Econometric Theory, **40**, 2024, 320–359.
doi:10.1017/S0266466622000457

A MOLLIFIER APPROACH TO THE DECONVOLUTION OF PROBABILITY DENSITIES

THORSTEN HOHAGE
Institute for Numerical and Applied Mathematics, University of Göttingen

PIERRE MARÉCHAL
Mathematical Institute of Toulouse, Paul Sabatier University

LÉOPOLD SIMAR
Institut de Statistique, Biostatistique et sciences actuarielles, Université Catholique de Louvain-la-Neuve

ANNE VANHEMS
TBS Business School

We use mollification to regularize the problem of deconvolution of random variables. This regularization method offers a unifying and generalizing framework in order to compare the benefits of various filter-type techniques like deconvolution kernels, Tikhonov, or spectral cutoff methods. In particular, the mollifier approach allows to relax some restrictive assumptions required for the deconvolution kernels, and has better stabilizing properties compared with spectral cutoff or Tikhonov. We show that this approach achieves optimal rates of convergence for both finitely and infinitely smoothing convolution operators under Besov and Sobolev smoothness assumptions on the unknown probability density. The qualification can be arbitrarily high depending on the choice of the mollifier function. We propose an adaptive choice of the regularization parameter using the Lepskiĭ method, and we provide simulations to compare the finite sample properties of our estimator with respect to the well-known regularization methods.

1. INTRODUCTION

Deconvolution is a very classical issue in statistics and econometrics and is well known to be an ill-posed problem. Various methods have been studied to regularize it, among which the seminal deconvolution kernels (see Carroll and Hall, 1988; Devroye, 1989; Stefanski and Carroll, 1990; Fan, 1991a, 1991b, among others), the Tikhonov regularization (as in Carrasco, Florens, and Renault, 2007), and the spectral cutoff (see Mair and Ruymgaart, 1996; Johannes, 2009). Projection-based methods have also been investigated (as in Comte, Rozenholc, and Taupin,

Address correspondence to Anne Vanhems, TBS Business School, Toulouse, France; e-mail: a.vanhems@tbs-education.fr.

2006, 2007, or in Van Rooij and Ruymgaart, 1991; Efromovich, 1997; Pensky and Vidakovic, 1999, among others).

Optimal rates of convergence have been achieved for these estimators for various smoothness combinations between noise and density of interest (Fan, 1991b, Butucea, 2004, Butucea and Tsybakov, 2008), and also in multidimensional anisotropic settings (Comte and Lacour, 2013, Rebelles, 2016, Lepskiĭ and Willer, 2017, 2019). Among the classical assumptions required to proceed, authors often impose the noise density function to be strictly positive, although this assumption has been relaxed in Carrasco and Florens (2011), where the density may have isolated zeros, or in Hall and Meister (2007), Delaigle and Meister (2011), and Trong and Phuong (2015). Note also that an important stream of the literature has been devoted to unknown density error, as in Delaigle, Hall, and Meister (2008), Johannes (2009), or Delaigle and Hall (2016).

In this paper, we propose a different method to regularize the deconvolution problem, which uses a regularization principle introduced in the deterministic setting, and has been applied in several fields of signal and image processing such as deconvolution of images in astronomy and computerized tomography (as in Maréchal, Togane, and Celler, 2000). To the best of our knowledge, this regularization principle has never been applied in the stochastic setting. We refer to it as *regularization by mollification*, or merely as mollification.

Mollification belongs to a class of variational methods, like Tikhonov regularization, and has also some commonalities with deconvolution kernels. A brief history of the approach is given in Section 2.2. The present paper provides a conceptual framework that facilitates comparisons, from both theoretical and practical viewpoints, between mollification and competitors also defined as filter-type estimators, such as deconvolution kernels, Tikhonov, or spectral cutoff. In this framework, mollification appears to provide a better compromise between stability of the estimator and fidelity to the data simultaneously over the whole frequency range. This fact can be seen using both the variational definition of the regularization method and the filter-type definition of the estimator. It is illustrated later in the paper through simulations using a stability index tool.

We consider a multidimensional setting with a known isotropic error density, provide rates of convergence for our mollified estimator in the general Besov spaces setting, for ordinary smooth and supersmooth error cases, and show that the obtained rates correspond to optimal rates found in the literature. We also propose an empirical rule for the smoothing parameter selection based on the Lepskiĭ method and prove the convergence of the adaptive estimator.

The paper is organized as follows: In Section 2, we present the general setting of deconvolution, introduce the mollification method, and define our regularized estimator. In Sections 3 and 4, we study the convergence properties of our estimator under an a priori parameter selection rule and also using the Lepskiĭ method as an a posteriori parameter selection rule. In Section 5, we provide a framework for comparison with classical regularization methods and show, in particular, that mollification allows to relax some restrictive assumptions required by the deconvolution kernels, while having better stabilizing properties than the Tikhonov regularization, and better morphological properties than the spectral cutoff. At last, in Section 6, we provide simulations to compare the finite sample properties of our estimator with respect to the well-known methods.

2. SETTING AND THE MOLLIFICATION APPROACH

2.1. The Deconvolution Problem

2.1.1. *Assumptions.* Consider the equation $Y = X + \varepsilon$ in which Y is the observed random vector, X is the latent random vector, and ε is a random noise vector. All random vectors $Y, X,$ and ε have values in $\mathbb{R}^d$. Throughout, we make the following assumptions:

(A1) The random vectors X and ε are independent.
(A2) All random vectors Y, X, and ε have densities with respect to the Lebesgue measure, denoted, respectively, by $g, f,$ and γ.
(A3) Both f and g belong to $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$.

From assumptions (A1) and (A2), the equation $Y = X + \varepsilon$ gives rise to the relation $g = f * \gamma$ with the convolution $(f * \gamma)(x) := \int_{\mathbb{R}^d} f(x-y)\gamma(y)\,\mathrm{d}y$. Our objective is to recover the density function f by solving this deconvolution problem. We are mostly interested here in the case where γ is known, either from modeling or from empirical observations. The density g is unknown and given only approximately by the statistical sample $Y_1, \dots, Y_n$.

Assumption (A3) places us in the familiar framework of Hilbert spaces, which lend themselves naturally to variational methods. In this framework, finding the unknown density f entails inverting the convolution operator T_γ defined by

$$\begin{array}{llcl} T_\gamma : & L^2(\mathbb{R}^d) & \longrightarrow & L^2(\mathbb{R}^d) \\ & f & \longmapsto & T_\gamma f := f * \gamma. \end{array} \tag{1}$$

A classical way to proceed to recover the density function f is to apply the Fourier transform to the equation $g = f * \gamma$. The Fourier transform of an integrable function h is defined by $\hat{h}(\xi) := (Uh)(\xi) := \int e^{-2i\pi\langle x, \xi\rangle} h(x)\,\mathrm{d}x$. When dealing with square-integrable functions, we use the Fourier–Plancherel operator, obtained by *closure* of the previous integral transform, thanks to the Plancherel theorem (see, e.g., Rudin, 1970), denoted likewise. With the help of the Fourier convolution theorem (see, e.g., Rudin, 1970), the deconvolution problem can then be rewritten equivalently as $\hat{g} = \hat{f} \cdot \hat{\gamma}$.

Remark 1. In this paper, estimated quantities are indexed with the sample size n, whereas the *hat* is reserved for the Fourier transform.

2.1.2. *Ill-Posedness.* Recall the basic inequality $\|f * \gamma\| \le \|f\|\,\|\gamma\|_1 = \|f\|$, in which $\|\cdot\|$ denotes (and will denote throughout) the L^2-norm and $\|\cdot\|_1$ denotes the L^1-norm. Under the mild assumption that the set $\left\{\xi \in \mathbb{R}^d \,\middle|\, \hat{\gamma}(\xi) = 0\right\}$ has

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

Lebesgue measure zero, the operator T_γ is injective. As a matter of fact, $T_\gamma f = 0 \Leftrightarrow f * \gamma = 0 \Leftrightarrow \hat{f} \cdot \hat{\gamma} = 0 \Leftrightarrow \hat{f} = 0 \Leftrightarrow f = 0$, where equality in $L^2(\mathbb{R}^d)$ (that is, almost everywhere) is meant. The difficulty is that the Moore–Penrose pseudoinverse $T_\gamma^\dagger$ of T_γ is unbounded, which makes the problem ill-posed. As a matter of fact, we have $\inf_{\|f\|^2=1} \left\|T_\gamma f\right\|^2 = \inf_{\|Uf\|^2=1} \|U\gamma \cdot Uf\|^2 = 0$. Here, the first equality follows by the Plancherel theorem. As for the second equality, it is easily obtained from the Riemann–Lebesgue lemma, which says that $U\gamma$ is continuous and vanishes at infinity, by observing that, for every $\alpha \in \mathbb{R}^d$, $U\big(e^{2i\pi\langle\alpha,x\rangle} f(x)\big)(\xi) = Uf(\xi - \alpha)$.

Remark 2. In the one-dimensional case, the assumption that $\left\{\xi \in \mathbb{R} \,\middle|\, \hat{\gamma}(\xi) = 0\right\}$ has Lebesgue measure zero is much less stringent than imposing that $|\hat{\gamma}(\xi)| > 0$ for all ξ in $\mathbb{R}$, as it is assumed in Stefanski and Carroll (1990) or Johannes (2009). The strict positivity assumption of $\hat{\gamma}$ was obviously related to the shape of the kernel or spectral cutoff estimator and has been later generalized by Hall and Meister (2007), Delaigle et al. (2008), and Trong and Phuong (2015), as well as by Carrasco and Florens (2011), to allow for many isolated zeros. As we will see later, as soon as we consider variational regularization methods, the estimator can be defined in much more general settings.

The ill-posedness of the deconvolution problem makes it necessary to apply a regularization method to solve it, and we present in the next section an introduction to the variational approach to mollification that we will use in this work.

2.2. Mollification

In this section, we recall the historical roots of mollification and present the variational approach to mollification.

2.2.1. *Historical Background.* Mollifiers were introduced in the field of partial differential equations by Friedrichs (see Friedrichs, 1944; Wikipedia Contributors, 2020). To the best of our knowledge, the term *mollification* has been used in the field of inverse problems since the 80s. In the original works on the subject, mollifiers were used to smooth the data prior to inversion, whenever an explicit inversion formula was available. In his book, Murio (2011) presents this approach and its application to some classical inverse problems, and gives a rich bibliography on the subject. Let us mention, in particular, Hào (1994), in which *a sequence of "mollification operators" maps the improper data into well-posedness classes of the problem*, which provides a wide framework for the mollification approach. Louis and Maass (1990) proposed an alternative approach, based on inner product duality, and subsequently referred to as *the method of approximate inverses* (see Schuster, 2007). Their paper opened the way to the application of the concept to inverse problems in which the operator has no explicit inverse, but the adjoint equation has explicit solutions. A third approach, based on a variational formulation, also appeared in the same period of time. Lannes, Roques, and Casanove (1987) give such a formulation while studying the problems of

Fourier synthesis and deconvolution. To the best of our knowledge, this variational formulation was not studied further until the papers by Alibaud, Maréchal, and Saesor (2009) and by Bonnefond and Maréchal (2009), in which the convergence of the variational formulation was explored.

2.2.2. *The Mollification Approach.* We now present the mollification method and situate it among the various methodologies currently used for our deconvolution problem.

Redefining the Target Object. The mollification approach to solving ill-posed equations, such as the deconvolution problem, consists in redefining the *target object* by means of a convolution operator. Given an integrable function φ such that $\int \varphi = 1$, one defines the family of functions $(\varphi_\beta)_{\beta\in(0,1]}$ by letting $\varphi_\beta(x) := \frac{1}{\beta^d}\varphi\left(\frac{x}{\beta}\right)$, $x \in \mathbb{R}^d$. The corresponding family of operators $(C_\beta)_{\beta\in(0,1]}$, where $C_\beta : f \mapsto \varphi_\beta * f$, is then called an *approximate identity*. A standard approximation theorem states that, for every $f \in L^2(\mathbb{R}^d)$, $C_\beta f \to f$ in $L^2(\mathbb{R}^d)$ as $\beta \downarrow 0$. Although great generality can be reached in the above definition, it is customary to choose a function φ that is smooth, isotropic, positive, and sometimes compactly supported (see Wikipedia Contributors, 2020). So we will rather try to recover $C_\beta f$ in place of f, and β is our regularization parameter.

Mollification as a Filter-Type Method. As discussed in Section 2.1, thanks to the Fourier convolution theorem, we can write $UT_\gamma f = \hat{\gamma} \cdot Uf$. Recall that U is unitary, that is, $U^{-1} = U^*$ where U^* is the adjoint of U. The operator T_γ can then be written as

$$T_\gamma = U^*\left[\hat{\gamma}\right]U,$$

in which $\left[\hat{\gamma}\right]$ denotes the operator of multiplication by $\hat{\gamma}$. More precisely, consider the function $h \in L^\infty(\mathbb{R}^d)$ and let $[h] : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ be the *multiplication operator* defined by $([h]f)(\xi) = h(\xi)\cdot f(\xi), \quad \xi \in \mathbb{R}^d$. Then, the inverse of $T_\gamma : L^2(\mathbb{R}^d) \to \operatorname{ran} T_\gamma$ is given by

$$T_\gamma^{-1} = U^{-1}\left[\frac{1}{\hat{\gamma}}\right]U.$$

The unboundedness of $1/\hat{\gamma}$ yields that of T_γ^{-1} (and $T_\gamma^\dagger$). We call *filter-type method* any regularization method which acts explicitly in the Fourier domain by *bounding* the multiplication operation. The corresponding regularized solution f_{REG} is defined as $f_{\mathrm{REG}} = U^{-1}[\Phi]Ug$ or, equivalently, by $\hat{f}_{\mathrm{REG}} = \Phi\cdot\hat{g}$, in which the *filter* Φ depends on regularization parameters. The mollification solution to our deconvolution problem is defined by

$$f_{\mathrm{MO},\beta} := U^{-1}\left[\frac{\overline{\hat{\gamma}}\,\hat{\varphi}_\beta}{|\hat{\gamma}|^2 + |1-\hat{\varphi}_\beta|^2}\right]Ug \qquad \textbf{(2)}$$

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

or, equivalently, by $\hat{f}_{\text{MO},\beta}(\xi) = \frac{\overline{\hat{\gamma}(\xi)}\,\hat{\varphi}_\beta(\xi)}{|\hat{\gamma}(\xi)|^2+|1-\hat{\varphi}_\beta(\xi)|^2}\,\hat{g}(\xi)$. We observe right away that, whenever φ is even, the solution is well defined for every positive value of the regularization parameter β. As a matter of fact, in this case, the denominator of the filter in (2) remains bounded away from zero, as a consequence of the Riemann–Lebesgue lemma and of the fact that the real function $\xi \mapsto 1-\hat{\varphi}_\beta(\xi)$ vanishes only at $\xi = 0$.

Mollification as a Variational Method. It is readily seen that the above mollification solution is also the minimizer of the functional

$$\mathscr{F}(f) := \frac{1}{2}\left\|C_\beta g - T_\gamma f\right\|^2 + \frac{1}{2}\left\|(I-C_\beta)f\right\|^2, \tag{3}$$

in which C_β denotes the operator of convolution with φ_β. From this angle of view, the rationale for considering $\mathscr{F}$ is as follows. In the tautological decomposition $f = C_\beta f + (I-C_\beta)f$, $(I-C_\beta)f$ is the undesired component, and it appears natural to choose $\left\|(I-C_\beta)f\right\|^2$ as the penalty term. As for the fit term, the residual squared norm $\left\|C_\beta g - T_\gamma f\right\|^2$ promotes adequacy to the *mollified data*, that is, the data corresponding to the target object, since $T_\gamma C_\beta = C_\beta T_\gamma$.

Let us stress, at this point, that a major advantage of the variational approach is that additional restrictions can be introduced as constraints in the optimization problem. Since we are looking for a density, one may wish to search for minimizers over some closed convex subset of L^2, such as the subset of nonnegative functions. Such constraints can be handled by means of a *projection gradient method*, even in infinite-dimensional Hilbert spaces. See, for example, McCormick and Tapia (1972), an early reference on this subject. In the finite-dimensional implementation, projection onto the positive orthant is, of course, explicit, and convergent numerical schemes can be used. Such methods have a long history, and the interested reader may consult Bauschke and Combettes (2011) for a general purpose book on constrained optimization in general Hilbert spaces. However, we stress that adding constraints makes the convergence analysis considerably more difficult.

Remark 3. The above design of the fit term relies on the observation that convolution operators commute, giving rise to the simple *intertwining relationship* $T_\gamma C_\beta = C_\beta T_\gamma$. In a simpler version of mollification, one may merely require adequacy to the original data, yielding the solution

$$f_{\text{MM},\beta} = U^{-1}\left[\frac{\overline{\hat{\gamma}}}{|\hat{\gamma}|^2+|1-\hat{\varphi}_\beta|^2}\right]Ug. \tag{4}$$

Well-Posedness. Clearly, $f_{\text{MO},\beta}$ and $f_{\text{MM},\beta}$ depend continuously on g. As a matter of fact, the Fourier–Plancherel operator is an isometry and the multiplication operators by the bounded filter functions Φ in (2) and (4) have finite norms (see Proposition A.1 in Appendix A).

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

2.2.3. *Estimation.* The definition of the mollification given in (2) assumes the knowledge of the density g, whereas in practice g can only be estimated from a random i.i.d. sample $(Y_1,\dots,Y_n)$. In this formula, Ug may be replaced by the empirical estimator

$$\hat{g}_n(\xi) = \frac{1}{n}\sum_{j=1}^{n} e^{-2i\pi\langle \xi, Y_j\rangle}, \tag{5}$$

but then the variational interpretation would no longer hold, for $\hat{g}_n$ does not belong to $L^2(\mathbb{R}^d)$. By the way, it is the Fourier transform of the distribution

$$g_n := \frac{1}{n}\sum_{j=1}^{n} \delta_{Y_j}, \tag{6}$$

not that of a density. However, the formula (2) can be turned into a kernel-type estimator by rewriting it as

$$f_{\mathrm{MO},\beta} = U^{-1}\left[\frac{\overline{\hat{\gamma}}}{|\hat{\gamma}|^2+|1-\hat{\varphi}_\beta|^2}\right]U(\varphi_\beta * g) = U^{-1}\left[\frac{\overline{\hat{\gamma}}}{|\hat{\gamma}|^2+|1-\hat{\varphi}_\beta|^2}\right]\hat{\varphi}_\beta\cdot\hat{g},$$

in which $\hat{g}$ can now be replaced by the empirical estimator (5). Note that, under the mild additional constraint that $\varphi\in L^2(\mathbb{R}^d)$, the function $\xi\mapsto\hat{\varphi}_\beta(\xi)\cdot\hat{g}_n(\xi)$ is also in $L^2(\mathbb{R}^d)$ so that, in this case, the estimator corresponding to $f_{\mathrm{MO},\beta}$ is nothing but the estimator corresponding to $f_{\mathrm{MM},\beta}$ applied to a kernel estimation of g with the same bandwidth β^{-1}. In what follows, we denote this estimator by

$$f_{\beta,n} = U^{-1}\left[\frac{\overline{\hat{\gamma}}}{|\hat{\gamma}|^2+|1-\hat{\varphi}_\beta|^2}\right]U(\varphi_\beta * g_n). \tag{7}$$

Remark 4. We stress here that the variational interpretation of the estimator corresponding to $f_{\mathrm{MM},\beta}$ is possible only if preceded by an estimator g_n of g such that $Ug_n\in L^2(\mathbb{R})$. For example, a kernel estimator of Ug of the form $\xi\mapsto\hat{\phi}_{\beta'}(\xi)\cdot\hat{g}_n(\xi)$, in which ϕ is a kernel function and $\beta'>0$ is the bandwidth parameter, may be considered. In this case, the corresponding estimator of f would depend on the two parameters β and β', and would take the form $U^{-1}\left[\frac{\overline{\hat{\gamma}}}{|\hat{\gamma}|^2+|1-\hat{\varphi}_\beta|^2}\right]\hat{\phi}_{\beta'}\cdot\hat{g}_n = U^{-1}\left[\frac{\overline{\hat{\gamma}}\hat{\phi}_{\beta'}}{|\hat{\gamma}|^2+|1-\hat{\varphi}_\beta|^2}\right]\hat{g}_n$. A natural choice, which we make from now on, is a kernel estimator with kernel φ_β, i.e., $C_\beta g_n$.

3. CONVERGENCE ANALYSIS

3.1. Consistency

In this subsection, we present a general consistency result where g_n can be any consistent estimator of the true unknown density g. Optimal rates of convergence will be derived in the next subsections with g_n defined as in equation (6).

THEOREM 5. *Assume that g_n is a consistent nonparametric estimator of g, that is to say, $E(\|g_n - g\|^2) \to 0$ as n goes to infinity. Let $f_{n,\beta}$ denote the mollified solution corresponding to data g_n. There then exists a sequence $\beta_n \downarrow 0$ such that*

$$E\left\|f_{\beta_n,n} - f\right\|^2 \to 0 \quad \textit{as} \quad n \to \infty.$$

The proof of the theorem relies on the following simple lemma.

LEMMA 6. *Let $c\colon (0,1] \to \mathbb{R}_+$ be any function, and let (α_n) be any sequence of positive numbers which converges to zero. Then, there exists a sequence (β_n) such that:*

(1) $\beta_n \downarrow 0$ *as* $n \to \infty$.
(2) $c(\beta_n)\alpha_n \to 0$ *as* $n \to \infty$.

Proof. Let $(\beta^{(k)}) \in (0,1]^{\mathbb{N}^*}$ be strictly decreasing, converging to zero. Since (α_n) converges to zero, for every $k \in \mathbb{N}^*$, there exists $n_k \in \mathbb{N}^*$ such that $c(\beta^{(k)})\alpha_n \leq \beta^{(k)}$, for all $n \geq n_k$. Clearly, we can choose (n_k) to be strictly increasing. Define $(\beta_n) \in (0,1]^{\mathbb{N}^*}$ by $\beta_n = 1$ if $n < n_1$ and, for $k \geq 1$, $\beta_n = \beta^{(k)}$ if $n_k \leq n < n_{k+1}$. Then (β_n) has the desired properties. □

Proof of Theorem 5. Consider the classical decomposition: $f_{\beta,n} - f = f_{\beta,n} - f_{\mathrm{MO},\beta} + f_{\mathrm{MO},\beta} - f$. Let us first control the size of the deterministic part. By Parseval's theorem,

$$\begin{aligned}\left\|f_{\mathrm{MO},\beta} - f\right\|^2 &= \left\|\left(T_\gamma^* T_\gamma + (I - C_\beta)^*(I - C_\beta)\right)^{-1} T_\gamma^* C_\beta T_\gamma f - f\right\|^2 \\ &= \left\|\left(\frac{|\hat{\gamma}|^2\hat{\varphi}_\beta}{|\hat{\gamma}|^2 + |1-\hat{\varphi}_\beta|^2} - 1\right)\hat{f}\right\|^2 = \int\left|\frac{|\hat{\gamma}|^2\hat{\varphi}_\beta}{|\hat{\gamma}|^2 + |1-\hat{\varphi}_\beta|^2} - 1\right|^2\left|\hat{f}\right|^2 \mathrm{d}\xi.\end{aligned}$$

Since the integrand is dominated by the integrable function $\xi \mapsto |\hat{f}(\xi)|^2$, and since it converges pointwise to zero as $\beta \downarrow 0$ (recall that $\hat{\varphi}_\beta(\xi) = \hat{\varphi}(\beta\xi)$), Lebesgue's dominated convergence theorem shows that $\left\|f_{\mathrm{MO},\beta} - f\right\|^2 \longrightarrow 0$ as $\beta \to 0$. Let us now deal with the stochastic term. Using again Parseval's theorem, we have

$$\begin{aligned}E\left\|f_{\beta,n} - f_{\mathrm{MO},\beta}\right\|^2 &= E\left\|\left(T_\gamma^* T_\gamma + (I - C_\beta)^*(I - C_\beta)\right)^{-1} T_\gamma^* C_\beta (g_n - g)\right\|^2 \\ &= E\left\|\left(\frac{\bar{\hat{\gamma}}\hat{\varphi}_\beta}{|\hat{\gamma}|^2 + |1-\hat{\varphi}_\beta|^2}\right)(\hat{g}_n - \hat{g})\right\|^2 = c(\beta)\, E\,\|g_n - g\|^2,\end{aligned}$$

in which $c(\beta) := \sup_\xi \left(\frac{\bar{\hat{\gamma}}(\xi)\hat{\varphi}_\beta(\xi)}{|\hat{\gamma}(\xi)|^2 + |1-\hat{\varphi}_\beta(\xi)|^2}\right)^2$. Using the Riemann–Lebesgue lemma, it is easy to see that, for every fixed $\beta > 0$, the function inside the supremum takes the value 1 at $\xi = 0$, vanishes at infinity, and is continuous. Consequently, the supremum is always finite (and greater than or equal to 1). Now, applying Lemma 6 with $\alpha_n = E\,\|g_n - g\|^2$ yields the desired result. □

Remark 7. Note that this result mentions the existence of a sequence β_n, but does not specify the condition it should satisfy *jointly* with $E\|g_n - g\|^2$. Indeed, one must ensure that $c(\beta_n)E\|g_n - g\|^2$ goes to 0 as $n \to \infty$, which requires some conditions on the smoothing parameter defining the nonparametric estimator g_n. The factor $c(\beta)$ will be considered hereafter, where it will be proved that optimal rates can be reached with suitably chosen mollifier.

We now proceed to analyze the convergence rates of mollification. Let f denote the exact solution, and let $f_{\beta,n}$ denote the mollified estimated solution in (7).

In the sequel, we first consider the convergence rates in the case of power-type or exponential decay of $\widehat{\gamma}$, combined with Besov-smoothness condition for f (see Sections 3.3 and 3.4, respectively). Throughout, we assume that the *mollifier function* φ enjoys the following properties.

Assumption 8. There exists a strictly decreasing differentiable function $\Phi\colon [0,\infty) \to \mathbb{R}$ such that

$$\forall \xi \in \mathbb{R}^d, \quad \hat{\varphi}(\xi) = \Phi(|\xi|) \tag{8}$$

(in which $|\xi|$ denotes the euclidean norm of ξ) and positive constants $s, C_\Phi < \infty$ with the following properties:

$$\forall t \in [0,1], \quad \frac{1}{2} \leq \Phi(t) \leq 1, \tag{8a}$$

$$\forall t \in [0,1], \quad C_\Phi^{-1} t^s \leq 1 - \Phi(t) \leq C_\Phi t^s, \tag{8b}$$

$$\forall t \in [0,1], \quad \left|\Phi'(t)\right| \leq C_\Phi t^{s-1}, \tag{8c}$$

$$\int_0^\infty \Phi(t)^2 t^{d-1}\,\mathrm{d}t < \infty. \tag{8d}$$

Let us discuss these assumptions. A function Φ with the property (8) exists if and only if φ is isotropic, i.e., $\varphi(x)$ depends on $|x|$ only. In this case, $\hat{\varphi}$ is isotropic as well and real-valued. Note that $\Phi(0) = 1$ if and only if $\int \varphi(x)\,\mathrm{d}x = 1$. The first inequality in (8a) is a scaling condition on φ with respect to dilation. From condition (8b), the integer part of s, denoted by $\lfloor s \rfloor$, is the order of the root of $1 - \Phi$ at 0. Any density whose moments of order $\leq \lfloor s \rfloor - 1$ vanish satisfies this condition. Finally, (8d) is equivalent to $\varphi \in L^2(\mathbb{R}^d)$. Notice that the Levy kernels, defined by $\Phi(t) = \exp(-t^s)$, satisfy Assumption 8 and are positive whenever $s \in (0,2]$. Note also that Assumptions (8a) and (8b) are the same as in Fan (1991b).

Remark 9. For the sake of simplicity, we confined ourselves to isotropic *deconvolution*. This choice is by no mean necessary algorithmically, and probably also our analysis could be extended to anisotropic kernels as in Comte and Lacour (2013), Rebelles (2016), and Lepskiĭ and Willer (2019).

3.2. Besov–Nikolskiĭ spaces

We will describe the smoothness of f in terms of the scale of Besov–Nikolskiĭ spaces. These spaces will turn out to be the largest sets on which the bias of the MO

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

and MM estimators converge at a given rate. For readers not familiar with these spaces, we will recall the definition, collect a few basic properties, and provide some background in the following.

Definition 10. For a function $f \in L^2(\mathbb{R}^d)$, let $e_f(t) := \int_{|\xi|>t} \left|\widehat{f}\right|^2(\xi)\, d\xi,\ t > 0.$ We define the Besov–Nikolskiĭ space $B^u_{2,\infty}(\mathbb{R}^d)$ of smoothness index $u > 0$ as the set of all $f \in L^2(\mathbb{R}^d)$ for which

$$\|f\|_{B^u_{2,\infty}} := \left(\sup_{t>0}(1+t)^{2u} e_f(t)\right)^{1/2}$$

is finite.

It can be shown that $B^u_{2,\infty}(\mathbb{R}^d)$ equipped with the norm $\|f\|_{B^u_{2,\infty}}$ is a Banach space. For a concise introduction to more general Besov spaces $B^u_{p,q}(\mathbb{R}^d)$ with $p,q \in [1,\infty]$ and a discussion of properties relevant for statistics, we refer to the monograph Giné and Nickl (2021), in particular Sections 4.3.1 and 4.3.6.

Besov0-Nikolskiĭ spaces were introduced and analyzed by Nikol'skii (1951) in 1951 several years before the introduction of Sobolev–Slobodeckij spaces in 1958 and general Besov spaces in 1959. It was shown by Kerkyacharian and Picard (1993) that these are the largest sets in which density estimators by kernels or wavelets converge at given power-type rate. More recently, it was demonstrated in Hohage and Weidling (2017) that these spaces also characterize convergence rates of most spectral regularization methods for many important inverse problems. Note that anisotropic Besov spaces can also be defined (see Triebel, 2006).

Let us compare the Besov spaces $B^u_{2,\infty}(\mathbb{R}^d)$ to the more commonly used Sobolev spaces $H^u(\mathbb{R}^d)$ of functions $f \in L^2(\mathbb{R}^d)$ for which the norm

$$\|f\|_{H^u} := \left(\int (1+|\xi|)^{2u} |(Uf)(\xi)|^2\, \mathrm{d}\xi\right)^{1/2}$$

is finite. Under assumptions that will be introduced below, they coincide with spectral source sets $(T^*T)^\nu(L^2(\mathbb{R}^d))$, and the estimation of the bias in these spaces can conveniently be reduced to the estimation of the supremum norm of certain real-valued functions.

The Sobolev space $H^u(\mathbb{R}^d)$ is a subspace of $B^u_{2,\infty}(\mathbb{R}^d)$ since

$$(1+t)^{2u} e_f(t) \le \int_{|\xi|>t} (1+|\xi|)^{2u} |(Uf)(\xi)|^2\, \mathrm{d}\xi \le \|f\|^2_{H^u}.$$

An example of a function which belongs to $B^{1/2}_{2,\infty}(\mathbb{R})$, but not to $H^{1/2}(\mathbb{R})$ is $f(x) := \exp(-x)$ for $x \ge 0, f(x) := 0$ for $x < 0$ with Fourier transform $(Uf)(\xi) = \frac{-1}{2\pi i\xi+1}$ and $e_f(t) = O(t^{-1})$ as $t \to \infty$. By translation and superposition, any piecewise smooth function with a finite number of jumps and sufficient decay at infinity belongs to the difference set $B^{1/2}_{2,\infty}(\mathbb{R}) \setminus H^{1/2}(\mathbb{R})$. Analogously, $B^{3/2}_{2,\infty}(\mathbb{R}) \setminus H^{3/2}(\mathbb{R})$ contains piecewise smooth functions with kinks. By estimating the bias in Besov–Nikolskiĭ rather than Sobolev spaces, we can show optimal rates for such important classes of functions. Moreover, we show that these function spaces cannot be further increased.

3.3. Convergence Rates Under Power - Type Decay

In this section, we assume that the density γ of ε satisfies the following *ordinary smoothness condition*:

$$C^{-1}(1+|\xi|)^{-a} \leq \left|\hat{\gamma}(\xi)\right| \leq C(1+|\xi|)^{-a}, \quad \xi \in \mathbb{R}^d, \tag{9}$$

for some $a > 0$ and $C \geq 1$. In this case, the problem is mildly ill-posed. A classical example of density function γ satisfying condition (9) are symmetrized chi-square densities with a degrees of freedom. Another example for $a = 2$ is the Laplace distribution.[1]

Note that with condition (9), we impose strict positivity of $\hat{\gamma}$. We stress that such an assumption is only used for the purpose of deriving convergence rates. Note also that some papers have proposed modified versions of this condition to relax strict positivity, as in Hall and Meister (2007) and Delaigle et al. (2008).

We formulate below necessary and sufficient conditions for power-type convergence rates of the bias as the parameter β tends to 0.

THEOREM 11 (Approximation rates of MM). *Suppose Assumption 8 and condition (9) are satisfied. Then, for $0 < u < 2s + 2a$, the following statements are equivalent for $f \in L^2(\mathbb{R}^d)$:*

$$f \in B^u_{2,\infty}(\mathbb{R}^d), \tag{10a}$$

$$\sup_{0<\beta\leq 1} \beta^{-\frac{su}{s+a}} \|f - f_{\mathrm{MM},\beta}\| < \infty. \tag{10b}$$

Moreover, $\|f - f_{\mathrm{MM},\beta}\| = O(\beta^{2s})$ as $\beta \downarrow 0$ iff $f \in H^{2s+2a}(\mathbb{R}^d)$.

THEOREM 12 (Approximation rates of MO). *Suppose Assumption 8 and condition (9) are satisfied. Then, for $0 < u < s + a$, the following statements are equivalent for $f \in L^2(\mathbb{R}^d)$:*

$$f \in B^u_{2,\infty}(\mathbb{R}^d), \tag{11a}$$

$$\sup_{0<\beta\leq 1} \beta^{-\frac{su}{s+a}} \|f - f_{\mathrm{MO},\beta}\| < \infty. \tag{11b}$$

Moreover, $\|f - f_{\mathrm{MO},\beta}\| = O(\beta^{s})$ as $\beta \downarrow 0$ iff $f \in H^{s+a}(\mathbb{R}^d)$.

Note that in our proofs, we use both sides of the inequality in condition (9), although the first inequality might be enough. Note also that the rates of MO and

[1] Note that, in the anisotropic setting, this assumption could be adapted as in Comte and Lacour (2013).

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

MM agree for smoothness parameters $u < s+a$, but for MO the best rate is $O(\beta^s)$ for $u = s+a$, whereas MM can achieve rates up to $O(\beta^{2s})$ for higher smoothness of the solution f. This indicates that MM converges faster than MO for smooth solutions in a deterministic setting where both methods can be applied to the noisy data g. (Since the final aim is usually a rate of convergence as the deterministic data noise level tends to 0, a deterministic analysis of the propagated data error and a balancing of both error terms by a proper choice of β would be required to complete the picture.)

This discussion is substantiated by the following proposition asserting that the maximal rates $O(\beta^{2s})$ and $O(\beta^s)$ in Theorems 11 and 12 cannot be improved except for the trivial solution. This is the so-called "saturation effect" detailed below. Note, however, that the saturation bound can be increased arbitrarily by increasing the regularity of the mollifier. This result is a clear advantage of mollification over Tikhonov.

PROPOSITION 13 (Saturation). *Suppose Assumption 8 holds true.*

1. *If* $\|f - f_{\mathrm{MM},\beta}\| = o(\beta^{2s})$, *then* $f = 0$.
2. *If* $\|f - f_{\mathrm{MO},\beta}\| = o(\beta^{s})$, *then* $f = 0$.

Proof. 1. Suppose that $\|f - f_{\mathrm{MM},\beta}\| = o(\beta^{2s})$, i.e.,

$$0 = \lim_{\beta\downarrow 0} \beta^{-4s}\|f - f_{\mathrm{MM},\beta}\|^2 = \lim_{\beta\downarrow 0}\int_{\mathbb{R}^d}\left(\frac{\beta^{-2s}(1-\hat{\varphi}_\beta)^2}{|\hat{\gamma}|^2+(1-\hat{\varphi}_\beta)^2}\right)^2 |Uf|^2\,\mathrm{d}\xi,$$

the unitarity of the Fourier transform. Then, by Fatou's lemma and (8b),

$$0 = \int_{\mathbb{R}^d}\liminf_{\beta\downarrow 0}\left(\frac{\beta^{-2s}(1-\hat{\varphi}_\beta)^2}{|\hat{\gamma}|^2+(1-\hat{\varphi}_\beta)^2}\right)^2 |Uf|^2\,\mathrm{d}\xi \geq \int_{\mathbb{R}^d}\liminf_{\beta\downarrow 0}\left(\frac{\beta^{-2s}C_\Phi^{-2}(\beta|\xi|)^{2s}}{|\hat{\gamma}|^2+(1-\hat{\varphi}_\beta)^2}\right)^2 |Uf|^2\,\mathrm{d}\xi.$$

Therefore, the integrand on the right-hand side vanishes almost everywhere. Since the first factor is positive almost everywhere, it follows that $f = 0$.

2. We have

$$Uf - Uf_{\mathrm{MO},\beta} = \left(1 - \hat{\varphi}_\beta\frac{|\hat{\gamma}|^2}{|\hat{\gamma}|^2+(1-\hat{\varphi}_\beta)^2}\right)Uf = \left(v_\beta + \hat{\varphi}_\beta w_\beta\right)Uf$$

with $v_\beta := 1-\hat{\varphi}_\beta$ and $w_\beta := \frac{(1-\hat{\varphi}_\beta)^2}{|\hat{\gamma}|^2+(1-\hat{\varphi}_\beta)^2}$. The proof now follows along the lines of the first part by showing that

$$\begin{aligned}\liminf_{\beta\downarrow 0}\beta^{-s}\left(v_\beta(\xi)+\hat{\varphi}_\beta(\xi)w_\beta(\xi)\right) &= \liminf_{\beta\downarrow 0}[\beta^{-s}v_\beta(\xi)] + \lim_{\beta\downarrow 0}[\beta^s\hat{\varphi}_\beta(\xi)]\liminf_{\beta\downarrow 0}[\beta^{-2s}w_\beta(\xi)]\\ &\geq C_\Phi^{-1}\beta^{-s}(\beta|\xi|)^s + 0 = C_\Phi^{-1}|\xi|^s.\end{aligned}$$

□

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

Remark 14. Along the lines of the proof of Proposition 13, one can also show that $\|f-f_{\mathrm{MM},\beta}\| = O(\beta^{2s})$ implies $f \in H^{2a+2s}(\mathbb{R}^d)$ in Theorem 11 as

$$C_\Phi^{-2}\int_{\mathbb{R}^d} |\xi|^{4s}\,|(Uf)(\xi)|^2\,\mathrm{d}\xi \le \liminf_{\beta\downarrow 0}\beta^{-4s}\|f-f_{\mathrm{MM},\beta}\|^2 < \infty.$$

Similarly, one can show that $\|f-f_{\mathrm{MO},\beta}\| = O(\beta^s)$ implies $f \in H^{a+s}(\mathbb{R}^d)$ in Theorem 12. In other words, in the M0-limit case $u = s+a$, the Besov smoothness $f \in B^u_{2,\infty}(\mathbb{R}^d)$ is no longer sufficient for the rate $\|f-f_{\mathrm{MO},\beta}\| = O(\beta^{\frac{su}{s+a}})$, but we need the stronger Sobolev smoothness $f \in H^u(\mathbb{R}^d)$.

To obtain error bounds for observed data in terms of the sample size n, the bounds on the bias need to be complemented by bounds on the variance term.

PROPOSITION 15 (Bound on variance term). *With* $\Phi^{\mathrm{MO}}_\beta := \frac{\overline{\widehat{\gamma}}\widehat{\varphi}_\beta}{|\widehat{\gamma}|^2+(1-\widehat{\varphi}_\beta)^2}$, *we have*

$$\mathrm{E}\left(\left\|f_{\beta,n}-f_{\mathrm{MO},\beta}\right\|^2\right) \le \frac{2}{n}\|\Phi^{\mathrm{MO}}_\beta\|^2_{L^2}.$$

In particular, if Assumption 8 and condition (9) hold true and $4s \ge d-2a$, *then*

$$\mathrm{E}\left(\left\|f_{\beta,n}-f_{\mathrm{MO},\beta}\right\|^2\right) = O\left(\frac{1}{n}\beta^{-\frac{s(d+2a)}{s+a}}\right).$$

Proof. We compute

$$\mathrm{E}\left(\left\|f_{\beta,n}-f_{\mathrm{MO},\beta}\right\|^2\right) = \mathrm{E}\left(\left\|Uf_{\beta,n}-\mathrm{E}\left(Uf_{\beta,n}\right)\right\|^2\right) = \mathrm{E}\left(\left\|\Phi^{\mathrm{MO}}_\beta(\hat{g}_n-\hat{g})\right\|^2\right)$$
$$= \int_{\mathbb{R}^d}\left|\Phi^{\mathrm{MO}}_\beta(\xi)\right|^2\mathrm{Var}\left(\hat{g}_n(\xi)\right)\mathrm{d}\xi,$$

and in view of the independence of $Y_1,\dots,Y_n$, we have $\mathrm{Var}\left(\hat{g}_n(\xi)\right) = \frac{1}{n}\mathrm{Var}\left(e^{-2i\pi\langle\xi,Y_1\rangle}\right) \le \frac{2}{n}$. This implies the first inequality, and the second follows from Lemma A.6. □

Now, we can show the main result of this section, an order optimal bound on the mean integrated square error in terms of the sample size. Here and in the following, we write $\psi_1(x)\sim\psi_2(x)$ as $x\to x_0$ for two positive functions ψ_1 and ψ_2 if $\liminf_{x\to x_0}\frac{\psi_1(x)}{\psi_2(x)} > 0$ and $\limsup_{x\to x_0}\frac{\psi_1(x)}{\psi_2(x)} < \infty$.

THEOREM 16 (Convergence rate). *Suppose that Assumption 8 and condition (9) hold true, that* $4s \ge d-2a$, *and that* $f \in B^u_{2,\infty}(\mathbb{R}^d)$ *for some* $0 < u < s+a$ *or* $f \in H^{s+a}(\mathbb{R}^d)$ *for* $u = s+a$. *Then, for* $\beta \sim n^{-\frac{s+a}{2su+s(d+2a)}}$, *we obtain the optimal rate*

$$\mathrm{E}\left(\left\|f_{\beta,n}-f\right\|^2\right) = O\left(n^{-\frac{u}{u+a+d/2}}\right) \qquad \text{as } n\to\infty. \tag{12}$$

Proof. Using the bias-variance decomposition, $\mathrm{E}\left(f_{\beta,n}\right) = f_{\mathrm{MO},\beta}$, Proposition 15, and Theorem 12, we obtain

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

$$\begin{aligned}\mathrm{E}\left(\left\|f_{\beta,n}-f\right\|^2\right)&=\mathrm{E}\left(\left\|f_{\beta,n}-f_{\mathrm{MO},\beta}\right\|^2\right)+\left\|f_{\mathrm{MO},\beta}-f\right\|^2\\&=O\left(\frac{1}{n}\beta^{-\frac{s(d+2a)}{s+a}}+\beta^{\frac{su}{s+a}}\right)=O\left(n^{-\frac{u}{u+a+d/2}}\right).\end{aligned}$$ □

The rates we obtain are optimal in a minimax sense, i.e., they cannot be improved by any other method. In fact, it has been shown that even on smaller smoothness classes, no better error bounds can be achieved, e.g., on univariate Hölder classes (Fan, 1991b) or multivariate Sobolev classes (Johannes, 2009) or multivariate Hölder classes (see Comte and Lacour, 2013, which also treats Sobolev spaces and anisotropic smoothness classes).

3.4. Convergence Rates Under Exponential Decay

In this section, we assume that the density γ of ε satisfies the following *super-smoothness condition*. More precisely, we assume that there exist constants $a,\kappa>0$ and $C\geq 1$ such that

$$C^{-1}\exp\left(-\kappa\,|\xi|^a\right)\leq|\widehat{\gamma}(\xi)|^2\leq C\exp\left(-\kappa\,|\xi|^a\right),\quad \xi\in\mathbb{R}^d. \tag{13}$$

In this case, the problem is severely ill-posed. Note that $a=2$ corresponds to Gaussian errors ε and $a=1$ to Cauchy errors.

Note that with condition (13), as for condition (9), we impose strict positivity of $\hat{\gamma}$. This condition could be relaxed as in Hall and Meister (2007) and adapted to the anisotropic setting (as in Comte and Lacour, 2013).

When $\hat{\gamma}$ satisfies condition (13), we only obtain logarithmic convergence rates for the bias under finite smoothness assumptions on *f*. Furthermore, no saturation effects occur in this setting.

THEOREM 17 (Approximation rates). *Suppose that Assumption 8 and condition (13) hold true with $s>\frac{1}{2}$. Then, the following statements are equivalent for $u>0$:*

$$f\in B^u_{2,\infty}(\mathbb{R}^d), \tag{14a}$$

$$\sup_{0<\beta<1}(-\ln\beta)^{u/a}\left\|f-f_{\mathrm{MM},\beta}\right\|<\infty, \tag{14b}$$

$$\sup_{0<\beta<1}(-\ln\beta)^{u/a}\left\|f-f_{\mathrm{MO},\beta}\right\|<\infty. \tag{14c}$$

Note that, as for condition (9), we again use in the proofs both sides of condition (13), although the first inequality might be enough. For the variance term, a rather coarse estimate is sufficient.

PROPOSITION 18 (Bound on variance term). *If condition (13) holds true, then, for any $b>4s$, the statistical error satisfies*

$$\mathrm{E}\left(\left\|f_{\beta,n}-\mathrm{E}\left(f_{\beta,n}\right)\right\|^2\right)=O\left(\frac{1}{n}\beta^{-2b}\right).$$

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

Proof. Due to Proposition 15, we can reduce the proof to showing that

$$\int_{\mathbb{R}^d} |\zeta_\beta(\xi)|^2 |\hat{\varphi}_\beta(\xi)|^2 \,\mathrm{d}\xi = O(\beta^{-b}) \quad \text{with} \quad \zeta_\beta(\xi) := \frac{\overline{\widehat{\gamma}(\xi)}}{|\widehat{\gamma}(\xi)|^2 + (1-\widehat{\varphi}_\beta(\xi))^2}.$$

Due to assumption (8d), it suffices to show that

$$\|\zeta_\beta\|_\infty = O(\beta^{-b}). \tag{15}$$

On the one hand, we have

$$\frac{|\zeta_\beta(\xi)|}{\beta^{-b}} \le \frac{1}{\beta^{-b}\,|\hat{\gamma}(\xi)|} \le C\beta^b \exp(\kappa\beta^a) \qquad \text{if } |\xi| \le \beta,$$

and, on the other hand, using the monotonicity of Φ and (8b), we have

$$\frac{|\zeta_\beta(\xi)|}{\beta^{-b}} \le \frac{\beta^b}{(1-\Phi(\beta\,|\xi|))^2} \le \frac{\beta^b}{(1-\Phi(\beta^2))^2} \le C_\Phi \beta^{b-4s} \qquad \text{if } |\xi| > \beta.$$

As $b > 4s$, these bounds imply (15). □

Combining these results yields the following logarithmic convergence rates with respect to the sample size.

THEOREM 19 (Convergence rates). *Suppose that Assumption 8 and condition (13) hold true with $s > \frac{1}{2}$. Let $f \in B^u_{2,\infty}(\mathbb{R}^d)$ for some $u > 0$, and let $\beta = \frac{1}{n}$. Then*

$$\mathrm{E}\left(\left\|f_{\beta,n} - f\right\|^2\right) = O\left((\ln n)^{-2u/a}\right) \qquad \text{as } n \to \infty.$$

Proof. This follows from the bias-variance decomposition

$$\mathrm{E}\left(\left\|f_{\beta,n} - f\right\|^2\right) = \mathrm{E}\left(\left\|f_{\beta,n} - f_{\mathrm{MO},\beta}\right\|^2\right) + \left\|f_{\mathrm{MO},\beta} - f\right\|^2$$

using $f_{\mathrm{MO},\beta} = f_{\mathrm{MO},\beta}$, Theorem 17, and Proposition 18. □

The rates in Theorem 19 are again minimax, and similar to Fan (1991b) or Comte and Lacour (2013) (for the multidimensional setting assuming isotropy), but only logarithmic.

To achieve faster than logarithmic rates for smooth noise distributions satisfying (13), one would have to assume that f belongs to smoothness classes consisting of infinitely differentiable or even analytic functions (see Pensky and Vidakovic, 1999; Butucea, 2004; Butucea and Tsybakov, 2008).

4. REGULARIZATION PARAMETER SELECTION

In this section, we propose fully data-driven choice of the regularization parameter and show that it achieves a bias-variance compromise. This empirical method will be implemented in Section 6.

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

4.1. The Lepskiĭ Balancing Principle

In the previous convergence results, the choice of the regularization parameter β requires the a priori knowledge of the smoothness of the unknown function f. To adapt to the unknown smoothness of f, we need to use an a posteriori parameter choice rule. In what follows, we present the Lepskiĭ principle, developed in Lepskiĭ (1991, 1992, 1993), and simplified in the context of inverse problems in Mathé and Pereverzev (2003) or Mathé (2006). The method starts with the well-known error decomposition

$$\left\| f_{\beta,n} - f \right\| \leq \left\| f_{\beta,n} - f_{\mathrm{MO},\beta} \right\| + \left\| f_{\mathrm{MO},\beta} - f \right\|, \quad \text{for all} \quad \beta > 0. \tag{16}$$

We seek to choose β such that the right-hand side of (16) is minimal. The main issue is that the bias $\left\| f_{\mathrm{MO},\beta} - f \right\|$, which is increasing in β, is typically unknown, whereas for the variance term $\mathrm{E}\left(\left\| f_{\beta,n} - f_{\mathrm{MO},\beta} \right\|^2 \right)$, which is decreasing in β, we have established a bound in Proposition 15. We confine β to a finite equidistant grid on a logarithmic scale given by $B_n := \{\beta_0 q^j : j = 0, \dots, J_n\}$, for some $0 < q < 1$ and some $\beta_0 > 0$, and J_n is chosen as the smallest integer such that $\sqrt{2/n}\| \Phi^{\mathrm{MO}}_{\min B_n} \| > 1$. The Lepskiĭ principle attempts to balance the two terms on the right-hand side of (16) by determining the largest $\tilde{\beta} \in B_n$ such that all differences $f_\beta - f_{\tilde{\beta}}$, for $\beta \in B_n$ with $\beta < \tilde{\beta}$, are dominated in norm by the noise bound from Proposition 15:

$$\beta^\star := \max \left\{ \tilde{\beta} \in B_n : \left\| f_{\beta,n} - f_{\tilde{\beta},n} \right\| \leq \frac{2\sqrt{2}\theta}{\sqrt{n}} \left\| \Phi^{\mathrm{MO}}_{\beta} \right\| \text{ for all } \beta \in B_n \text{ with } \beta < \tilde{\beta} \right\} \tag{17}$$

for some parameter $\theta > 0$.

4.2. Convergence Result

To prove error bounds for the Lepskiĭ balancing principle, we need a large deviation inequality for the variance term.

PROPOSITION 20. *Let C_{MO} be the constant from Lemma A.6, and let $C_{\mathrm{c}} := \frac{\sqrt{8}+34.5}{2\sqrt{2}}$. Then, for all $\rho > 1$ and for all $n \in \mathbb{N}$, we have*

$$\mathrm{P}\left[\sup_{0<\beta\leq\beta_0} \beta^{\frac{s(a+d/2)}{s+a}} \| f_{\beta,n} - f_{\mathrm{MO},\beta} \right] \| \geq \frac{\rho 2\sqrt{2} C_{\mathrm{MO}}}{\sqrt{n}} \leq \exp\left(-\frac{\rho}{C_{\mathrm{c}}} \right). \tag{18}$$

With this, we can prove order optimal rates up to a logarithmic factor.

THEOREM 21. *Under the assumptions of Theorem 16, the expected error for the regularization parameter β^* determined by the Lepskiĭ balancing principle (17) with $\theta \geq 2$ satisfies*

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

$$\mathrm{E}\left(\|f_{\beta^*,n}-f\|^2\right)=O\left(\left(\frac{\ln n}{\sqrt{n}}\right)^{\frac{2u}{u+a+d/2}}\right) \qquad \textit{as } n\to\infty.$$

Proof. We define an event *good event* A_n that $\sup_{0<\beta\leq\beta_0}\beta^{\frac{s(a+d/2)}{s+a}}\|f_{\beta,n}-f_{\mathrm{MO},\beta}\|\leq\frac{C_c(\ln n)2\sqrt{2}C_{\mathrm{MO}}}{\sqrt{n}}$. Then, by Proposition 20 with $\rho=\rho(n)=C_c\ln n$, its complement A_n^c, i.e., the *bad event*, has probability $\mathrm{P}\left[A_n^c\right]\leq\frac{1}{n}$.

Note that, at least for sufficiently large n, the minimum of the right-hand side of (16) is attained at a parameter β in $[\min B_n, \max B_n]$. Therefore, the deterministic oracle inequality in Mathé (2006, Cor. 1) shows that, in the event A_n, the error is bounded by

$$\|f_{\beta^*,n}-f\|=O\left(\left(\frac{\ln n}{\sqrt{n}}\right)^{\frac{u}{u+a+d/2}}\right),$$

where the logarithm appears in comparison to Theorem 16 since in the noise bound $1/\sqrt{n}$ is replaced by $\ln n/\sqrt{n}$ compared to Proposition 15 due to our choice of $\rho(n)$.

Using (18), it is easy to see that

$$\mathrm{E}\left(\|f_{\min B_n,n}-f_{\mathrm{MO},\,\min B_n}\|\,\middle|A_n^c\right)\leq C\frac{\ln n}{\sqrt{n}}(\min B_n)^{\frac{-s(a+d/2)}{s+a}}\leq C'\ln n,$$

where the definition of B_n was used in the second inequality. Using the error decomposition (16) and the monotonicity of the error components, we finally obtain

$$\begin{aligned}
&\mathrm{E}\left(\|f_{\min B_n,n}-f_{\mathrm{MO},\,\min B_n}\|\right)\\
&=\mathrm{E}\left(\|f_{\min B_n,n}-f_{\mathrm{MO},\,\min B_n}\|\,\middle|A_n\right)\mathrm{P}[A_n]+\mathrm{E}\left(\|f_{\min B_n,n}-f_{\mathrm{MO},\,\min B_n}\|\right)\left|A_n^c\,\mathrm{P}\left[A_n^c\right]\right.\\
&\leq O\left(\left(\frac{\ln n}{\sqrt{n}}\right)^{\frac{u}{u+a+d/2}}\right)+\left(\|f_{\mathrm{MO},\beta_0}-f\|+C'\ln n\right)\frac{1}{n}\;=\;O\left(\left(\frac{\ln n}{\sqrt{n}}\right)^{\frac{u}{u+a+d/2}}\right).
\end{aligned}$$

□

5. FRAMEWORK FOR COMPARISON

The purpose of this section is to put the various approaches (mollification, Tikhonov, deconvolution kernels, and spectral cut-off) in a unified framework and, by doing so, to derive theoretical and numerical tools for their comparison and assessment.

5.1. The Filtering Viewpoint

5.1.1. *Deconvolution Kernels.* The deconvolution kernels were introduced by Stefanski and Carroll (1990) in the late 80s, in the case $d=1$. In essence, the deconvolution kernel estimator stabilizes the reconstruction by *bounding* the function $1/\hat{\gamma}$. More precisely, using our notation system, and ignoring (temporarily) the discrete aspects carried by the estimation process, the reconstructed density can

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

be written as

$$f_{\mathrm{DK}} = U^{-1}\left[\frac{\hat{\varphi}_h}{\hat{\gamma}}\right]Ug \tag{19}$$

or, equivalently, by

$$\hat{f}_{\mathrm{DK}}(\xi) = \frac{\hat{\varphi}(h\xi)}{\hat{\gamma}(\xi)}\hat{g}(\xi) = \frac{\hat{\varphi}_h(\xi)}{\hat{\gamma}(\xi)}\hat{g}(\xi), \tag{20}$$

where φ_h is defined as $\varphi_h(x) := \frac{1}{h}\varphi\left(\frac{x}{h}\right)$ with $h > 0$. Here, φ is the *deconvolution kernel* (denoted by K in the original paper by Stefanski and Carroll, (1990), and in the subsequent literature) and h is the regularization parameter. The corresponding filter Φ is here defined by $\Phi(\xi) = \hat{\varphi}_h(\xi)/\hat{\gamma}(\xi)$. We see right away that, for the solution to be well defined and stable, it is necessary that:

(a) the function $\hat{\gamma}$ does not vanish;
(b) for every $h > 0$, the function $\xi \longmapsto \hat{\varphi}(h\xi)/\hat{\gamma}(\xi)$ is bounded.

Notice that, whenever γ is symmetric, $\hat{\gamma}$ is real, and hence condition (a) entails strict positivity of $\hat{\gamma}$. As a matter of fact, the Riemann–Lebesgue lemma tells us that $\hat{\gamma}$ is continuous and vanishes at infinity (and of course $\hat{\gamma}(0) = 1$). At all events, the above assumptions are somewhat restrictive, obviously, and constitute a serious limitation of the methodology.

Remark 22. In the original paper by Stefanski and Carroll (1990), it was also requested that, for every $h > 0$, $\int \left|\frac{\hat{\varphi}_h(\xi)}{\hat{\gamma}(\xi)}\right| \mathrm{d}\xi < \infty$. This assumption ensures that the function $\big(\hat{\varphi}_h(\xi)/\hat{\gamma}(\xi)\big)\hat{g}(\xi)$ in equation (20) is integrable, allowing for application of the inverse Fourier *integral* transform. However, the latter function is square-integrable anyways, allowing for the application of the inverse Fourier–Plancherel operator. This is why this assumption can be relaxed.

Remark 23. The deconvolution kernels can be regarded as a particular instance of the *approximate inverses*, a regularization method that was introduced, in the deterministic setting, in Louis and Maass (1990), and gave rise to a large amount of developments. See Appendix B.

5.1.2. *Spectral Cutoff.* In the spectral cutoff method, usually designed for the case $d = 1$ (see, e.g., Johannes, 2009, and the references therein), the Fourier transform of the reconstructed density is merely truncated whenever $\hat{\gamma}(\xi)$ falls below a threshold $a > 0$, which plays the role of the regularization parameter. The spectral cutoff may be regarded as a special case of the deconvolution kernels, in which the regularization parameter is *indexed* on the convolution kernel γ: the solution is defined as

$$f_{\mathrm{SC}} = U^{-1}\left[\frac{1_{|\hat{\gamma}|^2 \geq a}}{\hat{\gamma}}\right]Ug$$

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

or, equivalently, as $\hat{f}_{\mathrm{SC}}(\xi) = \frac{\mathbb{1}_{\{|\hat{\gamma}|^2 \geq a\}}(\xi)}{\hat{\gamma}(\xi)} \hat{g}(\xi)$. Here, $\mathbb{1}_S$ denotes the indicator function of the set S, that is, the function which takes the value 1 if the argument belongs to S and 0 otherwise. Therefore, the spectral cutoff is also a filter-type regularization method, with

$$\Phi(\xi) = \frac{\mathbb{1}_{\{|\hat{\gamma}|^2 \geq a\}}(\xi)}{\hat{\gamma}(\xi)}.$$

The spectral cutoff produces a solution that belongs to the class of *band-limited* functions. The inverse Fourier transform of the kernel $\mathbb{1}_{\{|\hat{\gamma}|^2 \geq a\}}$ is not a density function, and its behavior is similar to the well-known sinc function. We will see in Section 6 that this specific choice introduces additional perturbations (a Gibbs-like phenomenon) in the reconstruction. Note that the numerator of the filter Φ could be replaced by a smooth cutoff that smoothly goes to zero beyond a. That would remove the Gibbs phenomena and place us back into the framework of kernel methods.

5.1.3. *Tikhonov Regularization.* The Tikhonov regularization (see Tikhonov and Arsenin, 1977) has been applied and studied in the context of econometrics in Carrasco et al. (2007) and Carrasco and Florens (2011). The Tikhonov solution is defined by $f_{\mathrm{TK}} = \left(T_\gamma^* T_\gamma + \alpha I\right)^{-1} T_\gamma^* g$ or, equivalently, by

$$f_{\mathrm{TK}} = U^{-1} \left[\frac{\bar{\hat{\gamma}}}{|\hat{\gamma}|^2 + \alpha} \right] Ug.$$

Expressed in the Fourier domain, we get $\hat{f}_{\mathrm{TK}}(\xi) = \frac{\overline{\hat{\gamma}(\xi)}}{|\hat{\gamma}(\xi)|^2 + \alpha} \hat{g}(\xi)$. The solution is well defined for every positive value of the regularization parameter α. Therefore, the Tikhonov method can be regarded as a filter-type technique (for deconvolution), with

$$\Phi(\xi) = \frac{\overline{\hat{\gamma}(\xi)}}{|\hat{\gamma}(\xi)|^2 + \alpha}.$$

Note that Carrasco and Florens (2011) define their solution in weighted L^2-spaces in order to recover the compactness of the operator T_γ, which we do not need in our context.

5.2. The Variational Viewpoint

For simplicity of the notation, T_γ is now denoted by T. Consider the generic functional

$$\mathscr{F}(f) := \frac{1}{2} \|Pg - Tf\|^2 + \frac{\alpha}{2} \|Hf\|^2,$$

in which P and H are bounded operators. By specifying the operators P and H as well as the parameter α, we can retrieve the aforementioned methods. Recall

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

first that, under the following *complementation condition* (see Morozov, 1984, also mentioned in Engl, Hanke, and Neubauer, 1996): $\exists\mu > 0\colon\ \forall f \in F,\ \|Tf\|^2 + \|Qf\|^2 \geq \mu\,\|f\|^2$, the operator $T^*T + H^*H$ admits a bounded inverse, and that, consequently, the unique minimizer of $\mathscr{F}$, namely

$$\bar{f} = \left(T^*T + \alpha H^*H\right)^{-1} T^*Pg$$

depends continuously on g. Since $T = U^{-1}\left[\hat{\gamma}\right]U$, it is readily seen that $T^* = U^{-1}\left[\overline{\hat{\gamma}}\right]U$ and $T^*T = U^{-1}\left[\left|\hat{\gamma}\right|^2\right]U$. Therefore:

(1) Letting $H = P = I$ obviously yields the Tikhonov functional.
(2) The choice $H = I - C_\beta$, $P = C_\beta = U^{-1}\left[\hat{\varphi}_\beta\right]U$, and $\alpha = 1$ yields the mollification functional.
(3) Letting $\alpha = 0$ and $P = C_h$ (where C_h denotes the operator of convolution with φ_h) yields the functional $\mathscr{F}(f) = \|C_h g - Tf\|^2/2$, whose unique minimizer is

$$T^\dagger C_h g = (T^*T)^{-1} T^* C_h g = U^{-1}\left[\frac{\hat{\varphi}_h}{\hat{\gamma}}\right]Ug,$$

which turns out to be the deconvolution kernel solution f_{DK}.
(4) Letting $\alpha = 0$ and P be the convolution by $U^{-1}\mathbb{1}_{|\hat{\gamma}|^2 \geq a}$ yields the spectral cutoff solution f_{SC}.

Notice first that the convolution kernels corresponding to the above three filters need not be real (if γ is not symmetric) or even positive, in case they are real. Notice also that, in the Tikhonov case, $\Phi(0) = (1+\alpha)^{-1} \neq 1$, which implies the additional drawback that the integral of the reconstructed function will not be equal to one, as one should expect from a density. We may then consider the modified Tikhonov filter

$$\Phi(\xi) = \frac{(1+\alpha)\overline{\hat{\gamma}}(\xi)}{\left|\hat{\gamma}(\xi)\right|^2 + \alpha}.$$

In this modified version, we merely let $P = [(1+\alpha)\mathbb{1}]$ instead of $P = I = [\mathbb{1}]$, in which $\mathbb{1}$ denotes the function identically equal to 1. As for the mollification approach, it would also make sense to consider a version with $P = I$, letting the regularization be operated by $H = I - C_\beta$ only. The corresponding filter is easily shown to be

$$\Phi(\xi) = \frac{\overline{\hat{\gamma}}(\xi)}{\left|\hat{\gamma}(\xi)\right|^2 + \left|1 - \hat{\varphi}_\beta(\xi)\right|^2},$$

and we shall refer to the corresponding method as the *modified mollification*. Finally, notice that the particular kernel corresponding to the spectral cutoff method is the inverse Fourier transform of the function $\mathbb{1}_{\{|\hat{\gamma}|^2 \geq a\}}/\hat{\gamma}$, and that the regularization parameter is now $a > 0$. Table 1 gives an overview of the functionals and filters associated with each regularization method.

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

Table 1. Overview of regularization methods for the deconvolution problem

	Functional $\mathscr{F}$	Filter Φ
TK	$\frac{1}{2}\|g-\gamma*f\|^2+\frac{\alpha}{2}\|f\|^2$	$\frac{\overline{\hat{\gamma}}}{\|\hat{\gamma}\|^2+\alpha}$
MT	$\frac{1}{2}\|(1+\alpha)g-\gamma*f\|^2+\frac{\alpha}{2}\|f\|^2$	$\frac{(1+\alpha)\overline{\hat{\gamma}}}{\|\hat{\gamma}\|^2+\alpha}$
MO	$\frac{1}{2}\left\|\varphi_\beta*g-\gamma*f\right\|^2+\frac{1}{2}\left\|f-\varphi_\beta*f\right\|^2$	$\frac{\overline{\hat{\gamma}}\hat{\varphi}_\beta}{\|\hat{\gamma}\|^2+\|1-\hat{\varphi}_\beta\|^2}$
MM	$\frac{1}{2}\|g-\gamma*f\|^2+\frac{1}{2}\left\|f-\varphi_\beta*f\right\|^2$	$\frac{\overline{\hat{\gamma}}}{\|\hat{\gamma}\|^2+\|1-\hat{\varphi}_\beta\|^2}$
DK	$\frac{1}{2}\|\varphi_h*g-\gamma*f\|^2$	$\frac{\hat{\varphi}_h}{\hat{\gamma}}$
SC	$\frac{1}{2}\left\|\check{\mathbb{1}}_{\{\|\hat{\gamma}\|^2\geq a\}}*g-\gamma*f\right\|^2$	$\frac{\mathbb{1}_{\{\|\hat{\gamma}\|^2\geq a\}}}{\hat{\gamma}}$

Notes: TK stands for Tikhonov, MT for modified Tikhonov, MO for mollification, MM for modified mollification, DK for the deconvolution kernels, and SC for the spectral cutoff.

5.3. Comparisons

We have seen that mollification reaches optimal convergence rates, along with its competitors: the deconvolution kernels, the spectral cutoff, and the Tikhonov regularization. The purpose of this section is to compare mollification to the other approaches from other viewpoints.

5.3.1. *Deconvolution Kernels Versus Mollification.* We emphasize that, obviously, the regularization parameters h and β have the same interpretation. The first obvious limitation of the deconvolution kernels is the restriction imposed by the decrease of $\hat{\gamma}$ and its strict positivity. For example, if γ is Gaussian, the decrease of $\hat{\varphi}$ at infinity should be faster, which discards many deconvolution kernels. An even more extreme example is provided by the convolution kernel $\gamma(x)=\operatorname{sinc}^2(\pi x)$. Its Fourier transform is the triangle function

$$\hat{\gamma}(\xi)=\begin{cases}\xi+1, & \text{if } \xi\in[-1,0),\\ -\xi+1, & \text{if } \xi\in[0,1),\\ 0, & \text{elsewhere.}\end{cases}$$

Here, the convolution operator T_γ fails to be injective, as a consequence of the fact that the support of $\hat{\gamma}$ is the interval $[-1,1]$, and the deconvolution kernel solution cannot be defined. By contrast, the mollification solution is defined for any even mollifier $\varphi\in L^1(\mathbb{R})$ and any positive value of β. Moreover, in the variational formulation of the deconvolution kernels, the regularization appears only in the fit term, so that the optimization problem remains ill-posed. This may be an obstacle to the introduction of additional constraints, such as the positivity of the reconstruction, since such a constraint can be introduced only in the variational form. On the contrary, the variational form of the mollification approach is stable, thanks to the regularization term, and such a constraint may therefore be safely introduced. The only price to be paid would then be a different numerical strategy (based on optimization). We emphasize here that deconvolution kernels and mollification both aim at reconstructing an explicit object, which we refer to as the *target object*, namely, $\varphi_\beta * f$. This is why both estimators depend on one smoothing parameter only. In addition, as will be illustrated in Section 6, whenever the deconvolution kernels regularization is well-defined, the performance of both the deconvolution kernels and mollification approaches are similar, in terms of the tradeoff between fidelity and stability.

5.3.2. *Spectral Cutoff Versus Mollification.* Unlike the case of deconvolution kernels, the spectral cutoff solution remains defined when $\hat{\gamma}$ vanishes. The target object, in the sense defined above, is $\psi_a * f$, with $\psi_a = U^{-1}\mathbb{1}_{\{|\hat{\gamma}|^2 \geq a\}}$. The function ψ_a can be regarded as a *target* impulse response of the reconstruction. Its definition relies not only on the regularization parameter a, but also on the shape of $\hat{\gamma}$. We stress here that, as the inverse Fourier transform of some indicator function, this impulse response may have poor morphological properties, and may induce Gibbs-like oscillations in the reconstructed density. These oscillations may incidentally produce significant negative parts, which is a serious drawback for probability densities. By contrast, the mollification approach enables to choose an *apodized* target impulse response, by avoiding sharp edges in the Fourier domain. This will be illustrated in Section 6. Notice at last that, as in the case of the deconvolution kernels, the variational form of the spectral cutoff has no regularization term. Again, this may be an obstacle to the introduction of additional constraints, such as the positivity of the reconstruction (see the discussion in the previous paragraph).

5.3.3. *Tikhonov Versus Mollification.* Unlike deconvolution kernels, spectral cutoff, and mollification, Tikhonov regularization does not appeal to any target object, which is a conceptual drawback. The regularization is *uniformly* exercised in the Fourier domain, as can be seen from the variational formulation. Using the well-known fact that the Fourier–Plancherel operator is an isometry, the Tikhonov solution is readily seen to be the minimizer of $\mathscr{F}_{TK}(f) = \frac{1}{2}\left\|\hat{g} - \hat{\gamma}\cdot\hat{f}\right\|^2 + \frac{\alpha}{2}\left\|\hat{f}\right\|^2$. The penalty term attracts $\hat{f}$ toward zero everywhere. This contradicts the action of the fit term in the low-frequency domain, where both $\hat{g}$ and $\hat{\gamma}$ are not expected to be close to zero. This opposition between the fit and regularization terms may induce an unfavorable tradeoff between stability and fidelity to the model. The mollification approach avoids this pitfall by introducing a *smooth disjunction* of the realms of action of the fit and regularization terms, as can be seen from the transposition of the mollification functional in the Fourier domain: $\mathscr{F}_{MO}(f) = \frac{1}{2}\left\|\hat{\varphi}_\beta\cdot\hat{g} - \hat{\gamma}\cdot\hat{f}\right\|^2 + \frac{1}{2}\left\|(1-\hat{\varphi}_\beta)\hat{f}\right\|^2$. The resulting improvement in the tradeoff between stability and fidelity to the initial model equation will be illustrated by

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

means of simulations in Section 6. A nice aspect of Tikhonov, that is shared with mollification, is that, unlike the deconvolution kernels and the spectral cutoff, the variational formulation is stabilized, which opens the way to the introduction of the positivity constraint.

Remark 24. Hybrid kernel ridge regularization has been considered in Hall and Meister (2007), Delaigle et al. (2008), and Trong and Phuong (2015) so as to relax strict positivity of the error density. In particular, in Delaigle et al. (2008), the corresponding estimator combines the deconvolution kernel approach with a Tikhonov-type estimator. This hybrid method entails the choice of two regularization parameters. In the variational approach of mollification as defined in equation (3), the two parts of the functional to be minimized contribute to the same objective: that of reconstructing a smoothed version of the object of interest. As a result, mollification has only one regularization parameter.

6. SIMULATIONS

Having derived, in the previous section, a common framework for all reconstruction methods under consideration, we now proceed to develop tools for their assessment and comparison, in terms of the tradeoff between stability and fidelity. In all cases, $f_{\mathrm{REG}} = U^{-1}[\Phi]Ug$, in which the filter Φ depends on regularization parameters. If g is replaced by its estimation g_n, the corresponding reconstruction is denoted by $f_{\mathrm{REG},n}$. Otherwise expressed, $f_{\mathrm{REG},n} := U^{-1}[\Phi]Ug_n$.

6.1. Assessment of the Various Regularization Methods

We now define the quantities to be used for the assessment of the various regularization methods. Concerning the fidelity, a meaningful quantity is the *reconstruction error*

$$\begin{aligned} f_{\mathrm{REG},n} - f &= \left(f_{\mathrm{REG},n} - f_{\mathrm{REG}}\right) + \left(f_{\mathrm{REG}} - f\right) \\ &= U^{-1}[\Phi]U(g_n - g) + \left(f_{\mathrm{REG}} - f\right). \end{aligned} \quad \textbf{(21)}$$

In the right-hand side, the first and second terms will be, respectively, called the *statistical error* and *regularization error*. The L^2-norm of the reconstruction error, referred to as the *reconstruction*-rise (Root Integrated Square Error), will be one important indicator to compare the performances of the main four regularization methods: deconvolution kernels, spectral cutoff, Tikhonov regularization, and mollification.

Obviously, the reconstruction- rise depends on the value of the regularization parameter of the considered regularization method (either h, a, α, or β), and their range of variations as well as their impact on the solution (through the chosen regularization method) may not be easily comparable. That is why we have also introduced a common criterion to evaluate the stability of the reconstruction, which we define below.

In our setting, the reconstructed density depends linearly on the data g, and the error on the data $g_n - g$ is potentially amplified by the action of the reconstruction operator $U^{-1}[\Phi]U$ by a factor equal to its operator norm (see equation (21) as an illustration). Since U is unitary, the operator norm of the reconstruction operator $U^{-1}[\Phi]U$ is equal to the L^∞-norm of Φ (see Proposition A.1 in Appendix A). The stability of the reconstruction can then be estimated via the computation of $\|\Phi\|_\infty$ as a function of the regularization parameters. We may refer to $\|\Phi\|_\infty$ as an *instability index*.

Therefore, in Section 6.2, we study the performance of the regularization methods by comparing the variation of the reconstruction- rise with respect to the *instability index*, whose level can be arbitrarily fixed for all regularization methods. To any fixed level of the instability index corresponds a fixed value of the regularization parameter. Notice that the reconstruction error depends on the true value f, unobserved in practice. The *residual* $g_n - T_\gamma f_{\mathrm{REG},n}$ may then serve the purpose of evaluating the fidelity to the original model. Some additional plots to evaluate the performance of the regularization methods are provided using the *residual* rise with respect to the instability index.

We also provide in Section 6.3 a comparison using the optimal *oracle choice* of the regularization parameter for each method, which corresponds to the setting where the estimators should perform the best. Under this *oracle framework*, we perform Monte Carlo (MC) simulations and check that, on average, our estimator still outperforms the other methods. We also check that our empirical selection rule for the regularization parameter is working well in practice.

One way to compare the performances of the four approaches is to plot the various components of the rise and the residual rise versus the reached instability index. The results are shown in Figure 2. Note that the top-left panel allows to find for each method the optimal instability index minimizing the reconstruction rise and is only available in a simulated data framework (the true f is known), whereas the bottom-right panel with the residuals is available in a real data framework and can be used to find the optimal method for a given level of the instability index.

6.2. Comparison Using the Instability Index

We illustrate the comparison between the various regularization approaches with an example. The signal f is a Beta$(3,2)$ density function, rescaled for having the support $[-2,2]$ (note that its standard deviation is $\sigma_f = 0.8$). The sample size is $n = 500$, and the noise γ is a Cauchy$(0,\sigma)$, where σ is the scale, i.e., Cauchy$(0,\sigma) = \sigma t_{(1)}$, where $t_{(1)}$ is a student-t with one degree of freedom.[2] Since the standard deviation does not exist, we choose the scale $\sigma = 0.20$ such that this Cauchy has an interquartile range (IQR) equal to 0.40.

We choose for the mollification and the deconvolution kernel techniques a normal density with regularization parameters β and h, respectively. So, in this

[2]Note that other cases have been implemented, for example, when noise γ is an $\mathrm{N}(0,\sigma^2)$ where $\sigma = 0.5 * \sigma_f = 0.40$. Since the results were similar, and in order to save space, we do not present these additional results in the paper.

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

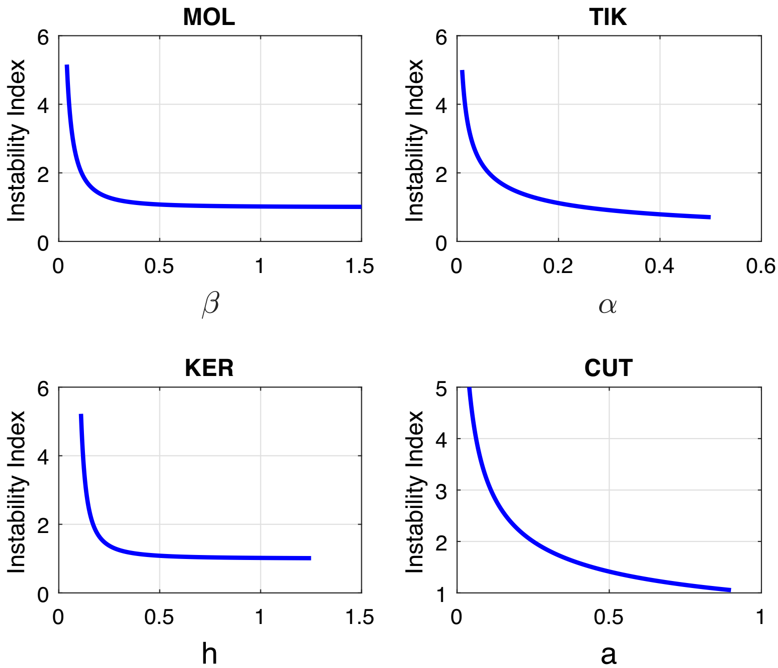

FIGURE 1. The instability index as a function of the regularization parameters.

case, where the four approaches described above can be used, we expect a quite similar behavior between the mollification and deconvolution kernel approaches. Figure 1 shows how the instability index varies with the regularization parameter in the four methods. In all the cases, as it should, the methods are more stable when increasing the regularization parameter.

6.3. Comparison Using Oracle Regularization Parameter Choice

The Root Integrated Squared Errors (rise) of the total error has a minimum for an optimal oracle value of the parameters. This oracle choice for the regularization parameters is obviously not reachable in practice since the rise depends on the true unknown function f. So, in practice, it is mandatory to use an empirical selection rule, such as the Lepskiĭ method proposed in Section 4. In what follows, we use this oracle framework as another way to assess the various regularization methods.

First, direct and precise calculations for this particular sample of size $n = 500$ give the values shown in Table 2. We observe that the best performance is achieved by the mollification approach. For this particular sample of size $n = 500$, it is also interesting to show how the function f can be reconstructed in the various approaches (with the optimal oracle values of the regularization parameters). This

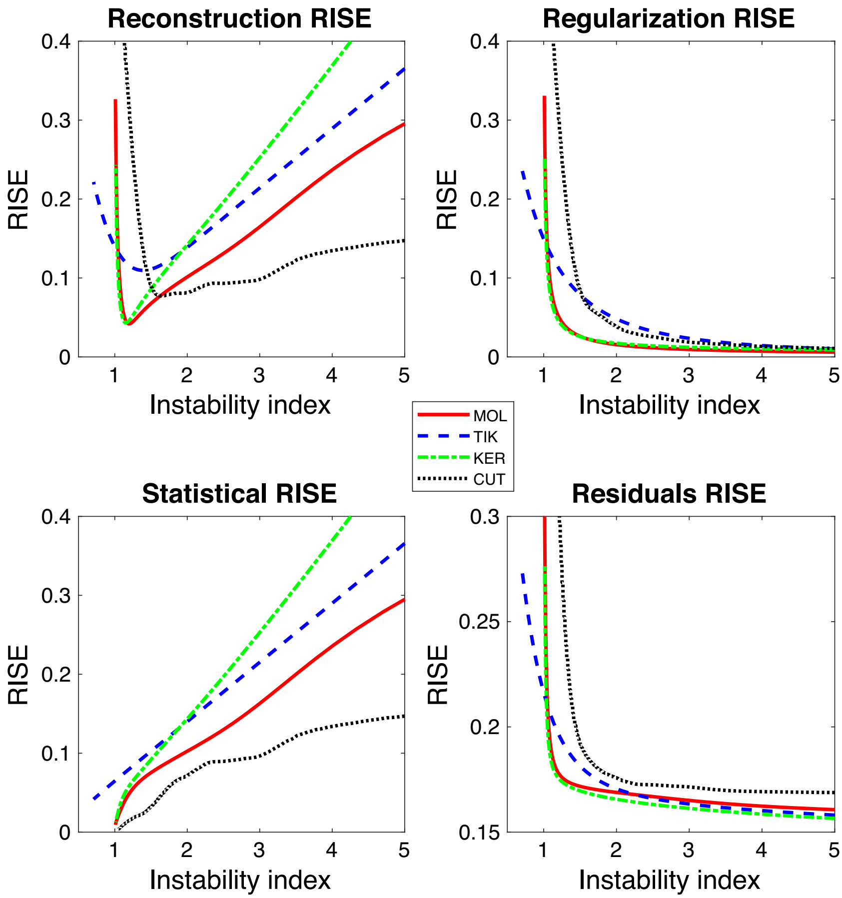

FIGURE 2. Various components of the rise and the corresponding values of the residual rise as a function of the instability index.

TABLE 2. Reconstruction- rise for the four methods and values of optimal regularization parameter.

Method	Parameter	Rec- rise
MOL	$\beta = 0.30346$	0.042010
TIK	$\alpha = 0.12912$	0.109681
KER	$h = 0.37571$	0.043268
CUT	$a = 0.36546$	0.077186

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

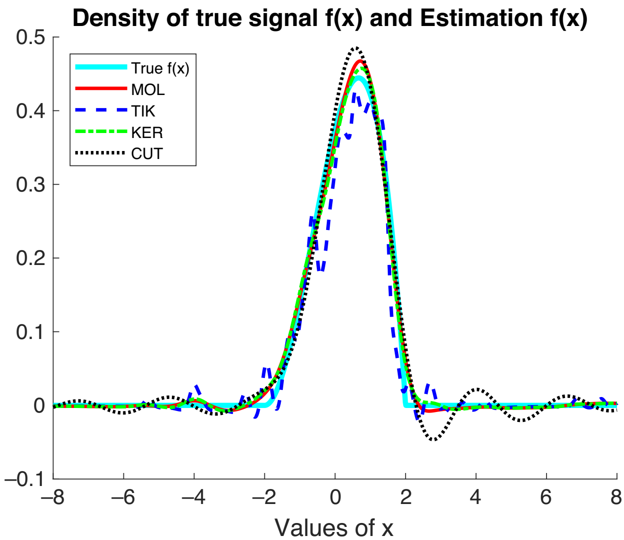

FIGURE 3. Different estimates of the signal f and its true value.

is shown in Figure 3, where the various estimates are displayed along with the true value of the density f. We see how irregular is the estimate in the Tikhonov approach and the Gibbs effect of the cutoff approach for values of $|x|$ greater than 2. Again, we see that the mollification and deconvolution kernel approaches give very similar and good results.

Then, for the case illustrated previously, we perform an MC experiment by simulating a large number of times, say M, a sample of size n, and compare the performance of the different approaches by recording the achieved minimal reconstruction rise for each approach, and then averaging over the M simulated samples. The results are displayed in Table 3. The table provides the MC estimator of the average optimal rise for the reconstruction error of the signal f, defined in (21), computed over $M = 1{,}000$ simulated samples, i.e., $\text{ARISE} = \frac{1}{M}\sum_{m=1}^{M}\left\|f_{\text{REG},n,m} - f\right\|$, where $f_{\text{REG},n,m}$ is the reconstruction obtained with the sample m of size n computed with the optimal regularization parameter obtained by minimizing the reconstruction rise. To appreciate if the differences are significant, we also provide the MC standard deviation of this estimator, i.e., $\text{Std}_{\text{ARISE}} = \sqrt{\frac{1}{M(M-1)}\sum_{m=1}^{M}\left(\left\|f_{\text{REG},n,m} - f\right\| - \text{ARISE}\right)^2}$, which corresponds to the standard-deviation formulas of the average estimator arise (which requires to divide by $\sqrt{M}$). The table also gives for each case the average of the M optimal values of the regularization parameters. At last, in addition to the results obtained for the optimal values of the regularization parameters, we also compute values using the Lepskiĭ method for the mollification in order to check that the results obtained were close to the optimal setting. More specifically, we fix $q = 0.99$, $J_n = 500$,

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

Table 3. MC performances of the four methods over $M = 1{,}000$ replications of samples of size $n = 100, 500,$ and $1{,}000$, respectively

Method	MOL(Lepskiĭ)	MOL	TIK	KER	CUT
	$n = 100$				
ARISE	0.195860	0.100038	0.209369	0.106937	0.105383
$\text{Std}_{\text{ARISE}}$	0.001213	0.000941	0.000898	0.000952	0.000813
⟨RegPar⟩	0.810334	0.358822	0.238026	0.460649	0.373794
	$n = 500$				
ARISE	0.114902	0.064398	0.134029	0.070904	0.068135
$\text{Std}_{\text{ARISE}}$	0.000637	0.000470	0.000543	0.000481	0.000450
⟨RegPar⟩	0.524592	0.257730	0.124067	0.344160	0.272226
	$n = 1{,}000$				
ARISE	0.091830	0.052165	0.108078	0.058255	0.055289
$\text{Std}_{\text{ARISE}}$	0.000417	0.000355	0.000426	0.000366	0.000359
⟨RegPar⟩	0.444509	0.221208	0.094325	0.303763	0.225243

Notes: The quantity arise is the MC average of the reached optimal reconstruction rise, $\text{Std}_{\text{ARISE}}$ is its MC standard deviation, and ⟨RegPar⟩ is the mean of the respective optimal values of the regularization parameters, except for the first column where the regularization parameter is chosen using the Lepskiĭ method.

$\beta_0 = 10$, and $\theta = 0.75$. Although our analysis, which is based on a deterministic argument, only covers the case $\theta \geq 2$, it is well known (see Bauer and Lukas, 2011) that, for stochastic errors, better results can be achieved for smaller values of θ.

From the table, we see that for each method, as expected, the RISE decreases when the sample size increases. The mollification approach provides better results than the other regularization techniques, and the difference is significant (compared with the respective standard errors). The cutoff method seems to be better than the deconvolution kernel (when available) and than the Tikhonov, which appears to be the less reliable method. Of course, these general comments apply only for the chosen scenario, but added with the theoretical comparisons made above, this MC exercise seems to advocate for using mollification techniques in deconvolution problems.

7. CONCLUSION

We have introduced the mollification approach to the deconvolution of probability densities. We have established the consistency of the corresponding estimator. By placing mollification in the framework of *filter-type* methods, we have compared it, both theoretically and numerically, with various other methods, namely the deconvolution kernels, the spectral cutoff, and the Tikhonov regularization. This comparison reveals notably that the mollification enables to substantially extend the domain of applicability of the deconvolution kernels, while providing better performances than all methods under consideration in terms of the tradeoff between fidelity and stability of the reconstruction. mollification inherits advantages of both the deconvolution kernels (in particular, a *target object* is clearly defined) and the Tikhonov regularization (in particular, the flexibility brought by the variational formulation).

We have studied the convergence rates of mollification for the problem of deconvolution of probability densities. It has been shown that optimal rates are reached by mollification both in the power decay and exponential decay settings under general Besov regularity assumptions. There is a saturation effect in the case of power decay, but this saturation bound can be made as large as desired by suitable choice of the mollifier. This is a definite advantage compared to Tikhonov regularization and explains our computational result. There is no saturation effect in the case of exponential decay. Hence, both our theoretical and our computational results suggest that mollification is a very promising regularization method for the deconvolution of probability densities.

APPENDIX A. Proofs of the Results in Sections 3.3, 3.4, and 4.2

We start this section by a result on the operator norm of the multiplication operator, which will be useful later in the proofs.

PROPOSITION A.1. *Let φ be in $L^\infty(\mathbb{R})$, and let $M\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$ be the multiplication operator defined by $Mf = \varphi \cdot f$. Then, $\|M\| = \|\varphi\|_\infty$, in which $\|M\|$ denotes the operator norm of M.*

Proof. We can assume without loss of generality that $\|\varphi\|_\infty > 0$. Let λ denote the Lebesgue measure on $\mathbb{R}$. Clearly, $\int |\varphi f|^2 \,\mathrm{d}\lambda \le \|\varphi\|_\infty^2 \int |f|^2 \,\mathrm{d}\lambda = \|\varphi\|_\infty^2 \|f\|^2$, so that $\|M\| \le \|\varphi\|_\infty$. In order to obtain the opposite inequality, let $\varepsilon > 0$ be fixed. The Lebesgue measure being σ-finite, one can find $A \subset \mathbb{R}$ measurable such that:

(i) $0 < \lambda(A) < \lambda(\mathbb{R}) = \infty$;
(ii) for every $x \in A$, $\varphi(x) \ge \|\varphi\|_\infty - \varepsilon$.

As a matter of fact, let $A_\circ := \left\{x \in \mathbb{R} \,\middle|\, |\varphi(x)| \ge \|\varphi\|_\infty - \varepsilon\right\}$. Then $\lambda(A_\circ) > 0$ (for otherwise one would have $\|\varphi\|_\infty \le \|\varphi\|_\infty - \varepsilon$). If $\lambda(A_\circ) < \infty$, then just take $A = A_\circ$. Otherwise, consider an increasing sequence (B_n) such that $\lambda(B_n) > 0$ for all n and $\cup_n B_n = \mathbb{R}$. The sequence $(A_\circ \cap B_n)$ is increasing, with limit $A_\circ$. For $n_\circ$ sufficiently large, $\lambda(A_\circ \cap B_{n_\circ}) > 0$, since $\lambda(A_\circ \cap B_n) \to \lambda(A_\circ) > 0$. Thus, take $A = A_\circ \cap B_{n_\circ}$ in this case.

Now, let $f = \mathbb{1}_A/\sqrt{\lambda(A)}$. Then $f \in L^2(\mathbb{R})$ and $\|f\|_2 = 1$, so that

$$\|M\|^2 \ge \|Mf\|^2 = \int |\varphi f|^2 \,\mathrm{d}\lambda = \frac{1}{\lambda(A)} \int_A |\varphi|^2 \,\mathrm{d}\lambda \ge \left(\|\varphi\|_\infty - \varepsilon\right)^2.$$

Since ε can be chosen arbitrarily small, we have shown that $\|M\| \ge \|\varphi\|_\infty$. □

The following lemma shows that bounding the regularization error over Sobolev balls coincides with bounding the supremum norm of certain functions.

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

LEMMA A.2. *For any $u > 0$, we have*

$$\sup_{\|f\|_{H^u} \le 1} \|f - f_{\mathrm{MM},\beta}\| = \|\tau_{\beta,u}\|_\infty \quad \text{with } \tau_{\beta,u}(\xi) := \frac{(1-\widehat{\varphi}_\beta(\xi))^2(1+|\xi|)^{-u}}{|\widehat{\gamma}(\xi)|^2 + (1-\widehat{\varphi}_\beta(\xi))^2}.$$

Proof. Note that

$$Uf - Uf_{\mathrm{MM},\beta} = \left(1 - \frac{\hat{\gamma}\overline{\hat{\gamma}}}{|\widehat{\gamma}(\xi)|^2 + (1-\widehat{\varphi}_\beta(\xi))^2}\right) Uf = \frac{(1-\widehat{\varphi}_\beta(\xi))^2}{|\widehat{\gamma}(\xi)|^2 + (1-\widehat{\varphi}_\beta(\xi))^2} Uf.$$

Furthermore, recall that $\|f\|_{H^u} = \|w\|$ with $w(\xi) := (1+|\xi|)^u Uf(\xi)$. This together with the unitarity of the Fourier transform implies $\sup_{\|f\|_{H^u}\le 1} \|f - f_{\mathrm{MM},\beta}\| = \sup_{\|w\|\le 1} \|\tau_{\beta,u} \cdot w\| = \|\tau_{\beta,u}\|_\infty$ since that operator norm of a multiplication operator in L^2 is given by the supremum norm of the multiplier function (see Proposition A.1). □

LEMMA A.3 (Asymptotic properties of intersection points). *Let $\Gamma, \Phi \colon \mathbb{R}_+ \to \mathbb{R}_+^*$ be strictly decreasing functions that go to zero at infinity, and let $C > 0$. Then, for every $\beta > 0$, the equation*

$$C\Gamma(t) = \Phi(0) - \Phi(\beta t) \tag{A.1}$$

has a unique solution t_β^* which satisfies $t_\beta^* \to \infty \quad \text{and} \quad \beta t_\beta^* \to 0 \quad \text{as} \quad \beta \downarrow 0.$

Proof. Without loss of generality, assume that $\Gamma(0) = 1$, so that Γ is a decreasing bijection from $\mathbb{R}_+$ to $(0,1]$ and its inverse Γ^{-1} is a decreasing bijection from $(0,1]$ to $\mathbb{R}_+$. Similarly, $\Phi(0) - \Phi$ is an increasing bijection from $\mathbb{R}_+$ onto $[0, \Phi(0))$, and its inverse $(\Phi(0) - \Phi)^{-1}$ is an increasing bijection from $[0, \Phi(0))$ onto $\mathbb{R}_+$. It is then easy to see that

$$\begin{aligned} \psi\colon \quad (a,\infty) &\longrightarrow \mathbb{R}_+^* \\ t &\longmapsto \psi(t) := \big(\Phi(0) - \Phi\big)^{-1}\big(C\Gamma(t)\big), \end{aligned}$$

with $a := \Gamma^{-1}\big(C^{-1}\Phi(0)\big)$ is a decreasing bijection. Then the solution to equation (A.1) coincides with the solution to the equation $\beta t = \psi(t)$. Now, existence and uniqueness of t_β^* follows from strict monotonicity of $t \mapsto \beta t - \psi(t)$ and the intermediate value theorem. Furthermore, the desired limits follow immediately from the above fixed point characterization of t_β^*. □

LEMMA A.4 (Approximation error of mollifier). *If $f \in B^u_{2,\infty}(\mathbb{R}^d)$ and Assumption 8 holds true, there exists a constant M independent of f such that $\|f - C_\beta f\| \le M \|f\|_{B^u_{2,\infty}} \beta^{\min(s,u)}$ for all β sufficiently small.*

Proof. We have

$$\begin{aligned} \|f - C_\beta f\|^2 &= \int_{\mathbb{R}^d} (1 - \Phi(\beta|\xi|))^2 \big|\widehat{f}(\xi)\big|^2 \,\mathrm{d}\xi \;=\; \int_0^\infty (1-\Phi(\beta t))^2 (-\mathrm{d}e_f(t)) \\ &= \int_0^\infty e_f(t) d(1-\Phi(\beta t))^2 \,\mathrm{d}t. \end{aligned}$$

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

In order to estimate the integral, we treat the subintervals $\left(0,\beta^{-1}\right)$ and $\left(\beta^{-1},\infty\right)$ separately. For the first interval, we use Assumption 8 to obtain

$$\begin{aligned}
\int_0^{\frac{1}{\beta}} e_f(t) d\left(1-\Phi(\beta t)\right)^2 \,\mathrm{d}t &= 2\beta \int_0^{\frac{1}{\beta}} e_f(t)\left(1-\Phi(\beta t)\right)\left(-\Phi'(\beta t)\right)\mathrm{d}t \\
&\le 2C_\Phi^2 \|f\|^2_{B^u_{2,\infty}} \int_0^{\frac{1}{\beta}} (1+t)^{-2u}\beta(\beta t)^{s+(s-1)}\,\mathrm{d}t \\
&\le 2C_\Phi^2 \|f\|^2_{B^u_{2,\infty}} \beta^{2s}\left(\int_0^1 t^{2s-1}\,\mathrm{d}t + \int_1^{\frac{1}{\beta}} (1+t)^{-2u} t^{2s-1}\,\mathrm{d}t\right) \\
&= \|f\|^2_{B^u_{2,\infty}} O\left(\beta^{2s}+\beta^{2u}\right).
\end{aligned}$$

On the second interval, we estimate

$$\begin{aligned}
\int_{\frac{1}{\beta}}^\infty e_f(t) d\left(1-\Phi(\beta t)\right)^2 \,\mathrm{d}t &\le e_f\left(\frac{1}{\beta}\right)\left|\left(1-\Phi(\beta\cdot)\right)^2\right|_{\mathrm{TV}} \\
&\le \|f\|^2_{B^u_{2,\infty}}\left|\left(1-\Phi(\beta\cdot)\right)^2\right|_{\mathrm{TV}}\beta^{2u}.
\end{aligned}$$

□

Under Assumptions 8 and condition (9), we have

$$\frac{1}{C^2}\Theta_{\beta,u}(|\xi|) \le \tau_{\beta,u}(\xi) \le C^2\Theta_{\beta,u}(|\xi|) \quad \text{with} \quad \Theta_{\beta,u}(t) := \frac{\left(1-\Phi(\beta t)\right)^2(1+t)^{-u}}{(1+t)^{-2a}+\left(1-\Phi(\beta t)\right)^2},$$

for all $\xi \in \mathbb{R}^d$. For the proof of Theorem 11, we need the following bound on these functions, which is equivalent to convergence rate in Sobolev balls:

Proposition A.5 (Bias bounds in Sobolev balls). *Suppose that Assumption 8 and condition (9) hold true, and let $\tau_{\beta,u}$ be defined as in Lemma A.2. Then*

$$\|\tau_{\beta,u}\|_\infty \sim \beta^{\min\left(\frac{su}{s+a},2s\right)} \quad \textit{as } \beta\downarrow 0.$$

In particular, $\sup_{\|f\|_{H^u}\le 1}\|f-f_{\mathrm{MM},\beta}\| \sim \beta^{\min\left(\frac{su}{s+a},2s\right)}$.

Proof. By Lemma A.3, the equation

$$(1+t)^{-a} = 1-\Phi(\beta t) \tag{A.2}$$

has a unique solution $\bar t_\beta$ which satisfies $\bar t_\beta \to \infty$ and $\beta\bar t_\beta \to 0$ as $\beta\downarrow 0$. Therefore, $(1+\bar t_\beta)^{-a} \sim (\bar t_\beta)^{-a}$ and $1-\Phi(\beta\bar t_\beta) \sim (\beta\bar t_\beta)^s$ as $\beta\downarrow 0$. We deduce that $(\bar t_\beta)^{-a} \sim (\beta\bar t_\beta)^s$, or

$$\bar t_\beta \sim \beta^{-\frac{s}{s+a}} \quad \text{as} \quad \beta\downarrow 0. \tag{A.3}$$

Omitting the smaller of the two terms in the denominator of $\Theta_{\beta,u}(t)$ yields the upper bound

$$\frac{\overline{\Theta}_{\beta,u}(t)}{2} \le \Theta_{\beta,u}(t) \le \overline{\Theta}_{\beta,u}(t) \quad \text{with } \overline{\Theta}_{\beta,u}(t) := \begin{cases} \left(1-\Phi(\beta t)\right)^2(1+t)^{2a-u}, & 0\le t\le \bar t_\beta, \\ (1+t)^{-u}, & t\ge \bar t_\beta. \end{cases}$$

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

Caseu $\in [0,2a]$: In this case, $\overline{\Theta}_{\beta,u}$ is increasing for $t \le \bar{t}_\beta$ and decreasing for $t \ge \bar{t}_\beta$, and therefore

$$\|\tau_{\beta,u}\|_\infty \sim \|\Theta_{\beta,u}\|_\infty \sim \|\overline{\Theta}_{\beta,u}\|_\infty = \overline{\Theta}_{\beta,u}(\bar{t}_\beta) = (1+\bar{t}_\beta)^{-u} \sim \beta^{\frac{su}{s+a}}.$$

Caseu $> 2a$: In this case, we need further work to bound $\overline{\Theta}_{\beta,u}$ on the interval $[0,\bar{t}_\beta]$. Using assumption (8b) and the inequalities $\frac{1}{2\max(1,t)} \le \frac{1}{1+t} \le \frac{1}{\max(1,t)}$, we obtain

$$C_\Phi^{-2} 2^{2a-u} \overline{\overline{\Theta}}_{\beta,u}(t) \le \overline{\Theta}_{\beta,u}(t) \le C_\Phi^2 \overline{\overline{\Theta}}_{\beta,u}(t) \quad \text{for } 0 \le t \le \bar{t}_\beta$$

with

$$\overline{\overline{\Theta}}_{\beta,u}(t) := \begin{cases} (\beta t)^{2s}, & 0 \le t \le \min(1,\bar{t}_\beta), \\ \beta^{2s} t^{2s+2a-u}, & 1 < t \le \bar{t}_\beta. \end{cases}$$

Since $\overline{\Theta}$ is continuous, positive, and decreasing on $(\bar{t}_\beta,\infty)$, we have

$$\begin{aligned} \|\overline{\Theta}_{\beta,u}\|_{L^\infty([0,\infty))} = \|\overline{\Theta}_{\beta,u}\|_{L^\infty([0,\bar{t}_\beta])} &\sim \|\overline{\overline{\Theta}}_{\beta,u}\|_{L^\infty([0,\bar{t}_\beta])} \\ &= \begin{cases} \overline{\overline{\Theta}}_{\beta,u}(\bar{t}_\beta) \sim \beta^{\frac{su}{s+a}}, & u \le 2s+2a, \\ \overline{\overline{\Theta}}_{\beta,u}(1) \sim \beta^{2s}, & u > 2s+2a. \end{cases} \end{aligned}$$

This completes the proof. □

Proof of Theorem 11. The last statement on the limit case $u = 2a+2s$ follows from Proposition A.5.

(10a)⇒(10b): Note that $Uf - Uf_{\mathrm{MM},\beta} = \frac{(1-\hat{\varphi}_\beta)^2}{|\hat{\gamma}|^2+(1-\hat{\varphi}_\beta)^2} Uf$. Therefore, estimating the square of the fraction by

$$r_\beta(t) := \left(\frac{(1-\Phi(\beta t))^2}{(1+t)^{-2a} + (1-\Phi(\beta t))^2} \right)^2 \tag{A.4}$$

with Φ from Assumption 8, we obtain with C from condition (9) that

$$\begin{aligned} C^{-2} \left\| f - f_{\mathrm{MM},\beta} \right\|^2 &\le \int_{ml^d} r_\beta(|\xi|) |\widehat{f}(\xi)|^2 \, \mathrm{d}\xi \\ &= \int_0^\infty r_\beta(t) \, \mathrm{d}(-e_f)(t) \;=\; \int_0^\infty e_f(t) \, \mathrm{d}r_\beta(t). \end{aligned} \tag{A.5}$$

In the third line, we have used partial integration and the fact that $r_\beta(0) = 0$ and $\lim_{t\to\infty} e_f(t) = 0$. For the numbers $\bar{t}_\beta$ defined by (A.2), we have $\bar{t}_\beta \sim \beta^{-\frac{s}{s+a}}$ (see (A.3)) and

$$r_\beta(\bar{t}_\beta) = \frac{1}{4}, \tag{A.6}$$

for all $\beta \in (0,1]$. We split the integral on the right-hand side of (A.5) into two parts:

$$C^{-2} \left\| f - f_{\mathrm{MM},\beta} \right\|^2 \le \int_0^{\bar{t}_\beta} e_f(t) \, \mathrm{d}r_\beta(t) + \int_{\bar{t}_\beta}^\infty e_f(t) \, \mathrm{d}r_\beta(t).$$

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

As e_f is decreasing and $\lim_{t\to\infty} r_\beta(t) = 1$, the second integral is bounded by

$$\int_{\bar{t}_\beta}^{\infty} e_f(t)\,\mathrm{d}r_\beta(t) \le e_f(\bar{t}_\beta)(1 - r_\beta(\bar{t}_\beta)) \le \frac{3}{4}\|f\|^2_{B^u_{2,\infty}}(1+\bar{t}_\beta)^{-2u} \le C_1^2\|f\|^2_{B^u_{2,\infty}}\beta^{\frac{2su}{s+a}}$$

for some C_1 independent of f and β sufficiently small. For the first integral, we introduce some $\mu \in \left(\frac{u}{2s+2a}, 1\right)$ and estimate

$$\begin{aligned}
\int_0^{\bar{t}_\beta} e_f(t)\,\mathrm{d}r_\beta(t) &\le \|f\|^2_{B^u_{2,\infty}} \int_0^{\bar{t}_\beta} (1+t)^{-2u}\,\mathrm{d}r_\beta(t)\\
&= \|f\|^2_{B^u_{2,\infty}} \int_0^{\bar{t}_\beta} (1+t)^{-2u} r_\beta(t)^\mu \frac{1}{r_\beta(t)^\mu}\,\mathrm{d}r_\beta(t)\\
&\le \|f\|^2_{B^u_{2,\infty}} \sup_{t>0}\left[(1+t)^{-2u/\mu} r_\beta(t)\right]^\mu \int_0^{\bar{t}_\beta} \frac{1}{r_\beta(t)^\mu}\,\mathrm{d}r_\beta(t)\\
&= \|f\|^2_{B^u_{2,\infty}} \|\Theta^2_{\beta,u/\mu}\|^\mu_\infty \frac{r_\beta(\bar{t}_\beta)^{1-\mu}}{1-\mu} \le C_2^{2\mu}\|f\|^2_{B^u_{2,\infty}}\beta^{\frac{2su}{s+a}}\frac{4^{\mu-1}}{1-\mu},
\end{aligned}$$

where $C_2 := \sup_{0<\beta\le 1}\beta^{-\frac{s}{s+a}\frac{u}{\mu}}\|\Theta_{\beta,u/\mu}\|_\infty$ is finite for $u < 2a+2s$ by Proposition A.5. Putting these estimates together shows that

$$\|f - f_{\mathrm{MM},\beta}\| \le C\left(C_1 + C_2^\mu \frac{2^{\mu-1}}{\sqrt{1-\mu}}\right)\|f\|_{B^u_{2,\infty}}\beta^{\frac{su}{s+a}}.$$

(10b)⇒(10a): It follows from (A.5) and (A.6) that

$$C^2\|f - f_{\mathrm{MM},\beta}\|^2 \ge \int_{\bar{t}_\beta}^{\infty} r_\beta(t)\,\mathrm{d}(-e_f)(t) \ge r_\beta(\bar{t}_\beta)\int_{\bar{t}_\beta}^{\infty}\mathrm{d}(-e_f)(t) = \frac{1}{4}e_f(\bar{t}_\beta). \tag{A.7}$$

If M denotes the value of the supremum in (10b), this shows that $e_f(\bar{t}_\beta) \le 4(CM)^2\beta^{\frac{2su}{s+a}} \le C_3\bar{t}_\beta^{-2u}$ for some $C_3 > 0$. As $\lim_{\beta\to 0}\bar{t}_\beta = \infty$ and $e_f(0) = \|f\|^2 < \infty$, this shows that $\|f\|_{B^u_{2,\infty}} < \infty$. □

Proof of Theorem 12. The last statement on the limit case $u = a+s$ follows from the triangle inequality

$$\begin{aligned}
\|f - f_{\mathrm{MO},\beta}\| &\le \|f - C_\beta f\| + \|C_\beta f - f_{\mathrm{MO},\beta}\| \\
&\le \|f - C_\beta f\| + \|C_\beta\|\,\|f - f_{\mathrm{MM},\beta}\|,
\end{aligned} \tag{A.8}$$

along with Lemma A.4 and Proposition A.5.

(11a)⇒(11b): We use again the triangle inequality (A.8). The second term on the right-hand side is of order $O\left(\beta^{\frac{su}{s+a}}\right)$ by Theorem 11. By Lemma A.4, the first term is of order $O\left(\beta^{\min(u,s)}\right)$, which is of higher order $o\left(\beta^{\frac{su}{s+a}}\right)$ as long as $u < s+a$.

(11b)⇒(11a): We have

$$\begin{aligned}
\|f_{\mathrm{MO},\beta} - f\| &\ge \|f_{\mathrm{MO},\beta} - C_\beta f\| - \|C_\beta f - f\|\\
&= \|C_\beta f_{\mathrm{MM},\beta} - C_\beta f\| - \|C_\beta f - f\|.
\end{aligned}$$

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

As in (A.7), we can show that

$$C^2\left\|C_\beta f_{\mathrm{MM},\beta}-C_\beta f\right\|^2 \geq r_\beta(\bar{t}_\beta)^2\Phi(\beta\bar{t}_\beta)^2 e_f(\bar{t}_\beta)=\frac{1}{4}\Phi(\beta\bar{t}_\beta)^2 e_f(\bar{t}_\beta).$$

As $\lim_{\beta\to 0}\beta\bar{t}_\beta=0$, this together with (11b) and Lemma A.4 yields

$$\begin{aligned} e_f(\bar{t}_\beta) &\leq \frac{4C^2}{\Phi(\beta\bar{t}_\beta)^2}\left\|C_\beta f_{\mathrm{MM},\beta}-C_\beta f\right\|^2 \\ &\leq \frac{8C^2}{\Phi(\beta\bar{t}_\beta)^2}\left(\left\|f_{\mathrm{MO},\beta}-f\right\|^2+\left\|C_\beta f-f\right\|^2\right) \\ &= O\left(\beta^{\frac{2su}{s+a}}+\beta^{2\min(u,s)}\right)=O\left(\beta^{\frac{2su}{s+a}}\right)=O\left(\bar{t}_\beta^{2u}\right), \end{aligned}$$

and we conclude that $f\in B^u_{2,\infty}(\mathbb{R}^d)$. □

Lemma A.6. *Under the assumptions of Proposition 15, there exists a constant C_{MO} for all $\beta_0>0$ such that $\|\Phi_\beta^{\mathrm{MO}}\|\leq C_{\mathrm{MO}}\beta^{-\frac{s(a+d/2)}{s+a}}$ for all $\beta\in(0,\beta_0]$.*

Proof. Due to Assumption 8 and condition (9), we have

$$\|\Phi_\beta^{\mathrm{MO}}\|^2 \leq \int_0^\infty \frac{C^2(1+t)^{-2a}\Phi(\beta t)^2}{\left|C^{-2}(1+t)^{-2a}+(1-\Phi(\beta t))^2\right|^2}t^{d-1}\,\mathrm{d}t.$$

We use the numbers $\bar{t}_\beta$ defined in (A.2) and estimate the integrand on the three subintervals $(0,\bar{t}_\beta)$, $(\bar{t}_\beta,\frac{1}{\beta})$, and $(\frac{1}{\beta},\infty)$ separately. Using the asymptotic behavior (A.3) of $\bar{t}_\beta$ as $\beta\downarrow 0$ and Assumption 8, we obtain

$$\begin{aligned} \int_0^{\bar{t}_\beta} \frac{C^2(1+t)^{-2a}\Phi(\beta t)^2}{\left|C^{-2}(1+t)^{-2a}+(1-\Phi(\beta t))^2\right|^2}t^{d-1}\,\mathrm{d}t &\leq C^6\int_0^{\bar{t}_\beta}(1+t)^{2a}t^{d-1}\,\mathrm{d}t \\ &= O\left(\bar{t}_\beta^{d+2a}\right)=O\left(\beta^{-\frac{s(d+2a)}{s+a}}\right), \end{aligned}$$

$$\begin{aligned} \int_{\bar{t}_\beta}^{\frac{1}{\beta}} \frac{C^2(1+t)^{-2a}\Phi(\beta t)^2}{\left|C^{-2}(1+t)^{-2a}+(1-\Phi(\beta t))^2\right|^2}t^{d-1}\,\mathrm{d}t &\leq C^2C_\Phi^4\int_{\bar{t}_\beta}^\infty \frac{(1+t)^{-2a}t^{d-1}}{(\beta t)^{4s}}\,\mathrm{d}t \\ &= O\left(\beta^{-\frac{s(d+2a)}{s+a}}\right), \end{aligned}$$

$$\begin{aligned} \int_{\frac{1}{\beta}}^\infty \frac{C^2(1+t)^{-2a}\Phi(\beta t)^2}{\left|C^{-2}(1+t)^{-2a}+(1-\Phi(\beta t))^2\right|^2}t^{d-1}\,\mathrm{d}t &\leq 4C^2\int_{\frac{1}{\beta}}^\infty(1+t)^{-2a}\Phi(\beta t)^2t^{d-1}\,\mathrm{d}t \\ &\leq 4C^2\beta^{2a-d}\int_1^\infty \tilde{t}^{-2a}\Phi(\tilde{t})^2\tilde{t}^{d-1}\,\mathrm{d}\tilde{t} \\ &= O\left(\beta^{2a-d}\right). \end{aligned}$$

Since $-\frac{s(d-2a)}{s+a}\leq 2a-d$ if and only if $4s\geq d-2a$, this yields the desired result. □

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

Proof of Theorem 17. (14a)⇒(14b): We proceed as in the proof of Theorem 11, replacing the definition (A.4) of r_β by $r_\beta(t) := \left(\frac{(1-\Phi(\beta t))^2}{\exp(-\kappa t^a)+(1-\Phi(\beta t))^2}\right)^2$. Moreover, we define $\bar{t}_\beta := \left(-\frac{1}{\kappa}\ln\beta\right)^{1/a}$ such that $\exp(-\kappa\bar{t}_\beta^a) = \beta$. It is clear that $\bar{t}_\beta \to \infty$ and $\beta\bar{t}_\beta \to 0$ as $\beta \downarrow 0$. As in the proof of Theorem 11, we estimate

$$C^{-2}\left\|f - f_{\mathrm{MM},\beta}\right\|^2 \le \int_0^{\bar{t}_\beta} e_f(t)\,\mathrm{d}r_\beta(t) + \int_{\bar{t}_\beta}^\infty e_f(t)\,\mathrm{d}r_\beta(t).$$

As e_f is decreasing and $\lim_{t\to\infty} r_\beta(t) = 1$, the second integral is bounded by

$$\begin{aligned}\int_{\bar{t}_\beta}^\infty e_f(t)\,\mathrm{d}r_\beta(t) &\le e_f(\bar{t}_\beta)(1 - r_\beta(\bar{t}_\beta)) \le \|f\|^2_{B^u_{2,\infty}}(1+\bar{t}_\beta)^{-2u}\\ &\le C_1\|f\|^2_{B^u_{2,\infty}}(-\ln\beta)^{-2u/a}\end{aligned}$$

for some C_1 independent of f and $\beta \in (0,1)$. As in the proof of Theorem 11, we choose some $\mu \in (0,1)$ and bound the first integral by

$$\begin{aligned}\int_0^{\bar{t}_\beta} e_f(t)\,\mathrm{d}r_\beta(t) &\le \|f\|^2_{B^u_{2,\infty}}\frac{r_\beta(\bar{t}_\beta)^{1-\mu}}{1-\mu}\sup_{0<t\le\bar{t}_\beta}\left[(1+t)^{-2u/\mu}r_\beta(t)\right]^\mu,\\ &\le \|f\|^2_{B^u_{2,\infty}}\frac{r_\beta(\bar{t}_\beta)^{1-\mu}}{1-\mu}\left((-\ln\beta)^{-2u/a}\right)C_2^{2\mu},\end{aligned}$$

where $C_2 := \sup_{0\le\beta\le1}\left((-\ln\beta)^{\frac{u}{a\mu}}\sup_{0<t\le\bar{t}_\beta}\left[(1+t)^{-u/\mu}r_\beta(t)^{1/2}\right]\right)$ is finite for all $u>0$. Indeed, we have, for $0 < t \le \bar{t}_\beta$,

$$r_\beta(t)^{1/2} \le \frac{(1-\Phi(\beta t))^2}{\exp(-\kappa t^a)} \le C_\Phi^2(\beta\bar{t}_\beta)^{2s}\exp(\kappa\bar{t}_\beta^a) \le C_\Phi^2\left(\frac{1}{\kappa}\right)^{2s/a}\beta^{2s-1}(-\ln\beta)^{2s/a}$$

such that $C_2 \le C_\Phi^2\left(\frac{1}{\kappa}\right)^{2s/a}\sup_{0\le\beta\le1}\beta^{2s-1}(-\ln\beta)^{\frac{2s}{a}+\frac{u}{a\mu}}$. As the last supremum is finite for $s > 1/2$, we have shown that

$$\int_0^{\bar{t}_\beta} e_f(t)\,\mathrm{d}r_\beta(t) = O\left((-\ln\beta)^{2u/a}\right).$$

Putting everything together shows that $\left\|f - f_{\mathrm{MM},\beta}\right\| = O\left((-\ln\beta)^{-u/a}\right)$ as $\beta \downarrow 0$.
(14b)⇒(14c): This follows again from (A.8).

(14c)⇒(14a): As in the proof of Theorem 12, we obtain

$$e_f(\bar{t}_\beta) \le \frac{8}{\Phi(\beta\bar{t}_\beta)^2}\left(\left\|f_{\mathrm{MO},\beta} - f\right\|^2 + \left\|C_\beta f - f\right\|^2\right).$$

Using $\lim_{\beta\to0}\beta\bar{t}_\beta = 0$, together with (14c) and Lemma A.4, this yields

$$e_f(\bar{t}_\beta) = O\left((-\ln\beta)^{-2u/a} + \beta^{\min(u,s)}\right) = O\left((-\ln\beta)^{-2u/a}\right) = O\left(\bar{t}_\beta^{-2u}\right).$$

This shows that $f \in B^u_{2,\infty}(\mathbb{R}^d)$. □

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

Proof of Proposition 20. Our proof is based on the following general large deviation inequality due to Massart based on the seminal work by Talagrand.

Theorem [Massart, 2000, Thm. 3] Let $\mathcal{F} \subset L^\infty(\mathbb{R}^d)$ be a countable family of functions with $\|\varphi\|_\infty \le b$ for all $\varphi \in \mathcal{F}$. Moreover, let

$$Z := n \sup_{\varphi \in \mathcal{F}} \left| \int_{\mathbb{R}^d} \varphi(\mathrm{d}g_n - g\mathrm{d}x) \right| \tag{A.9}$$

and $v := n \sup_{\varphi \in \mathcal{F}} \int_{\mathbb{R}^d} \varphi^2 g \,\mathrm{d}x$. Then

$$P\left[Z \ge (1+\epsilon)EZ + \sqrt{8v\xi} + \kappa(\epsilon) b\xi\right] \le \exp(-\xi), \tag{A.10}$$

for all $\epsilon, \xi > 0$, where $\kappa(\epsilon) = 2.5 + 32/\epsilon$.

To apply this to $Z := n\beta^{\frac{s(a+d/2)}{s+a}} \|f_{\beta,n} - f_{\mathrm{MO},\beta}\|$, we write the estimator in the form $\hat{f}_\beta = \check{\Phi}_\beta^{\mathrm{MO}} * g_n$ with g_n defined in (6) and $\check{\Phi}_\beta^{\mathrm{MO}}$ the inverse Fourier transforms of the functions Φ_β^{MO} in Proposition 15. Since the unit spheres of separable Hilbert spaces are separable, there exists a sequence $(h_m)_{m\in\mathbb{N}}$ which is dense in the unit sphere of $L^2(\mathbb{R}^d)$, and the variance term can be written in the form (A.9) as follows:

$$\begin{aligned}
&\|f_{\beta,n} - f_{\mathrm{MO},\beta}\| \\
&\quad = \sup_{m\in\mathbb{N}} |\langle h_m, f_{\beta,n} - f_{\mathrm{MO},\beta}\rangle| = \sup_{m\in\mathbb{N}} \left|\langle h_m, \check{\Phi}_\beta^{\mathrm{MO}} * (g_n - g)\rangle\right| \\
&\quad = \sup_{m\in\mathbb{N}} \left|\langle \overline{\check{\Phi}_\beta^{\mathrm{MO}}(-\cdot)} * h_m, (g_n - g)\rangle\right| = \beta^{-\frac{s(a+d/2)}{s+a}} \sup_{\varphi\in\mathcal{F}_\beta} \left|\int_{\mathbb{R}^d} \varphi(\mathrm{d}g_n - g\mathrm{d}x)\right|.
\end{aligned}$$

Here, $\mathcal{F}_\beta := \{\varphi_{m,\beta} : m \in \mathbb{N}\}$ and $\varphi_{m,\beta} := \beta^{\frac{s(a+d/2)}{s+a}} \overline{\check{\Phi}_\beta^{\mathrm{MO}}(-\cdot)} * h_m$, and we have

$$|\varphi_{m,\beta}(x)| \le \beta^{\frac{s(a+d/2)}{s+a}} \int_{\mathbb{R}^d} |\overline{\check{\Phi}_\beta^{\mathrm{MO}}(-x+y)} h_m(y)|\,\mathrm{d}y \le \beta^{\frac{s(a+d/2)}{s+a}} \|\Phi_\beta^{\mathrm{MO}}\|_{L^2} \|h_m\|_{L^2} \le C_{\mathrm{MO}}$$

by Lemma A.6. Let M_β be a separable dense subset of $(0, \beta_0]$ and set $\mathcal{F} := \bigcup_{\beta\in M_\beta} \mathcal{F}_\beta$. Then $b := \sup\{\|\varphi\|_\infty : \varphi \in \mathcal{F}\} \le C_{\mathrm{MO}}$, and by Proposition 15 we have

$$\mathrm{E}\left(n\|f_{\beta,n} - f_{\mathrm{MO},\beta}\|\right) \le n\left(E\|f_{\beta,n} - f_{\mathrm{MO},\beta}\|^2\right)^{1/2} \le C_{\mathrm{MO}}\sqrt{2n}\beta^{-\frac{s(a+d/2)}{a+s}}.$$

Finally, we use that the operator $R_\beta : g \mapsto \check{\Phi}_\beta^{\mathrm{MO}} * g$ is a Hilbert–Schmidt from the weighted space $L^2(g\mathrm{d}x)$ to $L^2(\mathbb{R}^d)$ where the norm of the former space is given by $\|h\|_{g\mathrm{d}x)} = (\int |h|^2 g\mathrm{d}x))^{1/2}$. Since the operator norm of an integral operator is bounded by the L^2-norm of its kernel, we obtain

$$\|R_\beta\|^2_{L^2(g\mathrm{d}x)\to L^2} \le \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |\check{\Phi}_\beta^{\mathrm{MO}}(x-y)|^2 \,\mathrm{d}x\, g(y)\,\mathrm{d}y = \|\check{\Phi}_\beta^{\mathrm{MO}}\|^2 \le C_{\mathrm{MO}}^2 \beta^{-\frac{s(2a+d)}{a+s}}$$

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

using Lemma A.6 again. Therefore,

$$v = n \sup_{v\in\mathcal{F}} \int_{\mathbb{R}^d} \varphi^2 g\, \mathrm{d}x \le n\beta^{\frac{s(d+2a)}{s+a}} \sup_m \|R_\beta^* h_m\|^2_{L^2(g\mathrm{d}x)} \le nC^2_{\mathrm{MO}}.$$

Plugging all this into (A.10) with $\varepsilon = 1$ yields

$$\mathrm{P}\left[\sup_{\beta\le\beta_0} \beta^{\frac{s(a+d/2)}{a+s}} \|f_{\beta,n} - f_{\mathrm{MO},\beta}\| \ge \frac{2\sqrt{2}C_{\mathrm{M0}}}{\sqrt{n}} + \frac{C_{\mathrm{M0}}\sqrt{8\xi}}{\sqrt{n}} + \frac{34.5C_{\mathrm{M0}}\xi}{n}\right] \le \exp(-\xi),$$

for all $\xi > 0$. As $\frac{1}{n} \le \frac{1}{\sqrt{n}}$ and $\sqrt{\xi} \le \xi$ for $\xi \ge 1$, this yields (18) with $\rho = C_{\mathrm{c}}\xi$. □

APPENDIX B. The Deconvolution Kernels as Approximate Inverses

Following Schuster (2007), we say that a family function $\psi_\beta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ indexed by a parameter $\beta \in (0, \beta_\circ]$ is a *mollifier* if:

(i) for every $\beta > 0$ and $y \in \mathbb{R}^d$, $\psi_\beta(\cdot, y) \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, and $\int_{\mathbb{R}} \psi_\beta(x,y)\,\mathrm{d}x = 1$;
(ii) for every $f \in L^2(\mathbb{R}^d)$, the function $f^{(\beta)}$ defined by $f^{(\beta)}(y) = \langle f, \psi_\beta(\cdot,y)\rangle$ converges to f in $L^2(\mathbb{R}^d)$ as $\beta \downarrow 0$.

For a general bounded linear operator T in $L^2(\mathbb{R}^d)$, assuming the existence of a family of functions $\big(v_\beta(\cdot,y)\big)$ such that

$$\forall\beta > 0, \quad \forall y \in \mathbb{R}^d, \quad T^* v_\beta(\cdot,y) = \psi_\beta(\cdot,y), \tag{B.1}$$

we see that $f^{(\beta)}$ is then given by $f^{(\beta)}(y) = \langle f, T^* v_\beta(\cdot,y)\rangle = \langle Tf, v_\beta(\cdot,y)\rangle = \langle g, v_\beta(\cdot,y)\rangle$. More generally, if $\psi_\beta(\cdot,y)$ belongs to $\mathscr{D}\big((T^*)\big) = \operatorname{ran} T^* + (\operatorname{ran} T^*)^\perp$, then the minimum norm least square solution to (B.1) is used instead, and denoted by $v_\beta(\cdot,y)$ again. The family of mappings

$$\begin{array}{llll} \tilde{T}_\beta : & L^2(\mathbb{R}^d) & \longrightarrow & L^2(\mathbb{R}^d) \\ & g & \longmapsto & \langle g, v_\beta(\cdot,y)\rangle \end{array}$$

is then called an *approximate inverse* of T. If $\psi_\beta(x,y) = \varphi_\beta(y-x)$ and T is the convolution operator in (1), then the function $f^{(\beta)}$ is a convolution of f:

$$f^{(\beta)}(y) = \int_{\mathbb{R}^d} f(x)\psi_\beta(x,y)\,\mathrm{d}x = \int_{\mathbb{R}^d} f(x)\varphi_\beta(y-x)\,\mathrm{d}x = (\varphi_\beta * f)(y).$$

The family of functions (φ_β) (indexed by β in some interval of the form $(0,\beta_\circ]$) emulates the Dirac distribution as $\beta \downarrow 0$). It is referred to as an *approximation of unity*. The standard way to produce such an approximation of unity is to choose an integrable function φ with $\int \varphi(x)\,\mathrm{d}x = 1$ and to define φ_β by $\varphi_\beta(x) := \frac{1}{\beta^d}\varphi\left(\frac{x}{\beta}\right), \quad x \in \mathbb{R}^d$. For the convolution operator T in (1), we have $(T^*)^{-1} = U^*\left[\frac{1}{\overline{\widehat{\gamma}}}\right]U$. Provided that $\psi_\beta(\cdot,y)$ belongs to $\operatorname{ran} T^*$, for all $\beta > 0$ and all $y \in \mathbb{R}^d$, equation (B.1) yields

$$Uv_\beta(\cdot,y)(\xi) = \left[\frac{1}{\overline{\widehat{\gamma}}} U\psi_\beta(\cdot,y)\right](\xi) = \frac{1}{\overline{\widehat{\gamma}(\xi)}} \int e^{-2i\pi\langle\xi,x\rangle}\varphi_\beta(y-x)\,\mathrm{d}x = e^{-2i\pi\langle\xi,y\rangle}\frac{\overline{\widehat{\varphi_\beta}(\xi)}}{\overline{\widehat{\gamma}(\xi)}}.$$

https://doi.org/10.1017/S0266466622000457 Published online by Cambridge University Press

Using the unitarity of U, the solution $f^{(\beta)}$ is then given by

$$f^{(\beta)}(y) := \langle g, v_\beta(\cdot, y)\rangle = \int \hat{g}(\xi) e^{2i\pi\langle \xi, y\rangle} \frac{\widehat{\varphi}_\beta(\xi)}{\widehat{\gamma}(\xi)}\,\mathrm{d}\xi = \left(U^* \left[\frac{\widehat{\varphi}_\beta}{\widehat{\gamma}}\right] Ug\right)(y).$$

Clearly, the latter estimate corresponds to the deconvolution kernel solution.

REFERENCES

Alibaud, N., P. Maréchal, & Y. Saesor (2009) A variational approach to the inversion of truncated Fourier operators. *Inverse Problems* 25(4), 045002.

Bauer, F. & M.A. Lukas (2011) Comparing parameter choice methods for regularization of ill-posed problems. *Mathematics and Computers in Simulation* 81(9), 1795–1841.

Bauschke, H.H. & P.L. Combettes (2011) *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics, vol. 408. Springer.

Bonnefond, X. & P. Maréchal (2009) A variational approach to the inversion of some compact operators. *Pacific Journal of Optimization* 5(1), 97–110.

Butucea, C. (2004) Deconvolution of supersmooth densities with smooth noise. *Canadian Journal of Statistics* 32(2), 181–192.

Butucea, C. & A.B. Tsybakov (2008) Sharp optimality in density deconvolution with dominating bias. I. *Theory of Probability & Its Applications* 52(1), 24–39.

Carrasco, M. & J.-P. Florens (2011) A spectral method for deconvolving a density. *Econometric Theory* 27(3), 546–581.

Carrasco, M., J.-P. Florens, & E. Renault (2007) Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of Econometrics* 6, 5633–5751.

Carroll, R.J. & P. Hall (1988) Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association* 83(404), 1184–1186.

Comte, F. & C. Lacour (2013) Anisotropic adaptive kernel deconvolution. *Annales de l'IHP Probabilités et Statistiques* 49, 569–609.

Comte, F., Y. Rozenholc, & M.-L. Taupin (2006) Penalized contrast estimator for adaptive density deconvolution. *Canadian Journal of Statistics* 34(3), 431–452.

Comte, F., Y. Rozenholc, & M.-L. Taupin (2007) Finite sample penalization in adaptive density deconvolution. *Journal of Statistical Computation and Simulation* 77(11), 977–1000.

Delaigle, A. & P. Hall (2016) Methodology for non-parametric deconvolution when the error distribution is unknown. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78(1), 231–252.

Delaigle, A., P. Hall, & A. Meister (2008) On deconvolution with repeated measurements. *Annals of Statistics* 36(2), 665–685.

Delaigle, A. & A. Meister (2011) Nonparametric function estimation under Fourier-oscillating noise. *Statistica Sinica* 21, 1065–1092.

Devroye, L. (1989) Consistent deconvolution in density estimation. *Canadian Journal of Statistics* 17(2), 235–239.

Efromovich, S. (1997) Density estimation for the case of supersmooth measurement error. *Journal of the American Statistical Association* 92(438), 526–535.

Engl, H.W., M. Hanke, & A. Neubauer (1996) *Regularization of Inverse Problems*. Mathematics and Its Application, vol. 375. Springer Science & Business Media.

Fan, J. (1991a) Global behavior of deconvolution kernel estimates. *Statistica Sinica* 1, 541–551.

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

Friedrichs, K.O. (1944) The identity of weak and strong extensions of differential operators. *Transactions of the American Mathematical Society* 55(1), 132–151.

Giné, E. & R. Nickl (2021) *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press.

Hall, P. & A. Meister (2007) A ridge-parameter approach to deconvolution. *Annals of Statistics* 35(4), 1535–1558.

Hào, D.N. (1994) A mollification method for ill-posed problems. *Numerische Mathematik* 68, 469–506.

Hohage, T. & F. Weidling (2017) Characterizations of variational source conditions, converse results, and maxisets of spectral regularization methods. *SIAM Journal on Numerical Analysis* 55(2), 598–620.

Johannes, J. (2009) Deconvolution with unknown error distribution. *Annals of Statistics* 37(5A), 2301–2323.

Kerkyacharian, G. & D. Picard (1993) Density estimation by kernel and wavelets methods: Optimality of Besov spaces. *Statistics & Probability Letters* 18(4), 327–336.

Lannes, A., S. Roques, & M.-J. Casanove (1987) Stabilized reconstruction in signal and image processing: I. Partial deconvolution and spectral extrapolation with limited field. *Journal of Modern Optics* 34(2), 161–226.

Lepskiĭ, O. (1991) On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability & Its Applications* 35(3), 454–466.

Lepskiĭ, O. (1992) Asymptotically minimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates. *Theory of Probability & Its Applications* 36(4), 682–697.

Lepskiĭ, O. (1993) Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation: Adaptive estimators. *Theory of Probability & Its Applications* 37(3), 433–448.

Lepskiĭ, O. & T. Willer (2017) Lower bounds in the convolution structure density model. *Bernoulli* 23(2), 884–926.

Lepskiĭ, O. & T. Willer (2019) Oracle inequalities and adaptive estimation in the convolution structure density model. *Annals of Statistics* 47(1), 233–287.

Louis, A.K. & P. Maass (1990) A mollifier method for linear operator equations of the first kind. *Inverse Problems* 6(3), 427.

Mair, B.A. & F.H. Ruymgaart (1996) Statistical inverse estimation in Hilbert scales. *SIAM Journal on Applied Mathematics* 56(5), 1424–1444.

Maréchal, P., D. Togane, & A. Celler (2000) A new reconstruction methodology for computerized tomography: FRECT (Fourier regularized computed tomography). *IEEE Transactions on Nuclear Science* 47(4), 1595–1601.

Massart, P. (2000) About the constants in Talagrand's concentration inequalities for empirical processes. *The Annals of Probability* 28(2), 863–884.

Mathé, P. (2006) The Lepskii principle revisited. *Inverse Problems* 22(3), L11.

Mathé, P. & S.V. Pereverzev (2003) Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems* 19(3), 789.

McCormick, G. & R. Tapia (1972) The gradient projection method under mild differentiability conditions. *SIAM Journal on Control* 10(1), 93–98.

Morozov, V.A. (1984) *Methods for Solving Incorrectly Posed Problems*. Springer Science & Business Media.

Murio, D.A. (2011) *The Mollification Method and the Numerical Solution of Ill-Posed Problems*. Wiley.

Nikol'skii, S.M. (1951) Inequalities for entire functions of finite degree and their application in the theory of differentiable functions of several variables. *Trudy Matematicheskogo Instituta imeni VA Steklova* 38, 244–278.

Pensky, M. & B. Vidakovic (1999) Adaptive wavelet estimator for nonparametric density deconvolution. *Annals of Statistics* 27(6), 2033–2053.

Rebelles, G. (2016) Structural adaptive deconvolution under $\lessdot_p$-losses. *Mathematical Methods of Statistics* 25(1), 26–53.

Rudin, W. (1970) *Real and Complex Analysis: Part 2*. McGraw-Hill.

Schuster, T. (2007) *The Method of Approximate Inverse: Theory and Applications*. Lecture Notes in Mathematics, vol. 1906. Springer.

Stefanski, L.A. & R.J. Carroll (1990) Deconvolving kernel density estimators. *Statistics* 21(2), 169–184.

Tikhonov, A.N. & V.Y. Arsenin (1977) *Methods for Solving Ill-Posed Problems*. Wiley.

Triebel, H. (2006) *Theory of Function Spaces III*. Monographs in Mathematics, vol. 100. Birkhäuser Basel.

Trong, D.D. & C.X. Phuong (2015) Ridge-parameter regularization to deconvolution problem with unknown error distribution. *Vietnam Journal of Mathematics* 43(2), 239–256.

Van Rooij, A.C.M. & F.H. Ruymgaart (1991) Regularized deconvolution on the circle and the sphere. In G. Roussas (ed.), *Nonparametric Functional Estimation and Related Topics*, pp 679–690. Springer, Dordrecht.

Wikipedia Contributors (2020) Mollifier—Wikipedia, the free encyclopedia (Online; accessed 17 April 2020).