DECONVOLUTION WITH UNKNOWN NOISE DISTRIBUTION IS POSSIBLE FOR MULTIVARIATE SIGNALS

BY ÉLISABETH GASSIAT[1], SYLVAIN LE CORFF[2] AND LUC LEHÉRICY[3]

[1]Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France, elisabeth.gassiat@universite-paris-saclay.fr

[2]Samovar, Télécom SudParis, département CITI, TIPIC, Institut Polytechnique de Paris, Palaiseau, France. sylvain.le_corff@telecom-sudparis.eu

[3]Laboratoire J. A. Dieudonné, Université Côte d'Azur, CNRS, 06108, Nice, France, luc.lehericy@univ-cotedazur.fr

Abstract This paper considers the deconvolution problem in the case where the target signal is multidimensional and no information is known about the noise distribution. More precisely, no assumption is made on the noise distribution and no samples are available to estimate it: the deconvolution problem is solved based only on observations of the corrupted signal. We establish the identifiability of the model up to translation when the signal has a Laplace transform with an exponential growth ρ smaller than 2 and when it can be decomposed into two dependent components. Then, we propose an estimator of the probability density function of the signal which is consistent for any unknown noise distribution with finite variance. We also prove rates of convergence and, as the estimator depends on ρ which is usually unknown, we propose a model selection procedure to obtain an adaptive estimator with the same rate of convergence as the estimator with a known tail parameter. This rate of convergence is known to be minimax when $\rho = 1$.

1. Introduction. Estimating the distribution of a signal corrupted by some additive noise, referred to as solving the *deconvolution problem*, is a long-standing challenge in non-parametric statistics. In such problems, the observation $\mathbf{Y}$ is given by

$$\mathbf{Y} = \mathbf{X} + \varepsilon\,, \tag{1}$$

where $\mathbf{X}$ is the signal and ε is the noise. Recovering the distribution of the signal using data contaminated by additive noise is a common problem in all fields of statistics, see [40] and the references therein. It has been applied in a large variety of disciplines and has stimulated a great research interest for instance in signal processing [41, 1], in image reconstruction [33, 8] or in astronomy [46].

Although a great deal of research effort has been devoted to design efficient estimators of the distribution of the signal and to derive optimal convergence rates, the results available in the literature suffer from a crucial limitation: they assume that the distribution of the noise is known. Estimators based on Fourier transforms are the most widespread in this setting as convolution with a known error density translates into a multiplication of the Fourier transform of the signal by the Fourier transform of the noise. However, this assumption may have a significant impact on the robustness of deconvolution estimators as pointed out in [38] where the author established that the mean integrated squared error of such an estimator can grow to infinity when the noise distribution is misspecified.

The aim of this paper is to solve the deconvolution problem without any assumption on the noise distribution and based only on a sample of observations $\mathbf{Y}_1, \dots, \mathbf{Y}_n$. In particular,

MSC 2010 subject classifications: Primary 62G05; secondary 62G07, 62G20, 62H12

Keywords and phrases: deconvolution, identifiability, nonparametric estimation, minimax rates, adaptivity

we do not assume that some samples with the same distribution as ε are available as in [28, 34]. We prove this is possible as soon as the signal $\mathbf{X}$ has a distribution with light enough tails and has at least two dimensions and may be decomposed into two subsets of random variables which satisfy some weak dependency assumption. We then propose an estimator of the density of its distribution which is shown to be minimax adaptive for the mean integrated squared error.

The main reason why it becomes possible to solve the deconvolution problem in this multivariate setting is the structural difference between signal and noise: the signal has dependent components while the noise has independent components. We prove that such a hidden structure may be discovered based only on observations $\mathbf{Y}_1, \dots, \mathbf{Y}_n$. A first step to establish the identifiability in deconvolution without any assumption on the noise was obtained by [24] with a dependency assumption on the signal, but under the restrictive assumption that the signal takes a finite number of values. This identifiability result was extended recently by [21] who proved the identifiability up to translation of the distributions of the signal and of the noise when the hidden signal is a hidden stationary Markov chain independent of the noise. Following these ideas, the first part of our paper establishes the identifiability up to translation of the deconvolution model when the signal $\mathbf{X}$ which lies in $\mathbb{R}^d$, $d \geqslant 2$, can be decomposed into two dependent components $X^{(1)} \in \mathbb{R}^{d_1}$, $d_1 \geqslant 1$, and $X^{(2)} \in \mathbb{R}^{d_2}$, $d_2 \geqslant 1$, with $d_1 + d_2 = d$:

$$\mathbf{Y} = \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix} + \begin{pmatrix} \varepsilon^{(1)} \\ \varepsilon^{(2)} \end{pmatrix} = \mathbf{X} + \boldsymbol{\varepsilon}\,. \tag{2}$$

The identifiability up to translation of the law of $\mathbf{X} \in \mathbb{R}^d$ and of $\boldsymbol{\varepsilon} \in \mathbb{R}^d$ based on the law of $\mathbf{Y}$ when the noise is independent of the signal only requires that the Laplace transform of the signal has an exponential growth smaller than 2 and some dependency assumption between $X^{(1)}$ and $X^{(2)}$.

The second objective of this paper is to propose an estimator of the probability density function of $\mathbf{X}$ which is consistent without any assumptions on the noise distribution provided it has finite variance, and to study the rate of convergence of this estimator. In the pioneering works on deconvolution for i.i.d. data, the distribution of $\mathbf{X}$ is recovered by filtering the received observations to compensate for the convolution using Fourier inversion and kernel based methods, see [17, 37, 47] for some early nonparametric deconvolution methods and [10, 20] for minimax rates. On the other hand, more recent works were dedicated to multivariate deconvolution problems such as [15] for kernel density estimators, [44] for a Bayesian approach or [18] for a multiscale based inference. In all these works, deconvolution is solved under two restrictive assumptions: (a) the distribution of the noise is assumed to be known and (b) this distribution is assumed to be such that its Fourier transform is nowhere vanishing.

An important step toward solving the deconvolution problem without such restrictions on the noise distribution was achieved in [39] for signals in $\mathbb{R}$ with a probability density function supported on a compact subset of $\mathbb{R}$. In [39], the estimation procedure only requires the Fourier transform of the noise to be known on a compact interval around 0. The procedure relies first on recovering as usual the Fourier transform of the signal by direct inversion on the compact interval where the noise distribution is known, and by choosing a polynomial expansion on this compact interval. Then, the Fourier transform is extended to larger intervals before using a Fourier inversion to provide a probability density estimator. Under standard smoothness assumptions, [39] established an upper bound for the mean integrated squared error which is shown to be optimal under a few additional assumptions.

In this paper, we propose an estimation procedure inspired from our identifiability proof. We provide an identification equation on Fourier transforms which can be used to build a contrast function to be minimized over a class of possible estimators of the unknown Fourier transform of the distribution of the signal. Once an estimator of the Fourier transform of the signal in a neighborhood of 0 is available, we use polynomial expansions of this estimator as in [39] to extend it to $\mathbb{R}^{d_1+d_2}$ before using a Fourier inversion to obtain an estimator of the density. To be able to get consistency and rates of convergence, one of the main hurdles to overcome is to relate the value of the contrast function to the error on the Fourier transform. In our opinion, this is far from obvious and it is the most difficult part of our work. Then, under common smoothness assumptions, we obtain consistency and we provide rates of convergence for the estimator of the probability density function of $\mathbf{X}$ depending on the lightness of its tail. Both the regularity and the tail lightness have an impact on the rates of convergence. Surprisingly, while this estimation procedure does not require any prior knowledge on the noise, we obtain the same rates as in [39] when the signal distribution has a compact support: not knowing the noise distribution does not affect these rates. Also, the lower bound proved in [39] applies in this case and the rate of convergence of our estimator is minimax.

We then propose a model selection method to obtain an estimator that is rate adaptive to the unknown lightness of the tail. Minimax rates of convergence in deconvolution problems may be found in [20], [6], [7] and in [40]. In most works on deconvolution, not only the distribution of the noise is assumed to be known (or estimated for instance as in [28] and [34]) but the rates of convergence depend on the decay of its Fourier transform (ordinary or super smooth). It is interesting to note that in our context where the noise is completely unknown, the rate of convergence depends only on the signal and not on the noise.

The paper is organized as follows. Section 2.1 displays the general identifiability result which establishes that the distributions of the signal and of the noise can be recovered from the observations up to a translation indeterminacy. This general result allows to identify submodels as illustrated in Section 2.2 with several common statistical frameworks. Section 3 describes the consistent estimator, the adaptive estimation procedure, and provides convergence rates. Section 4 suggests a few possibilities for future works and settings in which our results may contribute significantly. All proofs are postponed to the appendices.

2. Identifiability results.

2.1. *General theorem.*

The following assumption is assumed to hold throughout the paper.

H1 The signal $\mathbf{X}$ belongs to $\mathbb{R}^d$ with $d \geqslant 2$ and the observation model is given by (2) in which $\boldsymbol{\varepsilon}$ is independent of $\mathbf{X}$ and $\varepsilon^{(1)}$ is independent of $\varepsilon^{(2)}$.

Consider model (2) in which $\boldsymbol{\varepsilon}$ is independent of $\mathbf{X}$ and $\varepsilon^{(1)}$ is independent of $\varepsilon^{(2)}$. Let $\mathbb{P}_{R,Q}$ be the distribution of $\mathbf{Y}$ when $\mathbf{X}$ has distribution R and for $i \in \{1, 2\}$, $\varepsilon^{(i)}$ has distribution $Q^{(i)}$, with $Q = Q^{(1)} \otimes Q^{(2)}$. Denote by $R^{(1)}$ the distribution of $X^{(1)}$ and by $R^{(2)}$ the distribution of $X^{(2)}$. For any $\rho \geqslant 0$ and any integer $p \geqslant 1$, let $\mathcal{M}^p_\rho$ be the set of positive measures μ on $\mathbb{R}^p$ such that there exist $A, B > 0$ satisfying, for all $\lambda \in \mathbb{R}^p$,

$$\int \exp\left(\lambda^\top x\right) \mu(\mathrm{d}x) \leqslant A \exp\left(B\|\lambda\|^\rho\right) ,$$

where for a vector λ in a Euclidian space, $\|\lambda\|$ denotes its Euclidian norm and for any matrix C, $C^\top$ is the transpose matrix of C. When $R \in \mathcal{M}^d_\rho$, the characteristic function of R can be extended into a multivariate analytic function denoted by

$$\begin{aligned}\Phi_R : \mathbb{C}^{d_1} \times \mathbb{C}^{d_2} &\longrightarrow \mathbb{C}\\ (z_1, z_2) &\longmapsto \int \exp\left(iz_1^\top x_1 + iz_2^\top x_2\right) R(\mathrm{d}x_1, \mathrm{d}x_2) .\end{aligned}$$

Note that no assumption other than H1 is made on the noise ε, and that assumption H2 may be understood as a dependency assumption between the components $X^{(1)}$ and $X^{(2)}$ of $\mathbf{X}$ as discussed below.

H2 For any $z_0 \in \mathbb{C}^{d_1}$, $z \mapsto \Phi_R(z_0, z)$ is not the null function and for any $z_0 \in \mathbb{C}^{d_2}$, $z \mapsto \Phi_R(z, z_0)$ is not the null function.

Assumption H2 means that for any $z_1 \in \mathbb{C}^{d_1}$, there exists $z_2 \in \mathbb{C}^{d_2}$ such that $\Phi_R(z_1, z_2) \neq 0$ and for any $z_2 \in \mathbb{C}^{d_2}$, there exists $z_1 \in \mathbb{C}^{d_1}$ such that $\Phi_R(z_1, z_2) \neq 0$.

In the following, the assertion *$R = \widetilde{R}$ and $Q = \widetilde{Q}$ up to translation* means that there exists $m = (m_1, m_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ such that if X has distribution R and for $i \in \{1, 2\}$, ε_i has distribution Q_i, then $(X_i - m_i)_{i \in \{1,2\}}$ has distribution $\widetilde{R}$ and for $i \in \{1, 2\}$, $\epsilon_i + m_i$ has distribution $\widetilde{Q}_i$.

THEOREM 2.1. *Assume that R and $\widetilde{R}$ are probability distributions on $\mathbb{R}^d$ which satisfy assumption H2. Assume also that there exists $\rho < 2$ such that R and $\widetilde{R}$ are in $\mathcal{M}_\rho^d$. Then, $\mathbb{P}_{R,Q} = \mathbb{P}_{\widetilde{R},\widetilde{Q}}$ implies that $R = \widetilde{R}$ and $Q = \widetilde{Q}$ up to translation.*

One way to fix the "up to translation" indeterminacy when the noise has a first order moment is to assume that $\mathbb{E}[\varepsilon] = 0$. The proof of Theorem 2.1 is postponed to Appendix C.

2.1.0.1. Comments on the assumptions of Theorem 2.1.. First of all, Theorem 2.1 involves no assumption at all on the noise distribution. This noise can be deterministic and there is no assumption on the set where its characteristic function vanishes. In addition, there is no density or singularity assumption on the distribution of the hidden signal. The signal may have an atomic or a continuous distribution, and no specific knowledge about this is required. The only assumptions are on the tail of the signal distribution and assumption H2 which, as discussed below, is a dependency assumption.

The assumption that $R \in \mathcal{M}_\rho^d$ is an assumption on the tails of the distribution of $\mathbf{X}$. If R is compactly supported, then $R \in \mathcal{M}_1^d$, and if a probability distribution is in $\mathcal{M}_\rho^d$ for some ρ, then $\rho \geqslant 1$ except in case it is a Dirac mass at point 0. The assumption $\rho < 2$ means that R is required to have tails lighter than that of Gaussian distributions. It is useful to note that R is in $\mathcal{M}_\rho^d$ for some ρ if and only if $R^{(1)}$ is in $\mathcal{M}_\rho^{d_1}$ for some ρ and $R^{(2)}$ is in $\mathcal{M}_\rho^{d_2}$ for some ρ.

Let us now comment assumption H2. Hadamard's factorization theorem states that entire functions are completely determined by their set of zeros up to a multiplicative indeterminacy which is the exponential of a polynomial with degree at most the exponential growth of the function (here ρ). If $R \in \mathcal{M}_\rho$ for some $\rho < 2$, then a consequence of Hadamard's factorization theorem (arguing variable by variable) is that $\Phi_R(\cdot)$ has no complex zeros if and only if $R \in \mathcal{M}_\rho$ is a dirac mass. Since we are interested in non deterministic signals, in general $\Phi_R(\cdot, \cdot)$, $\Phi_R(\cdot, 0)$ and $\Phi_R(0, \cdot)$ will have complex zeros. Now, if the variables $X^{(1)}$ and $X^{(2)}$ are independent, then for all $z_1 \in \mathbb{C}^{d_1}$ and $z_2 \in \mathbb{C}^{d_2}$, $\Phi_R(z_1, z_2) = \Phi_R(z_1, 0)\,\Phi_R(0, z_2)$, so that $\Phi_R(z_1, \cdot)$ is identically zero as soon as z_1 is a complex zero of $\Phi_R(\cdot, 0)$. Thus, assumption H2 implies that the variables $X^{(1)}$ and $X^{(2)}$ are not independent except if they are deterministic. Moreover, if for $i \in \{1, 2\}$, $X^{(i)}$ can be decomposed as $X^{(i)} = \widetilde{X}^{(i)} + \eta_i$, with η_1 and η_2 independent variables independent of $\widetilde{\mathbf{X}} = (\widetilde{X}^{(1)}, \widetilde{X}^{(2)})$, and if for some z_1, $\mathbb{E}[e^{iz_1^\top \eta_1}] = 0$ or for some z_2, $\mathbb{E}[e^{iz_2^\top \eta_2}] = 0$, then H2 does not hold. In other words, H2 can hold only if all the additive noise has been removed from $\mathbf{X}$. Here, additive noise means a random variable with independent components. When the components $X^{(1)}$ and $X^{(2)}$ of the signal have each a finite support set of cardinality 2, Assumption H2 is even equivalent to the fact that $X^{(1)}$ and $X^{(2)}$ are not independent.

Other examples in which assumption H2 holds are provided in Section 2.2, showing that assumption H2 is a mild assumption which may hold for a large class of multivariate signals with dependent components.

2.2. *Identification of structured submodels.* This section displays examples to which Theorem 2.1 applies, and in particular, for each model, we provide conditions which ensure that assumption H2 holds. This means of course that such models are identifiable. But, since they are submodels of the general model, it also means that they may be recovered in this larger general model. Additional examples that could be investigated are discussed in Section 4.

2.2.1. *Noisy Independent Component Analysis.* Independent Component Analysis assumes that $\mathbf{Y} \in \mathbb{R}^d$ is a random vector such that there exist an unknown integer $q \geqslant 1$, an unknown matrix A of size $d \times q$, and two independent random vectors $\mathbf{S} \in \mathbb{R}^q$ and $\boldsymbol{\varepsilon} \in \mathbb{R}^d$ such that

$$\mathbf{Y} = A\mathbf{S} + \boldsymbol{\varepsilon}\,, \tag{3}$$

where all coordinates of the signal $\mathbf{S}$ are independent, centered and with variance one and all coordinates of the noise ε are independent. The statistical challenge is to estimate A and the probability distribution of $\mathbf{S}$ while only Y is observed. The *noise free* formulation of this problem, i.e. $\mathbf{Y} = A\mathbf{S}$, was proposed in the signal processing litterature, see for instance [29]. The identifiability of the noise free linear independent component analysis has been established in [14, 19] under the following (sufficient) conditions.

- The components S_i, $1 \leqslant i \leqslant q$, are not Gaussian random variables (with the possible exception of one component).
- $d \geqslant q$, i.e. the number of observations is greater than the number of independent components.
- The matrix A has full rank.

A noisy extension of the ordinary ICA model which implies further identifiability issues was considered for instance in [41]. A correct identification of the mixing matrix A can be obtained by assuming that the additive noise is Gaussian and independent of the signal sources which are non-Gaussian, see for instance [27]. In our paper, identifiability of the ICA model with unknown additive noise is established using Theorem 2.1 under some assumptions (discussed below). In the following, for any subset I of $\{1, \ldots, d\}$ and any matrix B of size $d \times q$, let B_I denote the $|I| \times q$ matrix whose lines are the lines of B with index in I, where $|C|$ is the number of element of any finite set C.

COROLLARY 2.2. *Let A and $\widetilde{A}$ be two matrices of size $d \times q$. Assume that there exists a partition $I \cup J = \{1, \ldots, d\}$ such that all columns of A_I, $\widetilde{A}_I$, A_J and $\widetilde{A}_J$ are nonzero. Assume also that $(S_j)_{1 \leqslant j \leqslant q}$ (resp. $(\widetilde{S}_j)_{1 \leqslant j \leqslant q}$) are independent and that there exists $\rho < 2$ such that the distributions of all S_j (resp. $\widetilde{S}_j$) are in $\mathcal{M}^1_\rho$. Denote by Q (resp. $\widetilde{Q}$) the distribution of $\boldsymbol{\varepsilon}$ (resp. $\widetilde{\boldsymbol{\varepsilon}}$) and by R (resp. $\widetilde{R}$) the distribution of $A\mathbf{S}$ (resp. $\widetilde{A}\widetilde{\mathbf{S}}$) in (3). Then, $\mathbb{P}_{R,P} = \mathbb{P}_{\widetilde{R},\widetilde{P}}$ implies that $R = \widetilde{R}$ and $Q = \widetilde{Q}$ up to translation.*

Corollary 2.2 is proved in Section D. Apart from the assumption that the independent components of the signal have distribution with light tails, the main assumption is that the observation $\mathbf{Y}$ may be splitted into two known parts so that the corresponding lines of the matrix A have a non zero entry in each column. Although this assumption is not common in the ICA literature, as explained in [43, Section 1.1.3], a wide range of applications require to design source separation techniques to deal with grouped data. Identifiability of such a group structured ICA is likely to rely on specific assumptions and we propose in Corollary 2.2 a set of assumptions which allow to apply Theorem 2.1.

2.2.2. *Repeated measurements.* In deconvolution problems with repeated measurements, the observation model is

$$Y^{(1)} = X^{(1)} + \varepsilon^{(1)} \quad \text{and} \quad Y^{(2)} = X^{(1)} + \varepsilon^{(2)}\,, \tag{4}$$

where $X^{(1)}$ has distribution $R^{(1)}$ on $\mathbb{R}^{d_1}$ and is independent of $\boldsymbol{\varepsilon} = (\varepsilon^{(1)}, \varepsilon^{(2)})^\top$ where $\varepsilon^{(1)}$ is independent of $\varepsilon^{(2)}$ and $\boldsymbol{\varepsilon}$ has distribution Q, see [16] for a detailed description of such models and all the references therein for the numerous applications. Let R be the distribution of $(X^{(1)}, X^{(1)})^\top$ on $\mathbb{R}^{2d_1}$.

COROLLARY 2.3. *Assume that there exists $\rho < 2$ such that $R^{(1)}$ and $\widetilde{R}^{(1)}$ are in $\mathcal{M}_\rho^{d_1}$. Then, $\mathbb{P}_{R,Q} = \mathbb{P}_{\widetilde{R},\widetilde{Q}}$ implies that $R = \widetilde{R}$ and $Q = \widetilde{Q}$ up to translation.*

PROOF. Assumption H2 holds since $\Phi_R(z_1, z_2) = \Phi_{R^{(1)}}(z_1 + z_2)$ for all $z_1 \in \mathbb{C}^{d_1}$ and $z_2 \in \mathbb{C}^{d_1}$, and $\Phi_{R^{(1)}}$ can not be identically zero since $\Phi_{R^{(1)}}(0) = 1$. We then apply Theorem 2.1. □

Therefore, deconvolution with at least two repetitions is identifiable without any assumption on the noise distribution, under the mild assumption that the distribution of the variable of interest has light tails. The model may also contain outliers with unknown probability and still be identifiable.

Corollary 2.3 may be compared to [31, Lemma 1], in which $\mathbf{Y}$ is assumed to have a non vanishing characteristic function, which implies that the characteristic functions of $X^{(1)}$ and of the noise are nowhere vanishing. Identifiability of model (4) has been proved by [35] under the assumption that the characteristic functions of $X^{(1)}$ and of the noise are not vanishing everywhere. In [16], kernel estimators where proved equivalent to those for deconvolution with known noise distribution when $X^{(1)}$ has a real characteristic function and for ordinary smooth errors and signal.

2.2.3. *Errors in variable regression models.* The observations of errors in variable regression models are defined as

$$Y^{(1)} = X^{(1)} + \varepsilon^{(1)} \quad \text{and} \quad Y^{(2)} = g(X^{(1)}) + \varepsilon^{(2)}\,, \tag{5}$$

where $g : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$, $X^{(1)}$ has distribution $R^{(1)}$ on $\mathbb{R}^{d_1}$ and is independent of $\boldsymbol{\varepsilon} = (\varepsilon^{(1)}, \varepsilon^{(2)})^\top$, $\varepsilon^{(1)}$ is independent of $\varepsilon^{(2)}$ and $\boldsymbol{\varepsilon}$ has distribution Q. Let R be the distribution of $(X^{(1)}, g(X^{(1)}))$ on $\mathbb{R}^{d_1+d_2}$. If the distribution of $(X^{(1)}, g(X^{(1)}))$ is identified, then its support is identified and the support of $(X^{(1)}, g(X^{(1)}))$ is the graph of the function g so that g is identified on the support of the distribution of $X^{(1)}$.

COROLLARY 2.4. *Assume that there exists $\rho < 2$ such that $R^{(1)}$ and $\widetilde{R}^{(1)}$ are in $\mathcal{M}_\rho^{d_1}$ and that $R^{(2)}$ and $\widetilde{R}^{(2)}$ are in $\mathcal{M}_\rho^{d_2}$. Assume also that the supports of $X^{(1)}$ and $g(X^{(1)})$ have a nonempty interior and that g is one-to-one on a subset of the support of X_1 with nonempty interior. Then, $\mathbb{P}_{R,Q} = \mathbb{P}_{\widetilde{R},\widetilde{Q}}$ implies that $R = \widetilde{R}$ and $Q = \widetilde{Q}$ up to translation.*

This identifiability relies on weaker assumptions on the errors in variable regression models than in [16] where the noise distribution is assumed to be ordinary-smooth (which implies in particular that its Fourier transform does not vanish on the real line) and where the distribution of $X^{(1)}$ is assumed to have a probability density with respect to the Lebesgue measure on $\mathbb{R}$. In [45], the authors also assumed a nowhere vanishing Fourier transform of the noise distribution and that the distribution of $X^{(1)}$ admits a probability density with respect to the Lebesgue measure uniformly bounded and supported on an open interval. In this setting (more restrictive on the noise and with different restrictions on the signal), the identification result in [45] is not comparable to ours.

PROOF. The proof boils down to establishing that Assumption H2 holds to apply Theorem 2.1. If Assumption H2 does not hold, then either there exists $z_0 \in \mathbb{C}^{d_1}$ such that for all $z \in \mathbb{C}^{d_2}$, $\mathbb{E}[\mathrm{e}^{z_0^\top X^{(1)} + z^\top g(X^{(1)})}] = 0$, or there exists $z_0 \in \mathbb{C}^{d_2}$ such that for all $z \in \mathbb{C}^{d_1}$, $\mathbb{E}[\mathrm{e}^{z^\top X^{(1)} + z_0^\top g(X^{(1)})}] = 0$. In the last case, since the support of $X^{(1)}$ has a nonempty interior, this is equivalent to $\mathbb{E}[\mathrm{e}^{z_0^\top g(X^{(1)})} | X^{(1)}] = 0$, which means that $e^{z_0^\top g(X^{(1)})} = 0$, which is impossible. Thus, since the support of $g(X^{(1)})$ has a nonempty interior (which is the case for instance if g is a continuous function), H2 does not hold if and only if for some z_0, $\mathbb{E}\left[\mathrm{e}^{z_0^\top X^{(1)}} \middle| g(X^{(1)})\right] = 0$. The error in variables regression model is then identifiable without knowing the distribution of the noise as soon as for all z_0,

$$\mathbb{E}\left[\mathrm{e}^{z_0^\top X^{(1)}} \middle| g(X^{(1)})\right] \neq 0 \,. \tag{6}$$

When g is one-to-one on a subset of the support of $X^{(1)}$ with nonempty interior, for all z_0, (6) is verified and the model is identifiable. □

3. Consistent estimation and rates of convergence.

In this section, we propose an estimator of the signal density that is adaptive in the tail parameter ρ and we study its rate of convergence. We first explain in Section 3.1 the construction of the estimator for a fixed tail parameter. We then study in Section 3.2 the consistency and the rates of convergence for the estimators with fixed tail parameter and give an upper bound for the maximum integrated squared error over a class of densities with fixed regularity and tail parameters. We provide in Section 3.3 a model selection method to choose the tail parameter based only on $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ and prove that the resulting estimator is rate adaptive over the previously considered classes of regularity and tail parameters.

Notations. In the following, the unknown distribution of the signal is denoted $R^\star$ and we assume that it admits a density $f^\star$ with respect to the Lebesgue measure. Likewise, the unknown distribution of the noise is written $Q^\star$. For all $h : \mathbb{C}^{d_1} \times \mathbb{C}^{d_2} \to \mathbb{C}$, write $h^{(1)} : (t_1, t_2) \mapsto h(t_1, 0)$ and $h^{(2)} : (t_1, t_2) \mapsto h(0, t_2)$ and for all $h_1 : \mathbb{C}^{d_1} \to \mathbb{C}$, $h_2 : \mathbb{C}^{d_2} \to \mathbb{C}$, write $h_1 \otimes h_2 : (t_1, t_2) \mapsto h_1(t_1) h_2(t_2)$. Define, for any positive integer p and any $\nu > 0$, $\mathsf{B}_\nu^p = [-\nu, \nu]^p$, and write $\mathbf{L}^2(\mathsf{B}_\nu^p)$ the set of square integrable functions on B_ν^p (possibly taking complex values) with respect to the Lebesgue measure. For all $h : \mathbb{C}^{d_1} \times \mathbb{C}^{d_2} \to \mathbb{C}$ and $\nu > 0$, we write $\|h\|_{2,\nu}$ the $\mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$-norm of h, $\|h\|_{1,\nu}$ the $\mathbf{L}^1(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$-norm of h and $\|h\|_{\infty,\nu}$ the $\mathbf{L}^\infty(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$-norm of h. We also write $\|h\|_2$ the $\mathbf{L}^2(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$-norm of h, $\|h\|_1$ the $\mathbf{L}^1(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$-norm of h and $\|h\|_\infty$ the $\mathbf{L}^\infty(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$-norm of h. For any discrete set A, $|A|$ denotes the number of elements in A. For any matrix B, $\|B\|_F$ denotes the Frobenius norm of B. For all $i \in \mathbb{N}^d$, $\|i\|_1 = \sum_{a=1}^d i_a$.

3.1. *Estimation procedure.* The first step of our procedure is to estimate the Fourier transform of $f^\star$. For all $\nu > 0$ and all measurable and bounded functions $\phi : \mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2} \to \mathbb{C}$, define

$$M_\star(\phi;\nu) = \left\| \left(\phi \Phi_{R^\star}^{(1)} \Phi_{R^\star}^{(2)} - \Phi_{R^\star} \phi^{(1)} \phi^{(2)} \right) \Phi_{Q^{\star,(1)}} \otimes \Phi_{Q^{\star,(2)}} \right\|_{2,\nu}^2 ,$$

where $\Phi_{Q^{\star,(1)}}$ (resp. $\Phi_{Q^{\star,(2)}}$) is the Fourier transform of the (unknown) distribution $Q^{\star,(1)}$ of ε_1 (resp. $Q^{\star,(2)}$ of ε_2). This contrast function is inspired by the identifiability proof, see equation (S.4). Indeed, following the identifiability proof, we know that for all $Q^\star$, if $R^\star$ satisfies the assumptions of Theorem 2.1, and if ϕ is a multivariate analytic function satisfying Assumption H2, such that there exist $A, B > 0$ and $\rho \in (0,2)$ such that for all $(z_1, z_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, $|\phi(iz_1, iz_2)| \leqslant A \exp(B\|(z_1,z_2)\|^\rho)$ and such that $\phi(0) = 1$ and for all $z \in \mathbb{R}^d$, $\overline{\phi(z)} = \phi(-z)$, then for any $\nu > 0$,

$$M_\star(\phi;\nu) = 0 \ \text{ if and only if } \ \phi = \Phi_{R^\star} . \tag{7}$$

In practice, $R^\star$ and $Q^\star$ are unknown. Choose first some fixed arbitrary $\nu_{\text{est}} > 0$. The estimator is defined by minimizing an empirical counterpart of $M_\star(\cdot, \nu_{\text{est}})$ over classes of analytic functions to be chosen later. For all $n \geqslant 0$, define

$$M_n(\phi) = \left\| \phi \, \widetilde{\phi}_n^{(1)} \, \widetilde{\phi}_n^{(2)} - \widetilde{\phi}_n \, \phi^{(1)} \, \phi^{(2)} \right\|_{2,\nu_{\text{est}}}^2 ,$$

where for all $(t_1, t_2) \in \mathbb{C}^{d_1} \times \mathbb{C}^{d_2}$,

$$\widetilde{\phi}_n(t_1, t_2) = \frac{1}{n} \sum_{\ell=1}^{n} \mathrm{e}^{it_1^\top Y_\ell^{(1)} + it_2^\top Y_\ell^{(2)}} .$$

For all $i \in \mathbb{N}^d$ and all analytic function ϕ defined on $\mathbb{C}^d$, write $\partial^i \phi$ the partial derivative of order i of ϕ: for all $x \in \mathbb{C}^d$, $\partial^i \phi(x) = \partial_{x_1}^{i_1} \ldots \partial_{x_d}^{i_d} \phi(x)$. For all $\kappa > 0$ and $S < \infty$, let

$$\Upsilon_{\kappa,S} = \Bigg\{ \phi \text{ analytic}; \forall z \in \mathbb{R}^d, \overline{\phi(z)} = \phi(-z), \phi(0) = 1 , \\ \forall i \in \mathbb{N}^d \setminus \{0\}, \left| \frac{\partial^i \phi(0)}{\prod_{a=1}^d i_a!} \right| \leqslant \frac{S^{\|i\|_1}}{\|i\|_1^{\kappa \|i\|_1}} \Bigg\} . \tag{8}$$

Note that for all $\kappa > 0$ and $S < \infty$, the elements of $\Upsilon_{\kappa,S}$ are equal to their Taylor series expansion. As shown in the following lemma, the sets $\Upsilon_{\kappa,S}$ and $\mathcal{M}_{1/\kappa}^d$ are equivalent in that the set of all characteristic functions in $\bigcup_S \Upsilon_{\kappa,S}$ is the set of characteristic functions of probability measures in $\mathcal{M}_{1/\kappa}^d$. Its advantage over $\mathcal{M}_{1/\kappa}^d$ is the more convenient characterization of its elements ϕ in terms of their Taylor expansion.

LEMMA 3.1. *For each $\rho \geqslant 1$ and probability measure $\mu \in \mathcal{M}_\rho^d$, there exists $S > 0$ such that $\lambda \mapsto \int \exp\left(i\lambda^\top x\right) \mu(\mathrm{d}x)$ is in $\Upsilon_{1/\rho,S}$. Conversely, for all $\kappa > 0$, there exists a constant c such that for any $S > 0$ and for any probability measure μ on $\mathbb{R}^d$ such that $\lambda \mapsto \int \exp\left(i\lambda^\top x\right) \mu(\mathrm{d}x)$ is in $\Upsilon_{\kappa,S}$, μ satisfies for all $\lambda \in \mathbb{R}^p$,*

$$\int \exp\left(\lambda^\top x\right) \mu(\mathrm{d}x) \leqslant c \left(1 + (S\|\lambda\|)^{\frac{d+1}{\kappa}}\right) \exp\left(\kappa (S\|\lambda\|)^{1/\kappa}\right) .$$

In particular, $\mu \in \mathcal{M}_{1/\kappa}^d$.

PROOF. The proof is postponed to Appendix E. □

Let now $\mathcal{H}$ be a set of functions $\mathbb{R}^d \to \mathbb{C}^d$ such that all elements of $\mathcal{H}$ satisfy H2 and which is closed in $\mathbf{L}^2(\mathsf{B}^d_{\nu_{\text{est}}})$. For all $\kappa > 0$, $S > 0$, $n \geqslant 1$, the Fourier transform $\Phi_{R^\star}$ of the distribution of $\mathbf{X}$ is estimated by

$$\widehat{\phi}_{\kappa,n} \in \underset{\phi \in \Upsilon_{\kappa,S} \cap \mathcal{H}}{\arg\min}\, M_n(\phi)\,. \tag{9}$$

To address possible measurability issues, note that we could take $\widehat{\phi}_{\kappa,n}$ as a measurable function such that $M_n(\widehat{\phi}_{\kappa,n}) \leqslant \inf_{\phi \in \Upsilon_{\kappa,S} \cap \mathcal{H}} M_n(\phi) + 1/n$, and all the following results would still hold.

Consistency of $\widehat{\phi}_{\kappa,n}$ in $\mathbf{L}^2(\mathsf{B}^{d_1}_\nu \times \mathsf{B}^{d_2}_\nu)$ for any $\nu \in (0, \nu_{\text{est}}]$ will follow from (7) and the compactness of $\Upsilon_{\kappa,S} \cap \mathcal{H}$. An estimator of the density $f^\star$ is then obtained by Fourier inversion. The first step is to truncate the polynomial expansion of $\widehat{\phi}_{\kappa,n}$. For all $m \in \mathbb{N}$, let $\mathbb{C}_m[X_1, \ldots, X_d]$ be the set of multivariate polynomials in d variables with (total) degree m and coefficients in $\mathbb{C}$. In the following, if ϕ is an analytic function defined in a neighborhood of 0 in $\mathbb{C}^d$ written as $\phi : x \mapsto \sum_{i \in \mathbb{N}^d} c_i \prod_{a=1}^d x_a^{i_a}$, define its truncation on $\mathbb{C}_m[X_1, \ldots, X_d]$ as

$$T_m\phi : x \mapsto \sum_{i \in \mathbb{N}^d : \|i\|_1 \leqslant m} c_i \prod_{a=1}^d x_a^{i_a}\,. \tag{10}$$

Then, for some integer $m_{\kappa,n}$ (to be chosen later), the estimator of $f^\star$ is defined as follows:

$$\widehat{f}_{\kappa,n}(x) = \frac{1}{(2\pi)^d} \int_{B^{d_1}_{\omega_{\kappa,n}} \times B^{d_2}_{\omega_{\kappa,n}}} \exp(-it^\top x) \left(T_{m_{\kappa,n}} \widehat{\phi}_{\kappa,n} \right)(t) \mathrm{d}t\,, \tag{11}$$

for some $\omega_{\kappa,n} > 0$ (to be chosen later).

3.2. *Consistency and rates of convergence.* In this section, we explain how to choose $(m_{\kappa,n})_{\kappa,n}$ and $(\omega_{\kappa,n})_{\kappa,n}$ to obtain the rate of convergence of $\widehat{f}_{\kappa,n}$ to $f^\star$ in $\mathbf{L}^2(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$. For any $\kappa \in (1/2, 1]$, define

$$m_{\kappa,n} = \left\lfloor \frac{1}{8\kappa} \frac{\log n}{\log(\log n/4)} \right\rfloor \tag{12}$$

and

$$\omega_{\kappa,n} = c_\omega m_{\kappa,n}^\kappa / S \tag{13}$$

for some constant $c_\omega \leqslant \nu_{\text{est}} \wedge 2\kappa \exp(-(3d+5)/2)$. The following assumption allows to control the regularity of the target density $f^\star$.

H3 We say that $\Phi_{R^\star}$ satisfies H3 for the constants $\beta, c_\beta > 0$ if

$$\int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} |\Phi_{R^\star}(t)|^2 (1 + \|t\|^2)^\beta \mathrm{d}t \leqslant c_\beta\,.$$

For all κ, $S > 0$, $\beta > 0$, $c_\beta > 0$, $\nu > 0$, $c_\nu > 0$ and $c_Q > 0$, consider the following notations.

- $\Psi(\kappa, S, \beta, c_\beta)$ is the set of functions in $\Upsilon_{\kappa,S}$ that can be written as Φ_R for some probability measure R on $\mathbb{R}^d$ and that satisfy H3 for β, c_β.
- $\mathbf{Q}(\nu, c_\nu, c_Q)$ is the class of probability measures of the form $Q^{(1)} \otimes Q^{(2)}$ where $Q^{(1)}$ (resp. $Q^{(2)}$) is a probability measure on $\mathbb{R}^{d_1}$ (resp. $\mathbb{R}^{d_2}$) such that $|\Phi_{Q^{(1)}}| \geqslant c_\nu$ on $[-\nu, \nu]^{d_1}$ and $|\Phi_{Q^{(2)}}| \geqslant c_\nu$ on $[-\nu, \nu]^{d_2}$, and such that if ε is a random variable with distribution Q, then $\mathbb{E}[\|\varepsilon\|^2] \leqslant c_Q$.

THEOREM 3.2. *For all $\kappa \in (1/2,1]$, $S>0$, $\beta>0$ and $c_\beta>0$, for all $\nu>0$, $c_\nu>0$ and $c_Q>0$,*

$$\limsup_{n\to+\infty} \sup_{\substack{Q^\star\in\mathbf{Q}(\nu,c_\nu,c_Q)\\ R^\star:\Phi_{R^\star}\in\Psi(\kappa,S,\beta,c_\beta)\cap\mathcal{H}}} \mathbb{E}_{R^\star,Q^\star}\left[\left(\frac{\log n}{\log\log n}\right)^{2\kappa\beta}\|\widehat{f}_{\kappa,n}-f^\star\|_2^2\right]<+\infty\,,$$

where $\mathcal{H}$ is introduced in the definition of $\widehat{\phi}_{\kappa,n}$, see (9).

For $\kappa=1$, the rate of convergence $(\log n/\log\log n)^{-2\beta}$ obtained in Theorem 3.2 is minimax optimal, see [39] where the situation in which the characteristic function of the noise is known on an open interval is investigated. For the general case of $\kappa\in(1/2,1]$ we conjecture that the rate of convergence $(\log n/\log\log n)^{-2\kappa\beta}$ is minimax optimal. Arguments to support the conjecture are detailed in [23, Section 4].

It is possible to obtain rates of convergence that enjoy uniformity properties in the tail parameter κ. Since such uniformity will be useful to prove adaptive rates of convergence for the adaptation procedure proposed in Section 3.3 (see Theorem 3.5), Theorem 3.2 is deduced as a corollary of the following theorem.

THEOREM 3.3. *For all $\kappa_0\in(1/2,1]$, $S>0$, $\beta>0$ and $c_\beta>0$, for all $\nu>0$, $c_\nu>0$ and $c_Q>0$,*

$$\limsup_{n\to+\infty}\sup_{\kappa\in[\kappa_0,1]} \sup_{\substack{Q^\star\in\mathbf{Q}(\nu,c_\nu,c_Q)\\ R^\star:\Phi_{R^\star}\in\Psi(\kappa,S,\beta,c_\beta)\cap\mathcal{H}}} \mathbb{E}_{R^\star,Q^\star}\left[\sup_{\kappa'\in[\kappa_0,\kappa]}\left\{\left(\frac{\log n}{\log\log n}\right)^{2\kappa'\beta}\|\widehat{f}_{\kappa',n}-f^\star\|_2^2\right\}\right]<+\infty\,,$$

where $\mathcal{H}$ is introduced in the definition of $\widehat{\phi}_{\kappa,n}$, see (9).

PROOF. The proof is postponed to Section A. □

It is important to note that the procedure does not require the knowledge of ν, which leads to the rate of convergence $(\log n/\log\log n)^{-2\kappa\beta}$ without any prior knowledge about the distribution of the noise, since for any $\nu_{\text{est}}>0$, there exists $\nu\in(0,\nu_{\text{est}}]$ and $c_\nu>0$ such that $|\Phi_{Q^{(1)}}|\geqslant c_\nu$ on $[-\nu,\nu]^{d_1}$ and $|\Phi_{Q^{(2)}}|\geqslant c_\nu$ on $[-\nu,\nu]^{d_2}$. Also, the assumption $\Phi_{R^\star}\in\Upsilon_{\kappa^\star,S}$ is not restrictive since by Lemma 3.1, $f^\star\in\mathcal{M}_\rho^d$ implies $\phi^\star\in\Upsilon_{1/\rho,S}$ for some $S>0$. The assumption $\kappa_0>1/2$ is required only to apply Theorem 2.1 and corresponds to the assumption $\rho<2$. If the identifiability theorem held for a wider range of ρ, Theorem 3.3 would be valid for the corresponding range of κ without any change in the proofs.

As a consequence of Theorem 3.2, the estimator is consistent without any assumption on the noise distribution provided it has finite variance.

COROLLARY 3.4. *Assume the noise has finite variance. Then as soon as $\Phi_{R^\star}\in\Psi(\kappa,S,\beta,c_\beta)\cap\mathcal{H}$ for some $\kappa\in(1/2,1]$, $S>0$, $\beta>0$ and $c_\beta>0$, the estimator $\widehat{f}_{\kappa,n}$ is a consistent estimator of $f^\star$ in $\mathbf{L}^2(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})$.*

The proof of Theorem 3.3 can be decomposed into the following steps.

(i) **Consistency.** The fist step consists in proving that there exists a constant c which depends on κ, S, d and ν_{est} such that for all $n\geqslant 1$ and all $x>0$, with probability at least $1-4\mathrm{e}^{-x}$,

$$\sup_{\phi\in\Upsilon_{\kappa,S}}|M_n(\phi)-M_\star(\phi;\nu_{\text{est}})|\leqslant c\left(\sqrt{\frac{1}{n}}\vee\sqrt{\frac{x}{n}}\vee\frac{x}{n}\right)\,.$$

This result is established in Lemma A.1. A key observation will be that for any $\nu \leqslant \nu_{\text{est}}$ and any ϕ,

$$M_\star(\phi;\nu) \leqslant M_\star(\phi;\nu_{\text{est}}) .$$

This is enough to establish that, for any $\nu \leqslant \nu_{\text{est}}$, all convergent subsequences of $(\widehat{\phi}_{\kappa,n})_{n\geqslant 1}$ have limit $\Phi_{R^\star}$ in $\mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$, provided $\Phi_{R^\star} \in \Upsilon_{\kappa,S}$. Since $\Upsilon_{\kappa,S}$ is a compact subset of $\mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$, this implies that $(\widehat{\phi}_{\kappa,n})_{n\geqslant 1}$ is a consistent estimator of $\Phi_{R^\star}$ in $\mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$, uniformly in κ and $R^\star$.

(ii) **Rates for the estimation of $\Phi_{R^\star}$.** Then, for a fixed $\nu \in (0,\nu_{\text{est}}]$, for h in a neighborhood of 0 in $\mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$, the risk $M_\star(\Phi_{R^\star}+h;\nu)$ is lower bounded as follows:

$$M_\star(\Phi_{R^\star}+h;\nu) \geqslant c\|h\|_{2,\nu}^4 , \tag{14}$$

where c depends on d and ν. This result is established in Proposition A.2 in Appendix A.2 and is obtained by decomposing $M_\star(\Phi_{R^\star}+h;\nu)$ into two terms, the first one involving the $\mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$ norm of $h^{(1)}h^{(2)}$ and the second part involving the $\mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$ norm of a linear term in h. The main challenge to prove equation (14) is to establish a lower bound of the first term and an upper bound of the second term for h in a neighorhood of 0 in $\mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$. Obtaining these two bounds requires many technicalities and they need to be balanced sharply to establish (14). Then, we show in Proposition A.3 that there exist constants c_1, c_2 and c_3 which depend on κ_0, ν, S, d and $\mathbb{E}[\|\mathbf{Y}\|^2]$ such that for all $x \geqslant 1$, for all $n \geqslant (1 \vee xc_1)/c_2$, with probability at least $1-4\mathrm{e}^{-x}$,

$$\sup_{\kappa\in[\kappa_0,\kappa^\star]} \|\widehat{\phi}_{\kappa,n} - \Phi_{R^\star}\|_{2,\nu} \leqslant c_3 \left(\sqrt{\frac{x}{n}} \vee \frac{x}{n}\right)^{1/4} . \tag{15}$$

(iii) **Rates for the estimation of $f^\star$.** Then, using assumption H3, the error term $\|\widehat{f}_{\kappa,n} - f^\star\|_2^2$ is upper bounded based on the Fourier inversion (11) as follows

$$\|\widehat{f}_{\kappa,n} - f^\star\|_2^2 \leqslant C\|T_{m_{\kappa,n}}\widehat{\phi}_{\kappa,n} - \Phi_{R^\star}\|_{2,\omega_{\kappa,n}}^2 + \frac{C}{(1+\omega_{\kappa,n}^2)^\beta} .$$

This allows to establish Theorem 3.3 by controlling the error between $T_{m_{\kappa,n}}\widehat{\phi}_{\kappa,n}$ and the truncation of $\phi^\star$ in $\mathbb{C}_{m_{\kappa,n}}[X_1,\ldots,X_d]$ using Legendre polynomials, and the distance between functions in $\Upsilon_{\kappa,S}$ and their truncations in $\mathbb{C}_{m_{\kappa,n}}[X_1,\ldots,X_d]$.

Comments on the practical computation of the estimator. In practice, computing the minimum over the infinite dimensional set defined in (9) requires to introduce a truncation parameter. In other words, instead of minimizing M_n over all elements ϕ of $\Upsilon_{\kappa,S} \cap \mathcal{H}$, we would minimize it over all $T_m\phi$, where m is the so-called truncation parameter. This truncation has no impact on the result proved in Theorem 3.3, i.e. on the rates of convergence derived in this paper, as long as this truncation parameter is chosen sufficiently large with respect to $m_{\kappa,n}$ to obtain the rates for the estimation of $\Phi_{R^\star}$. As observed just after equation (9), the result is an approximate minimizer of M_n. In the case where this new truncation parameter is at least greater than $2m_{\kappa,n}$, this allows in (15) to control the additional bias term and to balance it with the term $(\sqrt{x/n} \vee x/n)^{1/4}$. Although the estimator may be adapted to allow practical computations, this does not ensure a stable and numerically efficient result in real life learning frameworks. Moreover, designing a set $\mathcal{H}$ that is closed in $\mathbf{L}^2([-\nu_{\text{est}},\nu_{\text{est}}]^d)$ and whose elements satisfy H2 that is in addition rich enough for Theorem 3.3 to hold for a wide choice of $R^\star$ is complex and would be a significant practical contribution. Designing an efficient and stable implementation of the proposed algorithm is a challenge on its own and is left for future works, as described in Section 4. The focus of this paper is to derive theoretical properties of the deconvolution estimator without any assumption on the noise distribution.

3.3. *Adaptivity in κ.* In Section 3.2, we studied estimators built using the tail parameter κ. Unfortunately this tail parameter is typically unknown in practice. We now propose a data-driven model selection procedure to choose κ, and we prove that the resulting estimator has a rate corresponding to the largest κ such that $\Phi_{R^\star} \in \Upsilon_{\kappa,S}$ for some $S > 0$.

Our strategy is based on Goldenshluger and Lepski's methodology ([25, 26], see also [5] for a very clear introduction). Like in all model selection problems, the core idea is to perform a careful bias-variance tradeoff to select κ. While a variance bound is readily available thanks to Theorem 3.3, the bias is not so easily accessible. Goldenshluger and Lepski's methodology provides a way to compute a proxy of the bias, thus allowing selection of a proper $\widehat{\kappa}$. The variance bound (which can also be seen as a penalty term) is taken as

$$\sigma_n(\kappa') = c_\sigma \left(\frac{\log n}{\log \log n} \right)^{-\kappa' \beta},$$

for all $\kappa' \in [\kappa_0, 1]$ and for some constant $c_\sigma > 0$. While the selection procedure works as soon as this constant c_σ is large enough, the exact threshold depends on the true parameters. This is a usual problem of selection procedures based on penalization: the penalty is typically known only up to a constant. Approaches such as the slope heuristics or dimension jump heuristics have been proposed to solve this issue and proved to work in several settings, see [3] and references therein. The proxy for the bias is defined for all $\kappa' \in [\kappa_0, 1]$ as

$$A_n(\kappa') = 0 \vee \sup_{\kappa'' \in [\kappa_0, \kappa']} \left\{ \|\widehat{f}_{\kappa'',n} - \widehat{f}_{\kappa',n}\|_2 - \sigma_n(\kappa'') \right\}.$$

Finally, the tail parameter is selected as

$$\widehat{\kappa}_n \in \arg\min_{\kappa' \in [\kappa_0, 1]} \{A_n(\kappa') + \sigma_n(\kappa')\}.$$

When $\Phi_{R^\star} \in \Upsilon_{\kappa,S}$, $\widehat{f}_{\widehat{\kappa}_n,n}$ reaches the same rate of convergence as $\widehat{f}_{\kappa,n}$ for the integrated square risk.

THEOREM 3.5. *For all $\kappa_0 \in (1/2, 1)$, $S > 0$, $\beta > 0$ and $c_\beta > 0$, for all $\nu > 0$, $c_\nu > 0$ and $c_Q > 0$, there exists $c_\sigma > 0$ such that if $\sigma_n(\kappa') \geqslant c_\sigma (\log n / \log \log n)^{-\kappa' \beta}$ for all $\kappa' \in [\kappa_0, 1]$,*

$$\limsup_{n \to +\infty} \sup_{\kappa \in [\kappa_0, 1]} \sup_{\substack{R^\star : \Phi_{R^\star} \in \Psi(\kappa, S, \beta, c_\beta) \cap \mathcal{H} \\ Q^\star \in \mathbf{Q}(\nu, c_\nu, c_Q)}} \left(\frac{\log n}{\log \log n} \right)^{2\kappa\beta} \mathbb{E}_{R^\star, Q^\star} \left[\|\widehat{f}_{\widehat{\kappa}_n,n} - f^\star\|_2^2 \right] < +\infty,$$

where $\mathcal{H}$ is introduced in the definition of $\widehat{\phi}_{\kappa,n}$, see (9).

The proof of Theorem 3.5 is detailed in Section F. It is a consequence of deviation upper bounds developed to prove Theorem 3.3 showing that if $\Phi_{R^\star} \in \Upsilon_{\kappa,S}$, with probability at least $1 - 4/n$, for all $\kappa' \in [\kappa_0, \kappa]$, $\|\widehat{f}_{\kappa',n} - f^\star\|_2 \leqslant \sigma_n(\kappa')$.

4. Conclusion and future works. Recently, in [4], the authors summarized the standard assumptions on the noise distribution and their implications on the minimax risk of the estimator of the signal distribution. In particular, they pointed out that obtaining assumptions under which standard rates of convergence can be established when the Fourier transform of the noise can vanish have not received satisfactory solutions in the existing literature. In the direction of weakening the assumptions on the noise, such limitation has been completely overcome in this paper. The rate of convergence in our setting does not depend at all on the unknown noise. In another direction, it would be interesting to find if it is possible, in the context of unknown noise, to recover noise dependent minimax risk by restricting the set of possible unknown noises. One way could be to make in our methodology $\nu = \nu_{\text{est}}$ go to infinity and to study the square integrated risk with c_ν having a precise decreasing behavior. This can not be directly obtained by the proofs in this work in which we use the fact that ν is finite to derive equation (18) which is itself a basic step to establish Proposition A.3.

There are numerous avenues for future works. We specifically chose to focus on the theoretical properties of the deconvolution estimator obtained from the risk function M_n without assumption on the noise distribution, leaving mainly open the question of designing efficient numerical solutions. Recently, in this unknown noise setting, [21] provided two algorithms to compute nonparametric estimators of the law of the hidden process in a general state space translation model, i.e. when the hidden signal is a Markov chain. More thorough and scalable practical solutions remain to be developed. Although the estimator proposed in this paper enjoys interesting theoretical properties, designing a stable and numerically efficient algorithm remains mainly an open problem.

In a more applied perspective, the recent emergence of blind spot neural networks such as [2] or [32] represent a breakthrough in the field of blind image denoising. In these papers, the authors manage to improve state-of-the-art performance in signal prediction using mainly local (spatially) dependencies on the signal and assuming that the noise components are independent. See also [42]. Our results which in addition do not require any assumption on the noise are likely to provide new architectures or new loss functions to extend such works.

We are particularly interested in applying our results to widespread models such as noisy independent component analysis and nonlinear component analysis, see for instance [30]. As mentionned in [43], a wide range of applications require to design source separation techniques to deal with grouped data and structured signals. The identifiability of such a group structured ICA is likely to rely on specific assumptions similar to the one derived in our paper which should provide new insights to derive numerical procedures. Additive index models studied in [36, 48] could also benefit from this work to weaken the assumptions on the signal and on the functions involved in the mixture defining the observation.

As underlined in Section 2.2, submodels may be identified in the larger general deconvolution model studied in this paper. It could be of interest to study statistical testing of such structured submodels, for instance using the minimax non parametric hypothesis testing theory.

In another line of works referred to as topological data analysis (TDA), see [13], [12], the aim is to provide mathematical results and methods to infer, analyze and exploit the complex topological and geometric structures of the data. Despite fruitful developments, geometric inference from noisy data remains mainly an open problem. Although they appear to be concentrated around geometric shapes, real data are often corrupted by noise and outliers. Quantifying and distinguishing topological/geometric noise, which is difficult to model or unknown, from topological/geometric signal, to infer relevant geometric structures is a subtle problem. Ourwork is likely to be applied to multidimensional signals supported on manifolds and opens the way to find strategies to infer relevant topological and geometric information about signals additively corrupted with totally unknown noise. One way to proceed is to use the distance to measure strategy developed in [11]. It shows that it is possible to build robust methods to estimate geometric parameters of the supports of probability distributions from perturbed versions of it in Wasserstein's metric. This is the subject of an ongoing research project. In particular in [9], it is proved that distributions whose supports are closed regular curves in $\mathbb{R}^2$ satisfy H2.

APPENDIX A: PROOF OF THEOREM 3.3

A.1. Uniform consistency. By definition, for all $R^\star$ and all $Q^\star$ such that $\Phi_{R^\star} \in \Upsilon_{\kappa,S}$,

$$\begin{aligned} M_\star(\widehat{\phi}_{\kappa,n};\nu_{\text{est}}) &\leqslant M_n(\widehat{\phi}_{\kappa,n}) + \sup_{\phi\in\Upsilon_{\kappa,S}} |M_n(\phi) - M_\star(\phi;\nu_{\text{est}})| \,, \\ &\leqslant M_n(\Phi_{R^\star}) + \sup_{\phi\in\Upsilon_{\kappa,S}} |M_n(\phi) - M_\star(\phi;\nu_{\text{est}})| \,, \\ &\leqslant |M_n(\Phi_{R^\star}) - M_\star(\Phi_{R^\star};\nu_{\text{est}})| + \sup_{\phi\in\Upsilon_{\kappa,S}} |M_n(\phi) - M_\star(\phi;\nu_{\text{est}})| \,. \end{aligned} \tag{16}$$

Lemma A.1 provides a control on the deviation $|M_n(\phi) - M_\star(\phi;\nu_{\text{est}})|$ for $\phi \in \Upsilon_{\kappa,S}$.

LEMMA A.1. *For all $S > 0$, there exists $c > 0$ such that for all $\Delta > 0$, $n \geqslant 1$, $x > 0$, and probability measures $R^\star$ and $Q^\star$ on $\mathbb{R}^d$ such that $\mathbb{E}_{R^\star,Q^\star}[\|\mathbf{Y}\|^2] \leqslant \Delta$, with probability at least $1 - 4e^{-x}$ under $\mathbb{P}_{R^\star,Q^\star}$, for all $\kappa' \in [1/2, 1]$,*

$$\sup_{\phi\in\Upsilon_{\kappa',S}} |M_n(\phi) - M_\star(\phi;\nu_{est})| \leqslant c\left[\sqrt{\frac{\Delta}{n}} \vee \sqrt{\frac{x}{n}} \vee \frac{x}{n}\right] .$$

In particular, for all $S > 0$, $\nu \in (0, \nu_{est}]$ and $\Delta > 0$, there exists a constant c such that for all $\kappa \in [1/2, 1]$, $n \geqslant 1$ and $x > 0$,

$$\sup_{\substack{R^\star:\Phi_{R^\star}\in\Upsilon_{\kappa,S}\\ Q^\star:\mathbb{E}_{R^\star,Q^\star}[\|\mathbf{Y}\|^2]\leqslant\Delta}} \mathbb{P}_{R^\star,Q^\star}\left(\sup_{\kappa'\in[1/2,\kappa]} M_\star(\widehat{\phi}_{\kappa',n};\nu) \geqslant c\left(\sqrt{\frac{x}{n}} \vee \frac{x}{n}\right)\right) \leqslant 4e^{-x} \,. \tag{17}$$

PROOF. The proof of the first inequality is postponed to Section G in the supplementary material. The second follows from equation (16) (which requires $\Phi_{R^\star} \in \Upsilon_{\kappa',S}$, hence the assumption $\kappa' \leqslant \kappa$ since the family $(\Upsilon_{\kappa,S})_\kappa$ is nonincreasing in κ), and the observation that for all $\nu \leqslant \nu_{\text{est}}$,

$$M_\star(\widehat{\phi}_{\kappa',n};\nu) \leqslant M_\star(\widehat{\phi}_{\kappa',n};\nu_{\text{est}}) \,.$$

The proof is then completed by taking the supremum over $\kappa' \in [1/2, \kappa]$. □

Since $\sup_{R^\star:\Phi_{R^\star}\in\Upsilon_{\kappa,S}} \mathbb{E}_{R^\star}[\|\mathbf{X}\|^2]$ is bounded by a constant that depends only on κ and S, assuming $\mathbb{E}_{Q^\star}[\|\varepsilon\|^2] \leqslant \widetilde{\Delta}$ and $\Phi_{R^\star} \in \Upsilon_{\kappa,S}$ ensures that $\mathbb{E}_{R^\star,Q^\star}[\|\mathbf{Y}\|^2] \leqslant \Delta$ for some constant Δ depending on S and $\widetilde{\Delta}$. Thus, we may instead use the conditions $\Phi_{R^\star} \in \Upsilon_{\kappa,S}$ and $\mathbb{E}_{Q^\star}[\|\varepsilon\|^2] \leqslant \widetilde{\Delta}$ in equation (17).

For any $\nu > 0$, by the proof of Theorem 2.1 and Lemma 3.1, if $\Phi_{R^\star} \in \Upsilon_{\kappa,S} \cap \mathcal{H}$, the only zero of the contrast function $\phi \mapsto M_\star(\phi;\nu)$ on $\Upsilon_{\kappa,S} \cap \mathcal{H}$ is $\phi = \Phi_{R^\star}$ as soon as $1/\kappa < 2$ since all functions in $\mathcal{H}$ satisfy H2. Moreover, the mapping $(\phi, \Phi_{R^\star}, \Phi_{Q^\star}) \in \mathbf{L}^\infty(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})^3 \mapsto M_\star(\phi;\nu)$ is continuous and for all $\kappa > 0$, $S > 0$ and $\Delta > 0$, the sets $\Upsilon_{\kappa,S}$ and $\{\Phi_Q : Q \text{ s.t. } \mathbb{E}_Q[\|\varepsilon\|^2] \leqslant \Delta\}$ are compact in $\mathbf{L}^\infty(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$ by Arzelà–Ascoli's theorem (the second derivative of Φ_Q is bounded by the second moment of Q and likewise for Φ_R, so these sets are uniformly equicontinuous and all of their elements have value 1 at zero). Thus, for all $S, \nu > 0$, $\kappa_0 \in (1/2, 1]$, $\Delta > 0$ and $\eta > 0$,

$$\inf_{\substack{\phi,\Phi_{R^\star}\in\Upsilon_{\kappa_0,S}\cap\mathcal{H}\\ \|\phi-\Phi_{R^\star}\|_{2,\nu}\geqslant\eta\\ Q^\star:\mathbb{E}_{Q^\star}[\|\varepsilon\|^2]\leqslant\Delta}} M_\star(\phi;\nu) > 0 \,.$$

Let $S > 0$ and $\nu \in (0, \nu_{\mathrm{est}}]$. This equation and Lemma A.1 together with the fact that the family $(\Upsilon_{\kappa,S})_\kappa$ is nonincreasing in κ ensure that for all $\kappa_0 \in (1/2, 1]$, there exists $c > 0$ such that for all $\kappa \in [\kappa_0, 1]$, $\Delta > 0$, $\eta > 0$, $n \geqslant 1$ and $x \in (0, cn]$,

$$\sup_{\substack{R^\star : \Phi_{R^\star} \in \Upsilon_{\kappa,S} \cap \mathcal{H} \\ Q^\star : \mathbb{E}_{Q^\star}[\|\varepsilon\|^2] \leqslant \Delta}} \mathbb{P}_{R^\star, Q^\star} \left(\sup_{\kappa' \in [\kappa_0, \kappa]} \|\widehat{\phi}_{\kappa',n} - \Phi_{R^\star}\|_{2,\nu} \geqslant \eta \right) \leqslant 4e^{-x} . \tag{18}$$

In particular, the family of estimators $(\widehat{\phi}_{\kappa',n})_{\kappa'}$ is $\mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$-consistent uniformly in $\kappa' \in [\kappa, \kappa_0]$, and uniformly in the true parameters $R^\star$ and $Q^\star$.

A.2. Upper bound for the estimator of the Fourier transform of the signal distribution. Recall, for all bounded and measurable functions $h : \mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2} \to \mathbb{C}$, for any $\nu > 0$ and any probability measures $R^\star$ and $Q^\star$ on $\mathbb{R}^d$,

$$\begin{aligned} &M_\star(\Phi_{R^\star} + h; \nu) \\ &= \left\| \left(h\Phi_{R^\star}^{(1)}\Phi_{R^\star}^{(2)} - \Phi_{R^\star} h^{(1)}\Phi_{R^\star}^{(2)} - \Phi_{R^\star}\Phi_{R^\star}^{(1)} h^{(2)} - \Phi_{R^\star} h^{(1)} h^{(2)} \right) \Phi_{Q^{\star,(1)}} \otimes \Phi_{Q^{\star,(2)}} \right\|_{2,\nu}^2 . \end{aligned}$$

In addition, for all $Q \in \mathbf{Q}(\nu, c_\nu, c_Q)$, $\inf_{\mathsf{B}_\nu^{d_1}} |\Phi_{Q^{(1)}}| \wedge \inf_{\mathsf{B}_\nu^{d_2}} |\Phi_{Q^{(2)}}| \geqslant c_\nu$. Using that for all $(a,b) \in \mathbb{R}$, $(a-b)^2 \geqslant a^2/2 - b^2$ and $\|\Phi_{Q^{\star,(1)}}\|_\infty = \|\Phi_{Q^{\star,(2)}}\|_\infty = \|\Phi_{R^\star}\|_\infty = 1$ yields for all probability measures $R^\star$ and $Q^\star$ on $\mathbb{R}^d$ such that $Q^\star \in \mathbf{Q}(\nu, c_\nu, c_Q)$,

$$M_\star(\Phi_{R^\star} + h; \nu) \geqslant c_\nu^4 M^{\mathrm{lin}}(h, \Phi_{R^\star}; \nu)/2 - c_\nu^4 \|h^{(1)} h^{(2)}\|_{2,\nu}^2 , \tag{19}$$

where

$$M^{\mathrm{lin}}(h, \phi; \nu) = \left\| h\phi^{(1)}\phi^{(2)} - \phi h^{(1)}\phi^{(2)} - \phi\phi^{(1)} h^{(2)} \right\|_{2,\nu}^2 . \tag{20}$$

Section B provides an upper bound for $\|h^{(1)}h^{(2)}\|_{2,\nu}^2$ and a lower bound for $M^{\mathrm{lin}}(h, \Phi_{R^\star}; \nu)$ which allows to establish the lower bound given in Proposition A.2. When $\Phi_{R^\star} \in \Upsilon_{\kappa,S}$, the functions h such that $\Phi_{R^\star} + h \in \Upsilon_{\kappa,S}$ belong to the set

$$\mathcal{G}_{\kappa,S} = \{\phi - \phi' : \phi, \phi' \in \Upsilon_{\kappa,S}\} . \tag{21}$$

PROPOSITION A.2. *For all $S, \nu, c_\nu > 0$, there exist $\eta, c > 0$ such that for all $\kappa \in [1/2, 1]$ and all $h \in \mathcal{G}_{\kappa,S}$ such that $\|h\|_{2,\nu} \leqslant \eta$,*

$$\inf_{\substack{R^\star : \Phi_{R^\star} \in \Upsilon_{\kappa,S} \\ Q^\star \in \mathbf{Q}(\nu, c_\nu, +\infty)}} M_\star(\Phi_{R^\star} + h; \nu) \geqslant c\|h\|_{2,\nu}^4 .$$

PROOF. The proof is postponed to Section B. □

Using the above proposition for $\kappa = \kappa_0$ together with equations (17) and (18) is enough to establish Proposition A.3.

PROPOSITION A.3. *For all $\kappa_0 \in (1/2, 1]$, $\nu \in (0, \nu_{est}]$ and $S, c_\nu, c_Q > 0$, there exist $c, c' > 0$ such that for all $n \geqslant 1$, $x \in (0, cn]$ and $\kappa \in [\kappa_0, 1]$,*

$$\inf_{\substack{R^\star : \Phi_{R^\star} \in \Upsilon_{\kappa,S} \cap \mathcal{H} \\ Q^\star \in \mathbf{Q}(\nu, c_\nu, c_Q)}} \mathbb{P}_{R^\star, Q^\star} \left(\sup_{\kappa' \in [\kappa_0, \kappa]} \|\widehat{\phi}_{\kappa',n} - \Phi_{R^\star}\|_{2,\nu} \leqslant c' \left(\sqrt{\frac{x}{n}} \vee \frac{x}{n} \right)^{1/4} \right) \geqslant 1 - 4e^{-x} . \tag{22}$$

A.3. Upper bound for the estimator of the density of the signal distribution. Let $\kappa' \in (0,1]$. Assume H3 holds for the constants β, c_β. Then, by definition of $\widehat{f}_{\kappa',n}$ together with Plancherel's theorem,

$$\begin{aligned}\left\|\widehat{f}_{\kappa',n} - f^\star\right\|_2^2 &= \frac{1}{(4\pi^2)^d}\left\|\mathbb{1}_{\mathsf{B}^{d_1}_{\omega_{\kappa',n}}\times \mathsf{B}^{d_2}_{\omega_{\kappa',n}}} T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n} - \Phi_{R^\star}\right\|_2^2 ,\\ &= \frac{1}{(4\pi^2)^d}\left\|T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n} - \Phi_{R^\star}\right\|_{2,\omega_{\kappa',n}}^2 + \frac{1}{(4\pi^2)^d}\left\|\Phi_{R^\star}(t)\right\|^2_{\mathbf{L}^2((\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})\setminus(\mathsf{B}^{d_1}_{\omega_{\kappa',n}}\times \mathsf{B}^{d_2}_{\omega_{\kappa',n}}))} ,\\ &\leqslant \frac{1}{(4\pi^2)^d}\left\|T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n} - \Phi_{R^\star}\right\|_{2,\omega_{\kappa',n}}^2 + \frac{1}{(4\pi^2)^d}\frac{c_\beta}{(1+\omega^2_{\kappa',n})^\beta} ,\\ &\leqslant c\max\left\{\|T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n} - \Phi_{R^\star}\|^2_{2,\omega_{\kappa',n}} ; (1+\omega^2_{\kappa',n})^{-\beta}\right\} .\end{aligned}$$

for some constant $c>0$. Let $S,\nu>0$ be fixed in the remaining of the proof. For all $i\geqslant 0$, let P_i be the i-th Legendre polynomial and

$$P_i^{\text{norm}} = (i+1/2)^{1/2}\nu^{-1/2}P_i(X/\nu) \tag{23}$$

the normalized i-th Legendre polynomial on $[-\nu,\nu]$. For all positive integer p, define the orthonormal basis $(\mathbf{P}_i^{\text{norm}})_{i\in\mathbb{N}^p}$ of $\mathbb{C}[X_1,\dots,X_p]$ (seen as a subset of $\mathbf{L}^2(\mathsf{B}^p_\nu)$), where for all $i\in\mathbb{N}^p$,

$$\mathbf{P}_i^{\text{norm}}(X_1,\dots,X_p) = (P_{i_1}^{\text{norm}}\otimes\cdots\otimes P_{i_p}^{\text{norm}})(X_1,\dots,X_p) = \prod_{a=1}^p P_{i_a}^{\text{norm}}(X_a) . \tag{24}$$

Since $T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n}$ and $T_{m_{\kappa',n}}\Phi_{R^\star}$ are in $\mathbb{C}_{m_{\kappa',n}}[X_1,\dots,X_d]$, there exists a sequence $(a_i)_{i\in\mathbb{N}^d}$ such that $a_i=0$ if $\|i\|_1 > m_{\kappa',n}$ and $T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n} - T_{m_{\kappa',n}}\Phi_{R^\star} = \sum_{i\in\mathbb{N}^d} a_i\mathbf{P}_i^{\text{norm}}(X)$, where $\mathbf{P}_i^{\text{norm}}$ is defined in (24). By properties of the Legendre polynomials, see [39, page 11], for all $x\in\mathbb{R}$, $|P_i(x)|\leqslant(2|x|+2)^i$ so that $|P_i^{\text{norm}}(x)|\leqslant((2i+1)/(2\nu))^{1/2}(2|x/\nu|+2)^i$. Therefore, for all $i\in\mathbb{N}$,

$$\int_{-\omega_{\kappa',n}}^{\omega_{\kappa',n}}|P_i^{\text{norm}}(x)|^2\mathrm{d}x \leqslant \frac{1}{2}\left(2+2\frac{\omega_{\kappa',n}}{\nu}\right)^{2i+1} ,$$

and by Cauchy-Schwarz inequality,

$$\begin{aligned}&\|T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n} - T_{m_{\kappa',n}}\Phi_{R^\star}\|^2_{2,\omega_{\kappa',n}}\\ &\quad\leqslant \left(\sum_{i\in\mathbb{N}^d,\|i\|_1\leqslant m_{\kappa',n}}\prod_{a=1}^d\int_{-\omega_{\kappa',n}}^{\omega_{\kappa',n}}|P_{i_a}^{\text{norm}}(x)|^2\mathrm{d}x\right)\left(\sum_{i\in\mathbb{N}^d}|a_i|^2\right) ,\\ &\quad\leqslant (m_{\kappa',n}+1)^d 2^{-d}\left(2+2\frac{\omega_{\kappa',n}}{\nu}\right)^{2m_{\kappa',n}+d}\|T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n} - T_{m_{\kappa',n}}\Phi_{R^\star}\|^2_{2,\nu} ,\end{aligned}$$

$$\leqslant m_{\kappa',n}^d\left(2+2\frac{\omega_{\kappa',n}}{\nu}\right)^{2m_{\kappa',n}+d}\|T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n} - T_{m_{\kappa',n}}\Phi_{R^\star}\|^2_{2,\nu} . \tag{25}$$

Since $\Upsilon_{\kappa,S}\subset\Upsilon_{\kappa',S}$ when $\kappa'\leqslant\kappa$, by Lemma H.2 and Lemma H.3 in the supplementary material, when $\Phi_{R^\star}\in\Upsilon_{\kappa,S}$, there exists a constant c such that for all $\kappa'\in[1/2,\kappa]$ and $m_{\kappa',n}\geqslant 2d$,

$$\|\Phi_{R^\star} - T_{m_{\kappa',n}}\Phi_{R^\star}\|^2_{2,\omega_{\kappa',n}} \leqslant (8\omega_{\kappa',n})^d(S\omega_{\kappa',n})^{2m_{\kappa',n}}m_{\kappa',n}^{-2\kappa' m_{\kappa',n}+2d}f_{\kappa'}(S\omega_{\kappa',n})^2$$

$$\leqslant \mathrm{c}\,\omega_{\kappa',n}^{d+2m_{\kappa',n}+2/\kappa'}S^{2m_{\kappa',n}}m_{\kappa',n}^{-2\kappa' m_{\kappa',n}+2d}\exp(2\kappa'(S\omega_{\kappa',n})^{1/\kappa'}) , \tag{26}$$

Since $\Upsilon_{\kappa,S} \subset \Upsilon_{\kappa',S}$ when $\kappa' \leqslant \kappa$, by Lemma H.3 in the supplementary material, when $\Phi_{R^\star} \in \Upsilon_{\kappa,S}$, for all $\kappa' \in [1/2,\kappa]$ and $m_{\kappa',n} \geqslant 2d$,

$$\|\Phi_{R^\star} - T_{m_{\kappa',n}}\Phi_{R^\star}\|^2_{2,\omega_{\kappa',n}} \leqslant (8\omega_{\kappa',n})^d (S\omega_{\kappa',n})^{2m_{\kappa',n}} m_{\kappa',n}^{-2\kappa' m_{\kappa',n}+2d} f_{\kappa'}(S\omega_{\kappa',n})^2,$$

where the function $f_{\kappa'}$ is defined in (S.18), so that by Lemma H.2, there exists a constant c such that for all $\kappa \in [1/2,1]$ such that $\Phi_{R^\star} \in \Upsilon_{\kappa,S}$, for all $\kappa' \in [1/2,\kappa]$ and $m_{\kappa',n} \in \mathbb{N}^*$,

(27)
$$\|\Phi_{R^\star} - T_{m_{\kappa',n}}\Phi_{R^\star}\|^2_{2,\omega_{\kappa',n}} \leqslant \mathrm{c}\,\omega_{\kappa',n}^{d+2m_{\kappa',n}+2/\kappa'} S^{2m_{\kappa',n}} m_{\kappa',n}^{-2\kappa' m_{\kappa',n}+2d} \exp(2\kappa'(S\omega_{\kappa',n})^{1/\kappa'})\,,$$

and likewise

(28)
$$\|\widehat{\phi}_{\kappa',n} - T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n}\|^2_{2,\nu} \leqslant \mathrm{c}(S\nu)^{2m_{\kappa',n}+2/\kappa'} m_{\kappa',n}^{-2\kappa' m_{\kappa',n}+2d}$$

and

(29)
$$\|\Phi_{R^\star} - T_{m_{\kappa',n}}\Phi_{R^\star}\|^2_{2,\nu} \leqslant \mathrm{c}(S\nu)^{2m_{\kappa',n}+2/\kappa'} m_{\kappa',n}^{-2\kappa' m_{\kappa',n}+2d}\,.$$

Write

$$U(\omega_{\kappa',n}) = \mathrm{c}\,\omega_{\kappa',n}^{d+2m_{\kappa',n}+2/\kappa'} S^{2m_{\kappa',n}} m_{\kappa',n}^{-2\kappa' m_{\kappa',n}+2d} \exp(2\kappa'(S\omega_{\kappa',n})^{1/\kappa'})\,,$$
$$U(\nu) = \mathrm{c}(S\nu)^{2m_{\kappa',n}+2/\kappa'} m_{\kappa',n}^{-2\kappa' m_{\kappa',n}+2d}\,.$$

Then, equations (25) to (29) show that for all $\kappa' \in [1/2,\kappa]$,

$$\|T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n} - \Phi_{R^\star}\|^2_{2,\omega_{\kappa',n}} \leqslant 4U(\omega_{\kappa',n}) + 4m_{\kappa',n}^d\left(2 + 2\frac{\omega_{\kappa',n}}{\nu}\right)^{2m_{\kappa',n}+d}\left(2U(\nu) + \|\widehat{\phi}_{\kappa',n} - \Phi_{R^\star}\|^2_{2,\nu}\right)\,,$$

which is controlled by equation (22). Now, choose $\omega_{\kappa',n}$ and $m_{\kappa',n}$ as in (12) and (13), that is

$$\omega_{\kappa',n} = c_\omega m_{\kappa',n}^{\kappa'}/S \quad \text{and} \quad m_{\kappa',n} \leqslant \frac{1}{2\kappa'}\frac{\alpha \log n}{\log(\alpha \log n)}\,,$$

for some $c_\omega \in (0,1]$ and $\alpha > 0$ (note that $\alpha = 1/4$ in equation 12). Proposition A.3 shows that for all $\kappa_0 \in (1/2,1]$, $\nu \in (0,\nu_{\text{est}}]$ and $S, c_\nu, c_Q, \beta, c_\beta > 0$, there exist $c, c' > 0$ such that for all $n \geqslant 1$, $x \in (0, cn]$ and $\kappa \in [\kappa_0, 1]$,

$$\inf_{\substack{R^\star:\,\Phi_{R^\star}\in\Psi(\kappa,S,\beta,c_\beta)\\ Q^\star\in\mathbf{Q}(\nu,c_\nu,c_Q)}} \mathbb{P}_{R^\star,Q^\star}\left(\forall \kappa' \in [\kappa_0,\kappa],\ \|\widehat{f}_{\kappa',n} - f^\star\|_2^2 \leqslant c' \max\left\{m_{\kappa',n}^{-2\kappa'\beta}, \mathrm{e}^{-m_{\kappa',n}} v(x,n)\right\}\right) \geqslant 1 - 4e^{-x}\,,$$

where

$$v(x,n) = 1 \vee \frac{x^{1/4} n^\alpha}{n^{1/4}} \vee \frac{x^{1/2} n^\alpha}{n^{1/2}}\,.$$

Now, when $\alpha \leqslant 1/4$ and $(c_m \log n)/\log\log n \leqslant m_{\kappa,n} \leqslant (C_m \log n)/\log\log n$ for all κ and n for some constants $c_m > 0$ and $C_m > 0$ and take $x = \log n$. It follows that there exists n_0 such that for all $n \geqslant n_0$,

(30)
$$\sup_{\kappa\in[\kappa_0,1]} \inf_{\substack{R^\star:\,\Phi_{R^\star}\in\Psi(\kappa,S,\beta,c_\beta)\\ Q^\star\in\mathbf{Q}(\nu,c_\nu,c_Q)}} \mathbb{P}_{R^\star,Q^\star}\left(\sup_{\kappa'\in[\kappa_0,\kappa]}\left\{m_{\kappa',n}^{2\kappa'\beta}\|\widehat{f}_{\kappa',n} - f^\star\|_2^2\right\} \leqslant c'\right) \geqslant 1 - \frac{4}{n}\,.$$

Finally, note that $m_{\kappa',n}^{2\kappa'\beta}\|\widehat{f}_{\kappa',n} - f^\star\|_2^2 \leqslant (C_M \log n/\log\log n)^{2\beta}\mathrm{diam}(\Upsilon_{\kappa_0,S})^2$ by construction, so that Theorem 3.3 follows.

APPENDIX B: PROOF OF PROPOSITION A.2

By (19), Proposition A.2 may be proved by balancing a lower bound for $M^{\mathrm{lin}}(h,\phi;\nu)$ and an upper bound for $\|h^{(1)}h^{(2)}\|_{2,\nu}^2$. The lower bound on $M^{\mathrm{lin}}(h,\phi;\nu)$ is first obtained for polynomials with known degree m.

LEMMA B.1. *For all $S,\nu>0$, there exists $c>0$ and $C>1$ such that for all $\kappa\in[1/2,1]$, $m\in\mathbb{N}^*$, $\phi\in\Upsilon_{\kappa,S}$ and $h\in\mathcal{G}_{\kappa,S}$,*

$$M^{lin}(T_m h, T_m\phi;\nu)\geqslant c\,m^{-5d-3}C^{-m}\|T_m h\|_{2,\nu}^2\,,$$

where M^{lin}, $\Upsilon_{\kappa,S}$, $\mathcal{G}_{\kappa,S}$ and $T_m\phi$ are defined in (20), (8), (21) and (10).

PROOF. The proof is postponed to Section I in the supplementary material. □

Then, we extend this lower bound to all functions h and ϕ by controlling the difference between h and ϕ and their truncations to degree m.

LEMMA B.2. *For all $S,\nu>0$, there exist $c,c'>0$ and $C>1$ such that for all $\kappa\in[1/2,1]$, $m\in\mathbb{N}^*$, $\phi\in\Upsilon_{\kappa,S}$ and $h\in\mathcal{G}_{\kappa,S}$,*

$$M^{lin}(h,\phi;\nu)\geqslant c\,m^{-5d-3}C^{-m}\|h\|_{2,\nu}^2-c'(S\nu)^{2m}m^{-2\kappa m+2d}\,,$$

where $M^{lin}(h,\phi;\nu)$, $\Upsilon_{\kappa,S}$ and $\mathcal{G}_{\kappa,S}$ are defined in (20), (8) and (21).

PROOF. The proof is postponed to Section J in the supplementary material. □

Finally, a careful choice of m allows to show that $M^{\mathrm{lin}}(h,\phi;\nu)$ is lower bounded by $\|h\|_{2,\nu}^{2+o(1)}$ when $\|h\|_{2,\nu}$ is small enough.

PROPOSITION B.3. *For all $S,\nu>0$, there exist $\eta,\alpha,c>0$ such that for all $\kappa\in[1/2,1]$, $\phi\in\Upsilon_{\kappa,S}$ and $h\in\mathcal{G}_{\kappa,S}$ such that $\|h\|_{2,\nu}\leqslant\eta$,*

$$M^{lin}(h,\phi;\nu)\geqslant c\|h\|_{2,\nu}^2\left(\frac{\log\log(1/\|h\|_{2,\nu})}{\log(1/\|h\|_{2,\nu})}\right)^{5d+3}\|h\|_{2,\nu}^{\frac{\alpha}{\log\log(1/\|h\|_{2,\nu})}}\,,$$

where M^{lin}, $\Upsilon_{\kappa,S}$ and $\mathcal{G}_{\kappa,S}$ are defined in (20), (8) and (21).

PROOF. The proof is postponed to Section J in the supplementary material. □

The upper bound on $\|h^{(1)}h^{(2)}\|_{2,\nu}^2$ is likewise first obtained on polynomials with known degrees m then extended to any function h by controlling the difference between h and its truncation.

LEMMA B.4. *For all $S,\nu>0$, there exists $c>0$ such that for all $\kappa\in[1/2,1]$, $m\in\mathbb{N}^*$ and $h\in\mathcal{G}_{\kappa,S}$,*

$$\|h^{(1)}h^{(2)}\|_{2,\nu}^2\leqslant cm^d(\|h\|_{2,\nu}^4+(S\nu)^{4m}m^{-4\kappa m+4d})\,,$$

where $\mathcal{G}_{\kappa,S}$ is defined in (21).

PROOF. The proof is postponed to Section J in the supplementary material. □

Finally, a careful choice of m shows that this term is upper bounded by $\|h\|_{2,\nu}^{4-o(1)}$ when $\|h\|_{2,\nu}$ is small enough.

PROPOSITION B.5. *For all $S, \nu > 0$, there exist $\eta, c > 0$ such that for all $\kappa \in [1/2, 1]$ and $h \in \mathcal{G}_{\kappa,S}$ such that $\|h\|_{2,\nu} \leqslant \eta$,*

$$\|h^{(1)}h^{(2)}\|_{2,\nu}^2 \leqslant c\left(\frac{\log(1/\|h\|_{2,\nu})}{\log\log(1/\|h\|_{2,\nu})}\right)^d \|h\|_{2,\nu}^4\,,$$

where $\mathcal{G}_{\kappa,S}$ is defined in (21).

PROOF. The proof is postponed to Section J in the supplementary material. □

By Proposition B.3, Proposition B.5 and (19), for all $S, \nu, c_\nu > 0$, there exist constants $\eta, \alpha, \mathrm{c}, \mathrm{c}' > 0$ such that for all $\kappa \in [1/2, 1]$, for all $Q \in \mathbf{Q}(\nu, c_\nu, +\infty)$ and $R^\star$ such that $\Phi_{R^\star} \in \Upsilon_{\kappa,S}$ and for all $h \in \mathcal{G}_{\kappa,S}$ such that $\|h\|_{2,\nu} \leqslant \eta$,

$$M_\star(\Phi_{R^\star} + h; \nu) \geqslant \mathrm{c}\|h\|_{2,\nu}^2 \left(\frac{\log\log(1/\|h\|_{2,\nu})}{\log(1/\|h\|_{2,\nu})}\right)^{5d+3} \|h\|_{2,\nu}^{\overline{\frac{\alpha}{\log\log(1/\|h\|_{2,\nu})}}} - \mathrm{c}'\left(\frac{\log(1/\|h\|_{2,\nu})}{\log\log(1/\|h\|_{2,\nu})}\right)^d \|h\|_{2,\nu}^4\,.$$

Therefore, assuming

(31)
$$\mathrm{c}\left(\frac{\log\log(1/\|h\|_{2,\nu})}{\log(1/\|h\|_{2,\nu})}\right)^{5d+3} \|h\|_{2,\nu}^{\overline{\frac{\alpha}{\log\log(1/\|h\|_{2,\nu})}}} \geqslant 2\mathrm{c}'\left(\frac{\log(1/\|h\|_{2,\nu})}{\log\log(1/\|h\|_{2,\nu})}\right)^d \|h\|_{2,\nu}^2$$

yields

$$M_\star(\phi^\star; \nu) \geqslant \mathrm{c}'\left(\frac{\log(1/\|h\|_{2,\nu})}{\log\log(1/\|h\|_{2,\nu})}\right)^d \|h\|_{2,\nu}^4\,.$$

Note that (31) is implied by

$$\left(\frac{\log\log(1/\|h\|_{2,\nu})}{\log(1/\|h\|_{2,\nu})}\right)^{6d+3} \left(\frac{1}{\|h\|_{2,\nu}}\right)^{2-\overline{\frac{\alpha}{\log\log(1/\|h\|_{2,\nu})}}} \geqslant \frac{2\mathrm{c}'}{\mathrm{c}}\,,$$

which is true as soon as $\|h\|_{2,\nu} \leqslant \eta$ for some $\eta > 0$ depending only on α, c and c'.

SUPPLEMENTARY MATERIAL

Omitted proofs are provided in the supplementary material [22].
().

REFERENCES

[1] ATTIAS, H. and SCHREINER, C. E. (1998). Blind source separation and deconvolution: the dynamic component analysis algorithm. *Neural computation* **10** 1373–1424.

[2] BATSON, J. and ROYER, L. (2019). Noise2Self: Blind Denoising by Self-Supervision. *Proceedings of the 36th International Conference on Machine Learning (ICML).*

[3] BAUDRY, J.-P., MAUGIS, C. and MICHEL, B. (2012). Slope heuristics: overview and implementation. *Statistics and Computing* **22** 455–470.

[4] BELOMESTNY, D. and GOLDENSHLUGER, A. (2019). Density deconvolution under general assumptions on the distribution of measurement errors. *arXiv:1907.11024*.
[5] BERTIN, K., LACOUR, C. and RIVOIRARD, V. (2016). Adaptive pointwise estimation of conditional density function. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **52** 939–980. Institut Henri Poincaré.
[6] BUTUCEA, C. and TSYBAKOV, B. (2008). Sharp optimality in density deconvolution with dominating bias. I. *Theory of Probability and Its Applications* **52** 24-39.
[7] BUTUCEA, C. and TSYBAKOV, B. (2008). Sharp optimality in density deconvolution with dominating bias. II. *Theory of Probability and Its Applications* **52** 237-249.
[8] CAMPISI, P. and EGIAZARIAN, K. (2017). *Blind image deconvolution: theory and applications*. CRC press.
[9] CAPITAO MINICONI, J. (2021). Reconstruction of smooth curves from noisy measurements. Preprint.
[10] CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186. MR997599
[11] CHAZAL, F., COHEN-STEINER, D. and MÉRIGOT, Q. (2011). Geometric Inference for Probability Measures. *Journal on Foundations of Computational Mathematics* **11**.
[12] CHAZAL, F., FASY, B., LECCI, F., MICHEL, B., RINALDO, A., RINALDO, A. and WASSERMAN, L. (2017). Robust topological inference: Distance to a measure and kernel distance. *The Journal of Machine Learning Research* **18** 5845–5884.
[13] CHAZAL, F. and MICHEL, B. (2017). An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*.
[14] COMON, P. (1994). Independent component analysis: a new concept? *Signal Processing* **36** 287–314.
[15] COMTE, F. and LACOUR, C. (2013). Anisotropic adaptive kernel deconvolution. In *Annales de l'IHP Probabilités et statistiques* **49** 569–609.
[16] DELAIGLE, A., HALL, P. and MEISTER, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.* **36** 665–685. MR2396811
[17] DEVROYE, L. (1989). Consistent deconvolution in density estimation. *Canad. J. Statist.* **17** 235–239. MR1033106
[18] ECKLE, K., BISSANTZ, N. and DETTE, H. (2016). Multiscale inference for multivariate deconvolution. *arXiv:1611.05201*.
[19] ERIKSSON, J. and KOIVUNEN, V. (2004). Identifiability, separability, uniqueness of linear ICA models. *IEEE Signal Processing Letters* **11** 601–604.
[20] FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324
[21] GASSIAT, E., LE CORFF, S. and LEHÉRICY, L. (2020). Identifiability and Consistent Estimation of Nonparametric Translation Hidden Markov Models with General State Space. *Journal of Machine Learning Research* **21** 1-40.
[22] GASSIAT, E., LE CORFF, S. and LEHÉRICY, L. (2020). Deconvolution with unknown noise distribution is possible for multivariate signals - Supplement A.
[23] GASSIAT, E., LE CORFF, S. and LEHÉRICY, L. (2020). Deconvolution with unknown noise distribution is possible for multivariate signals. *arXiv:2006.14226*.
[24] GASSIAT, E. and ROUSSEAU, J. (2016). Nonparametric finite translation hidden Markov models and extensions. *Bernoulli* **22** 193-212.
[25] GOLDENSHLUGER, A. and LEPSKI, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** 1150–1190.
[26] GOLDENSHLUGER, A. and LEPSKI, O. (2013). General selection rule from a family of linear estimators. *Theory of Probability & Its Applications* **57** 209–226.
[27] HYVARINEN, A., KARHUNEN, J. and OJA, E. (2002). *Independent Component Analysis*. John Wiley & Sons.
[28] JOHANNES, J. (2009). Deconvolution with unknown error distribution. *The Annals of Statistics* **37** 2301–2323.
[29] JUTTEN, C. (1991). Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. *Signal Processing* **2** 1–10.
[30] KHEMAKHEM, I., KINGMA, D. P., PIO MONTI, R. and HYVARINEN, A. (2020). Variational Autoencoders and Nonlinear ICA: A Unifying Framework. *ArXiv:1907.04809*.
[31] KOTLARSKI, I. (1967). On characterizing the Gamma and the normal distribution. *Pacific Journal of Mathematics* **20** 69–76.
[32] KRULL, A., BUCHHOLZ, T. O. and JUG, F. (2019). Noise2Void - Learning Denoising From Single Noisy Images. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
[33] KUNDUR, D. and HATZINAKOS, D. (1996). Blind image deconvolution. *IEEE signal processing magazine* **13** 43–64.

[34] LACOUR, C. and COMTE, F. (2010). Pointwise deconvolution with unknown error distribution. *Comptes Rendus Mathematique de l'Academie des Sciences* **348** 323–326.

[35] LI, T. and VUONG, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivariate Anal.* **65** 139–165. MR1625869

[36] LIN, W. and KULASEKERA, K. (2007). Identifiability of single-index models and additive-index models. *Biometrika* **94** 496–501.

[37] LIU, M. C. and TAYLOR, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.* **17** 427–438. MR1047309

[38] MEISTER, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Canadian Journal of Statistics* **32** 439–449.

[39] MEISTER, A. (2007). Deconvolving compactly supported densities. *Mathematical Methods of Statistics* **16** 63–76.

[40] MEISTER, A. (2009). *Deconvolution problems in nonparametric statistics*. Springer.

[41] MOULINES, E., CARDOSO, J.-F. and GASSIAT, E. (1997). Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* **5** 3617–3620. IEEE.

[42] OLLION, J., OLLION, C., GASSIAT, E., LEHÉRICY, L. and LE CORFF, S. (2021). Joint self-supervised blind denoising and noise estimation. *arXiv:2102.08023*.

[43] PFISTER, N., WEICHWALD, S., BUHLMANN, B. and SCHOLKOPF, B. (2019). Robustifying Independent Component Analysis by Adjusting for Group-Wise Stationary Noise. *Journal of Machine Learning Research* **20** 1–50.

[44] SARKAR, A., PATI, D., CHAKRABORTY, A., MALLICK, B. K. and CARROLL, R. J. (2018). Bayesian semiparametric multivariate density deconvolution. *Journal of the American Statistical Association* **113** 401–416.

[45] SCHENNACH, S. M. and HU, Y. (2013). Nonparametric identification and semiparametric estimation of classical measurement error models without side information. *J. Amer. Statist. Assoc.* **108** 177–186.

[46] STARCK, J.-L., PANTIN, E. and MURTAGH, F. (2002). Deconvolution in astronomy: A review. *Publications of the Astronomical Society of the Pacific* **114** 1051.

[47] STEFANSKI, L. and CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184. MR1054861

[48] YUAN, M. (2011). On the identifiability of additive index models. *Statistica Sinica* **21** 1901–1911.