varphi}]_{-j}|^2)}{j\omega_j^+} < \frac{\log(n+2)}{n}\right\} - 1,$$

$$\widehat{M}_m := \min\left\{1 \leq j \leq m \,\middle|\, \min\left(\left|[\widehat{\varphi}]_j\right|^2, \left|[\widehat{\varphi}]_{-j}\right|^2\right) < m^{-1+b_m}\right\} - 1,$$

with $\widehat{N}_n := N_n^\circ$ and $\widehat{M}_m := m$ if the respective sets in the definition are empty.

We now define a data-driven choice of k, which, in contrast to $\widetilde{k}$, depends not on the sequences δ, N, or M, but rather on $\widehat{\delta}$, $\widehat{N}$, and $\widehat{M}$:

$$\widehat{k} := \operatorname*{argmin}_{0 \leq k \leq (\widehat{N}_n \wedge \widehat{M}_m)} \left[-\|\widehat{f}_k\|_\omega^2 + 600\frac{\widehat{\delta}_k}{n}\right]. \tag{3.2}$$

The constant 600 arising in the definition of $\widehat{k}$, although convenient for deriving the theory, may be far too large in practice and instead be determined by means of a simulation study, as done by Comte, Rozenholc and Taupin [12], for example.

In the proof of Theorem 3.3, we used $(N_n^\lambda \wedge M_m^\lambda) \leq (N_n \wedge M_m) \leq (N_n^u \wedge M_m^u)$ (Lemma A1). In the proof of the next theorem, we consider the event $\{(N_n^\lambda \wedge M_m^\lambda) \leq (\widehat{N}_n \wedge \widehat{M}_m) \leq (N_n^u \wedge M_m^u)\}$, on which we can imitate the proof of Theorem 3.3. To control the risk on the complement of this event, we need to bound its probability, which necessitates the following assumption.

Assumption A2. *Suppose that $m^7 \exp(-m\lambda_{M_m^u+1}/(72d)) \leq C(\lambda, d)$ for all $m \geq 1$.*

Theorem 3.7. *Under Assumptions A1 and A2, we have, for all $n, m \geq 1$,*

$$\sup_{f \in \mathcal{F}_\gamma^r} \sup_{\varphi \in \mathcal{E}_\lambda^d} \left\{\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2\right\} \leq C\left\{(r + d\zeta_d) \min_{0 \leq k \leq (N_n^\lambda \wedge M_m^\lambda)} \left[\max\left(\frac{\omega_k}{\gamma_k}, \frac{\delta_k^\lambda}{n}\right)\right] + rd\kappa_m\right\}$$
$$+ C(r, d, \lambda, \Sigma)\left[\frac{1}{m} + \frac{1}{n}\right].$$

Remark 3.8. Up to a change in the constant in front of the negligible terms, we obtain the same bound as for the partially adaptive estimator (Theorem 3.3). Compared with Theorem 3.3, the only additional assumption is A2. Note that in Lemma A2(ii) in the Appendix, we show that $m^7 \exp(-m\lambda_{M_m^u}/(72d)) \leq C(d)$ for all $m \geq 1$ using only Assumption A1. However, it is not obvious to us that Assumption A1 also implies the slightly stronger assertion $m^7 \exp(-m\lambda_{M_m^u+1}/(72d)) \leq C(d)$ for sufficiently large m, although in the illustrations below, we show that Assumption A2 is satisfied.

Comparing Theorem 3.7 with the lower bound from Corollary 2.4 shows that this upper bound does not necessarily ensure minimax optimality of the estimator $\widehat{f}_{\widehat{k}}$. However, as in the partially adaptive case (cf. Corollary 3.4), under the additional assumption $\eta^{\diamond} < \infty$, the next assertion establishes its optimality. Because this is an immediate consequence of Theorem 3.7, we omit the proof.

Corollary 3.9. *Under Assumptions* A1 *and* A2, *and if*

$$\eta^{\diamond} = \sup_{n,m\geq 1} \left\{ \psi^{\diamond}_{n,m} / \max(\psi_n, \kappa_m) \right\} < \infty,$$

we have, for all $n, m \geq 1$,

$$\sup_{f\in\mathcal{F}^r_\gamma} \sup_{\varphi\in\mathcal{E}^d_\lambda} \left\{ \mathbb{E} \| \widehat{f}_{\widehat{k}} - f \|^2_\omega \right\} \leq C\big(\eta^{\diamond}, \Sigma, r, d, \Lambda\big) \max(\psi_n, \kappa_m).$$

Conclusion

The minimax optimality of the estimator $\widehat{f}_{k^*_n}$ has been shown under Assumption A1 in Theorem 2.5, where the choice k^*_n of the dimension parameter depends on the deconvolution density f and the error density φ. We have developed a fully data-driven choice $\widehat{k}$. The foregoing results show that we need only the additional Assumptions A2 and $\eta^{\diamond} < \infty$ for the adaptive estimator $\widehat{f}_{\widehat{k}}$ to be minimax optimal as well.

Illustration: Estimation of derivatives (continued from Section 2)

The following result shows that without any prior knowledge on the error density φ, the adaptive penalized estimator automatically attains the optimal rate in the cases [o-s] and [s-o] and in the case [o-o] if $p - s > a$. Recall that the computation of the dimension parameter $\widehat{k}$ given in (3.2) involves the sequence N°, which in our illustration satisfies $N^{\circ}_n \asymp n^{1/(2s)}$.

Proposition 3.10. *Let* $(m_n)_{n\geq 1}$ *be a sequence of positive integers and suppose that the limits* $q_{\text{o-o}}$, $q_{\text{o-s}}$, *and* $q_{\text{s-o}}$ *defined in Proposition* 2.8 *exist in the respective cases.*

[o-o] *We have that*

$$\Delta^{\lambda}_k \sim k^{2a+2s}, \qquad \delta^{\lambda}_k \asymp k^{2a+2s+1}, \qquad \psi^{\diamond}_{n,m_n} \sim \big(k^*_n \wedge M^{\lambda}_{m_n}\big)^{-2(p-s)},$$
$$N^{\lambda}_n \asymp (n/\log n)^{1/(2a+2s+1)}, \qquad M^{\lambda}_{m_n} \asymp m_n^{(1-b_m)/(2a)}.$$

In the case wehere $p - s > a$, *the adaptive estimator* $\widehat{f}^{(s)}_{\widehat{k}}$ *attains the optimal rates* (*see Proposition* 2.8). *In the case where* $p - s \leq a$, *if* $q_{\text{o-o}} < \infty$, *then we have, supposing that* $q^b_{\text{o-o}} := \lim_{n\to\infty} n^{2a/(2p+2a+1)} m_n^{-1+b_{m_n}}$ *exists,*

$$\sup_{f\in\mathcal{F}^r_\gamma} \sup_{\varphi\in\mathcal{E}^d_\lambda} \left\{ \mathbb{E} \big\| \widehat{f}^{(s)}_{\widehat{k}} - f^{(s)} \big\|^2 \right\} = \begin{cases} \mathrm{O}\big(n^{-2(p-s)/(2p+2a+1)}\big) & \text{if } q^b_{\text{o-o}} < \infty, \\ \mathrm{O}\big(m_n^{-(p-s)/a} m_n^{b_{m_n}}\big) & \text{otherwise,} \end{cases}$$

whereas if $q_{\text{o-o}} = \infty$, *then we have*

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\{\mathbb{E}\|\widehat{f}_{\widehat{k}}^{(s)} - f^{(s)}\|^2\} = \mathrm{O}\big(m_n^{-(p-s)/a} m_n^{b_m}\big).$$

[s-o] *The sequences* Δ^λ, δ^λ, N^λ, *and* M^λ *are the same as above. We have that* $\psi_{n,m_n}^{\diamond} \sim (k_n^* \wedge M_{m_n}^\lambda)^{2s}\exp(-(k_n^* \wedge M_{m_n}^\lambda)^{2p})$, *and* $\widehat{f}_{\widehat{k}}^{(s)}$ *attains the optimal rates.*

[o-s] *We have that*

$$\Delta_k^\lambda = k^{2s}\exp\big(k^{2a}\big), \qquad \delta_k^\lambda \asymp k^{2a+2s+1}\exp\big(k^{2a}\big)(\log k)^{-1},$$
$$\psi_{n,m}^{\diamond} \sim \big(k_n^* \wedge M_{m_n}^\lambda\big)^{-2(p-s)},$$
$$N_n^\lambda \asymp \Big(\log\big(n/(\log n)^{(2a+2s+1)/(2a)}\big)\Big)^{1/(2a)}, \qquad M_{m_n}^\lambda \asymp \big((1-b_m)\log m_n\big)^{1/(2a)},$$

and the adaptive estimator $\widehat{f}_{\widehat{k}}^{(s)}$ *attains the optimal rates.*

The adaptive estimator always attains the minimal rates if $n \lesssim m_n$. We emphasize that this still holds when $m_n \lesssim n$, except in the case [o-s] when the error density is smoother than the sth derivative of the deconvolution density ($p - s \le a$) and when at the same time m_n grows far more slowly than n. The estimation of φ is negligible as soon as $m_n^{1-b_{m_n}}$ grows at least as fast as $n^{2a/(2p+2a+1)}$ in this situation, whereas in the nonadaptive case, only m_n must satisfy this condition. In the lossy case, the convergence rate differs from the optimal rate by a factor $m_n^{b_{m_n}}$ only; however, the exponent b_{m_n} tends to 0 as n tends to infinity.

If considering the [o-s] case only, we could replace the bound m^{-1+b_m} by $m^{-1}\log m$ in the definition of M^u (Definition 3.2). Using this definition, Assumption A2 would still hold, and applying Theorem 3.7, the adaptive estimator would miss the optimal rates by a logarithmic factor in the lossy case only. However, Assumption A2 is violated in the super smooth case under this definition of M^u.

Appendix: Proofs

A.1. Proofs of Section 2 (minimax theory)

Lower bounds

Proof of Theorem 2.1. Given $\zeta := \eta\min(r-1, 1/(8d\Gamma))$ and $\alpha_n := \psi_n(\sum_{0<|j|\le k_n^*}\omega_j/(\lambda_j \times n))^{-1}$, we consider the function $f := 1 + (\zeta\alpha_n/n)^{1/2}\sum_{0<|j|\le k_n^*}\lambda_j^{-1/2}e_j$. We show that for any $\theta := (\theta_j) \in \{-1,1\}^{2k_n^*}$, the function $f_\theta := 1 + \sum_{0<|j|\le k_n^*}\theta_j[f]_j e_j$ belongs to $\mathcal{F}_\gamma^r$ and thus is a possible candidate for the deconvolution density. For each θ, the Y density corresponding to the X density f_θ is given by $g_\theta := f_\theta * \varphi$. We denote by g_θ^n the joint density of an i.i.d. n sample from g_θ and by $\mathbb{E}_\theta$ the expectation with respect to the joint density g_θ^n. Furthermore, for $0 < |j| \le k_n^*$

and each θ, we introduce $\theta^{(j)}$ by $\theta_l^{(j)} = \theta_l$ for $j \neq l$ and $\theta_j^{(j)} = -\theta_j$. The key argument of this proof is the following reduction scheme. If $\widetilde{f}$ denotes an estimator of f, then we conclude that

$$\begin{aligned}\sup_{f\in\mathcal{F}_\gamma^r} \mathbb{E}\|\widetilde{f}-f\|_\omega^2 &\geq \sup_{\theta\in\{-1,1\}^{2k_n^*}} \mathbb{E}_\theta\|\widetilde{f}-f_\theta\|_\omega^2 \geq \frac{1}{2^{2k_n^*}} \sum_{\theta\in\{-1,1\}^{2k_n^*}} \mathbb{E}_\theta\|\widetilde{f}-f_\theta\|_\omega^2 \\ &\geq \frac{1}{2^{2k_n^*}} \sum_{\theta\in\{-1,1\}^{2k_n^*}} \sum_{0<|j|\leq k_n^*} \omega_j \mathbb{E}_\theta\big|[\widetilde{f}-f_\theta]_j\big|^2 \\ &= \frac{1}{2^{2k_n^*}} \sum_{0<|j|\leq k_n^*} \frac{\omega_j}{2} \sum_{\theta\in\{-1,1\}^{2k_n^*}} \big\{\mathbb{E}_\theta\big|[\widetilde{f}-f_\theta]_j\big|^2 + \mathbb{E}_{\theta^{(j)}}\big|[\widetilde{f}-f_{\theta^{(j)}}]_j\big|^2\big\},\end{aligned}$$

where for each $0 < |j| \leq k_n^*$ and any function $F : \{-1, 1\}^{2k_n^*} \to \mathbb{R}$, we have

$$\sum_{\theta\in\{-1,1\}^{2k_n^*}} F(\theta) = \sum_{\theta\in\{-1,1\}^{2k_n^*}} F\big(\theta^{(j)}\big).$$

Below we show that for all $n \geq 2$, we have

$$\big\{\mathbb{E}_\theta\big|[\widetilde{f}-f_\theta]_j\big|^2 + \mathbb{E}_{\theta^{(j)}}\big|[\widetilde{f}-f_{\theta^{(j)}}]_j\big|^2\big\} \geq \frac{\zeta\alpha_n}{4\lambda_j n}. \tag{A.1}$$

Combining the last lower bound and the reduction scheme gives

$$\sup_{f\in\mathcal{F}_\gamma^r} \mathbb{E}\|\widetilde{f}-f\|_\omega^2 \geq \frac{1}{2^{2k_n^*}} \sum_{\theta\in\{-1,1\}^{2k_n^*}} \sum_{0<|j|\leq k_n^*} \frac{\omega_j}{2}\frac{\zeta\alpha_n}{4\lambda_j n} = \frac{\zeta}{8}\alpha_n \sum_{0<|j|\leq k_n^*} \frac{\omega_j}{\lambda_j n}.$$

Thus, using the definition of ζ and α_n, we obtain the lower bound given in the theorem.

To conclude the proof, it remains to check (A.1) and $f_\theta \in \mathcal{F}_\gamma^r$ for all $\theta \in \{-1, 1\}^{2k_n^*}$. The latter is easily verified if $f \in \mathcal{F}_\gamma^r$. To show that $f \in \mathcal{F}_\gamma^r$, we first note that f integrates to 1. Moreover, f is nonnegative, because $|\sum_{0<|j|\leq k_n^*}[f]_j e_j| \leq 1$ and $\|f\|_\gamma^2 \leq r$, which can be realized as follows. Using the condition $\sum_{j\in\mathbb{Z}} \gamma_j^{-1} = \Gamma < \infty$, we have

$$\begin{aligned}\Big|\sum_{0<|j|\leq k_n^*} [f]_j e_j\Big| &\leq \sum_{0<|j|\leq k_n^*} \big|[f]_j\big| = \left(\frac{\zeta\alpha_n}{n}\right)^{1/2} \sum_{0<|j|\leq k_n^*} \lambda_j^{-1/2} \\ &\leq (\zeta\alpha_n)^{1/2}\Bigg(\sum_{0<|j|\leq k_n^*} \gamma_j^{-1}\Bigg)^{1/2}\Bigg(\sum_{0<|j|\leq k_n^*} \frac{\gamma_j}{n\lambda_j}\Bigg)^{1/2} \\ &\leq (\zeta\alpha_n\Gamma)^{1/2}\Bigg(\sum_{0<|j|\leq k_n^*} \frac{\gamma_j}{n\lambda_j}\Bigg)^{1/2}.\end{aligned}$$

Because ω/γ is nonincreasing, the definitions of ζ, α_n, and η imply that

$$\left|\sum_{0<|j|\le k_n^*}[f]_j e_j\right| \le (\zeta\Gamma)^{1/2}\left(\frac{\gamma_{k_n^*}}{\omega_{k_n^*}}\alpha_n \sum_{0<|j|\le k_n^*}\frac{\omega_j}{\lambda_j n}\right)^{1/2} \le \left(\frac{\zeta\Gamma}{\eta}\right)^{1/2} \le 1, \tag{A.2}$$

as well as $\|f\|_\gamma^2 \le 1+\zeta\frac{\gamma_{k_n^*}}{\omega_{k_n^*}}\alpha_n(\sum_{0<|j|\le k_n^*}\frac{\omega_j}{n\lambda_j}) \le 1+\zeta/\eta \le r$.

It remains to show (A.1). Consider the Hellinger affinity $\rho(g_\theta^n, g_{\theta^{(j)}}^n) = \int\sqrt{g_\theta^n}\sqrt{g_{\theta^{(j)}}^n}$. We then obtain that, for any estimator $\widetilde{f}$ of f,

$$\begin{aligned}\rho\big(g_\theta^n, g_{\theta^{(j)}}^n\big) &\le \int\frac{|[\widetilde{f}-f_{\theta^{(j)}}]_j|}{|[f_\theta - f_{\theta^{(j)}}]_j|}\sqrt{g_{\theta^{(j)}}^n}\sqrt{g_\theta^n} + \int\frac{|[\widetilde{f}-f_\theta]_j|}{|[f_\theta - f_{\theta^{(j)}}]_j|}\sqrt{g_\theta^n}\sqrt{g_{\theta^{(j)}}^n}\\ &\le \left(\int\frac{|[\widetilde{f}-f_{\theta^{(j)}}]_j|^2}{|[f_\theta - f_{\theta^{(j)}}]_j|^2}g_{\theta^{(j)}}^n\right)^{1/2} + \left(\int\frac{|[\widetilde{f}-f_\theta]_j|^2}{|[f_\theta - f_{\theta^{(j)}}]_j|^2}g_\theta^n\right)^{1/2}.\end{aligned}$$

Rewriting the last estimate, we obtain

$$\big\{\mathbb{E}_\theta\big|[\widetilde{f}-f_\theta]_j\big|^2 + \mathbb{E}_{\theta^{(j)}}\big|[\widetilde{f}-f_{\theta^{(j)}}]_j\big|^2\big\} \ge \tfrac{1}{2}\big|[f_\theta - f_{\theta^{(j)}}]_j\big|^2\rho^2\big(g_\theta^n, g_{\theta^{(j)}}^n\big). \tag{A.3}$$

Next, we bound from below the Hellinger affinity $\rho(g_\theta^n, g_{\theta^{(j)}}^n)$. Therefore, we first consider the Hellinger distance,

$$\begin{aligned}H^2(g_\theta, g_{\theta^{(j)}}) &:= \int(\sqrt{g}_\theta - \sqrt{g}_{\theta^{(j)}})^2\\ &= \int\frac{|g_\theta - g_{\theta^{(j)}}|^2}{(\sqrt{g}_\theta + \sqrt{g}_{\theta^{(j)}})^2} \le 4\|g_\theta - g_{\theta^{(j)}}\|^2 = 16\big|[f]_j\big|^2\big|[\varphi]_j\big|^2 \le \frac{16\zeta d}{\eta n},\end{aligned}$$

where we have used that $\alpha_n \le 1/\eta$, $\varphi \in \mathcal{E}_\lambda^d$, and $g_\theta \ge 1/2$ because $|\sum_{0<|j|\le k_n^*}[g_\theta]_j e_j| \le 1/2$, which can be realized as follows. Using the condition $\sum_{j\in\mathbb{Z}}\gamma_j^{-1} = \Gamma < \infty$ and $\varphi \in \mathcal{E}_\lambda^d$, we obtain, in analogy to the proof of (A.2), that

$$\left|\sum_{0<|j|\le k_n^*}[g_\theta]_j e_j\right| \le \sum_{0<|j|\le k_n^*}\big|[f]_j\big|\big|[\varphi]_j\big| \le \left(\frac{\zeta\alpha_n d}{n}\right)^{1/2}\sum_{0<|j|\le k_n^*}\lambda_j^{-1/2} \le \left(\frac{\zeta d\Gamma}{\eta}\right)^{1/2} \le 1/2.$$

Therefore, the definition of ζ implies $H^2(g_\theta, g_{\theta^{(j)}}) \le 2/n$. Using the independence, that is, $\rho(g_\theta^n, g_{\theta^{(j)}}^n) = \rho(g_\theta, g_{\theta^{(j)}})^n$, together with the identity $\rho(g_\theta, g_{\theta^{(j)}}) = 1 - \frac{1}{2}H^2(g_\theta, g_{\theta^{(j)}})$, it follows $\rho(g_\theta^n, g_{\theta^{(j)}}^n) \ge (1-n^{-1})^n \ge 1/4$ for all $n \ge 2$. Combining the last estimate with (A.3), we obtain (A.1), which completes the proof. □

Proof of Theorem 2.3. We construct for each $\theta \in \{-1, 1\}$ an error density $\varphi_\theta \in \mathcal{E}_\lambda^d$ and a deconvolution density $f_\theta \in \mathcal{F}_\gamma^r$, such that $g_\theta := f_\theta * \varphi_\theta$ satisfies $g_1 = g_{-1}$. To be more precise, define $k_m^* := \operatorname{argmax}_{|j|>0}[\omega_j\gamma_j^{-1}\min(1, m^{-1}\lambda_j^{-1})]$ and $\alpha_m := \zeta\min(1, m^{-1/2}\lambda_{k_m^*}^{-1/2})$ with

$\zeta := \min(1/(2\sqrt{d}), (1-d^{-1/4}))$. Observe that $1 \geq (1-\alpha_m)^2 \geq (1-(1-1/d^{1/4}))^2 \geq 1/d^{1/2}$ and $1 \leq (1+\alpha_m)^2 \leq (1+(1-1/d^{1/4}))^2 = (2-1/d^{1/4})^2 \leq d^{1/2}$, which implies $1/d^{1/2} \leq (1+\theta\alpha_m)^2 \leq d^{1/2}$. We use these inequalities below without further reference. By assumption, there is a density $\varphi \in \mathcal{E}_\lambda^{\sqrt{d}}$ such that $\varphi \geq 1/2$. We show below that for each θ, the function $f_\theta := 1 + (1-\theta\alpha_m)\frac{\min(\sqrt{r-1},1)}{d^{1/4}}\gamma_{k_m^*}^{-1/2} e_{k_m^*}$ belongs to $\mathcal{F}_\gamma^r$, and that the function $\varphi_\theta := \varphi + \theta\alpha_m[\varphi]_{k_m^*} e_{k_m^*}$ is an element of $\mathcal{E}_\lambda^{\sqrt{d}}$. Moreover, it is easily verified that $g_\theta = 1 + (1-\alpha_m^2)\frac{\min(\sqrt{r-1},1)}{d^{1/4}}\gamma_{k_m^*}^{-1/2}[\varphi]_{k_m^*} e_{k_m^*}$, and thus $g_1 = g_{-1}$. We denote by g_θ^n the joint density of an i.i.d. n-sample from g_θ and by φ_θ^m the joint density of an i.i.d. m sample from φ_θ. Because the samples are independent of one another, $p_\theta := g_\theta^n \varphi_\theta^m$ is the joint density of all observations, and we denote by $\mathbb{E}_\theta$ the expectation with respect to p_θ. Applying a reduction scheme, we deduce that for each estimator $\widetilde{f}$ of f,

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\mathbb{E}\|\widetilde{f}-f\|_\omega^2 \geq \max_{\theta\in\{-1,1\}}\mathbb{E}_\theta\|\widetilde{f}-f_\theta\|_\omega^2 \geq \frac{1}{2}\big\{\mathbb{E}_1\|\widetilde{f}-f_1\|_\omega^2 + \mathbb{E}_{-1}\|\widetilde{f}-f_{-1}\|_\omega^2\big\}.$$

Below we also show that for all $m \geq 2$, we have

$$\mathbb{E}_1\|\widetilde{f}-f_1\|_\omega^2 + \mathbb{E}_{-1}\|\widetilde{f}-f_{-1}\|_\omega^2 \geq \tfrac{1}{8}\|f_1-f_{-1}\|_\omega^2. \tag{A.4}$$

Moreover, we have $\|f_1 - f_{-1}\|^2 = 4\alpha_m^2\omega_{k_m^*}\gamma_{k_m^*}^{-1}\frac{(r-1)\wedge 1}{d^{1/2}} = 4\frac{(r-1)\wedge 1}{d^{1/2}}\zeta^2\omega_{k_m^*}\gamma_{k_m^*}^{-1}\min(1, \frac{1}{m\lambda_{k_m^*}})$. Combining the last lower bound, the reduction scheme, and the definition of k_m^* implies the result of the theorem.

To conclude the proof, it remains to check (A.4), $f_\theta \in \mathcal{F}_\gamma^r$, and $\varphi_\theta \in \mathcal{E}_\lambda^d$ for both θ. To show $f_\theta \in \mathcal{F}_\gamma^r$, we first observe that f_θ integrates to 1. Moreover, f_θ is nonnegative, because $|(1-\theta\alpha_m)\frac{1\wedge\sqrt{r-1}}{d^{1/4}}\gamma_{k_m^*}^{-1/2}| \leq \gamma_{k_m^*}^{-1/2} \leq 1$ and $\|f_\theta\|_\gamma^2 = 1 + \gamma_{k_m^*}|[f_\theta]_{k_m^*}|^2 \leq 1 + \gamma_{k_n^*}|(1-\theta\alpha_m)\frac{1\wedge\sqrt{r-1}}{d^{1/4}}\gamma_{k_m^*}^{-1/2}|^2 \leq r$. Consider φ_θ, which obviously integrates to 1. Furthermore, as $\varphi \geq 1/2$, the function $\varphi_\theta = \varphi + \theta\alpha_m[\varphi]_{k_m^*} e_{k_m^*}$ is nonnegative, because $|\theta\alpha_m[\varphi]_{k_m^*} e_{k_m^*}| \leq \alpha_m\lambda_{k_m^*}^{1/2} d^{1/2} \leq \zeta m^{-1/2}\sqrt{d} \leq 1/2$ by using the definition of α_m and ζ. To check that $\varphi_\theta \in \mathcal{E}_\lambda^d$, it remains to show that $1/d \leq [\varphi_\theta]_j^2/\lambda_j \leq d$ for all $|j| > 0$. Because $\varphi \in \mathcal{E}_\lambda^{\sqrt{d}}$, it follows from the definition of φ_θ that these inequalities are satisfied for all $j \neq k_m^*$, and, moreover, that $1/d \leq \frac{|[\varphi]_{k_m^*}|^2}{\sqrt{d}\lambda_{k_m^*}} \leq \frac{(1+\theta\alpha_m)^2|[\varphi]_{k_m^*}|^2}{\lambda_{k_m^*}} \leq \frac{\sqrt{d}|[\varphi]_{k_m^*}|^2}{\lambda_{k_m^*}} \leq d$. Finally, consider (A.4). As in the proof of Theorem 2.1, by using the Hellinger affinity $\rho(p_1, p_{-1})$, we obtain, for any estimator $\widetilde{f}$ of f, that

$$\big\{\mathbb{E}_1\|\widetilde{f}-f_1\|_\omega^2 + \mathbb{E}_{-1}\|\widetilde{f}-f_1\|_\omega^2\big\} \geq \tfrac{1}{2}\|f_1-f_{-1}\|_\omega^2\rho(p_1, p_{-1}).$$

Next, we bound from below the Hellinger affinity $\rho(p_1, p_{-1}) \geq 1/4$ for all $m \geq 2$, which proves (A.4). From the independence and the fact that $g_1 = g_{-1}$, it is readily seen that Hellinger affinity satisfies $\rho(p_1, p_{-1}) = \rho(g_1, g_{-1})^n\rho(\varphi_1, \varphi_{-1})^m = \rho(\varphi_1, \varphi_{-1})^m = (1 - \frac{1}{2}H^2(\varphi_1, \varphi_{-1}))^m$. Thus,

we conclude $\rho(p_1, p_{-1}) \geq (1 - 1/m)^m \geq 1/4$, for all $m \geq 2$, because

$$H^2(\varphi_1, \varphi_{-1}) \leq \int \frac{|\varphi_1 - \varphi_{-1}|^2}{\varphi_1 + \varphi_{-1}} = \int \frac{|\varphi_1 - \varphi_{-1}|^2}{\varphi} \leq 2 \int |\varphi_1 - \varphi_{-1}|^2$$
$$\leq 2 \int 4\alpha_m^2 \big|[\varphi]_{k_m^*}\big|^2 e_{k_m^*}^2 \leq 8d\alpha_m^2 \lambda_{k_m^*} = 8d\zeta^2 m^{-1} \leq 2m^{-1},$$

where we used that $\varphi \geq 1/2$ and the definition of α_m and ζ. This completes the proof. □

Upper bound

Proof of Theorem 2.5. We begin our proof with the observation that $\mathbb{V}\mathrm{ar}([\widehat{g}]_j) \leq 1/n$ and $\mathbb{V}\mathrm{ar}([\widehat{\varphi}]_j) \leq 1/m$ for all $j \in \mathbb{Z}$. Moreover, by applying Theorem 2.10 of Petrov [43], there exists a constant $C > 0$ such that $\mathbb{E}|[\widehat{\varphi}]_j - [\varphi]_j|^4 \leq C/m^2$ for all $j \in \mathbb{Z}$ and $m \in \mathbb{N}$. We use these results below without further reference. Now define $\widetilde{f} := 1 + \sum_{0<|j|\leq k_n^*} [f]_j \mathbb{1}\{|[\widehat{\varphi}]_j|^2 \geq 1/m\} e_j$ and decompose the risk into two terms,

$$\mathbb{E}\|\widehat{f}_{k_n^*} - f\|_\omega^2 \leq 2\mathbb{E}\|\widehat{f}_{k_n^*} - \widetilde{f}\|_\omega^2 + 2\mathbb{E}\|\widetilde{f} - f\|_\omega^2 =: A + B, \tag{A.5}$$

which we bound separately. First, consider A, which we decompose further,

$$\mathbb{E}\|\widehat{f}_{k_n^*} - \widetilde{f}\|_\omega^2 \leq 2 \sum_{0<|j|\leq k_n^*} \omega_j \mathbb{E}\left[\frac{|[\widehat{g}]_j - [g]_j|^2}{|[\widehat{\varphi}]_j|^2} \mathbb{1}\big\{\big|[\widehat{\varphi}]_j\big|^2 \geq 1/m\big\}\right]$$
$$+ 2 \sum_{0<|j|\leq k_n^*} \omega_j \big|[f]_j\big|^2 \mathbb{E}\left[\frac{|[\widehat{\varphi}]_j - [\varphi]_j|^2}{|[\widehat{\varphi}]_j|^2} \mathbb{1}\big\{\big|[\widehat{\varphi}]_j\big|^2 \geq 1/m\big\}\right] =: A_1 + A_2.$$

Using the elementary inequality $|[\varphi]_j/[\widehat{\varphi}]_j|^2 \leq 2|[\varphi]_j/[\widehat{\varphi}]_j - 1|^2 + 2$, the independence of $\widehat{\varphi}$ and $\widehat{g}$, and $\varphi \in \mathcal{E}_\lambda^d$, together with the definition of ψ_n given in (2.1), we obtain

$$A_1 \leq 4 \sum_{0<|j|\leq k_n^*} \omega_j \left\{\frac{m\,\mathbb{V}\mathrm{ar}([\widehat{g}]_j)\,\mathbb{V}\mathrm{ar}([\widehat{\varphi}]_j)}{|[\varphi]_j|^2} + \frac{\mathbb{V}\mathrm{ar}([\widehat{g}]_j)}{|[\varphi]_j|^2}\right\} \leq 8d \sum_{0<|j|\leq k_n^*} \frac{\omega_j}{n\lambda_j} \leq 8d\psi_n.$$

Moreover, we have $\mathbb{E}\frac{|[\widehat{\varphi}]_j - [\varphi]_j|^2}{|[\widehat{\varphi}]_j|^2} \mathbb{1}\{|[\widehat{\varphi}]_j|^2 \geq 1/m\} \leq \frac{2m\mathbb{E}|[\widehat{\varphi}]_j - [\varphi]_j|^4}{|[\varphi]_j|^2} + \frac{2\,\mathbb{V}\mathrm{ar}([\widehat{\varphi}]_j)}{|[\varphi]_j|^2} \leq \frac{2(C+1)}{m|[\varphi]_j|^2} \leq \frac{2(C+1)d}{m\lambda_j}$ and $\mathbb{E}\frac{|[\widehat{\varphi}]_j - [\varphi]_j|^2}{|[\widehat{\varphi}]_j|^2} \mathbb{1}\{|[\widehat{\varphi}]_j|^2 \geq 1/m\} \leq 1$, where we have again used the elementary inequality and $\varphi \in \mathcal{E}_\lambda^d$. Combining both bounds together with $f \in \mathcal{F}_\gamma^r$ and the definition of κ_m given in (2.2), we obtain

$$A_2 \leq 4(C+1)d \sum_{0<|j|\leq k_n^*} \omega_j \big|[f]_j\big|^2 \min\left(1, \frac{1}{m\lambda_j}\right) \leq 4(C+1)dr\kappa_m.$$

Now consider B, which we decompose further into

$$\mathbb{E}\|\widetilde{f}-f\|_\omega^2 = \sum_{0<|j|} \omega_j |[f]_j|^2 \big(1-\mathbb{1}\{0<|j|\le k_n^*\}\mathbb{1}\{|[\widehat{\varphi}]_j|^2 \ge 1/m\}\big)^2$$
$$= \sum_{|j|>k_n^*} \omega_j |[f]_j|^2 + \sum_{0<|j|\le k_n^*} \omega_j |[f]_j|^2 \mathbf{P}\big(|[\widehat{\varphi}]_j|^2 < 1/m\big) =: B_1 + B_2,$$

where $B_1 \le \|f\|_\gamma^2 \omega_{k_n^*}\gamma_{k_n^*}^{-1} \le r\psi_n$, because $f \in \mathcal{F}_\gamma^r$. Moreover, $B_2 \le 4dr\kappa_m$, using that

$$\mathbf{P}\big(|[\widehat{\varphi}]_j|^2 < 1/m\big) \le 4d \min\bigg(1, \frac{1}{m\lambda_j}\bigg), \tag{A.6}$$

which we show below. The result of the theorem now follows by combining the decomposition (A.5) and the estimates of A_1, A_2, B_1, and B_2.

To conclude, we prove (A.6). If $|[\varphi]_j|^2 \ge 4/m$, then, using Tchebychev's inequality, we deduce that

$$\mathbf{P}\big(|[\widehat{\varphi}]_j|^2 < 1/m\big) \le \mathbf{P}\big(|[\widehat{\varphi}]_j/[\varphi]_j| < 1/2\big) \le \mathbf{P}\big(|[\widehat{\varphi}]_j - [\varphi]_j| > |[\varphi]_j|/2\big)$$
$$\le 4\frac{\mathbb{V}\mathrm{ar}([\widehat{\varphi}]_j)}{|[\varphi]_j|^2} \le 4d/(m\lambda_j).$$

On the other hand, in the case where $|[\varphi]_j|^2 < 4/m$, the estimate $\mathbf{P}(|[\widehat{\varphi}]_j|^2 < 1/m) \le 4d/(m\lambda_j)$ also holds, because $1 \le 4/(m|[\varphi]_j|^2) \le 4d/(m\lambda_j)$. Combining the last estimates and $\mathbf{P}(|[\widehat{\varphi}]_j|^2 < 1/m) \le 1$, we obtain (A.6), which completes the proof. □

Illustration: *Estimation of derivatives*

Proof of Proposition 2.6. Because for each $0 \le s \le p$, we have $\mathbb{E}\|\widetilde{f}^{(s)} - f^{(s)}\|^2 \sim \mathbb{E}\|\widetilde{f}-f\|_\omega^2$, we intend to apply the general result given in Corollary 2.4. In both cases, the additional conditions formulated in Theorem 2.1 and 2.3 are readily verified. Thus, it is sufficient to evaluate the lower bounds ψ_n and κ_m given in (2.1) and (2.2), respectively. Note that the optimal dimension parameter, $k_n^* := \mathrm{argmin}_{j\in\mathbb{N}}[\max(\frac{\omega_j}{\gamma_j}, \sum_{0<|l|\le j}\frac{\omega_l}{n\lambda_l})]$ satisfies $n\omega_{k_n^*}/\gamma_{k_n^*} \sim \sum_{0<|l|\le k_n^*}\omega_l/\lambda_l$, because both sequences (γ_j/ω_j) and $(\sum_{0<|l|\le j}\frac{\omega_l}{n\lambda_l})$ are non-increasing.

[o-o] The well-known approximation $\sum_{j=1}^m j^r \sim m^{r+1}$ for $r>0$ implies that $(\gamma_{k_n^*}/\omega_{k_n^*}) \times \sum_{0<|l|\le k_n^*}\omega_l/\lambda_l \sim (k_n^*)^{2a+2p+1}$. It follows that $k_n^* \sim n^{1/(2p+2a+1)}$, and the first lower bound is $\psi_n \sim n^{-(2p-2s)/(2p+2a+1)}$. Moreover, we have $\kappa_m \sim m^{-([p-s]\wedge a)/a}$, because the minimum in $\kappa_m = \sup_{j\in\mathbb{Z}}\{|j|^{-2(p-s)}\min(1, |j|^{2a}/m)\}$ is equal to 1 for $|j| \ge m^{1/2a}$ and $|j|^{-2(p-s)}$ is non-increasing.

[s-o] Approximating the sum in the same way as above, we obtain $(\gamma_{k_n^*}/\omega_{k_n^*})\sum_{0<|l|\le k_n^*}\omega_l/\lambda_l \sim (k_n^*)^{2a+1}\exp(k_n^{*2p})$, and thus $k_n^* \sim (\log n)^{1/(2p)}$. The resulting rate is $\psi_n \sim n^{-1} \times (\log n)^{(2a+2s+1)/(2p)}$. Furthermore, we have $\kappa_m \sim m^{-1}$, because the supremum is taken over $j^{2s}\exp(-j^{2p})\min(1, j^{2a}/m)$, which takes its maximum at the border because of the dominating exponential term.

[o-s] Applying Laplace's method (see chap. 3.7 in Olver [41]), we have $(\gamma_{k_n^*}/\omega_{k_n^*}) \times \sum_{0<|l|\le k_n^*} \omega_l/\lambda_l \sim (k_n^*)^{2p+((2a-1)\vee 0)} \exp(|k_n^*|^{2a})$, which implies that $k_n^* \sim (\log n)^{1/(2a)}$ and that the first lower bound can be rewritten as $\psi_n \sim (\log n)^{-(p-s)/a}$. Furthermore, we have $\kappa_m \sim (\log m)^{-(p-s)/a}$, because the minimum in $\kappa_m = \sup_{j\in\mathbb{Z}}\{|j|^{-2(p-s)} \min(1, \exp(|j|^{2a})/m)\}$ is equal to 1 for $|j| \ge (\log m)^{(1/2a)}$ and $|j|^{-2(p-s)}$ is non-increasing. Consequently, the lower bounds in Proposition 2.6 follow by applying Corollary 2.4. □

Proof of Proposition 2.8. The result is an immediate consequence of Theorem 2.5 and Proposition 2.6. □

A.2. Proofs of Section 3

Partially adaptive estimation

We begin by defining and recalling notations to be used in the proof. Given $u \in L^2[0,1]$, we denote by $[u]$ the infinite vector of Fourier coefficients $[u]_j := \langle u, e_j\rangle$. In particular, we use the notations

$$\widehat{f}_k = \sum_{j=-k}^{k} \frac{\widehat{[g]}_j}{\widehat{[\varphi]}_j} \mathbb{1}\big\{\big|\widehat{[\varphi]}_j\big|^2 \ge 1/m\big\} e_j, \qquad \widetilde{f}_k := \sum_{j=-k}^{k} \frac{\widehat{[g]}_j}{[\varphi]_j} e_j, \qquad f_k := \sum_{j=-k}^{k} \frac{[g]_j}{[\varphi]_j} e_j,$$

$$\widehat{\Phi}_u := \sum_{j\in\mathbb{Z}} \frac{[u]_j}{\widehat{[\varphi]}_j} \mathbb{1}\big\{\big|\widehat{[\varphi]}_j\big|^2 \ge 1/m\big\} e_j, \qquad \widetilde{\Phi}_u := \sum_{j\in\mathbb{Z}} \frac{[u]_j}{[\varphi]_j} e_j.$$

Furthermore, let $\widehat{g}$ be the function with Fourier coefficients $[\widehat{g}]_j := \widehat{[g]}_j$. Given $0 \le k \le k'$, we then have, for all $t \in \mathcal{S}_k := \operatorname{span}\{e_{-k}, \dots, e_k\}$,

$$\langle t, f_{k'}\rangle_\omega = \langle t, \widetilde{\Phi}_g\rangle_\omega = \sum_{j=-k}^{k} \frac{\omega_j [t]_j [g]_j}{[\varphi]_j} = \sum_{j=-k}^{k} \omega_j [t]_j [f]_j = \langle t, f\rangle_\omega,$$

$$\langle t, \widetilde{f}_{k'}\rangle_\omega = \langle t, \widetilde{\Phi}_{\widehat{g}}\rangle_\omega = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=-k}^{k} e_j(-Y_i) \frac{\omega_j [t]_j}{[\varphi]_j} = \langle t, \widetilde{f}_k\rangle_\omega,$$

$$\langle t, \widehat{f}_{k'}\rangle_\omega = \langle t, \widehat{\Phi}_{\widehat{g}}\rangle_\omega = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=-k}^{k} e_j(-Y_i) \frac{\omega_j [t]_j}{\widehat{[\varphi]}_j} \mathbb{1}\big\{\big|\widehat{[\varphi]}_j\big|^2 \ge 1/m\big\} = \langle t, \widehat{f}_k\rangle_\omega.$$

Consider the function $\nu = \widehat{g} - g$ with Fourier coefficients $[\nu]_j = \widehat{[g]}_j - [g]_j = \widehat{[g]}_j - \mathbb{E}\widehat{[g]}_j$. We then have, for every $t \in \mathcal{S}_k$,

$$\begin{aligned}\langle t, \widehat{\Phi}_{\widehat{g}} - f\rangle_\omega &= \langle t, \widehat{\Phi}_{\widehat{g}} - \widetilde{\Phi}_g\rangle_\omega = \langle t, \widetilde{\Phi}_{\widehat{g}} - \widetilde{\Phi}_g\rangle_\omega + \langle t, \widehat{\Phi}_{\widehat{g}} - \widetilde{\Phi}_{\widehat{g}}\rangle_\omega \\ &= \langle t, \widetilde{\Phi}_\nu\rangle_\omega + \langle t, \widehat{\Phi}_{\widehat{g}} - \widetilde{\Phi}_{\widehat{g}}\rangle_\omega = \langle t, \widetilde{\Phi}_\nu\rangle_\omega + \langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu\rangle_\omega + \langle t, \widehat{\Phi}_g - \widetilde{\Phi}_g\rangle_\omega.\end{aligned} \tag{A.7}$$

At the end of this section, we prove some technical lemmas that are used in the following proof.

Proof of Theorem 3.3. We consider the contrast

$$\Upsilon(t) := \|t\|_\omega^2 - 2\langle t, \widehat{\Phi}_{\widehat{g}}\rangle_\omega \qquad \forall t \in L^2[0,1].$$

It obviously follows that, for all $t \in \mathcal{S}_k$, $\Upsilon(t) = \|t - \widehat{f}_k\|_\omega^2 - \|\widehat{f}_k\|_\omega^2$, and thus,

$$\arg\min_{t\in\mathcal{S}_k} \Upsilon(t) = \widehat{f}_k \qquad \forall k \geq 0. \tag{A.8}$$

Moreover, the adaptive choice of the dimension parameter from (3.1) can be rewritten as

$$\widetilde{k} = \operatorname*{argmin}_{0\leq k\leq (N_n\wedge M_m)} \left[\Upsilon(\widehat{f}_k) + 60\frac{\delta_k}{n}\right]. \tag{A.9}$$

Let $\mathrm{pen}(k) := 60\delta_k/n$; then, for all $1 \leq k \leq (N_n \wedge M_m)$, we have

$$\Upsilon(\widehat{f}_{\widetilde{k}}) + \mathrm{pen}(\widetilde{k}) \leq \Upsilon(\widehat{f}_k) + \mathrm{pen}(k) \leq \Upsilon(f_k) + \mathrm{pen}(k),$$

using first (A.9) and then (A.8). This inequality implies that

$$\|\widehat{f}_{\widetilde{k}}\|_\omega^2 - \|f_k\|_\omega^2 \leq 2\langle \widehat{f}_{\widetilde{k}} - f_k, \widehat{\Phi}_{\widehat{g}}\rangle_\omega + \mathrm{pen}(k) - \mathrm{pen}(\widetilde{k}),$$

and thus, using (A.7), we have, for all $1 \leq k \leq (N_n \wedge M_m)$,

$$\begin{aligned}\|\widehat{f}_{\widetilde{k}} - f\|_\omega^2 \leq{}& \|f - f_k\|_\omega^2 + \mathrm{pen}(k) - \mathrm{pen}(\widetilde{k})\\ &+ 2\langle \widehat{f}_{\widetilde{k}} - f_k, \widetilde{\Phi}_\nu\rangle_\omega + 2\langle \widehat{f}_{\widetilde{k}} - f_k, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu\rangle_\omega + 2\langle \widehat{f}_{\widetilde{k}} - f_k, \widehat{\Phi}_g - \widetilde{\Phi}_g\rangle_\omega.\end{aligned} \tag{A.10}$$

Consider the unit ball $\mathcal{B}_k := \{f \in \mathcal{S}_k\colon \|f\|_\omega \leq 1\}$ and, for arbitrary $\tau > 0$ and $t \in \mathcal{S}_k$, the elementary inequality

$$2\big|\langle t, h\rangle_\omega\big| \leq 2\|t\|_\omega \sup_{t\in\mathcal{B}_k}\big|\langle t, h\rangle_\omega\big| \leq \tau\|t\|_\omega^2 + \frac{1}{\tau}\sup_{t\in\mathcal{B}_k}\big|\langle t, h\rangle_\omega\big|^2 = \tau\|t\|_\omega^2 + \frac{1}{\tau}\sum_{j=-k}^{k}\omega_j\big|[h]_j\big|^2.$$

Combining the last estimate with (A.10) and $\widehat{f}_{\widetilde{k}} - f_k \in \mathcal{S}_{\widetilde{k}\vee k} \subset \mathcal{S}_{N_n\wedge M_m}$, we obtain

$$\begin{aligned}\|\widehat{f}_{\widetilde{k}} - f\|_\omega^2 \leq{}& \|f - f_k\|_\omega^2 + 3\tau\|\widehat{f}_{\widetilde{k}} - f_k\|_\omega^2 + \mathrm{pen}(k) - \mathrm{pen}(\widetilde{k})\\ &+ \frac{1}{\tau}\sup_{t\in\mathcal{B}_{k\vee\widetilde{k}}}\big|\langle t, \widetilde{\Phi}_\nu\rangle_\omega\big|^2 + \frac{1}{\tau}\sup_{t\in\mathcal{B}_{(N_n\wedge M_m)}}\big|\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu\rangle_\omega\big|^2\\ &+ \frac{1}{\tau}\sup_{t\in\mathcal{B}_{(N_n\wedge M_m)}}\big|\langle t, \widehat{\Phi}_g - \widetilde{\Phi}_g\rangle_\omega\big|^2.\end{aligned}$$

Note that $\|\widehat{f}_{\widetilde{k}} - f_k\|_\omega^2 \le 2\|\widehat{f}_{\widetilde{k}} - f\|_\omega^2 + 2\|f_k - f\|_\omega^2$ and that $\|f - f_k\|_\omega^2 \le r\omega_k/\gamma_k$ for all $f \in \mathcal{F}_\gamma^r$, because ω/γ is non-increasing. Setting $\tau := 1/8$, we obtain

$$\begin{aligned}\frac{1}{4}\|\widehat{f}_{\widetilde{k}} - f\|_\omega^2 \le{}& \frac{7}{4}(r\omega_k/\gamma_k) + \operatorname{pen}(k) - \operatorname{pen}(\widetilde{k}) \\ &+ 8\sup_{t\in\mathcal{B}_{k\vee\widetilde{k}}} \big|\langle t, \widetilde{\Phi}_\nu\rangle_\omega\big|^2 + 8\sup_{t\in\mathcal{B}_{(N_n\wedge M_m)}} \big|\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu\rangle_\omega\big|^2 \\ &+ 8\sup_{t\in\mathcal{B}_{(N_n\wedge M_m)}} \big|\langle t, \widehat{\Phi}_g - \widetilde{\Phi}_g\rangle_\omega\big|^2.\end{aligned} \tag{A.11}$$

Defining the event

$$\Omega_q := \left\{\forall 0 \le |j| \le M_m^u\colon \left|\frac{1}{\widehat{[\varphi]}_j} - \frac{1}{[\varphi]_j}\right| \le \frac{1}{2|[\varphi]_j|} \wedge \big|\widehat{[\varphi]}_j\big|^2 \ge 1/m\right\}, \tag{A.12}$$

consider the following decomposition of the risk:

$$\mathbb{E}\|\widehat{f}_{\widetilde{k}} - f\|_\omega^2 = \mathbb{E}\|\widehat{f}_{\widetilde{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_q\} + \mathbb{E}\|\widehat{f}_{\widetilde{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_q^c\}. \tag{A.13}$$

We bound these two terms separately. Consider the first term. By Lemma A1 below and $\mathbb{1}\{|\widehat{[\varphi]}_j|^2 \ge 1/m\}\mathbb{1}\{\Omega_q\} = \mathbb{1}\{\Omega_q\}$, it follows that for all $1 \le |j| \le (N_n \wedge M_m)$,

$$\left(\frac{[\varphi]_j}{\widehat{[\varphi]}_j}\mathbb{1}\big\{\big|\widehat{[\varphi]}_j\big|^2 \ge 1/m\big\} - 1\right)^2 \mathbb{1}\{\Omega_q\} = \big|[\varphi]_j\big|^2 \mathbb{1}\{\Omega_q\}\left|\frac{1}{\widehat{[\varphi]}_j} - \frac{1}{[\varphi]_j}\right|^2 \le \frac{1}{4}.$$

Thus, $\sup_{t\in\mathcal{B}_k} |\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu\rangle_\omega|^2 \mathbb{1}\{\Omega_q\} \le \frac{1}{4}\sup_{t\in\mathcal{B}_k} |\langle t, \widetilde{\Phi}_\nu\rangle_\omega|^2$ for all $0 \le k \le (N_n \wedge M_m)$, and (A.11) implies that

$$\begin{aligned}\frac{1}{4}\|\widehat{f}_{\widetilde{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_q\} \le{}& \frac{7}{4}(r\omega_k/\gamma_k) + 10\Big(\sup_{t\in\mathcal{B}_{k\vee\widetilde{k}}} \big|\langle t, \widetilde{\Phi}_\nu\rangle_\omega\big|^2 - (6\delta_{k\vee\widetilde{k}})/n\Big)_+ \\ &+ (60\delta_{k\vee\widetilde{k}})/n + \operatorname{pen}(k) - \operatorname{pen}(\widetilde{k}) + 8\sup_{t\in\mathcal{B}_{(N_n\wedge M_m)}} \big|\langle t, \widehat{\Phi}_g - \widetilde{\Phi}_g\rangle_\omega\big|^2.\end{aligned} \tag{A.14}$$

Moreover, we have that $60\delta_{k\vee\widetilde{k}}/n = \operatorname{pen}(k\vee\widetilde{k}) \le \operatorname{pen}(k) + \operatorname{pen}(\widetilde{k})$. Further note that

$$\Delta_k \le d\Delta_k^\lambda, \qquad \delta_k \le d\zeta_d\delta_k^\lambda, \quad \text{and} \quad \delta_k/\Delta_k \ge 2k\zeta_d^{-1}\frac{\log(\Delta_k^\lambda \vee (k+2))}{\log(k+2)}, \tag{A.15}$$

with $\zeta_d = \log(3d)/\log d$. From Lemma A1, it follows that

$$\begin{aligned}&\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d} \mathbb{E}\|\widehat{f}_{\widetilde{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_q\} \\ &\quad\le 480(r + d\zeta_d)\min_{0\le k\le N_n^\lambda\wedge M_m^\lambda}\big[\max\big(\omega_k/\gamma_k, \delta_k^\lambda/n\big)\big]\end{aligned}$$

$$+40 \sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\sum_{0\le k'\le(N_n^u\wedge M_m^u)}\mathbb{E}\Big(\sup_{t\in\mathcal{B}_{k'}}\big|\langle t,\widetilde{\Phi}_v\rangle_\omega\big|^2-(6\delta_{k'})/n\Big)_+$$

$$+32 \sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\mathbb{E}\Big[\sup_{t\in\mathcal{B}_{(N_n^u\wedge M_m^u)}}\big|\langle t,\widehat{\Phi}_g-\widetilde{\Phi}_g\rangle_\omega\big|^2\Big].$$

To bound the second term, we apply Lemma A4 with $\delta_k^*=\delta_k$ and $\Delta_k^*=\Delta_k$. By virtue of (A.15), we have, for all $k\ge 0$,

$$\mathbb{E}\Big(\sup_{t\in\mathcal{B}_k}\big|\langle t,\widetilde{\Phi}_v\rangle_\omega\big|^2-6\frac{\delta_k}{n}\Big)_+$$
$$\le C\bigg\{\frac{1}{n^2}\exp\big(-K_2\sqrt{n}\big)d\zeta_d\delta_k^\lambda$$
$$+\frac{\|\varphi\|^2\|f\|^2}{n}d\Delta_k^\lambda\exp\bigg(-\frac{k}{3\|\varphi\|^2\|f\|^2\zeta_d}\frac{\log(\Delta_k^\lambda\vee(k+2))}{\log(k+2)}\bigg)\bigg\}.$$

Owing to Lemmas A1 and A2(i) and the properties of the function Σ from Definition 3.1, we have

$$\sum_{k=0}^{N_n^u}\mathbb{E}\Big(\sup_{t\in\mathcal{B}_k}\big|\langle t,\widetilde{\Phi}_v\rangle_\omega\big|^2-6\frac{\delta_k}{n}\Big)_+\le\frac{C}{n}d\Sigma\big(\|\varphi\|^2\|f\|^2\zeta_d\big).$$

It can be readily verified that $\|\varphi\|^2\le d\Lambda$ for all $\varphi\in\mathcal{E}_\lambda^d$ and $\|f\|^2\le r$ for all $f\in\mathcal{F}_\gamma^r$. The remaining term can be controlled by virtue of Lemma A5, which shows that

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\mathbb{E}\|\widehat{f}_{\widetilde{k}}-f\|_\omega^2\mathbb{1}\{\Omega_q\}\le C\Big\{(r+d\zeta_d)\min_{0\le k\le(N_n^\lambda\wedge M_m^\lambda)}\big[\max\big(\omega_k/\gamma_k,\delta_k^\lambda/n\big)\big]$$
$$+rd\kappa_m+d\Sigma(rd\Lambda\zeta_d)n^{-1}\Big\}. \tag{A.16}$$

Consider the second term from (A.13). Let $\check{f}_k:=1+\sum_{0<|j|\le k}[f]_j\mathbb{1}\{|\widehat{[\varphi]}_j|^2\ge 1/m\}e_j$. It is easy to see that $\|\widehat{f}_k-\check{f}_k\|^2\le\|\widehat{f}_{k'}-\check{f}_{k'}\|^2$ for all $k\le k'$ and $\|\check{f}_k-f\|^2\le\|f\|^2$ for all $k\ge 0$. Thus, using that $0\le\widetilde{k}\le(N_n^\circ\wedge m)$, we can write

$$\mathbb{E}\|\widehat{f}_{\widetilde{k}}-f\|_\omega^2\mathbb{1}\big\{\Omega_q^c\big\}\le 2\big\{\mathbb{E}\|\widehat{f}_{\widetilde{k}}-\check{f}_{\widetilde{k}}\|_\omega^2\mathbb{1}\big\{\Omega_q^c\big\}+\mathbb{E}\|\check{f}_{\widetilde{k}}-f\|_\omega^2\mathbb{1}\big\{\Omega_q^c\big\}\big\}$$
$$\le 2\big\{\mathbb{E}\|\widehat{f}_{(N_n^\circ\wedge m)}-\check{f}_{(N_n^\circ\wedge m)}\|_\omega^2\mathbb{1}\big\{\Omega_q^c\big\}+\|f\|_\omega^2\mathbf{P}\big[\Omega_q^c\big]\big\}.$$

Moreover, applying Theorem 2.10 of Petrov [43],

$$\mathbb{E}\|\widehat{f}_{(N_n^\circ\wedge m)}-\check{f}_{(N_n^\circ\wedge m)}\|_\omega^2\mathbb{1}\big\{\Omega_q^c\big\}$$
$$\le 2m\sum_{0<|j|\le(N_n^\circ\wedge m)}\omega_j\big\{\mathbb{E}\big(\widehat{[g]}_j-[\varphi]_j[f]_j\big)^2\mathbb{1}\big\{\Omega_q^c\big\}+\mathbb{E}\big([\varphi]_j[f]_j-\widehat{[\varphi]}_j[f]_j\big)^2\mathbb{1}\big\{\Omega_q^c\big\}\big\}$$

$$\begin{aligned}
&\le 2m\Bigg\{\sum_{0<|j|\le(N_n^\circ\wedge m)}\omega_j\big[\mathbb{E}\big(\widehat{[g]}_j-[g]_j\big)^4\big]^{1/2}\mathbf{P}\big[\Omega_q^c\big]^{1/2}\\
&\qquad+\sum_{0<|j|\le(N_n^\circ\wedge m)}\omega_j\big|[f]_j\big|^2\big[\mathbb{E}\big(\widehat{[\varphi]}_j-[\varphi]_j\big)^4\big]^{1/2}\mathbf{P}\big[\Omega_q^c\big]^{1/2}\Bigg\}\\
&\le 2m\Big\{2m\Big(\max_{1\le j\le N_n^\circ}\omega_j\Big)\big(Cn^{-1}\big)+\big(Cm^{-1}\big)\|f\|_\omega^2\Big\}\mathbf{P}\big[\Omega_q^c\big]^{1/2},
\end{aligned}$$

which implies, using Definition 3.1(ii),

$$\begin{aligned}
\mathbb{E}\|\widehat{f_{\widehat{k}}}-f\|_\omega^2\mathbb{1}\big\{\Omega_q^c\big\}&\le 4C\big(m^2+\|f\|_\omega^2\big)\mathbf{P}\big[\Omega_q^c\big]^{1/2}+2\|f\|_\omega^2\mathbf{P}\big[\Omega_q^c\big]\\
&\le 6Cm^2\big(1+\|f\|_\omega^2\big)\mathbf{P}\big[\Omega_q^c\big]^{1/2}.
\end{aligned}\tag{A.17}$$

By Lemma A6, it follows that for all $m\in\mathbb{N}$,

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\mathbb{E}\|\widehat{f_{\widehat{k}}}-f\|_\omega^2\mathbb{1}\big\{\Omega_p^c\big\}\le C(d)(1+r)m^{-1}.\tag{A.18}$$

The result of the theorem follows from a combination of the last estimate and (A.16). □

Lemma A1. *Under Assumption* A1, *we have, for all* $n,m\in\mathbb{N}$,

$$N_n^\lambda\le N_n\le N_n^u\quad\textit{and}\quad M_m^\lambda\le M_m\le M_m^u.$$

Proof. We first prove that $N_n^\lambda\le N_n$. If $N_n^\lambda=0$ or $N_n=N_n^\circ$, then there is nothing to show. Noting that

$$N_n^\lambda=0\iff\max_{1\le j\le N_n^\circ}\frac{\lambda_j}{j\omega_j^+}<\frac{4d\log n}{n}\quad\text{and}\quad N_n=0\iff\max_{1\le j\le N_n^\circ}\frac{\lambda_j}{j\omega_j^+}<\frac{d\log n}{n},$$

we deduce that in the case where $N_n=0$, we also have $N_n^\lambda=0$. This also holds when $N_n^\lambda>0$ and $N_n^\circ>N_n>0$, which implies

$$\min_{1\le j\le N_n^\lambda}\frac{\lambda_j}{j\omega_j^+}\ge\frac{4d\log n}{n}\quad\text{and}\quad\frac{\log n}{n}>\frac{|[\varphi]_{N_n+1}|^2}{N_n\omega_{N_n+1}}\ge\frac{\lambda_{N_n+1}}{dN_n\omega_{N_n+1}^+}$$

and thus $N_n+1>N_n^\lambda$, which proves the claim.

We now prove $N_n\le N_n^u$. If $N_n=0$ or $N_n^u=n$, then this is trivial. On the other hand, if $n>N_n^u\ge0$ and $N_n^\circ\ge N_n>0$, then it follows from the definitions that

$$\min_{1\le j\le N_n}\frac{d\lambda_j}{j\omega_j^+}\ge\min_{1\le j\le N_n}\frac{|[\varphi]_j|^2}{j\omega_j^+}\ge\frac{\log n}{n}\quad\text{and}\quad\frac{\lambda_{N_n^\circ+1}}{(N_n^\circ+1)\omega_{N_n^\circ+1}^+}<\frac{\log n}{4dn},$$

which implies that $N_n^\circ+1>N_n$, and hence the claim. Similar arguments show the corresponding estimates in m. □

Lemma A2. *Under Assumption A1, we have that*

(i) $\delta_{N_n^u}/n \le 32d^2$ *for all* $n \ge 1$,

(ii) $m^7 \exp(-\frac{m\lambda_{M_m^u}}{72d}) \le C(d)$ *for all* $m \ge 1$,

and for $m \ge \exp(512\log(3d)^2)$ *that*

(iii) $\min_{1\le j\le M_m^u} |[\varphi]_j|^2 \ge \frac{2}{m}$.

Proof. (i) For $N_n^u = 0$, we have $\delta_{N_n^u} = 0$, and there is nothing to show. If $0 < N_n^u \le n$, then we can show that $\omega^+_{N_n^u}/\lambda_{N_n^u} \le 4dn/(N_n^u \log(n+2))$, which we use in the following computation:

$$\begin{aligned}
\delta_{N_n^u} &= N_n^u \frac{\omega^+_{N_n^u}}{\lambda_{N_n^u}} \frac{\log((\omega^+_{N_n^u}/\lambda_{N_n^u}) \vee (N_n^u+2))}{\log(N_n^u+2)} \\
&\le \frac{4dn}{\log(n+2)} \log\biggl(\frac{4dn}{N_n^u \log(n+2)} \vee \bigl(N_n^u+2\bigr)\biggr) \Big/ \log\bigl(N_n^u+2\bigr) \\
&\le n \begin{cases} 4d & \bigl(\log(n+2) \ge 4d\bigr), \\ 4d\bigl(4d+\log(4d)\bigr)/\bigl(\log(n+2)\bigr) & (\text{otherwise}), \end{cases}
\end{aligned}$$

which implies that $\delta_{N_n^u}/n \le 4d(4d+\log(4d)) \le 32d^2$ for all $n \ge 1$.

(ii) For $0 < M_m^u \le m$, we have $\lambda_{M_m^u} \ge m^{-1+b_m}(4d)^{-1}$. Thus,

$$m^7 \exp\biggl(-\frac{m\lambda_{M_m^u}}{72d}\biggr) \le \exp\biggl(-\frac{m^{b_m}}{288d^2} + 7\log m\biggr).$$

This proves the claim, because $\log m \lesssim m^{b_m}$. Note that $M_m^u = 0$ cannot occur, because we assume that $\lambda_1 = 1$.

(iii) We have that

$$\min_{1\le j\le M_m^u} \bigl|[\varphi]_j\bigr|^2 \ge \min_{1\le j\le M_m^u} \frac{\lambda_j}{d} \ge \frac{m^{b_m}}{4d^2 m} \ge \frac{2}{m},$$

where the last step holds for $m \ge \exp(128\log(8d^2)^2)$, as shown by some algebra. □

For the proof of Lemma A4 below, we need the following lemma, which can be found in Talagrand [46].

Lemma A3 (Talagrand's inequality). *Let* $T_1, \dots, T_n$ *be independent random variables, and let* $\nu_n^*(r) = (1/n)\sum_{i=1}^n [r(T_i) - \mathbb{E}[r(T_i)]]$, *for* r *belonging to a countable class* $\mathcal{R}$ *of measurable functions. Then,*

$$\mathbb{E}\Bigl[\sup_{r\in\mathcal{R}} \bigl|\nu_n^*(r)\bigr|^2 - 6H_2^2\Bigr]_+ \le C\biggl(\frac{v}{n}\exp\bigl(-\bigl(nH_2^2/6v\bigr)\bigr) + \frac{H_1^2}{n^2}\exp\bigl(-K_2(nH_2/H_1)\bigr)\biggr)$$

with numerical constants $K_2 = (\sqrt{2}-1)/(21\sqrt{2})$ *and* $C > 0$ *and with*

$$\sup_{r\in\mathcal{R}} \|r\|_\infty \le H_1, \qquad \mathbb{E}\Big[\sup_{r\in\mathcal{R}} \big|\nu_n^*(r)\big|\Big] \le H_2, \qquad \sup_{r\in\mathcal{R}} \frac{1}{n}\sum_{i=1}^n \mathbb{V}\mathrm{ar}\big(r(T_i)\big) \le v.$$

Lemma A4. *Let* δ^* *and* Δ^* *be sequences such that for all* $k \ge 1$,

$$\delta_k^* \ge \sum_{-k\le j\le k} \frac{\omega_j}{|[\varphi]_j|^2} \quad \textit{and} \quad \Delta_k^* \ge \max_{0\le|j|\le k} \frac{\omega_j}{|[\varphi]_j|^2}$$

and let $K_2 := (\sqrt{2}-1)/(21\sqrt{2})$. *Then, for all* $n, k \ge 1$,

$$\mathbb{E}\bigg[\bigg(\sup_{t\in\mathcal{B}_k} \big|\langle t, \widetilde{\Phi}_v\rangle_\omega\big|^2 - \frac{6\delta_k^*}{n}\bigg)_+\bigg]$$
$$\le C\bigg\{\frac{\|\varphi\|^2\|f\|^2}{n}\Delta_k^* \exp\bigg(-\frac{1}{6\|\varphi\|^2\|f\|^2}\big(\delta_k^*/\Delta_k^*\big)\bigg) + \frac{1}{n^2}\exp(-K_2\sqrt{n})\delta_k^*\bigg\}.$$

Proof. For $t \in \mathcal{S}_k$, define the function $r_t := \sum_{k\le j\le k} \omega_j [t]_j \overline{[\varphi]}_j^{-1} e_j$. Then it is readily seen that $\langle t, \widetilde{\Phi}_v\rangle_\omega = \frac{1}{n}\sum_{k=1}^n r_t(Y_k) - \mathbb{E}[r_t(Y_k)]$. We next compute constants H_1, H_2, and v verifying the three inequalities required in Lemma A3, which then implies the result.

First, consider H_1:

$$\sup_{t\in\mathcal{B}_k} \|r_t\|_\infty^2 = \sup_{y\in\mathbb{R}} \sum_{-k\le j\le k} \omega_j \big|[\varphi]_j\big|^{-2} \big|e_j(y)\big|^2 = \sum_{-k\le j\le k} \omega_j \big|[\varphi]_j\big|^{-2} \le \delta_k^* =: H_1^2.$$

Next, find H_2. Note that

$$\mathbb{E}\Big[\sup_{t\in\mathcal{B}_k} \big|\langle t, \widetilde{\Phi}_v\rangle_\omega\big|^2\Big] = \frac{1}{n}\sum_{-k\le j\le k} \omega_j \big|[\varphi]_j\big|^{-2} \mathbb{V}\mathrm{ar}\big(e_j(Y_1)\big).$$

Because $\mathbb{V}\mathrm{ar}(e_j(Y_1)) \le \mathbb{E}[|e_j(Y_1)|^2] = 1$, we define $\mathbb{E}[\sup_{t\in\mathcal{B}_k} |\langle t, \widetilde{\Phi}_v\rangle|^2] \le \delta_k^*/n =: H_2^2$.

Finally, consider v. Given $t \in \mathcal{B}_k$ and a sequence $(z_j)_{j\in\mathbb{Z}}$, let $\underline{[t]} := ([t]_{-k}, \dots, [t]_k)^T$ and denote by $D_k(z) := \mathrm{diag}[z_{-k}, \dots, z_k]$ the corresponding diagonal matrix. Define the Hermitian and positive semi-definite matrix $A_k := (\overline{[\varphi]}_j^{-1}[\varphi]_{j'}^{-1}[\varphi]_{j-j'}[f]_{j-j'})_{j,j'=-k,\dots,k}$. Straightforward algebra shows that $\sup_{t\in\mathcal{B}_k} \mathbb{V}\mathrm{ar}(r_t(Y_1)) \le \sup_{t\in\mathcal{B}_k} \langle A_k D_k(\omega)\underline{[t]}, D_k(\omega)\underline{[t]}\rangle_{\mathbb{C}^{2k+1}}$; thus,

$$\sup_{t\in\mathcal{B}_k} \frac{1}{n}\sum_{k=1}^n \mathbb{V}\mathrm{ar}\big(r_t(Y_k)\big) \le \sup_{t\in\mathcal{B}_k} \big\langle A_k^{1/2} D_k(\omega)\underline{[t]}, A_k^{1/2} D_k(\omega)\underline{[t]}\big\rangle_{\mathbb{C}^{2k+1}}$$
$$= \sup_{t\in\mathcal{B}_k} \big\| A_k^{1/2} D_k(\omega)\underline{[t]}\big\|_{\mathbb{C}^{2k+1}}^2 = \big\| D_k(\sqrt{\omega}) A_k D_k(\sqrt{\omega})\big\|_{\mathbb{C}^{2k+1}}.$$

Clearly, we have $A_k = D_k([\varphi]^{-1})B_k D_k(\overline{[\varphi]}^{-1})$, where $B_k := ([\varphi]_{j-k}[f]_{j-k})_{j,k=-k,\dots,k}$. Consequently,

$$\sup_{t\in\mathcal{B}_k}\frac{1}{n}\sum_{k=1}^{n}\mathbb{V}\mathrm{ar}\big(r_t(Y_k)\big)\le\big\|D_k\big(\sqrt{\omega}[\varphi]^{-1}\big)\big\|^2_{\mathbb{C}^{2k+1}}\|B_k\|_{\mathbb{C}^{2k+1}}.$$

We have that $\|D_k(\sqrt{\omega}[\varphi]^{-1})\|^2_{\mathbb{C}^{2k+1}} = \max_{0\le|j|\le k}\omega_j|[\varphi]_j|^{-2}\le\Delta_k^*$. It remains to show the boundedness of $\|B_k\|_{\mathbb{C}^{2k+1}}$. Let ℓ^2 be the space of square-summable sequences in $\mathbb{C}$, and define the operator $B:\ell^2\to\ell^2$ by $(Bz)_k := \sum_{j\in\mathbb{Z}}[\varphi]_{j-k}[f]_{j-k}z_j$, $k\in\mathbb{Z}$. Then it is easily verified that for any $z\in\ell^2$ with $\|z\|_{\ell^2}=1$, the Cauchy–Schwarz inequality yields $\|Bz\|^2_{\ell^2}\le\|\varphi\|^2\|f\|^2$, and thus $\|B\|^2_{\ell^2}\le\|\varphi\|^2\|f\|^2$. Given the orthogonal projection Π_k in ℓ^2 onto $\mathcal{S}_k$, the operator $\Pi_k B\Pi_k:\mathcal{S}_k\to\mathcal{S}_k$ has matrix representation B_k via the isomorphism $\mathcal{S}_k\cong\mathbb{C}^{2k+1}$, and hence $\|\Pi_k B\Pi_k\|_{\ell^2}=\|B_k\|_{\mathbb{C}^{2k+1}}$. Given orthogonal projections with a norm bounded by 1, we conclude that $\|B_k\|_{\mathbb{C}^{2k+1}}\le\|B\|_{\ell^2}$ for all $k\in\mathbb{N}$, which implies that $\sup_{t\in\mathcal{B}_k}\frac{1}{n}\sum_{k=1}^{n}\mathbb{V}\mathrm{ar}(r_t(Y_k))\le\|\varphi\|^2\|f\|^2\Delta_k^*=:v$, which completes the proof. □

Lemma A5. *For every $m\ge1$ and $k\ge0$, we have*

$$\sup_{f\in\mathcal{F}_\gamma^r}\mathbb{E}\Big[\sup_{t\in\mathcal{B}_k}\big|\langle t,\widehat{\Phi}_g-\widetilde{\Phi}_g\rangle_\omega\big|^2\Big]\le Cr\max_{j\in\mathbb{N}}\bigg\{\frac{\omega_j}{\gamma_j}\min\bigg(1,\frac{1}{m[\varphi]_j^2}\bigg)\bigg\}\le Cdr\kappa_m(\gamma,\lambda,\omega).$$

Proof. First, given that $f\in\mathcal{F}_\gamma^r$, it can be easily seen that

$$\mathbb{E}\Big[\sup_{t\in\mathcal{B}_k}\big|\langle t,\widehat{\Phi}_g-\widetilde{\Phi}_g\rangle_\omega\big|^2\Big]\le r\sup_{-k\le j\le k}\frac{\omega_j}{\gamma_j}\mathbb{E}\big[|R_j|^2\big],$$

where R_j is as defined by

$$R_j := \bigg(\frac{[\varphi]_j}{\widehat{[\varphi]}_j}\mathbb{1}\big\{\big|\widehat{[\varphi]}_j\big|^2\ge1/m\big\}-1\bigg).$$

In view of the definition (2.2) of κ_m, the result follows from $\mathbb{E}[|R_j|^2]\le C\min\{1,\frac{1}{m|[\varphi]_j|^2}\}$, which can be realized as follows. Consider the identity

$$\mathbb{E}|R_j|^2=\mathbb{E}\bigg[\bigg|\frac{[\varphi]_j}{\widehat{[\varphi]}_j}-1\bigg|^2\mathbb{1}\big\{\big|\widehat{[\varphi]}_j\big|^2\ge1/m\big\}\bigg]+\mathbf{P}\big[\big|\widehat{[\varphi]}_j\big|^2<1/m\big]=:R_j^{\mathrm{I}}+R_j^{\mathrm{II}}.$$

Trivially, $R_j^{\mathrm{II}}\le1$. If $1\le4/(m|[\varphi]_j|^2)$, then obviously $R_j^{\mathrm{II}}\le4\min\{1,\frac{1}{m|[\varphi]_j|^2}\}$. Otherwise, we have $1/m<|[\varphi]_j|^2/4$ and thus, using Tchebychev's inequality,

$$R_j^{\mathrm{II}}\le\mathbf{P}\big[\big|\widehat{[\varphi]}_j-[\varphi]_j\big|>\big|[\varphi]_j\big|/2\big]\le\frac{4\,\mathbb{V}\mathrm{ar}(\widehat{[\varphi]}_j)}{|[\varphi]_j|^2}\le4\min\bigg\{1,\frac{1}{m|[\varphi]_j|^2}\bigg\},$$

where $\mathbb{V}\mathrm{ar}(\widehat{[\varphi]}_j) \le 1/m$ for all j. Now consider R_j^{I}. We find that

$$R_j^{\mathrm{I}} = \mathbb{E}\left[\frac{|\widehat{[\varphi]}_j - [\varphi]_j|^2}{|\widehat{[\varphi]}_j|^2}\mathbb{1}\{|\widehat{[\varphi]}_j|^2 \ge 1/m\}\right] \le m\,\mathbb{V}\mathrm{ar}(\widehat{[\varphi]}_j) \le 1. \tag{A.19}$$

On the other hand, using that $\mathbb{E}[|\widehat{[\varphi]}_j - [\varphi]_j|^4] \le C/m^2$ (cf. Petrov [43], Theorem 2.10), we obtain

$$\begin{aligned} R_j^{\mathrm{I}} &\le \mathbb{E}\left[\frac{|\widehat{[\varphi]}_j - [\varphi]_j|^2}{|\widehat{[\varphi]}_j|^2}\mathbb{1}\{|\widehat{[\varphi]}_j|^2 \ge 1/m\}2\left\{\frac{|\widehat{[\varphi]}_j - [\varphi]_j|^2}{|[\varphi]_j|^2} + \frac{|\widehat{[\varphi]}_j|^2}{|[\varphi]_j|^2}\right\}\right] \\ &\le \frac{2m\mathbb{E}[|\widehat{[\varphi]}_j - [\varphi]_j|^4]}{|[\varphi]_j|^2} + \frac{2\,\mathbb{V}\mathrm{ar}(\widehat{[\varphi]}_j)}{|[\varphi]_j|^2} \le \frac{2C}{m|[\varphi]_j|^2} + \frac{2}{m|[\varphi]_j|^2}. \end{aligned}$$

Combining this result with (A.19) gives $R_j^{\mathrm{I}} \le 2(C+1)\min\{1, \frac{1}{m|[\varphi]_j|^2}\}$, which completes the proof. □

Lemma A6. *Under Assumption* A1, $\mathbf{P}[\Omega_q^c] \le C(d)m^{-6}$ *for all* $m \ge 1$.

Proof. The estimate is obvious for $m < \exp(512\log(3d)^2) =: m_0$. Consider the complement of Ω_q given by

$$\Omega_q^c = \left\{\exists 0 < |j| \le M_m^u\colon \left|\frac{[\varphi]_j}{\widehat{[\varphi]}_j} - 1\right| > \frac{1}{2} \vee |\widehat{[\varphi]}_j|^2 < 1/m\right\}.$$

Because of Lemma A2(iii), for all $m \ge m_0$ and $0 < |j| \le M_m^u$, we have $|[\varphi]_j|^2 \ge 2/m$. This yields

$$\Omega_q^c \subseteq \left\{\exists 0 < |j| \le M_m^u\colon \left|\frac{\widehat{[\varphi]}_j}{[\varphi]_j} - 1\right| > \frac{1}{3}\right\}.$$

By Hoeffding's inequality, for all $0 < |j| \le M_m^u$,

$$\mathbf{P}\big[|\widehat{[\varphi]}_j/[\varphi]_j - 1| > 1/3\big] \le 2\exp\left(-\frac{m|[\varphi]_j|^2}{72}\right) \le 2\exp\left(-\frac{m\lambda_{M_m^u}}{72d}\right), \tag{A.20}$$

which implies the result by virtue of Lemma A2(ii). □

Fully adaptive estimation

Proof of Theorem 3.7. We begin the proof by defining the event $\Omega_{qp} := \Omega_q \cap \Omega_p$, where Ω_q is given in (A.12) and

$$\Omega_p := \big\{\big(N_n^\lambda \wedge M_m^\lambda\big) \le (\widehat{N}_n \wedge \widehat{M}_m) \le \big(N_n^u \wedge M_m^u\big)\big\}. \tag{A.21}$$

Observe that on Ω_q, we have $(1/2)\Delta_k \le \widehat{\Delta}_k \le (3/2)\Delta_k$ for all $0 \le k \le M_m^u$, and thus $(1/2)[\Delta_k \vee (k+2)] \le [\widehat{\Delta}_k \vee (k+2)] \le (3/2)[\Delta_k \vee (k+2)]$, which implies that

$$\begin{aligned}(1/2)k\Delta_k\bigg(\frac{\log[\Delta_k \vee (k+2)]}{\log(k+2)}\bigg)\bigg(1 - \frac{\log 2}{\log(k+2)}\frac{\log(k+2)}{\log(\Delta_k \vee [k+2])}\bigg)\\ \le \widehat{\delta}_k \le (3/2)k\Delta_k\bigg(\frac{\log(\Delta_k \vee [k+2])}{\log(k+2)}\bigg)\bigg(1 + \frac{\log 3/2}{\log(k+2)}\frac{\log(k+2)}{\log(\Delta_k \vee [k+2])}\bigg).\end{aligned}$$

Using $\log(\Delta_k \vee (k+2))/\log(k+2) \ge 1$, we conclude from the previous estimate that

$$\begin{aligned}\delta_k/10 \le (\log 3/2)/(2\log 3)\delta_k \le (1/2)\delta_k\big[1 - (\log 2)/\log(k+2)\big] \le \widehat{\delta}_k\\ \le (3/2)\delta_k\big[1 + (\log 3/2)/\log(k+2)\big] \le 3\delta_k.\end{aligned}$$

Letting $\mathrm{pen}(k) := 60\delta_k n^{-1}$ and $\widehat{\mathrm{pen}}(k) := 600\widehat{\delta}_k n^{-1}$, it follows that on Ω_q,

$$\mathrm{pen}(k) \le \widehat{\mathrm{pen}}(k) \le 30\,\mathrm{pen}(k) \qquad \forall 0 \le k \le M_m^u.$$

On $\Omega_{qp} = \Omega_q \cap \Omega_p$, we have $\widehat{k} \le M_m^u$. Thus,

$$\begin{aligned}\big(\mathrm{pen}(k \vee \widehat{k}) + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k})\big)\mathbb{1}\{\Omega_{qp}\} &\le \big(\mathrm{pen}(k) + \mathrm{pen}(\widehat{k}) + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k})\big)\mathbb{1}\{\Omega_{qp}\}\\ &\le 31\,\mathrm{pen}(k) \qquad \forall 0 \le k \le M_m^u.\end{aligned} \tag{A.22}$$

Now consider the decomposition

$$\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 = \mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2\mathbb{1}\{\Omega_{qp}\} + \mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2\mathbb{1}\big\{\Omega_{qp}^c\big\}. \tag{A.23}$$

We now bound the two terms separately:

$$\begin{aligned}\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2\mathbb{1}\{\Omega_{qp}\} \le C\bigg\{\|f - f_k\|_\omega^2 + d\zeta_d\frac{\delta_k^\lambda}{n} + rd\kappa_m\\ + d\zeta_d\frac{\delta_1^\lambda + \Sigma(\zeta_d^{-1}\|\varphi\|^2\|f\|^2)}{n}\bigg\},\\ \mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2\mathbb{1}\big\{\Omega_p^c\big\} \le C\bigg(\frac{d}{\lambda_1}\bigg)^7\frac{(1 + \|f\|_\omega^2)}{m}.\end{aligned}$$

Consider the first term. Following the proof of (A.14) line by line, it is easily seen that for $0 \le k \le (N_n^\lambda \wedge M_m^\lambda)$, we have

$$\begin{aligned}(1/4)\|\widehat{f}_{\widehat{k}} - f\|_\omega^2\mathbb{1}\{\Omega_{qp}\}\\ \le (7/4)(r\omega_k/\gamma_k) + 10\sum_{j=0}^{N_n^u}\bigg(\sup_{t\in\mathcal{B}_j}\big|\langle t, \widetilde{\Phi}_\nu\rangle_\omega\big|^2 - 6\frac{\delta_j}{n}\bigg)_+\end{aligned}$$

$$
\begin{aligned}
&+8\sup_{t\in\mathcal{B}_{N_n^u\wedge M_m^u}}\bigl|\langle t,\widehat{\Phi}_g-\widetilde{\Phi}_g\rangle_\omega\bigr|^2+\bigl(\operatorname{pen}(k\vee\widehat{k})+\widehat{\operatorname{pen}}(k)-\widehat{\operatorname{pen}}(\widehat{k})\bigr)\mathbb{1}\{\Omega_{qp}\}\\
&\le(7/4)(r\omega_k/\gamma_k)+10\sum_{j=0}^{N_n^u}\biggl(\sup_{t\in\mathcal{B}_j}\bigl|\langle t,\widetilde{\Phi}_\nu\rangle_\omega\bigr|^2-6\frac{\delta_j}{n}\biggr)_+\\
&\quad+8\sup_{t\in\mathcal{B}_{N_n^u\wedge M_m^u}}\bigl|\langle t,\widehat{\Phi}_g-\widetilde{\Phi}_g\rangle_\omega\bigr|^2+31\operatorname{pen}(k),
\end{aligned}
$$

where the last inequality follows from (A.22). The second and third terms are controlled by Lemmas A4 and A5, respectively (cf. the proof of (A.16)). It follows that

$$
\begin{aligned}
\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\mathbb{E}\|\widehat{f_{\widehat{k}}}-f\|_\omega^2\mathbb{1}\{\Omega_{qp}\}\le C\Bigl\{&(r+d\zeta_d)\min_{0\le k\le(N_n^\lambda\wedge M_m^\lambda)}\bigl[\max\bigl(\omega_k/\gamma_k,\delta_k^\lambda/n\bigr)\bigr]\\
&+rd\kappa_m+d\Sigma(rd\Lambda\zeta_d)n^{-1}\Bigr\}.
\end{aligned}
\tag{A.24}
$$

Consider the second term of (A.23). Following the proof of (A.17) and replacing Ω_q^c by Ω_{qp}^c, we obtain

$$
\mathbb{E}\|\widehat{f_{\widehat{k}}}-f\|_\omega^2\mathbb{1}\bigl\{\Omega_{qp}^c\bigr\}\le Cm^2\bigl(1+\|f\|_\omega^2\bigr)\mathbf{P}\bigl[\Omega_{qp}^c\bigr]^{1/2}.
$$

It follows by Lemma A7 that for all $m\ge1$,

$$
\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\mathbb{E}\|\widehat{f_{\widehat{k}}}-f\|_\omega^2\mathbb{1}\bigl\{\Omega_{qp}^c\bigr\}\le C(\lambda,d)(1+r)m^{-1}.
$$

The result of the theorem follows by combining the last estimate with (A.18) and (A.24). □

Lemma A7. *Under Assumptions* A1 *and* A2, *the event* Ω_p *defined in* (A.21) *satisfies*

$$
\mathbf{P}\bigl(\Omega_p^c\bigr)\le C(\lambda,d)m^{-6}\qquad\forall n,m\ge1.
$$

Proof. Let $\Omega_{\mathrm{I}}:=\{(N_n^\lambda\wedge M_m^\lambda)>(\widehat{N}_n\wedge\widehat{M}_m)\}$ and $\Omega_{\mathrm{II}}:=\{(\widehat{N}_n\wedge\widehat{M}_m)>(N_n^u\wedge M_m^u)\}$. We then have $\Omega_p^c=\Omega_{\mathrm{I}}\cup\Omega_{\mathrm{II}}$. First, consider $\Omega_{\mathrm{I}}=\{\widehat{N}_n<(N_n^\lambda\wedge M_m^\lambda)\}\cup\{\widehat{M}_m<(N_n^\lambda\wedge M_m^\lambda)\}$. By the definition of N_n^λ, we have that $\min_{1\le|j|\le N_n^\lambda}\frac{|[\varphi]_j|^2}{|j|\omega_j^+}\ge\frac{4(\log n)}{n}$, which implies that

$$
\begin{aligned}
\bigl\{\widehat{N}_n<\bigl(N_n^\lambda\wedge M_m^\lambda\bigr)\bigr\}&\subset\biggl\{\exists1\le|j|\le\bigl(N_n^\lambda\wedge M_m^\lambda\bigr):\ \frac{|\widehat{[\varphi]}_j|^2}{|j|\omega_j^+}<\frac{\log n}{n}\biggr\}\\
&\subset\bigcup_{1\le|j|\le N_n^\lambda\wedge M_m^\lambda}\biggl\{\frac{|\widehat{[\varphi]}_j|}{|[\varphi]_j|}\le1/2\biggr\}\subset\bigcup_{1\le|j|\le N_n^\lambda\wedge M_m^\lambda}\biggl\{\biggl|\frac{\widehat{[\varphi]}_j}{[\varphi]_j}-1\biggr|\ge1/2\biggr\}.
\end{aligned}
$$

From $\min_{1\le|j|\le M_m^\lambda}|[\varphi]_j|^2 \ge 4m^{-1+b_m}$, it follows in the same way that

$$\big\{\widehat{M}_m < \big(N_n^\lambda \wedge M_m^\lambda\big)\big\} \subset \bigcup_{1\le|j|\le N_n^\lambda\wedge M_m^\lambda} \bigg\{\bigg|\frac{\widehat{[\varphi]}_j}{[\varphi]_j} - 1\bigg| \ge 1/2\bigg\}.$$

Therefore, $\Omega_{\mathrm{I}} \subset \bigcup_{1\le|j|\le M_m^u}\{|\widehat{[\varphi]}_j/[\varphi]_j - 1| \ge 1/2\}$, because $M_m^\lambda \le M_m^u$. Thus, applying Hoeffding's inequality and Lemma A2(ii) as in (A.20) yields

$$\mathbf{P}[\Omega_{\mathrm{I}}] \le \sum_{1\le|j|\le M_m^u} 2\exp\bigg(-\frac{m|[\varphi]_j|^2}{72}\bigg) \le C(d)m^{-6}. \tag{A.25}$$

Consider $\Omega_{\mathrm{II}} = \{\widehat{N}_n > (N_n^u \wedge M_m^u)\} \cap \{\widehat{M}_m > (N_n^u \wedge M_m^u)\}$. In the case where $(N_n^u \wedge M_m^u) = N_n^u$, use $\frac{\log n}{4n} \ge \max_{|j|\ge N_n^u+1} \frac{|[\varphi]_j|^2}{|j|\omega_j^+}$, such that

$$\Omega_{\mathrm{II}} \subset \big\{\widehat{N}_n > N_n^u\big\} \subset \bigg\{\forall 1 \le |j| \le N_n^u + 1\colon \frac{|\widehat{[\varphi]}_j|^2}{|j|\omega_j^+} \ge \frac{\log n}{n}\bigg\}$$

$$\subset \bigg\{\frac{|\widehat{[\varphi]}_{N_n^u+1}|}{|[\varphi]_{N_n^u+1}|} \ge 2\bigg\} \subset \big\{\big|\widehat{[\varphi]}_{N_n^u+1}/[\varphi]_{N_n^u+1} - 1\big| \ge 1\big\}.$$

In the case where $(N_n^u \wedge M_m^u) = M_m^u$, it follows analogously from $m^{-1+b_m} \ge 4 \times \max_{|j|\ge M_m^u+1}|[\varphi]_j|^2$ that

$$\Omega_{\mathrm{II}} \subset \big\{\widehat{M}_m > M_m^u\big\} \subset \big\{\big|\widehat{[\varphi]}_{M_m^u+1}/[\varphi]_{M_m^u+1} - 1\big| \ge 1\big\}.$$

Therefore, we have $\Omega_{\mathrm{II}} \subset \{|\widehat{[\varphi]}_{(N_n^u\wedge M_m^u)+1}/[\varphi]_{(N_n^u\wedge M_m^u)+1} - 1| \ge 1\}$. Applying Hoeffding's inequality as in (A.20) and using Assumption A2, we obtain, for all $m \ge 1$,

$$\mathbf{P}[\Omega_{\mathrm{II}}] \le 2\exp\bigg(-\frac{m|[\varphi]_{M_m^u+1}|^2}{72}\bigg) \le C(\lambda,d)m^{-7}. \tag{A.26}$$

Combining (A.25) and (A.26) implies the result. □

Illustration: Estimation of derivatives

Proof of Proposition 3.10. In light of the proof of Proposition 2.6, we apply Theorem 3.7, where in both cases we need only check the additional Assumption A2. The result then follows by an evaluation of the upper bound.

[o-o] It is easily seen that $(m\lambda_{M_m^u+1})^{-1}\log m = o(1)$ as $m \to \infty$. Thus, Assumption A2 is satisfied in this case. Because $k_n^* \sim n^{1/(2a+2p+1)}$, we have $k_n^* \lesssim N_n^\lambda$. Thus, the upper bound is

$$\big(k_n^* \wedge M_{m_n}^\lambda\big)^{-2(p-s)} + m_n^{-(1\wedge((p-s)/a))}. \tag{A.27}$$

We consider two cases. First, let $p - s > a$. Suppose that $n^{2(p-s)/(2p+2a+1)} = \mathrm{O}(m_n)$; then,

$$\frac{k_n^*}{M_{m_n}^\lambda} \sim \frac{n^{1/(2a+2p+1)}}{(m_n^{1-b_{m_n}})^{(1/2a)}} = \frac{n^{1/(2a+2p+1)}}{m_n^{1/2(p-s)}} \big(m_n^{-a+(p-s)(1-b_{m_n})}\big)^{1/(2(p-s)a)} = \mathrm{o}(1).$$

This means that $k_n^* \lesssim M_{m_n}^\lambda$, so the resulting upper bound is $(k_n^*)^{-2(p-s)} + m_n^{-1} \lesssim (k_n^*)^{-2(p-s)}$. Suppose now that $m_n = \mathrm{o}(n^{2(p-s)/(2p+2a+1)})$. If in addition $k_n^* = \mathrm{O}(M_{m_n}^\lambda)$, then the first summand in (A.27) reduces to $(k_n^*)^{-2(p-s)}$ and thus the upper bound is m_n^{-1}. On the other hand, if $M_{m_n}^\lambda / k_n^* = \mathrm{o}(1)$, then the first term is $(M_{m_n}^\lambda)^{-2(p-s)} \sim (m_n^{-a+(p-s)(1-b_{m_n})})^{1/a} m_n^{-1} \lesssim m_n^{-1}$, because $p - s > a$. Combining both cases, we obtain the result in the case where $p - s > a$.

Now assume that $p - s \le a$. First, suppose that $k_n^* = \mathrm{O}(M_{m_n}^\lambda)$. Then the first summand in (A.27) reduces to $(k_n^*)^{-2(p-s)}$, and, moreover, it follows that $n^{2a/(2p+2a+1)} = \mathrm{O}(m_n)$. Therefore, the upper bound is $(k_n^*)^{-2(p-s)}$. Now consider $M_{m_n}^\lambda = \mathrm{o}(k_n^*)$. Then (A.27) can be rewritten as $(m_n^{1-b_{m_n}})^{-(p-s)/a} + m_n^{-(p-s)/a}$, which results in the rate $(m_n^{1-b_m})^{-(p-s)/a}$. Combining both cases gives the result. More precisely, $m_n = \mathrm{o}(n^{2a/(2p+2a+1)})$ implies $M_{m_n}^\lambda = \mathrm{o}(k_n^*)$. In contrast, in the case where $n^{2a/(2p+2a+1)} = \mathrm{O}(m_n)$, if $k_n^* / M_{m_n}^\lambda = \mathrm{O}(1)$, then the rate is $(k_n^*)^{-2p}$, whereas if $M_{m_n}^\lambda / k_n^* = \mathrm{o}(1)$, then the rate is $(m_n^{1-b_m})^{-(p-s)/a}$.

[s-o] As in case [o-o], Assumption A2 is satisfied. Recall that $k_n^* \sim (\log n)^{1/(2p)}$. If $n(\log n)^{-(2a+2s+1)/(2p)} = \mathrm{O}(m_n)$, then $k_n^* \lesssim M_{m_n}^\lambda$ and $m_n^{-1} \lesssim \psi_{n,m_n}^\diamond \sim n^{-1}(\log n)^{(2a+2s+1)/(2p)}$. In the opposite case, we have $\psi_{n,m_n}^\diamond \lesssim m_n^{-1}$, which proves the result.

[o-s] To verify that Assumption A2 is satisfied in this setting, we can proceed as follows. Define the sequence $\widetilde{M}^u$ exactly as M^u, but replacing b_m by $a_m = b_m^{2^k}$. Then $\widetilde{M}^u$ satisfies assertion Lemma A2(ii), the proof being similar to that for M^u. In contrast, we can show that $\widetilde{M}_m^u - M_m^u \to \infty$ as $m \to \infty$, which amounts to showing Assumption A2.

We have $k_n^* \asymp (\log n)^{1/2a}$. The upper bound becomes $(k_n^* \wedge M_{m_n}^\lambda)^{-2(p-s)} + (\log m_n)^{-(p-s)/a} \sim (k_n^* \wedge M_{m_n}^\lambda)^{-2(p-s)}$. Distinguishing $k_n^* \lesssim M_m^\lambda$ and the opposite case shows the result. □

Acknowledgements

This work was supported by the IAP research network no. P6/03 of the Belgian Government (Belgian Science Policy) and by the "Fonds Spéciaux de Recherche" from the Université catholique de Louvain.

We are grateful to two referees for raising a number of points which helped us clarify several important issues. Finally, we thank the Associate Editor for constructive criticism and clear guidelines.

References

[1] Barron, A., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413. MR1679028

[2] Bigot, J. and Van Bellegem, S. (2009). Log-density deconvolution by wavelet thresholding. *Scand. J. Stat.* **36** 749–763. MR2573306
[3] Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *Canad. J. Statist.* **32** 181–192. MR2064400
[4] Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli* **11** 309–340. MR2132729
[5] Butucea, C. and Tsybakov, A.B. (2008). Sharp optimality in density deconvolution with dominating bias. I. *Theory Probab. Appl.* **52** 24–39.
[6] Butucea, C. and Tsybakov, A.B. (2008). Sharp optimality in density deconvolution with dominating bias. II. *Theory Probab. Appl.* **52** 237–249.
[7] Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186. MR0997599
[8] Cavalier, L. and Hengartner, N.W. (2005). Adaptive estimation for inverse problems with noisy operators. *Inverse Problems* **21** 1345–1361. MR2158113
[9] Cochran, W.W., Mouritsen, H. and Wikelski, M. (2004). Migrating songbirds recalibrate their magnetic compass daily from twilight cues. *Science* **304** 405–408.
[10] Comte, F. and Johannes, J. (2010). Adaptive estimation in circular functional linear models. *Math. Methods Statist.* **19** 42–63. MR2682854
[11] Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 601–627. MR2853732
[12] Comte, F., Rozenholc, Y. and Taupin, M.L. (2006). Penalized contrast estimator for density deconvolution. *Canadian Journal of Statistics* **34** 431–452.
[13] Comte, F. and Taupin, M.L. (2003). Adaptive density deconvolution for circular data. Prépublication MAP5 2003-10 report, Université Paris Descartes.
[14] Curray, J.R. (1956). The analysis of two-dimensional orientation data. *The Journal of Geology* **64** 117–131.
[15] Devroye, L. (1989). Consistent deconvolution in density estimation. *Canad. J. Statist.* **17** 235–239. MR1033106
[16] Diggle, P.J. and Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser. B* **55** 523–531. MR1224414
[17] Efromovich, S. (1997). Density estimation for the case of supersmooth measurement error. *J. Amer. Statist. Assoc.* **92** 526–535. MR1467846
[18] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324
[19] Fan, J. (1992). Deconvolution with supersmooth distributions. *Canad. J. Statist.* **20** 155–169. MR1183078
[20] Fan, J. and Koo, J.Y. (2002). Wavelet deconvolution. *IEEE Trans. Inform. Theory* **48** 734–747. MR1889978
[21] Fisher, N.I. (1993). *Statistical Analysis of Circular Data*. Cambridge: Cambridge Univ. Press. MR1251957
[22] Gill, J. and Hangartner, D. (2010). Circular data in political science and how to handle it. *Political Analysis* **18** 316–336.
[23] Goldenshluger, A. (1999). On pointwise adaptive nonparametric deconvolution. *Bernoulli* **5** 907–925. MR1715444
[24] Goldenshluger, A. (2002). Density deconvolution in the circular structural model. *J. Multivariate Anal.* **81** 360–375. MR1906385
[25] Hall, P. and Yao, Q. (2003). Inference in components of variance models with low replication. *Ann. Statist.* **31** 414–441. Dedicated to the memory of Herbert E. Robbins. MR1983536

[26] Horowitz, J.L. and Markatou, M. (1996). Semiparametric estimation of regression models for panel data. *Rev. Econom. Stud.* **63** 145–168. MR1372250
[27] Johannes, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist.* **37** 2301–2323. MR2543693
[28] Kawata, T. (1972). *Fourier Analysis in Probability Theory*. *Probability and Mathematical Statistics* **15**. New York: Academic Press. MR0464353
[29] Kim, P.T. and Koo, J.Y. (2002). Optimal spherical deconvolution. *J. Multivariate Anal.* **80** 21–42. MR1889831
[30] Koo, J.Y. and Park, B.U. (1996). *B*-spline deconvolution based on the EM algorithm. *J. Statist. Comput. Simulation* **54** 275–288. MR1701219
[31] Korostelëv, A.P. and Tsybakov, A.B. (1993). *Minimax Theory of Image Reconstruction*. *Lecture Notes in Statistics* **82**. New York: Springer. MR1226450
[32] Mardia, K.V. (1972). *Statistics of Directional Data*. *Probability and Mathematical Statistics* **13**. London: Academic Press. MR0336854
[33] Meister, A. (2006). Density estimation with normal measurement error with unknown variance. *Statist. Sinica* **16** 195–211. MR2256087
[34] Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. *Lecture Notes in Statistics* **193**. Berlin: Springer. MR2768576
[35] Mendelsohn, J. and Rice, J. (1982). Deconvolution of micro-fluorometric histograms with B-splines. *J. Amer. Statist. Assoc.* **77** 748–753.
[36] Natterer, F. (1984). Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Anal.* **18** 29–37. MR0762862
[37] Neubauer, A. (1988). An a posteriori parameter choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates. *SIAM J. Numer. Anal.* **25** 1313–1326. MR0972456
[38] Neubauer, A. (1988). When do Sobolev spaces form a Hilbert scale? *Proc. Amer. Math. Soc.* **103** 557–562. MR0943084
[39] Neumann, M.H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist.* **7** 307–330. MR1460203
[40] Neumann, M.H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.* **98** 1955–1968. MR2396948
[41] Olver, F.W.J. (1974). *Asymptotics and Special Functions*. New York: Academic Press.
[42] Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27** 2033–2053. MR1765627
[43] Petrov, V.V. (1995). *Limit Theorems of Probability Theory*: *Sequences of Independent Random Variables*. *Oxford Studies in Probability* **4**. New York: Clarendon. MR1353441
[44] Schwarz, M. and Van Bellegem, S. (2010). Consistent density deconvolution under partially known error distribution. *Statist. Probab. Lett.* **80** 236–241. MR2575451
[45] Stefanski, L.A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.* **9** 229–235. MR1045189
[46] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563. MR1419006
[47] Zhang, C.H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–831. MR1056338

Received January 2011 and revised August 2011