

Smoothness-Penalized Deconvolution (SPeD) of a Density Estimate

David Kent & David Ruppert

To cite this article: David Kent & David Ruppert (15 Sep 2023): Smoothness-Penalized Deconvolution (SPeD) of a Density Estimate, Journal of the American Statistical Association, DOI: [10.1080/01621459.2023.2259028](https://doi.org/10.1080/01621459.2023.2259028)

To link to this article: <https://doi.org/10.1080/01621459.2023.2259028>



View supplementary material [↗](#)



Accepted author version posted online: 15 Sep 2023.



Submit your article to this journal [↗](#)



Article views: 88



View related articles [↗](#)



View Crossmark data [↗](#)

Smoothness-Penalized Deconvolution (SPeD) of a Density Estimate

David Kent^{a,*†} and David Ruppert^{a,b,†}

^aDepartment of Statistics and Data Science, Cornell University

^bSchool of Operations Research and Information Engineering, Cornell University

*dk657@cornell.edu

[†]This work was supported by the National Science Foundation under Grant AST-1814840. The opinions, findings, and conclusions, or recommendations expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Abstract

This paper addresses the deconvolution problem of estimating a square-integrable probability density from observations contaminated with additive measurement errors having a known density. The estimator begins with a density estimate of the contaminated observations and minimizes a reconstruction error penalized by an integrated squared m -th derivative. Theory for deconvolution has mainly focused on kernel- or wavelet-based techniques, but other methods including spline-based techniques and this smoothness-penalized estimator have been found to outperform kernel methods in simulation studies. This paper fills in some of these gaps by establishing asymptotic guarantees for the smoothness-penalized approach. Consistency is established in mean integrated squared error, and rates of convergence are derived for Gaussian, Cauchy, and Laplace error densities, attaining some lower bounds already in the literature. The assumptions are weak for most results; the estimator can be used with a broader class of error densities than the deconvoluting kernel. Our application example estimates the density of the mean cytotoxicity of certain bacterial isolates under random sampling; this mean cytotoxicity can only be measured experimentally with additive error, leading to the deconvolution problem. We also describe a method for approximating the solution by a cubic spline, which reduces to a quadratic program.

Keywords: ill-posed problem, measurement error, density estimation, regularization

1 Introduction

A smoothness-penalized density deconvolution estimator was introduced in Yang et al. [2020] which is fast to compute, amenable to shape constraints, and in simulation studies has substantially improved finite-sample performance over the common deconvoluting kernel density estimator of Stefanski and Carroll [1990]. A spline-based Bayesian approach for a related problem in Staudenmayer et al. [2008] outperforms deconvoluting kernels in simulation studies as well, and Sarkar et al. [2014] does yet better. In spite of these appealing properties, these estimators have not yet received much attention. This is perhaps due to a lack of theoretical guarantees; most asymptotic results for deconvolution estimators focus on kernel- or wavelet-based (e.g. Pensky and Vidakovic [1999]) methods, while these other methods have only been addressed in simulations.

In this paper, we address a continuous version of the smoothness-penalized estimator in Yang et al. [2020] and provide some theoretical guarantees. We prove the consistency of the density estimates in L_2 and derive upper bounds for the rate of convergence, which are found to be optimal when compared to lower bounds already in the literature. We also prove in Theorem 7, under stronger smoothness conditions inspired by typical assumptions in the ill-posed problem literature, fast rates of convergence which hold for any error density, whether smooth or super-smooth. We are not aware of similar results for kernel-based deconvolution estimators. Along the way, we derive a representation of the estimator which is more convenient for theoretical work than the variational formulation in Yang et al. [2020] and investigate the finite-sample error for a few settings.

Suppose a real-valued random variable of interest X has pdf f , and we wish to estimate f . However, we instead observe independent copies of $Y = X + E$, a surrogate of X which has been contaminated with an independent error E . Suppose further that E has known pdf g . Under these conditions, the pdf h of Y is given by the *convolution* of g and f , i.e.

$$h(y) = g * f(y) = \int f(t)g(y-t)dt.$$

The task of estimating the density f from a sample $Y_1, \dots, Y_n$ of independent random variables with pdf $h = g * f$ is sometimes called a *deconvolution* problem, since we can think of the main goal as “undoing” the convolution with g .

The method in Yang et al. [2020], to be described shortly, begins with a density estimate of h , and proceeds to estimate f through this density estimate of h . To that end, we will introduce one more abstraction: we will assume that we have access to an $L_2(\mathbb{R})$ -consistent estimator of h , which we will denote h_n (Estimators of functions will be indicated by a subscript n rather than the customary “hat” to avoid clutter when taking Fourier transforms, which will be denoted by an overset twiddle). The mean integrated squared error (MISE) of h_n will be denoted

$\delta_n^2 = \mathbb{E} \|h_n - h\|^2 = \mathbb{E} \int (h_n - h)^2$. We will think of h_n as the “data” in this problem and express the performance of our estimator in terms of δ_n^2 . Note that since the estimator is consistent, we have $\delta_n^2 \rightarrow 0$.

This setting occurs whenever a density estimate is required, but the variable is measured with error; it is therefore a nearly ubiquitous phenomenon, but typically ignored when the measurement error is small. For a window into the meaning of “small” here, note that ignoring measurement error E means to estimate $g * f$ in place of f , incurring at a point x the error $g * f(x) - f(x) = \mathbb{E}[f(x-E) - f(x)]$, where E has pdf g . Thus wherever f has large curvature on the scale of E , $g * f$ and f will not be similar, and in such cases measurement error should not be ignored.

Because of the ubiquity of the setting, the application domains are diverse. In this paper, we estimate a conditional density that occurs when estimating the cytotoxicity of bacterial isolates; in Yang et al. [2021], the authors use deconvolution to estimate the density of a conditional expectation occurring in nested Monte Carlo simulations. Staudenmayer et al. [2008] apply deconvolution to nutritional data from a clinical trial involving a dietary supplement, and Stefanski and Carroll [1990] treat data on saturated fat intake.

The treatment of this problem makes great use of the Fourier transform. Following conventions in Folland [1992], for $v \in L_1(\mathbb{R}) \cup L_2(\mathbb{R})$ we will write

$$\tilde{v}(\omega) = \lim_{r \rightarrow \infty} \int_{-r}^r e^{-i\omega x} v(x) dx = \int e^{-i\omega x} v(x) dx \quad \text{for the Fourier transform of } v, \text{ with the second}$$

equality holding as long as $v \in L_1(\mathbb{R})$. Let $\tilde{P}_n(\omega) = \frac{1}{n} \sum_{j=1}^n e^{-iY_j \omega}$ be the Fourier transform of the empirical distribution. If $u(x) = v * w(x)$, then $\tilde{u}(\omega) = \tilde{v}(\omega) \tilde{w}(\omega)$, so that the Fourier transform reduces convolution to multiplication.

A well-known estimator of f in this setting is the deconvoluting kernel (density) estimator (DKE), introduced in Stefanski and Carroll [1990], which takes advantage of the reduction of convolution to multiplication. First, we form a kernel density

estimate $h_n^\lambda(x) = \frac{1}{n\lambda} \sum_{j=1}^n K((x - Y_j)/\lambda)$ of h , in which case $\tilde{h}_n^\lambda(\omega) = \tilde{P}_n(\omega) \tilde{K}(\lambda\omega)$. Then we divide by $\tilde{g}(\omega)$ and inverse transform:

$$f_n^\lambda(x) = \frac{1}{2\pi} \int e^{i\omega x} \tilde{P}_n(\omega) \tilde{K}(\lambda\omega) / \tilde{g}(\omega) d\omega. \quad (1)$$

In Stefanski and Carroll [1990], they find that if $K_\lambda^*(x) = (2\pi)^{-1} \int e^{i\omega x} \tilde{K}(\omega) / \tilde{g}(\omega/\lambda) d\omega$, then f_n^λ has representation

$$f_n^\lambda(x) = \frac{1}{n\lambda} \sum_{j=1}^n K_\lambda^*((x - Y_j)/\lambda)$$

and are able to borrow from results on standard kernel density estimators in their analysis. To ensure that the Fourier inversion in Equation (1) is well-defined,

Stefanski and Carroll [1990] require K to be chosen to satisfy $\sup_{\omega} |\tilde{K}(\omega) / \tilde{g}(\omega/\lambda)| < \infty$

and $\int |\tilde{K}(\omega) / \tilde{g}(\omega/\lambda)| d\omega < \infty$ for all $\lambda > 0$, suggesting band-limited kernels, including

$$K(x) = \frac{1}{\pi} (\sin(x)/x)^2, \quad \text{which has Fourier transform } \kappa(\omega) = 1_{[-2,2]}(1 - |\omega|/2). \quad \text{Note that}$$

in particular, $K(x) = g(x)$ typically cannot satisfy these conditions. Additionally, Stefanski and Carroll [1990] restrict attention to g for which $|\tilde{g}(\omega)| > 0$, since the estimator involves division by $\tilde{g}(\omega)$. Under appropriate choice of $\lambda_n \rightarrow 0$, the

estimator is consistent and attains optimal rates in several settings. In Fan [1991], optimal rates are addressed for f in a class of functions with m th derivative Hölder-continuous. In Zhang [1990], optimal rates are addressed over for f in a class of functions satisfying $\|\omega \tilde{f}(\omega)\|^2 < M < \infty$.

The density deconvolution technique introduced in Yang et al. [2020] discretizes both the functions and the convolution operator. The estimate h_n is approximated on a grid by a vector $\mathbf{h}_n$, and the convolution operator by a matrix $\mathbf{C}$, so that if $\mathbf{v}$ is a discrete approximation of a function v , then $\mathbf{C}\mathbf{v}$ is a discrete approximation of g^*v . Then a discrete approximation $\mathbf{f}_n^\alpha$ of f is computed by solving the matrix problem

$$\mathbf{f}_n^\alpha = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{C}\mathbf{x} - \mathbf{h}_n\|^2 + \alpha Q(\mathbf{x}), \quad (2)$$

where $Q(\cdot)$ is a quadratic penalty. (Vectors and matrices will always be typeset in boldface, and we overload $\|\cdot\|$ to denote the vector 2-norm when the argument is a vector.) For $Q(\cdot)$ the authors suggest, among other choices, the squared norm of a second-differencing operator applied to $\mathbf{x}$: $Q(\mathbf{x}) = \|\mathbf{D}_2 \mathbf{x}\|_2^2$. Heuristically, this approach yields an estimate $\mathbf{f}_n^\alpha$ whose convolution $\mathbf{C}\mathbf{f}_n^\alpha$ is close to the density estimate $\mathbf{h}_n$ (due to the first term), but which is not too wiggly (due to the second term). They observe that Equation (2) can be formulated as a quadratic program and solved efficiently using standard methods, and that linear constraints can be introduced as well.

In this paper, we analyze the exact, continuous version of the estimator introduced in Yang et al. [2020], with penalty $Q(v) = \|v^{(m)}\|^2$, where $v^{(m)}$ denotes the m th derivative of $v \in L_2(\mathbb{R})$, i.e.

$$f_n^\alpha = \underset{v}{\operatorname{argmin}} \|g^*v - h_n\|^2 + \alpha \|v^{(m)}\|^2. \quad (3)$$

The argument v is taken to range over the subset of $L_2(\mathbb{R})$ for which the objective function is well-defined, which we will make specific in Section 4. We will occasionally find it useful to use operator notation, with $T : L_2(\mathbb{R}) \rightarrow L_2(\mathbb{R})$, $T : v \mapsto g^*v$ and $L : \mathcal{D}(L) \rightarrow L_2(\mathbb{R})$, $L : v \mapsto v^{(m)}$, so that we can alternatively write

$$f_n^\alpha = \operatorname{argmin}_v \|Tv - h_n\|^2 + \alpha \|Lv\|^2.$$

In Section 6, we suggest an alternative to the discretization approach in Yang et al. [2020]. We instead solve Equation (3) out of an approximation space of piecewise polynomial spline functions. Calling this approximation s_n^α , we prove in Theorem 15 that the resulting approximation error $\mathbb{E} \|s_n^\alpha - f_n^\alpha\|^2$ can be made to decrease faster than the order of convergence of $\mathbb{E} \|f_n^\alpha - f\|^2$ by choosing a suitably rich approximation space. It follows that $\mathbb{E} \|s_n^\alpha - f\|^2$ has the same order of convergence.

One appealing property of this estimator is that the computational techniques proposed here and in Yang et al. [2020] can be quite fast, with the computational complexity determined primarily by the dimension of the discretization grid or spline basis. We will see that computing the spline approximation can be formulated as a quadratic program, and that many useful linear constraints can be imposed. Among these, positivity and integrate-to-one constraints are easily imposed, as are support constraints and some shape constraints. Yang et al. [2020] even suggest a method for imposing a unimodal constraint by a family of “unimodal at a point x_0 ” constraints, each of which can be imposed as a linear constraint.

Finally, the DKE approach requires g to have non-vanishing Fourier transform and therefore cannot be applied to, for example, uniformly distributed errors. The estimator addressed here has no such requirement; instead, the Fourier transform of g must only be non-vanishing almost-everywhere, which is also a necessary condition for identifiability in this model.

After an overview of the inherent difficulties of deconvolution in Section 2 and introducing the estimator in detail in Section 4, we prove global $L_2(\mathbb{R})$ -consistency, as well as rates of convergence in Section 5. In Section 6, we address the practical issue of computing the estimate, investigate its performance in finite samples, and apply it to a problem on bacterial cytotoxicity.

2 Ill-Posedness of the Problem

Deconvolving a density estimate is a typical “ill-posed” problem. We will see that ill-posedness means a naive solution to the deconvolution problem must fail to be consistent, and any consistent deconvolution estimator must reflect some aspect of regularization. A problem is said to be well-posed if [Engl et al., 1996, Chapter 2] the following conditions are met: “ 1. For all admissible data, a solution exists, 2. For all admissible data, the solution is unique, and 3. The solution depends continuously on the data,” and ill-posed otherwise.

For the deconvolution problem, consider the operator $T: L_2(\mathbb{R}) \rightarrow L_2(\mathbb{R})$ which convolves a function with g , i.e. $v \mapsto g * v$. Since $h = g * f$, plugging in $v = f$ clearly solves the following operator equation:

$$Tv = h. \quad (4)$$

However, we do not know h . We have an estimate h_n of h , and we would like to solve the analogous problem with our estimate h_n on the right-hand side, i.e.

$$Tv = h_n. \quad (5)$$

There is an immediate issue with this approach: there is no v solving this equation unless $h_n \in \mathcal{R}(T)$, i.e. $h_n = T\psi$ for some $\psi \in L_2(\mathbb{R})$. If h_n is unrestricted, the problem of solving Equation (5) violates Condition 1 of well-posedness. However, we can overcome this problem by using a generalized inverse of T , so we will set it aside for the moment.

Instead, we will focus on a more critical deficiency: the solution operator for Equation (5) is not continuous in h_n . This means that a small perturbation of the right-hand side can lead to arbitrarily large fluctuations in the solution, so that problem of solving $Tv = h_n$ is not a good approximation of solving $Tv = h$ no matter how well h_n approximates h . If we require $h_n \in \mathcal{R}(T)$ so that the solution operator is simply T^{-1} , then this discontinuity would entail that for any $\varepsilon > 0$ and $C > 0$, we can have $\|h_n - h\| < \varepsilon$, but $\|T^{-1}h - T^{-1}h_n\| = \|f - T^{-1}h_n\| > C$. No matter how good we require the estimate h_n of h to be, its exact deconvolution may yet be an arbitrarily bad estimate

of f . Let's prove it formally: the following proposition guarantees the existence of a function u so that taking $h_n = h + u$ creates the unhappy situation just described.

Proposition 1. *Assume that the Fourier transform $\tilde{g}$ of g is a.e. non-vanishing, so that $T: L_2(\mathbb{R}) \rightarrow L_2(\mathbb{R})$ is injective (Supplemental Fact A.1). Let $T^{-1}: \mathcal{R}(T) \rightarrow L_2(\mathbb{R})$ be the inverse of T from its range. Then, for any $M > 0$, there is some $u \in \mathcal{R}(T)$ for which $\|T^{-1}u\| > M \|u\|$.*

Proof of Proposition 1. We will construct a sequence $\phi_n \in \mathcal{R}(T)$ which is a Cauchy sequence in $L_2(\mathbb{R})$, but with the property that for $n \in \mathbb{N}$, we have $\|T^{-1}(\phi_n - \phi_{n+1})\| = 1$. Once we have this sequence, we can finish the proof in the following way. Fix $M > 0$.

Since ϕ_n is Cauchy, we can choose $n \in \mathbb{N}$ large enough that $\|\phi_n - \phi_{n+1}\| < \frac{1}{M}$. Then $u = \phi_n - \phi_{n+1}$ satisfies

$$\|T^{-1}(\phi_n - \phi_{n+1})\| > M \|\phi_n - \phi_{n+1}\|,$$

as needed.

Now, if we can find such a sequence ϕ_n , we are finished. To that end, let $\psi_n = n1_{[0, 1/n]}$. It can be checked that $\|\psi_n - \psi_{n+1}\| = 1$. Furthermore, the ψ_n constitute an "approximate identity," so that by Folland [1999, Theorem 8.14a], we have $\|g * \psi_n - g\| \rightarrow 0$ as $n \rightarrow \infty$. Now, let $\phi_n = g * \psi_n = T\psi_n$. To see that ϕ_n is Cauchy, apply the triangle inequality:

$$\|\phi_n - \phi_m\| \leq \|\phi_n - g\| + \|\phi_m - g\| = \|g * \psi_n - g\| + \|g * \psi_m - g\|.$$

For the other property, note that

$$\|T^{-1}(\phi_n - \phi_{n+1})\| = \|T^{-1}(T\psi_n - T\psi_{n+1})\| = \|\psi_n - \psi_{n+1}\| = 1,$$

finishing the proof. $\square$

Now, even if $h_n \notin \mathcal{R}(T)$, a generalized inverse like the Moore-Penrose inverse $T^\dagger$ may be used in place of T^{-1} , ensuring that Conditions 1 and 2 are met. But these

generalized inverses extend T^{-1} from $\mathcal{R}(T)$, so they too fail to be continuous by Proposition 1.

We turn to a method of regularization solution. A regularized solution of Equation (4) is a family of operators $\{R_\alpha\}_{\alpha>0}$ which approximate T^{-1} or an extension thereof, and which has the property that for each α , R_α is a continuous operator. For our choice of regularization by smoothness penalty, we will see in Theorem 2 that each R_α is a bounded operator. In Theorem 6, we will see that the regularization does approximate the exact solution to Equation (4), and in Theorems 7 & 9, we will see the rates of convergence under a few different conditions.

3 Assumptions

We assume throughout that f and g are probability densities, and that $h_n \in L_2(\mathbb{R})$. The following is a list of all further assumptions that recur in the theoretical results; in each statement we will name the assumptions required from this list. Assumptions that are used only for a single result are stated in that result. First, assumptions that will be made on the target density f : (F1) $f \in L_2(\mathbb{R})$; (F2) $\int |\omega^k \tilde{f}(\omega)|^2 d\omega < \infty$ for some $1 \leq k \leq 2m$.

Now, assumptions that will be made on the error density g : (G1) $\tilde{g}$ vanishes only on a set of Lebesgue measure zero; (G2) $g \in L_2(\mathbb{R})$; (G3) $\int |\tilde{g}(\omega)| d\omega < \infty$; Note that Assumptions (G1)- (G3) all hold for Normal, Cauchy, and Laplace errors. Note also that if $f \in L_2(\mathbb{R})$, then by Young's convolution inequality, $h \in L_2(\mathbb{R})$ as well.

4 The Estimator

One family of solution operators for Equation (5) which extend T^{-1} are those which map h_n to a *least-squares solution*, i.e. to a v minimizing $\|Tv - h_n\|$. The familiar Moore-Penrose generalized inverse $T^\dagger$ is a least-squares extension—it is the operator which maps h_n to the least-squares solution v for which v has minimal norm $\|v\|$. Classical Tikhonov regularization approximates $T^\dagger$ by the family of operators $\{S_\alpha\}_{\alpha>0}$ mapping

$$S_\alpha : h_n \mapsto \operatorname{argmin}_v \|Tv - h_n\|^2 + \alpha \|v\|^2.$$

Intuitively, the solution $S_\alpha h_n$ is a function which is reasonably small in $L_2(\mathbb{R})$ due to the second term, and for which $\|TS_\alpha h_n - h_n\|$ is reasonably small.

Here we address a similar approach, but rather than preferring a function which is small in $L_2(\mathbb{R})$, we prefer one which is smooth, in the sense that its m th derivative, $m \geq 1$, has small norm. Thus we have a family $\{R_\alpha\}_{\alpha>0}$ mapping

$$R_\alpha : h_n \mapsto \operatorname{argmin}_v \|Tv - h_n\|^2 + \alpha \|v^{(m)}\|^2.$$

This is a particular case of Tikhonov regularization with differential operators, which has been treated in an abstract, non-statistical framework in Locker and Prenter [1980], Engl et al. [1996, Chapter 8], and Nair et al. [1997].

Since our estimator will measure the smoothness of a possible estimate by the magnitude of its square-integrated m th derivative, the estimate must be chosen from among those functions for which this quantity is finite. To that end, let

$H^m(\mathbb{R}) = \{v \in L_2(\mathbb{R}) : v^{(k)} \in L_2(\mathbb{R}) \text{ for } 0 \leq k \leq m\}$ denote the Sobolev space of square-integrable functions with square-integrable weak derivatives up to order m . Assume throughout that $A = \{\omega \mid \tilde{g}(\omega) = 0\}$ has Lebesgue measure zero.

Definition 1. The **Tikhonov functional** with data $u \in L_2(\mathbb{R})$ and penalty parameter $\alpha > 0$ is a function defined by

$$\begin{aligned} G(\cdot; u, \alpha) : H^m(\mathbb{R}) &\rightarrow \mathbb{R} \\ v &\mapsto \|g * v - u\|^2 + \alpha \|v^{(m)}\|^2. \end{aligned}$$

Definition 2. Let h_n be a density estimate of h from the sample $Y_1, \dots, Y_n$, and let $\alpha > 0$. The **Smoothness-penalized deconvolution of h_n** or **Smoothness-penalized deconvolution estimate (SPeD) of f** is defined variationally by

$$\begin{aligned} f_n^\alpha &= \operatorname{argmin}_{v \in H^m(\mathbb{R})} G(v; h_n, \alpha) \\ &= \operatorname{argmin}_{v \in H^m(\mathbb{R})} \|g * v - h_n\|^2 + \alpha \|v^{(m)}\|^2. \end{aligned} \quad (6)$$

Remark. For a given $h_n \in L_2(\mathbb{R})$ and $\alpha > 0$, the estimator f_n^α in Definition 2 is uniquely defined [Locker and Prenter, 1980, Theorem 3.5]. Moreover, $f_n^\alpha \in H^{2m}(\mathbb{R})$.

4.1 Representations of the estimator

Since the variational characterization of f_n^α does not lend itself to easy analysis, in Theorem 2 we present an explicit representation for f_n^α , both in terms of h_n and the Fourier transform of h_n ; if a kernel density estimator is used for h_n , we will see that f_n^α can be computed as a kernel estimate as well, though this is not the approach we take in the sequel. The Fourier representation will make clear the manner in which the Tikhonov regularization approximates the ill-posed exact deconvolution problem.

Theorem 2. (*Representing the solution*) Let

$$\tilde{\varphi}_\alpha(\omega) = \frac{\overline{\tilde{g}(\omega)}}{|\tilde{g}(\omega)|^2 + \alpha\omega^{2m}} \quad \text{and} \quad \varphi_\alpha(x) = \lim_{r \rightarrow \infty} \frac{1}{2\pi} \int_{-r}^r e^{i\omega x} \tilde{\varphi}_\alpha(\omega) d\omega. \quad (7)$$

Then

- (i) $\tilde{f}_n^\alpha(\omega) = \tilde{\varphi}_\alpha(\omega) \tilde{h}_n(\omega)$,
- (ii) $f_n^\alpha(x) = \varphi_\alpha * h_n(x)$, and
- (iii) if h_n is a kernel density estimate with bandwidth v , then there is another kernel $K_{\alpha,v}$ for which

$$f_n^\alpha(x) = \frac{1}{n} \sum_{i=1}^n K_{\alpha,v}(x - Y_i).$$

Furthermore,

- (iv) $\sup_{\omega} |\tilde{\varphi}_\alpha(\omega)| \leq C\alpha^{-\frac{1}{2}}$ for all $\alpha < M$,
- (v) If $\varphi_\alpha \in L_1(\mathbb{R})$, then $\int \varphi_\alpha(x) dx = 1$, so that $\int f_n^\alpha(x) dx = 1$.
- (vi) $\|D^k f_n^\alpha\| \leq C\alpha^{-1}$ for all $\alpha < M$ and $0 \leq k < 2m$ with C depending only on g , m , and M . Under Assumption (G2), it holds for $k = 2m$ as well.

(vii) $\sup_x |D^k f_n^\alpha(x)| \leq C\alpha^{-1}$ for all $\alpha < M$ and $0 \leq k < 2m-1$ with C depending only on g , m , and M . Under Assumption (G3), it holds for all $0 \leq k \leq 2m$.

Remark. Before moving on to the proof, it is worth making a few observations.

- Except where mentioned in (vi) and (vii), Theorem 2 does not need any assumptions on g or h_n beyond the fact that they are probability densities.
- If g is an even function, then $\tilde{\varphi}_\alpha(\omega)$ is even and purely real.
- Theorem 2(iv) holds for any g , but can be made sharper with information about a particular choice of g . See, for example, the proof of Theorem 9(iii).
- Theorem 2(iv) equivalently says that, with R_α denoting the operator which maps $h_n \mapsto f_n^\alpha$, the operator norm has a bound $\|R_\alpha\| \leq C\alpha^{-\frac{1}{2}}$, showing that the solution operator for each α is bounded.
- If the density estimate is a kernel density estimate h_n^λ with kernel appropriate for the DKE (e.g. bandlimited), then for fixed data and bandwidth λ , if we let $\alpha \rightarrow 0$, we have that $f_n^\alpha \rightarrow f_n^\lambda$, i.e. we obtain the DKE defined in Equation (2).

Proof of Theorem 2.

(i):

By Theorem 3.1 of Locker and Prenter [1980], a function f_n^α minimizes the Tikhonov functional $G_n^\alpha(f)$ if and only if $f_n^\alpha \in \mathcal{D}(L^*L)$ and f_n^α satisfies the Euler-Lagrange equation $(T^*T + \alpha L^*L)f_n^\alpha = T^*h_n$. By Supplemental Fact A.2, this corresponds to $g \star g^* f_n^\alpha + \alpha L^*L f_n^\alpha = g \star h_n$, where $g \star u(t) = \int g(x-t)u(x)dx$, so taking Fourier transforms yields (see Supplemental Fact A.2 for details)

$$|\tilde{g}(\omega)|^2 \tilde{f}_n^\alpha(\omega) + \alpha \omega^{2m} \tilde{f}_n^\alpha(\omega) = \overline{\tilde{g}(\omega)} \tilde{h}_n(\omega),$$

and re-arranging gives

$$\tilde{f}_n^\alpha(\omega) = \frac{\overline{\tilde{g}(\omega)}}{|\tilde{g}(\omega)|^2 + \alpha \omega^{2m}} \tilde{h}_n(\omega) = \tilde{\varphi}_\alpha(\omega) \tilde{h}_n(\omega),$$

as needed.

(ii) and (iv):

It will be convenient to prove (iv) first. We prove the equivalent inequality that for all ω , $\sqrt{\alpha} |\tilde{\varphi}_\alpha(\omega)| \leq C$. We do this by demonstrating two facts: first, that for all ω , $\sqrt{\alpha} |\tilde{\varphi}_\alpha(\omega)| < \frac{1}{2} |\omega|^{-m}$, so that if we can bound $\sqrt{\alpha} |\tilde{\varphi}_\alpha(\omega)|$ on a neighborhood of zero, we are finished, since the bound decreases as $|\omega| \rightarrow \infty$. Second, we show that $\sqrt{\alpha} |\tilde{\varphi}_\alpha(\omega)| \leq \frac{\sqrt{M}}{|\tilde{g}(\omega)|}$, and that on a neighborhood $|\omega| \leq \varepsilon$ of zero, $\tilde{g}(\omega)$ is bounded away from zero: $0 < c \leq |\tilde{g}(\omega)| \leq 1$, and take $C = \max\{\sqrt{M}/c, \frac{1}{2}\varepsilon^{-m}\}$.

For the first, apply the inequality $x + y \geq 2\sqrt{xy}$ for $x, y > 0$ to the denominator of $\tilde{\varphi}_\alpha$:

$$\sqrt{\alpha} |\tilde{\varphi}_\alpha(\omega)| = \sqrt{\alpha} \frac{|\tilde{g}(\omega)|}{|\tilde{g}(\omega)|^2 + \alpha\omega^{2m}} \leq \sqrt{\alpha} \frac{|\tilde{g}(\omega)|}{2\sqrt{\alpha} |\tilde{g}(\omega)|^2 \omega^{2m}} = \frac{1}{2} |\omega|^{-m}.$$

For the second,

$$\sqrt{\alpha} |\tilde{\varphi}_\alpha(\omega)| = \sqrt{\alpha} \frac{|\tilde{g}(\omega)|}{|\tilde{g}(\omega)|^2 + \alpha\omega^{2m}} \leq \sqrt{\alpha} \frac{|\tilde{g}(\omega)|}{|\tilde{g}(\omega)|^2} \leq \sqrt{M} |\tilde{g}(\omega)|^{-1},$$

where the first inequality is because $\alpha\omega^{2m} > 0$, and the second inequality is by the assumption that $\alpha < M$. Finally, to see that $\tilde{g}$ is bounded away from zero on a neighborhood of zero, recall that $\tilde{g}$ is the Fourier transform of a probability density g . Thus $\tilde{g}(0) = 1$, and $\tilde{g}$ is continuous, proving (iv).

Now, we will demonstrate that then $\tilde{\varphi}_\alpha \in L_2(\mathbb{C})$, so that the Fourier inversion in Equation (16) is legitimate. By the arguments proving (iv), we have also found a

square-integrable function $b(\omega) = \alpha^{-\frac{1}{2}} (1_{|\omega| < \varepsilon} C + \frac{1}{2} 1_{|\omega| \geq \varepsilon} |\omega|^{-m})$, such that

$|b(\omega)| \geq |\tilde{\varphi}_\alpha(\omega)|$. Thus, $\int |\tilde{\varphi}_\alpha(\omega)|^2 d\omega \leq \int |b(\omega)|^2 d\omega < \infty$, and $\tilde{\varphi}_\alpha \in L_2(\mathbb{C})$. Now, (ii)

follows from (i) and the well-known properties of the Fourier transform.

(iii) and (v)-(vii): Deferred to Supplemental Proof B.2. $\square$

Regularized solutions are well-behaved approximations to a poorly behaved exact problem, and the Fourier view of our estimator gives a nice picture of the manner of

approximation. Suppose briefly that $\tilde{g}$ is even and non-vanishing. Taking the Fourier transform of Equation (5) reduces the convolution to multiplication, giving $\tilde{g}\tilde{v} = \tilde{h}_n$, so that we may write $\tilde{v} = \tilde{h}_n / \tilde{g}$. Thus in Fourier space, exact deconvolution of h_n corresponds to multiplying $\tilde{h}_n$ by $1/\tilde{g}$. In Theorem 2(i), we see that in Fourier space, our regularized solution $\tilde{f}_n^\alpha$ corresponds to multiplying $\tilde{h}_n$ by this $\tilde{\varphi}_\alpha$ function. Inspection of $\tilde{\varphi}_\alpha(\omega)$ shows that when $|\omega|$ is small, $\tilde{\varphi}_\alpha(\omega) \approx 1/\tilde{g}(\omega)$, but that when $|\omega|$ is large, the $\alpha\omega^{2m}$ term dominates the expression and $\varphi_\alpha(\omega) \approx 0$, since $\tilde{g}(\omega)$ is bounded. Thus multiplying by $\tilde{\varphi}_\alpha$ performs similarly to multiplying by $1/\tilde{g}$ at low frequencies, but $\tilde{\varphi}_\alpha$ prevents high-frequency features of h_n from transferring to $\tilde{f}_n^\alpha$. This is pictured in Figure 1 for Gaussian g and a variety of α .

4.2 Decomposing the error

To analyze the error $\tilde{f}_n^\alpha - f$, it is useful to introduce a non-random function f^α for which $\tilde{f}_n^\alpha - f^\alpha$ represents the systematic error induced by solving the α -regularized problem in place of the exact problem.

Definition 3. The α -smoothed f , denoted f^α , is given by $f^\alpha = \arg\min_{v \in H^m(\mathbb{R})} G(v; h, \alpha)$.

Remark. The α -smoothed f is the smoothness-penalized deconvolution of the exact data h . In Supplemental Proposition A.1, it is shown to have representations $f^\alpha = \varphi_\alpha * h$ and $\tilde{f}^\alpha = \tilde{\varphi}_\alpha \tilde{h}$, and approximates f in the sense that $\|f^\alpha - f\| \rightarrow 0$ as $\alpha \rightarrow 0$.

As the next lemma shows, an appealing property of f^α is that, for fixed α , $\mathbb{E} \|\tilde{f}_n^\alpha - f^\alpha\|^2$ becomes small when $\delta_n^2 = \mathbb{E} \|h_n - h\|^2$ gets smaller, in contrast to the issue with exact deconvolution outlined in Proposition 1.

Lemma 3. Assume (F1). There is a C depending only on g , such that for sufficiently small $\alpha > 0$, we have $\mathbb{E} \|\tilde{f}_n^\alpha - f^\alpha\|^2 \leq C\delta_n^2 / \alpha$

Proof of Lemma 3. By the Plancherel Theorem,

$$\begin{aligned}
\|f_n^\alpha - f^\alpha\|^2 &= \frac{1}{2\pi} \|\tilde{f}_n^\alpha - \tilde{f}^\alpha\|^2 \\
&= \frac{1}{2\pi} \int |\varphi_\alpha(\omega)|^2 |\tilde{h}_n(\omega) - \tilde{h}(\omega)|^2 d\omega \\
&\leq \sup_{\omega} |\tilde{\varphi}_\alpha(\omega)|^2 \|h_n - h\|^2 \leq C \|h_n - h\|^2 / \alpha
\end{aligned}$$

where the second inequality is by Theorem 2(iv). Taking expectations gives the result. $\square$

Corollary 4. *Assume (F1). For sufficiently small α , we have the upper bound*

$$\mathbb{E} \|f_n^\alpha - f\|^2 \leq C \delta_n^2 / \alpha + 2 \|f^\alpha - f\|^2.$$

Proof of Corollary 4. Note that $(a+b)^2 \leq 2a^2 + 2b^2$, which can be seen by expanding $0 \leq (a-b)^2$, adding $a^2 + b^2$ to both sides, and re-arranging. Then the result follows from the triangle inequality and Lemma 3. $\square$

The rate at which $\|f^\alpha - f\| \rightarrow 0$ with α depends intimately on the particular form of g . In Lemma 5, we present upper bounds for $\|f^\alpha - f\|^2$ in terms of α .

Lemma 5. *Assume (F1), (F2). Then, with $W(\cdot)$ denoting the principal branch of the Lambert W function,*

(i) *(Normal errors) If $g(x) = (2\pi)^{-1} e^{-x^2/2}$, then*

$$\|f^\alpha - f\|^2 \leq \frac{C}{m^k W(m^{-1} \alpha^{-\frac{1}{m}})^k} \sim \frac{C}{m^k \log(m^{-1} \alpha^{-\frac{1}{m}})^k}$$

(ii) *(Cauchy errors) If $g(x) = \frac{1}{\pi(1+x^2)}$, then*

$$\|f^\alpha - f\|^2 \leq \frac{C}{m^{2k} W(m^{-1} \alpha^{-\frac{1}{2m}})^{2k}} \sim \frac{C}{m^{2k} \log(m^{-1} \alpha^{-\frac{1}{2m}})^{2k}}$$

(iii) *(Laplace errors) If $g(x) = \frac{1}{2} e^{-|x|}$, then*

$$\|f^\alpha - f\|^2 \leq C \left(\frac{1}{4\alpha} \right)^{\frac{k}{m+1}},$$

The asymptotic equivalences in the first two parts follow from Supplemental Fact A.4(iv)

Proof deferred to Supplemental Proof B.5.

5 Asymptotics

5.1 Consistency and Rates of Convergence

If Assumption (G1) holds, then $\|f^\alpha - f\| \rightarrow 0$ as $\alpha \rightarrow 0$, and the upper bound in Corollary 4 provides a sufficient condition for $L_2(\mathbb{R})$ -consistency of f_n^α :

Theorem 6. ($L_2(\mathbb{R})$ consistency) Assume (F1), (G1). Assume that $\delta_n^2 \rightarrow 0$, and α_n is chosen so that $\delta_n^2 / \alpha_n \rightarrow 0$ and $\alpha_n \rightarrow 0$. Then

$$\lim_{n \rightarrow \infty} \mathbb{E} \|f_n^{\alpha_n} - f\|^2 = 0.$$

Proof of Theorem 6. This follows immediately from Corollary 4 and Supplemental Proposition A.1(iv). $\square$

In deriving rates of convergence for ill-posed problems, it is typically assumed that the solution f is drawn from a “source set,” assuming some *a priori* degree of smoothness [Engl et al., 1996, Section 3.2]. In Nair et al. [1997] Theorem 5.1 and Engl and Neubauer [1985] Theorem 3.5, an abstract version of this Tikhonov problem is analyzed, and they find fast $\delta^{4/3}$ rates of convergence (in a stronger norm) compared to the often logarithmic rates in the statistical literature. The price is that strong assumptions are made on the target density. In Engl and Neubauer [1985], it is assumed that $f \in \mathcal{D}(L^*L)$ and $L^*Lf \in \mathcal{R}(T^*T)$. With T the operator that convolves a function with g and L the m th-derivative operator, this assumption requires that $f \in H^{2m}(\mathbb{R})$, and $f^{(2m)} = g \star g^* \psi$ for some $\psi \in L_2(\mathbb{R})$, which we express in terms of the Fourier transforms in the theorem. Nothing is

required of g , since their result holds for any bounded operator T , and convolution with a probability measure is bounded on $L_2(\mathbb{R})$.

Below is an analogue of those abstract results, in an explicitly statistical framework, and with a novel proof. The proof in the present framework turns out to be quite simple.

Theorem 7. (Rates when f is very smooth) Suppose $\int |\omega^{2m} \tilde{f}(\omega)|^2 d\omega < \infty$ and $|\tilde{f}(\omega)| = |\tilde{g}(\omega)|^2 |\omega^{-2m} \tilde{\psi}(\omega)|$ for some $\psi \in L_2(\mathbb{R})$ (Note that this condition implies Assumption (F1)). Then for sufficiently small α_n ,

$$\mathbb{E} \|f_n^{\alpha_n} - f\|^2 \leq C_1 \delta_n^2 / \alpha_n + C_2 \alpha_n^2,$$

and if $\alpha_n = C_3 \delta_n^{\frac{2}{3}}$, then

$$\mathbb{E} \|f_n^{\alpha_n} - f\|^2 = O(\delta_n^{\frac{4}{3}}).$$

Proof of Theorem 7. Our task is to find the dependence of $\|f^\alpha - f\|^2$ on α .

$$\begin{aligned} \|f^\alpha - f\|^2 &= \frac{1}{2\pi} \|\tilde{f}^\alpha - \tilde{f}\|^2 \\ &= \frac{1}{2\pi} \int \left| \frac{\alpha_n \omega^{2m}}{|\tilde{g}(\omega)|^2 + \alpha_n \omega^{2m}} \right|^2 |\tilde{g}(\omega)|^2 |\omega^{-2m} \tilde{\psi}(\omega)|^2 d\omega, \\ &= \frac{\alpha_n^2}{2\pi} \int \left| \frac{|\tilde{g}(\omega)|^2}{|\tilde{g}(\omega)|^2 + \alpha_n \omega^{2m}} \right|^2 |\tilde{\psi}(\omega)|^2 d\omega \leq \alpha_n^2 \|\psi\|^2, \end{aligned}$$

which, combined with Corollary 4, gives the bound. The upper bound is minimized by $\alpha_n \propto \delta_n^{\frac{2}{3}}$, in which case the upper bound becomes $\mathbb{E} \|f_n - f\|^2 \leq C \delta_n^{\frac{4}{3}}$, as needed. $\square$

Remark. Note that Theorem 7 does not require Assumption (G1); identifiability issues are sidestepped by the second assumption on $\tilde{f}$, which guarantees that $\tilde{f}$ is zero whenever $\tilde{g}$ is zero.

Remark. If Z has pdf $\eta(z) \in H^{2m}(\mathbb{R})$, with $\omega^{2m}\tilde{\eta}(\omega) = \tilde{\psi}(\omega)$, and E_1 and E_2 are independent with pdf g , then the hypothesis of Theorem 7 is satisfied for the pdf f of $X = Z + (E_1 - E_2)$.

Corollary 8. Assume the conditions of Theorem 7, and assume that $\alpha_n = C\delta_n^{\frac{2}{3}}$.

If h_n is a kernel density estimate with optimal choice of bandwidth, then

$$\mathbb{E} \|f_n^{\alpha_n} - f\|^2 = O(n^{-8/15}).$$

If h_n is a histogram with optimal choice of bin widths, then

$$\mathbb{E} \|f_n^{\alpha_n} - f\|^2 = O(n^{-4/9}).$$

Proof of Corollary 8. If h_n is a kernel density estimate with optimal bandwidth, then $\delta_n^2 = \|h_n - h\|^2 = O(n^{-\frac{4}{5}})$ (Wand and Jones [1994], Section 2.5), and the result follows immediately. Similarly, if h_n is a histogram with optimal binwidth, then by the same section, $\delta_n^2 = \|h_n - h\|^2 = O(n^{-\frac{2}{3}})$. $\square$

The rates in Theorem 7 are appealing, but are found under conditions different than those typically assumed in the literature. Now, we will assume a particular form for g —either Gaussian, Cauchy, or Laplace—and leverage the approximation bounds for the α -smoothed f from Lemma 5 to derive rates of convergence under a weaker smoothness assumption on f , namely Assumption (F2) that $\int |\omega^k \tilde{f}(\omega)|^2 d\omega < \infty$. This is a slight weakening of the assumption in Zhang [1990].

Theorem 9. Assume (F2). Then,

(i) (Normal errors) If $g(x) = (2\pi)^{-1} e^{-x^2/2}$, then for α small enough,

$$\mathbb{E} \|f_n^\alpha - f\|^2 \leq C_1 \delta_n^2 / \alpha + \frac{C_2}{m^k \log(m^{-1} \alpha^{-\frac{1}{m}})^k}$$

and if $\alpha_n = \delta_n^2 W(\delta_n^{-\frac{2}{k}})^k$ then $\mathbb{E} \|f_n^{\alpha_n} - f\|^2 = O([\log \delta_n^{-1}]^{-k})$.

(ii) (Cauchy errors) If $g(x) = \frac{1}{\pi(1+x^2)}$, then for α small enough,

$$\mathbb{E} \|f_n^\alpha - f\|^2 \leq C_1 \delta_n^2 / \alpha + \frac{C}{m^{2k} \log(m^{-1} \alpha^{-\frac{1}{2m}})^{2k}}$$

and if $\alpha_n = \delta_n^2 W(\delta_n^{-\frac{1}{k}})^{2k}$, then $\mathbb{E} \|f_n^\alpha - f\|^2 = O([\log \delta_n^{-1}]^{-2k})$.

(iii) (Laplace errors) If $g(x) = \frac{1}{2} e^{-|x|}$, then for α small enough,

$$\mathbb{E} \|f_n^\alpha - f\|^2 \leq C_1 \delta_n^2 \alpha^{-\frac{2}{m+2}} + C_2 \alpha^{\frac{k}{m+2}},$$

and if $\alpha_n = \delta_n^{\frac{2(m+2)}{k+2}}$, then $\mathbb{E} \|f_n^\alpha - f\|^2 = O(\delta_n^{\frac{2k}{k+2}})$.

Proof deferred to Supplemental Proof B.9.

Corollary 10. Let $k = 1$. Then if h_n is a KDE or histogram estimate with optimal bandwidth or bin choice, we have, assuming the conditions of Theorem 9 hold,

(i) (Normal errors) $\mathbb{E} \|f_n^\alpha - f\|^2 = O([\log n]^{-1})$ for KDE and histogram.

(ii) (Cauchy errors) $\mathbb{E} \|f_n^\alpha - f\|^2 = O([\log n]^{-2})$ for KDE and histogram.

(iii) (Laplace errors) $\mathbb{E} \|f_n^\alpha - f\|^2 = O(n^{-\frac{4}{15}})$ for the KDE and $\mathbb{E} \|f_n^\alpha - f\|^2 = O(n^{-\frac{2}{9}})$ for the histogram.

For normal and Cauchy errors, Corollary 10 shows that the smoothness-penalized deconvolution estimate attains the optimal rates derived in Zhang [1990]. However, for Laplace errors, the upper bound here is slower than the rate $n^{-2/7}$ attained by the deconvoluting kernel density estimator in Zhang [1990]. However, the SPeD estimator *can* attain the $n^{-2/7}$ rate for a certain choice of estimator h_n . Recall (cf. Equation (1)) that the DKE can be thought of as involving a kernel estimate of h using a kernel $K(\cdot)$ which has quickly decaying Fourier transform; in Zhang [1990] the kernel is required to be band-limited. If we use as our h_n a kernel estimator

satisfying the conditions in Zhang [1990], then the SPeD estimator attains the $n^{-2/7}$ rate.

Proposition 11. Assume g is Laplace, and suppose $\int |\omega \tilde{f}(\omega)|^2 d\omega = C < \infty$. Let $k(\cdot)$ be a pdf satisfying $k(x) = k(-x)$, $\int x^2 k(x) dx < \infty$, $\int |xk'(x)| dx < \infty$, and $\tilde{k}(\omega) = 0$ for $\omega \notin [-1, 1]$. Suppose that h_n is a kernel density estimate with kernel k ,

i.e.
$$h_n(y) = \frac{1}{n\lambda} \sum_{j=1}^n k\left(\frac{y - Y_j}{\lambda}\right)$$
. Suppose $\lambda_n = c_0 n^{-\frac{1}{7}}$, and $\alpha_n = O(n^{-\frac{2(m+2)}{7}})$. Then

$$\mathbb{E} \|f_n^\alpha - f\|^2 = O(n^{-\frac{2}{7}}).$$

Proof deferred to Supplemental Proof B.11.

The following examples show that there is a kind of critical variance or width imposed by the conditions of Theorem 7, at least for a subclass of densities: if $E \sim N(0, \sigma^2)$, then a normal target density f with variance $2\sigma^2 + \varepsilon$ satisfies the conditions of Theorem 7, but a normal target density with variance $2\sigma^2 - \varepsilon$ does not. In contrast, notice that if $f(\cdot)$ satisfies the conditions of Theorem 9, then a re-scaling $f_\sigma(\cdot) = \frac{1}{\sigma} f(\frac{\cdot}{\sigma})$ satisfies them as well (possibly with a different constant for the rate).

Example 1. Suppose $E \sim N(0, \sigma^2)$, $X \sim N(0, 2\sigma^2 + \varepsilon)$, with $\varepsilon > 0$. Then the pdf f of X satisfies the conditions of both Theorem 9 and Theorem 7. For the former, it suffices

to note that $f \in H^k(\mathbb{R})$ for any $k \geq 0$. For the latter, letting
$$\nu(x) = \frac{1}{\sqrt{2\pi\varepsilon}} e^{-x^2/2\varepsilon}$$
, we can take $\psi(x) = (-1)^m \nu^{(2m)}(x)$.

Example 2. Now take $E \sim N(0, \sigma^2)$, but $X \sim N(0, 2\sigma^2 - \varepsilon)$, with $0 < \varepsilon < 2\sigma^2$. Then the pdf f of X satisfies the conditions of Theorem 9, but not Theorem 7. The former holds for the same reason as before. To see why the conditions for Theorem 7 cannot hold, suppose that there was a ψ s.t. $L^*Lf = T^*T\psi$. Then we would have $(-1)^m f^{(2m)} = g \star g^* \psi$, and taking Fourier transforms yields

$(-1)^m (i\omega)^{2m} e^{-(2\sigma^2 - \varepsilon)\omega^2/2} = e^{-\sigma^2\omega^2} \tilde{\psi}(\omega)$, so that $\tilde{\psi}(\omega) = (-1)^m (i\omega)^{2m} e^{\varepsilon\omega^2/2}$. But then $|\tilde{\psi}(\omega)|^2 \rightarrow \infty$ as $\omega \rightarrow \infty$, so $\psi \notin L_2(\mathbb{R})$.

5.2 Constrained Solution

We may wish to incorporate *a priori* knowledge about f into our estimate. Suppose we know that $f \in \mathcal{B}$, a closed, convex set. One easy-to-manage approach is to first solve the unconstrained problem and find an estimate f_n^α not necessarily belonging to $\mathcal{B}$, and then somehow project this unconstrained estimate onto $\mathcal{B}$. Define the projection operator $P_{\mathcal{B}}$ onto a closed, convex set $\mathcal{B} \subset L_2(\mathbb{R})$ by

$$P_{\mathcal{B}}u = \argmin_{v \in \mathcal{B}} \|u - v\|. \quad (8)$$

In words, $P_{\mathcal{B}}$ maps u to the $L_2(\mathbb{R})$ -nearest element of $\mathcal{B}$. The projection operator onto a closed convex set is non-expansive (Engl et al. [1996], Section 5.4), meaning that for all $u, v \in L_2(\mathbb{R})$, $\|P_{\mathcal{B}}u - P_{\mathcal{B}}v\| \leq \|u - v\|$. An immediate consequence is that if $f \in \mathcal{B}$, then projecting f_n^α to $\mathcal{B}$ has error at least as small as f_n^α . Remembering that $P_{\mathcal{B}}f = f$, we have

$$\|P_{\mathcal{B}}f_n^\alpha - f\| = \|P_{\mathcal{B}}f_n^\alpha - P_{\mathcal{B}}f\| \leq \|f_n^\alpha - f\|$$

Now, we know *a priori* that f is a probability density function, so we ought to ensure that our estimate is a probability density function as well. Consider the set

$\mathcal{C} = \{v \in L_2(\mathbb{R}) : \int v(t)dt = 1, v(t) \geq 0 \forall t \in \mathbb{R}\}$; this is the set of square-integrable probability density functions, and now we can express this requirement as $f_n^\alpha \in \mathcal{C}$.

Unfortunately, while $\mathcal{C}$ is convex, it is not closed. To see this, note that the zero

function is a limit point of $\mathcal{C}$: let $\psi_n = \frac{1}{n} 1_{[0,n]}$, and note that $\|\psi_n - 0\| = \|\psi_n\| = n^{-\frac{1}{2}} \rightarrow 0$

as $n \rightarrow \infty$. Indeed, any non-negative function v with $\int v < 1$ is a limit point of $\mathcal{C}$. Thus the minimum in Equation (23) may not be attained, and the projection operator $P_{\mathcal{C}}$ is not well-defined. Instead, we can work with approximations to $\mathcal{C}$. Let

$\mathcal{C}_a = \{v \in \mathcal{C}, v(t) = 0 \forall t \notin [-a, a]\}$ be the subset of $\mathcal{C}$ of functions with support contained in $[-a, a]$.

Lemma 12. *For fixed a , the set $\mathcal{C}_a$ is closed and convex.*

Proof deferred to Supplemental Proof B.12.

Let the unconstrained estimator f_n^α projected to $\mathcal{C}_a$ be denoted $\overset{\circ}{f}_n^\alpha = P_{\mathcal{C}_a} f_n^\alpha$. The non-expansiveness of the projection suggests that $\overset{\circ}{f}_n^\alpha$ may inherit the asymptotics of f_n^α . If $f \in \mathcal{C}_a$ for some a and $a \rightarrow \infty$, then this is immediate from the earlier argument. If $f \notin \mathcal{C}_a$ for all a , we need to do a little more work, and for that we will need to know the size of $\|P_{\mathcal{C}_a} f - f\|$ in terms of a .

Lemma 13. *Assume (F1), $\mathbb{E}[|X|^\beta] < \infty$ and that $f(t) = o(1)$ as $|t| \rightarrow \infty$. Then for large enough a , $\|P_{\mathcal{C}_a} f - f\| \leq 2\mathbb{E}[|X|^\beta]a^{-\beta}$. If also $\mathbb{E}[e^{\beta|X|}] < \infty$, then $\|P_{\mathcal{C}_a} f - f\| \leq 2\mathbb{E}[e^{\beta|X|}]e^{-\beta a}$.*

Proof deferred to Supplemental Proof B.13.

With this in hand, we can say that our constrained estimator will be as good (in an asymptotic sense) as the unconstrained estimator, as long as we let a grow fast enough that the first term dominates:

Lemma 14. *Assume $\mathbb{E}[|X|^\beta] < \infty$ and that $f(t) = o(1)$ as $|t| \rightarrow \infty$. Then*

$$\|\overset{\circ}{f}_n^\alpha - f\| \leq \|f_n^\alpha - f\| + Ca^{-\beta}.$$

$$\text{If } \mathbb{E}[e^{\beta|X|}] < \infty, \text{ then } \|\overset{\circ}{f}_n^\alpha - f\| \leq \|f_n^\alpha - f\| + Ce^{-\beta a}.$$

Proof of Lemma 14. Add and subtract $P_{\mathcal{C}_a} f$:

$$\begin{aligned} \|\overset{\circ}{f}_n^\alpha - f\| &= \|P_{\mathcal{C}_a} f_n^\alpha - P_{\mathcal{C}_a} f + P_{\mathcal{C}_a} f - f\| \\ &\leq \|P_{\mathcal{C}_a} f_n^\alpha - P_{\mathcal{C}_a} f\| + \|P_{\mathcal{C}_a} f - f\| \leq \|f_n^\alpha - f\| + 2\mathbb{E}[|X|^\beta]a^{-\beta}. \end{aligned}$$

A similar approach yields the exponential version. $\square$

6 The Estimator in Practice

In this section we deal with using the estimator in practice, and compare its performance to the deconvoluting kernel density estimator in finite samples.

6.1 Computing the Estimate

The forms of f_n^α in Theorem 2(i)-(iii) are useful, but not the most practical for work on the computer; we need a convenient way to project our estimate to the set of pdfs as described in Section 5.2, and to impose other shape constraints as desired. Instead, we compute the estimate in Equation (6) out of an approximation space $X_n \subset H^m(\mathbb{R})$ of splines of degree $r > m$. We will find that this turns out to be a quadratic program, so that linear constraints are easily imposed.

Before discussing the details of the computations, we present a Theorem showing that this is a legitimate approximation to make. If we denote the spline approximation by s_n^α , Theorem 15 says that if the parameters of the spline space are selected appropriately, then $s_n^\alpha - f_n^\alpha$ is of a smaller order than the rate of convergence we found in Theorem 9. This means that asymptotically, s_n^α and f_n^α are the same estimator. As a consequence, the spline approximation s_n^α attains the same rate of convergence as the exact estimator f_n^α .

Theorem 15. *Suppose X_n is the space of r th-order splines, $r > m$, with uniform knot spacing γ on $[a, b]$ and uniform knot spacing γ^* on $[a - \gamma^*m, a]$ and $[b, b + \gamma^*m]$, with the condition that for all $s \in X_n$, and for $0 \leq k \leq m-1$, we have $s^{(k)}(a - \gamma^*m) = s^{(k)}(b + \gamma^*m) = 0$. Take our spline estimate to be*

$$s_n^\alpha = \operatorname{argmin}_{s \in X_n} \|g^*s - h_n\|^2 + \alpha \|s^{(2)}\|^2. \text{ Suppose also that } \mu_{\hat{\gamma}} = \mathbb{E} \int_{-\infty}^{\infty} |x| h_n(x) dx = O(1).$$

Adopt the assumptions of either Theorem 7 or Theorem 9, and let $r_n = \mathbb{E} \|f_n^\alpha - f\|^2$ denote the resulting rate of convergence of the exact estimator. Choose γ, γ^, α , and*

b so that $\alpha^{-4}\gamma^{2(r-m)} = o(r_n)$, $\alpha^{-4}\gamma^*(1+\gamma) = o(r_n)$, and $\alpha^{-4}(|a| \wedge |b|)^{-1} = o(r_n)$. Then $\mathbb{E} \|s_n^\alpha - f_n^\alpha\|^2 = o(r_n)$. It follows also that $\mathbb{E} \|s_n^\alpha - f\|^2 = O(r_n)$.

Proof deferred to Supplemental Proof C.15.

Now we describe how we compute s_n^α in concrete terms. In all of the following, unless otherwise stated, we fix $m = 2$. Fix $r = 3$, and let $\mathfrak{S}_q = \mathfrak{S}_{(r=3, \xi_1, \dots, \xi_{q+4})}$ denote the space of cubic splines (cf. [Powell, 1981, Chapter 3]) with knots $\xi_1, \dots, \xi_{q+4}$, with $\xi_1 < Y_{(1)}$ and $\xi_{q+4} > Y_{(n)}$, evenly spaced knots, no knots of multiplicity larger than one, and end conditions $s^{(k)}(\xi_1) = s^{(k)}(\xi_{q+4}) = 0$ for $k = 0, 1, 2$. The end conditions specify that members of $\mathfrak{S}_q$ vanish outside the interval $[\xi_1, \xi_{q+4}]$ and are twice continuously-differentiable at the boundary. This space $\mathfrak{S}_q$ has as a basis the collection of q unit-integral B-splines $\{b_i\}_{i=1}^q$, so that if $s \in \mathfrak{S}_q$, then $s(x) = \sum_{i=1}^q \theta_i b_i(x)$. Note that $\mathfrak{S}_q \subset H^m(\mathbb{R})$.

We now take as our estimate $s_n^\alpha = \operatorname{argmin}_{s \in \mathfrak{S}_q} \|g * s - h_n\|^2 + \alpha \|s^{(2)}\|^2$. If $s(x) = \sum_{i=1}^q \theta_i b_i(x)$, then $\|g * s - h_n\|^2 + \alpha \|s^{(2)}\|^2 = \theta^T \mathbf{M} \theta - 2\theta^T \mathbf{d} + \|h_n\|^2 + \alpha \theta^T \mathbf{P} \theta$, where θ is the vector of coefficients θ_i , and $\mathbf{M}$, $\mathbf{d}$, and $\mathbf{P}$ are a $q \times q$ matrix, $q \times 1$ vector, and $q \times q$ matrix respectively, with typical entries $M_{ij} = \int (g * b_i)(g * b_j)$, $d_i = \int (g * b_i) h_n$, and $P_{ij} = \int b_i^{(2)} b_j^{(2)}$.

With this matrix representation, we can see, using standard techniques, and noting that $\|h_n\|^2$ does not depend on s_n^α , that the coefficients of s_n^α are $\theta_n^\alpha = (\mathbf{M} + \alpha \mathbf{P})^{-1} \mathbf{d}$,

so that $s_n^\alpha(x) = \sum_{i=1}^q \theta_{n,i}^\alpha b_i(x)$. Analogous to the exact solution, s_n^α need not be a pdf. To produce a pdf, we now solve

$$s_n^\alpha = \operatorname{argmin}_{\substack{s=1 \\ s \geq 0}} \|s - s_n^\alpha\|^2, \quad (9)$$

At this stage, other linear constraints may be introduced by expressing them against the B-spline basis. If $\mathbf{G}$ is a matrix with typical entry $G_{ij} = \int b_i b_j$, and, letting $\xi_1 = x_1 < x_2, \dots, x_{n_x} = \xi_{q+4}$ be a grid of evenly spaced values on the support of ϕ^q , with $\mathbf{B}_x$ the $n_x \times q$ matrix with i, j th entry $b_j(x_i)$, the coefficients of the solution to Equation (9) are given (approximately) by the linearly-constrained quadratic program

$$\theta_n^\alpha = \underset{\substack{\theta \in \mathbb{R}^q \\ \mathbf{B}_x \theta \geq 0}}{\operatorname{argmin}} (\theta - \theta_n^\alpha)^T \mathbf{G} (\theta - \theta_n^\alpha). \quad (10)$$

The reason this is approximate is that the convex constraint $s(x) \geq 0$ for all x is

approximated by the collection of linear constraints $s(x_i) = \sum_{j=1}^q \theta_j b_j(x_i) \geq 0$, $i = 1, \dots, n_x$.

Equation (10) is a quadratic program with q -dimensional objective and $n_x + 1$ linear constraints.

The entries of $\mathbf{G}$ and $\mathbf{P}$ can be computed by hand from the piecewise-polynomial representation of the B-splines. Computing the entries of $\mathbf{M}$ and $\mathbf{d}$ benefits from the Fourier representation

$$M_{ij} = \frac{1}{2\pi} \int \tilde{b}_i(\omega) \overline{\tilde{b}_j(\omega)} |\tilde{g}(\omega)|^2 d\omega \quad \text{and} \quad d_i = \frac{1}{2\pi} \int \tilde{g}(\omega) \tilde{b}_i(\omega) \overline{\tilde{h}_n(\omega)} d\omega,$$

which can then be computed by an appropriate quadrature, bypassing the problem of dealing with the convolutions. When h_n is a kernel density estimate or a histogram, $\tilde{h}_n$ is not difficult to compute, and the $\tilde{b}_i$ are straightforward to compute, as B-spline basis functions can be represented as shifted, scaled self-convolutions of $1_{[0,1]}$.

6.2 Finite Sample Behavior

In Wand [1998], the author points out that while asymptotic rates for deconvolution are very slow no matter the size of the measurement error (cf. Theorem 9 here, Stefanski [1990], Zhang [1990], Fan [1991]), there is another side of the coin: for very small measurement error we ought to expect to be able to estimate f with MISE quite close to that of the error-free setting. For example, we could simply ignore

measurement error and increase our MISE by at most $\|g * f - f\|^2$, which becomes arbitrarily small as the measurement error decreases. Thus, we might expect that the pessimistic picture given by asymptotic rates is limited to truly large samples, especially when measurement error is small, and a direct investigation into small-sample behavior is required for a better understanding of deconvolution estimators.

To get a handle on the small-sample behavior, Wand [1998] creates two products for the deconvoluting kernel estimator: a log-log plot of the minimum attainable MISE, i.e. $\inf_{\lambda>0} \mathbb{E} \|f_n^\lambda - f\|^2$, against the sample size, as well as a table listing the smallest sample size required for the minimum attainable MISE in deconvolution to be at least as small as the minimum attainable MISE in the no-measurement-error case with some fixed sample size.

We will investigate these same properties for analogous quantity, the minimum attainable MISE for the SPeD estimator, given by $\inf_{\alpha>0} \mathbb{E} \|f_n^\alpha - f\|^2$. Supplemental Figure B.2 shows a plot of $\mathbb{E} \|f_n^\alpha - f\|^2$ as a function of α . Since the MISE involves unknown quantities, in practice α will have to be chosen from the data, and the search for a good data-driven choice of α is ongoing; in Section 6.3, we use what is essentially an iterated bootstrap, but at this point do not claim that it is optimal.

The settings addressed in Wand [1998], which we will use here as well, are as follows. The target random variable X has one of the following densities; (i) standard

normal, (ii) normal mixture $\frac{2}{3} N(0, \sigma=1) + \frac{1}{3} N(0, \sigma=\frac{1}{5})$, (iii) Gamma($\zeta=4, \beta=1$), (iv) gamma mixture $\frac{2}{5} \text{Gamma}(\zeta=5, \beta=1) + \frac{3}{5} \text{Gamma}(\zeta=13, \beta=1)$, with ζ and β the shape and rate parameters, respectively. We will consider normal measurement error E with $\text{Var}(E) = p \cdot \text{Var}(Y)$, with various choices of p .

To investigate these properties for the smoothness-penalized deconvolution estimator, we must compute the MISE of our estimator, $\text{MISE}(f_n^\alpha) = \mathbb{E} \int (f_n^\alpha - f)^2$. From Theorem 2(i), we have that the Fourier transform of the estimate is given by $\tilde{\varphi}_\alpha(\omega) \tilde{h}_n(\omega)$; to simplify calculations, we approximate $\tilde{h}_n$ by the Fourier transform of

the empirical distribution, $\tilde{P}_n(\omega) = \frac{1}{n} \sum_{j=1}^n e^{-i\omega Y_j}$, using instead

$$\tilde{f}_n^\alpha(\omega) = \tilde{\varphi}_\alpha(\omega) \tilde{P}_n(\omega) = \sum_{j=1}^n \tilde{\varphi}_\alpha(\omega) e^{-i\omega Y_j}.$$

Even though we have replaced the density

estimate h_n by the empirical distribution, which has no density at all, the approximation is quite good; see Figure B.2 in the supplemental material. The resulting MISE, derived in Supplemental Fact A.3, is

$$\text{MISE}(\alpha) = \frac{1}{2\pi} \left[\int |\tilde{\varphi}_\alpha(\omega) \tilde{g}(\omega) - 1|^2 |\tilde{f}(\omega)|^2 d\omega + \frac{1}{n} \int |\tilde{\varphi}_\alpha(\omega)|^2 (1 - |\tilde{g}(\omega) \tilde{f}(\omega)|^2) d\omega \right]$$

which we will evaluate numerically in the following.

In Figure 2, we show plots of best-attainable MISE, i.e. $\inf_{\alpha>0} \mathbb{E} \|\tilde{f}_n^\alpha - f\|^2$, for the SPeD (computed via Equation (11)), and the same, but with infimum over the bandwidth, for the DKE and a conventional kernel estimator on the non-contaminated X 's for reference (both computed via the expressions in Wand [1998]). For the deconvoluting kernel density estimate, we use a base kernel K_{DKE} with Fourier transform $\kappa_{\text{DKE}}(\omega) = 1_{|\omega|<1} (1 - \omega^2)^3$; this is κ_1 in Wand [1998], and is the default choice in the `deconvolve` R package Delaigle et al.. For the error-free kernel estimator, we use kernel with Fourier transform $\kappa_{\text{ef}}(\omega) = (1 + \omega^4)^{-1}$. This relates to the smoothness-penalized deconvolution estimator in the following sense: the error-free setting is equivalent to the measurement error problem where E is a point-mass at zero. In that case, $\tilde{g}(\omega) = 1$, and then $\tilde{\varphi}_\alpha(\omega) = (1 + \alpha\omega^{2m})^{-1}$. If we replace $\tilde{h}$ by $\tilde{P}_n$ again in

Theorem 2(iv), we have a kernel estimator with $K_{\text{ef}}(x) = \frac{1}{2\pi} \int e^{i\omega x} \kappa_{\text{ef}}(\omega) d\omega$. Note that $\int K_{\text{ef}} = 1$, but K_{ef} is not non-negative. In fact, when $m = 2$, K_{ef} is a fourth-order kernel.

In Figure 2, the smoothness-penalized deconvolution estimator gives a much more optimistic picture of the deconvolution problem in finite samples compared to the deconvoluting kernel estimator. The SPeD has nearly uniformly lower MISE, excepting a small range of n in setting (iv). In setting (i), which satisfies the

conditions of Theorem 7, the SPeD under 30% measurement error has better optimal MISE than the DKE under 10% measurement error, for sample sizes small enough to be commonly encountered in practice.

Table 1 lists sample sizes required for the deconvolution estimators to attain MISE as small as the error-free setting. We can see that in every case listed in the table, the SPeD requires fewer samples than the DKE; in some cases the difference is dramatic. To achieve the same MISE as estimating the Gamma mixture density in setting (iv) in the error-free setting with a sample of size $n = 1,000$ when there is 10% measurement error, the SPeD would require 7,963 samples, while the DKE would require 388,770 samples. In practice, this may mean the difference between an expensive experiment and an impossible one. Another takeaway is how strongly the required n varies with the target density. In setting (i), the problem does not seem so bad; in setting (ii), it seems all but impossible.

6.3 Application to Cytotoxicity Data

Bacillus cereus sensu lato (s.l.) is a group of closely-related bacteria with diverse relationships to humans, including *B. thuringiensis*, which is used on crops as a pesticide, *B. anthracis*, which can cause anthrax disease, and others which can cause other illness and spoil food Ceuppens et al. [2013]. These bacteria are ubiquitous in many environments, their taxonomy is “complex and equivocal,” Ceuppens et al. [2013], and distinguishing between members of *B. cereus s.l.* with typical methods can be difficult. Scientists are therefore interested in developing practical laboratory tests which can readily discriminate between harmful representatives of this group and those less likely to cause harm.

As one element of that investigation, a colleague requires a density estimate of a certain conditional expectation. Suppose i is an isolate of *B. cereus s.l.*, sampled from a large collection. Suppose it is cultured under certain conditions, centrifuged, and the supernatant is applied to human cells. Let X_i denote the mean normalized cytotoxicity of isolate i , and $C_{ij} = X_i + \varepsilon_{ij}$ denote the cytotoxicity observed the j th time this procedure is applied to isolate i and further assume that the ε_{ij} , are i.i.d., have mean zero and are independent of X_i . We are interested in the density f of X_i as i

varies over the collection of isolates. However, the investigator only has access to a

sample approximation $Y_i = \frac{1}{k} \sum_{j=1}^k C_{ij}$ of X_i obtained by fixing i and repeatedly

measuring the cytotoxicity. With $E_i = \frac{1}{k} \sum_{j=1}^k \varepsilon_{ij}$, we are in the setting described in the introduction. We do not know the density of g of E_i exactly, as assumed for the theory; however, we may approximate it by $N(0, \sigma_\varepsilon^2 / k)$ as long as the ε_{ij} are not too skewed. We then only need to estimate σ_ε^2 , which can be done at parametric rates much faster than the rates involved in deconvolution.

We have been provided preliminary data, which comprise a table of measured cytotoxicity C_{ij} from $j = 1, \dots, k = 6$ replicates of isolates $i = 1, \dots, n = 313$. We have estimated σ_ε^2 by fitting the linear model $C_{ij} = X_i + \varepsilon_{ij}$ in R and extracting the residual standard error. Tuning parameter α was chosen by picking an arbitrary provisional α_0 , seeking α_i which minimizes $\mathbb{E} \|s_n^\alpha - s_n^{\alpha_{i-1}}\|^2$ assuming the X_i have pdf $s_n^{\alpha_{i-1}}$, and iterating until convergence. The results are shown in Figure 3, along with a standard kernel density estimate of the Y_i . This example has a relatively small amount of measurement error, with proportion $p = \text{Var}(E) / \text{Var}(Y) \approx 0.044$. To illustrate SPeD with greater measurement error and to see if the number of replicates may be reduced in future experiments, we have also split the replicates randomly into two groups (i) and (ii), and re-fit the estimator as if there were only three available replicates. This yields $p \approx 0.082$ and $p \approx 0.086$ for groups (i) and (ii), respectively. The two modes present in the full data are blurred to one mode in the reduced data, but our estimator does recover two modes in one of the two reduced data settings.

7 Acknowledgements

The authors thank Professor Jasna Kovac for sharing with us the *B. cereus* cytotoxicity data, and Professor Kengo Kato for several helpful conversations.

References

Siele Ceuppens, Nico Boon, and Mieke Uyttendaele. Diversity of *Bacillus cereus* group strains is reflected in their broad range of pathogenicity and diverse ecological lifestyles. *FEMS Microbiology Ecology*, 84(3):433–450, June 2013. ISSN 0168-6496. doi: 10.1111/1574-6941.12110.

Aurore Delaigle, Timothy Hyndman, and Tianying Wang. *deconvolve: Deconvolution Tools for Measurement Error Problems*.

Heinz W. Engl and Andreas Neubauer. Optimal Discrepancy Principles for the Tikhonov Regularization of Integral Equations of the First Kind. In G. Hämmerlin and K.-H. Hoffmann, editors, *Constructive Methods for the Practical Treatment of Integral Equations: Proceedings of the Conference Mathematisches Forschungsinstitut Oberwolfach, June 24–30, 1984*, International Series of Numerical Mathematics, pages 120–141. Birkhäuser, Basel, 1985. ISBN 978-3-0348-9317-6. doi: 10.1007/978-3-0348-9317-6\10.

Heinz W Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*. Kluwer Academic Publishers, Dordrecht, 1996. ISBN 978-0-7923-4157-4.

Jianqing Fan. On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems. *The Annals of Statistics*, 19(3):1257–1272, September 1991. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1176348248.

G. B Folland. *Fourier analysis and its applications*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, Calif., 1992. ISBN 978-0-534-17094-3.

G. B Folland. *Real analysis: modern techniques and their applications*. Wiley, New York, 2nd ed. edition, 1999. ISBN 978-0-471-31716-6.

John Locker and P.M Prenter. Regularization with differential operators. I. General theory. *Journal of Mathematical Analysis and Applications*, 74(2):504–529, April 1980. ISSN 0022247X. doi: 10.1016/0022-247X(80)90145-6.

M. Thamban Nair, Markus Hegland, and Robert S. Anderssen. The trade-off between regularity and stability in Tikhonov regularization. *Mathematics of Computation*, 66(217):193–207, January 1997. ISSN 0025-5718. doi: 10.1090/S0025-5718-97-00811-9.

Marianna Pensky and Brani Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, 27(6):2033–2053, December 1999. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1017939249.

M. J. D Powell. *Approximation theory and methods*. Cambridge University Press, Cambridge [Eng.], 1981. ISBN 978-0-521-22472-7.

Abhra Sarkar, Bani K. Mallick, John Staudenmayer, Debdeep Pati, and Raymond J. Carroll. Bayesian Semiparametric Density Deconvolution in the Presence of Conditionally Heteroscedastic Measurement Errors. *Journal of Computational and Graphical Statistics*, 23(4):1101–1125, October 2014. ISSN 1061-8600. doi: 10.1080/10618600.2014.899237.

John Staudenmayer, David Ruppert, and John P Buonaccorsi. Density Estimation in the Presence of Heteroscedastic Measurement Error. *Journal of the American Statistical Association*, 103 (482):726–736, June 2008. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214508000000328.

Leonard Stefanski and Raymond J. Carroll. Deconvoluting Kernel Density Estimators. *Statistics*, 21(2):169–184, January 1990. ISSN 0233-1888. doi: 10.1080/02331889008802238.

Leonard A. Stefanski. Rates of convergence of some estimators in a class of deconvolution problems. *Statistics & Probability Letters*, 9(3):229–235, March 1990. ISSN 0167-7152. doi: 10.1016/0167-7152(90)90061-B.

M. P. Wand. Finite sample performance of deconvolving density estimators. *Statistics & Probability Letters*, 37(2):131–139, 1998.

M.P. Wand and M.C. Jones. *Kernel Smoothing*. CRC Press, Boca Raton, FL, first edition. edition, 1994. ISBN 978-0-429-17059-1.

Ran Yang, Daniel W. Apley, Jeremy Staum, and David Ruppert. Density Deconvolution With Additive Measurement Errors Using Quadratic Programming. *Journal of Computational and Graphical Statistics*, 29(3):580–591, July 2020. ISSN 1061-8600. doi: 10.1080/10618600.2019.1704294.

Ran Yang, David Kent, Daniel W. Apley, Jeremy Staum, and David Ruppert. Bias-corrected Estimation of the Density of a Conditional Expectation in Nested Simulation Problems. *ACM Transactions on Modeling and Computer Simulation*, 31(4):22:1–22:36, July 2021. ISSN 1049-3301. doi: 10.1145/3462201.

Cun-Hui Zhang. Fourier Methods for Estimating Mixing Densities and Distributions. *The Annals of Statistics*, 18(2):806–831, 1990. ISSN 0090-5364.

Accepted Manuscript

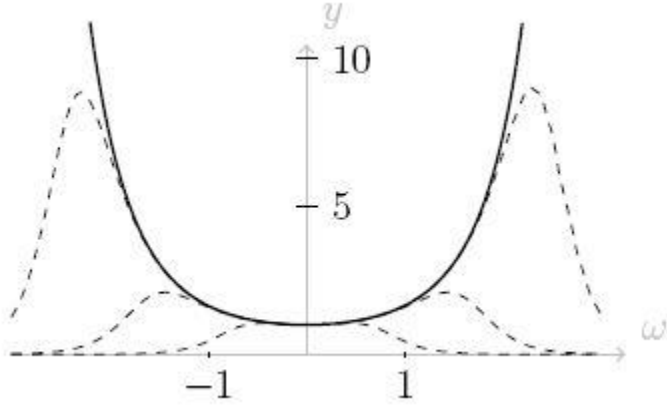


Fig. 1 For $\tilde{g}$ corresponding to $N(0,1)$. Thick line is $1/\tilde{g}$, while dashed lines are, from lower to upper, the multiplier $\tilde{\varphi}_\alpha$ in Theorem 2(i) for $\alpha = 1, 10^{-2}, 10^{-4}$.

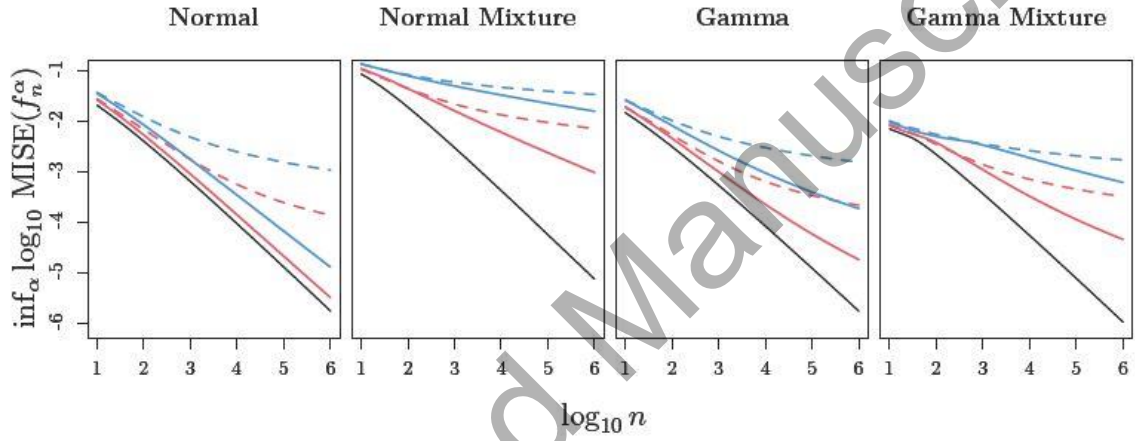


Fig. 2 MISE all under oracle choice of tuning parameter, densities (i)-(iv), left-to-right. Solid black is MISE for kernel estimator in the error-free setting. Solid lines are SPeD, and dashed lines are DKE. Red and blue lines have $p = 0.1, 0.3$, where $\text{Var}(E) = p \cdot \text{Var}(Y)$.

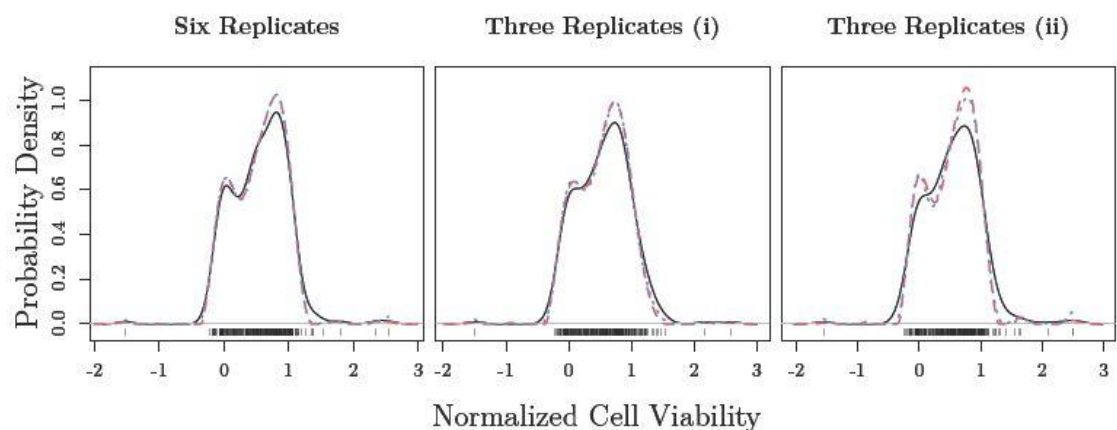


Fig. 3 Density estimates of cytotoxicity data described in Section 6.3. Standard Gaussian kernel density estimate of the Y_i as solid black line. Smoothness-penalized density estimate of the X_i as dashed red line; QP estimator as dotted blue line. Individual data locations marked below plot. Leftmost panel is full data; right two panels each use only three of the available six replicates for each measurement.

Table 1 Minimum sample sizes for stated estimator, with ρ the proportion of measurement error, to achieve MISE as small as error-free kernel density estimation on the X 's with kernel K_{ef} . The analogous value with respect to kernel K_{DKE} is in parentheses.

	Sample size $n = 100$		$n = 1,000$	
ρ	Tikhonov	DKE	Tikhonov	DKE
(i) Standard normal density				
10%	146 (102)	243 (156)	1,525 (1,001)	7,386 (2,931)
30%	303 (204)	1,761 (788)	4,170 (2,239)	$> 10^6$ ($> 10^6$)
50%	1,221 (747)	924,510 (103,089)	34,945 (15,566)	$> 10^6$ ($> 10^6$)
(ii) Normal mixture density				
10%	687 (604)	1,798 (1,415)	56,150 (35,361)	$> 10^6$ ($> 10^6$)
30%	303,719 (211,484)	$> 10^6$ ($> 10^6$)	$> 10^6$ ($> 10^6$)	$> 10^6$ ($> 10^6$)
50%	$> 10^6$ ($> 10^6$)	$> 10^6$ ($> 10^6$)	$> 10^6$ ($> 10^6$)	$> 10^6$ ($> 10^6$)
(iii) Gamma(4) density				
10%	179 (140)	266 (197)	2,548 (1,721)	17,342 (7,863)
30%	695 (499)	8,016 (3,620)	42,254 (21,561)	$> 10^6$ ($> 10^6$)
50%	9,451 (5,551)	$> 10^6$ ($> 10^6$)	$> 10^6$ ($> 10^6$)	$> 10^6$ ($> 10^6$)
(iv) Gamma mixture density				
10%	284 (270)	300 (282)	7,963 (6,020)	388,770 (151,942)
30%	5,522 (4,993)	53,740 (41,039)	$> 10^6$ ($> 10^6$)	$> 10^6$ ($> 10^6$)
50%	$> 10^6$ ($> 10^6$)	$> 10^6$ ($> 10^6$)	$> 10^6$ ($> 10^6$)	$> 10^6$ ($> 10^6$)