Adaptive circular deconvolution by model selection under unknown error distribution

Jan Johannes * Maik Schwarz *

Université catholique de Louvain

23 February 2010

We consider a circular deconvolution problem, where the density f of a circular random variable X has to be estimated nonparametrically based on an iid. sample from a noisy observation Y of X. The additive measurement error is supposed to be independent of X. The objective of this paper is the construction of a fully data-driven estimation procedure when the error density φ is unknown. However, we suppose that in addition to the iid. sample from Y, we have at our disposal an additional iid. sample independently drawn from the error distribution.

First, we develop a minimax theory in terms of both sample sizes. We propose an orthogonal series estimator attaining the minimax rates but requiring an 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 our 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 supposing the degree of ill-posedness of the underlying inverse problem to be known, which amounts to assuming partial knowledge of the error distribution. We show that this data-driven estimator can attain the lower risk bound up to a constant in both sample sizes n and m over a wide range of density classes covering in particular ordinary and super smooth densities. Finally, by randomizing the penalty and the collection of models, we modify the estimator such that it does not require any prior knowledge of the error distribution anymore. 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.

AMS 2000 subject classifications: Primary 62G07; secondary 62G20, 42A85.
Keywords: Circular deconvolution, Orthogonal series estimation, Spectral cut-off, Model selection, Adaptive density estimation, Minimax theory.

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.

*Institut de Statistique, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgique, e-mail: {jan.johannes|maik.schwarz}@uclouvain.be

1. Introduction

This work deals with the estimation of the density of a circular random variable from noisy observations. Such data occur in various fields of natural science, as for example in geology and biology, to mention but two. Curray (1956) discusses the analysis of directional data in the context of geological research, where it is often useful to measure and analyze the orientations of various features. More recently, Cochran et al. (2004) investigated migrating songbirds' navigation abilities. They fitted birds with radio transmitters and placed them in outdoor cages in an artificially turned magnetic field. The observations consisted of the directions the birds departed in when released. Such directional observations can be represented as points on a compass rose and hence on the circle. For a more general and detailed discussion of the particularities of circular data we refer to Mardia (1972) and Fisher (1993).

Let X be the circular random variable whose density f we are interested in and ε an independent additive circular error with unknown density φ. Denote by Y the contaminated observation data and by g its density. Throughout this work we will 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 $Y = X+\varepsilon \mod [0,1)$. Then, we have

$$g(y) = (f * \varphi)(y) := \int_{[0,1)} f((y-s) - \lfloor y-s\rfloor)\,\varphi(s)\,ds, \quad y \in [0,1),$$

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 $\langle f,g\rangle = \int_{[0,1)} f(x)\overline{g(x)}dx$ where $\overline{g(x)}$ denotes the complex conjugate of $g(x)$. In this work we suppose that f and φ, and hence g, belong to the subset $\mathcal{D}$ of all densities in L^2. As a consequence, they admit representations as discrete Fourier series with respect to the exponential basis $\{e_j\}_{j\in\mathbb{Z}}$ of L^2, where $e_j(x) := \exp(-i2\pi jx)$ 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 j-th 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 is assumed from now on, we have

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

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 the error density is fully known, a very popular approach based on kernel methods has been considered by Carroll and Hall (1988), Devroye (1989), Fan (1991, 1992), Stefanski (1990), Zhang (1990), Goldenshluger (1999, 2000) and Kim and Koo (2002)), to name but a few. Mendelsohn and Rice (1982) and Koo and Park (1996), for example, have studied spline-based methods, while a wavelet decomposition has been used by Pensky and Vidakovic (1999), Fan and Koo (2002) and Bigot and Van Bellgem (2009), for instance. Situations with only partial knowledge about the error density have also been considered (c.f. Butucea and Matias (2005), Meister (2004, 2006), or Schwarz and Van Bellegem (2009)). Consistent deconvolution without prior knowledge of the error distribution is also possible in the case of panel data (c.f. Horowitz and Markatou (1996), Hall and Yao (2003) or Neumann (2007)) or by assuming an additional sample from the error distribution (c.f. Diggle and Hall (1993), Neumann (1997), Johannes (2009) or Comte and Lacour (2009)). For a broader overview on deconvolution problems the reader may refer to the recent monograph by Meister (2009).

Let us return to the circular case. In this paper we suppose that we do not know the error density φ, but that we have at our disposal in addition to the iid. sample $(Y_k)_{k=1}^n$ of size $n \in \mathbb{N}$ from g an independent iid. sample $(\varepsilon_k)_{k=1}^m$ of size $m \in \mathbb{N}$ from φ. Our purpose is to establish a fully data-driven estimation procedure for the deconvolution density f which 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 shall 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 $\|\cdot\|_\omega := \sum_{j\in\mathbb{Z}}\omega_j|[\cdot]_j|^2$, where $\omega := (\omega_j)_{j\in\mathbb{Z}}$ is a strictly positive sequences of weights. This allows us to quantify the estimation accuracy in terms of the mean integrated square error (MISE) not only of f itself, but as well 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 sequence 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. In the standard context of an ordinary smooth deconvolution density for example, i.e. when $([f])_{j\in\mathbb{Z}}$ decays polynomially, logarithmic rates of convergence appear when the error density is super smooth, i.e., $([\varphi])_{j\in\mathbb{Z}}$ has a exponential decay. This special case is treated in Efromovich (1997), for example. However, this situation and many others are covered by the density classes

$$\mathcal{F}_\gamma^r := \Big\{p \in \mathcal{D} \;:\; \sum_{j\in\mathbb{Z}} \gamma_j|[p]_j|^2 =: \|p\|_\gamma^2 \leqslant r\Big\} \text{ and}$$

$$\mathcal{E}_\lambda^d := \Big\{p \in \mathcal{D} \;:\; 1/d \leqslant \frac{|[p]_j|^2}{\lambda_j} \leqslant d \quad \forall\, j \in \mathbb{Z}\Big\},$$

where $r, d \geqslant 1$ 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 show a lower bound of the maximal weighted risk which is essentially determined by the sequences γ, λ and ω. This lower bound is composed of two main terms, each of them depending on the size of one sample, but not on the other. Let us define an orthogonal series estimator by replacing the unknown Fourier coefficients in (1.1) by empirical counterparts, that is,

$$\widehat{f}_k := 1 + \sum_{0<|j|\leqslant k} \frac{\widehat{[g]}_j}{\widehat{[\varphi]}_j}\mathbb{1}\{|\widehat{[\varphi]}_j|^2 \geqslant 1/m\}e_j \quad \text{with}$$

$$\widehat{[g]}_j := \frac{1}{n}\sum_{k=1}^n e_j(-Y_k) \quad \text{and} \quad \widehat{[\varphi]}_j := \frac{1}{m}\sum_{k=1}^m e_j(-\varepsilon_k). \quad (1.2)$$

Again, things work out similarly in deconvolution on the real line, where one only has to replace the empirical Fourier coefficients by the corresponding values of the empirical characteristic functions. Similar estimators have already been studied by Neumann (1997) on the real line and by Efromovich (1997) in the circular case, for example. We show below that the estimator $\widehat{f}_k$ attains the lower bound and is hence 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 is a polynomial in n, i.e. $m_n = n^r$ for any $r > 0$.

Of course, minimax optimality is only achieved as long as the dimension parameter k is chosen in an optimal way. In general, this optimal choice of k depends among others on 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 but not on γ. Hence, the estimator is automatically adaptive with respect to γ under the optimal choice of k. In this situation Efromovich (1997) provides an estimator which is also adaptive with respect to the super smooth error density. On the contrary, Cavalier and Hengartner (2005), deriving oracle inequalities in an indirect regression problem based on a circular convolution contaminated by Gaussian white noise, treat the ordinary smooth case only. As in our setting, their observation scheme involves two independent samples. It is worth to note that in order to apply these estimators, one has to know in advance at least if the error density is ordinary or super smooth. We provide in this work a unified estimation procedure which can attain minimax rates in either of the both cases, that is, which is adaptive over a class including both ordinary and super smooth error densities. This fully adaptive method to choose the parameter k, only depends on the observations and not on characteristics of neither f nor φ. The central result of the present paper states that for this automatic choice $\widehat{k}$, the estimator $\widehat{f}_{\widehat{k}}$ attains the lower bound up to a constant, and is thus minimax-optimal, over a wide range of sequences γ and λ, covering in particular both ordinary and super smooth error densities.

As far as the two sample sizes are concerned, the assumption made by Cavalier and Hengartner (2005) 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 is also used by Efromovich (1997). However, as mentioned above, without changing the minimax rates, the ε-sample size can be reduced to m_n, which can be far smaller than n. This is a desirable property, as the observation of the additional sample from ε may be expensive in practise. Nevertheless, the minimal choice of m depends among others on the sequences γ and λ and is hence unknown in general. In spite of the minimax rate being eventually deteriorated by 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 has to be paid for adaptivity. Surprisingly, even in the cases where the optimal rate is not attained anymore, the deterioration is only of logarithmic order as far as the error density is either ordinary or super smooth.

The adaptive choice of k is motivated by the general model selection strategy developed in Barron et al. (1999). Concretely, following Comte and Taupin (2003), who treat the case of a known error density only, $\widehat{k}$ is the minimizer of a penalized contrast

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

Note that we can compute $\|\widehat{f}_k\|_\omega^2 = 1 + \sum_{0<|j|\leqslant k} \omega_j |[\widehat{g}]_j|^2 |[\widehat{\varphi}]_j|^{-2} \mathbb{1}\{|[\widehat{\varphi}]_j|^2 \geqslant 1/m\}$. As in case of a known error density, it turns out that the penalty function $\operatorname{pen}(\cdot)$ as well as the upper bound K needed for the right 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, assuming this parameter to be known, we show an upper risk bound for this partially adaptive estimator $\widehat{f}_{\widehat{k}}$. We prove that over a wide range of sequences γ and λ, the adaptive choice of k yields the same upper risk bound as the optimal choice, up to a constant Finally, we drop the requirement that the degree of ill-posedness is known. In order to choose k adaptively even in this case, we replace $\operatorname{pen}(\cdot)$ and K by estimates only depending 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. It is noteworthy that even though the proofs are more intricate in this case, the result strongly resembles its analogon in the case of known degree of ill-posedness.
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 only exploited in the proof of the lower bound and in order to guarantee the existence of the discrete representation in (1.1), which still holds in case of a compactly supported deconvolution density.
This article 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 we derive the optimal convergence rates in the ordinary and in the super smooth case. Section 3 is devoted to the construction of the adaptive estimator in the case of known degree of ill-posedness. An upper risk bound is shown and convergence rates for the ordinary and super smooth case are compared to the minimax optimal ones. The last section provides the fully adaptive generalization of this method. All proofs are deferred to the appendix.

2. Minimax optimal estimation

In this section we develop the minimax theory for the estimation of a circular deconvolution density under unknown error density when two independent samples from Y and ε are available. A lower bound depending on both sample sizes is derived and it is shown that the orthogonal series estimator $\widehat{f}_k$ defined in (1.2) attains this lower bound up to a constant. All results in this paper are derived under the following minimal regularity conditions.

Assumption 2.1 *Let $\gamma := (\gamma_j)_{j\in Z}$, $\omega := (\omega_j)_{j\in Z}$ and $\lambda := (\lambda_j)_{j\in Z}$ be strictly positive symmetric sequences of weights with $\gamma_0 = \omega_0 = \lambda_0 = 1$ such that $(\omega_n/\gamma_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ are non-increasing, respectively.*

Remark that λ_j is even a null sequence as $|j|$ tends to infinity as we suppose φ to be a density in L^2. The assumption that ω/γ is non-increasing ensures that the weighted risk is well defined.

Lower bounds The next assertion provides a lower bound in 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 case of an unknown error density.

Theorem 2.2 *Suppose an iid. Y-sample of size n and that the error density φ is known. Consider sequences ω, γ, and λ satisfying Assumption 2.1 such that $\sum_{j\in\mathbb{Z}}\gamma_j^{-1} = \Gamma < \infty$ and such that $\varphi \in \mathcal{E}_\lambda^d$ for some $d \geqslant 1$. Define for all $n \geqslant 1$*

$$\psi_n := \psi_n(\gamma,\lambda,\omega) := \min_{k\in\mathbb{N}}\Big\{\max\Big(\frac{\omega_k}{\gamma_k}, \sum_{0<|j|\leqslant k}\frac{\omega_j}{n\lambda_j}\Big)\Big\} \text{ and}$$
$$k_n^* := k_n^*(\gamma,\lambda,\omega) := \operatorname*{argmin}_{k\in\mathbb{N}}\Big\{\max\Big(\frac{\omega_k}{\gamma_k}, \sum_{0<|j|\leqslant k}\frac{\omega_j}{n\lambda_j}\Big)\Big\}. \quad (2.1)$$

If in addition $\eta := \inf_{n\geqslant 1}\{\psi_n^{-1}\min(\omega_{k_n^*}\gamma_{k_n^*}^{-1}, \sum_{0<|l|\leqslant k_n^*}\omega_l(n\lambda_l)^{-1})\} > 0$, then for all $n \geqslant 2$ and for any estimator $\widetilde{f}$ of f we have

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

The proof of the last assertion is based on Assuoad's cube technique (c.f. Korostolev and Tsybakov (1993)), where we construct $2^{2k_n^*}$ candidates of deconvolution densities which have the largest possible $\|\cdot\|_\omega$-distance but are still statistically non distinguishable. It is worth to note that the additional assumption $\sum_{j\in\mathbb{Z}}\gamma_j^{-1} = \Gamma < \infty$ is only used to ensure that these candidates are densities. Observe further that in case $r = 1$, the lower bound is equal to zero, because in this situation the set $\mathcal{F}_\gamma^r$ reduces to a singleton containing 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 *Suppose independent iid. samples from Y and ε of size n and m, respectively. Consider sequences ω, γ, and λ satisfying Assumption 2.1. For all $m \geqslant 2$, let*

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

If in addition there exists a density in $\mathcal{E}_\lambda^{\sqrt{d}}$ which is bounded from below by $1/2$, then, for all $m \geqslant 2$ and for any estimator $\widetilde{f}$ of f we have

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

The proof of the last assertion takes its inspiration from a proof given in Neumann (1997). In contrast to the proof of Theorem 2.2 we only have to compare two candidates of error densities which are still statistically non distinguishable. However, to ensure that these candidates are densities, we impose the additional condition. It is easily seen that this condition is satisfied if $\Lambda := \sum_{j\in\mathbb{Z}}\lambda_j^{-1/2} < \infty$ and $\sqrt{d} \geqslant \max(4\Lambda^2, 1)$. It is worth to note that in case $d = 1$, the set $\mathcal{E}_\lambda^d$ of possible error densities reduces to a singleton, and hence the lower bound is equal to zero. Finally, by combination of both lower bounds we obtain the next corollary.

Corollary 2.4 *Under the assumptions of Theorem 2.2 and 2.3 we have for any estimator* $\widetilde{f}$ *of* f *and for all* $n, m \geqslant 2$ *that*

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\Big\{\mathbb{E}\|\widetilde{f}-f\|_\omega^2\Big\} \geqslant \frac{\eta\min(r-1,(8d\Gamma)^{-1})\min(d^{-1/2},4(1-d^{1/4})^2)}{16d}\{\psi_n+\kappa_m\}.$$

Upper bound The next theorem summarizes sufficient conditions to ensure the optimality of the orthogonal series estimator $\widehat{f}_k$ defined in (1.2) provided the dimension parameter k is chosen appropriately. To be more precise, we use the value k_n^* defined in (2.1) which obviously depends on the sequences ω, γ and λ but surprisingly not on the ε-sample size m. However, under this choice the estimator attains the lower bound given in Corollary 2.4 up to a constant and hence it is minimax-optimal.

Theorem 2.5 *Suppose independent iid. sample from* Y *and* ε *of size* n *and* m*, respectively. Consider sequences* ω*,* γ *and* λ *satisfying Assumption 2.1. Let* $\widehat{f}_{k_n^*}$ *be the estimator given in (1.2) with* k_n^* *defined in (2.1). Then, there exists a numerical constant* $C > 0$ *such that for all* $n, m \geqslant 1$ *we have*

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

Note that under slightly stronger conditions on the sequences ω, γ and λ than Assumption 2.1 it can be shown that in case of equally large samples from Y and ε we have always the rate as in case of known error density. However, below we show that in special cases the required ε-sample size can be much smaller than the Y-sample size.

2.1. Illustration: estimation of derivatives.

To illustrate the previous results we assume in the following that the deconvolution density f is an element of the Sobolev space of periodic functions $\mathcal{W}_p$, $p \in \mathbb{N}$, given by

$$\mathcal{W}_p = \Big\{f \in H_s : f^{(j)}(0) = f^{(j)}(1), \quad j = 0, 1, \dots, p-1\Big\},$$

where $H_p := \{f \in L^2[0,1] : f^{(p-1)}$ absolutely continuous, $f^{(p)} \in L^2[0,1]\}$ is a Sobolev space (c.f. Neubauer (1988a,b)). However, if we consider the sequence of weights

$$\gamma_0 = 1 \quad \text{and} \quad \gamma_j = |j|^{2p}, \qquad |j| > 0,$$

then, the Sobolev space $\mathcal{W}_p$ of periodic functions coincides with $\mathcal{F}_w$. Therefore, let us denote by $\mathcal{W}_p^r := \mathcal{F}_w^r$, $r > 0$, an ellipsoid in the Sobolev space $\mathcal{W}_p$. In this illustration, we shall consider the estimation of derivatives of the deconvolution density f. Therefore, it is interesting to recall that, up to a constant, for any function $h \in \mathcal{W}_p^r$ the weighted norm $\|h\|_\omega$ with

$$\omega_0 = 1 \quad \text{and} \quad \omega_j = |j|^{2s}, \qquad |j| > 0,$$

equals the L^2-norm of the s-th weak derivative $h^{(s)}$ for each integer $0 \leqslant s \leqslant p$. By virtue of this relation, the results in the previous section imply also a lower as well as an upper bound of the L^2-risk for the estimation of the s-th weak derivative of f. Finally, we restrict our attention to error densities being either

[os] *ordinary smooth*, that is, the sequence λ is polynomially decreasing, i.e., $\lambda_0 = 1$ and $\lambda_j = |j|^{-2a}$, $|j| > 0$, for some $a > 1/2$, or

[ss] *super smooth*, that is, the sequence λ is exponentially decreasing, i.e., $\lambda_0 = 1$ and $\lambda_j = \exp(-|j|^{2a})$, $|j| > 0$, for some $a > 0$.

It is easily seen that the minimal regularity conditions given in Assumption 2.1 are satisfied. Moreover, the additional conditions used in Theorems 2.2 and 2.3, i.e., $\Gamma = \sum_{j\in\mathbb{Z}} \gamma_j^{-1} < \infty$ and that there there exists $\varphi \in \mathcal{E}_\lambda^{\sqrt{d}}$ with $\varphi \geqslant 1/2$, are satisfied in the super smooth case **[ss]** if $p > 1/2$ and in the ordinary smooth case **[os]** if in addition $a > 1$. Roughly speaking, this means that both the deconvolution density and the error density are at least continuous. The lower bound presented in the next assertion follows now directly from Corollary 2.4. Here and subsequently, we write $a_n \lesssim b_n$ when there exists $C > 0$ such that $a_n \leqslant C\, b_n$ 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.

Proposition 2.6 *Suppose independent iid. sample from Y and ε of size n and m, respectively. Then we have for any estimator $\widetilde{f}^{(s)}$ of $f^{(s)}$*

[os] *in the ordinary smooth case, for all $p > 1/2$ and $a > 1$ that*

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

[ss] *in the super smooth case, for all $p > 1/2$ that*

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

As an estimator of $f^{(s)}$, we shall consider, the s-th weak derivative of the estimator $\widehat{f}_k$ defined in (1.2). Given the exponential basis $\{e_j\}_{j\in\mathbb{Z}}$, we recall that for each integer $0 \leqslant s \leqslant p$ the s-th derivative in a weak sense of the estimator $\widehat{f}_k$ is

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

Applying Theorem 2.5, the rates of the lower bound given in the last assertion provide, up to a constant, also an upper bound of the L^2-risk of the estimator $\widehat{f}_k^{(s)}$, which is summarized in the next proposition. We have thus proved that these rates are optimal and the proposed estimator $\widehat{f}_k^{(s)}$ is minimax optimal in both cases. Furthermore, it is of interest to characterize the minimal size m of the additional sample from ε needed to attain the same rate as in case of a known error density. Hence, we let the ε-sample size depend on the Y-sample size n, too.

Proposition 2.7 *Suppose independent iid. sample from Y and ε of size n and m, respectively. Consider the estimator $\widehat{f}_k^{(s)}$ given in (2.3).*

[os] *In the ordinary smooth case, with dimension parameter $k \sim n^{1/(2p+2a+1)}$ we have*

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

and for any sequence $(m_n)_{n\geqslant 1}$ *follows as* $n \to \infty$

$$\sup_{f\in\mathcal{W}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\Big\{\mathbb{E}\|\widehat{f}_k^{(s)}-f^{(s)}\|^2\Big\} = \begin{cases} O(n^{-2(p-s)/(2p+2a+1)}) & \text{if } n^{2((p-s)\vee a)/(2p+2a+1)} = O(m_n)\\ O(m_n^{-((p-s)\wedge a)/a}) & \text{otherwise.}\end{cases}$$

[ss] *In the super smooth case, with dimension parameter* $k \sim (\log n)^{1/(2a)}$ *we have*

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

and for any sequence $(m_n)_{n\geqslant 1}$ *follows as* $n \to \infty$

$$\sup_{f\in\mathcal{W}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\Big\{\mathbb{E}\|\widehat{f}_k^{(s)}-f^{(s)}\|^2\Big\} = \begin{cases} O((\log n)^{-(p-s)/a}) & \text{if } \log n = O(\log m_n)\\ O((\log m_n)^{-(p-s)/a}) & \text{otherwise.}\end{cases}$$

Note that in the ordinary smooth case we obtain the rate of known error density whenever $n^{2((p-s)\vee a)/(2p+2a+1)} = O(m_n)$ which is much less than $n = m$. This is even more visible in the super smooth case, here the rate of known error density is attained even if $m_n = n^r$ for arbitrary small $r > 0$. Moreover, we shall emphasize the influence of the parameter a which characterizes the rate of the decay of the Fourier coefficients of the error density φ. Since a smaller value of a leads to faster rates of convergence, this parameter is often called *degree of ill-posedness* (c.f. Natterer (1984)).

3. A model selection approach: known degree of ill-posedness

Our objective is to construct an adaptive estimator of the deconvolution density f. Adaptation means that in spite of the unknown error density, the estimator should attain the optimal rate of convergence over the ellipsoid $\mathcal{F}_\gamma^r$ for a wide range of different weight sequences γ. However, in this section partial information about the error density φ is supposed to be available. To be precise, we assume that the sequence λ and the value d such that $\varphi \in \mathcal{E}_\lambda^d$ are given in advance. Roughly speaking, this means that the degree of ill-posedness of the underlying inverse problem is known. In what follows, the orthogonal series estimator $\widehat{f}_k$ defined in (1.2) is considered and a procedure to choose the dimension parameter k based on a model selection approach via penalization is constructed. This procedure will only involve the data and λ, d, and ω. First, we introduce sequences of weights which are used below.

Definition 3.1

(i) For all $k \geqslant 1$*, define* $\Delta_k := \max_{0\leqslant|j|\leqslant k}\omega_j/\lambda_j$, $\tau_k := \max_{0\leqslant|j|\leqslant k}(\omega_j)_{\vee 1}/\lambda_j$ *with* $(q)_{\vee 1} := \max(q,1)$ *and*

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

Let further Σ *be a non-decreasing function such that for all* $C > 0$

$$\sum_{k\geqslant 1} C\,\tau_k\exp\Big(-\frac{k\log(\tau_k\vee(k+2))}{3C\log(k+2)}\Big) \leqslant \Sigma(C) < \infty. \tag{3.1}$$

(ii) Define two sequences N and M as follows,

$$\begin{aligned} N_n &:= N_n(\lambda) := \max\{1 \leqslant N \leqslant n \mid \delta_N/n \leqslant \delta_1\}, \\ M_m &:= M_m(\lambda, d) \\ &:= \max\left\{1 \leqslant M \leqslant m \,\middle|\, m^7 \exp\Big(-\frac{m\,\lambda_M}{72d}\Big) \leqslant \Big(\frac{504\,d}{\lambda_1}\Big)^7\right\}. \end{aligned}$$

It is easy to see that there exists always a function Σ satisfying condition (3.1). Consider the orthogonal series estimator $\widehat{f}_k$ defined in (1.2). The adaptive estimator $\widehat{f}_{\widehat{k}}$ is now obtained by choosing the dimension parameter $\widehat{k}$ such that

$$\widehat{k} := \underset{1\leqslant k\leqslant (N_n\wedge M_m)}{\operatorname{argmin}} \left\{-\|\widehat{f}_k\|_\omega^2 + 60\,d\,\frac{\delta_k}{n}\right\}. \tag{3.2}$$

Next, we derive an upper bound for the risk of this adaptive estimator. To this end, we need the following assumption.

Assumption 3.2 *The sequence M satisfies $d^{-1}\min_{1\leqslant|j|\leqslant M_m}\lambda_j \geqslant 2/m$ for all $m \geqslant 1$.*

By construction, this condition is always satisfied for sufficiently large m.

Theorem 3.3 *Assume that we have independent iid. Y- and ε-samples of size n and m, respectively. Consider sequences ω, γ, and λ satisfying Assumption 2.1. Let δ, Δ, N, and M as in Definition 3.1 and suppose that Assumption 3.2 holds. Consider the estimator $\widehat{f}_{\widehat{k}}$ defined in (1.2) with $\widehat{k}$ given by (3.2). Then, there exists a numerical constant $C > 0$ such that for all $n, m \geqslant 1$*

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\left\{\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2\right\} \leqslant C\left\{(d+r)\min_{1\leqslant k\leqslant (N_n\wedge M_m)}\{\max(\omega_k/\gamma_k, \delta_k/n)\} + d\,r\,\kappa_m + d\left[\frac{\delta_1\,(d/\lambda_1)^{7/2}}{m} + \frac{\delta_1 + \Sigma(rd\Lambda)}{n}\right]\right\},$$

where $\Lambda := \sum_{j\in\mathbb{Z}}\lambda_j$ and κ_m is defined in Theorem 2.3.

Comparing the last assertion with the lower bound given in Corollary 2.4, we immediately obtain the following corollary.

Corollary 3.4 *Suppose in addition to the assumptions of Theorem 3.3 that the optimal dimension parameter k_n^* given in Theorem 2.2 is smaller than $N_n \wedge M_m$. If further $\xi := \sup_{k\geqslant 1}\{\delta_k/(\sum_{0<|j|\leqslant k}\omega_j/\lambda_j)\} < \infty$, then there is a numerical constant $C > 0$ such that*

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\left\{\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2\right\} \leqslant C\left\{\xi(d+r)\psi_n + dr\,\kappa_m + d\left[\frac{\delta_1\,(d/\lambda_1)^{7/2}}{m} + \frac{\delta_1 + \Sigma(rd\Lambda)}{n}\right]\right\},$$

where $\Lambda := \sum_{j\in\mathbb{Z}}\lambda_j$ and κ_m is defined in Theorem 2.3.

Under the additional conditions the last assertion establishes the minimax-optimality of the partially adaptive estimator, since its upper risk-bound differs from the optimal one given in Corollary 2.4 only by a constant and negligible terms. However, these additional conditions are not necessary as shown below.

3.1. Illustration: estimation of derivatives (continued)

In section 2.1, we described two different cases where we could choose the model k such that the resulting estimator reached the minimax optimal rate of convergence. The following result shows that in case of unknown error density $\varphi \in \mathcal{E}_\lambda^d$ with a-priori known λ and d, the adaptive estimator automatically attains the optimal rate over a wide range of values for the smoothness parameters.

Proposition 3.5 *Assume that we have independent iid. Y- and ε-samples of size n and m_n, respectively. Consider the estimator $\widehat{f}^{(s)}_{\widehat{k}}$ given in (2.3) with $\widehat{k}$ defined by (3.2).*

[**os**] *In the ordinary smooth case, we have*

$$\Delta_k = k^{2a+2s}, \quad \delta_k \sim k^{2a+2s+1}, \quad N_n \sim n^{1/(2a+2s+1)}, \quad M_{m_n} \sim \Big(\frac{m_n}{\log m_n}\Big)^{1/(2a)}.$$

In case $p - s > a$ we obtain

$$\sup_{f\in\mathcal{W}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\Big\{\mathbb{E}\|\widehat{f}^{(s)}_{\widehat{k}} - f^{(s)}\|^2\Big\} = \begin{cases} O(n^{-2(p-s)/(2p+2a+1)}) & \text{if } n^{2(p-s)/(2p+2a+1)} = O(m_n) \\ O(m_n^{-1}) & \text{otherwise,} \end{cases}$$

and in case $p - s \leqslant a$, if $n^{2a/(2p+2a+1)} = O(m_n)$

$$\sup_{f\in\mathcal{W}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\Big\{\mathbb{E}\|\widehat{f}^{(s)}_{\widehat{k}} - f^{(s)}\|^2\Big\} = \begin{cases} O(n^{-2(p-s)/(2p+2a+1)}) & \text{if } n^{2a/(2p+2a+1)} = O(m_n/\log m_n) \\ O(m_n^{-(p-s)/a}(\log m_n)^{(p-s)/a}) & \text{otherwise,} \end{cases}$$

while if $m_n = o(n^{2a/(2p+2a+1)})$

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

[**ss**] *In the super smooth case, we have*

$$\Delta_k = k^{2s}\exp(k^{2a}), \qquad \delta_k \sim k^{2a+2s+1}\exp(k^{2a})(\log k)^{-1},$$
$$N_n \sim \Big(\log\frac{n\,\log\log n}{(\log n)^{(2a+2s+1)/(2a)}}\Big)^{1/(2a)}, \qquad M_{m_n} \sim \Big(\log\frac{m_n}{\log m_n}\Big)^{1/(2a)}$$

and

$$\sup_{f\in\mathcal{W}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\Big\{\mathbb{E}\|\widehat{f}^{(s)}_{\widehat{k}} - f^{(s)}\|^2\Big\} = \begin{cases} O((\log n)^{-(p-s)/a}) & \text{if } \log n = O(\log m_n) \\ O((\log m_n)^{-(p-s)/a}) & \text{otherwise.} \end{cases}$$

Compare this result with Proposition 2.7. In case [**ss**], the adaptive estimator mimics exactly the behavior of the minimax optimal non-adaptive estimator, even though $\delta_k/(\sum_{0<|j|\leqslant k}\omega_j/\lambda_j) \sim k^{2a+1}/\log k$ is not bounded and hence the assumptions of Corollary 3.4 are violated. In case [**os**], if additionally $p - s > a$, the adaptive estimator still behaves like its minimax optimal non-adaptive counterpart. However, if $p - s \leqslant a$, the sequence $(m_n)_{n\geqslant 1}$ must grow a little faster than in the non-adaptive case. Otherwise, the convergence is slowed down by a logarithmic factor.

4. Unknown degree of ill-posedness

In this section, we dispense with any knowledge about the error density φ, that is, λ and d are not known anymore. We construct an adaptive estimator in this situation as well. Recall that in the previous section, the dimension parameter k was chosen using a criterion function that involved the sequences N, M, and δ which depend on λ and d. We circumvent this problem by defining empirical versions of these three sequences at the beginning of this section. The adaptive estimator is then defined analogously to the one from Section 3, but uses the estimated rather than the original sequences.

Definition 4.1 *Let $\widehat{\delta} := (\widehat{\delta}_k)_{k\geqslant 1}$, $\widehat{N} := (\widehat{N}_n)_{n\geqslant 1}$, and $\widehat{M} := (\widehat{M}_m)_{m\geqslant 1}$ be as follows.*

(i) Given $\widehat{\Delta}_k := \max_{0\leqslant |j|\leqslant k} \frac{\omega_j}{|\widehat{[\varphi]}_j|^2} \mathbb{1}\{|[\widehat{\varphi}]_j|^2 \geqslant 1/m\}$ and $\widehat{\tau}_k := \max_{0\leqslant |j|\leqslant k} \frac{(\omega_j)_{\vee 1}}{|\widehat{[\varphi]}_j|^2} \mathbb{1}\{|[\widehat{\varphi}]_j|^2 \geqslant 1/m\}$ let

$$\widehat{\delta}_k := k\widehat{\Delta}_k \frac{\log(\widehat{\tau}_k \vee (k+2))}{\log(k+2)}.$$

(ii) Given $N_n^u := \operatorname{argmax}_{0<N\leqslant n} \left\{ \max_{0<j\leqslant N} \omega_j/n \leqslant 1 \right\}$ let

$$\widehat{N}_n := \operatorname*{argmin}_{0<|j|\leqslant N_n^u} \left\{ \frac{|\widehat{[\varphi]}_j|^2}{|j|(\omega_j)_{\vee 1}} < \frac{\log n}{n} \right\}, \quad \text{and} \quad \widehat{M}_m := \operatorname*{argmin}_{0<|j|\leqslant m} \left\{ |\widehat{[\varphi]}_j|^2 < \frac{(\log m)^2}{m} \right\}.$$

It worth to stress that all these sequences do not involve any a-priori knowledge about neither the deconvolution density f nor the error density φ. Now, we choose $\widehat{k}$ as

$$\widehat{k} := \operatorname*{argmin}_{0<k\leqslant(\widehat{N}_n\wedge\widehat{M}_m)} \left\{ -\|\widehat{f}_k\|_\omega^2 + 600\,\frac{\widehat{\delta}_k}{n} \right\}. \tag{4.1}$$

Note that $\widehat{k}$ in contrast to the previous section, this choice does not depend on the sequences δ, N, or M, but only on $\widehat{\delta}$, $\widehat{N}$, and $\widehat{M}$, which can be computed from the observed data samples. This choice of the regularization parameter is hence fully data-driven. The constant 600 arising in the definition of $\widehat{k}$, though convenient for deriving the theory, may be far too large in practice and instead be determined by means of a simulation study as in Comte et al. (2006), for example.

In order to show an upper risk bound, we need the following assumption.

Assumption 4.2

(i) The sequences N and M from Definition 3.1 (ii) satisfy the additional conditions

$$\max_{j\geqslant N_n} \frac{\lambda_j}{j(\omega_j)_{\vee 1}} \leqslant \frac{\log n}{4dn} \qquad \text{and} \qquad \max_{j\geqslant M_m} \lambda_j \leqslant \frac{(\log m)^2}{4dm}.$$

(ii) For all $n \in \mathbb{N}$, N_n^u given in Definition 4.1 (ii) fulfills $N_n \leqslant N_n^u \leqslant n$.

By construction, these conditions are always satisfied for sufficiently large n and m. We are now able to state the main result of this paper providing an upper risk bound for the fully adaptive estimator.

Theorem 4.3 *Assume that we have independent iid. Y- and ε-samples of size n and m, respectively. Consider sequences ω, γ, and λ satisfying Assumption 2.1. Let the sequences δ, N, and M be as in Definition 3.1 and suppose that Assumptions 3.2 and 4.2 hold. Define further $N_n^l := \underset{1\leqslant j\leqslant N_n}{\operatorname{argmax}}\{\frac{\lambda_j}{j(\omega_j)\vee 1} \geqslant \frac{4d\,\log n}{n}\}$ and $M_m^l := \underset{1\leqslant j\leqslant M_m}{\operatorname{argmax}}\{\lambda_j \geqslant \frac{4d\,(\log m)^2}{m}\}$. Consider the estimator $\widehat{f}_{\widehat{k}}$ defined in (1.2) with $\widehat{k}$ given by (4.1). Then there exists a numerical constant C such that for all $n, m \geqslant 1$*

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\Big\{\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2\Big\} \leqslant C\bigg\{(r+d\zeta_d)\min_{1\leqslant k\leqslant(N_n^l\wedge M_m^l)}\{\max(\omega_k/\gamma_k, \delta_k/n)\} + d\,r\,\kappa_m + d\,\zeta_d\Big[\frac{(\delta_1+r)(d/\lambda_1)^7}{m} + \frac{\delta_1+\Sigma(rd\Lambda\zeta_d)}{n}\Big]\bigg\},$$

where $\Lambda := \sum_{j\in\mathbb{Z}}\lambda_j$, $\zeta_d := \log 3d/\log 3$, and κ_m is defined in Theorem 2.3.

Comparing the last assertion with Theorem 3.3, we assert that surprisingly, the estimation of the sequences δ, N, and M essentially changes the upper bound only by replacing N and M by N^l and M^l, respectively. Therefore, in analogy to the results in section 3, we have the following corollary.

Corollary 4.4 *Suppose that in addition to the assumptions of Theorem 4.3 we have that the optimal dimension parameter k_n^* given in Theorem 2.2 is smaller than $N_n^l \wedge M_m^l$. If further $\xi := \sup_{k\geqslant 1}\{\delta_k/(\sum_{0<|j|\leqslant k}\omega_j/\lambda_j)\} < \infty$, then there is a numerical constant $C > 0$ such that*

$$\sup_{f\in\mathcal{F}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\Big\{\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2\Big\} \leqslant C\bigg\{\xi(d\zeta_d+r)\psi_n + dr\,\kappa_m + d\,\zeta_d\Big[\frac{(\delta_1+r)(d/\lambda_1)^7}{m} + \frac{\delta_1+\Sigma(rd\Lambda\zeta_d)}{n}\Big]\bigg\},$$

where $\Lambda := \sum_{j\in\mathbb{Z}}\lambda_j$, $\zeta_d := \log 3d/\log 3$, and κ_m is defined in Theorem 2.3.

Under the additional conditions the last assertion establishes the minimax-optimality of the fully adaptive estimator, since its upper risk-bound differs from the optimal one given in Corollary 2.4 only by a constant and negligible terms. However, these additional conditions are not necessary as shown below.

4.1. Illustration: estimation of derivatives (continued)

The following result shows that even without any prior knowledge on the error density φ, the fully adaptive penalized estimator automatically attains the optimal rate in the super smooth case and in the ordinary smooth case as far as $p-s \geqslant a$. Recall that the computation of the dimension parameter $\widehat{k}$ given in (4.1) involves the sequence $(N_n^u)_{n\geqslant 1}$, which in our illustration satisfies $N_n^u \sim n^{1/(2s)}$ since $\omega_j = |j|^{2s}$, $j \geqslant 1$.

Proposition 4.5 *Assume that we have independent iid. Y- and ε-samples of size n and m, respectively. Consider the estimator $\widehat{f}_{\widehat{k}}^{(s)}$ given in (2.3) with $\widehat{k}$ defined by (4.1).*

[os] *In the ordinary smooth case with $p - s > a$ we obtain*

$$\sup_{f\in\mathcal{W}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\Big\{\mathbb{E}\|\widehat{f}_{\widehat{k}}^{(s)} - f^{(s)}\|^2\Big\} = \begin{cases} O(n^{-2(p-s)/(2p+2a+1)}) & \text{if } n^{2(p-s)/(2p+2a+1)} = O(m_n) \\ O(m_n^{-1}) & \text{otherwise,} \end{cases}$$

and with $p-s \leqslant a$, if $n^{2a/(2p+2a+1)} = O(m_n)$

$$\sup_{f\in\mathcal{W}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\Big\{\mathbb{E}\|\widehat{f}^{(s)}_{\widehat{k}} - f^{(s)}\|^2\Big\} = \begin{cases} O(n^{-2(p-s)/(2p+2a+1)}) & \text{if } n^{2a/(2p+2a+1)} = O(m_n/(\log m_n)^2) \\ O(m_n^{-(p-s)/a}(\log m_n)^{2(p-s)/a}) & \text{otherwise,}\end{cases}$$

while if $m_n = o(n^{2a/(2p+2a+1)})$

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

[ss] *In the super smooth case, we have*

$$\sup_{f\in\mathcal{W}_\gamma^r}\sup_{\varphi\in\mathcal{E}_\lambda^d}\Big\{\mathbb{E}\|\widehat{f}^{(s)}_{\widehat{k}} - f^{(s)}\|^2\Big\} = \begin{cases} O((\log n)^{-(p-s)/a}) & \text{if } \log n = O(\log m_n) \\ O((\log m_n)^{-(p-s)/a}) & \text{otherwise.}\end{cases}$$

Notice that the last result differs from Proposition 3.5 solely in case **[os]** with $p-s \leqslant a$, where $(\log m_n)$ is replaced by $(\log m_n)^2$. Hence, in all other cases the fully adaptive estimator attains the minimax optimal rate. In particular, it is not necessary to know in advance if the error density is ordinary or super smooth. Moreover, as long as $m_n \sim n$, the fully adaptive estimator always attains the same optimal rate as in case of known error density. However, over a wide range of values for the smoothness parameters, the minimax optimal rate is still obtained even when m_n grows slower than n.

A. Proofs

A.1. Proofs of section 2

Lower bounds

Proof of Theorem 2.2. Given $\zeta := \eta\min(r-1, 1/(8d\Gamma))$ and $\alpha_n := \psi_n(\sum_{0<|j|\leqslant k_n^*}\omega_j/(\lambda_j n))^{-1}$ we consider the function $f := 1 + (\zeta\alpha_n/n)^{1/2}\sum_{0<|j|\leqslant k_n^*}\lambda_j^{-1/2}e_j$. We will show that for any $\theta := (\theta_j) \in \{-1,1\}^{2k_n^*}$, the function $f_\theta := 1 + \sum_{0<|j|\leqslant k_n^*}\theta_j[f]_j e_j$ belongs to $\mathcal{F}_\gamma^r$ and is hence a possible candidate of 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 $0 < |j| \leqslant k_n^*$ 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

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

Below we show furthermore that for all $n \geqslant 2$ we have

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

Combining the last lower bound and the reduction scheme gives

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

Hence, employing 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 $f \in \mathcal{F}_\gamma^r$. In order to show that $f \in \mathcal{F}_\gamma^r$, we first notice that f integrates to one. Moreover, f is non-negative because $|\sum_{0<|j|\leqslant k_n^*}[f]_j e_j| \leqslant 1$, and $\|f\|_\gamma^2 \leqslant r$, which can be realized as follows. By employing the condition $\sum_{j\in\mathbb{Z}}\gamma_j^{-1} = \Gamma < \infty$ we have

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

Since ω/γ is non-increasing the definition of ζ, α_n and η implies

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

as well as $\|f\|_\gamma^2 \leqslant 1 + \zeta\frac{\gamma_{k_n^*}}{\omega_{k_n^*}}\alpha_n\Big(\sum_{0<|j|\leqslant k_n^*}\frac{\omega_j}{n\lambda_j}\Big) \leqslant 1 + \zeta/\eta \leqslant r$.
It remains to show (A.1). Consider the Hellinger affinity $\rho(g_\theta^n, g_{\theta^{(j)}}^n) = \int\sqrt{g_\theta^n}\sqrt{g_{\theta^{(j)}}^n}$, then we obtain for any estimator $\widetilde{f}$ of f that

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

Rewriting the last estimate we obtain

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

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

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

where we have used that $\alpha_n \leqslant 1/\eta$, $\varphi \in \mathcal{E}_\lambda^d$ and $g_\theta \geqslant 1/2$ because $|\sum_{0<|j|\leqslant k_n^*}[g_\theta]_j e_j| \leqslant 1/2$, which can be realized as follows. By 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

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

Therefore, the definition of ζ implies $H^2(g_\theta, g_{\theta(j)}) \leqslant 2/n$. By using the independence, i.e., $\rho(g_\theta^n, g_{\theta(j)}^n) = \rho(g_\theta, g_{\theta(j)})^n$, together with the identity $\rho(g_\theta, g_{\theta(j)}) = 1 - \frac{1}{2}H^2(g_\theta, g_{\theta(j)})$ it follows $\rho(g_\theta^n, g_{\theta(j)}^n) \geqslant (1-n^{-1})^n \geqslant 1/4$ for all $n \geqslant 2$. By combination of 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 $1 \geqslant (1-\alpha_m)^2 \geqslant (1-(1-1/d^{1/4}))^2 \geqslant 1/d^{1/2}$ and $1 \leqslant (1+\alpha_m)^2 \leqslant (1+(1-1/d^{1/4}))^2 = (2-1/d^{1/4})^2 \leqslant d^{1/2}$, which implies $1/d^{1/2} \leqslant (1+\theta\alpha_m)^2 \leqslant d^{1/2}$. These inequalities will be used below without further reference. By assumption there is a density $\varphi \in \mathcal{E}_\lambda^{\sqrt{d}}$ such that $\varphi \geqslant 1/2$. 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 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 hence $g_1 = g_{-1}$. We denote by g_θ^n the joint density of an i.i.d. n-sample from g_θ and φ_θ^m the joint density of an i.i.d. m-sample from φ_θ. Since the samples are independent from each other, $p_\theta := g_\theta^n\varphi_\theta^m$ 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

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

Below we show furthermore that for all $m \geqslant 2$ we have

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

Moreover, we have $\|f_1 - f_{-1}\|^2 = 4\alpha_m^2\omega_{k_m^*}\gamma_{k_m^*}^{-1}\frac{(r-1)\wedge 1}{d^{1/2}} = 4\frac{(r-1)\wedge 1}{d^{1/2}}\zeta^2\omega_{k_m^*}\gamma_{k_m^*}^{-1}\min\Big(1, \frac{1}{m\lambda_{k_m^*}}\Big)$. 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. In order to show $f_\theta \in \mathcal{F}_\gamma^r$, we first observe that f_θ integrates to one. Moreover, f_θ is non-negative because $|(1-\theta\alpha_m)\frac{1\wedge\sqrt{r-1}}{d^{1/4}}\gamma_{k_m^*}^{-1/2}| \leqslant \gamma_{k_m^*}^{-1/2} \leqslant 1$ and $\|f_\theta\|_\gamma^2 = 1 + \gamma_{k_m^*}|[f_\theta]_{k_m^*}|^2 \leqslant 1 + \gamma_{k_n^*}|(1-\theta\alpha_m)\frac{1\wedge\sqrt{r-1}}{d^{1/4}}\gamma_{k_m^*}^{-1/2}|^2 \leqslant r$. Consider φ_θ which obviously integrates to one. Furthermore, as $\varphi \geqslant 1/2$ the function $\varphi_\theta = \varphi + \theta\alpha_m[\varphi]_{k_m^*}e_{k_m^*}$ is non-negative since $|\theta\alpha_m[\varphi]_{k_m^*}e_{k_m^*}| \leqslant \alpha_m\lambda_{k_m^*}^{1/2}d^{1/2} \leqslant \zeta m^{-1/2}\sqrt{d} \leqslant 1/2$ by using the definition of α_m and ζ. To check that $\varphi_\theta \in \mathcal{E}_\lambda^d$, it remains to show that $1/d \leqslant [\varphi_\theta]_j^2/\lambda_j \leqslant d$ for all $|j| > 0$. Since $\varphi \in \mathcal{E}_\lambda^{\sqrt{d}}$, it follows from the definition of φ_θ that these inequalities are satisfied for all $j \neq k_m^*$ and moreover that

$1/d \leqslant \frac{|[\varphi]_{k_m^*}|^2}{\sqrt{d}\lambda_{k_m^*}} \leqslant \frac{(1+\theta\alpha_m)^2|[\varphi]_{k_m^*}|^2}{\lambda_{k_m^*}} \leqslant \frac{\sqrt{d}|[\varphi]_{k_m^*}|^2}{\lambda_{k_m^*}} \leqslant d$. Finally consider (A.4). As in the proof of Theorem 2.2 by employing the Hellinger affinity $\rho(p_1, p_{-1})$ we obtain for any estimator $\widetilde{f}$ of f that

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

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

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

where we used that $\varphi \geqslant 1/2$ 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 $\mathbb{V}\mathrm{ar}([\widehat{g}]_j) \leqslant 1/n$ and $\mathbb{V}\mathrm{ar}([\widehat{\varphi}]_j) \leqslant 1/m$ for all $j \in \mathbb{Z}$. Moreover, by applying Theorem 2.10 in Petrov (1995) there exists a constant $C > 0$ such that $\mathbb{E}|[\widehat{\varphi}]_j - [\varphi]_j|^4 \leqslant C/m^2$ for all $j \in \mathbb{Z}$ and $m \in \mathbb{N}$. These results are used below without further reference. Define now $\widetilde{f} := 1 + \sum_{0<|j|\leqslant k_n^*} [f]_j \mathbb{1}\{|[\widehat{\varphi}]_j|^2 \geqslant 1/m\} e_j$ and decompose the risk into two terms,

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

which we bound separately. Consider first A which we decompose further,

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

By using the elementary inequality $1/2 \leqslant |[\widehat{\varphi}]_j/[\varphi]_j - 1|^2 + |[\widehat{\varphi}]_j/[\varphi]_j|^2$, 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

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

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

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

Consider now B which we decompose further into

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

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

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

which we will show below. The result of the theorem follows now by combination of the decomposition (A.5) and the estimates of A_1, A_2, B_1 and B_2.
To conclude, let us prove (A.6). If $|[\varphi]_j|^2 \geqslant 4/m$, then we deduce by employing Tchebychev's inequality that

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

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

Illustration: estimation of derivatives

Proof of Proposition 2.6. Since for each $0 \leqslant s \leqslant p$ 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 Corollary 2.4. In both cases the additional conditions formulated in Theorem 2.2 and 2.3 are easily verified. Therefore, 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 $k_n^* := \mathrm{argmin}_{j\in\mathbb{N}}\{\max(\frac{\omega_j}{\gamma_j}, \sum_{0<|l|\leqslant j} \frac{\omega_l}{n\lambda_l})\}$ satisfies $n\omega_{k_n^*}/\gamma_{k_n^*} \sim \sum_{0<|l|\leqslant k_n^*} \omega_l/\lambda_l$ since both sequences (γ_j/ω_j) and $(\sum_{0<|l|\leqslant j} \frac{\omega_l}{n\lambda_l})$ are non-increasing.
[os] The well-known approximation $\sum_{j=1}^m j^r \sim m^{r+1}$ for $r > 0$ implies
$(\gamma_{k_n^*}/\omega_{k_n^*}) \sum_{0<|l|\leqslant k_n^*} \omega_l/\lambda_l \sim (k_n^*)^{2a+2p+1}$. It follows that $k_n^* \sim n^{1/(2p+2a+1)}$ and the first lower bound writes $\psi_n \sim n^{-(2p-2s)/(2p+2a+1)}$. Moreover, we have $\kappa_m \sim m^{-([p-s]\wedge a)/a}$, since the minimum in $\kappa_m = \sup_{j\in\mathbb{Z}}\{|j|^{-2(p-s)} \min(1, |j|^{2a}/m)\}$ is equal to one for $|j| \geqslant m^{1/2a}$ and $|j|^{-2(p-s)}$ is non-increasing.
[ss] Applying Laplace's Method (c.f. chapter 3.7 in Olver (1974)) we have
$(\gamma_{k_n^*}/\omega_{k_n^*}) \sum_{0<|l|\leqslant k_n^*} \omega_l/\lambda_l \sim (k_n^*)^{2p} \exp(|k_n^*|^{2a})$ 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

$\kappa_m \sim (\log m)^{-(p-s)/a}$ since the minimum in $\kappa_m = \sup_{j\in\mathbb{Z}}\{|j|^{-2(p-s)} \min(1, \exp(|j|^{2a})/m)\}$ is equal to one for $|j| \geqslant (\log m)^{(1/2a)}$ and $|j|^{-2(p-s)}$ is non-increasing. Consequently, the lower bounds in Proposition 2.7 follow by applying Corollary 2.4. □

Proof of Proposition 2.7. Since in both cases the condition on the dimension parameter k ensures that $k \sim k_n^*$ (see the proof of Proposition 2.6) the result follows from Theorem 2.5. □

A.2. Proofs of section 3

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

$$\widehat{f}_k = \sum_{j=-k}^{k} \frac{\widehat{[g]}_j}{\widehat{[\varphi]}_j}\mathbb{1}\{|\widehat{[\varphi]}_j|^2 \geqslant 1/m\}e_j, \quad \widetilde{f}_k := \sum_{j=-k}^{k} \frac{\widehat{[g]}_j}{[\varphi]_j}e_j, \quad f_k := \sum_{j=-k}^{k} \frac{[g]_j}{[\varphi]_j}e_j,$$
$$\widehat{\Phi}_u := \sum_{j\in\mathbb{Z}} \frac{[u]_j}{\widehat{[\varphi]}_j}\mathbb{1}\{|\widehat{[\varphi]}_j|^2 \geqslant 1/m\}e_j, \quad \widetilde{\Phi}_u := \sum_{j\in\mathbb{Z}} \frac{[u]_j}{[\varphi]_j}e_j.$$

Furthermore, let $\widehat{g}$ be the function with Fourier coefficients $[\widehat{g}]_j := \widehat{[g]}_j$. Given $1 \leqslant k \leqslant k'$ we have then for all $t \in \mathcal{S}_k := \operatorname{span}\{e_{-k}, \dots, e_k\}$

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

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

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

At the end of this section we will prove three technical Lemmata (A.2, A.4 and A.3) which are used in the following proof.

Proof of Theorem 3.3. We consider the contrast

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

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

$$\arg\min_{t\in\mathcal{S}_k} \Upsilon(t) = \widehat{f}_k, \quad \forall\, k \geqslant 1. \quad \text{(A.8)}$$

Moreover, the adaptive choice $\widehat{k}$ of the dimension parameter can be rewritten as

$$\widehat{k} = \underset{1\leqslant k\leqslant (N_n\wedge M_m)}{\operatorname{argmin}} \left\{ \Upsilon(\widehat{f}_k) + 60\frac{d\delta_k}{n} \right\}. \tag{A.9}$$

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

$$\Upsilon(\widehat{f}_{\widehat{k}}) + \operatorname{pen}(\widehat{k}) \leqslant \Upsilon(\widehat{f}_k) + \operatorname{pen}(k) \leqslant \Upsilon(f_k) + \operatorname{pen}(k),$$

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

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

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

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

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

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

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

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

Decompose $|\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu\rangle_\omega|^2 = |\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu\rangle_\omega|^2 \mathbb{1}\{\Omega_q\} + |\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu\rangle_\omega|^2 \mathbb{1}\{\Omega_q^c\}$ further using

$$\Omega_q := \left\{ \forall\, 0 < |j| \leqslant M_m \;:\; \left|\frac{1}{\widehat{[\varphi]}_j} - \frac{1}{[\varphi]_j}\right| \leqslant \frac{1}{2|[\varphi]_j|} \;\wedge\; |\widehat{[\varphi]}_j|^2 \geqslant 1/m \right\}. \tag{A.11}$$

Since $\mathbb{1}\{|\widehat{[\varphi]}_j|^2 \geqslant 1/m\}\mathbb{1}\{\Omega_q\} = \mathbb{1}\{\Omega_q\}$, it follows that for all $1 \leqslant |j| \leqslant (N_n \wedge M_m)$ we have

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

Hence, $\sup_{t\in\mathcal{B}_k} |\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu\rangle_\omega|^2\, \mathbb{1}\{\Omega_q\} \leqslant \frac{1}{4}\sup_{t\in\mathcal{B}_k} |\langle t, \widetilde{\Phi}_\nu\rangle_\omega|^2$ for all $1 \leqslant k \leqslant (N_n \wedge M_m)$. Letting $\tau := 1/8$ it follows from $\|\widehat{f}_{\widehat{k}} - f_k\|_\omega^2 \leqslant 2\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 + 2\|f_k - f\|_\omega^2$ that

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

Since ω/γ is non-increasing we obtain $\|f - f_k\|_\omega^2 \leqslant r\omega_k/\gamma_k$ for all $f \in \mathcal{F}_\gamma^r$. Furthermore, notice that $60\, d\, \delta_{k\vee\widehat{k}}/n = \mathrm{pen}(k \vee \widehat{k}) \leqslant \mathrm{pen}(k) + \mathrm{pen}(\widehat{k})$. By taking the expectation on both sides we conclude that there exists a numerical constant $C > 0$ such that

$$\sup_{f\in\mathcal{F}_\gamma^r} \sup_{\varphi\in\mathcal{R}_\lambda^d} \mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \leqslant C\,(d+r) \min_{1\leqslant k\leqslant(N_n\wedge M_m)} \Big\{\max\Big(\frac{\omega_k}{\gamma_k}, \frac{\delta_k}{n}\Big)\Big\}$$
$$+ C \sup_{f\in\mathcal{F}_\gamma^r} \sup_{\varphi\in\mathcal{R}_\lambda^d} \sum_{1\leqslant|k'|\leqslant(N_n\wedge M_m)} \mathbb{E}\Big(\sup_{t\in\mathcal{B}_{k'}} |\langle t, \widetilde{\Phi}_\nu\rangle_\omega|^2 - \big(6\, d\delta_{k'}\big)/n\Big)_+$$
$$+ C \sup_{f\in\mathcal{F}_\gamma^r} \sup_{\varphi\in\mathcal{R}_\lambda^d} \mathbb{E}\Big[\sup_{t\in\mathcal{B}_{(N_n\wedge M_m)}} |\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu\rangle_\omega|^2\, \mathbb{1}\{\Omega_q^c\}\Big]$$
$$+ C \sup_{f\in\mathcal{F}_\gamma^r} \sup_{\varphi\in\mathcal{R}_\lambda^d} \mathbb{E}\Big[\sup_{t\in\mathcal{B}_{(N_n\wedge M_m)}} |\langle t, \widehat{\Phi}_g - \widetilde{\Phi}_g\rangle_\omega|^2\Big].$$

In order to bound the second term, apply Lemma A.2 with $\delta_k^* = d\,\delta_k$ and $\Delta_k^* = d\,\Delta_k$. Due to the properties of N_n and of the function Σ from Definition 3.1, there is a numerical constant $C > 0$ such that

$$\sum_{k=1}^{N_n} \mathbb{E}\Big(\sup_{t\in\mathcal{B}_k} |\langle t, \widetilde{\Phi}_\nu\rangle_\omega|^2 - 6\,\frac{d\,\delta_k}{n}\Big)_+ \leqslant \frac{Cd}{n}\Big\{\delta_1 + \Sigma(\|\varphi\|^2\, \|f\|^2)\Big\}.$$

It is readily verified that $\|\varphi\|^2 \leqslant d\Lambda$ for all $\varphi \in \mathcal{E}_\lambda^d$ and $\|f\|^2 \leqslant r$ for all $f \in \mathcal{F}_\gamma^r$. The result follows now by virtue of Lemma A.3, A.4, A.5, and Definition 3.1 (i). $\square$

In the proof of Lemma A.2 below we will need the following Lemma, which can be found in Comte et al. (2006).

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

$$\mathbb{E}[\sup_{r\in\mathcal{R}} |\nu_n^*(r)|^2 - 6H_2^2]_+ \leqslant C\left(\frac{v}{n}\exp(-(nH_2^2/6v)) + \frac{H_1^2}{n^2}\exp(-K_2(nH_2/H_1))\right)$$

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

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

Lemma A.2 *Let $(\delta_k^*)_{k\in\mathbb{Z}}$ and $(\Delta_k^*)_{k\in\mathbb{Z}}$ be sequences such that*

$$\delta_k^* \geqslant \sum_{0\leqslant|j|\leqslant k} \frac{\omega_j}{|[\varphi]_j|^2} \qquad \text{and} \qquad \Delta_k^* \geqslant \max_{0\leqslant|j|\leqslant k} \frac{\omega_j}{|[\varphi]_j|^2}$$

and let $K_2 := (\sqrt{2}-1)/(21\sqrt{2})$. Then, there is a numerical constant $C > 0$ such that

$$\sum_{k=1}^{N_n} \mathbb{E}\Big[\Big(\sup_{t\in\mathcal{B}_k} |\langle t, \widetilde{\Phi}_\nu\rangle_\omega|^2 - \frac{6\,\delta_k^*}{n}\Big)_+\Big]$$
$$\leqslant C\left\{\frac{\|\varphi\|^2\,\|f\|^2}{n}\sum_{k=1}^{N_n} \Delta_k^* \exp\left(-\frac{1}{6\,\|\varphi\|^2\,\|f\|^2}(\delta_k^*/\Delta_k^*)\right) + \frac{1}{n^2}\exp\left(-K_2\sqrt{n}\right)\sum_{k=1}^{N_n}\delta_k^*\right\}.$$

Proof. For $t \in \mathcal{S}_k$ define the function $r_t := \sum_{0\leqslant|j|\leqslant k} \omega_j [t]_j \overline{[\varphi]_j}^{-1} e_j$, then it is readily seen that $\langle t, \widetilde{\Phi}_\nu \rangle_\omega = \frac{1}{n}\sum_{k=1}^n r_t(Y_k) - \mathbb{E}[r_t(Y_k)]$. Next, we compute constants H_1, H_2, and v verifying the three inequalities required in Lemma A.1, which then implies the result. Consider H_1 first:

$$\sup_{t\in\mathcal{B}_k} \|r_t\|_\infty^2 = \sup_{y\in\mathbb{R}} \sum_{0\leqslant|j|\leqslant k} \omega_j |[\varphi]_j|^{-2}\ |e_j(y)|^2 = \sum_{0\leqslant|j|\leqslant k} \omega_j |[\varphi]_j|^{-2} \leqslant \delta_k^* =: H_1^2.$$

Next, find H_2. Notice that

$$\mathbb{E}[\sup_{t\in\mathcal{B}_k} |\langle t, \widetilde{\Phi}_\nu \rangle_\omega|^2] = \frac{1}{n} \sum_{0\leqslant|j|\leqslant k} \omega_j |[\varphi]_j|^{-2}\ \mathbb{V}\mathrm{ar}(e_j(Y_1)).$$

As $\mathbb{V}\mathrm{ar}(e_j(Y_1)) \leqslant \mathbb{E}[|\, e_j(Y_1)\,|^2] = 1$, we define $\mathbb{E}[\sup_{t\in\mathcal{B}_k} |\langle t, \widetilde{\Phi}_\nu \rangle|^2] \leqslant \delta_k^*/n =: H_2^2$. 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 $A_k := \left(\overline{[\varphi]_j}^{-1} [\varphi]_{j'}^{-1} [\varphi]_{j-j'}\, [f]_{j-j'}\right)_{j,j'=-k,\dots,k}$. Straightforward algebra shows $\sup_{t\in\mathcal{B}_k} \mathbb{V}\mathrm{ar}(r_t(Y_1)) \leqslant \sup_{t\in\mathcal{B}_k} \langle A_k D_k(\omega)\ \underline{[t]}, D_k(\omega)\underline{[t]}\rangle_{\mathbb{C}^{2k+1}}$, hence

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

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

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

We have that $\|D_k(\sqrt{\omega}\ [\varphi]^{-1})\|_{\mathbb{C}^{2k+1}}^2 = \max_{0\leqslant|j|\leqslant k} \omega_j |[\varphi]_j|^{-2} \leqslant \Delta_k^*$. 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 $\|Bz\|_{\ell^2}^2 \leqslant \|\varphi\|^2\ \|f\|^2$, and hence $\|B\|_{\ell^2}^2 \leqslant \|\varphi\|^2\ \|f\|^2$. 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 the matrix representation B_k via the isomorphism $\mathcal{S}_k \cong \mathbb{C}^{2k+1}$ and hence $\|\Pi_k\, B\, \Pi_k\|_{\ell^2} = \|B_k\|_{\mathbb{C}^{2k+1}}$. Orthogonal projections having a norm bounded by 1, we conclude that $\|B_k\|_{\mathbb{C}^{2k+1}} \leqslant \|B\|_{\ell^2}$ for all $k \in \mathbb{N}$, which implies $\sup_{t\in\mathcal{B}_k} \frac{1}{n}\sum_{k=1}^n \mathbb{V}\mathrm{ar}(r_t(Y_k)) \leqslant \|\varphi\|^2\ \|f\|^2\ \Delta_k^* =: v$ and completes the proof. □

Lemma A.3 *There is a numerical constant $C > 0$ such that for every $k, m \in \mathbb{N}$*

$$\mathbb{E}\left[\sup_{t\in\mathcal{B}_k} |\langle t, \widehat{\Phi}_g - \widetilde{\Phi}_g \rangle_\omega|^2\right] \leqslant C\ d\ r\ \kappa_m(\gamma, \lambda, \omega).$$

Proof. Firstly, as $f \in \mathcal{F}_\gamma^r$, it is easily seen that

$$\mathbb{E}\left[\sup_{t\in\mathcal{B}_k^0} |\langle t, \widehat{\Phi}_g - \widetilde{\Phi}_g \rangle_\omega|^2\right] \leqslant r \sup_{0<|j|\leqslant k} \frac{\omega_j}{\gamma_j}\, \mathbb{E}[|R_j|^2],$$

where R_j is defined by

$$R_j := \left(\frac{[\varphi]_j}{\widehat{[\varphi]}_j} \mathbb{1}\{|\widehat{[\varphi]}_j|^2 \geqslant 1/m\} - 1 \right). \tag{A.12}$$

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

$$\mathbb{E}|R_j|^2 = \mathbb{E}\left[\left| \frac{[\varphi]_j}{\widehat{[\varphi]}_j} - 1 \right|^2 \mathbb{1}\{|\widehat{[\varphi]}_j|^2 \geqslant 1/m\} \right] + \mathbf{P}[|\widehat{[\varphi]}_j|^2 < 1/m] =: R_j^I + R_j^{II}. \tag{A.13}$$

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

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

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

$$R_j^I = \mathbb{E}\left[\frac{|\widehat{[\varphi]}_j - [\varphi]_j|^2}{|\widehat{[\varphi]}_j|^2}\ \mathbb{1}\{|\widehat{[\varphi]}_j|^2 \geqslant 1/m\} \right] \leqslant m\, \mathbb{V}\mathrm{ar}(|\widehat{[\varphi]}_j) \leqslant 1. \tag{A.14}$$

On the other hand, using that $\mathbb{E}[|\widehat{[\varphi]}_j - [\varphi]_j|^4] \leqslant c/m^2$ for some numerical constant $c > 0$ (cf. Petrov (1995), Theorem 2.10), we obtain

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

Combining with (A.14) gives $R_j^I \leqslant 2(c+1) \min\left\{1, \frac{1}{m|[\varphi]_j|^2}\right\}$, which completes the proof. □

Lemma A.4 *There is a numerical constant $C > 0$ such that*

$$\mathbb{E}\left[\sup_{t \in \mathcal{B}_{(N_n \wedge M_m)}} |\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu \rangle_\omega \mathbb{1}\{\Omega_q^c\}|^2 \right] \leqslant C d \delta_1 (\mathbf{P}[\Omega_q^c])^{(1/2)}.$$

Proof. Given with R_j from (A.12) we begin our proof observing that

$$\mathbb{E}\left[\sup_{t \in \mathcal{B}_{M_m}} |\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu \rangle_\omega\, \mathbb{1}\{\Omega_q^c\}|^2 \right] \leqslant \frac{1}{n} \sum_{0<|j|\leqslant (N_n \wedge M_m)} \frac{\omega_j}{|[\varphi]_j|^2}\, \mathbb{E}[|R_j|^2 \mathbb{1}\{\Omega_q^c\}],$$

and using the independence of the two samples and $\mathbb{V}\mathrm{ar}(\widehat{[g]}_j) \leqslant n^{-1}$. Since $d\delta_k \geqslant \sum_{0<|j|\leqslant k} \frac{\omega_j}{|[\varphi]_j|^2}$ for all $\varphi \in \mathcal{E}_\lambda^d$, the Cauchy-Schwarz inequality yields

$$\mathbb{E}\left[\sup_{t \in \mathcal{B}_{M_m}} |\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu \rangle_\omega\, \mathbb{1}\{\Omega_q^c\}|^2 \right] \leqslant\ d\,(\mathbf{P}[\Omega_q^c])^{1/2}\ \frac{\delta_{N_n}}{n} \max_{0<|j|\leqslant N_n} (\mathbb{E}[|R_j|^4])^{1/2}.$$

Proceeding analogously to (A.13) and (A.14), there exists a numerical constant C such that $\mathbb{E}[|R_j|^4] \leqslant C$. The result follows then by Definition 3.1 (ii). $\square$

Lemma A.5 *Consider the event Ω_q defined in (A.11). We have $\mathbf{P}[\Omega_q^c] \leqslant 4(504\, d/\lambda_1)^7 m^{-6}$ for all $m \geqslant 1$.*

Proof. Consider the complement of Ω_q given by

$$\Omega_q^c = \Big\{ \exists\, 0 < |j| \leqslant M_m \;:\; \Big| \frac{[\varphi]_j}{\widehat{[\varphi]}_j} - 1 \Big| > \frac{1}{2} \;\vee\; |\widehat{[\varphi]}_j|^2 < 1/m \Big\}.$$

It follows from Assumption 3.2 that $|[\varphi]_j|^2 \geqslant 2/m$ for all $0 < |j| \leqslant M_m$. This yields

$$\Omega_q^c \subseteq \Big\{ \exists\, 0 < |j| \leqslant M_m \;:\; \Big| \frac{\widehat{[\varphi]}_j}{[\varphi]_j} - 1 \Big| > \frac{1}{3} \Big\}.$$

By Hoeffding's inequality,

$$\mathbf{P}[|\widehat{[\varphi]}_j/[\varphi]_j - 1| > 1/3] \leqslant 2 \exp\Big(- \frac{m\, |[\varphi]_j|^2}{72} \Big) \tag{A.15}$$

which implies the result by employing the definition of M_m. $\square$

Illustration: estimation of derivatives

Proof of Proposition 3.5. In the light of the proof of Proposition 2.6 we apply Theorem 3.3, where in both cases the additional conditions are easily verified and the result follows by an evaluation of the upper bound.

[os] Let $k_n^* := n^{1/(2a+2p+1)}$ and note that $k_n^* \lesssim N_n$. Thus, the upper bound is

$$(k_n^* \wedge M_{m_n})^{-2(p-s)} + m_n^{-(1 \wedge ((p-s)/a))}. \tag{A.16}$$

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

$$\frac{k_n^*}{M_{m_n}} = \frac{n^{1/(2a+2p+1)}}{\Big(\frac{m_n}{\log m_n}\Big)^{(1/2a)}} = \frac{n^{1/(2a+2p+1)}}{m_n^{1/2(p-s)}} \frac{m_n^{1/2(p-s)} (\log m_n)^{1/2a}}{m_n^{1/2a}} = o(1).$$

This means that $k_n^* \lesssim M_{m_n}$, so the resulting upper bound is $(k_n^*)^{-2(p-s)} + m_n^{-1} \lesssim (k_n^*)^{-2(p-s)}$. Suppose now that $m_n = o(n^{2(p-s)/(2p+2a+1)})$. If in addition $k_n^* = O(M_{m_n})$, then the first summand in (A.16) reduces to $(k_n^*)^{-2(p-s)}$ and hence the upper bound is m_n^{-1}. On the other hand, if $M_{m_n}/k_n^* = o(1)$, then the first term is $(M_n)^{-2(p-s)} \lesssim M_{m_n}^{-2a} (\log m_n)^{-1} = m_n^{-1}$, since $p - s > a$. Combining both cases, we obtain the result in case $p - s > a$.
Now assume $p - s \leqslant a$. First, suppose that $k_n^* = O(M_{m_n})$. Then, then the first summand in (A.16) reduces to $(k_n^*)^{-2(p-s)}$ and moreover $n^{2a/(2p+2a+1)} = O(m_n)$. Therefore, the upper bound is $(k_n^*)^{-2(p-s)}$. Consider now $M_{m_n} = o(k_n^*)$. Then (A.16) can be rewritten as $(m_n/\log m_n)^{-(p-s)/a} + m_n^{-(p-s)/a}$ which results in the rate $(m_n/\log m_n)^{-(p-s)/a}$. Combining both cases gives the result. More precisely, $m_n = o(n^{2a/(2p+2a+1)})$ implies $M_{m_n} = o(k_n^*)$. On the other hand, in case $n^{2a/(2p+2a+1)} = O(m_n)$, if $k_n^*/M_{m_n} = O(1)$, then the rate is $(k_n^*)^{-2p}$, while if $M_{m_n}/k_n^* = o(1)$, we have the rate $(m_n/\log m_n)^{-p/a}$.

[ss] Choose $k_n^* \sim (\log n)^{1/2a}(1+o(1))$. And note that $N_n \sim (\log n)^{1/2a}(1+o(1))$ and $M_{m_n} \sim (\log m_n)^{1/2a}(1+o(1))$. The upper risk bound is now $(k_n^* \wedge M_{m_n})^{-2p} + (\log m_n)^{p/a}$. Consider two cases. Firstly, $\log n/\log m_n = O(1)$. This implies $N_n/M_{m_n} = O(1)$ and hence $k_n^*/M_{m_n} = O(1)$. This means that the upper bound is in fact $(k_n^*)^{-2p} + (\log m_n)^{-p/a} \sim (\log n)^{-p/a}$. In the case $\log m_n/\log n = o(1)$, an analogous argument proves the claim, which completes the proof. □

A.3. Proofs of section 4

Proof of Theorem 4.3. Define $\Delta_k^\varphi := \max_{0\leqslant|j|\leqslant k}\omega_j/|[\varphi]_j|^2$, $\tau_k^\varphi := \max_{0\leqslant|j|\leqslant k}(\omega_j)_{\vee 1}/|[\varphi]_j|^2$, and $\delta_k^\varphi := 2k\Delta_k^\varphi\left\{\log(\tau_k^\varphi \vee (k+2)) \,/\, \log(k+2)\right\}$. Then, it is easily seen that

$$\delta_k^\varphi \leqslant \delta_k\, d\,\frac{\log(3d)}{\log 3} = \delta_k\, d\,\zeta_d \qquad \forall\, k \geqslant 1. \tag{A.17}$$

with $\zeta_d = (\log(3d))/(\log 3)$. Moreover, define the event $\Omega_{qp} := \Omega_q \cap \Omega_p$ where Ω_q is given in (A.11) and

$$\Omega_p := \Big\{(N_n^l \wedge M_m^l) \leqslant (\widehat{N}_n \wedge \widehat{M}_m) \leqslant (N_n \wedge M_m)\Big\}. \tag{A.18}$$

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

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

Using $\log(\Delta_k^\varphi \vee (k+2))/\log(k+2) \geqslant 1$, we conclude from the last estimate that

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

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

$$\mathrm{pen}(k) \leqslant \widehat{\mathrm{pen}}(k) \leqslant 30\,\mathrm{pen}(k) \qquad \forall\, 1 \leqslant k \leqslant M_m.$$

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

$$\begin{aligned}\Big(\mathrm{pen}(k\vee\widehat{k}) + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k})\Big)\mathbb{1}\{\Omega_{qp}\} \leqslant \Big(\mathrm{pen}(k) + \mathrm{pen}(\widehat{k}) + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k})\Big)\mathbb{1}\{\Omega_{qp}\}\\ \leqslant 31\,\mathrm{pen}(k) \qquad \forall 1 \leqslant k \leqslant M_m.\end{aligned} \tag{A.19}$$

Furthermore, we have $\widehat{\Delta}_k \leqslant \Delta_k^\varphi m$ for every $k \geqslant 1$, which implies $\widehat{\delta}_k \leqslant m\,(1+\log m)\delta_k^\varphi$. Consequently, $\widehat{\mathrm{pen}}(k) \leqslant 10\,m\,(1+\log m)\,\mathrm{pen}(k) \leqslant 600\,m\,(1+\log m)d\,\zeta_d\,\delta_1$ for all $1 \leqslant k \leqslant N_n$ by employing (A.17) and the definition of N_n. Therefore, on $\Omega_q^c \cap \Omega_p$, where $\widehat{k} \leqslant N_n$, we have $\mathrm{pen}(k \vee \widehat{k}) \leqslant 60\,d\,\zeta_d\,\delta_1$ for all $1 \leqslant k \leqslant N_n$, and hence

$$(\mathrm{pen}(k\vee\widehat{k}) + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k}))\mathbb{1}\{\Omega_q^c \cap \Omega_p\} \leqslant 60\,d\,\zeta_d\,\delta_1(1+10\,m\,(1+\log m))\mathbb{1}\{\Omega_q^c \cap \Omega_p\}.$$

(A.20)

Now consider the decomposition

$$\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 = \mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_{qp}\} + \mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_q^c \cap \Omega_p\} + \mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_p^c\}.$$

Below we show that there exist a numerical constant C such that for all $n, m \geqslant 1$ and all $1 \leqslant k \leqslant N_n^l \wedge M_m^l$ we have

$$\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_{qp}\} \leqslant C\Big\{\|f - f_k\|_\omega^2 + d\,\zeta_d \frac{\delta_k}{n} + r\,d\,\kappa_m + d\,\zeta_d \frac{\delta_1 + \Sigma(\zeta_d \|\varphi\|^2 \|f\|^2)}{n}\Big\}, \quad (A.21)$$

$$\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_q^c \cap \Omega_p\} \leqslant C\Big\{\|f - f_k\|_\omega^2 + r\,d\,\kappa_m + d\,\zeta_d \Big[\frac{\delta_1 + \Sigma(\zeta_d \|\varphi\|^2 \|f\|^2)}{n} + \frac{\delta_1 \left(d/\lambda_1\right)^7}{m}\Big]\Big\}, \quad (A.22)$$

$$\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_p^c\} \leqslant C \left(\frac{d}{\lambda_1}\right)^7 \frac{(1 + \|f\|_\omega^2)}{m}. \quad (A.23)$$

The desired upper bound follows for every $1 \leqslant k \leqslant (N_n^l \wedge M_m^l)$ by virtue of Definition 4.1 and Assumption 4.2.

Proof of (A.21). Following the proof in case of known degree of ill-posedness (Section A.2) line by line, it is easily seen that for $1 \leqslant k \leqslant (N_n^l \wedge M_m^l)$,

$$\begin{aligned}(1/2)\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_{qp}\} &\leqslant (3/2)\|f - f_k\|_\omega^2 + 10 \sum_{k=1}^{N_n} \Big(\sup_{t \in \mathcal{B}_k} |\langle t, \widetilde{\Phi}_\nu\rangle_\omega|^2 - 6\frac{\delta_k^\varphi}{n}\Big)_+ \\ &\quad + 8 \sup_{t \in \mathcal{B}_{N_n \wedge M_m}} |\langle t, \widehat{\Phi}_g - \widetilde{\Phi}_g\rangle_\omega|^2 + \Big(\mathrm{pen}(k \vee \widehat{k}) + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k})\Big)\mathbb{1}\{\Omega_{qp}\} \\ &\leqslant (3/2)\|f - f_k\|_\omega^2 + 10 \sum_{k=1}^{N_n} \Big(\sup_{t \in \mathcal{B}_k} |\langle t, \widetilde{\Phi}_\nu\rangle_\omega|^2 - 6\frac{\delta_k^\varphi}{n}\Big)_+ \\ &\quad + 8 \sup_{t \in \mathcal{B}_{N_n \wedge M_m}} |\langle t, \widehat{\Phi}_g - \widetilde{\Phi}_g\rangle_\omega|^2 + 31\,\mathrm{pen}(k),\end{aligned}$$

where the last inequality follows from (A.19). The third term is bounded by employing Lemma A.3. In order to control the second term, apply Lemma A.2 with $\delta_k^* = \delta_k^\varphi$ and $\Delta_k^* = \Delta_k^\varphi$. Using (A.17), $\Delta_k^\varphi \leqslant d\,\tau_k$, $\zeta_d \log(\tau_k^\varphi \vee (k+2)) \geqslant \log(\tau_k \vee (k+2))$ and the definition of Σ, we conclude with Assumption 3.2 that there exists a numerical constant $C > 0$ such that

$$\sum_{k=1}^{N_n} \mathbb{E}\Big(\sup_{t \in \mathcal{B}_k^0} |\langle t, \widetilde{\Phi}_\nu\rangle_\omega|^2 - 6\,\frac{\delta_k^\varphi}{n}\Big)_+ \leqslant \frac{C d\,\zeta_d}{n}\Big\{\delta_1 + \Sigma(\|\varphi\|^2 \|f\|^2 \zeta_d)\Big\}. \quad (A.24)$$

Consequently, combining these estimates proves inequality (A.21).

Proof of (A.22). On $\Omega_q^c \cap \Omega_p$, we have $N_n^l \wedge M_m^l \leqslant \widehat{N}_n \wedge \widehat{M}_m \leqslant N_n \wedge M_m$. Applying (A.20), it follows in analogy to proof of Theorem 3.3 that for all $1 \leqslant k \leqslant N_n^l \wedge M_m^l$

$$\begin{aligned}(1/2)\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_q^c \cap \Omega_p\} &\leqslant (3/2)\|f - f_k\|_\omega^2 + 10 \sum_{k=1}^{N_n} \left(\sup_{t \in \mathcal{B}_k} |\langle t, \widetilde{\Phi}_\nu \rangle_\omega|^2 - 6\frac{\delta_k^\varphi}{n} \right)_+ \\ &\quad + 8 \sup_{t \in \mathcal{B}_{N_n^l \wedge M_m^l}} |\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu \rangle_\omega \mathbb{1}\{\Omega_q^c\}|^2 + 8 \sup_{t \in \mathcal{B}_{N_n^l \wedge M_m^l}} |\langle t, \widehat{\Phi}_g - \widetilde{\Phi}_g \rangle_\omega|^2 \\ &\quad + \left(\mathrm{pen}(k \vee \widehat{k}) + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k})\right) \mathbb{1}\{\Omega_q^c \cap \Omega_p\} \\ &\leqslant (3/2)\|f - f_k\|_\omega^2 + 10 \sum_{k=1}^{N_n} \left(\sup_{t \in \mathcal{B}_k} |\langle t, \widetilde{\Phi}_\nu \rangle_\omega|^2 - 6\frac{\delta_k^\varphi}{n} \right)_+ \\ &\quad + 8 \sup_{t \in \mathcal{B}_{N_n^l \wedge M_m^l}} |\langle t, \widehat{\Phi}_\nu - \widetilde{\Phi}_\nu \rangle_\omega \mathbb{1}\{\Omega_q^c\}|^2 + 8 \sup_{t \in \mathcal{B}_{N_n^l \wedge M_m^l}} |\langle t, \widehat{\Phi}_g - \widetilde{\Phi}_g \rangle_\omega|^2 \\ &\quad + 60\, d\, \zeta_d\, \delta_1 (1 + 10\, m\, (1 + \log m)) \mathbb{1}\{\Omega_q^c \cap \Omega_p\}.\end{aligned}$$

Due to Lemma A.3, A.4, and (A.24), there exists a numerical constant C such that

$$\begin{aligned}\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_q^c \cap \Omega_p\} &\leqslant C \Big\{ \|f - f_k\|_\omega^2 + d r \kappa_m \\ &\quad + d\, \zeta_d \left[\frac{\delta_1 + \Sigma(\zeta_d \|\varphi\|^2 \|f\|^2)}{n} + \delta_1 (\mathbf{P}[\Omega_q^c])^{(1/2)} + \delta_1\, m\, (1 + \log m) \mathbf{P}[\Omega_q^c] \right] \Big\}.\end{aligned}$$

Employing Lemma A.5 now proves (A.22).

Proof of (A.23). Let $\breve{f}_k := 1 + \sum_{0<|j|\leqslant k} [f]_j \mathbb{1}\{|\widehat{[\varphi]}_j|^2 \geqslant 1/m\} e_j$. It is easy to see that $\|\widehat{f}_k - \breve{f}_k\|^2 \leqslant \|\widehat{f}_{k'} - \breve{f}_{k'}\|^2$ for all $k' \leqslant k$ and $\|\breve{f}_k - f\|^2 \leqslant \|f\|^2$ for all $k \geqslant 1$. Thus, using that $1 \leqslant \widehat{k} \leqslant (N_n^u \wedge m)$, we can write

$$\begin{aligned}\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_p^c\} &\leqslant 2\{\mathbb{E}\|\widehat{f}_{\widehat{k}} - \breve{f}_{\widehat{k}}\|_\omega^2 \mathbb{1}\{\Omega_p^c\} + \mathbb{E}\|\breve{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_p^c\}\} \\ &\leqslant 2\Big\{ \mathbb{E}\|\widehat{f}_{(N_n^u \wedge m)} - \breve{f}_{(N_n^u \wedge m)}\|_\omega^2 \mathbb{1}\{\Omega_p^c\} + \|f\|_\omega^2\, \mathbf{P}[\Omega_p^c] \Big\}.\end{aligned}$$

Moreover, applying Theorem 2.10 in Petrov (1995) we conclude

$$\begin{aligned}&\mathbb{E}\|\widehat{f}_{(N_n^u \wedge m)} - \breve{f}_{(N_n^u \wedge m)}\|_\omega^2 \mathbb{1}\{\Omega_p^c\} \\ &\quad\leqslant 2m \sum_{0<|j|\leqslant (N_n^u \wedge m)} \omega_j \Big\{ \mathbb{E}(\widehat{[g]}_j - [\varphi]_j [f]_j)^2 \mathbb{1}\{\Omega_p^c\} + \mathbb{E}([\varphi]_j [f]_j - \widehat{[\varphi]}_j [f]_j)^2 \mathbb{1}\{\Omega_p^c\} \Big\} \\ &\quad\leqslant 2m \Big\{ \sum_{0<|j|\leqslant (N_n^u \wedge m)} \omega_j \left[\mathbb{E}\left(\widehat{[g]}_j - [g]_j \right)^4 \right]^{1/2} \mathbf{P}[\Omega_p^c]^{1/2} \\ &\qquad + \sum_{0<|j|\leqslant (N_n^u \wedge m)} \omega_j |[f]_j|^2 [\mathbb{E}(\widehat{[\varphi]}_j - [\varphi]_j)^4]^{1/2} \mathbf{P}[\Omega_p^c]^{1/2} \Big\} \\ &\quad\leqslant 2m \Big\{ (2m \max_{1\leqslant j \leqslant N_n^u} \omega_j)(cn^{-1}) + (cm^{-1})\|f\|_\omega^2 \Big\}\, \mathbf{P}[\Omega_p^c]^{1/2},\end{aligned}$$

which implies, using Definition 4.1 (ii),

$$\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_p^c\} \leqslant C \Big\{ \left(m^2 + \|f\|_\omega^2\right) \mathbf{P}[\Omega_p^c]^{1/2} + \|f\|_\omega^2\, \mathbf{P}[\Omega_p^c] \Big\}.$$

Lemma A.6 below together with Definition 3.1 (ii) yields, for some numeric $C > 0$,

$$\mathbb{E}\|\widehat{f}_{\widehat{k}} - f\|_\omega^2 \mathbb{1}\{\Omega_p^c\} \leqslant C\Big\{ \frac{(d/\lambda_1)^{7/2}}{m} + \frac{(d/\lambda_1)^{7/2}\,\|f\|_\omega^2}{m^3} + \frac{(d/\lambda_1)^{7}\,\|f\|_\omega^2}{m^6} \Big\}$$

which completes the proof. □

Lemma A.6 *Consider the event Ω_p defined in (A.18). Then we have*

$$\mathbf{P}(\Omega_p^c) \leqslant 6\,(504\,d/\lambda_1)^7\,m^{-6} \qquad \forall\, n, m \geqslant 1.$$

Proof. Let $\Omega_I := \{(N_n^l \wedge M_m^l) > (\widehat{N}_n \wedge \widehat{M}_m)\}$ and $\Omega_{II} := \{(\widehat{N}_n \wedge \widehat{M}_m) > (N_n \wedge M_m)\}$. Then we have $\Omega_p^c = \Omega_I \cup \Omega_{II}$. Consider $\Omega_I = \{\widehat{N}_n < (N_n^l \wedge M_m^l)\} \cup \{\widehat{M}_m < (N_n^l \wedge M_m^l)\}$ first. By definition of N_n^l, we have that $\min_{1\leqslant|j|\leqslant N_n^l} \frac{|[\varphi]_j|^2}{|j|(\omega_j)\vee 1} \geqslant \frac{4(\log n)}{n}$, which implies

$$\{\widehat{N}_n < (N_n^l \wedge M_m^l)\} \subset \Big\{ \exists 1 \leqslant |j| \leqslant (N_n^l \wedge M_m^l) : \frac{|\widehat{[\varphi]}_j|^2}{|j|(\omega_j)\vee 1} < \frac{\log n}{n} \Big\}$$
$$\subset \bigcup_{1\leqslant|j|\leqslant N_n^l\wedge M_m^l} \Big\{ \frac{|\widehat{[\varphi]}_j|}{|[\varphi]_j|} \leqslant 1/2 \Big\} \subset \bigcup_{1\leqslant|j|\leqslant N_n^l\wedge M_m^l} \Big\{ |\widehat{[\varphi]}_j/[\varphi]_j - 1| \geqslant 1/2 \Big\}.$$

One can see that from $\min_{1\leqslant|j|\leqslant M_m^l} |[\varphi]_j|^2 \geqslant \frac{4(\log m)^2}{m}$ it follows in the same way that

$$\Big\{ \widehat{M}_m < (N_n^l \wedge M_m^l) \Big\} \subset \bigcup_{1\leqslant|j|\leqslant N_n^l\wedge M_m^l} \Big\{ |\widehat{[\varphi]}_j/[\varphi]_j - 1| \geqslant 1/2 \Big\}.$$

Therefore, $\Omega_I \subset \bigcup_{1\leqslant|j|\leqslant M_m} \Big\{ |\widehat{[\varphi]}_j/[\varphi]_j - 1| \geqslant 1/2 \Big\}$, since $M_m^l \leqslant M_m$. Hence, as in (A.15) applying Hoeffding's inequality together with the definition of M_m gives

$$\mathbf{P}[\Omega_I] \leqslant \sum_{1\leqslant|j|\leqslant M_m} 2\,\exp\Big(-\frac{m\,|[\varphi]_j|^2}{72} \Big) \leqslant 4\,(504\,d/\lambda_1)^7\,m^{-6}. \qquad \text{(A.25)}$$

Consider $\Omega_{II} = \{\widehat{N}_n > (N_n \wedge M_m)\} \cap \{\widehat{M}_m > (N_n \wedge M_m)\}$. In case $(N_n \wedge M_m) = N_n$, use $\frac{\log n}{4n} \geqslant \max_{|j|\geqslant N_n} \frac{|[\varphi]_j|^2}{|j|(\omega_j)\vee 1}$ due to Assumption 4.2, such that

$$\Omega_{II} \subset \{\widehat{N}_n > N_n\} \subset \Big\{ \forall 1 \leqslant |j| \leqslant N_n : \frac{|\widehat{[\varphi]}_j|^2}{|j|(\omega_j)\vee 1} \geqslant \frac{\log n}{n} \Big\}$$
$$\subset \Big\{ \frac{|\widehat{[\varphi]}_{N_n}|}{|[\varphi]_{N_n}|} \geqslant 2 \Big\} \subset \Big\{ |\widehat{[\varphi]}_{N_n}/[\varphi]_{N_n} - 1| \geqslant 1 \Big\}.$$

In case $(N_n \wedge M_m) = M_m$, it follows analogously from $\frac{(\log m)^2}{4m} \geqslant \max_{|j|\geqslant M_m} |[\varphi]_j|^2$ that

$$\Omega_{II} \subset \{\widehat{M}_m > M_m\} \subset \Big\{ |\widehat{[\varphi]}_{M_m}/[\varphi]_{M_m} - 1| \geqslant 1 \Big\}.$$

Therefore, $\Omega_{II} \subset \left\{ |\widehat{[\varphi]}_{N_n \wedge M_m}/[\varphi]_{N_n\wedge M_m} - 1| \geqslant 1 \right\}$ and hence as in (A.15) applying Hoeffding's inequality together with the definition of M_m gives

$$\mathbf{P}[\Omega_{II}] \leqslant 2 \exp\left(- \frac{m\,|[\varphi]_{N_n\wedge M_m}|^2}{72} \right) \leqslant 2(504\, d/\lambda_1)^7\, m^{-7}. \tag{A.26}$$

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

Illustration: estimation of derivatives

Proof of Proposition 4.5. We start our proof with the observation that in both cases the sequences δ, Δ, N and M are the same as in Proposition 3.5 and it is easily verified that the additional Assumption 4.2 is satisfied. Moreover in case **[os]** we have $N_n^l \sim (n/(\log n))^{1/(2a+2s+1)}$ and $M_m^l \sim (m/(\log m)^2)^{1/(2a)}$. Let $k_n^* := n^{1/(2a+2p+1)}$ and note that still $k_n^* \lesssim N_n^l$. In case **[ss]** we have $N_n^l \sim \{\log(n/(\log n)^{(2p+2a+1)/(2a)})\}^{1/(2a)} = (\log n)^{1/(2a)}(1+o(1))$ and $M_m^l \sim \{\log(m/(\log m)^3)\}^{1/(2a)} = (\log m)^{1/(2a)}(1+o(1))$. The rest of the proof in both cases is almost identical to the one of proposition 3.5 but uses N_n^l and M_m^l rather than N_n and M_m, and we omit the details. □

References

Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113(3):301–413.

Bigot, J. and Van Bellgem, S. (2009). Log-density deconvolution by wavelet thresholding. *Scandinavian Journal of Statistics*, 36:749–763.

Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric deconvolution model. *Bernoulli*, 11:309–340.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83:1184–1186.

Cavalier, L. and Hengartner, N. (2005). Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21:1345–1361.

Cochran, W. W., Mouritsen, H., and Wikelski, M. (2004). Migrating songbirds recalibrate their magnetic compass daily from twylight cues. *Science*, 304:405.

Comte, F. and Lacour, C. (2009). Data driven density estimation in presence of unknown convolution operator. Prépublication map5 2008-15, Université Paris Descartes.

Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for density deconvolution. *Canadian Journal of Statistics*, 37(3).

Comte, F. and Taupin, M.-L. (2003). Adaptive density deconvolution for circular data. Prépublication map5 2003-10, Université Paris Descartes.

Curray, J. R. (1956). The analysis of two-dimensional orientation data. *The Journal of Geology*, 64(2):117–131.

Devroye, L. (1989). Consistent deconvolution in density estimation. *Canadian Journal of Statistics*, 17:235–239.

Diggle, P. J. and Hall, P. (1993). A fourier approach to nonparametric deconvolution of a density estimate. *Journal of the Royal Statistical Society. Series B*, 55(2):523–531.

Efromovich, S. (1997). Density estimation for the case of supersmooth measurement error. *Journal of the American Statistical Association*, 92:526–535.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics*, 19:1257–1272.

Fan, J. (1992). Deconvolution with supersmooth distributions. *Canadian Journal of Statistics*, 20:734–747.

Fan, J. and Koo, J. Y. (2002). Wavelet deconvolution. *IEEE Transactions on Information Theory*, 48:734–747.

Fisher, N. (1993). *Statistical analysis of circular data.* Cambridge University Press.

Goldenshluger, A. (1999). On pointwise adaptive nonparametric deconvolution. *Bernoulli*, 5:907–925.

Goldenshluger, A. (2000). Density deconvolution in the circular structural model. *Journal of Multivariate Analysis*, 81:360–375.

Hall, P. and Yao, Q. (2003). Inference in components of variance models with low replication. *Annals of Statistics*, 31(2):414–441.

Horowitz, J. L. and Markatou, M. (1996). Semiparametric estimation of regression models for panel data. *Review of Economic Studies*, 63:145–168.

Johannes, J. (2009). Deconvolution with unknown error distribution. Forthcoming in Annals of Statistics.

Kim, P. T. and Koo, J. Y. (2002). Optimal spherical deconvolution. *Journal of Multivariate Analysis*, 80:21–42.

Koo, J. Y. and Park, B. U. (1996). B-spline deconvolution based on the EM algorithm. *Journal of Statistical Computation and Simulation.*, 54:275–288.

Korostolev, A. P. and Tsybakov, A. B. (1993). *Minimax Theory for Image Reconstruction.*, volume 82 of *Lecture Notes in Statistics.* Springer-Verlag.

Mardia, K. (1972). *Statistics of directional data.* Probability and Mathematical Statistics. A Series of Monographs and Textbooks. Vol. 13. Academic Press.

Meister, A. (2004). Deconvolution density estimation with a testing procedure for the error distribution. Preprint., Univ. Stuttgart.

Meister, A. (2006). Density estimation with normal measurement error with unknown variance. *Statistica Sinica*, 16:195–211.

Meister, A. (2009). *Deconvolution problems in nonparametric statistics.* Lecture Notes in Statistics 193. Berlin: Springer.

Mendelsohn, J. and Rice, J. (1982). Deconvolution of micro-fluorometric histograms with B-splines. *Journal of the American Statistical Association*, 77:748–753.

Natterer, F. (1984). Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Analysis*, 18:29–37.

Neubauer, A. (1988a). An a posteriori parameter choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates. *SIAM Journal on Numerical Analysis*, 25(6):1313–1326.

Neubauer, A. (1988b). When do Sobolev spaces form a Hilbert scale? *Proceedings of the American Mathematical Society*, 103(2):557–562.

Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7:307–330.

Neumann, M. H. (2007). Deconvolution from panel data with unknown error distribution. *Journal of Multivariate Analysis*, 98(10):1955–1968.

Olver, F. (1974). *Asymptotics and special functions.* Academic Press.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Annals of Statistics*, 27:2033–2053.

Petrov, V. V. (1995). *Limit theorems of probability theory. Sequences of independent random variables.* Oxford Studies in Probability. Clarendon Press., Oxford, 4. edition.

Schwarz, M. and Van Bellegem, S. (2009). Consistent density deconvolution under partially known error distribution. *Statistics and Probability Letters*, In Press, Corrected Proof.

Stefanski, L. A. (1990). Rate of convergence of some estimators in a class of deconvolution problems. *Statistics and Probability Letters*, 9:229–235.

Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Annals of Statistics*, 18:806–831.