Deconvolution with unknown noise distribution is possible for multivariate signals

Élisabeth Gassiat[⋆], Sylvain Le Corff[†], and Luc Lehéricy[‡]

[⋆]Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France.
[†]Samovar, Télécom SudParis, département CITI, TIPIC, Institut Polytechnique de Paris, Palaiseau, France.
[‡]Laboratoire J. A. Dieudonné, Université Côte d'Azur, CNRS, 06100, Nice, France.

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 $\rho = 1$. We conjecture that it remains minimax for $\rho > 1$ and give elements of proof in this direction.

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 nonparametric 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 [Meister, 2009] 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 [Moulines et al., 1997, Attias and Schreiner, 1998], in image reconstruction [Kundur and Hatzinakos, 1996, Campisi and Egiazarian, 2017] or in astronomy [Starck et al., 2002].

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 [Meister, 2004] 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 $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$. In particular, we do not assume that some samples with the same distribution as ε are available as in [Johannes, 2009, Lacour and Comte, 2010]. 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 $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$. A first step to establish the identifiability in deconvolution without any assumption on the noise was obtained by [Gassiat and Rousseau, 2016] 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 [Gassiat et al., 2020a] 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 [Devroye, 1989, Liu and Taylor, 1989, Stefanski and Carroll, 1990] for some early nonparametric deconvolution methods and [Carroll and Hall, 1988, Fan, 1991] for minimax rates. On the other hand, more recent works were dedicated to multivariate deconvolution problems such as [Comte and Lacour, 2013] for kernel density estimators, [Sarkar et al., 2018] for a Bayesian approach or [Eckle et al., 2016] 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 [Meister, 2007] for signals in $\mathbb{R}$ with a probability density function supported on a compact subset of $\mathbb{R}$. In [Meister, 2007], 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, [Meister, 2007] 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 [Meister, 2007] 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 [Meister, 2007] when the signal distribution has a compact support: not knowing the noise distribution does not affect these rates. Also, the lower bound proved in [Meister, 2007] 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. Finally, we provide a conjecture for the lower bound on the minimax rate of convergence that matches the upper bound, as well as elements of proof to justify it. Minimax rates of convergence in deconvolution problems may be found in [Fan, 1991], [Butucea and Tsybakov, 2008a], [Butucea and Tsybakov, 2008b] and in [Meister, 2009]. In most works on deconvolution, not only the distribution of the noise is assumed to be known (or estimated for instance as in [Johannes, 2009] and [Lacour and Comte, 2010]) 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 states the lower bound conjecture on the minimax rates of convergence and Section 5 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 $\mathcal{M}_\rho^p$ 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 $R \in \mathcal{M}_\rho^d$, 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 $\boldsymbol{\varepsilon}$, 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 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 $\mathbb{E}[\boldsymbol{\varepsilon}] = 0$. The proof of Theorem 1 is postponed to Appendix C.

Comments on the assumptions of Theorem 1. First of all, Theorem 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 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 5.

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 $\boldsymbol{\varepsilon}$ 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 [Jutten, 1991]. The identifiability of the noise free linear independent component analysis has been established in [Comon, 1994, Eriksson and Koivunen, 2004] 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 [Moulines et al., 1997]. 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 [Hyvarinen et al., 2002]. In our paper, identifiability of the ICA model with unknown additive noise is established using Theorem 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. *Let A and $\widetilde{A}$ be two matrices of size $d \times q$. Assume that there exists a partition $I \cup J = \{1, \dots, d\}$ 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 $\mathcal{M}_\rho^1$. 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 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 [Pfister et al., 2019, 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 a set of assumptions which allow to apply Theorem 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 [Delaigle et al., 2008] 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 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 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 3 may be compared to [Kotlarski, 1967, 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 [Li and Vuong, 1998] under the assumption that the characteristic functions of $X^{(1)}$ and of the noise are not vanishing everywhere. In [Delaigle et al., 2008], 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 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 [Delaigle et al., 2008] 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 [Schennach and Hu, 2013], 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 [Schennach and Hu, 2013] is not comparable to ours.

Proof. The proof boils down to establishing that Assumption H2 holds to apply Theorem 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. We further study in Section 4 a lower bound of the minimax rate indicating that our final estimator is rate minimax adaptive.

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 (41). Indeed, following the identifiability proof, we know that for all $Q^\star$, if $R^\star$ satisfies the assumptions of Theorem 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

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

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 5. *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, $\mu \in \mathcal{M}^d_{1/\kappa}$.

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

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

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 6. *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 6 is minimax optimal, see [Meister, 2007] 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 [Gassiat et al., 2020b, 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 9), Theorem 6 is deduced as a corollary of the following theorem.

Theorem 7. *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 5, $f^\star \in \mathcal{M}_\rho^d$ implies $\phi^\star \in \Upsilon_{1/\rho,S}$ for some $S > 0$. The assumption $\kappa_0 > 1/2$ is required only to apply Theorem 1 and corresponds to the assumption $\rho < 2$. If the identifiability theorem held for a wider range of ρ, Theorem 7 would be valid for the corresponding range of κ without any change in the proofs.

As a consequence of Theorem 6, the estimator is consistent without any assumption on the noise distribution provided it has finite variance.

Corollary 8. *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 7 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 $1 - 4\mathrm{e}^{-x}$,

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

This result is established in Lemma 17. A key observation will be that for any $\nu \leqslant \nu_{\mathrm{est}}$ and any ϕ,

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

This is enough to establish that, for any $\nu \leqslant \nu_{\mathrm{est}}$, 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 $\nu \in (0, \nu_{\mathrm{est}}]$, 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, ν and c_ν. This result is established in Proposition 18 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 19 that there exist constants c_1, c_2 and c_3 which depend on κ_0, ν, ν_{est}, S, d and $\mathbb{E}[\|\mathbf{Y}\|^2]$ (and c_ν for c_3, and on the compact set $\mathcal{H}$ from equation (9) for c_1 and c_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 7 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 7, 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 7 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 5. 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 ([Goldenshluger and Lepski, 2008, Goldenshluger and Lepski, 2013], see also [Bertin et al., 2016] 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 7, 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 [Baudry et al., 2012] 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

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

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 9. *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 9 is detailed in Section F. It is a consequence of deviation upper bounds developed to prove Theorem 7 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 Lower bounds

In this section, we provide a lower bound showing that the rate of convergence $(\log n/\log\log n)^{-2\kappa\beta}$ obtained in Theorem 7 and in Theorem 9 is minimax optimal. The lower bound in [Meister, 2007] holds for $\kappa = 1$, so in the following we only consider $\kappa \in (0,1)$. In this section (and only this section), we use the notation $\mathcal{F}[h]$ (resp. $\mathcal{F}[Q]$) for the Fourier transform of the probability density function h (resp. the probability measure Q). Our lower bound is stated in Theorem 10.

The proof of Theorem 10 is based on Le Cam's method, also known as the *two-points* method, see [Le Cam, 2012], one of the most widespread technique to derive lower bounds. The minimax risk based on n observations is lower bounded by considering observations from model (3) i.e. assuming that $\mathbf{Y} = A\mathbf{S} + \boldsymbol{\varepsilon}$ where $\mathbf{S} \in \mathbb{R}^d$ with $d = d_1 + d_2$ in which the coordinates S_j, $j = 1, \dots, d$, of $\mathbf{S}$ are independent. Let f_0 and f_n be the probability densities of $A\mathbf{S}$ associated with different choices of densities for the distributions of S_j, $j = 1, \dots, d$, and Q be the distribution of the noise ε. Then, following Le Cam's method, the minimax risk is lower bounded by

$$\frac{1}{4}\|f_0 - f_n\|^2_{\mathbf{L}^2(\mathbb{R}^d)}\left[1 - \frac{1}{2}\left\|(f_0 * Q)^{\otimes n} - (f_n * Q)^{\otimes n}\right\|_{\mathbf{L}^1(\mathbb{R}^d)^n}\right], \tag{16}$$

where $*$ denotes the convolution operator. The goal is then to find two functions f_0 and f_n such that the right most term is greater than 1/2 while the left most term is as large as possible. In this lower bound, we consider a closed set $\mathcal{H}$ of functions from $\mathbb{R}^d$ to $\mathbb{C}^d$ such that all elements of $\mathcal{H}$ satisfy H2 and which contains the probability densities of the form given by f_0 and f_n. This is the starting point of the proof of Theorem 10 which also relies on a technical conjecture (Conjecture 12) which is strongly supported by numerical experiments, see Section L in the supplementary material.

Theorem 10. *Assume that Conjecture 12 is true. Then for all $\kappa \in (0,1)$, $\beta > 0$, $c_\beta > 0$, $c_Q > 0$ and $\nu > 0$, there exists a constant $c > 0$ such that*

$$\inf_{\widehat{f}} \sup_{\substack{R^\star : \Phi_{R^\star} \in \Psi(\kappa^\star, S, \beta, c_\beta) \cap \mathcal{H} \\ Q^\star \in \mathbf{Q}(\nu, c_\nu, c_Q)}} \mathbb{E}_{R^\star, Q^\star}\left[\|\widehat{f}_{\kappa,n} - f^\star\|^2_{\mathbf{L}^2(\mathbb{R}^d)}\right] \geqslant c\left(\frac{\log n}{\log\log n}\right)^{-2\kappa\beta}.$$

The infimum is taken over all estimators $\widehat{f}$, that is all measurable functions of $\mathbf{Y}_1, \dots, \mathbf{Y}_n$.

Let a be a (small) real number, in the following it is assumed that A is the $(d_1 + d_2) \times (d_1 + d_2)$ matrix

$$A = \left(\begin{array}{ccccc|cccc} 1 & 0 & 0 & \cdots & 0 & a & a & \cdots & a \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & \ddots & & \vdots & & \ddots & \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ \hline a & a & a & \cdots & a & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & & \vdots & & \ddots & \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{array}\right) .$$

Assume that the coordinates of ε are independent identically distributed with density (see [Ehm et al., 2004])

$$g : x \mapsto c_g \frac{1 + \cos(cx)}{(\pi^2 - (cx)^2)^2}$$

for some $c > 0$, where c_g is such that g is a probability density, with characteristic function

$$\mathcal{F}[g] : t \mapsto \left[\left(1 - \left|\frac{t}{c}\right|\right) \cos\left(\pi \frac{t}{c}\right) + \frac{1}{\pi} \sin\left(\pi \left|\frac{t}{c}\right|\right) \right] \mathbf{1}_{-c \leqslant t \leqslant c} .$$

With an adequate choice of c, $Q \in \mathbf{Q}(\nu, c_\nu, c_Q)$. Consider the probability density $u : x \in \mathbb{R} \mapsto c_u \cdot \exp(-1/(1 - x^2))\mathbf{1}_{[-1,1]}(x)$ with the appropriate $c_u > 0$ so that the integral of u equals one. For all $b > 0$ and $x \in \mathbb{R}$, write $u_b(x) = bu(bx)$.

Lemma 11. *Let $\kappa \in (0, 1)$, $c, T > 0$ and $\tau \geqslant 0$. Then, there exists $x_0 > 0$ such that the following holds. Let $h_\kappa = c_h \exp(-(\sqrt{[1 + (x/x_0)^2]/2})^{1/(1-\kappa)})$ where c_h is such that h_κ is a probability density. For all $b \geqslant 1/x_0$, any probability density ζ such that $\zeta \leqslant c([x \mapsto h_\kappa(x)(1+(x/x_0)^2)^\tau] * u_b)$ satisfies $\mathcal{F}[\zeta] \in \Upsilon_{\kappa,T}$.*

Proof. The proof is postponed to Appendix K in the supplementary material. □

Let x_0 and h_κ be as in Lemma 11. Since h_κ is infinitely differentiable with square integrable derivatives, for all $\beta > 0$, there exists L such that for all $b > 0$, $\int |\mathcal{F}[h_\kappa * u_b](t)|^2 (1 + t^2)^\beta \mathrm{d}t \leqslant L$. Let $(P_K)_{K \geqslant 0}$ be the family of orthonormal polynomials for the scalar product $\langle f, g \rangle = \int f(x)g(x)h_\kappa(x)^2 \mathrm{d}x$ such that $\deg(P_K) = K$. Consider the following conjecture on the properties on these polynomials.

Conjecture 12. *There exists an nonnegative envelope function F_{env} that has at most polynomial growth at infinity such that the family $(P_K)_{K \geqslant 1}$ satisfies $\sup_{K \geqslant 1} K^{(1-\kappa)/2} \|P_K h_\kappa / F_{env}\|_\infty < \infty$ and there exists constants c_0, c_1, c_2 such that for all K large enough, there exists at least $c_0 K^\kappa$ intervals of length at least $c_1 K^{-\kappa}$ in $[-1, 1]$ on which $|P_K h_\kappa| \geqslant c_2 K^{(\kappa-1)/2}$.*

Let us comment on the different elements of this conjecture. As discussed above, our objective is to construct two probability densities f_0 and f_n that are as far from each other as possible while the resulting distributions of $\mathbf{Y}$ are as close as possible, see equation (16). The boundedness of $P_K h_\kappa / F_{\text{env}}$ ensures that the densities we construct are nonnegative, and the assumption on the intervals is used to prove Corollary 13, which controls how large $\|f_0 - f_n\|$ is.

For the sake of simplicity, assume in the following that $\|F_{\text{env}} h_\kappa\|_{\mathbf{L}^1(\mathbb{R})} = 1$. Then there exists $c > 0$ and $\tau \geqslant 0$ such that $F_{\text{env}}(x) \leqslant c(1 + (x/x_0)^2)^\tau$ for all $x \in \mathbb{R}$, thus making it possible to use Lemma 11. To use the lemma, it is important to note that this c does not depend on the choice of x_0.

Another conjecture that gives a better idea of the behaviour of these functions is that there exists a shape function F_{shape} such that

$$\sup_{K \geqslant 1} \left\| x \mapsto \frac{(P_K h_\kappa)(x)}{K^{(\kappa-1)/2} F_{\text{shape}}(K^{\kappa-1}x)} \right\|_\infty < \infty \,. \tag{17}$$

This function F_{shape} diverges at x_0 and $-x_0$ for some finite $x_0 \geqslant 1$, as illustrated in Figure L.1 of Section L. As K grows, the peak of $P_K h_\kappa$ comes closer to this divergence point, but slowly enough that F_{env} only grows polynomially.

Corollary 13. *Assume Conjecture 12 is true, then there exist c_b, c_3, c_4 such that for K large enough, for all $b \geqslant c_b K^\kappa$,*

$$c_3 K^{\kappa-1} \geqslant \|P_K h_\kappa^2\|^2_{\mathbf{L}^2(\mathbb{R})} \geqslant \|(P_K h_\kappa^2) * u_b\|^2_{\mathbf{L}^2(\mathbb{R})} \geqslant c_4 K^{\kappa-1} \,.$$

Proof. The proof is postponed to Appendix K in the supplementary material. □

Note that in the limit $\kappa = 1$, h_κ is the indicator function of $[-1, 1]$, and the orthonormal polynomials P_K are the (normalized) Legendre polynomials. In this setting, Conjecture 12 (and therefore Equation 17) have been proved with $F_{\text{env}} = F_{\text{shape}} = \mathbf{1}_{[-1,1]}$, see Lemma 1 of [Meister, 2007].

In the limit $\kappa = 1/2$, h_κ is a normal density, and the functions $P_K h_\kappa$ are the Hermite functions. Approximations of Hermite funtions close to zero and near the turning points are known and corroborate our conjecture, see for instance [Boyd, 2018, Section A.11]: the behaviour near zero is approximately a trigonometric function times a shape function, validating equation (17) (near zero) and the second part of Conjecture 12. Near the turning points, they are best approximated by Airy functions with a scaling corresponding to $F_{\text{env}}(x) = O(x^{1/3})$.

Let $(\alpha_n)_{n\geqslant 1}$ be a sequence of nonnegative real numbers with limit zero, $(K_n)_{n\geqslant 1}$ a sequence of integers tending to infinity and $(b_n)_{n\geqslant 1}$ a sequence of real numbers tending to infinity. Define f_0 as the density of $\mathbf{X}$ when for all $1 \leqslant j \leqslant d$, $s_j = \zeta_0 = (F_{\text{env}} h_\kappa) * u_b$, and f_n as the density of $\mathbf{X}$ when S_1 has density

$$\zeta_n = (F_{\text{env}} h_\kappa + \alpha_n P_{K_n} h_\kappa^2) * u_{b_n} = \zeta_0 + \alpha_n (P_{K_n} h_\kappa^2) * u_{b_n} \tag{18}$$

and $S_2, \ldots, S_d$ have density ζ_0. The function ζ_n is nonnegative as soon as $\alpha_n \leqslant (\|P_{K_n} h_\kappa / F_{\text{env}}\|_\infty)^{-1}$, which is of order $K_n^{(1-\kappa)/2}$ by Conjecture 12. Its integral equals one for $K_n \geqslant 1$ since by definition the function $P_{K_n} h_\kappa^2$ is orthogonal to P_0 (which is a constant function) in $\mathbf{L}^2(\mathbb{R})$, so that the integral of $P_{K_n} h_\kappa^2 * u_{b_n}$ is zero. Therefore, ζ_n is a probability density. In addition, $\mathcal{F}[\zeta_0] \in \Upsilon_{\kappa,T}$ and $\mathcal{F}[\zeta_n] \in \Upsilon_{\kappa,T}$ follow immediately from Lemma 11.

Lemma 14. *The probability densities f_0 and f_n are in $\Upsilon_{\kappa,T}$.*

Proof. The proof is postponed to Appendix K in the supplementary material. □

Lemma 15. *For all $\kappa \in (0, 1]$, $\beta > 0$ and $c_\beta > 0$, there exist $x_0 > 0$ and $c_h > 0$ such that $\mathcal{F}[f_0]$ and $\mathcal{F}[f_n]$ belong to $\Psi(\kappa, T, \beta, c_\beta)$ as soon as the two following assumptions are met:*

$$\alpha_n \leqslant \|P_{K_n} h_\kappa\|_\infty^{-1} \,, \tag{19}$$

$$\alpha_n^2 \|P_{K_n} h_\kappa^2\|^2_{\mathbf{L}^2(\mathbb{R})} \leqslant c_h b_n^{-2\beta} \,. \tag{20}$$

Proof. The proof is postponed to Appendix K in the supplementary material. □

Following [Meister, 2007], it is straightforward to establish that

$$1 - \frac{1}{2}\|(f_0 * Q)^{\otimes n} - (f_n * Q)^{\otimes n}\|_{\mathbf{L}^1(\mathbb{R}^d)^n} \geqslant \left(1 - \frac{1}{2}\|(f_0 * Q) - (f_n * Q)\|_{\mathbf{L}^1(\mathbb{R}^d)}\right)^n .$$

Then, by (16), the minimax risk based on n observations is lower bounded by $c\|f_0 - f_n\|^2_{\mathbf{L}^2(\mathbb{R}^d)}$ for some constant $c > 0$ if $(\alpha_n)_{n\geqslant 1}$, $(b_n)_{n\geqslant 1}$ and $(K_n)_{n\geqslant 1}$ are chosen such that

$$\int_{\mathbb{R}^d} |(f_0 * Q)(x) - (f_n * Q)(x)| \,\mathrm{d}x = O\left(\frac{1}{n}\right) . \tag{21}$$

Lemma 16. *Assume that* (20) *holds and that*

$$K_n = \frac{c_h}{\kappa}\left(\frac{\log n}{\log\log n}\right) . \tag{22}$$

Then, (21) *holds.*

Proof. The proof is postponed to Appendix K. □

Therefore, the minimax risk based on n observations is lower bounded by $c\|f_0 - f_n\|^2_{\mathbf{L}^2(\mathbb{R}^d)}$ for some constant $c > 0$. In addition, by definition of f_0 and f_n, for all $u \in \mathbb{R}^d$,

$$f_0(u) = \mathrm{Det}(A)^{-1}\prod_{j=1}^d \zeta_0((A^{-1}u)_j) \quad \text{and} \quad f_n(u) = \mathrm{Det}(A)^{-1}\zeta_n((A^{-1}u)_1)\prod_{j=2}^d \zeta_0((A^{-1}u)_j) .$$

Therefore, there exists a constant $c > 0$ such that,

$$\|f_0 - f_n\|^2_{\mathbf{L}^2(\mathbb{R}^d)} \geqslant c\|\zeta_0\|^{2(d-1)}_{\mathbf{L}^2(\mathbb{R})}\|\zeta_0 - \zeta_n\|^2_{\mathbf{L}^2(\mathbb{R})} \geqslant c\alpha_n^2 K_n^{\kappa-1} ,$$

by Corollary 13 and (20). Then, choosing $b_n = c_b K_n^{\kappa}$, $\alpha_n^2 \propto K_n^{-2\kappa\beta}/\|P_{K_n} h_\kappa^2\|^2_{\mathbf{L}^2(\mathbb{R})} \propto K_n^{1-\kappa-2\kappa\beta}$ (by Corollary 13) and K_n as in (22) yields

$$\|f_0 - f_n\|^2_{\mathbf{L}^2(\mathbb{R}^d)} \geqslant c b_n^{-2\beta} \geqslant c K_n^{-2\kappa\beta} \geqslant c\left(\frac{\log n}{\log\log n}\right)^{-2\kappa\beta} .$$

The condition $\alpha_n \leqslant (\|P_{K_n} h_\kappa / F_{\mathrm{env}}\|_\infty)^{-1}$ corresponds to $K_n^{(1-\kappa)/2-\kappa\beta} = O(K_n^{(1-\kappa)/2})$, which is always true.

5 Conclusion and future works

Recently, in [Belomestny and Goldenshluger, 2019], 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 (25) which is itself a basic step to establish Proposition 19.

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, [Gassiat et al., 2020a] 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 [Batson and Royer, 2019] or [Krull et al., 2019] 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 [Ollion et al., 2021]. 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 [Khemakhem et al., 2020]. As mentionned in [Pfister et al., 2019], 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 [Lin and Kulasekera, 2007, Yuan, 2011] 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 [Chazal and Michel, 2017], [Chazal et al., 2017], 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 [Chazal et al., 2011]. 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 [Capitao Miniconi, 2021], it is proved that distributions whose supports are closed regular curves in $\mathbb{R}^2$ satisfy H2.

A Proof of Theorem 7

A.1 Uniform consistency

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

$$\begin{aligned} M_\star(\widehat{\phi}_{\kappa,n};\nu_{\text{est}}) &\leqslant M_n(\widehat{\phi}_{\kappa,n}) + \sup_{\phi\in\Upsilon_{\kappa,S}} |M_n(\phi) - M_\star(\phi;\nu_{\text{est}})| \;, \\ &\leqslant M_n(\Phi_{R^\star}) + \sup_{\phi\in\Upsilon_{\kappa,S}} |M_n(\phi) - M_\star(\phi;\nu_{\text{est}})| \;, \\ &\leqslant |M_n(\Phi_{R^\star}) - M_\star(\Phi_{R^\star};\nu_{\text{est}})| + \sup_{\phi\in\Upsilon_{\kappa,S}} |M_n(\phi) - M_\star(\phi;\nu_{\text{est}})| \;. \end{aligned} \tag{23}$$

Lemma 17 provides a control on the deviation $|M_n(\phi) - M_\star(\phi;\nu_{\text{est}})|$ for $\phi \in \Upsilon_{\kappa,S}$.

Lemma 17. *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 $x > 0$,

$$\sup_{\substack{R^\star:\Phi_{R^\star}\in\Upsilon_{\kappa,S}\\ Q^\star:\mathbb{E}_{R^\star,Q^\star}[\|\mathbf{Y}\|^2]\leqslant\Delta}} \mathbb{P}_{R^\star,Q^\star}\left(\sup_{\kappa'\in[1/2,\kappa]} M_\star(\widehat{\phi}_{\kappa',n};\nu) \geqslant c\left(\sqrt{\frac{x}{n}} \vee \frac{x}{n}\right)\right) \leqslant 4e^{-x} . \tag{24}$$

Proof. The proof of the first inequality is postponed to Section G in the supplementary material. The second follows from equation (23) (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 (24).

For any $\nu > 0$, by the proof of Theorem 1 and Lemma 5, 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 $\nu \in (0, \nu_{\text{est}}]$. This equation and Lemma 17 together with the fact that the family $(\Upsilon_{\kappa,S})_\kappa$ is nonincreasing in κ ensure that for all $\eta > 0$, $\kappa_0 \in (1/2, 1]$ and $\Delta > 0$, there exists $c > 0$ such that for all $\kappa \in [\kappa_0, 1]$, $n \geqslant 1$ and $x \in (0, cn]$,

$$\sup_{\substack{R^\star : \Phi_{R^\star} \in \Upsilon_{\kappa,S} \cap \mathcal{H} \\ Q^\star : \mathbb{E}_{Q^\star}[\|\varepsilon\|^2] \leqslant \Delta}} \mathbb{P}_{R^\star, Q^\star} \left(\sup_{\kappa' \in [\kappa_0, \kappa]} \|\widehat{\phi}_{\kappa',n} - \Phi_{R^\star}\|_{2,\nu} \geqslant \eta \right) \leqslant 4e^{-x} . \tag{25}$$

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 $\mathbb{R}^d$,

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

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 $Q^\star \in \mathbf{Q}(\nu, c_\nu, c_Q)$,

$$M_\star(\Phi_{R^\star} + h; \nu) \geqslant c_\nu^4 M^{\text{lin}}(h, \Phi_{R^\star}; \nu)/2 - c_\nu^4 \|h^{(1)}h^{(2)}\|_{2,\nu}^2 , \tag{26}$$

where

$$M^{\text{lin}}(h, \phi; \nu) = \left\| h\phi^{(1)}\phi^{(2)} - \phi h^{(1)}\phi^{(2)} - \phi\phi^{(1)}h^{(2)} \right\|_{2,\nu}^2 . \tag{27}$$

Section B provides an upper bound for $\|h^{(1)}h^{(2)}\|_{2,\nu}^2$ and a lower bound for $M^{\text{lin}}(h, \Phi_{R^\star}; \nu)$ which allows to establish the lower bound given in Proposition 18. 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

$$\mathcal{G}_{\kappa,S} = \{\phi - \phi' : \phi, \phi' \in \Upsilon_{\kappa,S}\} . \tag{28}$$

Proposition 18. *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 (24) and (25) is enough to establish Proposition 19.

Proposition 19. *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 $\kappa \in [\kappa_0, 1]$,*

$$\inf_{\substack{R^\star : \Phi_{R^\star} \in \Upsilon_{\kappa,S} \cap \mathcal{H} \\ Q^\star \in \mathbf{Q}(\nu, c_\nu, c_Q)}} \mathbb{P}_{R^\star, Q^\star} \left(\sup_{\kappa' \in [\kappa_0, \kappa]} \|\widehat{\phi}_{\kappa',n} - \Phi_{R^\star}\|_{2,\nu} \leqslant c' \left(\sqrt{\frac{x}{n}} \vee \frac{x}{n} \right)^{1/4} \right) \geqslant 1 - 4e^{-x} . \tag{29}$$

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,

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

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

$$P_i^{\text{norm}} = (i+1/2)^{1/2}\nu^{-1/2}P_i(X/\nu) \tag{30}$$

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 $\mathbb{C}[X_1,\ldots,X_p]$ (seen as a subset of $\mathbf{L}^2(\mathsf{B}^p_\nu)$), where for all $i \in \mathbb{N}^p$,

$$\mathbf{P}_i^{\text{norm}}(X_1,\ldots,X_p) = (P_{i_1}^{\text{norm}}\otimes\cdots\otimes P_{i_p}^{\text{norm}})(X_1,\ldots,X_p) = \prod_{a=1}^p P_{i_a}^{\text{norm}}(X_a) . \tag{31}$$

Since $T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n}$ and $T_{m_{\kappa',n}}\Phi_{R^\star}$ are in $\mathbb{C}_{m_{\kappa',n}}[X_1,\ldots,X_d]$, 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 (31). By properties of the Legendre polynomials, see [Meister, 2007, 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,

$$\begin{aligned}\|T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n} - &T_{m_{\kappa',n}}\Phi_{R^\star}\|^2_{2,\omega_{\kappa',n}}\\ &\leqslant \left(\sum_{i\in\mathbb{N}^d, \|i\|_1\leqslant m_{\kappa',n}}\prod_{a=1}^d\int_{-\omega_{\kappa',n}}^{\omega_{\kappa',n}}|P_{i_a}^{\text{norm}}(x)|^2\mathrm{d}x\right)\left(\sum_{i\in\mathbb{N}^d}|a_i|^2\right) ,\\ &\leqslant (m_{\kappa',n}+1)^d 2^{-d}\left(2 + 2\frac{\omega_{\kappa',n}}{\nu}\right)^{2m_{\kappa',n}+d}\|T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n} - T_{m_{\kappa',n}}\Phi_{R^\star}\|^2_{2,\nu} ,\\ &\leqslant m^d_{\kappa',n}\left(2 + 2\frac{\omega_{\kappa',n}}{\nu}\right)^{2m_{\kappa',n}+d}\|T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n} - T_{m_{\kappa',n}}\Phi_{R^\star}\|^2_{2,\nu} .\end{aligned} \tag{32}$$

Since $\Upsilon_{\kappa,S} \subset \Upsilon_{\kappa',S}$ when $\kappa' \leqslant \kappa$, by Lemma 28 in the 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 (55), so that by Lemma 27, 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 $m_{\kappa',n} \in \mathbb{N}^*$,

$$\|\Phi_{R^\star} - T_{m_{\kappa',n}}\Phi_{R^\star}\|_{2,\omega_{\kappa',n}}^2 \leqslant \mathrm{c}\omega_{\kappa',n}^{d+2m_{\kappa',n}+2/\kappa'} S^{2m_{\kappa',n}} m_{\kappa',n}^{-2\kappa' m_{\kappa',n}+2d} \exp(2\kappa'(S\omega_{\kappa',n})^{1/\kappa'})\,, \qquad (33)$$

and likewise

$$\|\widehat{\phi}_{\kappa',n} - T_{m_{\kappa',n}}\widehat{\phi}_{\kappa',n}\|_{2,\nu}^2 \leqslant \mathrm{c}(S\nu)^{2m_{\kappa',n}+2/\kappa'} m_{\kappa',n}^{-2\kappa' m_{\kappa',n}+2d} \qquad (34)$$

and

$$\|\Phi_{R^\star} - T_{m_{\kappa',n}}\Phi_{R^\star}\|_{2,\nu}^2 \leqslant \mathrm{c}(S\nu)^{2m_{\kappa',n}+2/\kappa'} m_{\kappa',n}^{-2\kappa' m_{\kappa',n}+2d}\,. \qquad (35)$$

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 (32) to (35) show that for all $\kappa' \in [1/2, \kappa]$,

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

which is controlled by equation (29). 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 19 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 $n \geqslant n_0$,

$$\sup_{\kappa\in[\kappa_0,1]} \inf_{\substack{R^\star:\Phi_{R^\star}\in\Psi(\kappa,S,\beta,c_\beta)\\ Q^\star\in\mathbf{Q}(\nu,c_\nu,c_Q)}} \mathbb{P}_{R^\star,Q^\star}\left(\sup_{\kappa'\in[\kappa_0,\kappa]}\left\{m_{\kappa',n}^{2\kappa'\beta}\|\widehat{f}_{\kappa',n} - f^\star\|_2^2\right\} \leqslant c'\right) \geqslant 1 - \frac{4}{n}\,. \qquad (36)$$

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 7 follows.

B Proof of Proposition 18

By (26), Proposition 18 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 20. *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 (27)*,* (8)*,* (28) *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 21. *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 (27)*,* (8) *and* (28)*.*

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 22. *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 $\|h\|_{2,\nu} \leqslant \eta$,*

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

where M^{lin}, $\Upsilon_{\kappa,S}$ and $\mathcal{G}_{\kappa,S}$ are defined in (27)*,* (8) *and* (28)*.*

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

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

where $\mathcal{G}_{\kappa,S}$ is defined in (28)*.*

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 24. *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 (28).

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

By Proposition 22, Proposition 24 and (26), 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 $\|h\|_{2,\nu} \leqslant \eta$,

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

Therefore, assuming

$$\mathrm{c} \left(\frac{\log\log(1/\|h\|_{2,\nu})}{\log(1/\|h\|_{2,\nu})} \right)^{5d+3} \|h\|_{2,\nu}^{\frac{\alpha}{\log\log(1/\|h\|_{2,\nu})}} \geqslant 2\mathrm{c}' \left(\frac{\log(1/\|h\|_{2,\nu})}{\log\log(1/\|h\|_{2,\nu})} \right)^d \|h\|_{2,\nu}^2 \qquad (37)$$

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 (37) is implied by

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

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

C Proof of Theorem 1

The proof follows the same lines as that of Theorem 1 in [Gassiat et al., 2020a]. The following statement is used repeatedly.

Lemma 25. *If a multivariate function is analytic on the whole multivariate complex space and is the null function on an open set of the multivariate real space or on an open set of the multivariate purely imaginary space, then it is the null function on the whole multivariate complex space.*

Proof. We prove the statement by induction on the number d of variables. If h is analytic on $\mathbb{C}$ and is not the null function, then h has isolated zeros, see Theorem 10.18 in [**?**], so that Lemma 25 holds for $d = 1$. Assume that the lemma holds for analytic functions on $\mathbb{C}^d$ and let h be an analytic function on $\mathbb{C}^{d+1}$ which is the null function on an open set A of $\mathbb{R}^{d+1}$. Then, there exists open sets $B_1, \ldots, B_{d+1}$ of $\mathbb{R}$ such that $B_1 \times \cdots \times B_{d+1} \subset A$. For any $t \in B_{d+1}$, let $h_t : \mathbb{C}^d \to \mathbb{C}$ such that $h_t(\cdot) = h(\cdot, t)$, then h_t is analytic on $\mathbb{C}^d$ and is the null function on $B_1 \times \cdots \times B_d$ so that by the induction hypothesis, for all $z \in \mathbb{C}^d$, $h_t(z) = 0$, that is $h(z,t) = 0$ for all $z \in \mathbb{C}$ and for all $t \in B_{d+1}$. Therefore, for any $z \in \mathbb{C}^d$, the function $h(z, \cdot)$ is analytic on $\mathbb{C}$ and is the null function on B_{d+1} so that for any $z_0 \in \mathbb{C}$, $h(z, z_0) = 0$ and h is the null function. The proof when h is the null function on an open set of the multivariate purely imaginary space is similar. □

Assume $\mathbb{P}_{R,Q} = \mathbb{P}_{\widetilde{R},\widetilde{Q}}$ and let ϕ_i (resp. $\widetilde{\phi}_i$) be the characteristic function of $Q^{(i)}$ (resp. $\widetilde{Q}^{(i)}$) for $i \in \{1,2\}$. Since the distribution of $Y^{(1)}$ and $Y^{(2)}$ are the same under $\mathbb{P}_{R,Q}$ and $\mathbb{P}_{\widetilde{R},\widetilde{Q}}$, for any $t \in \mathbb{R}^{d_1}$,

$$\phi_1(t)\, \Phi_R(t, 0) = \widetilde{\phi}_1(t)\, \Phi_{\widetilde{R}}(t, 0) \tag{38}$$

and for any $t \in \mathbb{R}^{d_2}$,

$$\phi_2(t)\, \Phi_R(0, t) = \widetilde{\phi}_2(t)\, \Phi_{\widetilde{R}}(0, t)\ . \tag{39}$$

Since the distribution of $\mathbf{Y}$ is the same under $\mathbb{P}_{R,Q}$ and $\mathbb{P}_{\widetilde{R},\widetilde{Q}}$, for any $(t_1, t_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$,

$$\phi_1(t_1)\, \phi_2(t_2)\, \Phi_R(t_1, t_2) = \widetilde{\phi}_1(t_1)\, \widetilde{\phi}_2(t_2)\, \Phi_{\widetilde{R}}(t_1, t_2)\ . \tag{40}$$

There exists a neighborhood V of 0 in $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ such that for all $t = (t_1, t_2) \in V$, $\phi_1(t_1) \neq 0$, $\phi_2(t_2) \neq 0$, $\widetilde{\phi}_1(t_1) \neq 0$, $\widetilde{\phi}_2(t_2) \neq 0$, so that (38), (39) and (40) imply that for any $(t_1, t_2) \in V^2$,

$$\Phi_R(t_1, t_2)\, \Phi_{\widetilde{R}}(t_1, 0)\, \Phi_{\widetilde{R}}(0, t_2) = \Phi_{\widetilde{R}}(t_1, t_2)\, \Phi_R(t_1, 0)\, \Phi_R(0, t_2)\ . \tag{41}$$

Since $(z_1, z_2) \mapsto \Phi_R(z_1, z_2)\, \Phi_{\widetilde{R}}(z_1, 0)\, \Phi_{\widetilde{R}}(0, z_2) - \Phi_{\widetilde{R}}(z_1, z_2)\, \Phi_R(z_1, 0)\, \Phi_R(0, z_2)$ is a multivariate analytic function of $d_1 + d_2$ variables which is zero on a purely real neighborhood of 0, then it is the null function on the whole multivariate complex space so that for any $z_1 \in \mathbb{C}^{d_1}$ and $z_2 \in \mathbb{C}^{d_2}$,

$$\Phi_R(z_1, z_2)\, \Phi_{\widetilde{R}}(z_1, 0)\, \Phi_{\widetilde{R}}(0, z_2) = \Phi_{\widetilde{R}}(z_1, z_2)\, \Phi_R(z_1, 0)\, \Phi_R(0, z_2)\ . \tag{42}$$

Fix $(u_2, \ldots, u_{d_1}) \in \mathbb{C}^{d_1 - 1}$ and let $\mathcal{Z}$ be the set of zeros of $u \mapsto \Phi_R(u, u_2, \ldots, u_{d_1}, 0)$ and $\widetilde{\mathcal{Z}}$ be the set of zeros of $u \mapsto \Phi_{\widetilde{R}}(u, u_2, \ldots, u_{d_1}, 0)$. Let $u_1 \in \mathcal{Z}$. Write $z_1 = (u_1, u_2, \ldots, u_{d_1})$ so that by (42), for any $z_2 \in \mathbb{C}^{d_2}$,

$$\Phi_R(z_1, z_2)\, \Phi_{\widetilde{R}}(z_1, 0)\, \Phi_{\widetilde{R}}(0, z_2) = 0\,. \tag{43}$$

Using H2, $z_2 \mapsto \Phi_R(z_1, z_2)$ is not the null function. Thus, there exists $z_2^\star$ in $\mathbb{C}^{d_2}$ such that $\Phi_R(z_1, z_2^\star) \neq 0$ and by continuity, there exists an open neighborhood of $z_2^\star$ such that for all z_2 in this open set, $\Phi_R(z_1, z_2) \neq 0$. Since $z \mapsto \Phi_{\widetilde{R}}(0, z)$ is not the null function and is analytic on $\mathbb{C}^{d_2}$, it can not be null all over this open set, so that there exists z_2 such that simultaneously $\Phi_R(z_1, z_2) \neq 0$ and $\Phi_{\widetilde{R}}(0, z_2) \neq 0$. Then (43) leads to $\Phi_{\widetilde{R}}(z_1, 0) = 0$, so that $\mathcal{Z} \subset \widetilde{\mathcal{Z}}$. A symmetric argument yields $\widetilde{\mathcal{Z}} \subset \mathcal{Z}$ so that $\mathcal{Z} = \widetilde{\mathcal{Z}}$. Moreover, the analytic functions $u \mapsto \Phi_R(u, u_2, \ldots, u_d, 0)$ and $u \mapsto \Phi_{\widetilde{R}}(u, u_2, \ldots, u_d, 0)$ have exponential growth order less than 2, so that using Hadamard's factorization Theorem, see [Stein and Shakarchi, 2003, Chapter 5, Theorem 5.1], there exists a polynomial function s with degree at most 1 (and coefficients depending on $(u_2, \ldots, u_d)$) such that for all $u \in \mathbb{C}$,

$$\Phi_R(u, u_2, \ldots, u_d, 0) = \mathrm{e}^{s(u)} \Phi_{\widetilde{R}}(u, u_2, \ldots, u_d, 0)\,.$$

Arguing similarly for all variables, there exists a function F on $\mathbb{C}^{d_1}$, which is, for any $i = 1, \ldots, d_1$, polynomial with degree at most 1 in u_i, and such that for all $(u_1, \ldots, u_{d_1}) \in \mathbb{C}^{d_1}$,

$$\Phi_R(u_1, u_2, \ldots, u_{d_1}, 0) = \mathrm{e}^{\mathsf{F}(u_1, u_2, \ldots, u_{d_1})} \Phi_{\widetilde{R}}(u_1, u_2, \ldots, u_{d_1}, 0)\,. \tag{44}$$

In other words, there exist complex functions a_i, b_i on $\mathbb{C}^{d_1-1}$ such that, if we denote $u^{(-i)}$ the (d_1-1)-dimensional complex vectors with the same coordinates as u except that u_i is not included in the coordinates, then

$$\mathsf{F}(u_1, u_2, \ldots, u_{d_1}) = a_i(u^{(-i)})u_i + b_i(u^{(-i)}),\ i = 1, \ldots, d_1\,.$$

But, for $i \neq j$, the fact that $a_i(u^{(-i)})u_i + b_i(u^{(-i)}) = a_j(u^{(-j)})u_j + b_i(u^{(-j)})$ implies that $a_i(u^{(-i)})$ and $b_i(u^{(-i)})$ are polynomial functions with degree at most 1 in u_j (this may be seen for instance by taking complex derivatives), and by induction we get that F is a polynomial function which is, for any $i = 1, \ldots, d_1$, polynomial with degree at most 1 in u_i.

Since $\Phi_R(0, \ldots, 0) = \Phi_{\widetilde{R}}(0, \ldots, 0) = 1$, the constant term of the polynomial F is 0. We are now going to prove that the polynomial F has total degree at most 1. Note that the fact that F has degree at most 1 in each variable is not enough to deduce that F is linear: for instance, u_1u_2 has degree at most 1 in each variable but has total degree 2.

Assume that $\widetilde{R}^{(1)}$ is not supported by 0. Then there exist $a = (a_1, \ldots, a_{d_1}) \in \mathbb{R}^{d_1}$, $\alpha > 0$ and $\delta > 0$ such that

$$0 \notin \prod_{j=1}^{d_1}[a_j - \alpha, a_j + \alpha] \quad \text{and} \quad \widetilde{R}^{(1)}\left(\prod_{j=1}^{d_1}[a_j - \alpha, a_j + \alpha]\right) \geqslant \delta\,,$$

which gives, for all $u \in \mathbb{R}^{d_1}$,

$$\Phi_{\widetilde{R}}(-iu, 0) \geqslant \delta \mathrm{e}^{\sum_{j=1}^{d_1} \inf_{x_j \in [a_j - \alpha, a_j + \alpha]} u_j x_j}\,,$$

so that using (44), for all $u \in \mathbb{R}^{d_1}$,

$$\Phi_R(-iu, 0) \geqslant \delta \mathrm{e}^{\mathsf{F}(u)} \mathrm{e}^{\sum_{j=1}^{d_1} \inf_{x_j \in [a_j - \alpha, a_j + \alpha]} u_j x_j}\,.$$

If F has total degree at least 2, then there exist $i \neq j$ and polynomial functions c_1 on $\mathbb{C}^{d_1-2}$ and c_2, c_3 on $\mathbb{C}^{d_1-1}$ with degree at most one in each variable such that, if we denote $u^{(-i,-j)}$ the (d_1-2)-dimensional complex vectors with the same coordinates as u except that u_i and u_j are not included in the coordinates, then $\mathsf{F}(u) = c_1(u^{(-i;-j)})u_iu_j + c_2(u^{(-i)}) + c_3(u^{(-j)})$. Without loss of generality say that $i = 1$ and $j = 2$. Then it is possible to find $u \in \mathbb{R}^{d_1}$ and $\widetilde{\delta} > 0$ such that for all $t \geqslant 0$, $\mathsf{F}(-itu_1, -itu_2, -iu_3, \ldots, -iu_{d_1}) \geqslant \widetilde{\delta}t(u_1^2 + u_2^2)$ leading to

$$\forall t \geqslant 0,\ \Phi_R(-itu_1, -itu_2, -iu_3, \ldots, -iu_{d_1}, 0) \geqslant \delta \mathrm{e}^{\widetilde{\delta}t(u_1^2+u_2^2)} \mathrm{e}^{\sum_{j=1}^{d_1} \inf_{x_j \in [a_j - \alpha, a_j + \alpha]} u_j x_j}\,,$$

contradicting the assumption that $R^{(1)} \in \mathcal{M}_\rho$ for some $\rho < 2$. Thus, F has total degree at most 1 and there exists $m \in \mathbb{C}^{d_1}$ such that for all $z \in \mathbb{C}^{d_1}$,

$$\Phi_R(z, 0) = \mathrm{e}^{im_1^\top z} \Phi_{\widetilde{R}}(z, 0)\,. \tag{45}$$

On the other hand, if $\widetilde{R}^{(1)}$ is supported by 0 then (44) leads to $\Phi_R(-iu, 0) = \mathrm{e}^{\mathsf{F}(-iu)}$ for all $u \in \mathbb{R}^{d_1}$ and the same argument leads to (45).

As for all $z \in \mathbb{R}^{d_1}$, $\Phi_R(-z, 0) = \overline{\Phi_R(z, 0)}$ and $\Phi_{\widetilde{R}}(-z, 0) = \overline{\Phi_{\widetilde{R}}(z, 0)}$, $m_1 \in \mathbb{R}^d$. Arguing similarly for the function $\Phi_R(0, z_2)$, there exists $m_2 \in \mathbb{R}^{d_2}$ such that for all $z \in \mathbb{C}^{d_2}$,

$$\Phi_R(0, z) = \mathrm{e}^{im_2^\top z} \Phi_{\widetilde{R}}(0, z)\,. \tag{46}$$

Combining (45) and (46) with (42) yields, for all $(t_1, t_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$,

$$\Phi_R(t_1, t_2) = e^{im_1^\top t_1 + im_2^\top t_2} \Phi_{\widetilde{R}}(t_1, t_2) . \tag{47}$$

Then, using (38), for all $t \in \mathbb{R}^{d_1}$ such that $\Phi_R(t, 0) \neq 0$, $\phi_1(t) = \mathrm{e}^{-im_1^\top t}\widetilde{\phi}_1(t)$. Since the set of zeros of $t \mapsto \Phi_R(t, 0)$ has empty interior, for each t such that $\Phi_R(t, 0) = 0$ it is possible to find a sequence $(t_n)_{n\geqslant 1}$ such that t_n tends to t and for all n, $\Phi_{R_K}(t_n, 0) \neq 0$. But ϕ_1 and $\widetilde{\phi}_1$ are continuous functions, so that for all $t \in \mathbb{R}^{d_1}$,

$$\phi_1(t) = \mathrm{e}^{-im_1^\top t}\widetilde{\phi}_1(t) . \tag{48}$$

Similarly using (39), we get that for all $t \in \mathbb{R}^{d_2}$,

$$\phi_2(t) = \mathrm{e}^{-im_2^\top t}\widetilde{\phi}_2(t) . \tag{49}$$

The proof is concluded by noting that (47), (48) and (49) imply that $R = \widetilde{R}$ and $Q = \widetilde{Q}$ up to translation.

D Proof of Corollary 2

The noisy ICA model may be written as

$$\begin{pmatrix} Y_I \\ Y_J \end{pmatrix} = \begin{pmatrix} A_I S \\ A_J S \end{pmatrix} + \begin{pmatrix} \varepsilon_I \\ \varepsilon_J \end{pmatrix} .$$

Then, the ICA model fits the setting of Theorem 1 with $Y^{(1)} = Y_I$, $Y^{(2)} = Y_J$, $X^{(1)} = A_I S$ and $X^{(2)} = A_J S$. Write $d_1 = |I|$ and $d_2 = d - |I|$ and denote by R the joint distribution of $(A_I S, A_J S)$. Note first that if for all $1 \leqslant j \leqslant q$ the distribution of all S_j is in $\mathcal{M}_\rho^1$ then the distribution of $A_I S$ is in $\mathcal{M}_\rho^{d_1}$ as for all $\lambda \in \mathbb{R}^{d_1}$,

$$\mathbb{E}\left[\exp\left(\lambda^\top A_I S\right)\right] = \prod_{j=1}^q \Psi_j\left((\lambda^\top A_I)_j\right) ,$$

where for all $1 \leqslant j \leqslant q$ and all $z \in \mathbb{C}$, $\Psi_j(z) = \mathbb{E}\left[\exp\left(zS_j\right)\right]$. Then, by assumption, and the Cauchy-Schwarz inequality, there exist $A_j \in \mathbb{R}$ and $B_j \in \mathbb{R}$ such that

$$\Psi_j\left((\lambda^\top A_I)_j\right) \leqslant A_j \mathrm{e}^{B_j |\langle \lambda; A_I(j)\rangle|^\rho} \leqslant A_j \mathrm{e}^{B_j \|A_I(j)\|^\rho \|\lambda\|^\rho} ,$$

where $A_I(j)$ is the j-th column of A_I. Therefore, $A_I S$ is in $\mathcal{M}_\rho^{d_1}$ with constants given by $A = \prod_{j=1}^q A_I(j)$ and $B = \sum_{j=1}^q B_j \|A_I(j)\||^\rho$. Similarly, $A_J S$ is in $\mathcal{M}_\rho^{d_2}$. Then, for any $(z_0, z) \in \mathbb{C}^{d_1} \times \mathbb{C}^{d_2}$,

$$\begin{aligned}\Phi_R(z_0, z) = \mathbb{E}\left[\exp\left((iz_0^\top A_I + iz^\top A_J)S\right)\right] &= \prod_{j=1}^q \mathbb{E}\left[\exp\left((iz_0^\top A_I + iz^\top A_J)_j S_j\right)\right] , \\ &= \prod_{j=1}^q \Psi_j\left((iz_0^\top A_I + iz^\top A_J)_j\right) .\end{aligned}$$

For all $1 \leqslant j \leqslant q$, the function $z \mapsto \Psi_j((z_0^\top A_I + z^\top A_J)_j)$ is analytic therefore $z \mapsto \Phi_R(z_0, z)$ is the null function if and only if there exists $1 \leqslant j \leqslant q$ such that the function $z \mapsto \Psi_j((iz_0^\top A_I + iz^\top A_J)_j)$ is null (as $\Phi_R(z_0,\cdot)$ is a finite product of analytic functions). As all columns of A_J are nonzero, for all $1 \leqslant j \leqslant q$, $z \mapsto \Psi_j((iz_0^\top A_I + iz^\top A_J)_j)$ is not the null function. Similarly, for any $(z,z_0) \in \mathbb{C}^{d_1} \times \mathbb{C}^{d_2}$,

$$\Phi_R(z,z_0) = \prod_{j=1}^{q} \Psi_j\left((iz^\top A_I + iz_0^\top A_J)_j\right) ,$$

and the proof that $\Phi_R(z,z_0)$ is not the null function follows the same steps.

E Links between $\mathcal{M}_\rho^d$ and $\Upsilon_{1/\rho,S}$: proof of Lemma 5

First implication. First, note that for all $n \geqslant 0$, $n! \geqslant (n/\mathrm{e})^n$, so that by the concavity of $x \mapsto \log x$, for all $j \in \mathbb{N}^d$,

$$\left(\prod_{a=1}^{d} j_a!\right)^{-1} \leqslant \mathrm{e}^{\|j\|_1} \exp\left(-\sum_{a=1}^{d} j_a \log j_a\right) \leqslant \mathrm{e}^{\|j\|_1} \exp(-\|j\|_1 \log(\|j\|_1/d)) \leqslant \left(\frac{\mathrm{e}d}{\|j\|_1}\right)^{\|j\|_1} . \quad (50)$$

Let $\rho \geqslant 1$ and $\mu \in \mathcal{M}_\rho^d$. Write $\varphi_\mu : \lambda \in \mathbb{R}^d \mapsto \int \exp\left(i\lambda^\top x\right)\mu(\mathrm{d}x)$. Then $\varphi_\mu(0) = 1$ and for all $j \in \mathbb{N}^d \setminus \{0\}$, if X has distribution μ, by the inequality of arithmetic and geometric means and by convexity of $x \longmapsto x^{\|j\|_1}$ on $\mathbb{R}_+$,

$$\begin{aligned}|\partial^j \varphi_\mu(0)| = \left|\mathbb{E}\left[\prod_{a=1}^{d} X_a^{j_a}\right]\right| &\leqslant \mathbb{E}\left[\left(\sum_{a=1}^{d} \frac{j_a}{\|j\|_1}|X_a|\right)^{\|j\|_1}\right] ,\\ &\leqslant \mathbb{E}\left[\sum_{a=1}^{d} \frac{j_a}{\|j\|_1}|X_a|^{\|j\|_1}\right] \leqslant \max_{1\leqslant a\leqslant d} \mathbb{E}\left[|X_a|^{\|j\|_1}\right] .\end{aligned}$$

Since $\mu \in \mathcal{M}_\rho^d$ by assumption, there exist A and B such that for all $\lambda \in \mathbb{R}^d$,

$$\mathbb{E}[e^{\lambda^\top X}] \leqslant A\mathrm{e}^{B\|\lambda\|_2^\rho} . \quad (51)$$

Hence, by Markov's inequality, for all $a \in \{1,\dots,d\}$, $t > 0$ and $\lambda > 0$,

$$\mathbb{P}(X_a \geqslant t) \leqslant \frac{\mathbb{E}[\mathrm{e}^{\lambda X_a}]}{\mathrm{e}^{\lambda t}} \leqslant A\exp(B\lambda^\rho - \lambda t) .$$

Thus, if $\rho = 1$, then $|X_a| \leqslant B$ almost surely, and therefore for all $j \in \mathbb{N}^d \setminus \{0\}$,

$$|\partial^j \varphi_\mu(0)| \leqslant B^{\|j\|_1} ,$$

which concludes the proof together with equation (50). In the following, assume $\rho > 1$, so that

$$\mathbb{P}(X_a \geqslant t) \leqslant A\exp(-Ct^{\rho/(\rho-1)}) ,$$

where $C = (\rho-1)B(\rho B)^{-\rho/(\rho-1)} > 0$. Therefore, writing $\gamma = \rho/(\rho-1) > 1$ yields

$$\mathbb{P}(|X_a| \geqslant t) \leqslant 2A\exp(-Ct^\gamma) .$$

Let $j \in \mathbb{N}^d \setminus \{0\}$, then

$$\begin{aligned}\mathbb{E}\left[|X_a|^{\|j\|_1}\right] = \int_{\varepsilon \geqslant 0} \mathbb{P}(|X_a|^{\|j\|_1} \geqslant \varepsilon)\mathrm{d}\varepsilon = \|j\|_1 \int_{t\geqslant 0} \mathbb{P}(|X_a| \geqslant t) t^{\|j\|_1 - 1}\mathrm{d}t\,, \\ \leqslant 2A\|j\|_1 \int_{t\geqslant 0} t^{\|j\|_1-1}\mathrm{e}^{-Ct^\gamma}\mathrm{d}t\,, \\ \leqslant 2A\|j\|_1 \left(1 + \int_{t\geqslant 1} t^{\|j\|_1-1}\mathrm{e}^{-Ct^\gamma}\right)\mathrm{d}t\,.\end{aligned}$$

For all $x \in \mathbb{R}$, note that $J_x = \int_{t\geqslant 1} t^x \mathrm{e}^{-Ct^\gamma}\mathrm{d}t = (\gamma C)^{-1}(\mathrm{e}^{-C} + (x-\gamma+1)J_{x-\gamma})$. Since for $x \leqslant 0$, $J_x \leqslant \int_{t\geqslant 1}\mathrm{e}^{-Ct}\mathrm{d}t \leqslant \mathrm{e}^{-C}/C$ as $\gamma > 1$

$$J_x \leqslant \frac{\mathrm{e}^{-C}}{\gamma C}\sum_{\ell=0}^{\lceil x/\gamma\rceil - 1}\left(\frac{x}{\gamma C}\right)^{\ell} + \left(\frac{x}{\gamma C}\right)^{\lceil x/\gamma \rceil}\frac{\mathrm{e}^{-C}}{C}\,.$$

Thus, if $x \geqslant \gamma C/2$, $J_x \leqslant 2(\mathrm{e}^{-C}/C)(4x/(\gamma C))^{\lceil x/\gamma\rceil}$, and if $x \leqslant \gamma C/2$, $J_x \leqslant 2\mathrm{e}^{-C}/C$, so that

$$J_x \leqslant 2\frac{\mathrm{e}^{-C}}{\gamma C}\left(1 + \left(\frac{4x}{\gamma C}\right)^{\lceil x/\gamma\rceil}\right)\,,$$

and as a consequence

$$\begin{aligned}\mathbb{E}\left[|X_a|^{\|j\|_1}\right] \leqslant 2A\|j\|_1(1 + J_{\|j\|_1-1}) \leqslant 2A\|j\|_1\left(1 + 2\frac{\mathrm{e}^{-C}}{C} + 2\frac{\mathrm{e}^{-C}}{C}\left(\frac{4(\|j\|_1 - 1)}{\gamma C}\right)^{\lceil(\|j\|_1-1)/\gamma\rceil}\right)\,, \\ \leqslant 2A\|j\|_1\left(1 + 2\frac{\mathrm{e}^{-C}}{C} + 2\frac{\mathrm{e}^{-C}}{C}\left(\frac{4\|j\|_1}{\gamma C}\right)^{(\|j\|_1-1)/\gamma+1}\right)\,.\end{aligned}$$

Hence, since $\|j\|_1 \geqslant 1$, there exist constants $c, c' > 0$ which only depend on ρ, A and B such that

$$\begin{aligned}|\partial^j\varphi_\mu(0)| &\leqslant c\|j\|_1^{2-1/\gamma}\left(1 + \left(\frac{4\|j\|_1}{\gamma C}\right)^{\|j\|_1/\gamma}\right)\,, \\ &\leqslant 2c\mathrm{e}^{\|j\|_1(2-1/\gamma)}\left(\left(\frac{4}{\gamma C}\vee 1\right)\|j\|_1\right)^{\|j\|_1/\gamma}\,, \\ &\leqslant (c'\|j\|_1)^{\|j\|_1/\gamma}\,.\end{aligned}$$

Bringing the above inequality together with equation (50) implies for all $j \in \mathbb{N}^d \setminus \{0\}$,

$$\left|\frac{\partial^j\varphi_\mu(0)}{\prod_{a=1}^d j_a!}\right| \leqslant (2\mathrm{e}d(c')^{1/\gamma})^{\|j\|_1}\|j\|_1^{\|j\|_1(1/\gamma-1)} \leqslant S^{\|j\|_1}\|j\|_1^{-\|j\|_1/\rho}\,,$$

where $S = \mathrm{e}d(c')^{1/\gamma}$, which concludes the proof.

Second implication. Let $S, \kappa > 0$ and let μ be a probability measure on $\mathbb{R}^d$ such that $\phi : \lambda \in \mathbb{R}^d \mapsto \int \exp(i\lambda^\top x)\mu(\mathrm{d}x) \in \Upsilon_{\kappa,S}$. Then, ϕ can be extended to $\mathbb{C}^d$ and is equal to its Taylor expansion. In particular, for all $\lambda \in \mathbb{R}^d$,

$$\phi(-i\lambda) \leqslant 1 + \sum_{j \in \mathbb{N}^d \setminus \{0\}} S^{\|j\|_1} \|j\|_1^{-\kappa\|j\|_1} \prod_{a=1}^{d} \lambda_a^{j_a} \leqslant 1 + \sum_{m \geqslant 1} m^d (S\|\lambda\|)^m m^{-\kappa m} .$$

By Lemma 26, there exists a constant $x_0 > 0$ depending only on κ and d such that for all $\lambda \in \mathbb{R}^d$,

$$\phi(-i\lambda) \leqslant 1 + 6(S\|\lambda\| \vee x_0)^{\frac{d+1}{\kappa}} \exp\left(\kappa(S\|\lambda\| \vee x_0)^{1/\kappa}\right) ,$$

which implies that there exists a constant c depending only on κ and d such that for all $\lambda \in \mathbb{R}^d$,

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

F Proof of Theorem 9

By equation (36) from the proof of Theorem 7, taking $m_{\kappa,n}$ as in equation (12), for all $\kappa_0 \in (1/2, 1]$, $S > 0$, $\beta > 0$, $\nu \in (0, \nu_{\text{est}}]$, $c_\nu > 0$, $c_Q > 0$ and $c_\beta > 0$, there exist $c' > 0$ and n_0 such that for all $n \geqslant n_0$,

$$\sup_{\kappa \in [\kappa_0, 1]} \inf_{\substack{R^\star : \Phi_{R^\star} \in \Psi(\kappa, S, \beta, c_\beta) \\ Q^\star \in \mathbf{Q}(\nu, c_\nu, c_Q)}} \mathbb{P}_{R^\star, Q^\star}\left(\sup_{\kappa' \in [\kappa_0, \kappa]} \left\{ \left(\frac{\log n}{\log \log n} \right)^{\kappa'\beta} \|\widehat{f}_{\kappa',n} - f^\star\|_2 \right\} \leqslant c' \right) \geqslant 1 - \frac{4}{n} . \quad (52)$$

Write $\sigma_n(\kappa') = c' \left(\log n / \log\log n\right)^{-\kappa'\beta}$, we will show that

$$\sup_{\kappa \in [\kappa_0, 1]} \inf_{\substack{R^\star : \Phi_{R^\star} \in \Psi(\kappa, S, \beta, c_\beta) \\ Q^\star \in \mathbf{Q}(\nu, c_\nu, c_Q)}} \mathbb{P}_{R^\star, Q^\star}\left(\|\widehat{f}_{\widehat{\kappa}_n,n} - f^\star\|_2 \leqslant 5\sigma_n(\kappa) \right) \geqslant 1 - \frac{4}{n} , \quad (53)$$

and since $\|\widehat{f}_{\widehat{\kappa}_n,n} - f^\star\|_2^2 \leqslant \mathrm{diam}(\Upsilon_{\kappa_0,S})^2$ by construction, Theorem 9 follows. Fix κ, $R^\star$ and Q^* and assume we are in the event of probability at least $1 - 4/n$ of equation (52) where $\|\widehat{f}_{\kappa',n} - f^\star\|_2^2 \leqslant \sigma_n(\kappa')$ for all $\kappa' \in [\kappa_0, \kappa]$. By the triangular inequality, for all $\kappa' \in [\kappa_0, \kappa]$,

$$\begin{aligned}
\|\widehat{f}_{\widehat{\kappa}_n,n} - f^\star\|_2 &\leqslant \|\widehat{f}_{\kappa',n} - f^\star\|_2 + \|\widehat{f}_{\widehat{\kappa}_n,n} - \widehat{f}_{\kappa',n}\|_2 , \\
&\leqslant \sigma_n(\kappa') + \begin{cases} A_n(\widehat{\kappa}_n) + \sigma_n(\kappa') & \text{if } \widehat{\kappa}_n \geqslant \kappa', \\ A_n(\kappa') + \sigma_n(\widehat{\kappa}_n) & \text{otherwise,} \end{cases} \\
&\leqslant \sigma_n(\kappa') + A_n(\widehat{\kappa}_n) + \sigma_n(\kappa') + A_n(\kappa') + \sigma_n(\widehat{\kappa}_n) \leqslant 2A_n(\kappa') + 3\sigma_n(\kappa')
\end{aligned}$$

by definition of $\widehat{\kappa}_n$ and since $A_n \geqslant 0$ and $\sigma_n \geqslant 0$. Recall that

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

so that as $\kappa' \leqslant \kappa$,

$$\begin{aligned}
A_n(\kappa') &\leqslant 0 \vee \sup_{\kappa'' \in [\kappa_0, \kappa]} \{\|\widehat{f}_{\kappa'',n} - f^\star\|_2 + \|\widehat{f}_{\kappa',n} - f^\star\|_2 - \sigma_n(\kappa'')\}\,, \\
&= \|\widehat{f}_{\kappa',n} - f^\star\|_2 + \left(0 \vee \sup_{\kappa'' \in [\kappa_0, \kappa]} \{\|\widehat{f}_{\kappa'',n} - f^\star\|_2 - \sigma_n(\kappa'')\}\right)\,, \\
&\leqslant \|\widehat{f}_{\kappa',n} - f^\star\|_2 \leqslant \sigma_n(\kappa')\,.
\end{aligned}$$

Equation (53) follows by taking $\kappa' = \kappa$.

G Proof of Lemma 17

To simplify the notations, write $\Phi^\star : (t_1, t_2) \mapsto \Phi_{R^\star}(t_1, t_2)\Phi^\star_{Q^{\star,(1)}}(t_1)\Phi^\star_{Q^{\star,(2)}}(t_2)$ the characteristic function of $\mathbf{Y}$ under the parameters $(R^\star, Q^\star)$. By definition of M and M_n and since for any complex numbers a and b, $||a|^2 - |b|^2| \leqslant |a-b|(|a|+|b|)$, for any $\kappa > 0$, $S > 0$, $\phi \in \Upsilon_{\kappa,S}$, $R^\star$ and $Q^\star$,

$$\begin{aligned}
&|M_n(\phi) - M_\star(\phi;\nu_{\text{est}})| \\
&\qquad \leqslant \Bigg\| \Big\{ (\phi\widetilde{\phi}_n^{(1)}\widetilde{\phi}_n^{(2)} - \widetilde{\phi}_n\phi^{(1)}\phi^{(2)}) - (\phi\Phi^{\star,(1)}\Phi^{\star,(2)} - \Phi^\star\phi^{(1)}\phi^{(2)}) \Big\}\,, \\
&\qquad\qquad \times \left(\left| (\phi\widetilde{\phi}_n^{(1)}\widetilde{\phi}_n^{(2)} - \widetilde{\phi}_n\phi^{(1)}\phi^{(2)}) \right| + \left| (\phi\Phi^{\star,(1)}\Phi^{\star,(2)} - \Phi^\star\phi^{(1)}\phi^{(2)}) \right| \right) \Bigg\|_{1,\nu_{\text{est}}} \\
&\qquad \leqslant 2\|\phi\|_{\infty,\nu_{\text{est}}}(1 + \|\phi\|_{\infty,\nu_{\text{est}}} \Big\| \phi(\widetilde{\phi}_n^{(1)}\widetilde{\phi}_n^{(2)} - \Phi^{\star,(1)}\Phi^{\star,(2)}) - (\widetilde{\phi}_n - \Phi^\star)\phi^{(1)}\phi^{(2)} \Big\|_{1,\nu_{\text{est}}}\,, \\
&\qquad \leqslant 4\|\phi\|^2_{\infty,\nu_{\text{est}}}(1 + \|\phi\|_{\infty,\nu_{\text{est}}})^2 \nu_{\text{est}}^d \|\widetilde{\phi}_n - \Phi^\star\|_{\infty,\nu_{\text{est}}}\,, \\
&\qquad \leqslant 16\|\phi\|^4_{\infty,\nu_{\text{est}}} \nu_{\text{est}}^d \|\widetilde{\phi}_n - \Phi^\star\|_{\infty,\nu_{\text{est}}}\,,
\end{aligned}$$

since $\|\Phi^\star\|_\infty \leqslant 1$ and $\|\widetilde{\phi}_n\|_\infty \leqslant 1$ by definition. Thus, by (58) in Lemma 27, for any $S > 0$, there exists $\mathrm{c} > 0$ such that for all $\kappa \in [1/2, 1]$, $n \geqslant 1$ and any probability measures $R^\star$ and $Q^\star$ on $\mathbb{R}^d$,

$$\sup_{\phi \in \Upsilon_{\kappa,S}} |M_n(\phi) - M_\star(\phi;\nu_{\text{est}})| \leqslant \mathrm{c}\|\widetilde{\phi}_n - \Phi^\star\|_{\infty,\nu_{\text{est}}}\,.$$

Let $N_\star^{\text{Re}}(\epsilon, \nu_{\text{est}})$ (resp. $N_\star^{\text{Im}}(\epsilon, \nu_{\text{est}})$) be the number of brackets of size ϵ required to cover $\{y \in \mathbb{R}^d \mapsto \text{Re}(\mathrm{e}^{it^\top y}), t \in \mathsf{B}^d_{\nu_{\text{est}}}\}$ (resp. with Im instead of Re), where the size of the bracket $[u, v]$ is $\mathbb{E}_{R^\star,Q^\star}[(v-u)^2(Y)]^{1/2}$. Since all these functions take values in $[-1, 1]$ and for all $y, t, t' \in \mathbb{R}^d$,

$$|\mathrm{e}^{it^\top y} - \mathrm{e}^{it'^\top y}| \leqslant |(t-t')^\top y| \leqslant \sqrt{d}\|t - t'\|_\infty \|y\|\,,$$

it is possible to obtain a bracket of size ϵ for each of these two sets from a bracket of size $\epsilon/(\sqrt{d\mathbb{E}_{R^\star,Q^\star}[\|\mathbf{Y}\|^2]})$ of $(\mathsf{B}^d_{\nu_{\text{est}}}, \|\cdot\|_\infty)$, which means that

$$N_\star^{\text{Re}}(\epsilon, \nu_{\text{est}}) \vee N_\star^{\text{Im}}(\epsilon, \nu_{\text{est}}) \leqslant \left(\frac{4\nu_{\text{est}}\sqrt{d\mathbb{E}_{R^\star,Q^\star}[\|\mathbf{Y}\|^2]}}{\epsilon} \vee 1 \right)^d\,.$$

Thus, by [Massart, 2007], Theorem 6.8 and Corollary 6.9, there exists a numerical constant C such that for all $x > 0$, $R^\star$ and $Q^\star$,

$$\mathbb{P}_{R^\star,Q^\star}\left(\|\widetilde{\phi}_n - \Phi^\star\|_{\infty,\nu_{\text{est}}} \geqslant C\left[\frac{E(R^\star,Q^\star)}{n} + \sqrt{\frac{x}{n}} + 2\frac{x}{n}\right]\right) \leqslant 4e^{-x}\,,$$

where

$$\begin{aligned}
E(R^\star,Q^\star) &= \sqrt{n}\int_0^1 \sqrt{n \wedge d\log\frac{\sqrt{1\vee 16\nu_{\text{est}}^2 d\mathbb{E}_{R^\star,Q^\star}[\|\mathbf{Y}\|^2]}}{u}}\mathrm{d}u + \frac{3}{2}d\log(1\vee 16\nu_{\text{est}}^2 d\mathbb{E}_{R^\star,Q^\star}[\|\mathbf{Y}\|^2])\,,\\
&= \sqrt{n}A\int_0^{1/A}\sqrt{n\wedge d\log\frac{1}{v}}\mathrm{d}u + 3d\log A\,, \qquad \text{where } A = \sqrt{1\vee 16\nu_{\text{est}}^2 d\mathbb{E}_{R^\star,Q^\star}[\|\mathbf{Y}\|^2]}\,,\\
&= \sqrt{n}A\left(\sqrt{n}\mathrm{e}^{-n} + d\int_A^{\mathrm{e}^n}\frac{\sqrt{\log x}}{x^2}\mathrm{d}x\right) + 3d\log A\,,\\
&\leqslant d\sqrt{n}A\left(\mathrm{e}^{-1} + \int_1^{+\infty}\frac{\sqrt{x}}{x^2}\mathrm{d}x\right) + 3dA\,,\\
&\leqslant 6d\sqrt{n}A\,,\\
&= 6d\sqrt{n}\sqrt{1\vee 16\nu_{\text{est}}^2 d\mathbb{E}_{R^\star,Q^\star}[\|\mathbf{Y}\|^2]}\,.
\end{aligned}$$

Hence, for all $n \geqslant 1$, $x > 0$, $R^\star$ and $Q^\star$, with probability at least $1 - 4e^{-x}$ under $\mathbb{P}_{R^\star,Q^\star}$,

$$\|\widetilde{\phi}_n - \Phi^\star\|_{\infty,\nu_{\text{est}}} \leqslant C\left[6d\sqrt{\frac{1\vee 16\nu_{\text{est}}^2 d\mathbb{E}_{R^\star,Q^\star}[\|\mathbf{Y}\|^2]}{n}} + \sqrt{\frac{x}{n}} + 2\frac{x}{n}\right]\,,$$

and finally for all $S > 0$, there exists $\mathrm{c} > 0$ such that for all $\kappa \in [1/2, 1]$, $n \geqslant 1$ and $x > 0$, with probability at least $1 - 4e^{-x}$,

$$\sup_{\phi\in\Upsilon_{\kappa,S}}|M_n(\phi) - M_*(\phi;\nu_{\text{est}})| \leqslant \mathrm{c}\left[\sqrt{\frac{\mathbb{E}_{R^\star,Q^\star}[\|\mathbf{Y}\|^2]}{n}} \vee \sqrt{\frac{1}{n}} \vee \sqrt{\frac{x}{n}} \vee \frac{x}{n}\right]\,,$$

which concludes the proof.

H Technical results

Lemma 26. *For all $d \geqslant 0$, $x > 0$ and $\kappa > 0$, let $\psi_{x,d}$ be the function defined on $\mathbb{R}_+^*$ by $\psi_{x,d} : u \mapsto u^d x^u u^{-\kappa u}$. Let $x_0 = ((d+4/3)/\kappa)^\kappa$, then for all $x > 0$,*

$$\sum_{m\geqslant 1}\psi_{x,d}(m) \leqslant 6(x\vee x_0)^{\frac{d+1}{\kappa}}\exp(\kappa(x\vee x_0)^{1/\kappa})\,.$$

Proof. For all $x > 0$, there exists $u_\star(x)$ such that $\psi_{x,d}$ is nondecreasing on $(0, u_\star(x)]$ and nonincreasing on $[u_\star(x), +\infty)$. This real number satisfies

$$u_\star(x) = \sup\left\{u > 0 : u\log(x) - \kappa u\log u - \kappa u + d \geqslant 0\right\}\,.$$

Hence, for all $x > 0$, $u_\star(x) \geqslant \mathrm{e}^{-1}x^{1/\kappa}$ and for all $x \geqslant (d/\kappa)^\kappa$, $u_\star(x) \leqslant x^{1/\kappa}$, so that

$$\psi_{x,d}(u_\star(x_0)) = u_\star(x)^d \left(\frac{x^{1/\kappa}}{u_\star(x)}\right)^{\kappa u_\star(x)} \leqslant x^{d/\kappa}\exp(\kappa x^{1/\kappa})\,.$$

Thus,

$$\begin{aligned}\sum_{m\geqslant 1}\psi_{x,d}(m) &\leqslant \sum_{m=1}^{\lfloor u_\star(x)\rfloor -1}\psi_{x,d}(m) + \psi_{x,d}(\lfloor u_\star(x)\rfloor) + \psi_{x,d}(\lceil u_\star(x)\rceil) + \sum_{m\geqslant \lceil u_\star(x)\rceil+1}\psi_{x,d}(m)\,,\\ &\leqslant \int_1^{\lfloor u_\star(x)\rfloor}\psi_{x,d}(u)\mathrm{d}u + 2\psi_{x,d}(u_\star(x)) + \int_{\lceil u_\star(x)\rceil}^{\infty}\psi_{x,d}(u)\mathrm{d}u\,,\end{aligned}$$

so that for all $x \geqslant (d/\kappa)^\kappa$,

$$\begin{aligned}\sum_{m\geqslant 1}\psi_{x,d}(m) &\leqslant 2x^{d/\kappa}\exp\left(\kappa x^{1/\kappa}\right) + \int_{u\geqslant 1}\psi_{x,d}(u)\mathrm{d}u\,,\\ &\leqslant 2x^{d/\kappa}\exp\left(\kappa x^{1/\kappa}\right) + \int_{u\geqslant 1}u^d\left(\frac{x}{u^\kappa}\right)^u\mathrm{d}u\,.\end{aligned}$$

Let $u = x^{1/\kappa}v$, then

$$\begin{aligned}\int_{u\geqslant 1}u^d\left(\frac{x}{u^\kappa}\right)^u\mathrm{d}u &\leqslant x^{\frac{d+1}{\kappa}}\int_{v\geqslant x^{-1/\kappa}}v^{d-\kappa x^{1/\kappa}v}\mathrm{d}v\,,\\ &\leqslant x^{\frac{d+1}{\kappa}}\left(\int_{0\leqslant v\leqslant 1}\exp(-\kappa x^{1/\kappa}v\log v)\mathrm{d}v + \int_{v\geqslant 1}v^{d-\kappa x^{1/\kappa}v}\mathrm{d}v\right)\,,\\ &\leqslant x^{\frac{d+1}{\kappa}}\left(\exp\left(\kappa x^{1/\kappa}\frac{1}{\mathrm{e}}\right) + \int_{v\geqslant 1}v^{d-\kappa x^{1/\kappa}}\mathrm{d}v\right)\,,\\ &\leqslant x^{\frac{d+1}{\kappa}}\left(\exp(\kappa x^{1/\kappa}) + \frac{1}{\kappa x^{1/\kappa}-(d+1))}\right)\,,\\ &\leqslant x^{\frac{d+1}{\kappa}}\left(\exp(\kappa x^{1/\kappa})+3\right)\,,\end{aligned}$$

when $\kappa x^{1/\kappa} \geqslant d + 4/3$, so that for all $x \geqslant \left(\frac{d+4/3}{\kappa}\right)^\kappa$,

$$\sum_{m\geqslant 1}\psi_{x,d}(m) \leqslant 3x^{\frac{d+1}{\kappa}}(1+\exp(\kappa x^{1/\kappa})) \leqslant 6x^{\frac{d+1}{\kappa}}\exp(\kappa x^{1/\kappa})\,.$$

The case $x \leqslant ((d+4/3)/\kappa)^\kappa$ follows from the fact that $x \mapsto \psi_{x,d}(m)$ is nondecreasing for all positive integer m and all $d \geqslant 0$. □

The results established in this section involve the following quantities.

$$C_\Upsilon(\kappa, S, \nu) = \sup_{\phi\in\Upsilon_{\kappa,S}}\|\phi\|_{\infty,\nu}\,, \tag{54}$$

$$f_\kappa : u \mapsto \sum_{m\geqslant 1}(m+d/\kappa)^{-\kappa m}u^m\,, \tag{55}$$

$$g : (\kappa, S) \mapsto \sup_{x\geqslant 1}\left\{(\max(S,1)\mathrm{e}^{d+2}2^\kappa)^x x^{-\kappa x+1}\right\}\,, \tag{56}$$

where $d = d_1 + d_2$.

Lemma 27. *Let $\kappa > 0$ and $u_0 = (4/(3\kappa))^\kappa$, then for all $u > 0$,*

$$f_\kappa(u) \leqslant 6(u \vee u_0)^{1/\kappa} \exp(\kappa(u \vee u_0)^{1/\kappa}) , \tag{57}$$

where f_κ is defined in (55)*. Let $\kappa, S > 0$ and $x_0 = 1 \vee ((d + 4/3)/\kappa)^\kappa$, then for all $\nu > 0$,*

$$C_\Upsilon(\kappa, S, \nu) \leqslant 7(S\nu \vee x_0)^{\frac{d+1}{\kappa}} \exp(\kappa(S\nu \vee x_0)^{1/\kappa}) , \tag{58}$$

where C_Υ is defined in (54)*. For all $\kappa, S > 0$,*

$$1 \leqslant g(\kappa, S) \leqslant 2\mathrm{e}^{(d+2)/\kappa}(S \vee 1)^{1/\kappa} \exp\left(2\kappa \mathrm{e}^{(d+2)/\kappa}(S \vee 1)^{1/\kappa}\right) , \tag{59}$$

where g is defined in (56)*.*

Proof. The inequality (57) follows from Lemma 26 and the fact that $f_\kappa(u) \leqslant \sum_{m\geqslant 1} m^{-\kappa m}u^m$ for all $u > 0$. The inequality (58) follows exactly the same proof as the second implication of Lemma 5. To prove (59), write, for all $\kappa, S > 0$, $\beta(\kappa, S) = (S\vee 1)\mathrm{e}^{d+2}2^\kappa$ and consider the function $\psi : x \mapsto \beta(\kappa, S)^x x^{-\kappa x+1} = \psi_{\beta(\kappa,S),1}(x)$ where $\psi_{x,d}$ is defined in Lemma 26. By definition, $g(\kappa, S) = \sup_{x\geqslant 1} \psi(x)$. In the proof of Lemma 26, it is shown that this function is upper bounded on $\mathbb{R}_+^*$ by $\beta(\kappa, S)^{1/\kappa} \exp(\kappa\beta(\kappa, S)^{1/\kappa})$ as soon as $\beta(\kappa, S) \geqslant (1/\kappa)^\kappa$, which is always true since $\beta(\kappa, S)^{1/\kappa} \geqslant 2e^{2/\kappa} \geqslant 2 \times 2/\kappa \geqslant 1/\kappa$. □

Lemma 28. *For all $\kappa > 0$, $S < \infty$, $\nu > 0$, $\phi \in \Upsilon_{\kappa,S}$ and $m \geqslant d/\kappa$,*

$$\|\phi - T_m\phi\|_{\infty,\nu} \leqslant 2^d(S\nu)^m m^{-\kappa m+d} f_\kappa(S\nu) ,$$

where $\Upsilon_{\kappa,S}$, f_κ and $T_m\phi$ are defined in (8), (55) *and* (10)*. In particular, for all $S, \nu > 0$, there exists $c > 0$ such that for all $\kappa \in [1/2, 1]$, $m \geqslant 1$ and $\phi \in \Upsilon_{\kappa,S}$,*

$$\|\phi - T_m\phi\|_{\infty,\nu} \leqslant c(S\nu)^m m^{-\kappa m+d} .$$

Proof. Let $\kappa > 0$, $S < \infty$ and $\phi \in \Upsilon_{\kappa,S}$. By definition of $\Upsilon_{\kappa,S}$, for all $x \in \mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2}$,

$$\phi(x) = \sum_{i\in\mathbb{N}^d} c_i \prod_{a=1}^d x_a^{i_a} , \quad \text{where} \quad |c_i| \leqslant S^{\|i\|_1} \|i\|_1^{-\kappa\|i\|_1} .$$

Then, for all $x \in \mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2}$,

$$(\phi - T_m\phi)(x) = \sum_{i\in\mathsf{I}_d^{m,+}} c_i \prod_{a=1}^d x_a^{i_a} ,$$

where, for all $m \in \mathbb{N}$,

$$\mathsf{I}_d^m = \left\{i \in \mathbb{N}^d : \sum_a i_a = m\right\} , \quad \mathsf{I}_d^{m,+} = \left\{i \in \mathbb{N}^d : \sum_a i_a > m\right\} . \tag{60}$$

This yields

$$\|\phi - T_m\phi\|_{\infty,\nu} \leqslant \sum_{i\in\mathsf{I}_d^{m,+}} \|i\|_1^{-\kappa\|i\|_1}(S\nu)^{\|i\|_1}$$

and

$$\|\phi - T_m\phi\|_{\infty,\nu} \leqslant \sum_{m'>m} (m')^{-\kappa m'} (S\nu)^{m'} |\mathsf{I}_d^{m'}| \leqslant \sum_{m'>m} (m')^{-\kappa m'} (S\nu)^{m'} (1+m')^d \,, \\ \leqslant 2^d \sum_{m'>m} (m')^{-\kappa m'} (S\nu)^{m'} (m')^d \,.$$

Therefore, for $m \geqslant d/\kappa$, as for all $m' > m$, $(m')^{-\kappa m + d} \leqslant (m+1)^{-\kappa m + d}$,

$$\|\phi - T_m\phi\|_{\infty,\nu} \leqslant 2^d (S\nu)^m (m+1)^{-\kappa m + d} \sum_{m'>m} (m')^{-\kappa(m'-m)} (S\nu)^{(m'-m)} \,, \\ \leqslant (S\nu)^m m^{-\kappa m + d} 2^d \sum_{m'>0} (m' + d/\kappa)^{-\kappa m'} (S\nu)^{m'} \,,$$

which concludes the proof by definition of f_κ, see (55). □

I Proof of Lemma 20

Let $\kappa, \nu > 0$, $S < \infty$, $m \in \mathbb{N}^*$, $\phi \in \Upsilon_{\kappa,S}$ and $h \in \mathcal{G}_{\kappa,S}$.

$$\begin{aligned} M^{\mathrm{lin}}(T_m h, T_m\phi; \nu) = \int_{\mathsf{B}_\nu^d} &|(T_m h)(t_1,t_2)(T_m\phi)(t_1,0)(T_m\phi)(0,t_2) \\ &- (T_m\phi)(t_1,t_2)(T_m h)(t_1,0)(T_m\phi)(0,t_2) \\ &- (T_m\phi)(t_1,t_2)(T_m\phi)(t_1,0)(T_m h)(0,t_2)|^2 \mathrm{d}t_1 \mathrm{d}t_2 \,, \\ = \|\mathcal{A}(\phi,m)h\|_{2,\nu}^2 \,, \end{aligned} \tag{61}$$

where $\mathcal{A}(\phi,m)$ is a linear operator onto $\mathbf{L}^2(\mathsf{B}_\nu^d)$. Write

$$P_1 = (T_m\phi)(\cdot,0)(T_m\phi)(0,\cdot) \,, \quad P_2 = -(T_m\phi)(\cdot,0)(T_m\phi) \,, \quad P_3 = -(T_m\phi)(0,\cdot)(T_m\phi) \,,$$

so that

$$\mathcal{A}(\phi,m)h = (T_m h)P_1 + (T_m h)(0,\cdot)P_2 + (T_m h)(\cdot,0)P_3 \,.$$

Let H be the vector of coordinates of h in the canonical basis of $\mathbb{C}[X_1,\ldots,X_{2d}]$, for all $(x,y) \in \mathsf{B}_\nu^d$,

$$h(x,y) = \sum_{i,j\in\mathbb{N}^d} H_{(i,j)} \prod_{a,b=1}^d x_a^{i_a} y_b^{j_b} \,.$$

Then $h = H^\top \mathfrak{M}$ where $\mathfrak{M}$ is the vector such that for all $i \in \mathbb{N}^{2d}$,

$$\mathfrak{M}_i = \prod_{a=1}^{2d} X_a^{i_a} \,. \tag{62}$$

Let A be the matrix such that the coordinates of $\mathcal{A}(\phi,m)h$ in the canonical basis are, for all $i \in \mathbb{N}^{2d}$,

$$\mathcal{A}_i = \sum_{j\in\mathbb{N}^{2d}} A_{i,j} H_j \,. \tag{63}$$

Likewise, let J_m be the matrix of the operator T_m in the canonical basis: for all $(i,j) \in \mathbb{N}^{2d} \times \mathbb{N}^{2d}$,

$$(J_m)_{i,j} = \mathbf{1}_{\|i\|_1 \leqslant m} \mathbf{1}_{i=j} \,. \tag{64}$$

Let f, $(P_{1,i})_{i\in\mathbb{N}^{2d}}$, $(P_{2,i})_{i\in\mathbb{N}^{2d}}$ and $(P_{3,i})_{i\in\mathbb{N}^{2d}}$ the vector of coordinates of ϕ, P_1, P_2 and P_3 in the canonical basis. Then, for all $(i,j) \in \mathbb{N}^{2d} \times \mathbb{N}^{2d}$,

$$A_{i,j} = (P_{1,i-j} + P_{2,i-j} + P_{3,i-j})\mathbf{1}_{\|j\|_1 \leqslant m} \,,$$

with the convention $P_{1,i} = P_{2,i} = P_{3,i} = 0$ if there exists $a \in \{1,\ldots,2d\}$ such that $i_a < 0$. For all $(i,j) \in (\mathbb{N}^d)^2$,

$$\begin{cases} P_{1,(i,j)} = f_{(i,0)} f_{(0,j)} \mathbf{1}_{\|i\|_1 \leqslant m} \mathbf{1}_{\|j\|_1 \leqslant m} \,, \\ P_{2,(i,j)} = -\displaystyle\sum_{u\in\mathbb{N}^d : u \leqslant i} f_{(u,j)} f_{(i-u,0)} \mathbf{1}_{\|u\|_1 \leqslant m} \mathbf{1}_{\|i-u\|_1 \leqslant m} \mathbf{1}_{\|j\|_1 \leqslant m} \,, \\ P_{3,(i,j)} = -\displaystyle\sum_{v\in\mathbb{N}^d : v \leqslant j} f_{(i,v)} f_{(0,j-v)} \mathbf{1}_{\|i\|_1 \leqslant m} \mathbf{1}_{\|v\|_1 \leqslant m} \mathbf{1}_{\|j-v\|_1 \leqslant m} \,. \end{cases}$$

Lemma 29. *For all $(i,j) \in \mathbb{N}^d \times \mathbb{N}^d$,*

- *i) $A_{i,j} = 0$ if there exists $a \in \{1,\ldots,2d\}$ such that $j_a \geqslant i_a$ (A is lower triangular) ;*
- *ii) $A_{i,j} = 0$ if $\|j\|_1 > m$, so that $AJ_m = A$;*
- *iii) $A_{i,j} = 0$ if $\|i\|_1 > \|j\|_1 + 2m$, so that $AJ_{m'} = J_{m'+2m}AJ_{m'}$ for all $m' \in \mathbb{N}$;*
- *iv) $A_{i,i} = -\phi(0)^2 = -1$;*
- *v) the coefficient $A_{i,j}$ is upper bounded as follows:*

$$|A_{i,j}| \leqslant S^{\|i-j\|_1} (\|i-j\|_1/2)^{-\kappa\|i-j\|_1} \left\{1 + (1 + \|i_1 - j_1\|_1)^{d_1+1} + (1 + \|i_2 - j_2\|_1)^{d_2+1}\right\} \,,$$

where $i = (i_1, i_2) \in \mathbb{N}^{d_1} \times \mathbb{N}^{d_2}$ and $j = (j_1, j_2) \in \mathbb{N}^{d_1} \times \mathbb{N}^{d_2}$.

Proof. Items i) to iv) are direct consequences of the definitions. Let $(i,j) \in (\mathbb{N}^d)^2$, by definition of $\Upsilon_{\kappa,S}$, $|f_k| \leqslant S^{\|k\|_1} \|k\|_1^{-\kappa\|k\|_1}$ for all $k \in \mathbb{N}^{2d}$. Then, by concavity of $x \mapsto x \log x$,

$$\begin{aligned} |P_{1,(i,j)}| &\leqslant S^{\|i\|_1 + \|j\|_1} \exp\left(-2\kappa\left[\frac{1}{2}\|i\|_1 \log\|i\|_1 + \frac{1}{2}\|j\|_1 \log\|j\|_1\right]\right) \,, \\ &\leqslant S^{\|i\|_1 + \|j\|_1} \exp\left(-\kappa(\|i\|_1 + \|j\|_1) \log\left(\frac{\|i\|_1 + \|j\|_1}{2}\right)\right) \,, \\ &\leqslant S^{\|i\|_1 + \|j\|_1} \left(\frac{\|i\|_1 + \|j\|_1}{2}\right)^{-\kappa(\|i\|_1 + \|j\|_1)} \,. \end{aligned}$$

Similarly, using definition (60),

$$
\begin{aligned}
|P_{2,(i,j)}| &\leqslant S^{\|i\|_1+\|j\|_1} \sum_{u\in\mathbb{N}^d : u\leqslant i} (\|u\|_1+\|j\|_1)^{-\kappa(\|u\|_1+\|j\|_1)} \|i-u\|_1^{-\kappa\|i-u\|_1}\,, \\
&\leqslant S^{\|i\|_1+\|j\|_1} \sum_{k=0}^{\|i\|_1} |\mathsf{I}_d^k| (k+\|j\|_1)^{-\kappa(k+\|j\|_1)} (\|i\|_1-k)^{-\kappa(\|i\|_1-k)}\,, \\
&\leqslant S^{\|i\|_1+\|j\|_1} \sum_{k=0}^{\|i\|_1} |\{0,\ldots,\|i\|_1\}^d| \exp\Big(-2\kappa\Big[\frac{1}{2}(k+\|j\|_1)\log(k+\|j\|_1) \\
&\qquad\qquad + \frac{1}{2}(\|i\|_1-k)\log(\|i\|_1-k)\Big]\Big)\,, \\
&\leqslant S^{\|i\|_1+\|j\|_1} (\|i\|_1+1)^{d+1} \exp\left(-\kappa(\|i\|_1+\|j\|_1)\log\left(\frac{\|i\|_1+\|j\|_1}{2}\right)\right)\,, \\
&\leqslant S^{\|i\|_1+\|j\|_1} (\|i\|_1+1)^{d+1} \left(\frac{\|i\|_1+\|j\|_1}{2}\right)^{-\kappa(\|i\|_1+\|j\|_1)}\,.
\end{aligned}
$$

and

$$
|P_{3,(i,j)}| \leqslant S^{\|i\|_1+\|j\|_1} (\|j\|_1+1)^{d_2+1} \left(\frac{\|i\|_1+\|j\|_1}{2}\right)^{-\kappa(\|i\|_1+\|j\|_1)}\,,
$$

which concludes the proof. □

For all $i \geqslant 0$, let P_i be the i-th Legendre polynomial and P_i^{norm} its normalized version defined as in (30). Let $\mathbf{B}$ be the change-of-basis matrix from the canonical basis formed by the monomials $(\mathfrak{M}_i)_{i\in\mathbb{N}^d}$, where $d = d_1 + d_2$ and $\mathfrak{M}_i$ is defined in (62), to the basis generated by the normalized Legendre polynomials: for all $i \in \mathbb{N}^d$,

$$
\mathbf{P}_i^{\text{norm}}(X_1,\ldots,X_d) = \sum_{j\in\mathbb{N}^d} \mathbf{B}_{i,j} \prod_{a=1}^{d} X_a^{j_a}\,, \tag{65}
$$

where $\mathbf{P}_i^{\text{norm}}$ is defined in (31). Then, $\mathbf{B}_{i,j} = 0$ if there exists $1 \leqslant a \leqslant d$ such that $j_a > i_a$ or such that $i_a - j_a$ is an odd integer. Otherwise, for all $k \in \mathbb{N}^d$ such that $k_a \leqslant i_a/2$, for all $a \in \{1,\ldots,d\}$,

$$
\begin{aligned}
\mathbf{B}_{i,i-2k} &= \nu^{-d/2} \left(\prod_{a=1}^{d}(i_a+1/2)\right)^{1/2} 2^{-\|i\|_1} \nu^{-\|i-2k\|_1} (-1)^{\|k\|_1} \prod_{a=1}^{d} \binom{i_a}{k_a}\binom{2i_a-2k_a}{i_a}\,, \\
&= \nu^{-d/2} \left(\prod_{a=1}^{d}(i_a+1/2)\right)^{1/2} 2^{-\|i\|_1} \nu^{-\|i-2k\|_1} (-1)^{\|k\|_1} \prod_{a=1}^{d} \binom{i_a-k_a}{k_a}\binom{2i_a-2k_a}{i_a-k_a}\,.
\end{aligned} \tag{66}
$$

Lemma 30. *Let $h \in \mathbf{L}^2(\mathsf{B}_\nu^{d_1}\times\mathsf{B}_\nu^{d_2})$ and X be the vector of coordinates of $T_m h$ in the Legendre polynomials basis. Then,*

$$
\|T_m h\|_{2,\nu}^2 = \|X\|^2
$$

and

$$
\|\mathcal{A}(\phi,m)h\|_{2,\nu}^2 = \|X^\top \mathbf{B} A^\top \mathbf{B}^{-1}\|^2 = \|X^\top J_m \mathbf{B} J_m A^\top J_{3m} \mathbf{B}^{-1} J_{3m}\|^2\,,
$$

where A, J_m and $\mathbf{B}$ are defined in (63), (64) *and* (65).

Proof. Let $h \in \mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$ and $\mathfrak{L}$ be the vector of Legendre polynomials. By definition of J_m, as $\mathfrak{L} = \mathbf{B}\mathfrak{M}$, by (65),

$$T_m h = X^\top \mathfrak{L} = X^\top \mathbf{B}\mathfrak{M} = (J_m H)^\top \mathfrak{M}\,.$$

Then, $H^\top J_m = X^\top \mathbf{B} = X^\top \mathbf{B} J_m$ and

$$\begin{aligned}\mathcal{A}(\phi, m)h = (AH)^\top \mathfrak{M} = (AJ_m H)^\top \mathfrak{M} = H^\top J_m A^\top (\mathbf{B}^{-1}\mathfrak{L}) &= X^\top \mathbf{B} A^\top \mathbf{B}^{-1}\mathfrak{L} \\ &= X^\top J_m \mathbf{B} J_m A^\top J_{3m} \mathbf{B}^{-1} J_{3m}\mathfrak{L}\end{aligned}$$

by Lemma 29 and the fact that $J_m\mathbf{B}^{-1} = J_m \mathbf{B}^{-1} J_m$ since $\mathbf{B}^{-1}$ is lower triangular, so that $X^\top = X^\top J_m$. The proof is concluded by noting that Legendre plolynomials form an orthonormal basis of $\mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$, the operator $\mathfrak{L}^\top : \mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2}) \to \mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$ is then norm preserving. □

A lower bound for $\|\mathcal{A}(\phi, m)h\|_{2,\nu}$ may then be obtained by lower bounding the smallest singular values of $J_m \mathbf{B} J_m$, $J_m A^\top J_{3m}$ and $J_{3m}\mathbf{B}^{-1} J_{3m}$ as

$$\begin{aligned}\inf_{h \in \mathcal{G}_{\kappa,S}} \frac{\|\mathcal{A}(\phi, m)h\|_{2,\nu}}{\|T_m h\|_{2,\nu}} &\geqslant \inf_{X \in \mathrm{Im}(J_m)} \frac{\|X^\top J_m \mathbf{B} J_m A^\top J_{3m}\mathbf{B}^{-1} J_{3m}\|}{\|X^\top\|}\,, \\ &\geqslant \sigma_{\mathrm{rk}(J_m)}(J_m \mathbf{B} J_m)\sigma_{\mathrm{rk}(J_m)}(J_{3m} A J_m)\sigma_{\mathrm{rk}(J_{3m})}(J_{3m}\mathbf{B}^{-1}J_{3m})\,, \\ &= \sigma_1(J_m \mathbf{B}^{-1})^{-1}\sigma_{\mathrm{rk}(J_m)}(AJ_m)\sigma_1(J_{3m}\mathbf{B})^{-1}\,. \end{aligned} \tag{67}$$

The following lemmas allow to control the three terms of equation (67).

Lemma 31. *Let $d = d_1 + d_2$. For all $m \in \mathbb{N}^*$ and all $\nu > 0$,*

$$\sigma_1(J_m \mathbf{B}) \leqslant \nu^{-d/2} m^d 4^m (\nu^{-1} \vee 1)^m\,,$$

where J_m and $\mathbf{B}$ are defined in (64) *and* (65).

Proof. For all $k, i \in \mathbb{N}$ such that $k \leqslant i$,

$$\binom{i}{k} \leqslant \binom{i}{i/2} \sim 2^i/\sqrt{\pi i/2} \quad \text{and} \quad \binom{2i-2k}{i} \leqslant \binom{2i}{i} \sim 4^i/\sqrt{\pi i}\,.$$

Thus, by (66), for all $\nu > 0$,

$$|\mathbf{B}_{i,i-2k}| \leqslant \nu^{-d/2} \prod_{a=1}^{d} (4/\nu)^{i_a}\nu^{2k_a} \leqslant \nu^{-d/2} 4^{\|i\|_1}\left(\nu^{-1} \vee 1\right)^{\|i\|_1 - 2\|k\|_1}\,.$$

Then,

$$\begin{aligned}\sigma_1(J_m \mathbf{B} J_m) &\leqslant |\{(i,j) \in \mathbb{N}^d \times \mathbb{N}^d : (J_m \mathbf{B} J_m)_{i,j} \neq 0\}|^{1/2} \|J_m \mathbf{B} J_m\|_\infty \\ &\leqslant \mathrm{rk}(J_m)\|J_m \mathbf{B} J_m\|_\infty \leqslant m^d \|J_m \mathbf{B} J_m\|_\infty\,,\end{aligned}$$

which yields $\sigma_1(J_m \mathbf{B} J_m) \leqslant \nu^{-d/2} m^d 4^m (\nu^{-1} \vee 1)^m$. □

Lemma 32. *Let $d = d_1 + d_2$. For all $m \in \mathbb{N}^*$ and all $\nu > 0$,*

$$\sigma_1(J_m \mathbf{B}^{-1}) \leqslant \sqrt{2} 2^d m^{(d+1)/2} \nu^{d/2} (\nu \vee 1)^m\,,$$

where J_m and $\mathbf{B}$ are defined in (64) *and* (65).

Proof. Write $\mathfrak{L}$ the vector of Legendre polynomials. By definition of $\mathfrak{M}$ and $\mathbf{B}$, see (62) and (65), $\mathfrak{L} = \mathbf{B}\mathfrak{M}$ and for all $i \in \mathbb{N}^d$,

$$\|\mathfrak{M}_i\|_{2,\nu}^2 = \left\| \sum_{j\in\mathbb{N}^d} (\mathbf{B}^{-1})_{i,j}\mathfrak{L}_j \right\|_{2,\nu}^2 = \sum_{j\in\mathbb{N}^d} (\mathbf{B}^{-1})_{i,j}^2 ,$$

as Legendre polynomials form an orthonormal basis of $\mathbf{L}^2(\mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2})$. Then, using that

$$\|\mathfrak{M}_i\|_{2,\nu}^2 = \prod_{a=1}^{d} \frac{2\nu^{2i_a+1}}{2i_a+1} ,$$

yields

$$\|J_m\mathbf{B}^{-1}J_m\|_F^2 \leqslant \begin{cases} 2^d(m+1)^d\nu^d & \text{if } \nu \leqslant 1 , \\ 2^d(m+1)^{d+1}\nu^{2m+d} & \text{if } \nu > 1 . \end{cases}$$

The proof is concluded by writing $\sigma_1(J_m\mathbf{B}^{-1}) = \sigma_1(J_m\mathbf{B}^{-1}J_m) \leqslant \|J_m\mathbf{B}^{-1}J_m\|_F$. □

Lemma 33. *Let $d = d_1 + d_2$. Then,*

$$\sigma_{rk(J_m)}(AJ_m) \geqslant 4^{-1}\,(2\sqrt{2})^{-d}m^{-d-1}(d\mathrm{e})^{-3m}g(\kappa, S, d_1, d_2)^{-3m} ,$$

where g is defined in (56).

Proof. By Lemma 29, $A = -J_m + N$ where N is a $(3m + 1)$-nilpotent strict lower triangular matrix. Let $D = -\sum_{k=0}^{3m} N^k$ with the convention $N^0 = J_m$, then $DA = J_m$, and

$$\sigma_{\mathrm{rk}(J_m)}(A) \geqslant \sigma_1(D)^{-1} .$$

Therefore,

$$\sigma_{\mathrm{rk}(J_m)}(A)^{-1} \leqslant \sigma_1\left(-\sum_{k=0}^{3m-1} N^k\right) \leqslant \sigma_1\left(J_{3m}\sum_{k=0}^{3m-1} N^k J_m\right) \leqslant (\mathrm{rk}(J_m)\mathrm{rk}(J_{3m}))^{1/2}\sum_{k=0}^{3m}\|N^k\|_\infty$$

and

$$\sigma_{\mathrm{rk}(J_m)}(A)^{-1} \leqslant ((3m+1)(m+1))^{d/2}\left(1 + \sum_{k=1}^{3m}\sup_{i,j\in\mathbb{N}^d}\left|\sum_{\substack{a^{(0)},a^{(1)},\dots,a^{(k)}\in\mathbb{N}^d \text{ distincts} \\ i=a^{(0)}\leqslant a^{(1)}\leqslant\dots\leqslant a^{(k)}=j}}\prod_{u=1}^{k} N_{a^{(u-1)},a^{(u)}}\right|\right) .$$

Let $k \in \mathbb{N}^*$ and $i = a^{(0)} \leqslant a^{(1)} \leqslant \dots \leqslant a^{(k)} = j$ distinct in $\mathbb{N}^d$. By Lemma 29, writing for all $u \geqslant 0$, $a^{(u)} = (a_1^{(u)}, a_2^{(u)}) \in \mathbb{N}^{d_1} \times \mathbb{N}^{d_2}$,

$$|N_{a^{(u-1)},a^{(u)}}| \leqslant S^{\|a^{(u)}-a^{(u-1)}\|_1}(\|a^{(u)} - a^{(u-1)}\|_1/2)^{-\kappa\|a^{(u)}-a^{(u-1)}\|_1} \times \left\{1 + (1 + \|a_1^{(u)} - a_1^{(u-1)}\|_1)^{d_1+1} + (1 + \|a_2^{(u)} - a_2^{(u-1)}\|_1)^{d_2+1}\right\} ,$$

so that

$$
\begin{aligned}
&\left|\prod_{u=1}^{k} N_{a^{(u-1)},a^{(u)}}\right| \\
&\quad \leqslant S^{\sum_{u=1}^{k}\|a^{(u)}-a^{(u-1)}\|_1} \exp\left(-\kappa k \sum_{u=1}^{k} \frac{1}{k}\|a^{(u)}-a^{(u-1)}\|_1 \log(\|a^{(u)}-a^{(u-1)}\|_1/2)\right) \\
&\qquad\qquad \times \prod_{u=1}^{k} (\|a_1^{(u)}-a_1^{(u-1)}\|_1+1)^{d_1+1}(\|a_2^{(u)}-a_2^{(u-1)}\|_1+1)^{d_2+1}\,, \\
&\quad \leqslant S^{\|j-i\|_1} \exp\left((d_1+1)\|j_1-i_1\|_1+(d_2+1)\|j_2-i_2\|_1\right) \exp\left(-\kappa\|j-i\|_1 \log\frac{\|j-i\|_1}{2k}\right)\,, \\
&\quad \leqslant S^{\|j-i\|_1} (\mathrm{e}^{d_1+1})^{\|j_1-i_1\|_1} (\mathrm{e}^{d_2+1})^{\|j_2-i_2\|_1} \left(\frac{\|j-i\|_1}{2k}\right)^{-\kappa\|j-i\|_1}\,,
\end{aligned}
$$

using that $x \mapsto x\log x$ is convex, $\log(1+x) \leqslant x$ for $x \geqslant 0$ and using

$$
\sum_{u=1}^{k} \|a^{(u)}-a^{(u-1)}\|_1 = \sum_{u=1}^{k} (\|a^{(u)}\|_1 - \|a^{(u-1)}\|_1) = \|j-i\|_1\,.
$$

It remains to count

$$
\mathsf{s}_{i,j}^k = \sum_{\substack{a^{(0)},a^{(1)},\dots,a^{(k)}\in\mathbb{N}^{2d} \text{ distincts} \\ i=a^{(0)}\leqslant a^{(1)}\leqslant\dots\leqslant a^{(k)}=j}} 1\,.
$$

This sum counts the number of paths connecting i and j, going away from 0 in $\mathbb{N}^d$ and made of k steps with non-zero length. Thus, it is upper bounded by the number of paths of length $\|j-i\|_1$ going away from zero in $\mathbb{N}^d$ and made of k steps with nonzero length. Such a path is entirely described by the direction of each step (d possibilities each) and the length of each step (or equivalently the distance travelled after each of the first $k-1$ steps, which is equivalent to choosing $k-1$ distinct integers in $\{1,\dots,\|j-i\|_1-1\}$). Therefore

$$
\mathsf{s}_{i,j}^k \leqslant d^k \binom{\|j-i\|_1-1}{k-1} \leqslant d^k\|j-i\|_1^k/k! \leqslant d^k\|j-i\|_1^k(\mathrm{e}/k)^k\,,
$$

and

$$
\begin{aligned}
\sigma_{\mathrm{rk}(J_m)}(AJ_m)^{-1} &\leqslant ((3m+1)(m+1))^{d/2}\left(1+\sum_{k=1}^{3m}(d\mathrm{e})^k \sup_{\ell\geqslant k}(S\mathrm{e}^{d_1+1}\mathrm{e}^{d_2+1}2^\kappa)^\ell\left(\frac{\ell}{k}\right)^{-\kappa\ell+k}\right)\,, \\
&\leqslant ((3m+1)(m+1))^{d/2}\left(1+\sum_{k=1}^{3m}\left(d\mathrm{e}\sup_{x\geqslant 1}(S\mathrm{e}^{d_1+1}\mathrm{e}^{d_2+1}2^\kappa)^x x^{-\kappa x+1}\right)^k\right)\,, \\
&\leqslant ((3m+1)(m+1))^{d/2}(3m+1)\max\left(1, d\mathrm{e}\sup_{x\geqslant 1}(S\mathrm{e}^{d_1+1}\mathrm{e}^{d_2+1}2^\kappa)^x x^{-\kappa x+1}\right)^{3m}\,, \\
&\leqslant 4(2\sqrt{2})^d m^{d+1}\max\left(1, d\mathrm{e}\sup_{x\geqslant 1}(S\mathrm{e}^{d_1+1}\mathrm{e}^{d_2+1}2^\kappa)^x x^{-\kappa x+1}\right)^{3m}\,,
\end{aligned}
$$

which concludes the proof by (56). □

The proof of Lemma 20 may then be completed. By equations (61) and (67) and the three above lemmas, there exists a numerical constant $\mathsf{c} > 0$ such that

$$\begin{aligned} M^{\mathrm{lin}}(T_m h, T_m \phi; \nu) &\\ &\geqslant \sigma_1(J_m \mathbf{B}^{-1})^{-2} \sigma_{\mathrm{rk}(J_m)}(A J_m)^2 \sigma_1(J_{3m}\mathbf{B})^{-2} \|T_m h\|_{2,\nu}^2 , \\ &\geqslant \mathsf{c}(4^{-d} m^{-d-1} \nu^{-d} (\nu \vee 1)^{-2m}) \times ((2\sqrt{2})^{-2d} m^{-2d-2} (d\mathrm{e})^{-6m} g(\kappa, S, d_1, d_2)^{-6m}) \\ &\quad \times (\nu^d (3m)^{-2d} 4^{-6m} \left(\nu^{-1} \vee 1\right)^{-6m}) \|T_m h\|_{2,\nu}^2 , \\ &\geqslant \mathsf{c}(4\sqrt{2})^{-2d} (4\mathrm{e})^{-6m} m^{-5d-3} (\nu \vee \nu^{-3})^{-2m} g(\kappa, S, d_1, d_2)^{-6m} d^{-6m} \|T_m h\|_{2,\nu}^2 . \end{aligned}$$

J Proofs of Section B

J.1 Proof of Lemma 21

Let $S, \nu > 0$, $\kappa \in [1/2, 1]$, $m \geqslant 1$, $\phi \in \Upsilon_{\kappa,S}$ and $h \in \mathcal{G}_{\kappa,S}$ and write $V = h - T_m h$ and $U = \phi - T_m \phi$. Using the inequality $|a+b|^2 \geqslant |a|^2/2 - |b|^2$ for all $(a,b) \in \mathbb{C}^2$,

$$M^{\mathrm{lin}}(h, \phi; \nu) \geqslant \frac{1}{2} M^{\mathrm{lin}}(T_m h, \phi; \nu) - 9(2\nu)^d \|V\|_{\infty,\nu}^2 \|\phi\|_{\infty,\nu}^4 .$$

By Lemma 28, $\|V\|_{\infty,\nu} \leqslant \mathrm{c}(S\nu)^m m^{-\kappa m + d}$ for some $\mathrm{c} > 0$ that only depends on S and ν, so that

$$M^{\mathrm{lin}}(h, \phi; \nu) \geqslant \frac{1}{2} M^{\mathrm{lin}}(T_m h, \phi; \nu) - 9(2\nu)^d 2^{2d} (S\nu)^{2m} m^{-2\kappa m + 2d} f_\kappa(S\nu)^2 C_\Upsilon^4(\kappa, S, \nu) .$$

$$M^{\mathrm{lin}}(h, \phi; \nu) \geqslant \frac{1}{2} M^{\mathrm{lin}}(T_m h, \phi; \nu) - \mathrm{c}'(S\nu)^{2m} m^{-2\kappa m + 2d}$$

for some $\mathrm{c}' > 0$ which only depends on S and ν. Similarly,

$$\begin{aligned} M^{\mathrm{lin}}(h, \phi; \nu) &\geqslant \frac{1}{2} M^{\mathrm{lin}}(h, T_m \phi; \nu) - (2\nu)^d \|U\|_{\infty,\nu}^2 \|h\|_{\infty,\nu}^2 (6\|T_m\phi\|_{\infty,\nu} + 3\|U\|_{\infty,\nu})^2 , \\ &\geqslant \frac{1}{2} M^{\mathrm{lin}}(h, T_m \phi; \nu) - (2\nu)^d 2^{2d} (S\nu)^{2m} m^{-2\kappa m + 2d} f_\kappa(S\nu)^2 (2C_\Upsilon(\kappa, S, \nu))^2 \\ &\qquad \times (6C_\Upsilon(\kappa, S, \nu) + 3 \times 2C_\Upsilon(\kappa, S, \nu))^2 , \\ &\geqslant \frac{1}{2} M^{\mathrm{lin}}(h, T_m \phi; \nu) - \mathrm{c}''(S\nu)^{2m} m^{-2\kappa m + 2d} , \end{aligned}$$

for some $\mathrm{c}'' > 0$ which only depends on S, ν and d. Then, by Lemma 20, there exists $\mathrm{c}, \mathrm{c}' > 0$ and $C > 1$ depending only on S and ν such that

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

Finally, we conclude with the inequalities

$$\|T_m h\|_{2,\nu}^2 \geqslant \|h\|_{2,\nu}^2 / 2 - \|h - T_m h\|_{2,\nu}^2 ,$$

and $\|h - T_m h\|_{2,\nu}^2 \leqslant \mathrm{c}(S\nu)^{2m} m^{-2\kappa m + 2d}$ by Lemma 28 for some constant c which only depends on S and ν.

J.2 Proof of Proposition 22

Let $S, \nu > 0$. Let $\mathrm{c}, \mathrm{c}' > 0$ and $C > 1$ be as in Lemma 21, so that for all $\kappa \in [1/2, 1]$, $m \geqslant 1$, $\phi \in \Upsilon_{\kappa,S}$ and $h \in \mathcal{G}_{\kappa,S}$,

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

The proposition will follow from a careful choice of m depending on $\|h\|_{2,\nu}$. Assume that

$$\frac{\mathrm{c}}{2} m^{-5d-3} C^{-m} \|h\|_{2,\nu}^2 \geqslant \mathrm{c}'(S\nu)^{2m} m^{-2\kappa m + 2d} , \tag{68}$$

then

$$M^{\mathrm{lin}}(h, \phi; \nu) \geqslant \frac{\mathrm{c}}{2} m^{-5d-3} C^{-m} \|h\|_{2,\nu}^2 . \tag{69}$$

Note that (68) is equivalent to

$$m^{\kappa m - a} b^m \geqslant \mathrm{c}'' \frac{1}{\|h\|_{2,\nu}} , \tag{70}$$

where $a = (7d+3)/2$, $b = C/(S\nu)$ and $\mathrm{c}'' = 2\mathrm{c}'/\mathrm{c}$. Thus, (70) hold when

$$\kappa m \log(\kappa m) - (\kappa m \log \kappa + \log(b^{-1}) m + a \log m + \log \mathrm{c}'') \geqslant \log(1/\|h\|_{2,\nu}) . \tag{71}$$

Equation (71) is satisfied when

$$\kappa m \left[\log(\kappa m) - \left(\log(b^{-1})/\kappa + a/\kappa + \log(\mathrm{c}'')/\kappa\right)\right] \geqslant \log(1/\|h\|_{2,\nu}) ,$$

which can be written

$$\kappa m \log\left(\left(\frac{b}{\mathrm{c}'' \mathrm{e}^a}\right)^{1/\kappa} \kappa m\right) \geqslant \log(1/\|h\|_{2,\nu}) . \tag{72}$$

Note that for all $A > 1$, the solution x of the equation $x \log x = A$ satisfies $x \leqslant 3A/(2\log A)$, so that choosing

$$m \geqslant \frac{3}{2\kappa} \frac{\log(1/\|h\|_{2,\nu})}{\log\left\{\left(1 \wedge \frac{b}{\mathrm{c}''\mathrm{e}^a}\right)^{1/\kappa} \log(1/\|h\|_{2,\nu})\right\}}$$

ensures that (72) holds as soon as

$$\left(1 \wedge \frac{b}{\mathrm{c}'' \mathrm{e}^a}\right)^{1/\kappa} \log(1/\|h\|_{2,\nu}) > 1 .$$

Assume that $\|h\|_{2,\nu}$ is small enough that $(1 \wedge b/(cst''\mathrm{e}^a))^{1/\kappa} \log(1/\|h\|_{2,\nu}) \geqslant \mathrm{e}$. Then, a valid choice of m is

$$m = \left\lfloor \frac{2}{\kappa} \frac{\log(1/\|h\|_{2,\nu})}{\log\left\{\left(1 \wedge \frac{b}{\mathrm{c}''\mathrm{e}^a}\right)^{1/\kappa} \log(1/\|h\|_{2,\nu})\right\}} \right\rfloor$$

provided $(2\kappa)^{-1} \log(1/\|h\|_{2,\nu}) \geqslant 1$. Assume in addition that $(1 \wedge b/(\mathrm{c}''\mathrm{e}^a))^{2/\kappa} \log(1/\|h\|_{2,\nu}) \geqslant 1$, then this choice of m implies

$$m \leqslant \frac{4}{\kappa} \frac{\log(1/\|h\|_{2,\nu})}{\log\log(1/\|h\|_{2,\nu})} .$$

Together with (69), this yields

$$M^{\text{lin}}(h, \phi; \nu) \geqslant \frac{\mathrm{c}}{2}(\kappa/4)^{5d+3}\left(\frac{\log\log(1/\|h\|_{2,\nu})}{\log(1/\|h\|_{2,\nu})}\right)^{5d+3}\|h\|_{2,\nu}^{\frac{4\log C}{\kappa\log\log(1/\|h\|_{2,\nu})}}\|h\|_{2,\nu}^2\,,$$

which concludes the proof.

J.3 Proof of Lemma 23

Let $S, \nu > 0$. For all $m \in \mathbb{N}^*$ and $h \in \mathbb{C}_m[X_1, \dots, X_d]$, there exists a unique matrix $H = (H_{i,j})_{i\in\mathbb{N}^{d_1}, j\in\mathbb{N}^{d_2}}$ such that for all $(x, y) \in \mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2}$, $h(x, y) = \sum_{i\in\mathbb{N}^{d_1}, j\in\mathbb{N}^{d_2}} H_{i,j}\mathbf{P}_i^{\text{norm}}(x)\mathbf{P}_j^{\text{norm}}(y)$, with $\mathbf{P}_i^{\text{norm}}$ defined in equation (31). Write in the following sentence 0_{d_2} for the zero of $\mathbb{R}^{d_2}$. Note that $\mathbf{P}_{0_{d_2}}^{\text{norm}}(x) = (2\nu)^{-d_2/2}$ for all $x \in \mathbb{R}^{d_1}$, so that for all $(x, y) \in \mathsf{B}_\nu^{d_1} \times \mathsf{B}_\nu^{d_2}$,

$$h^{(1)}(x, y) = \sum_{i\in\mathbb{N}^{d_1}}\left((2\nu)^{d_2/2}\sum_{j\in\mathbb{N}^{d_2}} H_{i,j}\mathbf{P}_j^{\text{norm}}(0)\right)\mathbf{P}_i^{\text{norm}}(x)\mathbf{P}_{0_{d_2}}^{\text{norm}}(y)\,.$$

Since $H_{i,j} = 0$ if $\|i\|_1 + \|j\|_1 > m$ (as $\deg(h) \leqslant m$), by Cauchy-Schwarz inequality,

$$\begin{aligned}
\|h^{(1)}\|_{2,\nu}^2 &= (2\nu)^{d_2}\sum_{i\in\mathbb{N}^{d_1}}\left|\sum_{j\in\mathbb{N}^{d_2}:\|j\|_1\leqslant m} H_{i,j}\mathbf{P}_j^{\text{norm}}(0)\right|^2 \\
&\leqslant (2\nu)^{d_2}\sum_{i\in\mathbb{N}^{d_1}}\sum_{j\in\mathbb{N}^{d_2}}|H_{i,j}|^2\sum_{j'\in\mathbb{N}^{d_2}:\|j'\|_1\leqslant m}\mathbf{P}_{j'}^{\text{norm}}(0)^2\,, \\
&= (2\nu)^{d_2}\sum_{i\in\mathbb{N}^{d_1}}\sum_{j\in\mathbb{N}^{d_2}}|H_{i,j}|^2\sum_{j'\in\mathbb{N}^{d_2}:\|j'\|_1\leqslant m}\prod_{a=1}^{d_2}\left(4^{-j'_a}\sqrt{\frac{j'_a+1/2}{\nu}}\binom{2j'_a}{j'_a}\right)^2.
\end{aligned}$$

By Stirling's formula, for all $j_a \in \mathbb{N}$,

$$4^{-j'_a}\sqrt{j'_a+1/2}\binom{2j'_a}{j'_a} \leqslant \sqrt{2/\pi}\,.$$

Then, there exists a numerical constant $c > 0$ such that

$$\|h^{(1)}\|_{2,\nu}^2 \leqslant \sum_{i\in\mathbb{N}^{d_1}}\sum_{j\in\mathbb{N}^{d_2}}|H_{i,j}|^2\sum_{j'\in\mathbb{N}^{d_2}:\|j'\|_1\leqslant m}(c/\nu)^{d_2} \leqslant c^{d_2}\|H\|_F^2|\{0,\dots,m\}^{d_2}| \leqslant (2cm)^{d_2}\|h\|_{2,\nu}^2\,.$$

Let $\mathrm{c}_T > 0$ be the constant from Lemma 28. Let $\kappa \in [1/2, 1]$, $m \in \mathbb{N}^*$. Assume now that $h \in \mathcal{G}_{\kappa,S}$. We no longer assume that h is a polynomial. Then,

$$\begin{aligned}
\|h^{(1)}\|_{2,\nu}^2 = \|T_m h^{(1)} + (h - T_m h)^{(1)}\|_{2,\nu}^2 &\leqslant 2\|T_m h^{(1)}\|_{2,\nu}^2 + 2\mathrm{c}_T(S\nu)^{2m}m^{-2\kappa m+2d}\,, \\
&\leqslant 2(2cm)^{d_2}\|T_m h\|_{2,\nu}^2 + 2\mathrm{c}_T(S\nu)^{2m}m^{-2\kappa m+2d}\,, \\
&\leqslant 2(2cm)^{d_2}\|h - (h - T_m h)\|_{2,\nu}^2 + 2\mathrm{c}_T(S\nu)^{2m}m^{-2\kappa m+2d}\,, \\
&\leqslant 4(2cm)^{d_2}\|h\|_{2,\nu}^2 + (4(2cm)^{d_2} + 2)\mathrm{c}_T(S\nu)^{2m}m^{-2\kappa m+2d}\,, \\
&\leqslant \mathrm{c}' m^{d_2}\left(\|h\|_{2,\nu}^2 + (S\nu)^{2m}m^{-2\kappa m+2d}\right)\,,
\end{aligned}$$

for some $c' > 0$ which only depends on S and ν. Following the same steps for $\|h^{(2)}\|_{2,\nu}^2$ yields

$$\begin{aligned}\|h^{(1)}h^{(2)}\|_{2,\nu}^2 &= (2\nu)^{-d}\|h^{(1)}\|_{2,\nu}^2\|h^{(2)}\|_{2,\nu}^2\,,\\ &\leqslant (2\nu)^{-d}(c')^2 m^d\left(\|h\|_{2,\nu}^2 + (S\nu)^{2m}m^{-2\kappa m+2d}\right)^2\,,\\ &\leqslant 2(2\nu)^{-d}(c')^2 m^d\left(\|h\|_{2,\nu}^4 + (S\nu)^{4m}m^{-4\kappa m+4d}\right)\,,\end{aligned}$$

which concludes the proof.

J.4 Proof of Proposition 24

Let $S, \nu > 0$. By Lemma 23, there exists $\mathsf{c} > 0$ such that for all $\kappa \in [1/2, 1]$, $m \in \mathbb{N}^*$ and $h \in \mathcal{G}_{\kappa,S}$,

$$\|h^{(1)}h^{(2)}\|_{2,\nu}^2 \leqslant \mathsf{c}m^d\left(\|h\|_{2,\nu}^4 + (S\nu)^{4m}m^{-4\kappa m+4d}\right)\,.$$

Assume that

$$\|h\|_{2,\nu} \geqslant (S\nu)^m m^{-\kappa m+d}\,, \tag{73}$$

then,

$$\|h^{(1)}h^{(2)}\|_{2,\nu}^2 \leqslant 2\mathsf{c}m^d\|h\|_{2,\nu}^4\,. \tag{74}$$

Assumption (73) can be written

$$m^{\kappa m-a}b^m \geqslant \frac{1}{\|h\|_{2,\nu}}\,, \tag{75}$$

where $a = (S\nu)^{-1}$. Following the same steps as for equation (70) yields that choosing

$$m = \left\lfloor \frac{2}{\kappa}\frac{\log(1/\|h\|_{2,\nu})}{\log\left\{\left(1 \wedge a/\mathrm{e}^d\right)^{1/\kappa}\log(1/\|h\|_{2,\nu})\right\}}\right\rfloor\,, \tag{76}$$

ensures that (75) holds as soon as

$$\left(1 \wedge a/\mathrm{e}^d\right)^{1/\kappa}\log(1/\|h\|_{2,\nu}) \geqslant \mathrm{e} \quad \text{and} \quad (2\kappa)^{-1}\log(1/\|h\|_{2,\nu}) \geqslant 1\,.$$

If moreover $(1 \wedge a/\mathrm{e}^d)^{2/\kappa}\log(1/\|h\|_{2,\nu}) \geqslant 1$, then the choice (76) implies

$$m \leqslant \frac{4}{\kappa}\frac{\log(1/\|h\|_{2,\nu})}{\log\log(1/\|h\|_{2,\nu})}\,.$$

Together with (74), this implies that

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

K Proofs of Section 4

K.1 Proof of Lemma 11

Let H be defined by $h_\kappa(x) = H(x/x_0)/x_0$. Then, $\zeta \leqslant c([x \mapsto h_\kappa(x)(1+(x/x_0)^2)^\tau] * u_b)$ is equivalent to $x_0\zeta(x_0 x) \leqslant c([z \mapsto H(z)(1+z^2)^\tau] * u_{bx_0})(x)$. In this proof, we show that there exists A and B such that for all $\lambda \in \mathbb{R}$ and $b \geqslant 1$,

$$\int \mathrm{e}^{\lambda x}([z \mapsto H(z)(1+z^2)^\tau] * u_b)(x)\mathrm{d}x \leqslant A\mathrm{e}^{B|\lambda|^{1/\kappa}} ,$$

in other words $[z \mapsto H(z)(1+z^2)^\tau] * u_b \in \mathcal{M}^1_{1/\kappa}$, which entails $(x \mapsto x_0\zeta(x_0 x)) \in \mathcal{M}^1_{1/\kappa}$ and thus $\mathcal{F}[x \mapsto x_0\zeta(x_0 x)] = \mathcal{F}[\zeta](\cdot/x_0) \in \Upsilon_{\kappa,T'}$ for some T' by Lemma 5. This ensures $\mathcal{F}[\zeta] \in \Upsilon_{\kappa,T'x_0}$ for all $b \geqslant 1/x_0$, which yields the result by choosing x_0 small enough. Let $c_H = c_h/x_0$ be the normalizing constant of H, then for all $b \geqslant 1$ and $\lambda \in \mathbb{R}$,

$$\begin{aligned}
&\int \mathrm{e}^{\lambda x}([z \mapsto H(z)(1+z^2)^\tau] * u_b)(x)\mathrm{d}x \\
&\qquad \leqslant \int \mathrm{e}^{\lambda x} \sup_{y\in[x-1/b,x+1/b]} H(y)(1+y^2)^\tau \mathrm{d}x \\
&\qquad \leqslant 2^\tau \frac{2c_H}{b} + 2c_H \int_{x\geqslant 0} \mathrm{e}^{|\lambda|(x+1/b)}(1+x^2)^\tau \mathrm{e}^{-([1+x^2]/2)^{1/(2(1-\kappa))}}\mathrm{d}x \\
&\qquad \leqslant 2^{1+\tau}c_H + 2c_H\mathrm{e}^{|\lambda|/b}\int_{x\geqslant 0}(1+x^2)^\tau \mathrm{e}^{|\lambda|x-(x/\sqrt{2})^{1/(1-\kappa)}}\mathrm{d}x \\
&\qquad \leqslant 2^{1+\tau}c_H + 2c_H\mathrm{e}^{|\lambda|}X_\lambda(1+X_\lambda^2)^\tau \mathrm{e}^{|\lambda|X_\lambda} \\
&\qquad\quad + 2c_H\mathrm{e}^{|\lambda|}\int_{x\geqslant X_\lambda} \mathrm{e}^{|\lambda|x+\tau\log(1+x^2)-(x/\sqrt{2})^{1/(1-\kappa)}}\mathrm{d}x
\end{aligned}$$

for all $X_\lambda > 0$. Let X_λ be such that $|\lambda|x + \tau\log(1+x^2) - (x/\sqrt{2})^{1/(1-\kappa)} \leqslant -(1/2)(x/\sqrt{2})^{1/(1-\kappa)}$ for all $x \geqslant X_\lambda$. Taking $X_\lambda = c_X|\lambda|^{-1+1/\kappa}$ works for λ large enough for an appropriate constant c_X. Then for λ large enough,

$$\begin{aligned}
\int e^{\lambda x}([z \mapsto H(z)(1+z^2)^\tau] * u_b)(x)\mathrm{d}x &\leqslant 2^{1+\tau}c_H + 2c_H\mathrm{e}^{|\lambda|}c_X|\lambda|^{-1+1/\kappa}(1+c_X^2|\lambda|^{-2+2/\kappa})^\tau e^{c_X|\lambda|^{1/\kappa}} \\
&\quad + 2c_H\mathrm{e}^{|\lambda|}\int_{x\geqslant 0}\mathrm{e}^{-2^{-1}(\frac{X_\lambda+x}{\sqrt{2}})^{1/(1-\kappa)}}\mathrm{d}x\,, \\
&\leqslant c\cdot\mathrm{e}^{\mathrm{cst}'|\lambda|^{1/\kappa}} + 2c_H\mathrm{e}^{|\lambda|}\exp(-2^{-1-1/(2(1-\kappa))}|\lambda|^{1/\kappa})\int_{x\geqslant 0}\mathrm{e}^{-2^{-1}(x/\sqrt{2})^{1/(1-\kappa)}}\mathrm{d}x\,, \\
&\leqslant A\cdot\mathrm{e}^{B|\lambda|^{1/\kappa}}\,,
\end{aligned}$$

by convexity of $x \mapsto x^{1/(1-\kappa)}$ for some constants A and B depending only on κ. Small values of λ are dealt with by changing A if necessary.

K.2 Proof of Corollary 13

The first inequality follows from the bound on $\|P_K h_\kappa / F_{\text{env}}\|_\infty$: there exists a constant c such that

$$\|P_K h_\kappa^2\|^2_{\mathbf{L}^2(\mathbb{R})} \leqslant cK^{\kappa-1} \|F_{\text{env}} h_\kappa\|^2_{\mathbf{L}^2(\mathbb{R})} ,$$

and the polynomial growth assumption on F_{env} ensures that $\|F_{\text{env}} h_\kappa\|^2_{\mathbf{L}^2(\mathbb{R})} < \infty$.

The second inequality is a consequence of Cauchy-Schwarz' inequality: for any function φ (here $P_K h_\kappa^2$),

$$\begin{aligned}
\|\varphi * u_b\|^2_{\mathbf{L}^2(\mathbb{R})} &= \int \left(\int \varphi(y) u_b(x-y) \mathrm{d}y \right)^2 \mathrm{d}x \\
&\leqslant \int \left(\int \varphi(y)^2 u_b(x-y) \mathrm{d}y \right) \left(\int u_b(x-y) \mathrm{d}y \right) \mathrm{d}x \\
&= \|\varphi\|^2_{\mathbf{L}^2(\mathbb{R})} .
\end{aligned}$$

For the third inequality, let c_0, c_1, c_2 be the constants of Conjecture 12. Write $(I_i)_i = ([s_i, t_i])_i$ the intervals of Conjecture 12. Assume $b \geqslant 2K^\kappa / c_1$, so that the support of u_b has length smaller than $c_1 K^{-\kappa}$. Then for all i and for all $x \in [s_i + b^{-1}, t_i - b^{-1}]$ (which are non-empty intervals by the assumption on b),

$$\begin{aligned}
((P_K h_\kappa^2) * u_b)(x) &= \int_{y \in [-b^{-1}, b^{-1}]} (P_K h_\kappa^2)(x-y) u_b(y) \mathrm{d}y \\
&\geqslant c_2 K^{(\kappa-1)/2} \left(\inf_{[-1,1]} h_\kappa \right) \int u_b(y) \mathrm{d}y \\
&= c_2 K^{(\kappa-1)/2} \left(\inf_{[-1,1]} h_\kappa \right)
\end{aligned}$$

so that

$$\begin{aligned}
\|(P_K h_\kappa^2) * u_b\|^2 &\geqslant \sum_i \int_{[s_i + b^{-1}, t_i - b^{-1}]} ((P_K h_\kappa^2) * u_b)^2(x) \mathrm{d}x \\
&\geqslant \sum_i (t_i - s_i - 2b^{-1}) c_2^2 K^{\kappa-1} \left(\inf_{[-1,1]} h_\kappa \right)^2 \\
&\geqslant c_0 K^\kappa (c_1 K^{-\kappa} - 2b^{-1}) c_2^2 K^{\kappa-1} \left(\inf_{[-1,1]} h_\kappa \right)^2 .
\end{aligned}$$

Taking $b \geqslant 4K^\kappa / c_1$ gives the desired inequality.

K.3 Proof of Lemma 14

By definition, f_0 is the density of $\mathbf{X}$ when for all $1 \leqslant j \leqslant d$, $s_j = \zeta_0 = h_\kappa * u_b$, and f_n is the density of $\mathbf{X}$ when S_1 has density $s_1 = \zeta_n$ and $S_2, \ldots, S_d$ have density $s_j = \zeta_0$. The derivative of $\mathcal{F}[f_0]$ is

$$\begin{aligned}
\partial^i \mathcal{F}[f_0] = \sum_{\substack{(j_1, \ldots, j_{d_1+1}) \in \mathbb{N}^{d_1+1} : \|j\|_1 = i_{d_1+1} \\ (k_1, k_{d_1+1}, \ldots, k_d) \in \mathbb{N}^{d_2+1} : \|k\|_1 = i_1}} & a^{i_{d_1+1} - j_{d_1+1}} a^{i_1 - k_1} \mathcal{F}[s_1]^{(k_1 + j_1)} \mathcal{F}[s_{d_1+1}]^{(j_{d_1+1} + k_{d_1+1})} \\
& \times \prod_{u=2}^{d_1} \mathcal{F}[s_u]^{(i_u + j_u)} \prod_{u=d_1+2}^{d} \mathcal{F}[s_u]^{(i_u + k_u)} ,
\end{aligned}$$

where the vector j corresponds to how $\partial_{d_1+1}^{i_{d_1+1}}$ is split among the $\mathcal{F}[s_u]$, $1 \leqslant u \leqslant d_1+1$, and k corresponds to how $\partial_1^{i_1}$ is split among the $\mathcal{F}[s_u]$, $u \in \{1, d_1+1, \dots, d\}$, so that

$$|\partial^i \mathcal{F}[f_0]| \leqslant T^{\|i\|_1} \sum_{\substack{(j_1,\dots,j_{d_1+1})\in\mathbb{N}^{d_1+1}:\|j\|_1=i_{d_1+1}\\ (k_1,k_{d_1+1},\dots,k_d)\in\mathbb{N}^{d_2+1}:\|k\|_1=i_1}} a^{i_{d_1+1}-j_{d_1+1}} a^{i_1-k_1} \frac{(k_1+j_1)!}{\|k_1+j_1\|_1^{\kappa\|k_1+j_1\|_1}} \frac{(j_{d_1+1}+k_{d_1+1})!}{\|j_{d_1+1}+k_{d_1+1}\|_1^{\kappa\|j_{d_1+1}+k_{d_1+1}\|_1}}$$
$$\times \prod_{u=2}^{d_1} \frac{(i_u+j_u)!}{\|i_u+j_u\|_1^{\kappa\|i_u+j_u\|_1}} \prod_{u=d_1+2}^{d} \frac{(i_u+k_u)!}{\|i_u+k_u\|_1^{\kappa\|i_u+k_u\|_1}}\,.$$

Using $(k/\mathrm{e})^k \leqslant k! \leqslant c(k/\mathrm{e})^k\sqrt{k}$ for some numerical constant c (for instance 5) and $(\prod_{a=1}^d i_a^{i_a})^{-1} \leqslant (\|i\|_1/d)^{-\|i\|_1}$ by convexity of $x \mapsto x\log x$,

$$\frac{|\partial^i \mathcal{F}[f_0]|}{\prod_{a=1}^d i_a!} \leqslant \left(\frac{T\mathrm{e}d}{\|i\|_1}\right)^{\|i\|_1} c^d \mathrm{e}^{-\|i\|_1}\|i\|_1^{d/2} \sum_{\substack{(j_1,\dots,j_{d_1+1})\in\mathbb{N}^{d_1+1}:\|j\|_1=i_{d_1+1}\\ (k_1,k_{d_1+1},\dots,k_d)\in\mathbb{N}^{d_2+1}:\|k\|_1=i_1}} a^{i_{d_1+1}-j_{d_1+1}} a^{i_1-k_1}\|k_1+j_1\|_1^{(1-\kappa)\|k_1+j_1\|_1}$$
$$\times \|j_{d_1+1}+k_{d_1+1}\|_1^{(1-\kappa)\|j_{d_1+1}+k_{d_1+1}\|_1}$$
$$\times \prod_{u=2}^{d_1} \|i_u+j_u\|_1^{(1-\kappa)\|i_u+j_u\|_1} \prod_{u=d_1+2}^{d} \|i_u+k_u\|_1^{(1-\kappa)\|i_u+k_u\|_1}\,,$$
$$\leqslant \left(\frac{Td}{\|i\|_1}\right)^{\|i\|_1} c^d \|i\|_1^{d/2} \sum_{\substack{(j_1,\dots,j_{d_1+1})\in\mathbb{N}^{d_1+1}:\|j\|_1=i_{d_1+1}\\ (k_1,k_{d_1+1},\dots,k_d)\in\mathbb{N}^{d_2+1}:\|k\|_1=i_1}} a^{i_{d_1+1}-j_{d_1+1}} a^{i_1-k_1} \frac{\|i\|_1^{\|i\|_1}}{(\|i\|_1/d)^{\kappa\|i\|_1}}\,,$$
$$\leqslant \frac{(Td^{1+\kappa})^{\|i\|_1}}{\|i\|_1^{\kappa\|i\|_1}} \frac{c^d}{(1-a)^2}\|i\|_1^{d/2}\,,$$
$$\leqslant \frac{(c'T)^{\|i\|_1}}{\|i\|_1^{\kappa\|i\|_1}}\,,$$

for some c' for all $i \neq 0$, which concludes the proof.

K.4 Proof of Lemma 15

Following [Meister, 2007], since without loss of generality $b_n \geqslant 1$ and by integration by part the quantity $c_{u,\beta} = \sup_{t\in\mathbb{R}} |\mathcal{F}[u](t)|^2(1+t^2)^\beta$ is finite,

$$\int |\mathcal{F}[\alpha_n(P_{K_n}h_\kappa^2) * u_{b_n}](t)|^2(1+t^2)^\beta \mathrm{d}t = \alpha_n^2 \int |\mathcal{F}[P_{K_n}h_\kappa^2](t)|^2 |\mathcal{F}[u_{b_n}](t)|^2 (1+t^2)^\beta \mathrm{d}t\,,$$
$$= \alpha_n^2 \int |\mathcal{F}[P_{K_n}h_\kappa^2](t)|^2 \left|\mathcal{F}[u]\left(\frac{t}{b_n}\right)\right|^2 (1+t^2)^\beta \mathrm{d}t\,,$$
$$\leqslant c_{u,\beta}\alpha_n^2 b_n^{2\beta} \int |\mathcal{F}[P_{K_n}h_\kappa^2](t)|^2 \mathrm{d}t\,,$$
$$\leqslant c_{u,\beta}\alpha_n^2 b_n^{2\beta} \|P_{K_n}h_\kappa^2\|_{\mathbf{L}^2(\mathbb{R})}^2$$

by Cauchy-Schwarz's inequality and using $\mathcal{F}[u_b](t) = \mathcal{F}[u](t/b)$. Thus, H3 holds for $\mathcal{F}[\zeta_n]$ if

$$\alpha_n^2 \|P_{K_n} h_\kappa^2\|^2_{\mathbf{L}^2(\mathbb{R})} = O(b_n^{-2\beta})\,.$$

K.5 Proof of Lemma 16

For any probability density m_0 on $\mathbb{R}$, by the Cauchy-Schwarz inequality,

$$\int_{\mathbb{R}^d} |(f_0 * Q)(x) - (f_n * Q)(x)|\,\mathrm{d}x \leqslant \left(\int_{\mathbb{R}^d} |((f_0 - f_n) * Q)(x)|^2 \prod_{i=1}^d m_0^{-1}(x_i)\mathrm{d}x \right)^{1/2} .$$

Choosing $m_0 : x \mapsto (\pi(1+x^2))^{-1}$, yields

$$\int_{\mathbb{R}^d} |(f_0 * Q)(x) - (f_n * Q)(x)|\,\mathrm{d}x \leqslant \pi^{d/2} \left(\int_{\mathbb{R}^d} |((f_0 - f_n) * Q)(x)|^2 \prod_{i=1}^d (1 + x_i^2)\mathrm{d}x \right)^{1/2} .$$

Note that for all $x \in \mathbb{R}^d$,

$$\mathcal{F}[f_0](x) = \frac{1}{\mathrm{Det}(A)} \int_{\mathbb{R}^d} \prod_{j=1}^d \zeta_0\left((A^{-1}t)_j\right) \mathrm{e}^{it^\top x}\mathrm{d}t = \int_{\mathbb{R}^d} \prod_{j=1}^d \zeta_0\left(t_j\right) \mathrm{e}^{it^\top A^\top x}\mathrm{d}t = \prod_{j=1}^d \mathcal{F}[\zeta_0]((A^\top x)_j)\,,$$

$$\mathcal{F}[f_n](x) = \mathcal{F}[\zeta_n]((A^\top x)_1) \prod_{j=2}^d \mathcal{F}[\zeta_0]((A^\top x)_j)\,.$$

By Parseval's identity, for all $\eta \in \mathbb{N}^d$,

$$\int_{\mathbb{R}^d} |((f_0 - f_n) * Q)(x)|^2 \prod_{j=1}^d x_j^{2\eta_j}\mathrm{d}x = \int_{\mathbb{R}^d} \left| \left(\prod_{j=1}^d \partial_{t_j}^{\eta_j} \right) ((\mathcal{F}[f_0] - \mathcal{F}[f_n])\mathcal{F}[Q])(t) \right|^2 \mathrm{d}t\,.$$

Let $A_c = \{A^\top x : x \in [-c,c]^d\} \subset [-(1+a)c, (1+a)c]^d$. Since $\mathcal{F}[g]$ and $\mathcal{F}[g]'$ are supported on $[-c,c]$, using the change of variables $v = A^\top t$, for all $\eta \in \{0,1\}^d$,

$$\begin{aligned}
&\int_{\mathbb{R}^d} |((f_0 - f_n) * Q)(x)|^2 \prod_{j=1}^d x_j^{2\eta_j}\mathrm{d}x \\
&\quad \leqslant c_d \sum_{0 \leqslant \eta' \leqslant \eta} \int_{[-c,c]^d} \left| \left(\prod_{j=1}^d \partial_{t_j}^{\eta'_j} \right) (\mathcal{F}[f_0] - \mathcal{F}[f_n])(t) \right|^2 \mathrm{d}t\,, \\
&\quad \leqslant c_d \sum_{0 \leqslant \eta' \leqslant \eta} \int_{[-c,c]^d} \left| \left(\prod_{j=1}^d \partial_{t_j}^{\eta'_j} \right) \left(t \mapsto (\mathcal{F}[\zeta_0] - \mathcal{F}[\zeta_n])((A^\top t)_1) \prod_{j=2}^d \mathcal{F}[\zeta_0]((A^\top t)_j) \right)(t) \right|^2 \mathrm{d}t\,, \\
&\quad \leqslant c'_d \sum_{0 \leqslant \eta' \leqslant \eta} \int_{A_c} |(\mathcal{F}[\zeta_0] - \mathcal{F}[\zeta_n])^{(\eta'_1)}(v_1)|^2 \prod_{j=2}^d |\mathcal{F}[\zeta_0]^{(\eta'_j)}(v_j)|^2 \mathrm{d}v\,,
\end{aligned}$$

for some constants c_d and c'_d, so that for some constant c''_d,

$$\|(f_0 * Q) - (f_n * Q)\|_{\mathbf{L}^1(\mathbb{R}^d)} \leqslant c''_d \left(\int_{-(1+a)c}^{(1+a)c} |\mathcal{F}[\zeta_0] - \mathcal{F}[\zeta_n]|(t)^2 \mathrm{d}t + \int_{-(1+a)c}^{(1+a)c} |(\mathcal{F}[\zeta_0] - \mathcal{F}[\zeta_n])'|(t)^2 \mathrm{d}t \right)^{1/2} .$$

Using that for all $t \in \mathbb{R}$, $\mathcal{F}[\zeta_0](t) - \mathcal{F}[\zeta_n](t) = \alpha_n \mathcal{F}[P_{K_n} h_\kappa^2](t) \mathcal{F}[u_{b_n}](t)$,

$$\begin{aligned} \int_{\mathbb{R}^d} |(f_0 * Q)(x) - (f_n * Q)(x)| \mathrm{d}x \leqslant c''_d \alpha_n \Bigg(& \int_{-c}^{c} |\mathcal{F}[P_{K_n} h_\kappa^2](t)|^2 \left| \mathcal{F}[u] \left(\frac{t}{b_n} \right) \right|^2 \mathrm{d}t \\ & + b_n^{-2} \int_{-c}^{c} |\mathcal{F}[P_{K_n} h_\kappa^2](t)|^2 \left| \mathcal{F}[u]' \left(\frac{t}{b_n} \right) \right|^2 \mathrm{d}t \\ & + \int_{-c}^{c} |\mathcal{F}[P_{K_n} h_\kappa^2]'(t)|^2 \left| \mathcal{F}[u] \left(\frac{t}{b_n} \right) \right|^2 \mathrm{d}t \Bigg)^{1/2} , \\ \leqslant c'''_d \alpha_n \left(\int_{-c}^{c} |\mathcal{F}[P_{K_n} h_\kappa^2](t)|^2 \mathrm{d}t + \int_{-c}^{c} |\mathcal{F}[P_{K_n} h_\kappa^2]'(t)|^2 \mathrm{d}t \right)^{1/2} & \end{aligned}$$

for some constant c'''_d. Then,

$$\mathcal{F}[P_{K_n} h_\kappa^2](t) = \int_{\mathbb{R}} P_{K_n}(x) h_\kappa^2(x) \sum_{j \geqslant 0} \frac{(ixt)^j}{j!} \mathrm{d}x = \sum_{j \geqslant K_n} \frac{(it)^j}{j!} \int_{\mathbb{R}} P_{K_n}(x) h_\kappa^2(x) x^j \mathrm{d}x \,,$$

since by definition $P_{K_n} h_\kappa^2$ is orthogonal to $x \mapsto x^j$ in $\mathbf{L}^2(\mathbb{R}^d)$ when $j \in \mathbb{N}$ and $j < K_n$. By Conjecture 12, there exists a nonnegative envelope function F_{env}, a constant c and a parameter $\alpha_\kappa \geqslant 0$ such that $|F_{\text{env}}(x)| \leqslant c(1 + |x|^{\alpha_\kappa})$ and such that the family $(P_K)_{K \geqslant 1}$ satisfies $\sup_{K \geqslant 1} K^{(1-\kappa)/2} \|P_K h_\kappa / F_{\text{env}}\|_\infty < \infty$. Then, there exists a constant c which depends on κ,

$$\begin{aligned} |\mathcal{F}[P_{K_n} h_\kappa^2](t)| &\leqslant \sup_K \|P_K h_\kappa / F_{\text{env}}\|_\infty \sum_{j \geqslant K_n} \frac{t^j}{j!} \int_{\mathbb{R}} h_\kappa(x) |F_{\text{env}}(x)| |x|^j \mathrm{d}x \\ &\leqslant c \sup_K \|P_K h_\kappa / F_{\text{env}}\|_\infty \sum_{j \geqslant K_n} \frac{t^j}{j!} \int_{\mathbb{R}_+} (x^j + x^{j+\alpha_\kappa}) e^{-x^{1/(1-\kappa)}} \mathrm{d}x \end{aligned}$$

and

$$|\mathcal{F}[P_{K_n} h_\kappa^2]'(t)| \leqslant c \sup_K \|P_K h_\kappa / F_{\text{env}}\|_\infty \sum_{j \geqslant K_n - 1} \frac{t^j}{j!} \int_{\mathbb{R}_+} (x^{j+1} + x^{j+\alpha_\kappa+1}) e^{-x^{1/(1-\kappa)}} \mathrm{d}x \,.$$

For all $j \in \mathbb{R}_+$, write $M_j = \int_{\mathbb{R}_+} x^j e^{-x^{1/(1-\kappa)}} \mathrm{d}x$. By integration by part with $u'(x) = \frac{1}{1-\kappa} x^{\frac{1}{1-\kappa}-1} e^{-x^{1/(1-\kappa)}}$ and $v(x) = (1-\kappa) x^{j+1-\frac{1}{1-\kappa}}$ and thus $u(x) = -e^{x^{1/(1-\kappa)}}$ and $v'(x) = (1-\kappa)(j+1-1/(1-\kappa)) x^{j-\frac{1}{1-\kappa}}$, for all $j > \frac{1}{1-\kappa} - 1$,

$$M_j = (1-\kappa) \left(j + 1 - \frac{1}{1-\kappa} \right) M_{j - \frac{1}{1-\kappa}} \,.$$

In particular, for all $j \geqslant 1$,

$$M_j \leqslant (1-\kappa)^{(1-\kappa)j-1} j^{(1-\kappa)j} \sup_{j' \in [0, 1/(1-\kappa))} M_{j'} \,.$$

Note that

$$\begin{aligned}\frac{(j+\alpha_\kappa)^{j+\alpha_\kappa}}{j^j} &\sim \frac{(j+\alpha_\kappa)!}{j!} \frac{e^{j+\alpha_\kappa}\sqrt{j+\alpha_\kappa}}{e^j \sqrt{j}} \\ &= O((j+\alpha_\kappa)^{\alpha_\kappa} e^{\alpha_\kappa} \sqrt{1+\alpha_\kappa/j}) \\ &= O((j+\alpha_\kappa)^{\alpha_\kappa}).\end{aligned}$$

Therefore, there exists a constant c such that for all $t \in \mathbb{R}$,

$$|\mathcal{F}[P_{K_n} h_\kappa^2](t)| \leqslant c \sum_{j \geqslant K_n} \frac{t^j}{j!} (M_j + M_{j+\alpha_\kappa}) \,,$$

and a similar upper bound for $|\mathcal{F}[P_{K_n} h_\kappa^2]'(t)|$. Note that for all $\alpha > 0$, there exists a constant c such that for all $t \in \mathbb{R}$, when $K_n \geqslant \alpha$,

$$\sum_{j \geqslant K_n} \frac{t^j}{j!} M_{j+\alpha} \leqslant c \sum_{j \geqslant K_n} \frac{(\mathrm{te}(1-\kappa)^{(1-\kappa)})^j}{j^j} j^{-1/2} (j+\alpha)^{(1-\kappa)\alpha} j^{(1-\kappa)j} \leqslant c \sum_{j \geqslant K_n} \frac{(\mathrm{te}(1-\kappa)^{(1-\kappa)})^j}{j^{\kappa j-(1-\kappa)\alpha}}$$

Therefore, there exists constants c and C such that

$$\int_{\mathbb{R}^d} |(f_0 * Q)(x) - (f_n * Q)(x)| \mathrm{d}x \leqslant c\alpha_n \left(\frac{C}{K_n}\right)^{\kappa K_n} \,.$$

Thus, equation (21) holds if K_n is chosen, for some large enough constant C', as

$$K_n = \frac{C'}{\kappa} \left(\frac{\log n}{\log \log n}\right) \,.$$

L Numerical illustration of Conjecture 12

In this section, we propose some numerical illustrations to support Conjecture 12. First note that the case $\kappa = 1$ is true as it boils down to the results established in [Meister, 2007] on Legendre Polynomials. The case $\kappa = 1/2$ is also strongly supported by properties of Hermite functions, see [Boyd, 2018, Section A.11].

The orthonormal polynomials used in Conjecture 12 were approximately computed using the Python package OrthoPoly[1] which allows to generate orthogonal polynomials with respect to any probability density functions. The Python code used in this numerical section is available online[2]. Figure L.1 displays the functions $x \mapsto K^{(1-\kappa)/2}(P_K h_\kappa)(K^{1-\kappa}x)$ for degrees $1 \leqslant K \leqslant 16$ and for $\kappa \in \{0.55, 0.6, 0.7, 0.8, 0.9, 0.95\}$. We chose to limit our simulations to $K \leqslant 16$ as for degrees larger than 18 the simulations faced some numerical instability to compute $P_K h_\kappa$. This figure illustrates Equation (17), i.e. the fact that there exists a function F_{shape} such that

$$\sup_{K \geqslant 1} \left\| x \mapsto \frac{(P_K h_\kappa)(x)}{K^{(\kappa-1)/2} F_{\text{shape}}(K^{\kappa-1}x)} \right\|_\infty < \infty \,.$$

[1] https://github.com/j-jith/orthopoly

[2] https://sylvainlc.github.io/project/algorithms/

Then, Figure L.2 illustrates the second part of the conjecture by displaying $x \mapsto K^{(1-\kappa)/2} P_K h_\kappa(K^{-\kappa} x)$ for the same values of κ and K as in Figure L.1.

Figure L.1: Graphical representation of $x \mapsto K^{(1-\kappa)/2} P_K h_\kappa(K^{1-\kappa}x)$ for several values of κ and K.

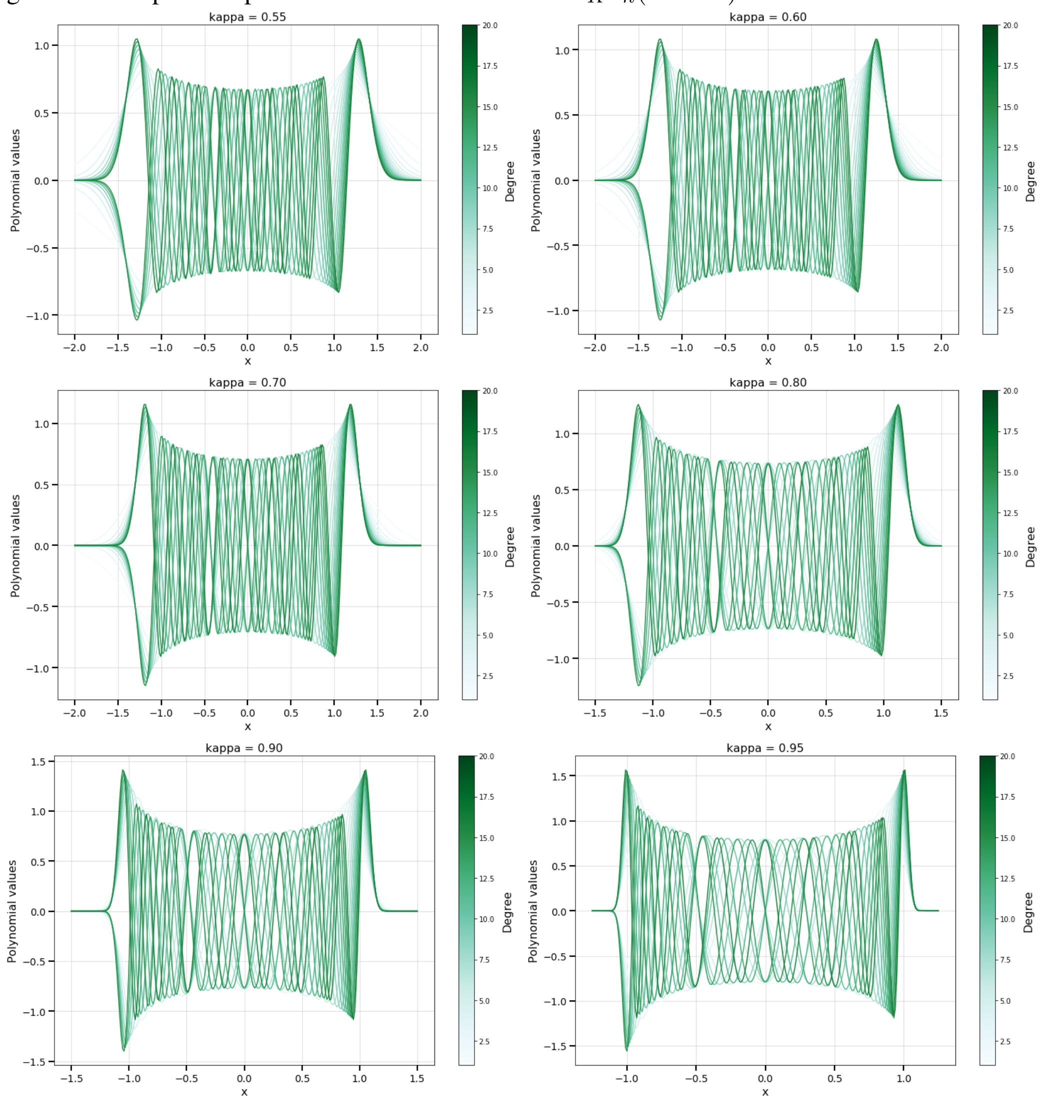
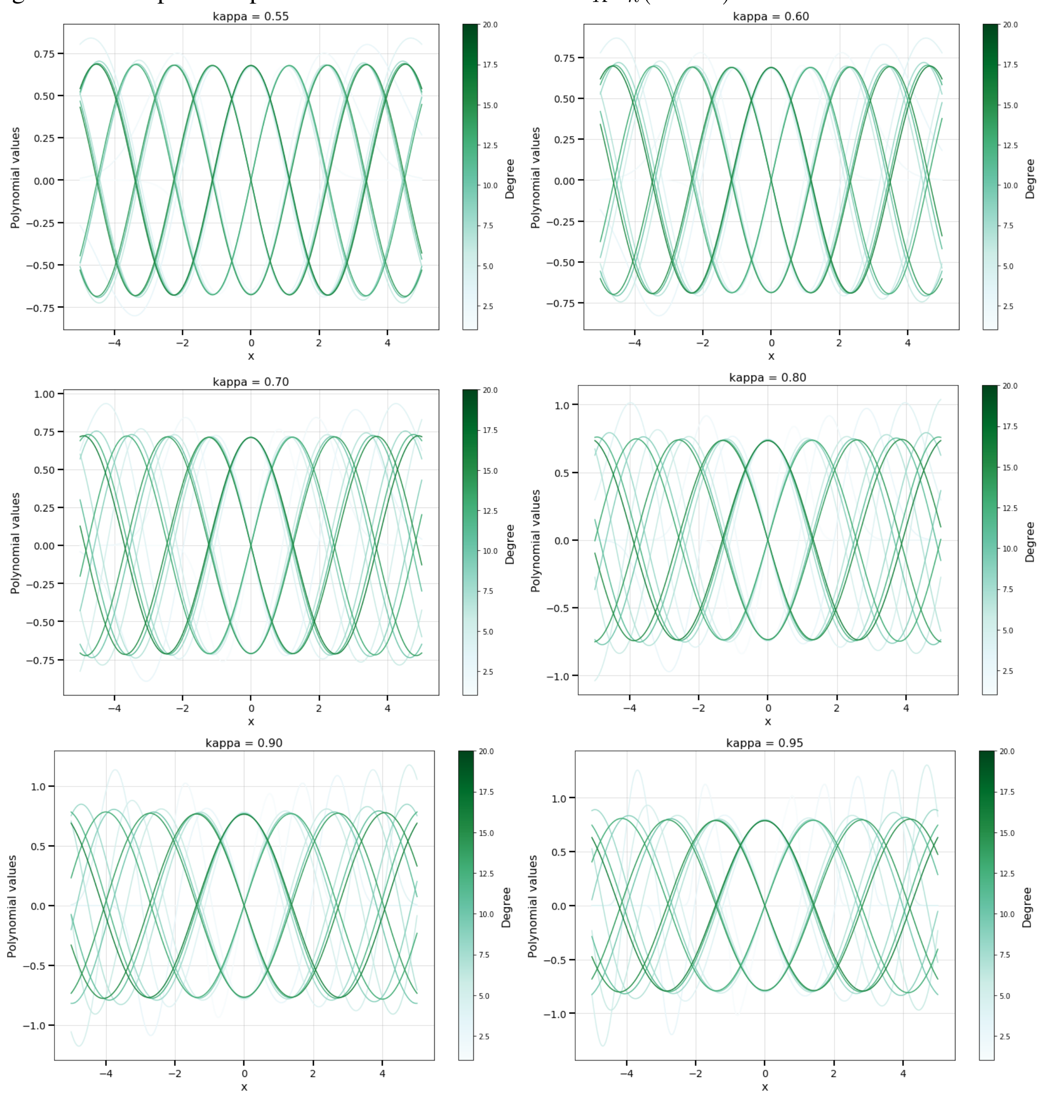


Figure L.2: Graphical representation of $x \mapsto K^{(1-\kappa)/2} P_K h_\kappa(K^{-\kappa} x)$ for several values of κ and K.



References

[Attias and Schreiner, 1998] Attias, H. and Schreiner, C. E. (1998). Blind source separation and deconvolution: the dynamic component analysis algorithm. *Neural computation*, 10(6):1373–1424.

[Batson and Royer, 2019] Batson, J. and Royer, L. (2019). Noise2self: Blind denoising by self-supervision. *Proceedings of the 36th International Conference on Machine Learning (ICML)*.

[Baudry et al., 2012] Baudry, J.-P., Maugis, C., and Michel, B. (2012). Slope heuristics: overview and implementation. *Statistics and Computing*, 22(2):455–470.

[Belomestny and Goldenshluger, 2019] Belomestny, D. and Goldenshluger, A. (2019). Density deconvolution under general assumptions on the distribution of measurement errors. *arXiv:1907.11024*.

[Bertin et al., 2016] 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*, volume 52, pages 939–980. Institut Henri Poincaré.

[Boyd, 2018] Boyd, J. P. (2018). *Dynamics of the equatorial ocean*. Springer.

[Butucea and Tsybakov, 2008a] Butucea, C. and Tsybakov, B. (2008a). Sharp optimality in density deconvolution with dominating bias. i. *Theory of Probability and Its Applications*, 52(1):24–39.

[Butucea and Tsybakov, 2008b] Butucea, C. and Tsybakov, B. (2008b). Sharp optimality in density deconvolution with dominating bias. ii. *Theory of Probability and Its Applications*, 52(2):237–249.

[Campisi and Egiazarian, 2017] Campisi, P. and Egiazarian, K. (2017). *Blind image deconvolution: theory and applications*. CRC press.

[Capitao Miniconi, 2021] Capitao Miniconi, J. (2021). Reconstruction of smooth curves from noisy measurements. Preprint.

[Carroll and Hall, 1988] Carroll, R. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83(404):1184–1186.

[Chazal et al., 2011] Chazal, F., Cohen-Steiner, D., and Mérigot, Q. (2011). Geometric inference for probability measures. *Journal on Foundations of Computational Mathematics*, 11(6).

[Chazal et al., 2017] 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(1):5845–5884.

[Chazal and Michel, 2017] Chazal, F. and Michel, B. (2017). An introduction to topological data analysis: fundamental and practical aspects for data scientists. *arXiv preprint arXiv:1710.04019*.

[Comon, 1994] Comon, P. (1994). Independent component analysis: a new concept? *Signal Processing*, 36:287–314.

[Comte and Lacour, 2013] Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. In *Annales de l'IHP Probabilités et statistiques*, volume 49, pages 569–609.

[Delaigle et al., 2008] Delaigle, A., Hall, P., and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.*, 36(2):665–685.

[Devroye, 1989] Devroye, L. (1989). Consistent deconvolution in density estimation. *Canad. J. Statist.*, 17(2):235–239.

[Eckle et al., 2016] Eckle, K., Bissantz, N., and Dette, H. (2016). Multiscale inference for multivariate deconvolution. *arXiv:1611.05201*.

[Ehm et al., 2004] Ehm, W., Gneiting, T., and Richards, D. (2004). Convolution roots of radial positive definite functions with compact support. *Transactions of the American Mathematical Society*, 356(11).

[Eriksson and Koivunen, 2004] Eriksson, J. and Koivunen, V. (2004). Identifiability, separability, uniqueness of linear ICA models. *IEEE Signal Processing Letters*, 11:601–604.

[Fan, 1991] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272.

[Gassiat et al., 2020a] Gassiat, E., Le Corff, S., and Lehéricy, L. (2020a). Identifiability and consistent estimation of nonparametric translation hidden markov models with general state space. *Journal of Machine Learning Research*, 21(115):1–40.

[Gassiat et al., 2020b] Gassiat, E., Le Corff, S., and Lehéricy, L. (2020b). Deconvolution with unknown noise distribution is possible for multivariate signals. *arXiv:2006.14226*.

[Gassiat and Rousseau, 2016] Gassiat, E. and Rousseau, J. (2016). Nonparametric finite translation hidden Markov models and extensions. *Bernoulli*, 22(1):193–212.

[Goldenshluger and Lepski, 2008] Goldenshluger, A. and Lepski, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14(4):1150–1190.

[Goldenshluger and Lepski, 2013] Goldenshluger, A. and Lepski, O. (2013). General selection rule from a family of linear estimators. *Theory of Probability & Its Applications*, 57(2):209–226.

[Hyvarinen et al., 2002] Hyvarinen, A., Karhunen, J., and Oja, E. (2002). *Independent Component Analysis*. John Wiley & Sons.

[Johannes, 2009] Johannes, J. (2009). Deconvolution with unknown error distribution. *The Annals of Statistics*, 37:2301–2323.

[Jutten, 1991] Jutten, C. (1991). Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 2(4):1–10.

[Khemakhem et al., 2020] Khemakhem, I., Kingma, D., Pio Monti, R., and Hyvarinen, A. (2020). Variational autoencoders and nonlinear ica: A unifying framework. *ArXiv:1907.04809*.

[Kotlarski, 1967] Kotlarski, I. (1967). On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics*, 20(1):69–76.

[Krull et al., 2019] 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)*.

[Kundur and Hatzinakos, 1996] Kundur, D. and Hatzinakos, D. (1996). Blind image deconvolution. *IEEE signal processing magazine*, 13(3):43–64.

[Lacour and Comte, 2010] Lacour, C. and Comte, F. (2010). Pointwise deconvolution with unknown error distribution. *Comptes Rendus Mathematique de l'Academie des Sciences*, 348(5-6):323–326.

[Le Cam, 2012] Le Cam, L. (2012). *Asymptotic methods in statistical decision theory*. Springer Science & Business Media.

[Li and Vuong, 1998] Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivariate Anal.*, 65(2):139–165.

[Lin and Kulasekera, 2007] Lin, W. and Kulasekera, K. (2007). Identifiability of single-index models and additive-index models. *Biometrika*, 94(2):496–501.

[Liu and Taylor, 1989] Liu, M. C. and Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.*, 17(4):427–438.

[Massart, 2007] Massart, P. (2007). *Concentration Inequalities and Model Selection : Ecole d'Eté de Probabilités de Saint-Flour XXXIII - 2003*. Berlin ; Heidelberg (DEU) ; New York : Springer.

[Meister, 2004] Meister, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Canadian Journal of Statistics*, 32(4):439–449.

[Meister, 2007] Meister, A. (2007). Deconvolving compactly supported densities. *Mathematical Methods of Statistics*, 16(1):63–76.

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

[Moulines et al., 1997] 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*, volume 5, pages 3617–3620. IEEE.

[Ollion et al., 2021] Ollion, J., Ollion, C., Gassiat, E., Lehéricy, L., and Le Corff, S. (2021). Joint self-supervised blind denoising and noise estimation. *arXiv:2102.08023*.

[Pfister et al., 2019] 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.

[Sarkar et al., 2018] 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(521):401–416.

[Schennach and Hu, 2013] 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(501):177–186.

[Starck et al., 2002] Starck, J.-L., Pantin, E., and Murtagh, F. (2002). Deconvolution in astronomy: A review. *Publications of the Astronomical Society of the Pacific*, 114(800):1051.

[Stefanski and Carroll, 1990] Stefanski, L. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics*, 21(2):169–184.

[Stein and Shakarchi, 2003] Stein, E. and Shakarchi, R. (2003). *Complex Analysis*. Princeton University Press, Princeton.

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