

# A NOTE ON A FIXED POINT METHOD FOR DECONVOLUTION

Céline Duval

► **To cite this version:**

Céline Duval. A NOTE ON A FIXED POINT METHOD FOR DECONVOLUTION. Statistics, Taylor & Francis: STM, Behavioural Science and Public Health Titles, 2017, 51 (2), pp.347-362. <10.1080/02331888.2016.1265967>. <hal-01199599v2>

**HAL Id: hal-01199599**

**<https://hal.archives-ouvertes.fr/hal-01199599v2>**

Submitted on 7 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A NOTE ON A FIXED POINT METHOD FOR DECONVOLUTION

C. DUVAL<sup>1</sup>

ABSTRACT. In this paper we study a particular multidimensional deconvolution problem. The distribution of the noise is assumed to be of the form  $G(dx) = (1 - \alpha)\delta(dx) + \alpha g(x)dx$ , where  $\delta$  is the Dirac mass at  $0 \in \mathbb{R}^d$ ,  $g : \mathbb{R}^d \rightarrow [0, \infty)$  is a density and  $\alpha \in [0, \frac{1}{2}[$ . We propose a new minimax estimation procedure, which is not based on a Fourier approach, but on a fixed point method. The performances of the procedure are studied over isotropic Besov balls for  $L_p$  loss functions,  $1 \leq p < \infty$ . A numerical study illustrates the method.

**Keywords.** Density deconvolution. Nonparametric estimation. Wavelets.

**AMS Classification.** 62G07, 62H12, 62G20.

## 1. INTRODUCTION

1.1. **Statistical setting.** Suppose we observe the following multivariate deconvolution model

$$(1) \quad Y_i = X_i + \varepsilon_i, \quad i = 1, \dots, n$$

where  $(X_i)$  are i.i.d. (independent and identically distributed) random variables on  $\mathbb{R}^d$ ,  $d \geq 1$ , with density  $f : \mathbb{R}^d \rightarrow [0, \infty)$  with respect to the Lebesgue measure and independent of  $(\varepsilon_i)$  which are  $d$  dimensional i.i.d. random variables with known distribution  $G$ . We aim at estimating  $f$  by a fixed point method from the indirect observations  $(Y_j)$  and under the following assumption **(H1)**: there exist a positive number  $0 \leq \alpha < 1/2$  and a density function  $g : \mathbb{R}^d \rightarrow [0, \infty)$ , with respect to the Lebesgue measure, such that

$$(H1) \quad G(dx) = (1 - \alpha)\delta(dx) + \alpha g(x)dx$$

where  $\delta$  is the Dirac mass concentrated at 0 and  $dx = dx_1 \dots dx_d$ . Assumption **(H1)** can be understood as follows. If  $\alpha = 0$ , it means that  $Y_i = X_i$  for all  $i$ , the estimation problem is direct. On the contrary,  $\alpha = 1$  corresponds to the classical deconvolution model (see references hereafter). Finally,  $1 > \alpha > 0$  means that  $\alpha$  percents of the dataset are blurred observations of  $X$  and the others are direct measurements, which is the case when observations are recorded with a device that is sometimes subject to measurement errors. However, we do not know from the dataset  $(Y_1, \dots, Y_n)$  which observation is blurred and which is not.

The law of the observations  $(Y_i)$  is  $f \star G$ , where  $\star$  denotes the convolution product. Then, to estimate  $f$  from the indirect observations  $(Y_i)$ , one needs to compute  $\mathbf{P}^{-1}$ , the inverse of the convolution operator:

$$\mathbf{P} : f \longrightarrow \mathbf{P}[f] = f \star G.$$

This model has been extensively studied in the literature assuming the distribution  $G$  to be absolutely continuous with respect to the Lebesgue measure ( $\alpha = 1$ ). Optimal rates of convergence and adaptive procedures are well known if  $d = 1$  (see *e.g.* Carroll and Hall [7], Stefanski [35], Stefanski and Carroll [36], Fan [19], Butucea [2], Butucea and Tsybakov [3, 4], Pensky and Vidakovic [33] or Comte *et al.* [6] for  $L_2$  loss functions or Lounici and Nickl [29] for the  $L_\infty$  loss). Results have also been established for multivariate anisotropic densities (see *e.g.* Comte

---

<sup>1</sup> Université Paris Descartes, MAP5, UMR CNRS 8145.

and Lacour [11] for  $L_2$  loss functions or Rebelles [34] for  $L_p$  loss functions,  $p \in [1, \infty]$ ). Deconvolution with unknown error distribution has also been studied (see *e.g.* Neumann [31], Delaigle *et al.* [12], Johannes [23] or Meister [30], if an additional error sample is available, or Comte and Lacour [10], Delattre *et al.* [13], Johannes and Schwarz [24], Comte and Kappus [9] or Kappus and Mabon [25] under other set of assumptions).

The spirit of all the afore-mentioned procedures is to transport the problem in the Fourier domain where the convolution product becomes a simple product, that can be easily inverted. Let  $\mathcal{F}[G]$  denote the Fourier transform of a distribution  $G$

$$\mathcal{F}[G](u) := \int e^{iuy} G(dy).$$

In the one dimensional setting we have

$$(2) \quad \mathcal{F}[G](u)\mathcal{F}[f](u) = \mathcal{F}[f \star G](u) \quad \forall u \in \mathbb{R}.$$

If the distribution  $G$  is absolutely continuous with respect to the Lebesgue measure ( $\alpha = 1$ ), its Fourier transform vanishes at infinity. Then, at the points where  $\mathcal{F}[G](u)$  gets small,  $\mathcal{F}[f](u)$  is badly recovered. It leads to specific minimax rates of convergence, slower than usual nonparametric rates, depending on how fast  $\mathcal{F}[G]$  goes to 0 at infinity (see *e.g.* Fan [19]).

A strategy based on (2) leads to an estimator of  $\mathcal{F}[f]$ , which needs to be sent back in the space of density functions to get an estimator of  $f$ . Working with a  $L_2$  loss function together with the use of Plancherel equality facilitate the study of the estimator. This methodology may also be adapted to general  $L_p$  loss functions (see *e.g.* Rebelles [34]) using a kernel density estimator with a suitably chosen kernel that takes advantage of the structure of the problem in the Fourier domain.

Under Assumption (H1), the ratio (2) is well defined as  $\mathcal{F}[G](u) = (1-\alpha) + \alpha\mathcal{F}[g](u)$ ,  $\forall u \in \mathbb{R}$ , does not vanish at infinity. Recently, Lepski and Willer [28] considered the same convolution model (1) with assumption (H1). For  $\alpha \in [0, 1]$ , they establish lower bounds for  $L_p$  loss functions,  $p \in [1, \infty]$ , over very general anisotropic Nikol'skii classes. The lower bounds of Lepski and Willer [28] suggests that under Assumption (H1), if  $\alpha < 1$ , there exists an estimator converging at usual nonparametric rate. Indeed standard deconvolution procedures based on a Fourier approach should attain usual nonparametric rates of convergence for  $L_2$  loss functions (see also Section 3.2). However, in their paper, if an outline on how to estimate  $f$  is suggested, no estimator is given. Therefore the upper bounds are not established.

This model is also related to the estimation of the jump density of a discretely observed jump process (see Section 3.3) and to the atomic distribution setting investigated by van Es *et al.* [18], Lee *et al.* [27] and Gugushvili *et al.* [20]. One observes independent realizations of  $X$  where

$$X = UV + Z$$

where  $(U, V, Z)$  are independent,  $V$  has density  $f_V$ ,  $U$  is a Bernoulli with unknown parameter  $1 - p$  and  $Z$  has a known density  $f_Z$ . The afore mentioned papers focus on the estimation of  $(p, f_V)$ . The distribution of  $X$  is  $pf_Z + (1 - p)f_Z \star f_V$ . The analogy with the present setting is that here we suppose that  $f_Z$  is unknown and to be estimated whereas  $(p, f_V)$  are known.

In this paper, we study a methodology based on a fixed point method to non parametrically estimate the density  $f$  for  $L_p$  loss functions,  $1 \leq p < \infty$  over isotropic Besov balls. For the fixed point procedure to work, we add some restrictions on the distribution of the noise  $G$ , namely  $\alpha \in [0, \frac{1}{2}]$ . The interest of the procedure is threefold. Firstly, we solve a particular deconvolution problem without relying on the specificity of the convolution product in the Fourier domain. Secondly, we provide a consistent estimator of the density  $f$ , that is easy to implement numerically. Thirdly, it matches the lower bounds of Lepski and Willer [28].

1.2. **Estimation strategy.** We have i.i.d. observations with distribution

$$f \star G(x) = \int_{\mathbb{R}^d} f(x-y)G(dy),$$

Formally, we introduce the convoluting operator  $\mathbf{P}$

$$(3) \quad \begin{aligned} \mathbf{P} : \mathcal{G}(\mathbb{R}^d) &\longrightarrow \mathcal{G}(\mathbb{R}^d) \\ f &\longrightarrow \mathbf{P}[f] := f \star G = G \star f \end{aligned}$$

where  $\mathcal{G}(\mathbb{R}^d)$  denotes the set of densities that are absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}^d$ . If we knew analytically the inverse  $\mathbf{P}^{-1}$  of  $\mathbf{P}$ , an estimator of  $f$  is

$$\widehat{f} = \mathbf{P}^{-1}[\widehat{\mathbf{P}[f]}]$$

where  $\widehat{\mathbf{P}[f]}$  denotes any estimator of  $f \star G$  build from the direct observations  $(Y_1, \dots, Y_n)$ . Even if we know that  $\mathbf{P}^{-1}$  exists, we do not have its analytic form. Here, we do not compute the inverse  $\mathbf{P}^{-1}$ , but we approximate it by a fixed point method. Consider the mapping  $\mathbf{T}$  (see also Duval [16])

$$(4) \quad \mathbf{T}[h] = \mathbf{P}[f] + h - \mathbf{P}[h], \quad h \in \mathcal{G}(\mathbb{R}^d).$$

We immediately check that  $f$  is a fixed point of  $\mathbf{T}$ . If moreover  $\mathbf{T}$  is contractant and  $f$  belongs to a given Banach space equipped with some norm  $\|\cdot\|_B$ , applying the Banach fixed point theorem we get  $\lim_{K \rightarrow \infty} \|f - \mathbf{T}^{\circ K}[h]\|_B = 0$ , for any density  $h$  and where  $\circ$  denotes the composition product and  $\mathbf{T}^{\circ K} = \mathbf{T} \circ \dots \circ \mathbf{T}$ ,  $K$  times. Given an estimator  $\widehat{\mathbf{P}[f]}$  of  $\mathbf{P}[f]$  an estimator of  $f$  is

$$\widehat{f}_K = \widehat{\mathbf{T}}^{\circ K}[\widehat{\mathbf{P}[f]}]$$

where  $\widehat{\mathbf{T}}$  is such that  $\widehat{\mathbf{T}}[h] = \widehat{\mathbf{P}[f]} + h - \mathbf{P}[h]$ ,  $h \in \mathcal{G}(\mathbb{R}^d)$ , and for some  $K \geq 1$ .

Predictably enough, to make  $\mathbf{T}$  contractant, we need to impose some conditions on  $G$ , that is why Assumption (H1) is introduced, with  $\alpha \in [0, \frac{1}{2}[$ . The introduction of a distribution  $G$  with a mass at 0 appears naturally since the Dirac mass  $\delta$  is the neutral element for the convolution product. This is not enough to make  $\mathbf{T}$  contractant, we need to impose  $0 < \alpha < \frac{1}{2}$  (see Proposition 1 hereafter). It means that the data set contains at least 50% of direct observations.

We investigate the nonparametric estimation of  $f$  on any compact set  $D$  of  $\mathbb{R}^d$ , under Assumption (H1) with  $\alpha < 1/2$ . We use wavelet threshold density estimators and study their performances uniformly over isotropic Besov balls for the following  $L_p$  loss function,  $1 \leq p < \infty$ ,

$$(5) \quad (\mathbb{E}[\|\widehat{f} - f\|_{L_p(D)}^p])^{1/p},$$

where  $\widehat{f}$  is an estimator of  $f$  and  $\|\cdot\|_{L_p(D)}$  denotes  $L_p$  loss over the compact set  $D$ . In the sequel we distinguish the norm  $\|\cdot\|_{L_p(\mathbb{R}^d)}$  from  $\|\cdot\|_{L_p(\nu)}$  as follows

$$\|f\|_{L_p(\mathbb{R}^d)} = \left( \int_{\mathbb{R}^d} |f(x)|^p dx \right)^{1/p} \quad \text{and} \quad \|G\|_{L_p(\nu)} = \left( \int_{\mathbb{R}^d} |G(x)|^p \nu(dx) \right)^{1/p}$$

where  $\nu(dy) = \delta(dy) + dy$  is a dominating measure.

## 2. ESTIMATION OF $f$

2.1. **Preliminary on Besov spaces and wavelet thresholding.** In the sequel we consider wavelet threshold density estimators and study their performance uniformly over isotropic Besov balls. In this paragraph we reproduce some classical results on Besov spaces, wavelet bases, wavelet-tensor products and wavelet threshold estimators (see Cohen [8], Donoho *et al.* [14] or Kerkyacharian and Picard [26]) that we use in the next sections.

*Wavelets and Besov spaces.* Let  $(\psi_\lambda)_\lambda$  be a regular wavelet basis adapted to the compact set  $D \subset \mathbb{R}^d$  (for a precise definition of  $(\psi_\lambda)_\lambda$  see hereafter). The multi-index  $\lambda$  concatenates the spatial index and the resolution level  $j = |\lambda|$ . Set  $\Lambda_j := \{\lambda, |\lambda| = j\}$  and  $\Lambda = \cup_{j \geq -1} \Lambda_j$ , for  $f : \mathbb{R}^d \rightarrow [0, \infty)$  in  $L_p(\mathbb{R}^d)$  we have

$$(6) \quad f = \sum_{j \geq -1} \sum_{\lambda \in \Lambda_j} \langle f, \psi_\lambda \rangle \psi_\lambda,$$

where  $j = -1$  incorporates the low frequency part of the decomposition and  $\langle \cdot, \cdot \rangle$  denotes the usual  $L_2$  inner product. Let  $s > 0$  and  $\pi \in (0, \infty]$ , a function  $f$  belongs to the Besov space  $\mathcal{B}_{\pi\infty}^s(D)$  if the norm

$$(7) \quad \|f\|_{\mathcal{B}_{\pi\infty}^s(D)} := \|f\|_{L_\pi(D)} + \|f^{(\lfloor s \rfloor)}\|_{L_\pi(D)} + \left\| \frac{w_\pi^2(f^{(\lfloor s \rfloor)}, t)}{t^a} \right\|_{L_\infty(D)}$$

is finite, where  $s = \lfloor s \rfloor + a$ ,  $\lfloor s \rfloor \in \mathbb{N}$  and  $a \in (0, 1]$ ,  $w$  is the modulus of continuity defined by

$$w_\pi^2(f, t) = \sup_{|h| \leq t} \|\mathbf{D}^h \mathbf{D}^h[f]\|_{L_\pi(D)}$$

and  $\mathbf{D}^h[f](x) = f(x - h) - f(x)$ . Equivalently we can define Besov space in term of wavelet coefficients (see Härdle *et al.* [22] or Kerkyacharian and Picard [26]),  $f$  belongs to  $\mathcal{B}_{\pi\infty}^s(D)$  if

$$\sup_{j \geq -1} 2^{j(s+d(1/2-1/\pi))} \left( \sum_{\lambda \in \Lambda_j} |\langle f, \psi_\lambda \rangle|^\pi \right)^{1/\pi} < \infty,$$

with usual modifications if  $\pi = \infty$ . We need additional properties on the wavelet basis  $(\psi_\lambda)_\lambda$ , which are listed in the following assumption.

**Assumption 1.** *Let  $p \geq 1$ , it holds that*

- For some  $C \geq 1$ ,

$$C^{-1} 2^{d|\lambda|(p/2-1)} \leq \|\psi_\lambda\|_{L_p(D)}^p \leq C 2^{d|\lambda|(p/2-1)}.$$

- For some  $C > 0$ ,  $\sigma > 0$  and for all  $s \leq \sigma$ ,  $J \geq 0$ ,

$$(8) \quad \left\| f - \sum_{j \leq J} \sum_{\lambda \in \Lambda_j} \langle f, \psi_\lambda \rangle \psi_\lambda \right\|_{L_p(D)} \leq C 2^{-Js} \|f\|_{\mathcal{B}_{\pi\infty}^s(D)}.$$

- If  $p \geq 1$ , for some  $C \geq 1$  and for any sequence of coefficients  $(u_\lambda)_{\lambda \in \Lambda}$ ,

$$(9) \quad C^{-1} \left\| \sum_{\lambda \in \Lambda} u_\lambda \psi_\lambda \right\|_{L_p(D)} \leq \left\| \left( \sum_{\lambda \in \Lambda} |u_\lambda \psi_\lambda|^2 \right)^{1/2} \right\|_{L_p(D)} \leq C \left\| \sum_{\lambda \in \Lambda} u_\lambda \psi_\lambda \right\|_{L_p(D)}.$$

- For any subset  $\Lambda_0 \subset \Lambda$  and for some  $C \geq 1$

$$(10) \quad C^{-1} \sum_{\lambda \in \Lambda_0} \|\psi_\lambda\|_{L_p(D)}^p \leq \int_D \left( \sum_{\lambda \in \Lambda_0} |\psi_\lambda(x)|^2 dx \right)^{p/2} \leq C \sum_{\lambda \in \Lambda_0} \|\psi_\lambda\|_{L_p(D)}^p.$$

Property (8) ensures that definition (7) of Besov spaces matches the definition in terms of linear approximation. Property (9) ensures that  $(\psi_\lambda)_\lambda$  is an unconditional basis of  $L_p$  and (10) is a super-concentration inequality (see Kerkyacharian and Picard [26] p.304 and p.306).

*Wavelet threshold estimator.* Let  $(\varphi, \psi)$  be a pair of scaling function and mother wavelet that generate a basis  $(\psi_\lambda)_\lambda$ . Denote  $\varphi_{0\mathbf{k}}(\cdot) = \varphi(\cdot - \mathbf{k})$ ,  $\psi_{j\mathbf{k}}(\cdot) = 2^{j/2}\psi(2^j \cdot - \mathbf{k})$ , the associated translated-dilated functions. Consider the triples  $(j, \mathbf{k}, A)$ , where  $j \in \mathbb{N}$ ,  $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d$  and  $A \in S_d$  the set of all non empty subsets of  $\{1, \dots, d\}$ . Let the functions  $\varphi^{\mathbf{k}} : \mathbb{R}^d \rightarrow \mathbb{R}$  and  $\psi_{(j, \mathbf{k}, A)} : \mathbb{R}^d \rightarrow \mathbb{R}$  defined by

$$\begin{aligned}\varphi^{\mathbf{k}}(x) &= \varphi^{\mathbf{k}}(x_1, \dots, x_d) = \prod_{i=1}^d \varphi_{0k_i}(x_i), \\ \psi_{(j, \mathbf{k}, A)}(x) &= \psi_{(j, \mathbf{k}, A)}(x_1, \dots, x_d) = \prod_{i \in A} \psi_{jk_i}(x_i) \prod_{i \in A^c} \varphi_{jk_i}(x_i).\end{aligned}$$

The system  $\{\varphi^{\mathbf{k}}, \mathbf{k} \in \mathbb{Z}^d, \psi_{(j, \mathbf{k}, A)}, j \in \mathbb{N}, \mathbf{k} \in \mathbb{Z}^d, A \in S_d\}$  is a wavelet-tensor product. If constructed on compactly supported wavelets  $(\varphi, \psi)$ , it satisfies Assumption 1 for some  $\sigma > 0$  (see Kerkyacharian and Picard [26] pp. 305, 306 and 314-315). To simplify notation, we write the basis  $\{\psi_{jk}, j \in \mathbb{N}, k \in A_j\}$  where  $A_j$  is a set of cardinality proportional to  $2^{jd}$  and incorporates boundary terms that we choose not to distinguish in the notation. Then, (6) becomes

$$f = \sum_{j \geq 0} \sum_{k \in A_j} \gamma_{jk} \psi_{jk},$$

where  $\gamma_{jk} = \int_{\mathbb{R}^d} \psi_{jk}(x) f(x) dx$ . We consider classical hard threshold estimators of the form

$$\hat{f}(\cdot) = \sum_{j=0}^J \sum_{k \in A_j} \hat{\gamma}_{jk} \mathbb{1}_{\{|\hat{\gamma}_{jk}| \geq \eta\}} \psi_{jk}(\cdot),$$

where  $\hat{\gamma}_{jk}$  is an estimator of  $\gamma_{jk}$ ,  $J$  and  $\eta$  are respectively the resolution level and the threshold, possibly depending on the data.

**2.2. Construction of the estimator.** We estimate densities  $f$  which satisfy a smoothness property in term of Besov balls

$$(11) \quad \mathcal{G}(s, \pi, M) = \{f \in \mathcal{G}(\mathbb{R}^d), \|f\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} \leq M\},$$

where  $M$  is a positive constant. The fact that  $f$  is in a Besov space  $\mathcal{B}_{\pi,\infty}^s(\mathbb{R}^d)$  is used to approximate  $\mathbf{P}^{-1}$  with a fixed point method. The fact that its Besov norm is bounded is used to control the risk of the estimator over the ball  $\mathcal{G}(s, \pi, M)$ .

*Construction of the inverse.* For  $\pi \geq 1$ , the space  $\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)$  is a Banach space if equipped with the Besov norm (7). Consider the mapping  $\mathbf{T}$ , for which  $f$  is a fixed point, defined for  $h$  in  $\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)$  by

$$(12) \quad \mathbf{T}[h] := \mathbf{P}[f] + h - \mathbf{P}[h],$$

where  $\mathbf{P}$  is defined in (3). The following Proposition 1 guarantees that the definition of the operator (12) matches the assumptions of the Banach fixed point theorem.

**Proposition 1.** *Let  $\pi \geq 1$  and  $0 < \alpha < \frac{1}{2}$ . Then, the mapping  $\mathbf{T}$  sends elements of  $\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)$  into itself and is a contraction. For all  $h_1, h_2 \in \mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)$  we have*

$$\|\mathbf{T}[h_1] - \mathbf{T}[h_2]\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} \leq 2\alpha \|h_1 - h_2\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)}.$$

Proposition 1 permits to apply the Banach fixed point theorem: let  $0 < \alpha < \frac{1}{2}$ , we derive that  $f$  is the unique fixed point of  $\mathbf{T}$  and from any initial point  $h_0$  in  $\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)$  we have

$$\|f - \mathbf{T}^{\circ K}[h_0]\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} \longrightarrow 0 \quad \text{as } K \rightarrow \infty.$$

We choose  $h_0 = \mathbf{P}[f]$  as initial point as we can construct an optimal estimator of  $\mathbf{P}[f]$  from the observations  $(Y_1, \dots, Y_n)$  (Lemma 1 in Section 5 ensures that  $\mathbf{P}[f]$  is in  $\mathcal{G}(s, \pi, M) \subset \mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)$ ).

**Remark 1.** *The restriction on  $\alpha$  is imposed by the factor 2 in the contraction property. This factor seems sharp: we control the difference of 2 singular probability measures, one absolutely continuous with respect to the Lebesgue measure and the other with the Dirac mass  $\delta$ . The operator  $\mathbf{T}$  seems useless to approximate  $\mathbf{P}^{-1}$  for  $\alpha \geq \frac{1}{2}$  (see also Section 3.2).*

**Proposition 2.** *Let  $K \geq 0$ , set  $(\delta - g)^{\star 0} = \delta$ , it holds that*

$$\mathbf{T}^{\circ K}[\mathbf{P}[f]] = \sum_{k=0}^K \alpha^k (\delta - g)^{\star k} \star \mathbf{P}[f] := \mathbf{H}_K[g] \star \mathbf{P}[f].$$

To establish Proposition 2 note that  $\mathbf{T}[h] = \mathbf{P}[f] + \alpha(\delta - g) \star h$ ,  $\forall h$ . Then, a recurrence directly leads to

$$\mathbf{T}^{\circ K}[h] = \alpha^K (\delta - g)^{\star K} \star h + \sum_{k=0}^{K-1} \alpha^k (\delta - g)^{\star k} \star \mathbf{P}[f].$$

Proposition 2 is useful from a numerical point of view to compute  $\mathbf{T}^{\circ K}[\mathbf{P}[f]]$ .

*Construction of an estimator of  $\mathbf{P}[f]$ .* Define the wavelet coefficients

$$(13) \quad \hat{\gamma}_{jk} = \frac{1}{n} \sum_{i=1}^n \psi_{jk}(Y_i).$$

Let  $\eta > 0$  and  $J \in \mathbb{N} \setminus \{0\}$ , define  $\hat{P}$  the estimator of  $\mathbf{P}[f]$  over  $D$  as

$$(14) \quad \hat{P}(x) = \sum_{j=0}^J \sum_{k \in A_j} \hat{\gamma}_{jk} \mathbf{1}_{\{|\hat{\gamma}_{jk}| \geq \eta\}} \psi_{jk}(x), \quad x \in D.$$

**Definition 1.** *Let  $\hat{f}_K$  be an estimator of  $f$  defined for  $K$  in  $\mathbb{N}$  and  $x$  in  $D$  as*

$$(15) \quad \hat{f}_K(x) = \mathbf{H}_K[G] \star \hat{P}(x)$$

where  $\mathbf{H}_K$  is defined in Proposition 2.

The estimator  $\hat{f}_K$  may be interpreted as follows, if  $\alpha = 0$  then  $f = \mathbf{P}[f]$ , which can be directly estimated, and one should take  $K = 0$ . However, if  $\alpha > 0$ , the dataset is contaminated with blurred observations that need to be counterbalanced by the addition of corrections. For instance, let  $\alpha = n^{-1/4}$ , straightforward computations lead to

$$T[\mathbf{P}[f]] = (1 - n^{-1/2})f + O(n^{-1/2})$$

whereas the direct approximation is  $\mathbf{P}[f] = (1 - n^{-1/4})f + O(n^{-1/4})$ . Then, applying  $\mathbf{T}$  permits to approximate  $f$  more rapidly than the crude approximation of  $\mathbf{P}[f]$ .

### 2.3. An upper bound.

**Theorem 1.** *We work under Assumption 1. Let  $p \geq \pi \geq 1$ ,  $\sigma > s > d/\pi$  and  $\hat{P}$  be the threshold wavelet estimator of  $\mathbf{P}[f]$  on  $D$  defined in (14). Take  $J$  such that  $2^{Jd} n^{-1} \log(n^{1/2}) \leq 1$  and  $\eta = \kappa n^{-1/2} \sqrt{\log(n^{1/2})}$ , for some  $\kappa > 0$ .*

(1) *The estimator  $\hat{P}$  of  $\mathbf{P}[f]$  satisfies for sufficiently large  $\kappa > 0$*

$$\sup_{\mathbf{P}[f] \in \mathcal{G}(s, \pi, M)} \left( \mathbb{E} \left[ \left\| \hat{P} - \mathbf{P}[f] \right\|_{L^p(D)}^p \right] \right)^{1/p} \leq C (\log(n))^\epsilon n^{-\delta(s, p, \pi)},$$

where  $C$  depends on  $s, \pi, p, M, \psi$ ,

$$(16) \quad \delta(s, p, \pi) = \min \left\{ \frac{s}{2s+d}, \frac{s+d/p-d/\pi}{2(s+d/2-d/\pi)} \right\}$$

and  $\mathfrak{c}$  is defined as follows

$$(17) \quad \mathfrak{c} = \begin{cases} \delta(s, p, \pi), & \text{if } \pi \neq \frac{dp}{(2s+d)} \\ \delta(s, p, \pi) + 1, & \text{otherwise.} \end{cases}$$

An explicit bound for  $\kappa$  is given in Lemma 3.

- (2) Suppose moreover that Assumption (H1) holds. The estimator  $\widehat{f}_K$  for  $K \in \mathbb{N}$  defined in (15) satisfies for sufficiently large  $\kappa > 0$

$$\sup_{f \in \mathcal{G}(s, \pi, M)} (\mathbb{E} [\|\widehat{f}_K - f\|_{L_p(D)}^p])^{1/p} \leq \max \left\{ \left( \frac{1 - (2\alpha)^{K+1}}{1 - 2\alpha} \right) (\log(n))^{\mathfrak{c}} n^{-\delta(s, p, \pi)}, (2\alpha)^{K+1} M \right\},$$

where  $C$  depends on  $s, \pi, p, M, \psi$  and  $\mathfrak{c}$  is defined in (17).

Proof of Theorem 1 is postponed to Section 5. Note that the estimator  $\widehat{f}_K$  with  $J$  and  $\eta$  chosen as in Theorem 1 is adaptive, recall that  $K$  is chosen by the practitioner.

### 3. DISCUSSION

**3.1. Discussion on the rates of convergence.** The upper bound of Theorem 1 is easy to interpret. The estimator cannot perform well if  $\mathbf{P}^{-1}$  is poorly approximated by  $\mathbf{T}^{\circ K}$ , it leads to the deterministic loss  $(2\alpha)^{K+1}$ . It also cannot perform better than the estimator  $\widehat{P}$  of  $\mathbf{P}[f]$ , which imposes the random error, up to the logarithmic factor,  $n^{-\delta(s, p, \pi)}$ , this is optimal.

Ideally, we should take  $K$  such that the deterministic error is negligible compared to the random error and would realize the tradeoff  $(1 - (2\alpha)^{K+1}) / (1 - 2\alpha)(n / \log(n))^{-\delta(s, p, \pi)} \asymp (2\alpha)^{K+1}$ . As  $\delta(s, p, \pi)$  is unknown, we use that  $\delta(s, p, \pi) \leq \frac{1}{2}$ , which leads to the adaptive choice  $K^*$ , the smallest integer such that  $(1 - (2\alpha)^{K+1}) / (1 - 2\alpha)(n / \log(n))^{-\frac{1}{2}} > (2\alpha)^{K+1}$ . The solution cannot be made explicit, but note that

$$\frac{1 - (2\alpha)^{K+1}}{1 - 2\alpha} = \sum_{k=0}^K (2\alpha)^k \in [0, K + 1].$$

Then, we choose  $K^*$  as the smallest integer such that  $(K + 1)(n / \log(n))^{-\frac{1}{2}} > (2\alpha)^{K+1}$ , that is

$$(18) \quad K^* = \left\lceil \frac{W\left(\sqrt{\frac{n}{\log(n)}} \log(1/(2\alpha))\right)}{\log(1/(2\alpha))} - 1 \right\rceil,$$

where  $W$  is the Lambert  $W$  function, defined as the inverse (for the composition product) of  $u \rightarrow ue^u$ . The function  $W$  is increasing, then, as expected  $K^*$  increases as with  $n$ . If  $\alpha < 1/2$  is fixed we can use the equivalent  $W(x) \sim x$  as  $x \rightarrow \infty$  to get

$$K^* \asymp \frac{\log(n)}{2 \log(1/(2\alpha))},$$

which increases with  $\alpha$  and  $n$ . Moreover, Theorem 1 states that  $\widehat{f}_{K^*}$  attains, up to an additional logarithmic factor, the lower bound of Lepski and Willer [28]. It follows that  $\widehat{f}_{K^*}$  is an adaptive minimax estimator of  $f$ .



**3.2. An alternative procedure.** The restriction on  $\alpha$  can also be understood in terms of Fourier transform. Indeed, we may rewrite the characteristic function of  $f$  as follows

$$\mathcal{F}[f](u) = \frac{1}{1-\alpha} \frac{\mathcal{F}[\mathbf{P}[f]](u)}{1 + \frac{\alpha}{1-\alpha} \mathcal{F}[G](u)} = \frac{1}{1-\alpha} \mathcal{F}[\mathbf{P}[f]](u) \sum_{k=0}^{\infty} (-1)^k \left( \frac{\alpha}{1-\alpha} \right)^k (\mathcal{F}[g](u))^k$$

which holds if  $\alpha < 1 - \alpha$  or equivalently  $\alpha < 1/2$ . It leads to the following series expansion:

$$f = \mathbf{P}[f] \star \sum_{k=0}^{\infty} (-1)^k \frac{\alpha^k}{(1-\alpha)^{k+1}} g^{\star k}.$$

Truncating the previous series at order  $K$  and replacing  $\mathbf{P}[f]$  by the estimator (14) also leads to a direct estimation procedure  $\tilde{f}_K$  that can be studied with a  $L_p$  loss function. This estimator would have the same properties as the one based on a fixed point argument since the remainders are of the same order.

**3.3. Link with the estimation of random sums.** Assumption (H1) appears naturally when one estimates the jump density of a jump process, *e.g.* a compound Poisson process or a renewal reward process, from high frequency observations. It can be formalized as a deconvolution problem. Let  $Z$  be a jump process with stationary increments defined by

$$Z_t = \sum_{i=1}^{N_t} \xi_i$$

where  $N$  is a counting process with stationary increments and independent of  $(\xi_i)$  which are i.i.d. with density  $f$ . Without loss of generality one can suppose that  $n$  nonzero increments of  $Z$  are observed at the sampling rate  $\Delta$ . The distribution of  $Z_\Delta | Z_\Delta \neq 0$  is

$$\begin{aligned} & \mathbb{P}(N_\Delta = 1 | N_\Delta \neq 0) f + \sum_{k=2}^{\infty} \mathbb{P}(N_\Delta = k | N_\Delta \neq 0) f^{\star k} \\ &= f \star \left( \mathbb{P}(N_\Delta = 1 | N_\Delta \neq 0) \delta + \sum_{k=2}^{\infty} f^{\star k-1} \mathbb{P}(N_\Delta = k | N_\Delta \neq 0) \right). \end{aligned}$$

We recover a similar setting as the one studied above. Note that the exact form of  $\mathbf{P}$  defined in (3) is unknown since it depends on the counting process  $N$  and on the density  $f$  itself. A fixed point approach has been investigated in Duval [16] (see also Duval [15] for an explicit computation of  $\mathbf{P}^{-1}$ ) in the particular case where the process is observed at high frequency, *i.e.*  $\Delta \rightarrow 0$ . Then, it is possible to estimate  $f$  at usual nonparametric rates. The constraint on the sampling rate  $\Delta$  entails that nonzero increments are *most of the time* realizations of  $f$ . If  $\Delta$  is small, we have  $\mathbb{P}(N_\Delta = 1 | N_\Delta \neq 0) \approx 1 - \Delta$ . We recover a condition similar to Assumption (H1).

In the case of a fixed sample size  $\Delta = 1$ , Assumption (H1) may not be satisfied, but it is possible to build estimators of  $f$ . For instance, a consistent estimator of the cumulative distribution function is studied in Buchmann and Grübel [1] and a consistent density estimator in van Es *et al.* [17]. The estimation of a Lévy measure from the discrete observation of a Lévy processes, which may have a Brownian component, is also possible. In that latter case the results of van Es *et al.* [18], Lee *et al.* [27] and Gugushvili *et al.* [20] apply (see also *e.g.* Neumann and Reiß [32], Comte and Genon-Catalot [5] or Gugushvili [21], which take advantage of the Lévy-Kintchine formula to derive minimax estimation procedures.).

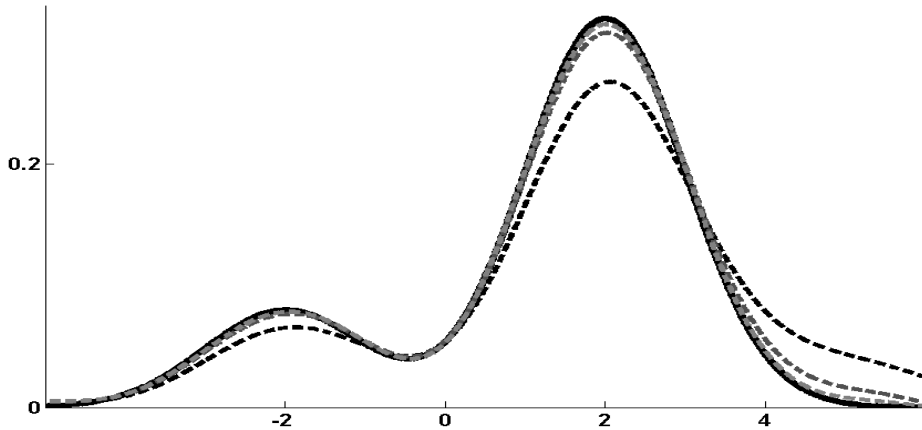


FIGURE 1. Estimators of the density  $f$  (plain black) for  $n = 10^6$  and  $\alpha = 0.25$ :  $\hat{f}_0$  (dotted black),  $\hat{f}_1$  (dotted dark grey) and  $\hat{f}_2$  (dashed light grey).

#### 4. NUMERICAL STUDY

In this Section we illustrate, in the univariate setting, how the method performs on simulated data and examine in particular its behavior when  $K$  increases and  $\alpha$  is varying. We also compare its performances with an oracle: the wavelet estimator we would compute in the idealized framework where direct observations  $(X_i, 1 \leq i \leq n)$  are available.

Wavelet estimators are based on the evaluation of the first wavelet coefficients. To perform those we use Symlets 16 wavelet functions, that are compactly supported and satisfy Assumption 1. Moreover we transform the data in an equispaced signal on a grid of length  $2^L$  with  $L = 8$ . It is the binning procedure (see Härdle *et al.* [22] Chap. 12). The threshold and the resolution level are chosen as in Theorem 1. The parameter  $\kappa$  is taken equal to 1. The estimators we compute take the form of a vector giving the estimated values of the density  $f$  on the uniform grid  $[a, b]$  with mesh 0.01, where  $a$  and  $b$  are adapted to the estimated density  $f$ . We use the wavelet toolbox of Matlab. To compute  $\hat{f}_K$  for  $K \geq 1$ , we compute  $((\delta - g)^{*k} \star \hat{P}, 1 \leq k \leq K - 1)$  with the function `conv` of Matlab.

Figure 1 represents the estimation procedure for different values of  $K \in \{0, 1, 2\}$ ,  $\alpha = 0.25$ ,  $g$  a Gaussian density with mean 2 and variance 2 and  $f$  being the following mixture:

$$0.2\mathcal{N}(-2, 1) + 0.8\mathcal{N}(2, 1).$$

All the estimators are evaluated on the same trajectory. They all manage to reproduce the shape of the density  $f$  but  $\hat{f}_0$  and  $\hat{f}_1$  are biased. Increasing  $K$  permits to reduce this bias. Even though the optimal choice for  $K$  given by (18) is 4, it seems that with  $K = 1$  or 2 we already have a good estimation of the mixture.

Evaluation of  $L_2$  risks confirms the former graphical observation. We approximate the  $L_2$  errors by Monte Carlo. For that we compute  $M = 1000$  times each estimator, for each iteration, the estimators  $(\hat{f}_K)_{K \geq 0}$  are computed on the same dataset  $(Y_j, 1 \leq j \leq n)$  and the oracle on the direct observations  $(X_i, 1 \leq i \leq n)$  used to compute the values  $(Y_j, 1 \leq j \leq n)$ . The results are reproduced in Tables 1, 2 and 3 and correspond to the values  $\alpha = 0.1, 0.25$  and 0.49 respectively. First, we see that the error  $g$  has little impact on the results. In Table 1 we

consider  $\alpha = 0.1$ , which is small, we observe that the  $L_2$  risks decrease with  $K$ , until  $K$  gets larger than  $K^*$ , afterwards, they get stable. In Tables 2 and 3 we see the procedure works for larger values of  $\alpha$  ( $\alpha \in \{0.25; 0.49\}$ ), even if  $\alpha$  gets close to the limiting value  $\alpha = 1/2$ . In every Tables, the risks associated to  $\widehat{f}_{K^*}$  match the oracle ones. Comparing the different Tables, we notice that the larger  $\alpha$  is, the larger  $n$  needs to be to get a risk close to the oracle one.

Numerical results are consistent with the theoretical results of Theorem 1. Also it appears that the most significant gains in the risk are observed at the first iterations of the fixed point method ( $K \in \{1, 2\}$ ) and that the gains is less important afterward, even though they still permit to improve the risks.

$g$	$n(K^*)$	Oracle	$\widehat{f}_0$	$\widehat{f}_1$	$\widehat{f}_2$	$\widehat{f}_{K^*}$
$\mathcal{N}(2, 2)$	$10^3$	$0.28 \times 10^{-2}$	$0.55 \times 10^{-2}$	$0.38 \times 10^{-2}$	$0.38 \times 10^{-2}$	$0.38 \times 10^{-2}$
	(1)	$(1.57 \times 10^{-3})$	$(2.18 \times 10^{-3})$	$(2.23 \times 10^{-3})$	$(2.24 \times 10^{-3})$	$(2.23 \times 10^{-3})$
	$10^4$	$0.05 \times 10^{-2}$	$0.19 \times 10^{-2}$	$0.08 \times 10^{-2}$	$0.08 \times 10^{-2}$	$0.08 \times 10^{-2}$
$\mathcal{U}([-1, 3])$	(1)	$(0.31 \times 10^{-3})$	$(0.47 \times 10^{-3})$	$(0.35 \times 10^{-3})$	$(0.35 \times 10^{-3})$	$(0.35 \times 10^{-3})$
	$10^5$	$0.01 \times 10^{-2}$	$0.12 \times 10^{-2}$	$0.01 \times 10^{-2}$	$0.01 \times 10^{-2}$	$0.01 \times 10^{-2}$
	(2)	$(0.05 \times 10^{-3})$	$(0.10 \times 10^{-3})$	$(0.04 \times 10^{-3})$	$(0.04 \times 10^{-3})$	$(0.04 \times 10^{-3})$
$\mathcal{U}([-1, 3])$	$10^3$	$0.27 \times 10^{-2}$	$0.38 \times 10^{-2}$	$0.31 \times 10^{-2}$	$0.31 \times 10^{-2}$	$0.31 \times 10^{-2}$
	(1)	$(1.57 \times 10^{-3})$	$(1.75 \times 10^{-3})$	$(1.76 \times 10^{-3})$	$(1.77 \times 10^{-3})$	$(1.76 \times 10^{-3})$
	$10^4$	$0.05 \times 10^{-2}$	$0.15 \times 10^{-2}$	$0.08 \times 10^{-2}$	$0.08 \times 10^{-2}$	$0.08 \times 10^{-2}$
$\mathcal{U}([-1, 3])$	(1)	$(0.33 \times 10^{-3})$	$(0.53 \times 10^{-3})$	$(0.45 \times 10^{-3})$	$(0.44 \times 10^{-3})$	$(0.45 \times 10^{-3})$
	$10^5$	$0.10 \times 10^{-3}$	$0.66 \times 10^{-3}$	$0.09 \times 10^{-3}$	$0.09 \times 10^{-3}$	$0.09 \times 10^{-3}$
	(2)	$(0.59 \times 10^{-4})$	$(0.80 \times 10^{-4})$	$(0.46 \times 10^{-4})$	$(0.46 \times 10^{-4})$	$(0.46 \times 10^{-4})$

TABLE 1. Mean of the  $L_2$ -risks for different values of  $K$  and the oracle estimator; standard deviation in parenthesis. In this case,  $f$  is  $0.2\mathcal{N}(-2, 1) + 0.8\mathcal{N}(2, 1)$ ,  $\alpha = 0.1$  and  $D = [-6, 6]$ . In bold it is the loss of the estimator  $\widehat{f}_{K^*}$ , where  $K^*$  is chosen as in (18).

$g$	$n(K^*)$	Oracle	$\widehat{f}_0$	$\widehat{f}_1$	$\widehat{f}_2$	$\widehat{f}_{K^*}$
$\mathcal{N}(2, 2)$	$10^3$	$0.27 \times 10^{-2}$	$1.32 \times 10^{-2}$	$0.63 \times 10^{-2}$	$0.63 \times 10^{-2}$	$0.63 \times 10^{-2}$
	(2)	$(1.57 \times 10^{-3})$	$(3.62 \times 10^{-3})$	$(3.61 \times 10^{-3})$	$(3.61 \times 10^{-3})$	$(3.61 \times 10^{-3})$
	$10^4$	$0.05 \times 10^{-2}$	$0.80 \times 10^{-2}$	$0.17 \times 10^{-2}$	$0.12 \times 10^{-2}$	$0.12 \times 10^{-2}$
$\mathcal{U}([-1, 3])$	(3)	$(0.32 \times 10^{-3})$	$(0.89 \times 10^{-3})$	$(0.62 \times 10^{-3})$	$(0.59 \times 10^{-3})$	$(0.59 \times 10^{-3})$
	$10^5$	$0.01 \times 10^{-2}$	$0.72 \times 10^{-2}$	$0.06 \times 10^{-2}$	$0.02 \times 10^{-2}$	$0.01 \times 10^{-2}$
	(4)	$(0.06 \times 10^{-3})$	$(0.29 \times 10^{-3})$	$(0.12 \times 10^{-3})$	$(0.08 \times 10^{-3})$	$(0.07 \times 10^{-3})$
$\mathcal{U}([-1, 3])$	$10^3$	$0.27 \times 10^{-2}$	$0.75 \times 10^{-2}$	$0.37 \times 10^{-2}$	$0.36 \times 10^{-2}$	$0.37 \times 10^{-2}$
	(2)	$(1.50 \times 10^{-3})$	$(2.83 \times 10^{-3})$	$(2.49 \times 10^{-3})$	$(2.52 \times 10^{-3})$	$(2.49 \times 10^{-3})$
	$10^4$	$0.05 \times 10^{-2}$	$0.50 \times 10^{-2}$	$0.12 \times 10^{-2}$	$0.10 \times 10^{-2}$	$0.10 \times 10^{-2}$
$\mathcal{U}([-1, 3])$	(3)	$(0.31 \times 10^{-3})$	$(0.69 \times 10^{-3})$	$(0.59 \times 10^{-3})$	$(0.58 \times 10^{-3})$	$(0.58 \times 10^{-3})$
	$10^5$	$0.10 \times 10^{-3}$	$3.50 \times 10^{-3}$	$0.30 \times 10^{-3}$	$0.10 \times 10^{-3}$	$0.10 \times 10^{-3}$
	(4)	$(0.60 \times 10^{-4})$	$(1.60 \times 10^{-4})$	$(0.78 \times 10^{-4})$	$(0.70 \times 10^{-4})$	$(0.70 \times 10^{-4})$

TABLE 2. Mean of the  $L_2$ -risks for different values of  $K$  and the oracle estimator; standard deviation in parenthesis. In this case,  $f$  is  $0.2\mathcal{N}(-2, 1) + 0.8\mathcal{N}(2, 1)$ ,  $\alpha = 0.25$  and  $D = [-6, 6]$ . In bold it is the loss of the estimator  $\widehat{f}_{K^*}$ , where  $K^*$  is chosen as in (18).

## 5. PROOFS

In the sequel  $C$  denotes a constant which may vary from line to line. Its dependency in other constants are sometimes given in subscripts.

$g$	$n$ ( $K^*$ )	Oracle	$\widehat{f}_0$	$\widehat{f}_1$	$\widehat{f}_2$	$\widehat{f}_{K^*}$
$\mathcal{N}(2, 2)$	$10^3$ (9)	$0.27 \times 10^{-2}$ ( $1.53 \times 10^{-3}$ )	$4.05 \times 10^{-2}$ ( $7.90 \times 10^{-3}$ )	$2.33 \times 10^{-2}$ ( $8.82 \times 10^{-3}$ )	$1.94 \times 10^{-2}$ ( $8.75 \times 10^{-3}$ )	$1.87 \times 10^{-2}$ ( $8.65 \times 10^{-3}$ )
	$10^4$ (21)	$0.05 \times 10^{-2}$ ( $0.31 \times 10^{-3}$ )	$2.90 \times 10^{-2}$ ( $2.12 \times 10^{-3}$ )	$0.99 \times 10^{-2}$ ( $2.03 \times 10^{-3}$ )	$0.51 \times 10^{-2}$ ( $1.76 \times 10^{-3}$ )	$0.35 \times 10^{-2}$ ( $1.50 \times 10^{-3}$ )
	$10^5$ (40)	$0.01 \times 10^{-2}$ ( $0.06 \times 10^{-3}$ )	$2.68 \times 10^{-2}$ ( $0.51 \times 10^{-3}$ )	$0.68 \times 10^{-2}$ ( $0.42 \times 10^{-3}$ )	$0.18 \times 10^{-2}$ ( $0.28 \times 10^{-3}$ )	$0.02 \times 10^{-2}$ ( $0.15 \times 10^{-3}$ )
$\mathcal{U}([-1, 3])$	$10^3$ (9)	$0.28 \times 10^{-2}$ ( $1.61 \times 10^{-3}$ )	$1.87 \times 10^{-2}$ ( $4.39 \times 10^{-3}$ )	$0.79 \times 10^{-2}$ ( $4.20 \times 10^{-3}$ )	$0.67 \times 10^{-2}$ ( $4.41 \times 10^{-3}$ )	$0.69 \times 10^{-2}$ ( $4.82 \times 10^{-3}$ )
	$10^4$ (21)	$0.05 \times 10^{-2}$ ( $0.31 \times 10^{-3}$ )	$1.43 \times 10^{-2}$ ( $1.50 \times 10^{-3}$ )	$0.32 \times 10^{-2}$ ( $1.16 \times 10^{-3}$ )	$0.15 \times 10^{-2}$ ( $0.94 \times 10^{-3}$ )	$0.11 \times 10^{-2}$ ( $0.87 \times 10^{-3}$ )
	$10^5$ (40)	$0.10 \times 10^{-3}$ ( $0.58 \times 10^{-4}$ )	$13.30 \times 10^{-3}$ ( $0.44 \times 10^{-4}$ )	$2.30 \times 10^{-3}$ ( $0.28 \times 10^{-4}$ )	$0.60 \times 10^{-3}$ ( $0.19 \times 10^{-4}$ )	$0.30 \times 10^{-3}$ ( $0.15 \times 10^{-4}$ )

TABLE 3. Mean of the  $L_2$ -risks for different values of  $K$  and the oracle estimator; standard deviation in parenthesis. In this case,  $f$  is  $0.2\mathcal{N}(-2, 1) + 0.8\mathcal{N}(2, 1)$ ,  $\alpha = 0.49$  and  $D = [-6, 6]$ . In bold it is the loss of the estimator  $\widehat{f}_{K^*}$ , where  $K^*$  is chosen as in (18).

5.1. **Preliminary.** We establish a technical lemma, which states that regularity assumptions on  $f$  transfer to  $\mathbf{P}[f]$ .

**Lemma 1.** *If  $f$  belongs to  $\mathcal{G}(s, \pi, M)$  then,  $\mathbf{P}[f]$  also belongs to  $\mathcal{G}(s, \pi, M)$ .*

*Proof of Lemma 1.* It is straightforward to derive  $\|\mathbf{P}[f]\|_{L_1(\nu)} = 1$ . The remainder of the proof is a consequence of the following result: Let  $f \in \mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)$  and  $g \in L_1(\mathbb{R}^d)$  we have

$$(19) \quad \|f \star g\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} \leq \|f\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} \|g\|_{L_1(\mathbb{R}^d)}.$$

To prove (19) we use the definition of the Besov norm (7); the result is a consequence of Young's inequality and elementary properties of the convolution product. First, Young's inequality gives

$$(20) \quad \|f \star g\|_{L_\pi(\mathbb{R}^d)} \leq \|f\|_{L_\pi(\mathbb{R}^d)} \|g\|_{L_1(\mathbb{R}^d)}.$$

Second, the differentiation property of the convolution product leads for  $n \geq 1$  to

$$(21) \quad \left\| \frac{d^n}{dx^n} (f \star g) \right\|_{L_\pi(\mathbb{R}^d)} = \left\| \left( \frac{d^n}{dx^n} f \right) \star g \right\|_{L_\pi(\mathbb{R}^d)} \leq \left\| \frac{d^n}{dx^n} f \right\|_{L_\pi(\mathbb{R}^d)} \|g\|_{L_1(\mathbb{R}^d)}.$$

Finally, translation invariance of the convolution product gives

$$(22) \quad \|\mathbf{D}^h \mathbf{D}^h [(f \star g)^{(n)}]\|_{L_\pi(\mathbb{R}^d)} = \|(\mathbf{D}^h \mathbf{D}^h [f^{(n)}]) \star g\|_{L_\pi(\mathbb{R}^d)} \leq \|\mathbf{D}^h \mathbf{D}^h [f^{(n)}]\|_{L_\pi(\mathbb{R}^d)} \|g\|_{L_1(\mathbb{R}^d)}.$$

Inequality (19) is then obtained by bounding  $\|f \star g\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)}$  using (20), (21) and (22). We now complete the proof of Lemma 1, using the triangle inequality and (19)

$$\|\mathbf{P}[f]\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} = \|G \star f\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} = \|(1 - \alpha)f + \alpha g \star f\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} \leq \|f\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} \leq M,$$

where  $\alpha$  and  $g$  are defined in (H1). The proof is now complete.  $\square$

5.2. **Proof of Proposition 1.** We show that  $\mathbf{T}$  is a contraction that sends elements of  $\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)$  into  $\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)$ . We have for all  $h_1, h_2 \in \mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)$

$$\mathbf{T}[h_1] - \mathbf{T}[h_2] = h_1 - h_2 - G \star (h_1 - h_2) = \alpha(h_1 - h_2) - \alpha g \star (h_1 - h_2).$$

It follows from Young's inequality and assumption (H1) that

$$\|\mathbf{T}[h_1] - \mathbf{T}[h_2]\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} \leq 2\alpha \|h_1 - h_2\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)}.$$

Finally, let  $h \in \mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)$ . The last assertion together with the fact that the null function is in  $\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)$  and Lemma 1 lead to

$$\begin{aligned} \|\mathbf{T}[h]\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} &\leq \|\mathbf{T}[0]\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} + \|\mathbf{T}[h] - \mathbf{T}[0]\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} \\ &\leq \|\mathbf{P}[f]\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} + 2\alpha\|h\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} < \infty. \end{aligned}$$

The proof is now complete.

### 5.3. Proof of Theorem 1.

*Proof of part 1) of Theorem 1.* To prove part 1) of Theorem 1 we apply the general results of Kerkyacharian and Picard [26]. For that we state some technical lemmas whose proof is based on classical Rosenthal's and Bernstein's inequalities.

**Lemma 2.** *Let  $2^{jd} \leq n$ , then for  $p \geq 1$  we have*

$$\mathbb{E}[|\widehat{\gamma}_{jk} - \gamma_{jk}|^p] \leq Cn^{-p/2},$$

where  $C$  depends on  $p$ ,  $\|\psi\|_{L_p(\mathbb{R}^d)}$ ,  $M$  and  $\widehat{\gamma}_{jk}$  is defined in (13) and

$$(23) \quad \gamma_{jk} = \int_{\mathbb{R}^d} \psi_{jk}(y) \mathbf{P}[f](y) dy.$$

*Proof of Lemma 2.* Let  $p \geq 2$ , the result is obtained applying Rosenthal's inequality: let  $(U_i)_i$  be centered independent real random variables such that  $\mathbb{E}[|U_i|^p] < \infty$ , then there exists  $C_p$  such that

$$(24) \quad \mathbb{E}\left[\left|\sum_{i=1}^n U_i\right|^p\right] \leq C_p \left\{ \sum_{i=1}^n \mathbb{E}[|U_i|^p] + \left(\sum_{i=1}^n \mathbb{E}[|U_i|^2]\right)^{p/2} \right\}.$$

Set  $Z_i = \psi_{jk}(Y_i)$ , for  $p \geq 2$  we have by convex inequality

$$\begin{aligned} \mathbb{E}[|Z_i - \mathbb{E}[Z_i]|^p] &\leq 2^p \mathbb{E}[|Z_i|^p] \leq 2^p 2^{jd p/2} \int_{\mathbb{R}^d} |\psi(2^j y - k)|^p \mathbf{P}[f](y) dy \\ &= 2^p 2^{jd(p/2-1)} \int_{\mathbb{R}^d} |\psi(z)|^p \mathbf{P}[f](2^{-j}(z - k)) dz \end{aligned}$$

where we made the substitution  $z = 2^j y - k$ . Lemma 1 and Sobolev embeddings (see [8, 14, 22])

$$(25) \quad \mathcal{B}_{\pi\infty}^s \hookrightarrow \mathcal{B}_{p\infty}^{s'} \quad \text{and} \quad \mathcal{B}_{\pi\infty}^{s'} \hookrightarrow \mathcal{B}_{\infty\infty}^s,$$

where  $p > \pi$ ,  $s\pi > d$  and  $s' = s - d/\pi + d/p$ , give  $\|\mathbf{P}[f]\|_{\infty} \leq M$ . It follows that

$$\mathbb{E}[|Z_i - \mathbb{E}[Z_i]|^p] \leq 2^p 2^{jd(p/2-1)} \|\psi\|_{L_p(\mathbb{R}^d)}^p M$$

and  $\mathbb{E}[|Z_i - \mathbb{E}[Z_i]|^2] \leq M$  since  $\|\psi\|_{L_2(\mathbb{R}^d)} = 1$ . Rosenthal's inequality (26) gives for  $p \geq 2$

$$\mathbb{E}[|\widehat{\gamma}_{jk} - \gamma_{jk}|^p] \leq C_p \left\{ 2^p \left(\frac{2^{jd}}{n}\right)^{\frac{p}{2}-1} \|\psi\|_{L_p(\mathbb{R}^d)}^p M + M^{p/2} \right\} n^{-\frac{p}{2}}.$$

Finally, since  $2^{jd} \leq n$

$$(26) \quad \mathbb{E}[|\widehat{\gamma}_{jk} - \gamma_{jk}|^p] \leq C_{p, \|\psi\|_{L_p(\mathbb{R}^d)}, M} n^{-p/2}.$$

Now if  $1 \leq p \leq 2$ , we apply Jensen's inequality to get

$$\mathbb{E}[|\widehat{\gamma}_{jk} - \gamma_{jk}|^p] \leq \left(\mathbb{E}[|\widehat{\gamma}_{jk} - \gamma_{jk}|^2]\right)^{p/2},$$

and applying (26) leads to  $\mathbb{E}[|\widehat{\gamma}_{jk} - \gamma_{jk}|^p] \leq C_{2, \|\psi\|_{L_2(\mathbb{R}^d)}, M} n^{-p/2}$ . This completes the proof.  $\square$

**Lemma 3.** Choose  $j$  and  $c$  such that

$$2^{jd}n^{-1} \log(n^{1/2}) \leq 1 \text{ and } c^2 \geq 8 \left( M + \frac{c \|\psi\|_\infty}{3} \right).$$

For all  $r \geq 1$ , let  $\kappa_r = cr$ . We have

$$\mathbb{P} \left( |\hat{\gamma}_{jk} - \gamma_{jk}| \geq \frac{\kappa_r}{2} n^{-1/2} \sqrt{\log(n^{1/2})} \right) \leq n^{-r/2},$$

where  $\hat{\gamma}_{jk}$  is defined in (13) and  $\gamma_{jk}$  in (23).

*Proof of Lemma 3.* The proof is obtained applying Bernstein's inequality. Let  $(U_i)_i$  be centered, bounded and independent real random variables such that  $|U_i| \leq S$  and set  $s_n^2 = \sum_{i=1}^n \mathbb{E}[U_i^2]$ . Then for any  $\lambda > 0$ ,

$$(27) \quad \mathbb{P} \left( \left| \sum_{i=1}^n U_i \right| > \lambda \right) \leq 2 \exp \left( - \frac{\lambda^2}{2(s_n^2 + \frac{\lambda S}{3})} \right).$$

We keep notation  $Z_i$  introduced in the proof of Lemma 2,  $\hat{\gamma}_{jk} - \gamma_{jk}$  is a sum of centered and identically distributed random variables bounded by  $2^{jd/2} \|\psi\|_\infty$  such that  $\mathbb{E}[|Z_i - \mathbb{E}[Z_i]|^2] \leq M$ . It follows from (27)

$$\begin{aligned} \mathbb{P} \left( |\hat{\gamma}_{jk} - \gamma_{jk}| \geq \frac{\kappa_r}{2} n^{-1/2} \sqrt{\log(n^{1/2})} \right) &= \mathbb{P} \left( \left| \sum_{i=1}^n Z_i - \mathbb{E}(Z_i) \right| \geq \frac{\kappa_r}{2} n^{1/2} \sqrt{\log(n^{1/2})} \right) \\ &\leq 2 \exp \left( - \frac{\kappa_r^2 \log(n^{1/2})}{8 \left( M + \frac{\kappa_r 2^{jd/2} n^{-1/2} \sqrt{\log(n^{1/2})} \|\psi\|_\infty}{3} \right)} \right). \end{aligned}$$

Using that  $2^{jd}n^{-1} \log(n^{1/2}) \leq 1$  we have

$$\mathbb{P} \left( |\hat{\gamma}_{jk} - \gamma_{jk}| \geq \frac{\kappa_r}{2} n^{-1/2} \sqrt{\log(n^{1/2})} \right) \leq 2 \exp \left( - \frac{c^2 r}{8 \left( M + \frac{\kappa_r \|\psi\|_\infty}{3} \right)} r \log(n^{1/2}) \right) \leq n^{-r/2},$$

if  $c^2 \geq 8 \left( M + \frac{c \|\psi\|_\infty}{3} \right)$  and  $r \geq 1$ . The proof is complete.  $\square$

*Proof of part 1) of Theorem 1.* It is a consequence of Lemma 1, 2, 3 and of the general theory of wavelet threshold estimators of Kerkyacharian and Picard [26]. Let  $J$ , such that  $2^{Jd} = n(\log(n))^{-1}$ . Conditions (5.1) and (5.2) of Theorem 5.1 of [26], are satisfied –Lemma 2 and 3– with  $c(n) = n^{-1/2} \sqrt{\log(n)}$  and  $\Lambda_n = c(n)^{-1}$  (with the notation of [26]), we can now apply Theorem 5.1, its Corollary 5.1 and Theorem 6.1 of [26] to obtain the result.  $\square$

*Completion of the proof of Theorem 1.* We decompose the  $L_p$  loss as follows, using notation of Proposition 2 and Definition 1

$$(28) \quad \left( \mathbb{E}[\|\hat{f}_K - f\|_{L_p(D)}^p] \right)^{\frac{1}{p}} \leq \left( \mathbb{E}[\|\hat{f}_K - \mathbf{H}_K[G] \star \mathbf{P}[f]\|_{L_p(D)}^p] \right)^{\frac{1}{p}} + \|\mathbf{H}_K[G] \star \mathbf{P}[f] - f\|_{L_p(D)}.$$

An upper bound for the first term is given by part 1) of Theorem 1, the triangle inequality and Young's inequality

$$\begin{aligned} \mathbb{E}[\|\hat{f}_K - \mathbf{H}_K[G] \star \mathbf{P}[f]\|_{L_p(D)}^p] &= \mathbb{E}[\|\mathbf{H}_K[G] \star (\hat{P} - \mathbf{P}[f])\|_{L_p(D)}^p] \\ &\leq \|\mathbf{H}_K[g]\|_{L_1(\nu)}^p \mathbb{E}[\|\hat{P} - \mathbf{P}[f]\|_{L_p(D)}^p] \\ (29) \quad &\leq C \left( \frac{1 - (2\alpha)^{K+1}}{1 - 2\alpha} \right)^p (\log(n))^{\epsilon} n^{-\delta(s,p,\pi)p}, \end{aligned}$$

where  $C$  depends on  $s, \pi, p, M, \varphi, \psi$ .

To bound the second term in (28) we use the fixed point theorem's approximation. First we have to relate the  $L_p$  norm with the Besov norm. The triangle inequality and Lemma 1 ensure that if  $f$  is in  $\mathcal{G}(s, \pi, M)$  then  $\mathbf{H}_K[G] \star \mathbf{P}[f] - f$  is in  $\mathcal{G}(s, \pi, M)$ . It follows from Sobolev embeddings (25), where  $p > \pi$  and  $s\pi > d$  that

$$\|\mathbf{H}_K[G] \star \mathbf{P}[f] - f\|_{L_p(D)} \leq \|\mathbf{H}_K[G] \star \mathbf{P}[f] - f\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)}.$$

We now use the approximation given by the Banach fixed point theorem

$$\|\mathbf{H}_K[G] \star \mathbf{P}[f] - f\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} \leq (2\alpha)^K \|\mathbf{H}_1[G] \star \mathbf{P}[f] - \mathbf{P}[f]\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)}.$$

After replacing  $\mathbf{H}_1[G] \star \mathbf{P}[f]$  by its expression, using Lemma 1, the triangle inequality and Young's inequality we have

$$\|\mathbf{H}_1[G] \star \mathbf{P}[f] - \mathbf{P}[f]\|_{\mathcal{B}_{\pi\infty}^s(\mathbb{R}^d)} \leq 2\alpha M,$$

which leads to

$$(30) \quad \|\mathbf{H}_K[G] \star \mathbf{P}[f] - f\|_{L_p(D)} \leq (2\alpha)^{K+1} M.$$

We conclude by injecting (29), (30) into (28). The proof of Theorem 1 is now complete.

#### ACKNOWLEDGEMENTS

The author is grateful to F. Comte, O. Lepski and T. Willer for stimulating discussions.

#### REFERENCES

- [1] B. Buchmann and R. Grübel. (2003). Decomposing: an estimation problem for Poisson random sums. *Ann. Statist.*, 31(4):1054–1074.
- [2] Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *The Canadian Journal of Statistics*, 32(2):181–192.
- [3] Butucea, C. and Tsybakov, A. (2008a). Sharp optimality in density deconvolution with dominating bias I. *Theory Proba. Appl.*, 52(1):24–39.
- [4] Butucea, C. and Tsybakov, A. (2008b). Sharp optimality in density deconvolution with dominating bias II. *Theory Proba. Appl.*, 52(2):237–249.
- [5] F. Comte and V. Genon-Catalot. (2011). Estimation for Lévy processes from high frequency data within a long time interval. *Ann. Statist.*, 39(2):803–837.
- [6] Comte, F., Rozenholc, Y. and Taupin M.-L. (2007). Finite sample penalization in adaptive density deconvolution. *J. Stat. Comput. Simul.*, 77(11-12):977–1000.
- [7] Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186.
- [8] Cohen, A. (2003). *Numerical Analysis of wavelet methods*. Studies in Mathematics and its Applications. Vol. 32.
- [9] Comte, F. and Kappus, J. (2014). Density deconvolution from repeated measurements without symmetry assumption on the errors. *Journal of Multivariate Analysis*, 140, 21–46, 2015.
- [10] Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society: Series B*, 73:601–627.
- [11] Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. H. Poincaré Probab. Statist.*, 49(2):569–609.
- [12] Delaigle, A., Hall, P. and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.*, 36(2):665–685.
- [13] Delattre, S., Hoffmann, M., Picard, D. and Vareschi, T. (2012). Blockwise SVD with error in the operator and application to blind deconvolution. *Electron. J. Stat.*, 6:2274–2308.
- [14] Donoho, D.L., Johnstone, I.M., Kerkycharian, G. and Picard, D. (1996). Density estimation by wavelet Thresholding. *The Annals of Statistics*. Vol. 24, No.2, 508–539.

- [15] Duval, C. (2013) Density estimation for compound Poisson processes from discrete data. *Stochastic Processes and their Applications*, 123, 3963–3986.
- [16] Duval, C. (2013). Nonparametric estimation of a renewal reward process from discrete data. *Mathematical Methods of Statistics*, 22(1), 28–56.
- [17] van Es, B., Gugushvili, S. and Spreij, P. (2007). A kernel type nonparametric density estimator for decompounding. *Bernoulli*, 13(3):672–694.
- [18] van Es, B., Gugushvili, S. and Spreij, P. (2008). Deconvolution for an atomic distribution. *Electron. J. Stat.*, 2:265–297.
- [19] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics* , 19(3):1257–1272.
- [20] Gugushvili, S., van Es, B. and Spreij, P. (2011). Deconvolution for an atomic distribution: rates of convergence. *J. Nonparametr. Stat.*, 23(4):1003–1029.
- [21] S. Gugushvili. (2012). Nonparametric inference for discretely sampled Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(1):282–307.
- [22] Härdle, W., Kerkycharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*. Lecture Notes in Statistics, 129. Springer.
- [23] Johannes, J. (2009). Deconvolution with unknown error distribution. *The Annals of Statistics* , 37(5a):2301–2323.
- [24] Johannes, J. and Schwarz, M. (2013). Adaptive circular deconvolution by model selection under unknown error distribution. *Bernoulli* , 19(5A):1576–1611.
- [25] Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution. *Electron. J. Statist.* 8, 2879–2904.
- [26] Kerkycharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well-concentrated bases. *Test*, Vol. 9, No. 2, 283–344.
- [27] M. Lee, H. Shen, C. Burch, and J. S. Marron. (2010). Direct deconvolution density estimation of a mixture distribution motivated by mutation effects distribution. *J. Nonparametr. Stat.*, 22(1-2):1–22.
- [28] Lepski, O.V. and Willer, T. (2015). Lower bounds in the convolution structure density model. *To appear in Bernoulli*.
- [29] Lounici, K. and Nickl, R. (2011). Uniform Risk Bounds and Confidence Bands in Wavelet Deconvolution. *Annals of Statistics*, 39, 201–231.
- [30] Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. Lecture Notes in Statistics. Springer.
- [31] Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics* , 7(4):307–330.
- [32] M. H. Neumann and M. Reiß. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15(1):223–248.
- [33] Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics* , 27(6):2033– 2053.
- [34] Rebelles, G. (2015) Structural adaptive deconvolution under  $L_p$ -losses. *Arxiv preprint*.
- [35] Stefanski, L. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statistics and Probability Letters* , 9(3):229–235.
- [36] Stefanski, S. and Carroll, R. (1990). Deconvoluting kernel density estimators. *Statistics* , 21(2):169–184