arXiv:2304.09452v1 [math.ST] 19 Apr 2023

Support and distribution inference from noisy data

Jérémie Capitao-Miniconi*, Élisabeth Gassiat* and Luc Lehéricy**

* Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405, Orsay, France,

** Université Côte d'Azur, CNRS, LJAD, France.

Abstract

We consider noisy observations of a distribution with unknown support. In the deconvolution model, it has been proved recently [19] that, under very mild assumptions, it is possible to solve the deconvolution problem without knowing the noise distribution and with no sample of the noise. We first give general settings where the theory applies and provide classes of supports that can be recovered in this context. We then exhibit classes of distributions over which we prove adaptive minimax rates (up to a $\log\log$ factor) for the estimation of the support in Hausdorff distance. Moreover, for the class of distributions with compact support, we provide estimators of the unknown (in general singular) distribution and prove maximum rates in Wasserstein distance. We also prove an almost matching lower bound on the associated minimax risk.

1 Introduction

1.1 Context and aim

It is a common observation that high dimensional data has a low intrinsic dimension. The computational geometry point of view gave rise to a number of interesting algorithms (see [6] and references therein) for the reconstruction of a non linear shape from a point cloud, and in the statistical community, past years have seen increasing interest for manifold estimation. The case of non noisy data, that is when the observations are sampled on the unknown manifold, is by now relatively well understood. When the loss is measured using the Hausdorff distance, minimax rates for manifold estimation are known and have been proved recently. The rates depend on the intrinsic dimension of the manifold and differ when the manifold has a boundary or does not have a boundary, due to the particular way points accumulate near boundaries (see [1] for the most recent results, together with an overview of the subject and references).

When considering the estimation of a distribution with unknown non linear low dimensional support, one has to choose a loss function. The Wasserstein distance allows to compare distributions that can be mutually singular, and is thus useful to compare distributions having possibly different supports. Moreover, approximating an unknown probability distribution μ by a good estimator $\hat{\mu}$ with respect to the Wasserstein metric allows to infer the topology of the support of μ, see [10]. When using non noisy data, one can look at [16] and [27] for the most recent results and for an overview of the references. However, despite these fruitful developments, geometric inference from noisy data remains a theoretical and practical widely open problem.

In this paper, we are interested in the estimation of possibly low dimensional supports, and of distributions supported on such supports, when the observations are corrupted with *unknown* noise. We aim at giving a new contribution on the type of noise which can affect the data without preventing to build consistent estimators of the support and of the law of the noisy signal.

1.2 Previous works: estimation of the support with noisy data

Some of the geometric ideas that have been developed to handle non noisy data can be applied, or adapted, to handle noisy data and build estimators with controlled risk. These works generally consider a noise that is normal to the unknown manifold, in which case the amplitude of the noise has to be bounded by the reach of the manifold (the reach is some regularity parameter of a manifold, see [17] for a precise definition). The upper bound on the risk contains a term depending on the amplitude of noise. Thus, the upper bound on the estimation risk is meaningful only when the bound on the noise is small, and the estimator is consistent when the noise tends to 0 with the amount of data tending to infinity. See [2], [1], [14], [18], see also [28] in which the noise can be non orthogonal to the manifold. In [3], the noise is not normal to the manifold but the data is uniformly sampled on a tubular neighborhood of the unknown manifold, which allows to take advantage of the fact that the manifold lies in the middle of the observations. The magnitude of the noise also has to be upper bounded by the reach. When the noise is not assumed very small, results are known in the specific setting of clutter noise, see [21], that is the situation where a proportion of data is uniformly sampled from a known compact set, and the remaining data is noiseless. The authors propose a clever idea to remove noise by comparing the way the empirical data concentrate near any regular shape, and they find a consistent estimator with upper bounded risk.

When we accept to consider noise with known distribution, a popular model for noisy data is the deconvolution model, in which the low dimensional data are corrupted with independent additive noise. In such models, all estimation procedures are roughly based on the fact that it is possible to get an estimator of the characteristic function of the non noisy data by dividing an estimator of the characteristic function of the noisy data by that (known) of the noise. In the deconvolution setting, the authors of [21] consider data corrupted with Gaussian noise, and propose as estimator of the manifold an upper level set of an estimator of a kernel smoothing density of the unknown distribution. With the Hausdorff loss, the authors prove that their estimator achieves a maximum risk (over some class of distributions) upper bounded by $(\sqrt{\log n})^{-1+\delta}$ for any positive δ, and prove a lower bound of order $(\log n)^{-1+\delta}$ for the minimax risk. Taking an upper level set of an estimated density had been earlier proposed to estimate a support based on non noisy data in [11]. In the context of full dimensional convex support and with additive Gaussian noise, [7] proposes an estimation procedure using convexity ideas. The authors prove an upper bound of order $\log\log n/\sqrt{\log n}$ and a lower bound of order $(\log n)^{-2/\tau}$ for the minimax Hausdorff risk, for any $\tau \in (0,1)$. Earlier work with known noise and with full dimensional support is [25], where the author first builds an estimator of the unknown density using deconvolution ideas, then samples from this estimated density and takes a union of balls centered on the sampled points, such as in [15].

1.3 Previous works: estimation of the distribution with noisy data

The case of unknown but small (and orthogonal to the unknown manifold) noise is handled in [16], the author proposes a kernel estimator and proves that it is minimax. The rate depends on the upper bound of the noise. Non parametric Bayesian methods have been explored in [5] for observations on a tubular neighborhood of the unknown manifold, that is again for bounded noise.

In the deconvolution problem, with known Gaussian noise, the authors of [13] prove matching upper and lower bounds for the minimax risk of the estimation of the unknown distribution using the Wasserstein distance. Results for other known noises, but limited to one dimensional observations, can be found in [12].

1.4 Contribution and main results

In this work, we consider the deconvolution problem *with totally unknown noise*. It has been proved recently [19] that, under very mild assumptions, it is possible to solve the deconvolution problem without knowing the noise distribution and with no sample of the noise. In [19], the authors consider the density estimation problem. Here, we are faced with the more general situation where the underlying non noisy data may have a distribution with a lower dimensional support than the ambient space, thus having no density with respect to Lebesgue measure.

Our main contributions are as follows.

- We first give general settings where the identifiability theory of [19] applies. We exhibit simple geometric properties of a support so that, whatever the distribution on such an (unknown) support (provided it does not have too heavy tails), the deconvolution problem can be solved without any knowledge regarding the noise, see Theorem 2. We also prove that these geometric properties almost always hold, in some sense developed in Section 2.4.

- We then exhibit classes of distributions over which we prove adaptive minimax rates (up to a log log factor) for the estimation of the support in Hausdorff distance, see Theorem 4, Theorem 6 and Theorem 5. Specifically, the minimax risk for the Hausdorff distance is upper bounded by $(\log\log n)^L/(\log n)^\kappa$ for some L, where $\kappa \in (1/2, 1]$ is a parameter depending on the tail of the distribution of the signal ($\kappa = 1$ corresponds to compactly supported distributions, and $\kappa = 1/2$ to sub-Gaussian distributions), while the minimax risk is lower bounded by $1/(\log n)^\kappa$ if $\kappa \in (1/2, 1)$ and $1/(\log n)^{1-\delta}$ if $\kappa = 1$, δ being any (small) positive number. Adaptation is with respect to κ.

- We finally consider the estimation of the unknown (in general singular) distribution of the hidden non noisy data itself when it has a compact support. We prove almost matching upper and lower bounds of order $1/(\log n)$ for the estimation risk of the distribution in Wasserstein distance, see Theorem 7 and Theorem 8.

Although we exhibit estimators, let us insist on the fact that our goal is mainly theoretical. We do not pretend to propose easy to compute estimation procedures, but to give precise answers about minimax adaptive rates for support and distribution estimation with noisy data in a very general deconvolution setting, where the noise is unknown and can have any distribution.

1.5 Organisation of the paper

Section 2 is devoted to the identifiability question. We first recall in Section 2.2 the identifiability result proved in [19]. We then exhibit in Section 2.3 geometric conditions under which this identifiability result applies, and the genericity of such conditions is considered in Section 2.4.

We focus on support estimation in Section 3. We first refine an estimation result of the characteristic function of the signal in 3.1, which is the basic step of any of the estimation procedures we propose. In Section 3.2, we propose an estimator of the support as an upper-level set of an estimated density following ideas of [21], the main difference being with the smoothing kernel we choose. Indeed, with this kernel, no prior knowledge on the intrinsic dimension is needed to build the estimator. The upper bound on the risk depends on the tail of the distribution of the signal, and adaptive estimation using Lepski's method is detailed in Section 3.4. We prove in Section 3.3 an almost matching lower bound.

Section 4 is devoted to the estimation of the distribution when it is compactly supported. Lower bounds are proved using the usual two-points method. Here, the points for the lower bound in [21] and in [13], [12], can not be used because of our tail assumption on the signal. Detailed proofs are given in Section 6.

1.6 Notations

The Euclidean norm (in any dimension) will be denoted $\|\cdot\|_2$, and the operator norm of a linear operator will be denoted $\|\cdot\|_{op}$. If A is a subset of $\mathbb{R}^D$, we write $\mathrm{Diam}(A)$ its diameter $\sup\{\|x-y\|_2 \mid x, y \in A\}$, and for any $x \in \mathbb{R}^D$, $d(x, A) = \inf\{\|x-y\|_2 \mid y \in A\}$. For any $\eta > 0$, A_η will denote the η-offset of A, that is the set of all points x in $\mathbb{R}^D$ such that $d(x, A) \leqslant \eta$. For any dimension d, any $x \in \mathbb{R}^d$ and $r > 0$, $B(x, r)$ will denote the Euclidean open ball centered on x of radius r and $\bar{B}(x, r)$ the closure of $B(x, r)$ in $\mathbb{R}^d$. For $k, l \in \{1, \ldots, D\}$ with $k \leqslant l$, write $\pi^{(k:l)}$ the projection $\pi^{(k:l)} : (x_1, \ldots, x_D) \in \mathbb{R}^D \mapsto (x_k, \ldots, x_l) \in \mathbb{R}^{l-k+1}$ and $\pi^{(k)} = \pi^{(k:k)}$.

We shall denote $d_H(A_1, A_2)$ the Hausdorff distance between A_1 and A_2 subsets of $\mathbb{R}^D$. It is defined as

$$d_H(A_1, A_2) = \sup_{x \in A_1 \cup A_2} |d(x, A_1) - d(x, A_2)|.$$

For any $r > 0$, we write $B_r = (-r, r)$ and for any measurable function f on B_r^D, we write $\|f\|_{\infty,r}$ the essential supremum of f over B_r^D and

$$\|f\|_{2,r} = \left(\int_{B_r^D} |f(u)|^2 du \right)^{1/2}.$$

When f is an integrable function from $\mathbb{R}^D$ to $\mathbb{R}$, we denote by $\mathcal{F}[f]$ (resp. $\mathcal{F}^{-1}[f]$) the (resp. inverse) Fourier transform of f defined, for all $y \in \mathbb{R}^d$, by

$$\mathcal{F}[f](y) = \int e^{it^\top y} f(t)dt \;\text{ and }\; \mathcal{F}^{-1}[f](y) = (\frac{1}{2\pi})^D \int e^{-it^\top y} f(t)dt.$$

For any $p \in [1, +\infty)$ and any two probability measures μ and ν on $\mathbb{R}^D$, we write $W_p(\mu, \nu)$ the Wasserstein distance of order p between μ and ν, that is

$$W_p(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \left(\int_{\mathbb{R}^D \times \mathbb{R}^D} \|x - y\|_2^p d\pi(x, y) \right)^{1/p},$$

where $\Pi(\mu, \nu)$ is the set of probability measures on $\mathbb{R}^D \times \mathbb{R}^D$ that have marginals μ and ν.

2 The identifiability Theorem and general applications

In this section, we first recall the general identifiability Theorem proved in [19]. We then provide geometrical conditions on the support of the signal that suffice to obtain identifiability of the model (2), whatever the distribution of the signal may be. We also show that the conditions on the signal distribution of the identifiability theorem hold generically.

2.1 Setting

We consider independent and identically distributed observations Y_i, $i = 1, \ldots, n$ coming from the model

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

in which the signal X and the noise ε are independent random variables. We assume that the observation has dimension at least two, and that its coordinates can be partitioned in such a way that the corresponding blocks of noise variables are independently distributed, that is

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

in which $Y^{(1)}, X^{(1)}, \varepsilon^{(1)} \in \mathbb{R}^{d_1}$ and $Y^{(2)}, X^{(2)}, \varepsilon^{(2)} \in \mathbb{R}^{d_2}$, for $d_1, d_2 \geqslant 1$ with $d_1 + d_2 = D$, and we assume that the noise components $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are independent random variables. We write G the distribution of X and $\mathcal{M}_G$ its support. For $i \in \{1, 2\}$, we write $\mathbb{Q}^{(i)}$ the distribution of $\varepsilon^{(i)}$, so that $\mathbb{Q} = \mathbb{Q}^{(1)} \otimes \mathbb{Q}^{(2)}$ is the distribution of ε.

We shall not make any more assumption on the distribution of the noise ε, and we shall not assume that its distribution is known. Indeed in [19], it is proved that under very mild conditions on the distribution of the signal X, model (2) is fully identifiable, that is one can recover G, and thus its support, and $\mathbb{Q}$ from $G * \mathbb{Q}$.

2.2 Identifiability Theorem

Let us introduce the assumptions on the distribution of the signal we shall use. The first one is about the tail of G. Let ρ be a positive real number.

A(ρ) There exists $a, b > 0$ such that for all $\lambda \in \mathbb{R}^D$, $\mathbb{E}\left[\exp\left(\lambda^\top X\right)\right] \leqslant a \exp\left(b\|\lambda\|_2^\rho\right)$.

Proposition 1. • *A random variable X satisfies $A(1)$ if and only if its support is compact.*

- *A random variable X satisfies $A(\rho)$ for $\rho > 1$ if and only if there exists constants $c, d > 0$ such that for any $t \geqslant 0$,*

$$\mathbb{P}(\|X\| \geqslant t) \leqslant c \exp(-dt^{\rho/(\rho-1)}).$$

The proof of Proposition 1 is detailed in Section 6.1.

Under A(ρ), the characteristic function of the signal can be extended into the multivariate analytic function

$$\begin{array}{rcl} \Phi_X : \mathbb{C}^{d_1} \times \mathbb{C}^{d_2} & \longrightarrow & \mathbb{C} \\ (z_1, z_2) & \longmapsto & \mathbb{E}\left[\exp\left(iz_1^\top X^{(1)} + iz_2^\top X^{(2)}\right)\right]. \end{array}$$

The second assumption is a mild dependence assumption (see the discussion after Theorem 2.1 in [19]).

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

Obviously, if no centering constraint is put on the signal or on the noise, it is possible to translate the signal by a fixed vector $m \in \mathbb{R}^D$ and the noise by $-m$ without changing the observation. The model can thus be identifiable only up to translation.

Theorem 1 (from [19])**.** *If the distribution of the signal satisfies $A(\rho)$ and (Adep), then the distribution of the signal and the distribution of the noise can be recovered from the distribution of the observations up to translation.*

The proof of this theorem is based on recovering Φ_X. The arguments show that knowing the characteristic function of the observations in a neighborhood of the origin allows to recover Φ_X in a neighborhood of the origin, and then over the whole multidimensional complex plane. Similarly, our estimators for the distribution of the signal or its support will start with the estimation of Φ_X, which is detailed in Section 3.1.

The end of the section is devoted to some geometric understanding of assumption (Adep). We first provide simple but useful properties.

Proposition 2. *The following holds.*

(i) Let U and V independent random variables satisfying $A(\rho)$. Then U and V satisfy (Adep) if and only if $U + V$ satisfies (Adep).

(ii) Let $U = \begin{pmatrix} U^{(1)} \\ U^{(2)} \end{pmatrix}$ be a random variable such that $U^{(1)} \in \mathbb{R}^{d_1}$ and $U^{(2)} \in \mathbb{R}^{d_2}$. Let $A \in GL_{d_1}(\mathbb{C})$, $B \in GL_{d_2}(\mathbb{C})$, $m_1 \in \mathbb{C}^{d_1}$ and $m_2 \in \mathbb{C}^{d_2}$. Define $V = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} U^{(1)} \\ U^{(2)} \end{pmatrix} + \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$. Then U satisfies $A(\rho)$ if and only if V satisfies $A(\rho)$. Moreover, U satisfies $A(\rho)$ and (Adep) if and only if V satisfies $A(\rho)$ and (Adep).

(iii) Let $U^{(1)}$ and $U^{(2)}$ be two independent random variables in $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$ respectively that satisfy $A(\rho)$ for some $\rho \geqslant 1$, then $U = \begin{pmatrix} U^{(1)} \\ U^{(2)} \end{pmatrix}$ satisfies (Adep) if and only if $U^{(1)}$ and $U^{(2)}$ are Gaussian or Dirac random variables.

The proof of Proposition 2 is detailed in section 6.2.

Point (i) of Proposition 2 makes it possible to transfer a proof of (Adep) for a support with full dimension D to a support with dimension $d < D$. Indeed, if U is a random variable with support of dimension $d < D$, by introducing an independent random variable V with support of full dimension D, proving that $U + V$ (whose support has full dimension) satisfies (Adep) ensures that U satisfies (Adep) as well. For instance, Theorem 2 below shows that a random variable having support the centered Euclidean ball with radius $\eta > 0$ satisfies A(1) and (Adep). Thus geometric conditions such as those proposed in Section 2.3 can be transposed from one dimension to another.

Point (ii) shows that the fact that A(ρ) and (Adep) hold is not modified by linear transformations of each component of the signal.

Finally, Point (iii) shows that to verify (Adep), outside of trivial cases, the two signal components cannot be independent. Even further, combined with Point (i), this shows that it is not possible to write the signal as the sum of two independent signals where one of them has independent components: such independent sub-signals with independent components must be part of the noise.

2.3 Sufficient geometrical conditions for (Adep) to hold

In [9], the authors prove that (Adep) holds for random variables supported on a sphere. In such a context, they prove that the radius of the sphere can be estimated at almost parametric rate. Here we give much more general conditions on the support of a random variable that are sufficient for (Adep) to hold.

We define the following assumptions (H1) and (H2).

(H1) For any $\Delta > 0$, there exists $A_\Delta \subset \mathbb{R}^{d_2}$ and $B_\Delta \subset \mathbb{R}^{d_1}$ such that $\mathbb{P}(X^{(2)} \in A_\Delta) > 0$, $\lim_{\Delta \to 0} \mathrm{Diam}(B_\Delta) = 0$ and $\mathbb{P}(X^{(1)} \in B_\Delta \,|\, X^{(2)} \in A_\Delta) = 1$.

(H2) For any $\Delta > 0$, there exists $A_\Delta \subset \mathbb{R}^{d_1}$ and $B_\Delta \subset \mathbb{R}^{d_2}$ such that $\mathbb{P}(X^{(1)} \in A_\Delta) > 0$, $\lim_{\Delta \to 0} \mathrm{Diam}(B_\Delta) = 0$ and $\mathbb{P}(X^{(2)} \in B_\Delta \,|\, X^{(1)} \in A_\Delta) = 1$.

It is showed in Theorem 2 that these assumptions are sufficient to ensure identifiability provided that A(ρ) is satisfied.

Theorem 2. *Assume that the distribution of X satisfies $A(\rho)$, (H1) and (H2). Then X satisfies $A(\rho)$ and (Adep).*

The proof of Theorem 2 is detailed in Section 6.3.

One can interpret the assumptions (H1) and (H2) geometrically as shown in Figure 1. In essence, it means that there exists a slice (along the first d_1, resp. last d_2, coordinates, with base A_Δ) such that the random variable belongs to this slice with positive probability and such that on this slice, the support of the distribution is contained in an orthogonal slice (along the last d_2, resp. first d_1, coordinates) of diameter smaller than Δ.

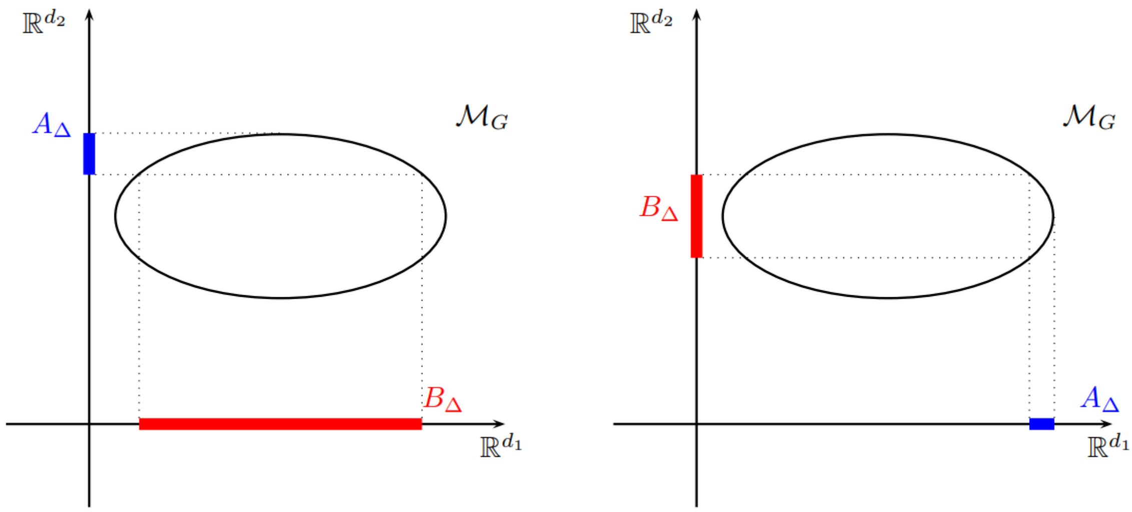

Figure 1: Left : Assumption (H1). Right : Assumption (H2).

A reformulation of (H1) and (H2) based on the support of the signal is as follows. Let

$$\mathcal{A}_1(\Delta,\varepsilon) = \{\mathcal{M} \subset \mathbb{R}^D \mid \text{There exists } x=(x_1,x_2)\in\mathcal{M} \text{ such that } \operatorname{Diam}\left(\pi^{(1:d_1)}\left[\mathcal{M}\cap(\mathbb{R}^{d_1}\times\bar{B}(x_2,\varepsilon))\right]\right) < \Delta\}$$

and

$$\mathcal{A}_2(\Delta,\varepsilon) = \{\mathcal{M} \subset \mathbb{R}^D \mid \text{There exists } x=(x_1,x_2)\in\mathcal{M} \text{ such that } \operatorname{Diam}\left(\pi^{(d_1+1:D)}\left[\mathcal{M}\cap(\bar{B}(x_1,\varepsilon)\times\mathbb{R}^{d_2})\right]\right) < \Delta\}.$$

The proof of the following proposition is straightforward.

Proposition 3. *Let $\mathcal{M} \in (\cap_{\Delta>0}\cup_{\varepsilon>0}\mathcal{A}_1(\Delta,\varepsilon)) \cap (\cap_{\Delta>0}\cup_{\varepsilon>0}\mathcal{A}_2(\Delta,\varepsilon))$. Then any random variable with support $\mathcal{M}$ satisfies (H1) and (H2).*

We now propose sets of compact subsets of $\mathbb{R}^D$ for which Proposition 3 holds. Define the sets $\mathcal{B}_1$ and $\mathcal{B}_2$ as

$$\mathcal{B}_1 = \{\mathcal{M}\subset\mathbb{R}^D \text{ compact} \mid \exists x_1\in\mathbb{R}^{d_1}, \operatorname{Card}((\{x_1\}\times\mathbb{R}^{d_2})\cap\mathcal{M})=1\},$$

and

$$\mathcal{B}_2 = \{\mathcal{M}\subset\mathbb{R}^D \text{ compact} \mid \exists x_2\in\mathbb{R}^{d_2}, \operatorname{Card}((\mathbb{R}^{d_1}\times\{x_2\})\cap\mathcal{M})=1\}.$$

Proposition 4. *Let $\mathcal{M}$ be a subset of $\mathbb{R}^D$ such that $\mathcal{M}\in\mathcal{B}_1\cap\mathcal{B}_2$. If X is a random variable with support $\mathcal{M}$, then X satisfies A(1), (H1) and (H2).*

The proof of Proposition 4 is detailed in Section 6.4.

For instance, any closed Euclidean ball, and more generally any strictly convex compact set in $\mathbb{R}^D$, is in $\mathcal{B}_1\cap\mathcal{B}_2$. To see this, consider the points of the set with maximal first (resp. last) coordinate: they are unique by strict convexity, which ensures that the set is in $\mathcal{B}_1$ (resp. $\mathcal{B}_2$). The same holds for the boundary of any strictly convex compact set.

2.4 Genericity

The main purpose of this subsection is to show that hypotheses (H1) and (H2) are verified generically.

First, we show that while the set of supports satisfying Proposition 3 is dense in the set of closed sets of $\mathbb{R}^D$, its complement is also dense.

Proposition 5. *The set $(\cap_{\Delta>0}\cup_{\varepsilon>0}\mathcal{A}_1(\Delta,\varepsilon)) \bigcap (\cap_{\Delta>0}\cup_{\varepsilon>0}\mathcal{A}_2(\Delta,\varepsilon))$ and its complement are dense in the set of closed subsets of $\mathbb{R}^D$ endowed with the Hausdorff distance.*

The proof of Proposition 5 is detailed in Section 6.5.

This result shows that any support $\mathcal{M}$ can be altered by a small perturbation to produce both supports that satisfy (H1) and (H2) and supports that satisfy neither. *A fortiori*, the same is true for (Adep), as on one hand (H1), (H2) and A(ρ) ensure (Adep) by Theorem 2 and on the other hand a small perturbation of the signal is enough to no longer satisfy (Adep) by Point (i) of Proposition 2.

Therefore, we need a stronger notion than topological density to assess the genericity of (H1) and (H2). Similarly to how "almost everywhere" (with respect to the Lebesgue measure) is a strong indication of genericity in $\mathbb{R}^D$, we construct a random and small perturbation of $\mathbb{R}^D$ such that any compact set is almost surely transformed into a compact set in $\mathcal{B}_1\cap\mathcal{B}_2$.

More precisely, for any $\varepsilon>0$, we define a (random) continuous bijection $f:\mathbb{R}^D\longrightarrow\mathbb{R}^D$ such that almost surely, $|f(x)-x|\leqslant\varepsilon$ for all $x\in\mathbb{R}^D$, and such that if $\mathcal{M}$ is compact, then $f(\mathcal{M})$ is in $\mathcal{B}_1\cap\mathcal{B}_2$ almost surely. This random bijection does not depend on which support

$\mathcal{M}$ is considered, and can for instance be seen as a modeling of the imperfections of "realistic" supports, or as a way to introduce a Bayesian prior on the support. In that sense, compact supports are almost surely in $\mathcal{B}_1 \cap \mathcal{B}_2$, and thus compactly supported random variables almost surely satisfy (Adep).

There is no canonical way to define a random perturbation of $\mathbb{R}^D$. Our approach is to tile the space with simplices, then add a small perturbation to each vertex of the tiling, keeping the transformation linear inside each simplex.

Simplicial tiling of $\mathbb{R}^D$. Let us recall a few definitions about simplicial complexes. For any $k \in \{0, \ldots, D\}$, a k-simplex of $\mathbb{R}^D$ is the convex hull of $(k+1)$ affinely independent points of $\mathbb{R}^D$. A simplicial complex $\mathcal{P}$ is a set of simplices such that every face of a simplex from $\mathcal{P}$ is also in $\mathcal{P}$, and the non-empty intersection of any two simplices $F_1, F_2 \in \mathcal{P}$ is a face of both F_1 and F_2. $\mathcal{P}$ is a homogeneous simplicial D-complex if each simplex of dimension less than D of $\mathcal{P}$ is the face of a D-simplex of of $\mathcal{P}$. For any simplex F, we write $\mathrm{relint}(F)$ its relative interior. Finally, a homogeneous simplicial D-complex $\mathcal{P}$ is called a simplicial tiling of $A \subset \mathbb{R}^D$ if the relative interior of its simplices form a partition of A. Note that the facets of $\mathcal{P}$, that is, its D-simplices, do not necessarily form a partition of A: two facets can have a non-empty intersection when they share a face.

First, consider a finite simplicial tiling of the hypercube $[0,1]^D$, and extend it to $\mathbb{R}^D$ by mirroring it along the hyperplanes orthogonal to the canonical axes crossing them at integer coordinates. Formally, for any $k = (k_1, \ldots, k_D) \in \mathbb{Z}^D$, the hypercube $\prod_{i=1}^D [k_i, k_i+1]$ contains the tiling of $[0,1]^D$, mirrored along axis i if and only if k_i is odd. The faces of the hypercubes defined in this way match, as each pair of hypercubes sharing a face are mirrors of each other with respect to that face. Thus, the resulting tiling $\mathcal{P}$ is a simplicial tiling of $\mathbb{R}^D$.

Let $(x_n)_{n\in\mathbb{N}}$ be the sequence of vertices of the simplicial tiling $\mathcal{P}$ (i.e. its 0-simplices). We identify each simplex $F \in \mathcal{P}$ with the set of its 0-dimensional faces $\{x_i\}_{i\in I}$, and write F_I in that case. Note that the set I is unique for any given simplex F and characterizes F.

Perturbation of the tiling. Fix a small $r > 0$. Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. uniform variables on $[-r,r]^D$, and define $\mathcal{P}^\varepsilon$ the simplicial complex defined by

$$\mathcal{P}^\varepsilon = \{\{x_i + \varepsilon_i\}_{i\in I} : \{x_i\}_{i\in I} \in \mathcal{P}\}.$$

Note that since the original tiling of $[0,1]^D$ was finite, there exists $r_0 > 0$ such that for any $(\varepsilon_n)_{n\in\mathbb{N}} \in ([-r_0, r_0]^D)^{\mathbb{N}}$, the vertices of any simplex in $\mathcal{P}$ are still affinely independent after being moved according to ε and any two simplices $F, F' \in \mathcal{P}$ sharing a face F'' (resp. with no intersection) are transformed into two simplices of $\mathcal{P}^\varepsilon$ that share exactly the transformation of F'' (resp. with no intersection), so that $\mathcal{P}^\varepsilon$ is indeed a simplicial complex. Finally, $\mathcal{P}^\varepsilon$ still covers $\mathbb{R}^D$ (as seen when moving each vertex in $[-1,2]^D$ one after the other along a continuous path, showing that no hole is created in the covering of $[0,1]^D$ at any point in time), so for any $r \in (0, r_0]$, $\mathcal{P}^\varepsilon$ is almost surely a simplicial tiling of $\mathbb{R}^D$.

Since the relative interiors of the simplices of $\mathcal{P}$ define a partition of $\mathbb{R}^D$, for each $z \in \mathbb{R}^D$, there exists exactly one face $F_I \in \mathcal{P}$ such that $z \in \mathrm{relint}(F_I)$. Writing $z = \sum_{i\in I} \alpha_i x_i$ (for $\alpha \in (0,1]^{|I|}$ such that $\sum_{i\in I} \alpha_i = 1$), we define the image of z by the perturbation as $f^\varepsilon(z) = \sum_{i\in I} \alpha_i (x_i + \varepsilon_i)$. In other words, each simplex is deformed according to the linear transformation given by the perturbation of its vertices.

The mapping f^ε is a (random) bijective and continuous transformation of $\mathbb{R}^D$ that is "small", in the sense that almost surely, $\sup_{z\in\mathbb{R}^D} \|z - f^\varepsilon(z)\| \leqslant r$.

Note that the transformation f^ε can be made with arbitrarily small granularity: the same approach works when considering tilings of $[0,\delta]^D$ for any $\delta > 0$ instead of $[0,1]^D$ (up to changing r). We may also iterate several random independent transformations $f^{\varepsilon^{(1)}} \circ \cdots \circ f^{\varepsilon^{(m)}}$ for $m \geqslant 1$, and the transformation of $\mathcal{M}$ will still almost surely belong to $\mathcal{B}_1 \cap \mathcal{B}_2$.

Theorem 3. *Let $r \in (0, r_0]$ with r_0 as above, $\varepsilon = (\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. uniform r.v. on $[-r,r]^D$, $\delta > 0$, and f^ε be the bijective transformation of $\mathbb{R}^D$ defined above.*

Then for any (random) continuous mapping $G : \mathbb{R}^D \to \mathbb{R}^D$ that is independent of ε, the mapping $F : z \longmapsto \delta f^\varepsilon(\frac{G(z)}{\delta})$ satisfies: for any compact set $\mathcal{M} \subset \mathbb{R}^D$, $F(\mathcal{M}) \in \mathcal{B}_1 \cap \mathcal{B}_2$ a.s..

The proof of Theorem 3 is detailed in Section 6.6.

This shows that for any compact set $\mathcal{M} \in \mathbb{R}^D$, a small change into the set $F(\mathcal{M})$ where F is a transformation of $\mathbb{R}^D$ of the type described in the Theorem almost surely results in a set in $\mathcal{B}_1 \cap \mathcal{B}_2$.

3 Estimation of the support

As explained after Theorem 1, the estimation of the characteristic function of the signal will be the first step to derive efficient estimators. In Section 3.1, we describe the estimator of the characteristic function used in all our procedures, and we give its properties. In Section 3.2, we provide an estimator of the support of the signal when ρ is known, and prove an upper bound for the maximum risk in Hausdorff distance. In Section 3.3, we prove a lower bound which shows that our estimator is minimax up to some power of $\log \log n$ for all $\rho \in (1, 2)$ and up to any small power of $\log n$ for $\rho = 1$. Section 3.4 is devoted to the construction of an adaptive estimator of the support for unknown ρ.

3.1 Estimation of the characteristic function

We shall need sets of multivariate analytic functions for which A(ρ) and (Adep) hold. For any $S > 0$, let $\Upsilon_{\rho,S}$ be the subset of multivariate analytic functions from $\mathbb{C}^D$ to $\mathbb{C}$ defined as follows.

$$\Upsilon_{\rho,S} = \left\{ \phi \text{ analytic s.t. } \forall z \in \mathbb{R}^D, \overline{\phi(z)} = \phi(-z), \phi(0) = 1 \text{ and } \forall i \in \mathbb{N}^D \setminus \{0\}, \left| \frac{\partial^i \phi(0)}{\prod_{a=1}^d i_a!} \right| \leqslant \frac{S^{\|i\|_1}}{\|i\|_1^{\|i\|_1/\rho}} \right\}$$

where $\|i\|_1 = \sum_{a=1}^D i_a$. If the distribution of X satisfies A(ρ), then there exists S such that $\Phi_X \in \Upsilon_{\rho,S}$, and the converse also holds, see Lemma 3.1 in [19].

Let $\Phi_{\varepsilon^{(i)}}$ be the characteristic function of $\varepsilon^{(i)}$, $i = 1, 2$, and define for all $\phi \in \Upsilon_{\rho,S}$ and any $\nu > 0$,

$$M(\phi; \nu|\Phi_X) = \int_{B_\nu^{d_1} \times B_\nu^{d_2}} |\phi(t_1, t_2)\Phi_X(t_1, 0)\Phi_X(0, t_2) - \Phi_X(t_1, t_2)\phi(t_1, 0)\phi(0, t_2)|^2 |\Phi_{\varepsilon^{(1)}}(t_1)\Phi_{\varepsilon^{(2)}}(t_2)|^2 dt_1 dt_2.$$

It follows from the proof of Theorem 1, see [19], that for any $\nu > 0$, if $\phi \in \Upsilon_{\rho,S}$ satisfies (Adep), then $M(\phi; \nu|\Phi_X) = 0$ if and only if $\phi = \Phi_X$ (up to translation). The estimator of the characteristic function of the signal can then be defined as a minimizer of an empirical estimator M_n of M. Fix some $\nu_{\text{est}} > 0$, and define M_n for any ϕ as follows

$$M_n(\phi) = \int_{B_{\nu_{\text{est}}}^{d_1} \times B_{\nu_{\text{est}}}^{d_2}} |\phi(t_1, t_2)\tilde{\phi}_n(t_1, 0)\tilde{\phi}_n(0, t_2) - \tilde{\phi}_n(t_1, t_2)\phi(t_1, 0)\phi(0, t_2)|^2 dt_1 dt_2,$$

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

$$\tilde{\phi}_n(t_1, t_2) = \frac{1}{n} \sum_{\ell=1}^n \exp\left\{ i t_1^\top Y_\ell^{(1)} + i t_2^\top Y_\ell^{(2)} \right\}.$$

Define now, for all $m \in \mathbb{N}$, the set $\mathbb{C}_m[X_1, \dots, X_D]$ of multivariate polynomials in D variables with total degree m and coefficients in $\mathbb{C}$, simply written $\mathbb{C}_m[X]$ in the following. If ϕ is an analytic function defined in a neighborhood of 0 in $\mathbb{C}^D$ written as $\phi : x \mapsto \sum_{(i_1,\dots,i_D) \in \mathbb{N}^D} c_i \prod_{a=1}^D x_a^{i_a}$, define its truncation on $\mathbb{C}_m[X]$ as

$$T_m \phi : x \mapsto \sum_{(i_1,\dots,i_D) \in \mathbb{N}^D \,:\, \|i\|_1 \leqslant m} c_i \prod_{a=1}^D x_a^{i_a}.$$

Let $\mathcal{H}$ be a subset of functions $\mathbb{C}^D \to \mathbb{C}^D$ such that all elements of $\mathcal{H}$ satisfy (Adep) and such that the set of the restrictions to $[-\nu_{\text{est}}, \nu_{\text{est}}]^D$ of functions in $\mathcal{H}$ is closed in $L^2([-\nu_{\text{est}}, \nu_{\text{est}}]^D)$. We are now ready to define our estimator of Φ_X:

For any integer m and any $\rho > 1$, let $\widehat{\Phi}_{n,m,\rho}$ be a (up to $1/n$) measurable minimizer of the functional $\phi \mapsto M_n(T_m\phi)$ over $\Upsilon_{\rho,S} \cap \mathcal{H}$.

For good choices of m, $\widehat{\Phi}_{n,m,\rho}$ is a consistent estimator of Φ_X in $L^2([-\nu,\nu]^d)$ at almost parametric rate. The constants will depend on the signal through ρ and S, and on the noise through its second moment and the following quantity:

$$c_\nu = \inf\{|\Phi_{\varepsilon^{(1)}}(t)|,\ t \in [-\nu,\nu]^{d_1}\} \wedge \inf\{|\Phi_{\varepsilon^{(2)}}(t)|,\ t \in [-\nu,\nu]^{d_2}\}. \tag{3}$$

Note that for any noise distribution, for small enough ν, c_ν is a positive real number. For any $\nu > 0$, $c(\nu) > 0$, $E > 0$, define $\mathcal{Q}^{(D)}(\nu, c(\nu), E)$ the set of distributions $\mathbb{Q} = \otimes_{j=1}^D \mathbb{Q}_j$ on $\mathbb{R}^D$ such that $c_\nu \geqslant c(\nu)$ and $\int_{\mathbb{R}^D} \|x\|^2 d\mathbb{Q}(x) \leqslant E$.

Proposition 6 (Variant of Proposition 1 in [9])**.** *For all $\rho_0 < 2$, $\nu \in (0, \nu_{est}]$, $S, c(\nu), E, C > 0$ and $\delta, \delta', \delta'' \in (0,1)$ with $\delta' > \delta$, there exist positive constants c and n_0 such that the following holds: let $\rho \in [1, \rho_0]$, for all $\Phi_X \in \Upsilon_{\rho,S} \cap \mathcal{H}$ and $\mathbb{Q} \in \mathcal{Q}^{(D)}(\nu, c(\nu), E)$, for all $n \geqslant n_0$ and $x \in [1, n^{1-\delta'}]$, with probability at least $1 - 2e^{-x}$,*

$$\sup_{\rho' \in [\rho,\rho_0],\ m \in [2\rho' \frac{\log n}{\log\log n}, C\frac{\log n}{\log\log n}]} \int_{B_\nu^{d_1} \times B_\nu^{d_2}} |\widehat{\Phi}_{n,m,\rho'}(t) - \Phi_X(t)|^2 dt \leqslant c\left(\frac{x}{n^{1-\delta}}\right)^{1-\delta''}.$$

Moreover, the same result holds when replacing $\widehat{\Phi}_{n,m,\rho'}$ by $T_m\widehat{\Phi}_{n,m,\rho'}$.

Note that the constants c and n_0 do not depend on the distribution of X or ε. The proof of Proposition 6 is based on results in [9] and [19] and is detailed in Section 6.7. For sake of simplicity, we denote $\widehat{\Phi}_{n,\rho}$ the estimator $\widehat{\Phi}_{n,m,\rho}$ in which $m = \lceil 4\frac{\log n}{\log\log n}\rceil$. Note that this is a valid choice of m for any $\rho \in [1, 2)$.

3.2 Estimation of the support: upper bound

We are now ready to provide an estimator of the support of the signal. The idea is the following. Define $\bar{g}$ a probability density which is the convolution of G, the unknown distribution of the signal X, with a kernel $\Psi_{A,h}$ defined later, in which h is a bandwidth parameter. Then, the multiplication of the estimator of the characteristic function of the signal with the Fourier transform of the kernel will give a good estimator $\widehat{g}_n$ of $\bar{g}$. When h tends to 0, $\bar{g}$ becomes larger on $\mathcal{M}$ and tends to 0 outside of it. Thus, by letting h tend to 0 with n and choosing an appropriate threshold λ_n, the set of points y for which $\widehat{g}_n \geqslant \lambda_n$ should be a good estimator of $\mathcal{M}$. Figure 2 illustrates this idea.

We now define the class over which we will prove an upper bound for the maximum risk in Hausdorff distance.

For any compact set $\mathcal{K}$ of $\mathbb{R}^D$, and for any positive constants a, d and r_0, we define $St_{\mathcal{K}}(a, d, r_0)$ as the set of positive measures G such that for all $x \in \mathcal{K}$, for all $r \leqslant r_0$, $G(B(x,r)) \geqslant ar^d$. The distributions in $St_{\mathcal{K}}(a, d, r_0)$ are called (a, d)-standard. It is commonly used for inferring topological information, see for instance [6].

Remark 1.

- *If a measure μ (for instance the d-dimensional Hausdorff measure on a manifold) is (a, d')-standard for some positive constants a and d', and if G admits a density g with respect to μ such that g is lower bounded by $c > 0$, then G is (ac, d')-standard.*
- *We do not make any assumptions on the reach of the support of G (see [17]) since it is not necessary here, although it provides a convenient way to check the (a, d)-standard assumption: if $\mathcal{M}$ is a Riemannian manifold of dimension d with reach$(\mathcal{M}) \geqslant \tau_{\min} > 0$, then the d-dimensional Hausdorff measure restricted to $\mathcal{M}$ is (a, d)-standard for some $a > 0$ (see Lemma 32 of [21]).*

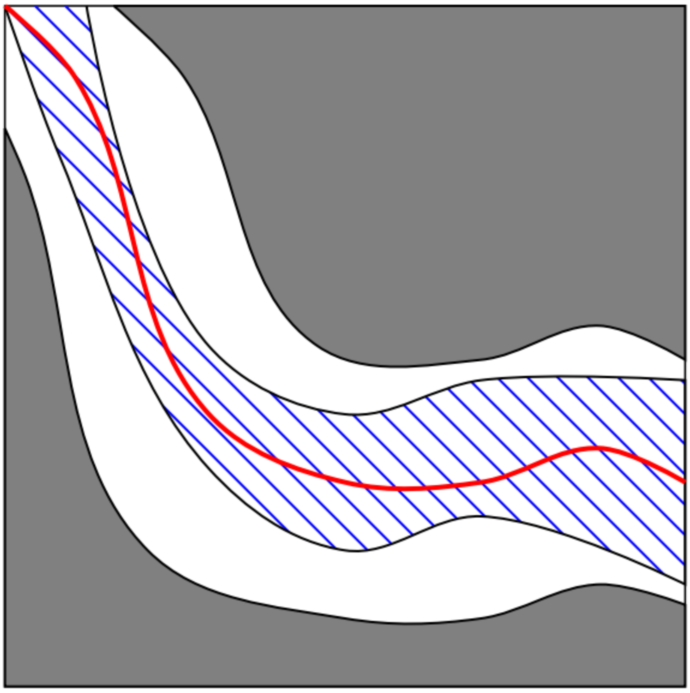

Figure 2: In red, the support of the signal distribution $\mathcal{M}$, the blue hatched area represents the set $\{\bar{g} \geqslant \lambda_n + \|\bar{g} - \widehat{g}_n\|_\infty\}$ and the gray area represents the set $\{\bar{g} \leqslant \lambda_n - \|\bar{g} - \widehat{g}_n\|_\infty\}$, so that the estimator of the support lies in between the gray and the blue areas.

As in [19], it will be convenient to use $\kappa = 1/\rho$. We denote $\mathcal{L}(\kappa, S, \mathcal{H})$ the set of distributions G such that, if X is a random variable with distribution G, then $\Phi_X \in \mathcal{H} \cap \Upsilon_{1/\kappa,S}$.

When $\kappa < 1$ and $G \in \mathcal{L}(\kappa, S, \mathcal{H})$, the support of G is not compact. Since we allow the support to be a non-compact set, we define a truncated loss function as in [21]. For any $\mathcal{K}$ compact subset of $\mathbb{R}^D$ and for any S_1, S_2 subsets of $\mathbb{R}^D$, the truncated loss function is

$$H_{\mathcal{K}}(S_1, S_2) = d_H(S_1 \cap \mathcal{K}, S_2 \cap \mathcal{K}).$$

We now introduce the kernel we shall use for our construction. For any $A > 0$, define, for all $y \in \mathbb{R}$,

$$u_A(y) = \exp\left\{-\frac{1}{(1-2y)^A} - \frac{1}{(1+2y)^A}\right\} 1|_{[-\frac{1}{2},\frac{1}{2}]}(y)$$

and

$$\tilde{\psi}_A(y) = I(A)\ \mathcal{F}^{-1}[u_A * u_A](y) \text{ with } I(A) = \frac{1}{\int \mathcal{F}^{-1}[u_A * u_A](x)dx}.$$

We shall extend $\tilde{\psi}_A$ to $\mathbb{R}^D$ as an isotropic function. For $y \in \mathbb{R}^D$, we write

$$\psi_A(y) = I(A)\ \mathcal{F}^{-1}[u_A * u_A](\|y\|_2) \text{ with } I(A) = \frac{1}{\int \mathcal{F}^{-1}[u_A * u_A](\|x\|_2)dx}.$$

For $h > 0$ and $x \in \mathbb{R}^D$, we write $\psi_{A,h}(x) = h^{-D}\psi_A(x/h)$, hence $\mathcal{F}[\psi_{A,h}](t) = \mathcal{F}[\psi_A](th)$. The following properties of ψ_A and $\mathcal{F}[\psi_A]$ hold

(I) The support of $\mathcal{F}[\psi_A]$ is the unit ball $\{y \in \mathbb{R}^D\ :\ \|y\|_2 \leqslant 1\}$.

(II) $\psi_A > 0$ and $\mathcal{F}[\psi_A] \geqslant 0$.

(III) There exist constants $c_A > 0$ and $d_A > 0$ such that for all $x \in \{y \in \mathbb{R}^D\ :\ \|y\|_2 \leqslant c_A\}$, $\psi_A(x) \geqslant d_A$.

(IV) ψ_A and $\psi_{A,h}$ are probability densities on $\mathbb{R}^D$.

(V) (Lemma in [30]) For all $A > 0$, there exists $\beta_A > 0$ such that

$$\lim_{\|t\|_2 \to \infty} \exp\{\beta_A \|t\|_2^{\frac{A}{A+1}}\}\psi_A(t) = 0 \tag{4}$$

(VI) It holds

$$\|\psi_{A,h}\|_2 = \frac{I(A)}{h^{D/2}}\|u_A * u_A\|_2. \tag{5}$$

Fix $A > 0$ and define the convoluted density of the signal, $\bar{g}$ by

$$\forall y \in \mathbb{R}^D,\ \bar{g}(y) = (\frac{1}{2\pi})^D \int e^{-it^\top y}\mathcal{F}[\psi_A](ht)\ \Phi_X(t)dt,$$

which may be rewritten using usual Fourier calculus, for all $y \in \mathbb{R}^D$, as

$$\bar{g}(y) = (\psi_{A,h} * G)(y) = \frac{1}{h^D}\int_{\mathbb{R}^D} \psi_A(\frac{\|y-u\|_2}{h})dG(u).$$

The density $\bar{g}$ is a kernel smoothing of the distribution G. The bandwidth parameter h will be chosen appropriately in Theorem 4 below.

We now construct an estimator of $\bar{g}$ by truncating $\widehat{\Phi}_{n,1/\kappa}$ depending on κ. Adaptation with respect to κ is handled in Section 3.4. For some integer $m_\kappa > 0$ to be chosen later, let

$$\forall y \in \mathbb{R}^D,\ \widehat{g}_{n,\kappa}(y) = \left(\frac{1}{2\pi}\right)^D \int e^{-it^\top y}\mathcal{F}[\psi_A](ht)\ T_{m_\kappa}\widehat{\Phi}_{n,1/\kappa}(t)dt.$$

Since for all $t \in \mathbb{R}^D$, $T_{m_\kappa}\widehat{\Phi}_{n,1/\kappa}(-t) = T_{m_\kappa}\widehat{\Phi}_{n,1/\kappa}(t)$, the function $\widehat{g}_{n,\kappa}$ is real valued. Finally, define an estimator of the support of the signal as the upper level set

$$\widehat{\mathcal{M}}_\kappa = \left\{y \in \mathbb{R}^D \mid \widehat{g}_{n,\kappa}(y) > \lambda_{n,\kappa}\right\},$$

for some $\lambda_{n,\kappa}$. The main theorem of this section gives an upper bound of the maximum risk.

Theorem 4. *Let $\kappa \in (1/2, 1]$, $a > 0$, $d \leqslant D$, $r_0 > 0$. For $c_h \geqslant \exp(2D + 2)$ and $\ell \in (0, 1)$, define m_κ and h as*

$$m_\kappa = \left\lfloor \frac{1}{4\kappa}\frac{\log(n)}{\log\log(n)} \right\rfloor, \quad h = c_h S m_\kappa^{-\kappa}$$

and $\lambda_{n,\kappa}$ depending whether $d < D$ or $d = D$ as

- *if $d < D$,*

$$\lambda_{n,\kappa} = \left(\frac{1}{h}\right)^\ell,$$

- *if $d = D$,*

$$\lambda_{n,\kappa} = \frac{1}{4}ac_A^D d_A.$$

Then for any $\kappa_0 \in (1/2, 1]$, $\nu \in (0, \nu_{est}]$, $c(\nu) > 0$, $E > 0$ $S > 0$, there exists n_0 and $C > 0$ such that for all $n \geqslant n_0$,

$$\sup_{\kappa \in [\kappa_0, 1]} \sup_{\substack{G \in St_\mathcal{K}(a,d,r_0)\cap\mathcal{L}(\kappa,S,\mathcal{H}) \\ \mathbb{Q}\in\mathcal{Q}^{(D)}(\nu,c(\nu),E)}} \mathbb{E}_{(G*\mathbb{Q})^{\otimes n}}[H_\mathcal{K}(\mathcal{M}_G, \widehat{\mathcal{M}}_\kappa)] \leqslant C\frac{\log(\log(n))^{\kappa+\frac{A+1}{A}}}{\log(n)^\kappa}.$$

Remark 2.

- *We prove in the next section a nearly matching lower bound. Thus, the minimax rate of convergence of the support in truncated Hausdorff distance depends on κ, that is on the way the distribution of the signal behaves at infinity. This rate deteriorates when the distribution of the signal has heavier tails. Indeed, since the distribution of the noise is unknown, taking into account distant observation points to build the estimator of the support becomes more difficult.*

- *When $d < D$, thanks to the use of the kernel ψ_A, our estimator does not require the knowledge of d, which has to be compared with the estimator in [21] where prior knowledge of d is needed.*

- *In [21], the upper bound on the rate is of order $1/\sqrt{\log n}$. Here we get a bound of order $1/(\log n)^{\kappa}$ depending on the tail of the distribution of the signal. We do not need to know the distribution of the noise, contrarily to [21] where the distribution of the noise is used in the construction of the estimator, as usual in the classical deconvolution litterature.*

- *It may be seen from the proof of Theorem 4 that the choice $\lambda_{n,\kappa} = \frac{1}{4}ac_A^D d_A$ is valid for any d. However, this requires the knowledge of a.*

- *Note that there are two truncation steps: the first one in the construction of $\hat{\Phi}_{n,1/\kappa}$ (chosen at the end of Section 3) and the second one in the definition of $\widehat{g}_{n,\kappa}$. This second truncation is necessary to control the error of $\hat{\Phi}_{n,1/\kappa}$ on $B^D_{1/h}$ (see Lemma 3, compared to the error on B^D_{ν} in Proposition 6), and the degree m_κ in the second truncation is always smaller than the degree m used in the construction of $\hat{\Phi}_{n,1/\kappa}$.*

The proof of Theorem 4 is detailed in section 6.11. As in [21], the idea is to lower bound $\bar{g}$ on the support $\mathcal{M}_G$ when the bandwidth parameter h becomes small, and to upper bound it on every points further than a small distance (depending on h) from that support, see Lemma 1 and 2 below.

Lemma 1. *Assume $G \in St_{\mathcal{K}}(a, d, r_0)$, then for any $h \leqslant r_0/c_A$,*

$$\inf_{y \in \mathcal{M}_G \cap \mathcal{K}} \bar{g}(y) \geqslant a c_A^d d_A \left(\frac{1}{h}\right)^{D-d},$$

where c_A and d_A are defined in property (III) of ψ_A.

The proof of Lemma 1 is detailed in Section 6.8.

Lemma 2. *For any $C_1 > 0$, there exists $h_0 > 0$ depending only on C_1, D and A such that for any $h \leqslant h_0$,*

$$\sup \left\{ \bar{g}(y) \mid y \in \mathcal{K},\ d(y, \mathcal{M}_G) > h \left[\frac{D}{\beta_A} \log\left(\frac{1}{h}\right)\right]^{\frac{A+1}{A}} \right\} \leqslant C_1.$$

The proof of Lemma 2 is detailed in Section 6.9.

The last ingredient is to control the difference between the convoluted density and its estimator, defined as $\Gamma_{n,\kappa} = \|\widehat{g}_{n,\kappa} - \bar{g}\|_\infty = \sup_{y \in \mathbb{R}^D} |\widehat{g}_{n,\kappa}(y) - \bar{g}(y)|$. We first relate it to $\|T_{m_\kappa} \widehat{\Phi}_{n,1/\kappa} - \Phi_X\|_{2,1/h}$.

Lemma 3. *Let $h > 0$ and $m > 0$. For any $A > 0$,*

$$\Gamma_{n,\kappa} \leqslant I(A) \frac{\|u_A * u_A\|_2}{h^{D/2}} \|T_{m_\kappa} \widehat{\Phi}_{n,1/\kappa} - \Phi_X\|_{2,1/h}.$$

The proof of Lemma 3 is detailed in Section 6.10. The parameters m_κ and h are chosen so that $\Gamma_{n,\kappa}$ tends to 0 with high probability, and the threshold $\lambda_{n,\kappa}$ is chosen using Lemmas 1 and 2.

3.3 Lower bound

The aim of this subsection is to prove a lower bound for the minimax risk of the estimation of $\mathcal{M}_G$ using the distance $H_{\mathcal{K}}$ as loss function. The proof of Theorem 5 is based on Le Cam's two-points method, see [31], one of the most widespread technique to derive lower bounds. Note that we can not use the lower bound proved in [21] since the two distributions they use for the signal X in their two-points proof have Gaussian tails, for which $\kappa = 1/2$.

Theorem 5. *For any $\kappa \in (1/2, 1]$, there exists $S_\kappa > 0$, $a_\kappa > 0$ and $\mathcal{H}_\kappa^\star$ a set of complex functions satisfying (Adep) such that the set of the restrictions of its elements to $[-\nu, \nu]^D$ is closed in $L_2([-\nu, \nu]^D)$ for any $\nu > 0$, and such that for all $S \geqslant S_\kappa$, $a \leqslant a_\kappa$, $d \geqslant 1$, $0 < r_0 < 1$, $E > 0$ and $\nu \in (0, \nu_{est}]$ such that $c(\nu) > 0$, there exists $C > 0$ depending only on a, D, S, E and ν, and there exists n_0, such that for all $n \geqslant n_0$,*

$$\inf_{\widehat{\mathcal{M}}} \sup_{\substack{G \in St_{\mathcal{K}}(a,d,r_0) \cap \mathcal{L}(\kappa, S, \mathcal{H}_\kappa^\star) \\ \mathbb{Q} \in \mathcal{Q}^{(D)}(\nu, c(\nu), E)}} \mathbb{E}_{(G*\mathbb{Q})^{\otimes n}}[H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}})] \geqslant \frac{C}{\log(n)^\kappa}, \tag{6}$$

and for any $\delta \in (0, 1)$, there exists $C > 0$ depending only on a, D, S, E, ν and δ, and there exists n_0, such that for all $n \geqslant n_0$,

$$\inf_{\widehat{\mathcal{M}}} \sup_{\substack{G \in St_{\mathcal{K}}(a,d,r_0) \cap \mathcal{L}(1, S, \mathcal{H}_1^\star) \\ \mathbb{Q} \in \mathcal{Q}^{(D)}(\nu, c(\nu), E)}} \mathbb{E}_{(G*\mathbb{Q})^{\otimes n}}[H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}})] \geqslant \frac{C}{\log(n)^{1+\delta}}, \tag{7}$$

where the infimum in (6) and (7) is taken over all possible estimators $\widehat{\mathcal{M}}$ of $\mathcal{M}_G$.

Remark 3.
- *The lower bound in Theorem 5 almost matches the upper bound for the maximum risk of our estimator in Theorem 4. Thus our work identifies the main factor in the minimax rate for the estimation of the support in Hausdorff loss.*
- *In [21], the lower bound does not match the upper bound by a larger power in the rate (almost twice).*
- *The sets of supports we consider are not the same as that considered in [21]. In [21], the authors assume that the support is a regular manifold with lower bounded reach. We do not assume regularity, we only assume that the distribution of the signal is (a, d)-standard.*

As usual for the two-points method, the idea is to find two distributions having support as far as possible in $H_{\mathcal{K}}$-distance, and a noise such that the joint distributions of the observations have total variation distance upper bounded by some $C < 1$.

Let us introduce the two distributions used in the two-points method, as well as the closed set $\mathcal{H}_\kappa^\star$. The rest of the proof of Theorem 5 is detailed in Section 6.17.

We shall consider the noise as in [19], with independent identically distributed coordinates having density q defined as

$$q : x \in \mathbb{R} \longmapsto c_q \frac{1 + \cos(cx)}{(\pi^2 - (cx)^2)^2}$$

for some $c > 0$, where c_q is such that q is a probability density, and with characteristic function

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

Let us now define the two distributions to apply the two-points method. For any $\kappa \in (1/2, 1]$, we first choose a density function f_κ according to the following Lemma.

Lemma 4. *For any $\kappa \in (1/2, 1)$, $p \geqslant 1$, there exists a continuous density function $f_\kappa : \mathbb{R} \to \mathbb{R}$ in $L_p(\mathbb{R})$, positive everywhere, and positive constants A, B such that for all $u \in \mathbb{R}$,*

$$|\mathcal{F}[f_\kappa](u)| \leqslant A \exp(-B|u|^{\frac{1}{\kappa}}) \quad \text{and} \quad |\mathcal{F}[f_\kappa]'(u)| \leqslant A \exp(-B|u|^{\frac{1}{\kappa}}).$$

For any $\delta \in (0,1)$, there exists a continuous compactly supported density function $f_1 : [-1,1] \to \mathbb{R}$ positive everywhere such that

$$|\mathcal{F}[f_1](u)| \leqslant A\exp(-B|u|^\delta) \quad \text{and} \quad |\mathcal{F}[f_1]'(u)| \leqslant A\exp(-B|u|^\delta).$$

The proof of Lemma 4 is detailed in Section 6.12.

Then, inspired by [21], for all $\gamma \in (0,1]$, define $\tilde{g}_\gamma : \mathbb{R} \to \mathbb{R}$ and $g_\gamma : \mathbb{R} \to \mathbb{R}^{D-d}$ for all $x \in \mathbb{R}$ as

$$\tilde{g}_\gamma(x) = \cos\left(\frac{x}{\gamma}\right) \quad \text{and} \quad g_\gamma(x) = (\tilde{g}_\gamma(x), 0, \ldots, 0)^\top .$$

Let $M_0(\gamma) = \{(u, \gamma g_\gamma(u)) : u \in \mathbb{R}\}$, $M_1(\gamma) = \{(u, -\gamma g_\gamma(u)) : u \in \mathbb{R}\}$, and for $\alpha \neq 0$, define the matrix $A_\alpha \in \mathbb{R}^{D\times D}$ by

$$A_\alpha = \left(\begin{array}{cc|ccc} \alpha & 0 & 0 & \ldots & 0 \\ \alpha & \alpha/2 & 0 & \ldots & 0 \\ \hline 0 & 0 & \alpha & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \alpha \end{array}\right).$$

For any $\kappa \in (1/2,1]$, let $U(\kappa)$ be the random variable in $\mathbb{R}$ having density f_κ defined in Lemma 4 and let $S_0(\kappa) = (U(\kappa), \gamma g_\gamma(U(\kappa)))$, $S_1(\kappa) = (U(\kappa), -\gamma g_\gamma(U(\kappa)))$. For $i \in \{0,1\}$, we shall denote $T_i(\kappa)$ the distribution of $S_i(\kappa)$. Finally we define $X_i(\kappa) = A_\alpha S_i(\kappa)$, $i = 0,1$ and $G_i(\kappa)$ the distribution of $X_i(\kappa)$.

To obtain the lower bound, the parameter γ will be chosen as large as possible while making sure that the joint distributions of the observations have total variation distance smaller than some $C < 1$.

Let us comment on dimensionality. The distributions used here are distributions with support of dimension 1. This is not an issue since the d in the definition of $St_{\mathcal{K}}(a,d,r_0)$ is an upper bound on the dimension of the support. We could also have used supports with dimension d by adding to $X_i(\kappa)$ an independent uniform distribution on a ball of a linear space of dimension d.

Lemma 5. *For any $i \in \{0,1\}$ and $\kappa \in (1/2,1]$, $X_i(\kappa)$ satisfies $A(1/\kappa)$.*

The proof of Lemma 5 is detailed in Section 6.13.

Lemma 6. *Let $\alpha > 0$. There exists $a_0 > 0$ such that for $i \in \{0,1\}$, for any $d \geqslant 1$, $r_0 < 1$ and $a \leqslant a_0$, $G_i(\kappa) \in St_{\mathcal{K}}(a,d,r_0)$.*

The proof of Lemma 6 is detailed in Section 6.14.

The support of $G_i(\kappa)$ is $A_\alpha M_i(\gamma)$, and the following lemma follows easily from the fact that for $i \in \{0,1\}$, $A_\alpha M_i(\gamma) = \gamma A_\alpha M_i(1)$.

Lemma 7. *For any $\gamma > 0$, and $\alpha > 0$,*

$$H_{\mathcal{K}}(A_\alpha M_0(\gamma), A_\alpha M_1(\gamma)) = \gamma H_{\mathcal{K}}(A_\alpha M_0(1), A_\alpha M_1(1)). \tag{8}$$

We finally exhibit a set of complex functions $\mathcal{H}^\star_\kappa$ such that all of its elements satisfy (Adep) and such that the set of the restrictions of its elements to $[-\nu,\nu]^D$ is closed in $L_2([-\nu,\nu]^D)$ for any $\nu > 0$.

We first define a class of such sets of complex functions, then choose $\mathcal{H}^\star_\kappa$ as one in that class. Let $(A^{(1)}_\Delta)_{\Delta>0}$ and $(B^{(1)}_\Delta)_{\Delta>0}$ be families of subsets of $\mathbb{R}$ and $(A^{(2)}_\Delta)_{\Delta>0}$ and $(B^{(2)}_\Delta)_{\Delta>0}$ be families of subsets of $\mathbb{R}^{D-1}$ such that

(i) For all $\Delta > 0$, $\mathrm{Diam}(B^{(2)}_\Delta) \leqslant \Delta$ and $\mathrm{Diam}(B^{(1)}_\Delta) \leqslant \Delta$,

(ii) The topological frontiers $\partial A^{(1)}_\Delta$ and $\partial A^{(2)}_\Delta$ in $\mathbb{R}$ are negligible with respect to the Lebesgue measure.

For $\kappa \in (1/2, 1]$, $S > 0$, $M > 0$ and $(c_\Delta)_{\Delta>0}$ a family of positive constants, let

$$\mathcal{H}(\kappa, S, M, (c_\Delta, A_\Delta^{(1)}, B_\Delta^{(1)}, A_\Delta^{(2)}, B_\Delta^{(2)})_{\Delta>0})$$

be the set of functions $\phi : \mathbb{C}^D \to \mathbb{C}$ in $\Upsilon_{\kappa,S}$ such that there exists a random variable X satisfying $A(1/\kappa)$ such that $\phi = \Phi_X$ and such that the following holds. Write $X^{(1)}$ the first coordinate of X and $X^{(2)}$ the vector of the last $D-1$ coordinates of X.

(iii) For all $\Delta > 0$, $\mathbb{P}[X^{(1)} \in A_\Delta^{(1)}] \geqslant c_\Delta$ and $\mathbb{P}[X^{(2)} \in B_\Delta^{(2)} | X^{(1)} \in A_\Delta^{(1)}] = 1$.

(iv) For all $\Delta > 0$, $\mathbb{P}[X^{(2)} \in A_\Delta^{(2)}] \geqslant c_\Delta$ and $\mathbb{P}[X^{(1)} \in B_\Delta^{(1)} | X^{(2)} \in A_\Delta^{(2)}] = 1$.

(v) All the coordinates of $X^{(2)}$ are null except the first one, and $X^{(1)}$ and the first coordinate of $X^{(2)}$ admit a continuous density with respect to Lebesgue measure which is upper bounded by M.

Lemma 8. *For any $\nu > 0$, the set $\mathcal{H}(\kappa, S, M, (c_\Delta, A_\Delta^{(1)}, B_\Delta^{(1)}, A_\Delta^{(2)}, B_\Delta^{(2)})_{\Delta>0})$ is closed in $L_2([-\nu, \nu]^D)$. Moreover, all elements of $\mathcal{H}(M, (c_\Delta, A_\Delta^{(1)}, B_\Delta^{(1)}, A_\Delta^{(2)}, B_\Delta^{(2)})_{\Delta>0})$ satisfy (Adep).*

Note that $\Phi_X \in \mathcal{H}(\kappa, S, M, (c_\Delta, A_\Delta^{(1)}, B_\Delta^{(1)}, A_\Delta^{(2)}, B_\Delta^{(2)})_{\Delta>0})$ implies that X satisfies (H1) and (H2) so that the second part of the Lemma is a consequence of Theorem 2. The remaining of the proof of Lemma 8 is detailed in Section 6.15.

Lemma 9. *For any $\kappa \in (1/2, 1]$, there exist S_κ, $M > 0$, $(c_\Delta)_{\Delta>0}$ a sequence of positive constants, and $(A_\Delta^{(1)}, B_\Delta^{(1)}, A_\Delta^{(2)}, B_\Delta^{(2)})_{\Delta>0}$ a sequence of sets such that for $i \in \{0, 1\}$, $\Phi_{X_i(\kappa)} \in \mathcal{H}(\kappa, S_\kappa, M, (c_\Delta, A_\Delta^{(1)}, B_\Delta^{(1)}, A_\Delta^{(2)}, B_\Delta^{(2)})_{\Delta>0})$.*

The proof of Lemma 9 is detailed in Section 6.16.

3.4 Adaptation to unknown κ

We now propose a data-driven model selection procedure to select κ such that the resulting estimator has the right rate of convergence. As usual, the idea is to perform a bias-variance trade off. Although we have an upper bound for the variance term, the bias is not easily accessible. We will use Goldenshluger and Lepski's method, see [22]. The variance bound is given as follows:

$$\sigma_n(\kappa) = c_\sigma \frac{(\log \log n)^{\kappa + \frac{A+1}{A}}}{(\log n)^\kappa}.$$

Fix some $\kappa_0 > 1/2$. The bias proxy is defined as

$$B_n(\kappa) = 0 \vee \sup_{\kappa' \in [\kappa_0, \kappa]} \left(H_{\mathcal{K}}(\widehat{\mathcal{M}}_\kappa, \widehat{\mathcal{M}}_{\kappa'}) - \sigma_n(\kappa') \right).$$

The estimator of κ is now given by

$$\widehat{\kappa}_n \in \arg\min \left\{ B_n(\kappa) + \sigma_n(\kappa),\ \kappa \in [\kappa_0, 1] \right\},$$

and the estimator of the support of the signal is $\widehat{\mathcal{M}}_{\widehat{\kappa}_n}$. The following theorem states that this estimator is rate adaptive.

Theorem 6. *For any $\kappa_0 \in (1/2, 1]$, $\nu \in (0, \nu_{est}]$, $c(\nu) > 0$, $E > 0$ $S > 0$, $a > 0$, $d \leqslant D$, there exists $c_\sigma > 0$ such that*

$$\limsup_{n \to +\infty} \sup_{\kappa \in [\kappa_0, 1]} \sup_{\substack{G \in St_{\mathcal{K}}(a,d) \cap \mathcal{L}(\kappa, S, \mathcal{H}) \\ \mathbb{Q} \in \mathcal{Q}^{(D)}(\nu, c(\nu), E)}} \frac{\log(n)^\kappa}{\log(\log(n))^{\kappa + \frac{A+1}{A}}} \mathbb{E}_{(G * \mathbb{Q})^{\otimes n}} [H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}}_{\widehat{\kappa}_n})] < +\infty.$$

The proof of Theorem 6 is detailed in Section 6.18.

4 Estimation of the distribution of the signal

In this section, we assume that the support $\mathcal{M}_G$ of G is a compact subset of $\mathbb{R}^D$. To estimate G, we shall consider the probability density $\bar{g}$ defined in Section 3.2 and define the probability distribution $P_{\psi_{A,h_n}}$ on $\mathbb{R}^D$ such that, for any $\mathcal{O}$ borelian set of $\mathbb{R}^D$,

$$P_{\psi_{A,h_n}}(\mathcal{O}) = \int_{\mathcal{O}} \bar{g}(y)dy.$$

This probability distribution can be considered as a convoluted approximation of G with kernel Ψ_A and smoothing parameter h_n. We then estimate $P_{\psi_{A,h_n}}$ using the estimation of $\bar{g}$ defined in Section 3.2 for $\kappa = 1$, $\widehat{g}_n := \widehat{g}_{n,1}$. Since $\widehat{g}_n$ can be non positive, we use $\widehat{g}_n^+ = \max\{0, \widehat{g}_n\}$ and renormalize it to get a probability distribution. We shall also estimate $P_{\psi_{A,h_n}}$ with a probability distribution having support on a (small) enlargement of the estimated support $\widehat{\mathcal{M}}$ restricted to the closed euclidean ball $\bar{B}(0, R_n)$, for some radius R_n that grows to infinity with n. Thus we fix some $\eta > 0$ and define $\widehat{P}_{n,\eta}$ such that, for any $\mathcal{O}$ borelian set of $\mathbb{R}^D$,

$$\widehat{P}_{n,\eta}(\mathcal{O}) = \frac{1}{\int_{(\widehat{\mathcal{M}}\cap\bar{B}(0,R_n))_\eta} \widehat{g}_n^+(y)dy} \int_{\mathcal{O}\cap(\widehat{\mathcal{M}}\cap\bar{B}(0,R_n))_\eta} \widehat{g}_n^+(y)dy = c_n \int_{\mathcal{O}\cap(\widehat{\mathcal{M}}\cap\bar{B}(0,R_n))_\eta} \widehat{g}_n^+(y)dy.$$

4.1 Upper bound for the Wasserstein risk

The aim of this subsection is to give an upper bound of the Wasserstein maximum risk for the estimation of G.

Theorem 7. *For all $\nu \in (0, \nu_{est}]$, $c(\nu) > 0$, $E > 0$, $S > 0$, $\eta > 0$, $a > 0$, $r_0 > 0$, $d \leqslant D$, define m_n, h_n and λ_n as in Theorem 4 for $\kappa = 1$. Assume that $\lim_{n\to+\infty} R_n = +\infty$ and that there exists $\delta \in (0, \frac{1}{2})$ such that $R_n \leqslant \exp(n^{1/2-\delta})$. Then there exist n_0 and $C > 0$ such that for all $n \geqslant n_0$,*

$$\sup_{\substack{G\in St_{\mathcal{K}}(a,d,r_0)\cap\mathcal{L}(1,S,\mathcal{H}) \\ \mathbb{Q}\in\mathcal{Q}^{(D)}(\nu,c(\nu),E)}} \mathbb{E}_{(G*\mathbb{Q})^{\otimes n}}[W_2(G, \widehat{P}_{n,\eta})] \leqslant C\frac{\log\log(n)}{\log(n)}.$$

The proof of Theorem 7 is detailed in Section 6.19. Note that the magnitude of η does not appear to be crucial when looking at the proof, at least in an asymptotic perspective.

Remark 4.
- *The lower bound in Theorem 8 almost matches the upper bound for the maximum risk of our estimator in Theorem 7. Thus our work identifies the main factor in the minimax rate for the estimation of the distribution in Wasserstein loss.*
- *Comparison with earlier results in the deconvolution setting [13] or [12] with known noise is not easy since the classes of signals they consider is much different than the ones we consider.*

4.2 Lower bound for the Wasserstein risk

The aim of this subsection is to establish a lower bound for the minimax Wasserstein risk of order p for any $p \geqslant 1$. Again, we can not use previous lower bounds proved in [13] or [12] since they use in the two-points method signals with distributions having too heavy tails.

Theorem 8. *For any $p \geqslant 1$, there exists $S_1 > 0$, $a_1 > 0$ and $\mathcal{H}_1^\star$ a set of complex functions satisfying (Adep) such that the set of the restrictions of its elements to $[-\nu,\nu]^D$ is closed in $L_2([-\nu,\nu]^D)$ for any $\nu > 0$, and such that for all $S \geqslant S_1$, $a \leqslant a_1$, $d \geqslant 1$, $0 < r_0 < 1$, $E > 0$ and $\nu \in (0, \nu_{est}]$ such that $c(\nu) > 0$, there exists $C > 0$ depending only on a, D, S, E and ν, and there exists n_0, such that for all $n \geqslant n_0$,*

$$\inf_{\widehat{P}_n} \sup_{\substack{G\in St_{\mathcal{K}}(a,d,r)\cap\mathcal{L}(1,S,\mathcal{H}_1^\star) \\ \mathbb{Q}\in\mathcal{Q}^{(D)}(\nu,c(\nu),E)}} \mathbb{E}_{(G*\mathbb{Q})^{\otimes n}}[W_p(G, \widehat{P}_n)] \geqslant C\frac{1}{\log(n)^{1+\delta}},$$

where the infimum is taken on all possible estimate $\widehat{P}_n$ of G.

As for Theorem 5, we use Le Cam's two-points method with the same two distributions $G_0(1)$ and $G_1(1)$. The proof essentially consists in showing that there exists a constant $C > 0$ independent of γ such that $W_p(G_0(1), G_1(1)) \geqslant CH_{\mathcal{K}}(M_0(\gamma), M_1(\gamma))$, that is $W_p(G_0(1), G_1(1)) \geqslant C\gamma$ for a constant $C > 0$. Once such an equality is established, the lower bound follows from taking γ as for Theorem 5.

The rest of the proof is detailed in Section 6.20.

5 Acknowledgments

Jeremie Capitao-Miniconi would like to acknowledge support from the UDOPIA-ANR-20-THIA-0013. Elisabeth Gassiat would like to acknowledge the Institut Universitaire de France and the ANR ASCAI : ANR-21-CE23-0035-02.

6 Proofs

6.1 Proof of Proposition 1

Case $\rho = 1$. It is clear that any compactly supported distribution satisfies A(1). Conversely, if $\mathbb{E}[e^{\langle \lambda, X\rangle}] \leqslant a \exp(b\|\lambda\|)$, then for any $\mu > 0$, we get, for any $b' > b$, if we denote $(e_j)_{1\leqslant j\leqslant D}$ the canonical basis of $\mathbb{R}^D$,

$$\begin{aligned}
\mathbb{P}(\|X\| \geqslant Db') &\leqslant \sum_{j=1}^{D} \mathbb{P}(|X_j| \geqslant b') \\
&= \sum_{j=1}^{D} \{\mathbb{P}(X_j \geqslant b') + \mathbb{P}(X_j \leqslant -b')\} \\
&= \sum_{j=1}^{D} \{\mathbb{P}(\langle \mu e_j, X\rangle \geqslant \mu b') + \mathbb{P}(-\langle \mu e_j, X\rangle \geqslant \mu b')\} \\
&\leqslant \sum_{j=1}^{D} \left\{ \frac{\mathbb{E}\left[\exp(\langle \mu e_j, X\rangle)\right]}{\exp(b'\mu)} + \frac{\mathbb{E}\left[\exp(-\langle \mu e_j, X\rangle)\right]}{\exp(b'\mu)} \right\} \quad \text{by Markov inequality} \\
&\leqslant 2D \frac{a\exp(b\mu)}{\exp(b'\mu)} \underset{\mu\to+\infty}{\longrightarrow} 0,
\end{aligned}$$

and hence $\|X\| \leqslant Db$ almost surely.

Case $\rho > 1$. Assume that for any $\lambda \in \mathbb{R}$, $\mathbb{E}[e^{\langle \lambda, X\rangle}] \leqslant a\exp(b\|\lambda\|^{\rho})$ for some $a, b > 0$. Then by using the same directional method as for $\rho = 1$, we get that for any $\mu, t \geqslant 0$,

$$\begin{aligned}
\mathbb{P}(\|X\| \geqslant t) &\leqslant 2Da\exp(b\mu^{\rho} - \mu t) \\
&= 2Da\exp\left(-\left(\frac{\rho}{b}\right)^{\frac{1}{\rho-1}} \left(1 - \rho^{-\frac{\rho+1}{\rho-1}}\right) t^{\frac{\rho}{\rho-1}} \right) \quad \text{by taking } \mu = \left(\frac{t}{b\rho}\right)^{\frac{1}{\rho-1}}.
\end{aligned}$$

Observe that since $\rho > 1$, $\left(1 - \rho^{-\frac{\rho+1}{\rho-1}}\right) > 0$ to get the result.

Now, assume that for any $t \geqslant 0$, $\mathbb{P}(\|X\| \geqslant t) \leqslant c\exp(-dt^{\rho/(\rho-1)})$ for some $c, d > 0$, then by the Cauchy-Schwarz inequality, for any $\lambda \in \mathbb{R}^D$,

$$\mathbb{E}[e^{\langle \lambda, X\rangle}] \leqslant \mathbb{E}[e^{\|\lambda\|\|X\|}].$$

Then, using that for any nonnegative random variable Y, $\mathbb{E}[Y] = \int_{t\geqslant 0} \mathbb{P}(Y \geqslant t)dt$,

$$\begin{aligned}
\mathbb{E}[e^{\langle\lambda,X\rangle}] &\leqslant 1 + \int_{t\geqslant 1} \mathbb{P}(e^{\|\lambda\|\|X\|} \geqslant t)dt \\
&\leqslant 1 + \int_{s\geqslant 0} \mathbb{P}(\|X\| \geqslant s)\|\lambda\| \exp(\|\lambda\|s)ds \quad \text{with } t = e^{\|\lambda\|s} \\
&\leqslant 1 + c\|\lambda\| \int_{s\geqslant 0} \exp(-ds^{\frac{\rho}{\rho-1}} + \|\lambda\|s)ds \\
&\leqslant 1 + c\|\lambda\|^{\rho} \int_{s'\geqslant 0} \exp(\|\lambda\|^{\rho}(-ds'^{\frac{\rho}{\rho-1}} + s'))ds' \quad \text{with } s = s'\|\lambda\|^{\rho-1}.
\end{aligned}$$

Note that $-ds'^{\frac{\rho}{\rho-1}} + s' \leqslant \frac{1}{\rho}(\frac{\rho-1}{\rho d})^{\rho-1}$ for any $s' \geqslant 0$, and $-ds'^{\frac{\rho}{\rho-1}} + s' \leqslant -s'$ when $s' \geqslant (\frac{2}{d})^{\rho-1}$. In particular,

$$\begin{aligned}
\mathbb{E}[e^{\langle\lambda,X\rangle}] &\leqslant 1 + c\|\lambda\|^{\rho} \int_{s'=0}^{(\frac{2}{d})^{\rho-1}} \exp(\|\lambda\|^{\rho}(-ds'^{\frac{\rho}{\rho-1}} + s'))ds' + c\|\lambda\|^{\rho} \int_{s'\geqslant(\frac{2}{d})^{\rho-1}} \exp(\|\lambda\|^{\rho}(-ds'^{\frac{\rho}{\rho-1}} + s'))ds' \\
&\leqslant 1 + c\|\lambda\|^{\rho} \left(\frac{2}{d}\right)^{\rho-1} \exp\left(\frac{\|\lambda\|^{\rho}}{\rho}\left(\frac{\rho-1}{\rho d}\right)^{\rho-1}\right) + c\exp\left(-\|\lambda\|^{\rho}\left(\frac{2}{d}\right)^{\rho-1}\right),
\end{aligned}$$

which proves that $A(\rho)$ holds.

6.2 Proof of Proposition 2

First, note that if U and V are independent random variables satisfying $\mathrm{A}(\rho)$ then $U+V$ satisfies also $\mathrm{A}(\rho)$ with the same constant ρ.

(i) If U and V are independent, then for all $(z_1, z_2) \in \mathbb{C}^{d_1} \times \mathbb{C}^{d_2}$,

$$\Phi_{U+V}(z_1, z_2) = \Phi_U(z_1, z_2)\Phi_V(z_1, z_2). \tag{9}$$

Assume first that U and V satisfy (Adep). Suppose that there exists $z_0 \in \mathbb{C}^{d_1}$ such that for all $z \in \mathbb{C}^{d_2}$, $\Phi_{U+V}(z_0, z) = 0$. Then for all $z \in \mathbb{C}^{d_2}$,

$$\Phi_U(z_0, z)\Phi_V(z_0, z) = 0.$$

If $Z_U^{(1)}(z_0) = \{z \in \mathbb{C}^{d_2} \mid \Phi_X(z_0, z) = 0\}$ and $Z_V^{(1)}(z_0) = \{z \in \mathbb{C}^{d_2} \mid \Phi_Y(z_0, z) = 0\}$, $Z_U^{(1)}(z_0) \cup Z_V^{(1)}(z_0) = \mathbb{C}^{d_2}$. Since $\Phi_U(z_0, \cdot)$ and $\Phi_V(z_0, \cdot)$ are not the null functions, Corollary 10 of [23], p. 9, implies that $Z_U^{(1)}(z_0) \cup Z_V^{(1)}(z_0)$ has zero $2d_2$-Lebesgue measure, which contradicts the fact that $Z_U^{(1)}(z_0) \cup Z_V^{(1)}(z_0) = \mathbb{C}^{d_2}$. If instead we suppose that there exists $z_0 \in \mathbb{C}^{d_2}$ such that for all $z \in \mathbb{C}^{d_1}$, $\Phi_{U+V}(z, z_0) = 0$, analogous arguments lead to a contradiction. Thus $U+V$ satisfies (Adep).

Assume now that $U+V$ satisfies (Adep). Then (9) implies that $\Phi_U(z_1, \cdot)$, $\Phi_V(z_1, \cdot)$, $\Phi_U(\cdot, z_2)$, $\Phi_V(\cdot, z_2)$ can not be the null function, so that U and V both satisfy (Adep).

(ii) Assume that U satisfies $A(\rho)$ with constants a and b. Then, for any $\lambda \in \mathbb{R}^D$,

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda^\top V\right)\right] &= \mathbb{E}\left[\exp\left(\lambda^\top \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} U^{(1)} \\ U^{(2)} \end{pmatrix} + \lambda^\top \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}\right)\right] \\
&\leqslant a\exp\left(b\left\|\lambda^\top \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\right\|_2^{\rho} + \|\lambda\|_2 \left\|\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}\right\|_2\right) \\
&\leqslant a\exp\left(b\left\|\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\right\|_{op}^{\rho} \|\lambda\|_2^{\rho} + \|\lambda\|_2 \left\|\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}\right\|_2\right).
\end{aligned}$$

Since $\rho \geqslant 1$, $\|\lambda\|_2 \leqslant \|\lambda\|_2^\rho$ for $\|\lambda\|_2 \geqslant 1$, so that if U satisfies $A(\rho)$ with constants a and b, then V satisfies $A(\rho)$ with constants $a \exp\left(\left\|\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}\right\|_2\right)$ and $b\left\|\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\right\|_{op}^\rho + \left\|\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}\right\|_2$. The converse follows from applying the direct proof to V with $-\begin{pmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{pmatrix}\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$ and $\begin{pmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{pmatrix}$.

Now, for all $(z_1, z_2) \in \mathbb{C}^{d_1} \times \mathbb{C}^{d_2}$,

$$\Phi_V(z_1, z_2) = \exp\left(\lambda\top \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}\right)\Phi_U(A\top z_1, B\top z_2)$$

and

$$\Phi_U(z_1, z_2) = \exp\left(-\lambda\top \begin{pmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{pmatrix}\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}\right)\Phi_V((A^{-1})\top z_1, (B^{-1})\top z_2),$$

so that U verifies (Adep) if and only if V verifies (Adep).

(iii) Since $U^{(1)}$ and $U^{(2)}$ are independent, for all $z_1 \in \mathbb{C}^{d_1}$ and $z_2 \in \mathbb{C}^{d_2}$, $\Phi_U(z_1, z_2) = \Phi_{U^{(1)}}(z_1)\Phi_{U^{(2)}}(z_2)$. Thus if $U^{(1)}$ and $U^{(2)}$ are deterministic or Gaussian random variables, U satisfies (Adep). Conversely, if U satisfies (Adep), then neither $\Phi_{U^{(1)}}$ nor $\Phi_{U^{(2)}}$ have any zero. By Hadamard's Theorem together with A(ρ), reasoning variable by variable we obtain that $\Phi_{U^{(1)}} = \exp(P_1)$ and $\Phi_{U^{(2)}} = \exp(P_2)$ for some polynomials P_1 and P_2 with degree bounded by ρ in each variable. Now, for $j = 1, 2$, for any $\lambda \in \mathbb{R}^{d_j}$, $t \mapsto \Phi_{U^{(j)}}(t\lambda)$ is the characteristic of the random variable $\langle \lambda, X^{(j)} \rangle$ and writes $\exp(P_j(t\lambda))$. But by Marcinkiewicz's theorem 2bis in [24], this implies that $t \mapsto P_j(t\lambda)$ is of degree at most two. Since this is true for any λ, we get that P_1 and P_2 are polynomials with total degree at most two. Thus the polynomials P_1 and P_2 are of the form $i\langle A, X\rangle - \frac{1}{2}X^\top BX$ for some symmetric matrix B since characteristic functions are equal to 1 at zero and $\Phi_U(-z) = \overline{\Phi_U(z)}$ for all $z \in \mathbb{R}^d$. Therefore the distribution of U_1 (resp. U_2) is a (possibly singular) Gaussian distribution.

6.3 Proof of Theorem 2

Consider a random variable X satisfying A(ρ). Theorem 2 is a direct consequence of the following Lemma. Indeed, for any $z_0 \in \mathbb{C}^{d_1}$ and $z \in \mathbb{C}^{d_2}$,

$$\mathbb{E}\left[\exp\left(iz_0^\top X^{(1)} + iz^\top X^{(2)}\right)\right] = \mathbb{E}\left[\mathbb{E}\left[\exp\left(iz_0^\top X^{(1)}\right) \mid X^{(2)}\right]\exp\left(iz^\top X^{(2)}\right)\right].$$

Usual arguments for multivariate analytic functions show that $z \mapsto \mathbb{E}[\exp(iz_0^\top X^{(1)} + iz^\top X^{(2)})]$ is the null function if and only if $\mathbb{E}[\exp(iz_0^\top X^{(1)}) \mid X^{(2)}]$ is zero $\mathbb{P}_{X^{(2)}}$-a.s. Likewise, for any $z_0 \in \mathbb{C}^{d_2}$, $z \mapsto \mathbb{E}[\exp(iz^\top X^{(1)} + iz_0^\top X^{(2)})]$ is the null function if and only if $\mathbb{E}[\exp(iz_0 \top X^{(2)}) \mid X^{(1)}]$ is zero $\mathbb{P}_{X^{(1)}}$-a.s.

Lemma 10. *Assume (H1) and (H2). Then, for all $z \in \mathbb{C}^{d_1}$, $\mathbb{E}[\exp(iz^\top X^{(1)}) \mid X^{(2)}]$ is not $\mathbb{P}_{X^{(2)}}$-a.s. the null random variable and for all $z \in \mathbb{C}^{d_2}$, $\mathbb{E}[\exp(iz^\top X^{(2)}) \mid X^{(1)}]$ is not $\mathbb{P}_{X^{(1)}}$-a.s. the null random variable*

Proof of Lemma 10 To begin with, by Proposition 2, we may assume without loss of generality that $0 \in B_\Delta$ in (H1) and (H2) (up to translation of X).

Let $z \in \mathbb{C}^{d_1}$ be such that $\mathbb{E}[\exp(iz^\top X^{(1)})|X^{(2)}]$ is $\mathbb{P}_{X^{(2)}}$-a.s. the null random variable. Then for any $\Delta > 0$, if we denote A_Δ a set given by (H1), $\mathbb{E}[\exp(iz^\top X^{(1)})|X^{(2)}]1|_{X^{(2)} \in A_\Delta} = 0$ $\mathbb{P}_{X_1^{(2)}}$ a.s., and taking the real part of this equation shows that

$$\mathbb{E}[\cos(\mathrm{Re}(z)^\top X^{(1)})\exp(-\mathrm{Im}(z)^\top X^{(1)})|X^{(2)}]1|_{X^{(2)} \in A_\Delta} = 0 \quad \mathbb{P}_{X^{(2)}}\,\text{a.s.} \tag{10}$$

Using (H1), we can fix $\Delta > 0$ small enough such that if $x \in B_\Delta$, $\cos(\mathrm{Re}(z)^\top x) > 0$. But for such Δ, Equation (10) can not hold since $\mathbb{P}(X^{(1)} \in B_\Delta \,|\, X^{(2)} \in A_\Delta) = 1$. Thus $\mathbb{E}[\exp(iz^\top X^{(1)})|X^{(2)}]$ is not $\mathbb{P}_{X^{(2)}}$-a.s. the null random variable.

The proof of the other part of Lemma 10 is analogous using (H2).

6.4 Proof of Proposition 4

Let $\mathcal{M}$ be a compact subset of $\mathbb{R}^D$. Let us first prove that the function $u \longmapsto \mathrm{Diam}(\{u\} \times \mathbb{R}^{d_2} \cap \mathcal{M})$ is upper semi-continuous.

Let $u \in \mathbb{R}^{d_1}$. Since $\mathcal{M}$ is compact, there exists sequences $u_n \to u$ and x_n, y_n in $(\{u_n\} \times \mathbb{R}^{d_2} \cap \mathcal{M})$ such that $\|x_n - y_n\|_2 = \mathrm{Diam}(\{u_n\} \times \mathbb{R}^{d_2} \cap \mathcal{M})$ and $\lim_{n\to+\infty} \|x_n - y_n\|_2 = \limsup_{v\to u} \mathrm{Diam}(\{v\} \times \mathbb{R}^{d_2} \cap \mathcal{M})$. Moreover, we may assume that there exists x, y in $(\{u\} \times \mathbb{R}^{d_2} \cap \mathcal{M})$ such that $x_n \to x$ and $y_n \to y$. Taking the limit along those sequences shows that $\mathrm{Diam}(\{u\} \times \mathbb{R}^{d_2} \cap \mathcal{M}) \geqslant \|x - y\| = \limsup_{v\to u} \mathrm{Diam}(\{v\} \times \mathbb{R}^{d_2} \cap \mathcal{M})$, proving the claimed upper-semi continuity.

Now, since $\mathcal{M}$ is compact, there exists $R > 0$ such that $\mathcal{M} \subset \bar{B}(0, R)$. If moreover $\mathcal{M} \in \mathcal{B}_1$, there exists $x_1 \in \mathbb{R}^{d_1}$ such that $\mathrm{Diam}(\{x_1\} \times \mathbb{R}^{d_2} \cap \mathcal{M}) = 0$. Using the upper semi-continuity shows that that $\mathcal{M} \in \cap_{n\geqslant 1}\mathcal{A}_2(1/n, R)$. Likewise, if $\mathcal{M} \in \mathcal{B}_2$, there exists $x_2 \in \mathbb{R}^{d_2}$ such that $\mathrm{Diam}(\mathbb{R}^{d_1} \times \{x_2\} \cap \mathcal{M}) = 0$ and $\mathcal{M} \in \cap_{n\geqslant 1}\mathcal{A}_1(1/n, R)$.

The end of the proof follows from Proposition 3 and the fact that any random variable with compact support satisfies A(1).

6.5 Proof of Proposition 5

First, let us show that the set $\mathcal{A} := (\cap_{\Delta>0} \cup_{\varepsilon>0} \mathcal{A}_1(\Delta, \varepsilon)) \bigcap (\cap_{\Delta>0} \cup_{\varepsilon>0} \mathcal{A}_2(\Delta, \varepsilon))$ is dense.

Let $\delta > 0$ and let $\mathcal{M}$ be a closed subset of $\mathbb{R}^D$, we show that there exists a closed $\mathcal{M}'$ in $\cap_{\Delta>0}(\mathcal{A}_1(\Delta, \delta) \cap \mathcal{A}_2(\Delta, \delta))$ (and thus in $\mathcal{A}$) such that $d_H(\mathcal{M}, \mathcal{M}') \leqslant 8\delta$.

Let $z = (z_1, z_2) \in \mathcal{M}$ with $z_1 = \pi^{(1:d_1)}(z)$ and $z_2 = \pi^{(d_1+1:D)}(z)$, $\mathcal{M}'$ is defined by cutting the space in half through z orthogonally to the space of the first d_1 cooordinates and spreading the two halves apart, connecting them by a single segment to ensure it is in $\mathcal{A}_2(\Delta, \delta)$, then cut and connect again orthogonally to the $(d_1 + 1)$-th axis to be in $\mathcal{A}_1(\Delta, \delta)$.

Formally, define $\mathcal{M}'$ as the union of:

- $\{y \,|\, y = (y_1, y_2) \in \mathcal{M}, \pi^{(1)}(y_1) \leqslant \pi^{(1)}(z_1) \text{ and } \pi^{(1)}(y_2) \leqslant \pi^{(1)}(z_2)\}$,
- $\{(y_1 + 4\delta(1, 0, \ldots, 0), y_2) \,|\, y = (y_1, y_2) \in \mathcal{M}, \pi^{(1)}(y_1) \geqslant \pi^{(1)}(z_1) \text{ and } \pi^{(1)}(y_2) \leqslant \pi^{(1)}(z_2)\}$,
- $\{(y_1, y_2 + 4\delta(1, 0, \ldots, 0)) \,|\, y = (y_1, y_2) \in \mathcal{M}, \pi^{(1)}(y_1) \leqslant \pi^{(1)}(z_1) \text{ and } \pi^{(1)}(y_2) \geqslant \pi^{(1)}(z_2)\}$,
- $\{(y_1+4\delta(1, 0, \ldots, 0), y_2+4\delta(1, 0, \ldots, 0)) \,|\, y = (y_1, y_2) \in \mathcal{M}, \pi^{(1)}(y_1) \geqslant \pi^{(1)}(z_1) \text{ and } \pi^{(1)}(y_2) \geqslant \pi^{(1)}(z_2)\}$,
- the segments between z and $(z_1+4\delta(1, 0, \ldots, 0), z_2)$ and between z and $(z_1, z_2+4\delta(1, 0, \ldots, 0))$.

An illustration of this construction is given in Figure 3.

By construction, the Hausdorff distance between this set $\mathcal{M}'$ and $\mathcal{M}$ is smaller than 8δ (the points in the first four sets have moved at most 8δ and the segments are at distance at most 8δ of z). $\mathcal{M}'$ is also closed, and taking $x = (z_1 + 2\delta(1, 0, \ldots, 0), z_2)$ and $x_2 = (z_1, z_2 + 2\delta(1, 0, \ldots, 0))$ in the definition of $\mathcal{A}_1(\Delta, \delta)$ and $\mathcal{A}_2(\Delta, \delta)$ is enough to check that $\mathcal{M}' \in \mathcal{A}_1(\Delta, \delta) \cap \mathcal{A}_2(\Delta, \delta)$ for any $\Delta > 0$.

To show that the complement of $\mathcal{A}$ is dense, let $\mathcal{M}$ be a closed subset of $\mathbb{R}^D$ and $\eta > 0$, and let $\mathcal{M}' = \{x + y \,|\, x \in \mathcal{M}, y \in [-\eta, \eta]^D\}$. Then $H(\mathcal{M}, \mathcal{M}') \leqslant \eta\sqrt{D}$ by construction, and for any $\Delta \leqslant 2\eta$ and $\varepsilon > 0$, $\mathcal{M}' \notin \mathcal{A}_1(\Delta, \varepsilon)$, and thus $\mathcal{M}' \in \mathcal{A}^\complement$.

Note that if $\mathcal{M}$ is the support of a random variable X, then $\mathcal{M}'$ is the support of $X + Y$, where Y is a uniform random variable on $[-\eta, \eta]^D$ that is independent of X. In that case, by Proposition 2 (i), $X + Y$ is a small perturbation of X that does not satisfy (Adep).

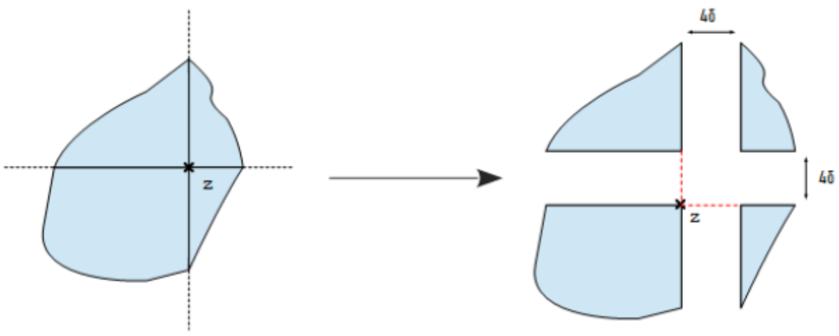

Figure 3: Transforming $\mathcal{M}$ into a set $\mathcal{M}' \in \mathcal{B}_1 \cap \mathcal{B}_2$

6.6 Proof of Theorem 3

Let $\mathcal{M}$ be a compact set of $\mathbb{R}^D$. Since ε and G are independent, and thus f^ε and $G(\mathcal{M})$ are independent, writing μ_G the distribution of $G(\mathcal{M})$:

$$\mathbb{P}(F(\mathcal{M}) \in \mathcal{B}_1 \cap \mathcal{B}_2) = \int \mathbb{P}\left(f^\varepsilon(\frac{g}{\delta}) \in \mathcal{B}_1 \cap \mathcal{B}_2\right) d\mu_G(g) = 1,$$

provided that for any compact set $\mathcal{M}' \in \mathbb{R}^D$, $f^\varepsilon(\mathcal{M}') \in \mathcal{B}_1 \cap \mathcal{B}_2$ a.s..

Thus, it suffices to show that for any compact set $\mathcal{M} \in \mathbb{R}^D$, almost surely, $f^\varepsilon(\mathcal{M})$ is in the set $\mathcal{B}_1$ from Proposition 4. The proof for $\mathcal{B}_2$ is identical.

We will show that $\mathrm{Card}(\arg\max_{z\in f^\varepsilon(\mathcal{M})} \pi^{(1)}(z)) = 1$, where $\pi^{(1)}(z)$ is the first coordinate of z. First, since $\mathcal{M}$ is compact and f^ε is continuous, $f^\varepsilon(\mathcal{M})$ is compact, therefore the supremum of $\pi^{(1)}$ is reached at least at one point.

Lemma 11. *The two following properties hold almost surely.*

1. *Let $F'_I, F'_J \in \mathcal{P}^\varepsilon$ be two different simplices, then at least one of the two following points holds:*
   - $\sup_{x\in f^\varepsilon(\mathcal{M})\cap relint(F'_I)} \pi^{(1)}(x) \neq \sup_{x\in f^\varepsilon(\mathcal{M})\cap relint(F'_J)} \pi^{(1)}(x)$
   - *$\pi^{(1)}$ does not reach its maximum on $f^\varepsilon(\mathcal{M}) \cap relint(F'_I)$ or does not reach its maximum on $f^\varepsilon(\mathcal{M}) \cap relint(F'_J)$.*
2. *Let $F'_I \in \mathcal{P}^\varepsilon$, then the supremum of $\pi^{(1)}$ on $f^\varepsilon(\mathcal{M}) \cap relint(F'_I)$ is reached at at most one point of $relint(F'_I)$.*

A consequence of this lemma is that almost surely, the maximizer of $\pi^{(1)}$ on $f^\varepsilon(\mathcal{M})$ is unique, as all maximizers of $\pi^{(1)}$ on $f^\varepsilon(\mathcal{M})$ belong to the relative interior of one simplex of $\mathcal{P}^\varepsilon$, which shows that $f^\varepsilon(\mathcal{M})$ is almost surely in $\mathcal{B}_1$.

Proof of Lemma 11. The following functions will be of use in the proof. For any finite $J \subset \mathbb{N}$ such that $F'_J = \{x_i + \varepsilon_i\}_{i\in J} \in \mathcal{P}^\varepsilon$, for any $j \in J$ and $\alpha \in (0,1]$, let

$$u_{\alpha,J} : e \in \mathbb{R} \longmapsto \sup \Bigg\{ \alpha(\pi^{(1)}(x_j) + e) + \sum_{k\in J\setminus\{j\}} \alpha_k \pi^{(1)}(x_k + \varepsilon_k), \text{ where}$$
$$z = \alpha x_j + \sum_{k\in J\setminus\{j\}} \alpha_k x_k \in \mathcal{M},\ \alpha_k \in (0,1] \text{ and } \alpha + \sum_k \alpha_k = 1 \Bigg\}.$$

In other words, u_{α_J} is the supremum of $\pi^{(1)}$ on the slice of $f^\varepsilon(\mathcal{M}) \cap \mathrm{relint}(F'_J)$ that gives weight α to the vertex $(x_j + \varepsilon_j)$. To simplify the notations, let $w_k : z \longmapsto \alpha_k$ be the "weight" functions. It is straightforward to check that

1. the function $u_{\alpha,J}$ is linear with slope α,

2. $\sup_{x\in f^\varepsilon(\mathcal{M})\cap \mathrm{relint}(F'_J)} \pi^{(1)}(x) = \sup_{\alpha\in(0,1]} u_{\alpha,J}(\pi^{(1)}(\varepsilon_j))$,

3. the function $h : \pi^{(1)}(\varepsilon_j) \longmapsto \sup_{x\in f^\varepsilon(\mathcal{M})\cap \mathrm{relint}(F'_J)} \pi^{(1)}(x)$ (all coordinates of all ε_k other than $\pi^{(1)}(\varepsilon_j)$ being fixed) is convex,

4. if the supremum of $\pi^{(1)}$ on the closure of $f^\varepsilon(\mathcal{M}) \cap \mathrm{relint}(F'_J)$ is reached at some point $z \in F'_J$ when $\pi^{(1)}(\varepsilon_j) = e$, then $w_j(z)$ is a sub-gradient of h at e,

5. since the number of points where the sub-gradient of a convex function on $\mathbb{R}$ is not unique is at most countable, almost surely (whether all coordinates of all ε_k other than $\pi^{(1)}(\varepsilon_j)$ are fixed or not), h has a unique sub-gradient at $\pi^{(1)}(\varepsilon_j)$.

Let us now prove the first point of the lemma. Let $F'_I = \{x_i+\varepsilon_i\}_{i\in I}$ and $F'_J = \{x_i+\varepsilon_i\}_{i\in J}$ be two different simplices of $\mathcal{P}^\varepsilon$, and let $j \in J \setminus I$ (by exchanging the two simplices, we may assume without loss of generality that J is not a subset of I).

Consider the following, conditionally to $(\varepsilon_n)_{n\neq j}$ and $\pi^{(2:D)}(\varepsilon_j)$. Assume that $h(\pi^{(1)}(\varepsilon_j)) = \sup_{x\in f^\varepsilon(\mathcal{M})\cap \mathrm{relint}(F'_I)} \pi^{(1)}(x)$ (otherwise we are in the first case of the first point of the lemma). We may assume without loss of generality (by point 5 above) that the sub-gradient of h at $\pi^{(1)}(\varepsilon_j)$ is unique. Two cases are possible:

- the sub-gradient of h at $\pi^{(1)}(\varepsilon_j)$ is 0. Then $\pi^{(1)}$ does not reach its maximum on $f^\varepsilon(\mathcal{M}) \cap \mathrm{relint}(F'_J)$, since if z is a maximizer of $\pi^{(1)}$, then $w_j(z) = 0$ by point 4,
- the sub-gradient of h at $\pi^{(1)}(\varepsilon_j)$ is positive, so there exists a single point e such that $h(e) = \sup_{x\in f^\varepsilon(\mathcal{M})\cap \mathrm{relint}(F'_I)} \pi^{(1)}(x)$. Since $\pi^{(1)}(\varepsilon_j)$ is uniform on $[-r, r]$ by construction, we almost surely have $\pi^{(1)}(\varepsilon_j) \neq e$, and thus this second case almost surely never happens.

For the second point of the lemma, by points 4 and 5, if the set of maximizers of $\pi^{(1)}$ on $f^\varepsilon(\mathcal{M}) \cap \mathrm{relint}(F'_J)$ is a non-empty set $\mathcal{Z}$, then for any $j \in J$, almost surely, w_j is constant on $\mathcal{Z}$. Since every point $z \in F'_J$ is characterized by the vector $(w_j(z))_{j\in J}$, this shows that $\mathcal{Z}$ contains a single point, which concludes the proof. □

6.7 Proof of Proposition 6

For any $\nu > 0$ and $h \in L^2([-\nu,\nu]^D)$ (resp. $L^\infty([-\nu,\nu]^D)$), write $\|h\|_{2,\nu}$ (resp. $\|h\|_{\infty,\nu}$) its L^2 (resp. L^∞) norm.

Let $\rho_0 \in [1, 2)$. Let us start with some preliminary results.

From [9], Section 7.1, for all $\nu > 0$, there exists $b > 0$, $\eta > 0$, $c_M > 0$ and $c_Z > 0$ such that, writing $\epsilon(u) = b/\log\log(1/u)$, the following properties hold for any $\rho' \in [1, \rho_0]$.

- For all $\phi \in \Upsilon_{\rho',S}$ and for all $h \in L^2([-\nu,\nu]^D)$ such that $\phi + h \in \Upsilon_{\rho',S}$ and $\|h\|_{2,\nu} \leqslant \eta$,
$$M(\phi + h; \nu|\phi) \geqslant c_\nu^4 \|h\|_{2,\nu}^{2+2\epsilon(\|h\|_{2,\nu})}. \tag{11}$$
- For all $n \geqslant 1$, writing $Z_n(t,\phi) = \sqrt{n}(\tilde{\phi}_n(t) - \phi(t)\Phi_{\varepsilon^{(1)}}(t_1)\Phi_{\varepsilon^{(2)}}(t_2))$, one has for all $\phi \in \Upsilon_{\rho',S}$ and $h \in L^2([-\nu,\nu]^D)$ such that $\phi + h \in \Upsilon_{\rho',S}$,
$$|M_n(\phi+h) - M(\phi+h;\nu_{\mathrm{est}}|\phi) - (M_n(\phi) - M(\phi;\nu_{\mathrm{est}}|\phi))| \leqslant c_M \frac{\|Z_n(\cdot,\phi)\|_{\infty,\nu_{\mathrm{est}}}}{\sqrt{n}} \|h\|_{2,\nu_{\mathrm{est}}}^{1-\epsilon(\|h\|_{2,\nu_{\mathrm{est}}})}. \tag{12}$$
- For all $x \in [1, n]$,
$$\mathbb{P}(\|Z_n(\cdot,\Phi_X)\|_{\infty,\nu_{\mathrm{est}}} \geqslant c_Z\sqrt{x}) \leqslant e^{-x}. \tag{13}$$

Moreover, from Lemma H.3 of [20], there exists a constant $c_T > 0$ such that for all $\rho' \in [1, \rho_0]$, $m \geqslant \rho' D$ and $\phi \in \Upsilon_{\rho',S}$,

$$\|\phi - T_m\phi\|_{\infty,\nu_{\text{est}}} \leqslant c_T (S\nu_{\text{est}})^m m^{-m/\rho'+D}.$$

Let $\rho' \in [1, \rho_0]$ and assume that $m \geqslant 2\rho' \frac{\log n}{\log\log n}$, then this equation becomes $\|\phi - T_m\phi\|_{\infty,\nu_{\text{est}}} = O(n^{-2+o_n(1)})$, where $o_n(1)$ denotes a sequence tending to 0 when n tends to infinity. In particular, there exists n_0 such that for all $n \geqslant n_0$,

$$\sup_{\rho' \in [1,\rho_0]} \sup_{\nu \in (0,\nu_{\text{est}}]} \sup_{m \geqslant 2\rho' \frac{\log n}{\log\log n}} \sup_{\phi \in \Upsilon_{\rho',S}} \|\phi - T_m\phi\|_{2,\nu} \leqslant \frac{1}{n} \tag{14}$$

and

$$\sup_{\rho' \in [1,\rho_0]} \sup_{m \geqslant 2\rho' \frac{\log n}{\log\log n}} \sup_{\phi \in \Upsilon_{\rho',S}} |M_n(\phi) - M_n(T_m\phi)| \leqslant c\|\phi - T_m\phi\|_{\infty,\nu_{\text{est}}} \leqslant \frac{1}{n}. \tag{15}$$

for some $c > 0$ that depends only on ν_{est}, ρ_0 and S, using that $\sup_{\phi \in \Upsilon_{\rho_0,S}} \|\phi\|_{\infty,\nu_{\text{est}}} < +\infty$. Finally, following the proof of equation (25) of Section A.3 of [19], for any $\nu' \geqslant \nu$, $m \geqslant 1$ and $\phi \in \mathbb{C}_m[X]$

$$\|\phi\|_{2,\nu'} \leqslant m^{D/2} (4\frac{\nu'}{\nu})^{m+D/2} \|\phi\|_{2,\nu}. \tag{16}$$

Let us now prove the proposition. Let $\rho \in [1, \rho_0]$ such that $\Phi_X \in \Upsilon_{\rho,S} \cap \mathcal{H}$. By definition, for any $m \geqslant 1$ and $\rho' \in [\rho, \rho_0]$, $\widehat{\Phi}_{n,m,\rho'}$ is such that $\widehat{\Phi}_{n,m,\rho'} \in \Upsilon_{\rho',S} \cap \mathcal{H}$ and

$$\begin{aligned} M_n(T_m\widehat{\Phi}_{n,m,\rho'}) &\leqslant \inf_{\phi \in \Upsilon_{\rho',S} \cap \mathcal{H}} M_n(T_m\phi) + \frac{1}{n} \\ &\leqslant \inf_{\phi \in \Upsilon_{\rho,S} \cap \mathcal{H}} M_n(T_m\phi) + \frac{1}{n} \\ &\leqslant M_n(T_m\Phi_X) + \frac{1}{n} \end{aligned}$$

and thus, by (15),

$$\sup_{\rho' \in [\rho,\rho_0]} \sup_{m \geqslant 2\rho' \frac{\log n}{\log\log n}} M_n(\widehat{\Phi}_{n,m,\rho'}) \leqslant M_n(\Phi_X) + \frac{3}{n}. \tag{17}$$

Therefore, by (12), for any $\nu \in (0, \nu_{\text{est}}]$, writing $h_{m,\rho'} = \widehat{\Phi}_{n,m,\rho'} - \Phi_X$,

$$\begin{aligned} M(\widehat{\Phi}_{n,m,\rho'}; \nu|\Phi_X) &\leqslant M(\widehat{\Phi}_{n,m,\rho'}; \nu_{\text{est}}|\Phi_X) \\ &\leqslant c_M \frac{\|Z_n(\cdot, \Phi_X)\|_{\infty,\nu_{\text{est}}}}{\sqrt{n}} \|h_{m,\rho'}\|_{2,\nu_{\text{est}}}^{1-\epsilon(\|h_{m,\rho'}\|_{2,\nu_{\text{est}}})} + \frac{3}{n}. \end{aligned} \tag{18}$$

Let us show that we may apply (11). Combining (17) with Lemma A.1 of [19] shows that for any $\delta > 0$, there exist $c_\eta > 0$ and n_0 (which do not depend on ρ) such that for all $n \geqslant n_0$, with probability at least $1 - 4e^{-c_\eta n}$,

$$\sup_{\rho' \in [\rho,\rho_0]} \sup_{m \geqslant 2\rho \frac{\log n}{\log\log n}} M(\widehat{\Phi}_{n,m,\rho'}; \nu_{\text{est}}|\Phi_X) \leqslant \delta.$$

In addition, since $\Upsilon_{\rho_0,S} \cap \mathcal{H}$ is compact in $L^2([-\nu_{\text{est}}, \nu_{\text{est}}]^D)$, $\phi \mapsto M(\phi; \nu_{\text{est}}|\Phi_X)$ is continuous on $L^2([-\nu_{\text{est}}, \nu_{\text{est}}]^D)$, and $M(\phi; \nu_{\text{est}}|\Phi_X) = 0$ implies $\phi = \Phi_X$ for all $\phi \in \mathcal{H} \cap \Upsilon_{\rho_0,S}$ by Theorem 1, there exists $\delta > 0$ such that

$$\inf_{\phi \in \Upsilon_{\rho_0,S} \cap \mathcal{H} \text{ s.t. } \|\phi - \Phi_X\|_{2,\nu_{\text{est}}} \geqslant \eta} M(\phi; \nu_{\text{est}}|\Phi_X) > \delta.$$

Therefore, there exist $c_\eta > 0$ and n_0 (which do not depend on ρ) such that for all $n \geqslant n_0$, with probability at least $1 - 4e^{-c_\eta n}$,

$$\sup_{\rho' \in [\rho, \rho_0]} \sup_{m \geqslant 2\rho \frac{\log n}{\log \log n}} \|h_{m,\rho'}\|_{2,\nu_{\text{est}}} \leqslant \eta, \tag{19}$$

which is what we need to apply (11).

Fix now $\nu \in (0, \nu_{\text{est}}]$, $c(\nu) > 0$ and $E > 0$ such that $\mathbb{Q} \in \mathcal{Q}^{(D)}(\nu, c(\nu), E)$. In particular, $c_\nu \geqslant c(\nu) > 0$. Then, by (11) and (18),

$$\|h_{m,\rho'}\|_{2,\nu}^{2+2\epsilon(\|h_{m,\rho'}\|_{2,\nu})} \leqslant \frac{2}{c(\nu)^4} \max\left(c_M \frac{\|Z_n(\cdot, \Phi_X)\|_{\infty,\nu_{\text{est}}}}{\sqrt{n}} \|h_{m,\rho'}\|_{2,\nu_{\text{est}}}^{1-\epsilon(\|h_{m,\rho'}\|_{2,\nu_{\text{est}}})}, \frac{3}{n} \right). \tag{20}$$

By (14) and (16), assuming $m \in [2\rho' \frac{\log n}{\log \log n}, C \frac{\log n}{\log \log n}]$ in the following series of inequalities (for some fixed $C > 2\rho'$),

$$\begin{aligned}
\|h_{m,\rho'}\|_{2,\nu_{\text{est}}} &\leqslant 2 \max(\|T_m h_{m,\rho'}\|_{2,\nu_{\text{est}}}, \frac{3}{n}) && \text{by (14)} \\
&\leqslant 2 \max\left(\|T_m h_{m,\rho'}\|_{2,\nu} m^{\frac{D}{2}} (4\frac{\nu_{\text{est}}}{\nu})^{m+\frac{D}{2}}, \frac{3}{n} \right) && \text{by (16)} \\
&\leqslant 4 \max\left(\|h_{m,\rho'}\|_{2,\nu} m^{\frac{D}{2}} (4\frac{\nu_{\text{est}}}{\nu})^{m+\frac{D}{2}}, \frac{3m^{\frac{D}{2}} (4\frac{\nu_{\text{est}}}{\nu})^{m+\frac{D}{2}}}{n} \right) && \text{by (14)} \\
&\leqslant n^{\epsilon(1/n)} \max\left(\|h_{m,\rho'}\|_{2,\nu}, \frac{1}{n} \right), && (21)
\end{aligned}$$

up to increasing the constant b in the definition of $u \mapsto \epsilon(u)$, which can be done without loss of generality. Together with (20) and (13), one gets for all $x \in [1, c_\eta n]$ (assuming $c_\eta \leqslant 1$ without loss of generality), with probability at least $1 - 4e^{-c_\eta n} - e^{-x} \geqslant 1 - 5e^{-x}$ (on the event where $\|Z(\cdot, \Phi_X)\|_{\infty,\nu_{\text{est}}} \leqslant c_Z \sqrt{x}$ and (19) holds) that for all $\rho' \in [\rho, \rho_0]$ and $m \in [2\rho' \frac{\log n}{\log \log n}, C \frac{\log n}{\log \log n}]$,

$$\|h_{m,\rho'}\|_{2,\nu}^{2+2\epsilon(\|h_{m,\rho'}\|_{2,\nu})} \leqslant c \max\left(\sqrt{\frac{x}{n}} n^{\epsilon(1/n)} \left(\|h_{m,\rho'}\|_{2,\nu} \vee \frac{1}{n} \right)^{1-\epsilon(\|h_{m,\rho'}\|_{2,\nu_{\text{est}}})}, \frac{1}{n} \right)$$

for some constant $c > 0$ that does not depend on ρ, ρ' or m. Since ϵ is increasing, recalling that $\|h_{m,\rho'}\|_{2,\nu_{\text{est}}} \leqslant \eta$ on the event considered, by (21),

$$\begin{aligned}
\epsilon(\|h_{m,\rho'}\|_{2,\nu_{\text{est}}}) &\leqslant \begin{cases} \max(\epsilon(\|h_{m,\rho'}\|_{2,\nu} n^{\epsilon(1/n)}), \epsilon(n^{-1+\epsilon(1/n)})) & \text{if } \|h_{m,\rho'}\|_{2,\nu} \leqslant n^{-2\epsilon(1/n)}, \\ \epsilon(\eta) & \text{always,} \end{cases} \\
&\leqslant \begin{cases} 2\epsilon(1/n) & \text{if } \|h_{m,\rho'}\|_{2,\nu} \leqslant n^{-2\epsilon(1/n)}, \\ \epsilon(\eta) & \text{always,} \end{cases} \\
&\leqslant [\epsilon(\eta) \textbf{ or } 2\epsilon(1/n)]
\end{aligned}$$

for n large enough (depending on b), up to decreasing η, where for compactness of notations, $[A \textbf{ or } B]$ means $\min(A, B)$ if $\|h_{m,\rho'}\|_{2,\nu} \leqslant n^{-2\epsilon(1/n)}$ and A otherwise in the following. Gathering the two previous equations shows that either

$$\|h_{m,\rho'}\|_{2,\nu}^{1+3[\epsilon(\eta) \textbf{ or } 2\epsilon(1/n)]} \leqslant c\sqrt{\frac{x}{n^{1-2\epsilon(1/n)}}}$$

or

$$\|h_{m,\rho'}\|_{2,\nu}^{2+2[\epsilon(\eta) \textbf{ or } 2\epsilon(1/n)]} \leqslant \frac{c}{n}.$$

Therefore, assuming $3\epsilon(\eta) \leqslant 1$ without loss of generality, $\|h_{m,\rho'}\|_{2,\nu} \leqslant n^{-\epsilon(1/n)}$ as soon as $x \leqslant n^{1-10\epsilon(1/n)}/c^2$ and thus, up to changing the constant c, for n large enough and for all

$x \in [1, n^{1-10\epsilon(1/n)}/c^2]$, with probability at least $1 - 4e^{-c_\eta n} - e^{-x}$, for all $\rho' \in [\rho, \rho_0]$ and $m \in [2\rho' \frac{\log n}{\log\log n}, C\frac{\log n}{\log\log n}]$,

$$\|h_{m,\rho'}\|_{2,\nu}^2 \leqslant c\left(\frac{x}{n^{1-2\epsilon(1/n)}}\right)^{1-6\epsilon(1/n)}.$$

Finally, note that $4e^{-c_\eta n}e^{n^{1-10\epsilon(1/n)}} \longrightarrow 0$, so that the probability that the last equation holds is larger than $1 - 2e^{-x}$ for n large enough, which concludes the proof for the version with $\widehat{\Phi}_{n,m,\rho'}$. The version for $T_m\widehat{\Phi}_{n,m,\rho'}$ follows from this and (14).

6.8 Proof of Lemma 1

Let $y \in \mathcal{M}_G \cap \mathcal{K}$. By property (III) of ψ_A,

$$\begin{aligned}\bar{g}(y) &= \frac{1}{h^D}\int \psi_A\left(\frac{\|y-u\|}{h}\right)dG(u)\\ &\geqslant \frac{1}{h^D}\int_{\|u-y\|_2\leqslant c_A h}\psi_A(\frac{\|y-u\|}{h})dG(u)\\ &\geqslant \frac{1}{h^D}d_A G^\star(B(y, c_A h))\\ &\geqslant \frac{1}{h^D}d_A a(c_A h)^d.\end{aligned}$$

6.9 Proof of Lemma 2

Recall the definition of $\bar{g}$: for all $y \in \mathbb{R}^D$,

$$\bar{g}(y) = \frac{1}{h^D}\int \psi_A(\frac{\|y-u\|}{h})dG(u).$$

Let $C_1 > 0$ and $\epsilon > 0$. By Property (V) of ψ_A and (4), there exists $T > 0$ (depending on A and C_1) such that for any $t \geqslant T$, $\psi_A(t) \leqslant C_1\exp(-\beta_A t^{A/(A+1)})$. Take $y \in \mathbb{R}^D$ such that $d(y, \mathcal{M}_G) > (\frac{\varepsilon}{\beta_A})^{\frac{A+1}{A}} h\log(\frac{1}{h})^{\frac{A+1}{A}}$, then for all $u \in \mathcal{M}_G$, $\frac{\|y-u\|}{h} \geqslant (\beta_A^{-1}\log(\frac{1}{h^\varepsilon}))^{\frac{A+1}{A}}$, therefore there exists $h_0 > 0$ depending only on ε, D, A and T (thus C_1 such that $h \leqslant h_0$ implies $\frac{\|y-u\|}{h} \geqslant T$ and thus

$$\begin{aligned}\psi_A(\frac{\|y-u\|}{h}) &\leqslant C_1\exp\{-\beta_A(\frac{\|y-u\|}{h})^{A/(A+1)}\}\\ &\leqslant C_1\exp\{-\log(\frac{1}{h^\varepsilon})\}\\ &= C_1 h^\varepsilon,\end{aligned}$$

and finally $\bar{g}(y) \leqslant C_1(\frac{1}{h})^{D-\varepsilon}$ since G is a probability distribution. Lemma 2 follows by taking $\varepsilon = D$.

6.10 Proof of Lemma 3

For $y \in \mathbb{R}^D$,

$$\widehat{g}_{n,\kappa}(y) - \bar{g}(y) = (\frac{1}{2\pi})^D\int e^{-it^\top y}\mathcal{F}[\psi_A](th)(T_{m_\kappa}\widehat{\Phi}_{n,1/\kappa}(t) - \Phi_X(t))dt.$$

Since $\mathcal{F}[\psi_A](th)$ is 0 for $\|t\|_2 > 1/h$,

$$\begin{aligned}\widehat{g}_{n,\kappa}(y) - \bar{g}(y) &= (\frac{1}{2\pi})^D\int e^{-it^\top y}\mathcal{F}[\psi_A](th)(T_{m_\kappa}\widehat{\Phi}_{n,1/\kappa}(t) - \Phi_X(t))1|_{\|t\|_2\leqslant 1/h}dt\\ &= \mathcal{F}^{-1}[\mathcal{F}[\psi_h]\{(T_{m_\kappa}\widehat{\Phi}_{n,1/\kappa} - \Phi_X)1|_{\|t\|_2\leqslant 1/h}\}](y)\\ &= \mathcal{F}^{-1}[\mathcal{F}[\psi_{A,h}]] * \mathcal{F}^{-1}[(T_{m_\kappa}\widehat{\Phi}_{n,1/\kappa} - \Phi_X)1|_{\|t\|_2\leqslant 1/h}](y). \qquad (22)\end{aligned}$$

By Young's convolution inequality,

$$\|\widehat{g}_{n,\kappa} - \bar{g}\|_\infty \leqslant \|\mathcal{F}^{-1}[\mathcal{F}[\psi_{A,h}]]\|_2 \|\mathcal{F}^{-1}[(T_{m_\kappa}\widehat{\Phi}_{n,1/\kappa} - \Phi_X)1|_{\|t\|_2 \leqslant 1/h}]\|_2.$$

Finally, using Parseval's equality and the fact that $\mathcal{F}^{-1}[\mathcal{F}[\psi_{A,h}]] = \psi_{A,h}$,

$$\|\widehat{g}_{n,\kappa} - \bar{g}\|_\infty \leqslant \|\psi_{A,h}\|_2 \|T_{m_\kappa}\widehat{\Phi}_{n,1/\kappa} - \Phi_X\|_{2,1/h},$$

and use (5) to conclude the proof.

6.11 Proof of Theorem 4

Let $\kappa_0 \in (1/2, 1]$, $\nu \in (0, \nu_{\text{est}}]$, $c(\nu) > 0$, $E > 0$, $S > 0$ and $C > 0$. Let $\kappa \in [\kappa_0, 1]$, $\mathbb{Q} \in \mathcal{Q}^{(D)}(\nu, c(\nu), E)$ and $G \in St_{\mathcal{K}}(a, d, r_0) \cap \mathcal{L}(\kappa, S, \mathcal{H})$.

Using inequalities analogous to (28)-(29) p.17 of [19], we get that for all $\kappa' \in [\kappa_0, \kappa]$ and all integer m,

$$\|T_m\widehat{\Phi}_{n,1/\kappa'} - \Phi_X\|_{2,1/h}^2 \leqslant 4U(h) + 4m^D(2 + 2\frac{1}{h\nu})^{2m+D}\left(2V(\nu) + \|\widehat{\Phi}_{n,1/\kappa'} - \Phi_X\|_{2,\nu}^2\right), \quad (23)$$

where

$$U(h) = ch^{-D-2m-2/\kappa'}S^{2m}m^{-2\kappa' m+2D}\exp(2\kappa'(S/h)^{1/\kappa'})$$
$$\text{and} \quad V(\nu) = c(S\nu)^{2m+2/\kappa'}m^{-2\kappa' m+2D}.$$

Thus, applying Lemma 3 and using $h = c_h S m_{\kappa'}^{-\kappa'}$, there exists $C > 0$ such that on the event where (23) holds:

$$\Gamma_{n,\kappa'}^2 \leqslant C(c_h S)^{-2D-2m_{\kappa'}-2/\kappa'} m_{\kappa'}^{2D(\kappa'+1)+2} S^{2m_{\kappa'}} \exp(2\kappa' c_h^{-1/\kappa'} m_{\kappa'})$$
$$+ C m_{\kappa'}^{D(1+\kappa')}(2 + 2\frac{m_{\kappa'}^{\kappa'}}{c_h S\nu})^{2m_{\kappa'}+D}\left((S\nu)^{2m_{\kappa'}+2/\kappa'} m_{\kappa'}^{-2\kappa' m_{\kappa'}+2D} + \|\widehat{\Phi}_{n,1/\kappa'} - \Phi_X\|_{2,\nu}^2\right). \quad (24)$$

The first term of the upper bound is upper bounded as follows.

$$(c_h S)^{-2D-2m_{\kappa'}-2/\kappa'} m_{\kappa'}^{2D(\kappa'+1)+2} S^{2m_{\kappa'}} \exp(2\kappa' c_h^{-1/\kappa'} m_{\kappa'})$$
$$= S^{-2D-2/\kappa'} \exp\{(-2D - 2m_{\kappa'} - 2/\kappa')\log(c_h) + (2D(\kappa'+1) + 2)\log(m_{\kappa'}) + 2\kappa' c_h^{-1/\kappa'} m_{\kappa'}\}$$
$$\leqslant C \exp\{(-2\log(c_h) + 1)m_{\kappa'} + 2(D(\kappa'+1) + 1)\log(m_{\kappa'})\} \quad (25)$$
$$\leqslant C \exp\{(-2\log(c_h) + 3 + 2D(\kappa'+1))m_{\kappa'}\}, \quad (26)$$

for another constant $C > 0$, where inequality (25) holds because $2\kappa c_h^{1/\kappa} > 1$ and inequality (26) holds because $\log(m_\kappa) \leqslant m_\kappa$. The second term of the upper bound is upper bounded by

$$m_{\kappa'}^{D(1+\kappa')}(2 + 2\frac{m_{\kappa'}^{\kappa'}}{c_h S\nu})^{2m_{\kappa'}+D}\left((S\nu)^{2m_{\kappa'}+2/\kappa'} m_{\kappa'}^{-2\kappa' m_{\kappa'}+2D} + \|\widehat{\Phi}_{n,1/\kappa'} - \Phi_X\|_{2,\nu}^2\right)$$
$$\leqslant C\ m_{\kappa'}^{D(1+2\kappa')}(2\kappa' m_{\kappa'})^{2\kappa' m_{\kappa'}}(2\kappa')^{-2\kappa' m_{\kappa'}}(c_h S\nu)^{-2m_{\kappa'}-D}$$
$$\times\left((S\nu)^{2m_{\kappa'}+2/\kappa'} m_{\kappa'}^{-2\kappa' m_{\kappa'}+2D} + \|\widehat{\Phi}_{n,1/\kappa'} - \Phi_X\|_{2,\nu}^2\right)$$
$$\leqslant C\Bigg(\exp\left\{(-2\log(c_h) + (3D + 2\kappa'))m_{\kappa'}\right\}$$
$$+ (2\kappa' m_{\kappa'})^{2\kappa' m_{\kappa'}} \exp\left\{(-2\log(c_h) + D(1 + 2\kappa'))m_{\kappa'}\right\}\|\widehat{\Phi}_{n,1/\kappa'} - \Phi_X\|_{2,\nu}^2\Bigg)$$

for another constant $C > 0$. Putting all together, we get that for yet another constant $C > 0$,

$$\Gamma^2_{n,\kappa'} \leqslant C \max\Bigg(\exp\left\{(-2\log(c_h) + 3 + 2D(\kappa'+1))m_{\kappa'}\right\}, \exp\left\{(-2\log(c_h) + (3D+2\kappa'))m_{\kappa'}\right\}, (2\kappa' m_{\kappa'})^{2\kappa' m_{\kappa'}} \exp\left\{(-2\log(c_h) + D(1+2\kappa'))m_{\kappa'}\right\} \|\widehat{\Phi}_{n,1/\kappa'} - \Phi_X\|^2_{2,\nu}\Bigg).$$

Choosing $c_h \geqslant \exp\{2D+2\}$ and $m_{\kappa'} = \frac{1}{4\kappa'}\frac{\log n}{\log\log n}$ for some $\gamma \in (0,1)$, it follows that

$$\Gamma^2_{n,\kappa'} \leqslant Ce^{-m_{\kappa'}}\left[1 \vee n^{1/2}\|\widehat{\Phi}_{n,1/\kappa'} - \Phi_X\|^2_{2,\nu}\right]. \tag{27}$$

By Proposition 6, taking $x = \log n$ and δ, δ'' such that $(1-\delta)(1-\delta'') > 1/2$, we obtain that with probability at least $1 - 2/n$, for all $\kappa' \leqslant \kappa$, $\Gamma^2_{n,\kappa'} \leqslant Ce^{-m_{\kappa'}} \longrightarrow 0$. Note that we could also take $x = n^{1/2-\delta'''}$ for any $\delta''' > 0$ and still have $\Gamma^2_{n,\kappa'} \leqslant Ce^{-m_{\kappa'}}$ with probability at least $1 - 2e^{-x}$, up to changing the constant C, by picking δ and δ'' small enough in Proposition 6.

Now, by Lemma 1, for any $h \leqslant (r_0/c_A) \wedge 1$,

$$\begin{aligned}\inf_{y\in\mathcal{M}_G\cap\mathcal{K}} \widehat{g}_{n,\kappa'}(y) &\geqslant \inf_{y\in\mathcal{M}_G\cap\mathcal{K}} \bar{g}(y) - \Gamma_{n,\kappa'} \\ &\geqslant c_A^d d_A a\left(\frac{1}{h}\right)^{D-d} - \Gamma_{n,\kappa'} \\ &\geqslant \frac{c_A^d d_A a}{2}\left(\frac{1}{h}\right)^{D-d}\end{aligned}$$

as soon as $\Gamma_{n,\kappa'} \leqslant \frac{c_A^d d_A a}{2}$, and this lower bound is strictly larger than $\lambda_{n,\kappa}$ for any d. This implies that on the event where $\Gamma_{n,\kappa'} \leqslant \frac{c_A^d d_A a}{2}$, $\mathcal{M}_G \cap \mathcal{K} \subset \widehat{\mathcal{M}}_{\kappa'} \cap \mathcal{K}$. Next,

$$\sup_{y\in\mathcal{K}, d(y,\mathcal{M}_G)\geqslant h\left[\frac{D}{\beta_A}\log\left(\frac{1}{h}\right)\right]^{\frac{A+1}{A}}} \widehat{g}_{n,\kappa'}(y) \leqslant \sup_{y\in\mathcal{K}, d(y,\mathcal{M}_G)\geqslant h\left[\frac{D}{\beta_A}\log\left(\frac{1}{h}\right)\right]^{\frac{A+1}{A}}} \bar{g}(y) + \Gamma_{n,\kappa'}.$$

Choosing $C_1 = \frac{c_A^d d_A a}{16}$ and applying Lemma 2 we get that, on the event where $\Gamma_{n,\kappa'} \leqslant \frac{c_A^d d_A a}{16}$,

$$\sup_{y\in\mathcal{K}, d(y,\mathcal{M}_G)\geqslant h\left[\frac{D}{\beta_A}\log\left(\frac{1}{h}\right)\right]^{\frac{A+1}{A}}} \widehat{g}(y) \leqslant 2C_1$$

for n large enough, and this upper bound is strictly less than $\lambda_{n,\kappa'}$ for any d. This implies that

$$\left\{y : y \in \mathcal{K}, d(y,\mathcal{M}_G) > h\left[\frac{D}{\beta_A}\log\left(\frac{1}{h}\right)\right]^{\frac{A+1}{A}}\right\} \cap \widehat{\mathcal{M}}_{\kappa'} = \varnothing.$$

We may now take h as in the statement of the Theorem. As a result, we have proved that: for all $\kappa_0 \in (1/2, 1]$, $S > 0$, $a > 0$ $d \leqslant D$, $\nu \in (0, \nu_{\text{est}}]$, $c(\nu) > 0$ and $E > 0$, there exists $c' > 0$ and n_0 such that for all $n \geqslant n_0$, for all $\kappa \in [\kappa_0, 1]$, $G \in St_{\mathcal{K}}(a, d, r_0) \cap \mathcal{L}(\kappa, S, \mathcal{H})$ and $\mathbb{Q} \in \mathcal{Q}^{(d)}(\nu, c(\nu), E)$, with $(G * \mathbb{Q})^{\otimes n}$-probability at least $1 - \frac{2}{n}$,

$$\sup_{\kappa'\in[\kappa_0,\kappa]} \frac{\log(n)^{\kappa'}}{\log(\log(n))^{\kappa' + \frac{A+1}{A}}} H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}}_{\kappa'}) \leqslant c'. \tag{28}$$

Using the fact that $H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}}_\kappa)$ is uniformly upper bounded, the theorem follows.

6.12 Proof of Lemma 4

Case $\kappa \neq 1$. This case is based on [26]. In the following, we will note constants that can change with upper case A, B and C. In [26], Theorem 2, the author defines for any positive constants $\mu > 0$, $q > 1$ and $a > 0$ a function $\zeta_{q,\mu,a}$, such that for $x \in \mathbb{R}$,

$$\zeta_{q,\mu,a}(x) = -i \int_{\mathcal{C}} z^{\mu} \exp(z^q - qax^2 z) dz,$$

where $\mathcal{C}$ is a curve in the complex plane so that the maximum of $|z^{\mu} \exp(z^q - qax^2 z)|$ for $z \in \mathcal{C}$ is attained on the positive real line. The author shows that $\zeta_{q,\mu,a}$ and $\zeta^2_{q,\mu,a}$ are integrable functions.

The author uses the saddle-point integration method to show that there exist $A > 0$ and $B > 0$ which depend on q, μ and a such that

$$|\mathcal{F}[\zeta_{q,\mu,a}](t)| \leqslant A \exp(-Bx^{\frac{2q}{q+1}}). \tag{29}$$

Finally, for $\kappa \in (1/2, 1)$, fix $\mu > 0$, $a > 0$, and define

$$f_\kappa = c_{f_\kappa} \mathrm{Re}[\zeta_{\frac{1}{2\kappa-1},\mu,a}]^2 * u_1,$$

where $u_1 : x \in \mathbb{R} \mapsto \exp(-\frac{1}{1-4x^2})1|_{(-1/2,1/2)}(x)$ and c_{f_κ} is a constant that ensures that f_κ is a density.

Let us first prove that there exist $A > 0$ and $B > 0$ positive constants such that $|\mathcal{F}[\mathrm{Re}[\zeta_{q,\mu,a}]^2](t)| \leqslant A \exp(-B|t|^{1/\kappa})$.

$$\begin{aligned}
|\mathcal{F}[\mathrm{Re}[\zeta_{\frac{1}{2\kappa-1},\mu,a}]^2](t)| &= |\mathcal{F}[\mathrm{Re}[\zeta_{\frac{1}{2\kappa-1},\mu,a}]] * \mathcal{F}[\mathrm{Re}[\zeta_{\frac{1}{2\kappa-1},\mu,a}]](t)| \\
&\leqslant A \int_{\mathbb{R}} \exp(-B|x-y|^{1/\kappa} - B|y|^{1/\kappa}) dy \\
&= \int_{|y-x| \geqslant |x|/2} \exp(-B|x-y|^{1/\kappa} - B|y|^{1/\kappa}) dy \\
&\quad + \int_{|y-x| < |x|/2} \exp(-B|x-y|^{1/\kappa} - B|y|^{1/\kappa}) dy \\
&\leqslant A \exp(-B|x|^{1/\kappa}).
\end{aligned} \tag{30}$$

Finally, for all $t \in \mathbb{R}$, using that $|\mathcal{F}[u](t)| \leqslant \|u\|_{1,1}$,

$$\begin{aligned}
|\mathcal{F}[f_\kappa](t)| &= |\mathcal{F}[\mathrm{Re}[\zeta_{\frac{1}{2\kappa-1},\mu,a}]^2](t)| \; |\mathcal{F}[u_1](t)| \\
&\leqslant A \exp(-B|x|^{\frac{1}{\kappa}}).
\end{aligned}$$

For $x \in \mathbb{R}$, $\mathcal{F}[f_\kappa(x)]' = \mathcal{F}[x \mapsto x f_\kappa(x)]$ and

$$x f_\kappa(x) = c_{f_\kappa} v * \mathrm{Re}[\zeta_{\frac{1}{2\kappa-1},\mu,a}]^2(x) + c_{f_\kappa} u_1 * \tilde{\zeta}(x),$$

where $v : x \in \mathbb{R} \mapsto x u_1(x)$ and $\tilde{\zeta} : x \in \mathbb{R} \mapsto x \mathrm{Re}[\zeta_{\frac{1}{2\kappa-1},\mu,a}]^2(x)$.

Following the same proof as Theorem 2 of [26], there exists $A > 0$ and $B > 0$ such that for all $t \in \mathbb{R}$, $|\mathcal{F}[x \mapsto x\mathrm{Re}[\zeta_{\frac{1}{2\kappa-1},\mu,a}](t)]| \leqslant A \exp(-B|t|^{1/\kappa})$, so that, following the proof of (30), $|\mathcal{F}[\tilde{\zeta}]|(t) \leqslant A \exp(-B|t|^{1/\kappa})$. Hence, there exists $A > 0$ and $B > 0$ such that $|\mathcal{F}[f_\kappa]'(t)| \leqslant A \exp(-B|t|^{1/\kappa})$.

Finally, note that f_κ is continuous as a convolution of an integrable function with a smooth function, and that for all $x \in \mathbb{R}$, $f_\kappa(x) > 0$ since $\mathrm{Re}[\zeta_{\frac{1}{2\kappa-1},\mu,a}]$ and u are not the null function almost everywhere.

Case $\kappa = 1$. Let $\delta \in (0,1)$ and define $f_1 : x \in \mathbb{R} \mapsto c_{f_1}(u_{\frac{1}{1-\delta}} * u_{\frac{1}{1-\delta}})(x)$, where c_{f_1} is a constant that ensures that f_1 is a probability density.

There exist $A > 0$ and $B > 0$ such that $\mathcal{F}[f_1](x) \leqslant A\exp(-B|x|^\delta)$, see Lemma in [30]. Moreover, $\mathcal{F}[f_1]'(x) = 2c_{f_1}\mathcal{F}[u_{\frac{1}{1-\delta}}](x)\mathcal{F}[u_{\frac{1}{1-\delta}}]'(x) \leqslant A\|x \mapsto xu_{\frac{1}{1-\delta}}(x)\|_1 \exp(-B|x|^\delta)$.

Finally, note that f_1 is continuous and does not vanish on its support.

6.13 Proof of Lemma 5

First, by Lemma 4, for any $\kappa \in (1/2, 1]$, $U(\kappa)$ satisfies $A(1/\kappa)$.

Let $i \in \{0,1\}$. For any $\lambda = (\lambda_1, \dots, \lambda_D) \in \mathbb{R}^D$,

$$\begin{aligned}\mathbb{E}[\exp(\lambda \top X_i(\kappa))] &= \mathbb{E}\left[\exp\left(\left(\alpha\lambda_1 + \frac{\alpha}{2}\lambda_2\right)U(\kappa) + (-1)^i\gamma\lambda_2\frac{\alpha}{2}\cos\left(\frac{U(\kappa)}{\gamma}\right)\right)\right] \\ &\leqslant e^{\gamma\frac{\alpha}{2}|\lambda_2|}\mathbb{E}\left[\exp\left(\left(\alpha\lambda_1 + \frac{\alpha}{2}\lambda_2\right)U(\kappa)\right)\right].\end{aligned} \tag{31}$$

Since $U(\kappa)$ satisfies $A(1/\kappa)$, there exist positive constants $A > 0$ and $B > 0$ such that for all $\lambda = (\lambda_1, \dots, \lambda_D) \in \mathbb{R}^D$, $\mathbb{E}[\exp(\lambda \top X_i(\kappa))] \leqslant A\exp(B|\lambda|^{\frac{1}{\kappa}})$. Applying this in (31),

$$\begin{aligned}\mathbb{E}[\exp(\lambda \top X_i(\kappa))] &\leqslant A\exp\left(\gamma\frac{\alpha}{2}|\lambda_2| + B\left|\left(\alpha\lambda_1 + \frac{\alpha}{2}\lambda_2\right)\right|^{\frac{1}{\kappa}}\right) \\ &\leqslant A'\exp(B'|\lambda|^{\frac{1}{\kappa}})\end{aligned}$$

for some other constants A' and B' since $1 \leqslant \kappa$, so that $X_i(\kappa)$ satisfies $A(1/\kappa)$.

6.14 Proof of Lemma 6

The proof is done in five steps.

1. We show that γg_γ is 1-lipschitz.
2. For $i \in \{0,1\}$ and $\kappa \in (\frac{1}{2}, 1]$, we compute the density p_i of $T_i(\kappa)$ with respect to the 1-dimensional Hausdorff measure μ_H and we show that for any compact set $\mathcal{K}$, there exists $b(\kappa, \mathcal{K}) > 0$ such that, for all $u \in M_i(\gamma) \cap \mathcal{K}$, $|p_i(u)| \geqslant b(\kappa, \mathcal{K})$.
3. We show that for $i \in \{0,1\}$, $\mu_H(\cdot \cap M_i(\gamma))$ is in $St_{\mathcal{K}}(2, d, r_0)$.
4. We deduce that for $i \in \{0,1\}$ and $d \geqslant 1$, T_i is in $St_{\mathcal{K}}(2b(\kappa, \mathcal{K}), d, r_0)$.
5. Finally, we show that for $i \in \{0,1\}$, $d \geqslant 1$ and a small enough, $G_i(\kappa) \in St_{\mathcal{K}}(a, d, r_0)$.

Proof of 1 For all $x \in \mathbb{R}$, $|\gamma\tilde{g}'_\gamma(x)| = |\sin(\frac{x}{\gamma})| \leqslant 1$, which implies that γg_γ is 1-Lipschitz.

Proof of 2 Let us first compute the density p_i of $T_i(\kappa)$ with respect to μ_H. For $i \in \{0,1\}$, denote $\zeta_i : x \in \mathbb{R} \mapsto (x, (-1)^i\gamma g_\gamma(x))$. Let $\mathcal{B}$ be an open subset of $\mathbb{R}^D$. For any $\kappa \in (\frac{1}{2}, 1]$,

$$T_i(\kappa)(\mathcal{B}) = \mathbb{P}[\zeta_i(U(\kappa)) \in \mathcal{B}] = \mathbb{P}[U(\kappa) \in \zeta_i^{-1}(\mathcal{B})] = \int_{\zeta_i^{-1}(\mathcal{B})} f_\kappa(u)du.$$

Let $J\zeta_i : u \in \mathbb{R} \mapsto \sqrt{1 + \gamma^2\tilde{g}_\gamma(u)^2}$ be the Jacobian of ζ_i. By the Area Formula (see equation (2.47) in [4]),

$$s_i(\kappa)(\mathcal{B}) = \int_{\zeta_i^{-1}(\mathcal{B})} \frac{f_\kappa(u)}{J\zeta_i(u)}J\zeta_i(u)du = \int_{\mathcal{B}\cap M_i(\gamma)} \frac{f_\kappa(\pi^{(1)}(u))}{J\zeta_i(\pi^{(1)}(u))}d\mu_H(u).$$

We then have that for all $x \in \mathbb{R}^D$,

$$p_i(x) = \frac{f_\kappa(\pi^{(1)}(x))}{J\zeta_i(\pi^{(1)}(x))}1|_{M_i(\gamma)}(x).$$

Since f_κ is continuous and does not vanish on its support, for any compact set $\mathcal{K}$, $M_i(\gamma)\cap\mathcal{K}$ is a compact subset of the support of f_κ. Thus, since f_κ is continuous and does not vanish on its support, for any compact set $\mathcal{K}$, there exists $c(\kappa,\mathcal{K})>0$ such that for all $u\in M_i(\gamma)\cap\mathcal{K}$, $f_\kappa(u)\geqslant c(\kappa,\mathcal{K})$. Moreover, for $i\in\{0,1\}$, $J\zeta_i(u)\leqslant\sqrt{2}$. Therefore, for all $x\in M_i(\gamma)\cap\mathcal{K}$, $|p_i(x)|\geqslant\frac{c(\kappa,\mathcal{K})}{\sqrt{2}}$.

Proof of 3 Recall that the 1-dimensional Hausdorff measure μ_H is defined as the limit $\lim_{\eta\to 0}\mu_H^\eta$, where for any set Z

$$\mu_H^\eta(Z)=\inf\left\{\sum_{i\in\mathbb{N}}\mathrm{Diam}(A_i):X\subset\bigcup_i A_i \text{ and } \forall i,\mathrm{Diam}(A_i)\leqslant\eta\right\}.$$

For any $z\in M_i(\gamma)$, there exists $x_0\in\mathbb{R}$ such that $z=(x_0,(-1)^i\gamma g_\gamma(x_0))$ and, for any $r>0$,

$$B(z,r)\cap M_i(\gamma)\supset\{(x,(-1)^i\gamma g_\gamma(x)),x\in B(x_0,r)\}$$

since $|x-x_0|\leqslant r$ implies $\|\gamma g_\gamma(x)-\gamma g_\gamma(x_0)\|_\infty\leqslant r$.

Let $(A_i)_{i\in\mathbb{N}}$ be a covering of $\{(x,(-1)^i\gamma g_\gamma(x)),x\in B(x_0,r)\}$, and $B_i=\pi^{(1)}(A_i)$, then B_i is a covering of $B(x_0,r)$. For all $\eta>0$,

$$\mu_H^\eta(\{(x,(-1)^i\gamma g_\gamma(x)),x\in B(x_0,r)\})\geqslant\mu_H^\eta(B(x_0,r)),$$

thus $\mu_H(B(z,r)\cap M_i(\gamma))\geqslant\mu_H(B(x_0,r))=2r$. If $r_0\leqslant 1$, then for any $r\leqslant r_0$,

$$\mu_H(B(z,r)\cap M_i(\gamma))\geqslant 2r^d,$$

which proves 3.

Proof of 4 Let $x_i\in M_i(\gamma)\cap\mathcal{K}$ and $r_0<1$. Then for all $r\leqslant r_0$,

$$T_i(B(x_i,r)\cap M_i(\gamma))=\int_{B(x_i,r)\cap M_i(\gamma)}p_i(u)d\mu_H(u)\geqslant b(\kappa,\mathcal{K})\mu_H(B(x_i,r)\cap M_i(\gamma))\geqslant 2b(\kappa,\mathcal{K})r^d.$$

Proof of 5 For $i\in\{0,1\}$, let $x_i\in A_\alpha M_i(\gamma)\cap\mathcal{K}$, $r_0<1$, and take $\tilde{\mathcal{K}}$ such that $A_\alpha^{-1}\mathcal{K}\subset\tilde{\mathcal{K}}$. For all $r\leqslant r_0$,

$$\begin{aligned}G_i(\kappa)(B(x_i,r))=\mathbb{P}[A_\alpha S_i(\kappa)\in B(x_i,r)]&\geqslant\mathbb{P}\left[S_i(\kappa)\in B\left(A_\alpha^{-1}x_i,\frac{r}{\|A_\alpha\|_{\mathrm{op}}}\right)\right]\\&\geqslant\frac{2b(\kappa,A_\alpha^{-1}\mathcal{K})}{\|A_\alpha\|_{\mathrm{op}}}r\\&\geqslant\frac{2b(\kappa,\tilde{\mathcal{K}})}{\alpha\|A_1\|_{\mathrm{op}}}r,\end{aligned}$$

so that for some a_0 depending on α and all $a\leqslant a_0$, $G_i(\kappa)(B(x_i,r))\geqslant ar^d$.

6.15 Proof of Lemma 8

Let $(\phi_n)_n$ be a sequence in $\mathcal{H}^\star(\kappa,S,(c_\Delta,A_\Delta^{(1)},B_\Delta^{(1)},A_\Delta^{(2)},B_\Delta^{(2)})_{\Delta>0})$ and $\phi^\star\in L_2([-\nu,\nu]^D)$ such that $\|\phi_n-\phi^\star\|_{2,\nu}\to 0$. For each n, there exists a random variable X_n such that $\phi=\Phi_{X_n}$. Without loss of generality, we can assume that ϕ_n converges almost everywhere to $\phi^\star$ on $[-\nu,\nu]^D$. Since $\Upsilon_{\kappa,S}$ is closed in $L_2([-\nu,\nu]^D)$, $\phi^\star\in\Upsilon_{\kappa,S}$. Let us show that $\phi^\star$ is the characteristic function of some random variable $X^\star$. Let $N\geqslant 1$, $(t_k)_{1\leqslant k\leqslant N}\subset\mathbb{R}^d$ and $(\lambda_k)_{1\leqslant k\leqslant N}\subset\mathbb{C}^D$, then

$$\sum_{k,l=1}^N\phi^\star(t_k-t_l)\lambda_k\bar{\lambda}_l=\lim_{n\to\infty}\sum_{k,l=1}^N\phi_n(t_k-t_l)\lambda_k\bar{\lambda}_l\geqslant 0.$$

Since $\phi^\star$ is continuous and $\phi^\star(0) = 1$, according to Bochner's Theorem, there exists $X^\star$ such that $\phi^\star = \Phi_{X^\star}$. Applying the Identity Theorem component-wise shows that for every $t \in \mathbb{R}^D$, $\phi_n(t) \to \Phi_{X^\star}(t)$, so that X_n converges in distribution to $X^\star$. Now, we have to show that $\Phi_{X^\star} \in \mathcal{H}^\star(\kappa, S, M, (c_\Delta, A_\Delta^{(1)}, B_\Delta^{(1)}, A_\Delta^{(2)}, B_\Delta^{(2)})_{\Delta>0})$.

Convergence in distribution of random vectors implies convergence in distribution of each coordinate, thus all the coordinates of $X^{\star,(2)}$ are null except the first one. Moreover, by Theorem 1 of [29], $X^\star$ satisfies (v). Let us prove that $X^\star$ satisfies (iii) using the Portmanteau Theorem. Since $A_\Delta^{(1)}$ is closed,

$$\mathbb{P}[(X^\star)^{(1)} \in A_\Delta^{(1)}] \geqslant \limsup \mathbb{P}[X_n^{(1)} \in A_\Delta^{(1)}] \geqslant c_\Delta,$$

and

$$\begin{aligned}
\mathbb{P}[(X^\star)^{(2)} \in B_\Delta^{(2)} | (X^\star)^{(1)} \in A_\Delta^{(1)}] &= \frac{\mathbb{P}[X^\star \in A_\Delta^{(1)} \times B^{(2)}]}{\mathbb{P}[(X^\star)^{(1)} \in A_\Delta^{(1)}]} \\
&\geqslant \frac{\limsup \mathbb{P}[X_n \in A_\Delta^{(1)} \times B_\Delta^{(2)}]}{\lim \mathbb{P}[(X_n)^{(1)} \in A_\Delta^{(1)}]} \text{ since } \mathbb{P}[(X^\star)^{(1)} \in \partial A_\Delta^{(1)}] = 0 \\
&\geqslant \limsup \frac{\mathbb{P}[X_n \in A_\Delta^{(1)} \times B_\Delta^{(2)}]}{\mathbb{P}[(X_n)^{(1)} \in A_\Delta^{(1)}]} = 1.
\end{aligned}$$

Let us prove that $\Phi_{X^\star}$ satisfies (iv). Since $A_\Delta^{(2)}$ is closed,

$$\mathbb{P}[(X^\star)^{(2)} \in A_\Delta^{(2)}] \geqslant \limsup \mathbb{P}[X_n^{(2)} \in A_\Delta^{(2)}] \geqslant c_\Delta.$$

Moreover,

$$\begin{aligned}
\mathbb{P}[(X^\star)^{(1)} \in B_\Delta^{(1)} | (X^\star)^{(2)} \in A_\Delta^{(2)}] &= \frac{\mathbb{P}[X^\star \in B_\Delta^{(1)} \times A^{(2)}]}{\mathbb{P}[(X^\star)^{(2)} \in A_\Delta^{(2)}]} \\
&\geqslant \frac{\limsup \mathbb{P}[X_n \in B_\Delta^{(1)} \times A_\Delta^{(2)}]}{\lim \mathbb{P}[(X_n)^{(2)} \in A_\Delta^{(2)}]} \text{ since } \mathbb{P}[(X_n)^{(2)} \in \partial A_\Delta^{(2)}] = 0 \\
&\geqslant \limsup \frac{\mathbb{P}[X_n \in B_\Delta^{(1)} \times A_\Delta^{(2)}]}{\mathbb{P}[(X_n)^{(2)} \in A_\Delta^{(2)}]} = 1.
\end{aligned}$$

Therefore, $\Phi_{X^\star} \in \mathcal{H}^\star(\kappa, S, M, (c_\Delta, A_\Delta^{(1)}, B_\Delta^{(1)}, A_\Delta^{(2)}, B_\Delta^{(2)})_{\Delta>0})$.

6.16 Proof of Lemma 9

Let us write $m_{i,\gamma}(x) = (x + (-1)^i \gamma \frac{\alpha}{2} \cos(\frac{x}{\alpha\gamma}), 0, \dots, 0)$, so that $X_i(\kappa) = (\alpha U(\kappa), m_{i,\gamma}(\alpha U(\kappa))$. For $i \in \{0, 1\}$, let $w_{i,\kappa,\gamma}$ be the density of the first coordinate of $m_{i,\gamma}(\alpha U(\kappa))$, then

$$M_\kappa = \sup_{x\in\mathbb{R},\gamma\in[0,1],i\in\{0,1\}} \{w_{i,\kappa,\gamma}(x) \vee \frac{1}{\alpha} f_\kappa(\frac{x}{\alpha})\}$$

is an upper bound of the density of $X_i(\kappa)^{(1)}$ and of the first coordinate of $X_i(\kappa)^{(2)}$ with respect to the Lebesgue measure. Let us show that M_κ is finite. First, note that $m_{i,\gamma}$ is one-to-one from $\mathbb{R}$ to $\mathbb{R} \times \{0\}^{D-2}$ and $m_{i,\gamma}^{-1}$ is Lipschitz with Lipschitz constant upper bounded by $1/2$. One can easily check that for all $x \in \mathbb{R}$,

$$w_{i,\kappa,\gamma}(x) = \frac{1}{\alpha} f_\kappa \left(\frac{(m_{i,\gamma})_1(x, 0, \dots, 0)}{\alpha} \right) \frac{1}{(m_{i,\gamma})'_1(m_{i,\gamma}^{-1}(x, 0, \dots, 0))},$$

where $(m_{i,\gamma})_1(x)$ is the first coordinate of $m_{i,\gamma}(x)$. Since $(m_{i,\gamma})'_1$ is lower bounded by $1/2$, $M_\kappa \leqslant \sup_{x\in\mathbb{R}} \frac{1}{\alpha} f_\kappa(\frac{x}{\alpha})$, which is finite.

For any $\Delta > 0$, define the sets:

$$A_\Delta^{(1)} = [-\Delta, \Delta] \quad \text{and} \quad B_\Delta^{(2)} = \bar{B}\left(0, (\frac{\alpha}{2}+2)\Delta\right) \cap (\mathbb{R} \times \{0\}^{D-2}),$$
$$A_\Delta^{(2)} = [-\Delta, \Delta] \times \{0\}^{D-2} \quad \text{and} \quad B_\Delta^{(1)} = \bar{B}(0, \Delta) \cap \mathbb{R}.$$

Define $c_{\Delta,\kappa,\alpha} = \mathbb{P}[\alpha U(\kappa) \in A_\Delta^{(1)}] \wedge \inf_{\gamma\in[0,1], i\in\{0,1\}} \mathbb{P}[m_{i,\gamma}(\alpha U(\kappa)) \in A_\Delta^{(2)}]$, and let us prove that $c_{\Delta,\kappa,\alpha} > 0$.

First, $\mathbb{P}[\alpha U(\kappa) \in A_\Delta^{(1)}] > 0$ since the density of $\alpha U(\kappa)$ is positive everywhere on its support. Then, for $i \in \{0, 1\}$,

$$\begin{aligned}
\mathbb{P}[m_{i,\gamma}(\alpha U(\kappa)) \in A_\Delta^{(2)}] &= \mathbb{P}\left(\alpha U(\kappa) + (-1)^i \frac{\alpha}{2}\gamma \cos\left(\frac{U(\kappa)}{\gamma}\right) \in [-\Delta, \Delta]\right) \\
&\geqslant \mathbb{P}\left(\alpha U(\kappa) \in [-\Delta/2, \Delta/2], (-1)^i \frac{\alpha}{2}\gamma \cos\left(\frac{U(\kappa)}{\gamma}\right) \in [-\Delta/2, \Delta/2]\right) \\
&\geqslant \mathbb{P}\left(\alpha U(\kappa) \in [-\Delta/2, \Delta/2], \cos\left(\frac{U(\kappa)}{\gamma}\right) \in [-\Delta/\gamma\alpha, \Delta/\gamma\alpha]\right) \\
&\geqslant \mathbb{P}\left(U(\kappa) \in \left[-\frac{\Delta}{2\alpha}, \frac{\Delta}{2\alpha}\right] \cap \left[\arccos\left(\frac{\Delta}{\alpha}\right), \pi - \arccos(\frac{\Delta}{\alpha})\right]\right),
\end{aligned}$$

which is positive.

It is clear that the sets satisfy (i) and (ii). It remains to prove that $X_i(\kappa)$ satisfies (iii) and (iv).

For any $\Delta > 0$ define $B_{\Delta,i,\gamma}^{(1)} = m_{i,\gamma}^{-1}(A_\Delta^{(2)})$. Then

$$\begin{aligned}
\text{Diam}(B_{\Delta,i,\gamma}^{(1)}) &= \sup_{x,y\in B_{\Delta,i,\gamma}^{(1)}} |x-y| \\
&= \sup_{x,y\in A_\Delta^{(2)}} |m_{i,\gamma}^{-1}(x) - m_{i,\gamma}^{-1}(y)| \\
&\leqslant \frac{1}{2} \sup_{x,y\in A_\Delta^{(2)}} \|x-y\| \leqslant \Delta.
\end{aligned}$$

Thus, $B_{\Delta,i,\gamma}^{(1)} \subset B_\Delta^{(1)}$, and

$$\mathbb{P}[(X_i(\kappa))^{(1)} \in B_{\Delta,i}^{(1)} | (X_i(\kappa))^{(2)} \in A_\Delta^{(2)}] \geqslant \mathbb{P}[(X_i(\kappa))^{(1)} \in B_{\Delta,i,\gamma}^{(1)} | (X_i(\kappa))^{(2)} \in A_\Delta^{(2)}] = 1.$$

Similarly, define $B_{\Delta,i,\gamma}^{(2)} = m_{i,\gamma}(A_\Delta^{(1)}) = \{(x + (-1)^i \gamma \frac{\alpha}{2} \cos(\frac{x}{\alpha\gamma}), 0, \dots, 0)\ ,\ x \in [-\Delta, \Delta]\}$, then

$$\begin{aligned}
\text{Diam}(B_{\Delta,i,\gamma}^{(2)}) &= \sup_{x,y\in A_\Delta^{(1)}} \left| x + (-1)^i \gamma \frac{\alpha}{2} \cos\left(\frac{x}{\alpha\gamma}\right) - y - (-1)^i \gamma \frac{\alpha}{2} \cos\left(\frac{y}{\alpha\gamma}\right) \right| \\
&\leqslant 2\Delta + \gamma \frac{\alpha}{2} \left| \cos\left(\frac{x}{\alpha\gamma}\right) - cos\left(\frac{y}{\alpha\gamma}\right) \right| \\
&\leqslant \left(\frac{\alpha}{2} + 2\right) \Delta.
\end{aligned}$$

Thus, $B_{\Delta,i,\gamma}^{(2)} \subset B_{\Delta,i}^{(2)}$, and

$$\mathbb{P}[(X_i(\kappa))^{(2)} \in B_{\Delta,i}^{(2)} | (X_i(\kappa))^{(1)} \in A_\Delta^{(1)}] \geqslant \mathbb{P}[(X_i(\kappa))^{(2)} \in B_{\Delta,i,\gamma}^{(2)} | (X_i(\kappa))^{(1)} \in A_\Delta^{(1)}] = 1.$$

6.17 Proof of Theorem 5

In the following, we will write A, B, C (with upper case letters) positive constants that can change from line to line. As in [19] and [21], we use the upper bound:

$$\|(G_0(\kappa) * Q)^{\otimes n} - (G_1(\kappa) * Q)^{\otimes n}\|_{TV} \leqslant 1 - (1 - \|(G_0(\kappa) * Q) - (G_1(\kappa) * Q)\|_{TV})^n,$$

where $\| \cdot - \cdot \|_{TV}$ denotes the total variation distance. Using Le Cam's two-points method, the minimax rate will be lower bounded by $H(A_\alpha M_0(\gamma), A_\alpha M_1(\gamma))$, that is γ, (see Lemma 7) provided that there exists a constant $C > 0$ such that $\|(G_0(\kappa) * Q)^{\otimes n} - (G_1(\kappa) * Q)^{\otimes n}\|_{TV} \leqslant C < 1$, so that we only need to find $C > 0$ such that

$$\int_{\mathbb{R}^D} |d(G_0(\kappa) * Q)(x) - d(G_1(\kappa) * Q)(x)| \leqslant \frac{C}{n}.$$

Since Q has a density q over $\mathbb{R}^D$, $G_0(\kappa) * Q$ and $G_1(\kappa) * Q$ also have a density over $\mathbb{R}^D$. We first prove that for $i \in \{0, 1\}$,

$$\int \prod_{j=1}^D x_j^2 \left| \frac{d(G_i(\kappa) * Q)}{dx}(x) \right|^2 dx < +\infty. \tag{32}$$

Indeed,

$$\int \prod_{j=1}^D x_j^2 \left| \frac{d(G_i(\kappa) * Q)}{dx}(x) \right|^2 dx \leqslant \left\| \frac{d(G_i(\kappa) * Q)}{dx} \right\|_\infty \int \prod_{j=1}^D x_j^2 d(G_i(\kappa) * Q)(x).$$

First, $\left\| \frac{d(G_i(\kappa) * Q)}{dx} \right\|_\infty \leqslant \|q\|_\infty^D < \infty$. Moreover, for $k \in \{1, \dots, D\}$, writing $X_i(\kappa)^{[k]}$ and $\varepsilon^{[k]}$ for the k-th coordinate of $X_i(\kappa)$ and ε,

$$\begin{aligned} \int \prod_{j=1}^D x_j^2 \, |d(G_i(\kappa) * Q)(x)| &= \mathbb{E}[\prod_{k=1}^D (X_i(\kappa)^{[k]} + \varepsilon^{[k]})^2] \\ &= \mathbb{E}[(X_i(\kappa)^{[1]} + \varepsilon^{[1]})^2 (X_i(\kappa)^{[2]} + \varepsilon^{[2]})^2] \prod_{k=3}^D \mathbb{E}[(\varepsilon^{[k]})^2]. \end{aligned} \tag{33}$$

We have that $(X_i(\kappa)^{[2]} + \varepsilon^{[2]})^2 \leqslant a^2 (X_i(\kappa)^{[1]})^2 + 2\gamma X_i(\kappa)^{[1]} + 2 X_i(\kappa)^{[1]} \varepsilon^{[2]} + (1+\gamma)(\varepsilon^{[2]})^2 + \gamma^2$, using (33) and the fact that $\varepsilon^{[2]}$ is independent of all other variables and that, for $k \in \{1, 2\}$, $X_i^{[1]}$ is independent of $\varepsilon^{[k]}$, we finally get that $\int \prod_{j=1}^D x_j^2 \, |d(G_i(\kappa) * Q)(x)|$ is upper bounded by product and sum of expectation of $((\varepsilon^{[j]})^2)_{j \in \{1, \dots, D\}}$, $(X_i(\kappa)^{[1]})^2$, $(X_i(\kappa)^{[1]})^3$ and $(X_i(\kappa)^{[1]})^4$ which are all finite thanks to Lemma 4.

By the Cauchy-Schwarz inequality,

$$\begin{aligned} &\int_{\mathbb{R}^D} |d(G_0(\kappa) * Q)(x) - d(G_1(\kappa) * Q)(x)| \\ &\qquad \leqslant \pi^{D/2} \left(\int \prod_{j=1}^D (1 + x_j^2) \left| \frac{d((G_0(\kappa) - G_1(\kappa)) * Q)}{dx}(x) \right|^2 dx \right)^{1/2}. \end{aligned} \tag{34}$$

By Parseval's identity, for all $\eta \in \{0, 1\}^D$,

$$\begin{aligned} \int_{\mathbb{R}^D} \prod_{j=1}^D x_j^{2\eta_j} \left| \frac{d((G_0(\kappa) - G_1(\kappa)) * Q)}{dx}(x) \right|^2 dx &= \int_{\mathbb{R}^D} \left| \left(\prod_{j=1}^D \partial_{t_j}^{\eta_j} \right) (\mathcal{F}[G_0(\kappa)] - \mathcal{F}[G_1(\kappa)])(t) \mathcal{F}[Q](t) \right|^2 dt \\ &= \int_{[-c,c]^D} \left| \left(\prod_{j=1}^D \partial_{t_j}^{\eta_j} \right) (\mathcal{F}[G_0(\kappa)] - \mathcal{F}[G_1(\kappa)])(t) \mathcal{F}[Q](t) \right|^2 dt, \end{aligned}$$

since $\mathcal{F}[Q]$ and for $\eta \in \{0, 1\}^D$, $\partial^\eta \mathcal{F}[Q]$ are supported on $[-c, c]^D$. Moreover, they are bounded functions, so that there exists a constant C (depending only on d) such that

$$\int_{\mathbb{R}^D} |d(G_0(\kappa)*Q)(x) - d(G_1(\kappa)*Q)(x)| \leqslant C \sum_{\eta\in\{0,1\}^D} \int_{[-c,c]^D} \left| \left(\prod_{j=1}^D \partial_{t_j}^{\eta_j}\right) (\mathcal{F}[G_0(\kappa)] - \mathcal{F}[G_1(\kappa)])(t) \right|^2 dt$$
$$= \sum_{\eta\in\{0,1\}^D} \int_{[-c,c]^D} \left| \left(\prod_{j=1}^D \partial_{t_j}^{\eta_j}\right) (t \mapsto \mathcal{F}[S_0] - \mathcal{F}[S_1])(A_a^\top t) \right|^2 dt.$$

Using the change of variable $u = A_\alpha^\top t$, and noticing that $\{A_\alpha^\top t\,;\ t\in[-c,c]^D\} \subset [-(1+\alpha)c,(1+\alpha)c]^D$, there exists a constant $C>0$ depending on d and a such that

$$\int_{\mathbb{R}^D} |d(G_0(\kappa)*Q)(x) - d(G_1(\kappa)*Q)(x)| \leqslant C \sum_{\eta\in\{0,1\}^D} \int_{[-(1+\alpha)c,(1+\alpha)c]^D} \left| \left(\prod_{j=1}^D \partial_{t_j}^{\eta_j}\right) (\mathcal{F}[S_0] - \mathcal{F}[S_1])(u) \right|^2 du.$$

For all $t=(t_1,\dots,t_D)\in\mathbb{R}^D$, for $i\in\{0,1\}$, $\mathcal{F}[T_i](t) = \mathcal{F}[\tilde{T}_i](t_1,t_2)$, where $\tilde{T}_i$ is the distribution of the 2 first coordinates of $S_i(\kappa)$ under T_i. There exists a constant $C>0$ such that

$$\int_{\mathbb{R}^D} |d(G_0(\kappa)*Q)(x) - d(G_1(\kappa)*Q)(x)|$$
$$\leqslant C \sum_{\eta\in\{0,1\}^2} \int_{[-(1+\alpha)c,(1+\alpha)c]^2} \left| \left(\prod_{j=1}^2 \partial_{t_j}^{\eta_j}\right) (\mathcal{F}[\tilde{T}_0] - \mathcal{F}[\tilde{T}_1])(t) \right|^2 dt. \quad (35)$$

Following the same approach as [21], we get that for all $t=(t_1,t_2)\in\mathbb{R}^2$,

$$\begin{aligned}(\mathcal{F}[\tilde{T}_0] - \mathcal{F}[\tilde{T}_1])(t) &= \int_{\mathbb{R}} \{e^{it_1u + i\gamma t_2\tilde{g}_\gamma(u)} - e^{it_1u - i\gamma t_2\tilde{g}_\gamma(u)}\} f_\kappa(u)du \\ &= 2i\int_{\mathbb{R}} e^{it_1u}\sin(t_2\gamma\tilde{g}_\gamma(u)) f_\kappa(u)du \\ &= 2i\int_{\mathbb{R}} e^{it_1u}\sum_{k=0}^\infty \frac{(-1)^k t_2^{2k+1}\gamma^{2k+1}}{(2k+1)!}\tilde{g}_\gamma^{2k+1}(u) f_\kappa(u)du.\end{aligned}$$

Since $\sum_{k=0}^\infty \int_{\mathbb{R}} \frac{|t_2|^{2k+1}\gamma^{2k+1}}{(2k+1)!}|\tilde{g}_\gamma^{2k+1}(u)| f_\kappa(u)du$ is finite, we can switch integral and sum thanks to Fubini Theorem, so that

$$\begin{aligned}(\mathcal{F}[\tilde{T}_0] - \mathcal{F}[\tilde{T}_1])(t) &= 2i\sum_{k=0}^\infty \frac{(-1)^k t_2^{2k+1}\gamma^{2k+1}}{(2k+1)!}\int_{\mathbb{R}} e^{it_1u}\tilde{g}_\gamma^{2k+1}(u) f_\kappa(u)du \\ &= 2i\sum_{k=0}^\infty \frac{(-1)^k t_2^{2k+1}\gamma^{2k+1}}{(2k+1)!} m_k(t_1),\end{aligned}$$

with for all $u\in\mathbb{R}$,

$$m_k(u) = \mathcal{F}[\tilde{g}^{2k+1}f_\kappa](u) = (\underbrace{\mathcal{F}[\tilde{g}]*\mathcal{F}[\tilde{g}]*\dots*\mathcal{F}[\tilde{g}]}_{2k+1 \text{ times}}*\mathcal{F}[f_\kappa])(u). \quad (36)$$

Since

$$\mathcal{F}[x\mapsto\cos(\frac{x}{\gamma})] = \frac{1}{2}\delta_{-\frac{1}{\gamma}} + \frac{1}{2}\delta_{\frac{1}{\gamma}},$$

for all $u\in\mathbb{R}$,

$$\begin{aligned}(\underbrace{\mathcal{F}[\tilde{g}]*\mathcal{F}[\tilde{g}]*\dots*\mathcal{F}[\tilde{g}]}_{2k+1 \text{ times}})(u) &= \underbrace{\mathcal{F}[\cos(\frac{\cdot}{\gamma})]*\dots*\mathcal{F}[\cos(\frac{\cdot}{\gamma})]}_{2k+1 \text{ times}} \\ &= \left(\frac{1}{2}\right)^{2k+1}\sum_{j=1}^{2k+1}\binom{2k+1}{j}\delta_{a_j},\end{aligned}$$

where $a_j = (2j-2k-1)/\gamma$. By (36),

$$m_k(u) = \left(\frac{1}{2}\right)^{2k+1} \sum_{j=0}^{2k+1} \binom{2k+1}{j} \mathcal{F}[f_\kappa](u-a_j).$$

Therefore,

$$\sup_{|t|\leqslant c} |m_k(t)| \leqslant \sup_{|t|\leqslant c, 0\leqslant j\leqslant 2k+1} \left| \mathcal{F}[f_\kappa]\left(t - \frac{2j-2k-1}{\gamma}\right)\right|$$

and

$$\sup_{|t|\leqslant c} |m_k'(t)| \leqslant \sup_{|t|\leqslant c, 0\leqslant j\leqslant 2k+1} \left| \mathcal{F}[f_\kappa]'\left(t - \frac{2j-2k-1}{\gamma}\right)\right|.$$

Assume first that $\kappa \in (1/2, 1)$**.** For γ that satisfies $\gamma \leqslant \frac{1}{2c}$, by Lemma 4, there exist two constants A, B independent of γ and k such that

$$\sup_{|t|\leqslant c, 0\leqslant j\leqslant 2k+1} \left| \mathcal{F}[f_\kappa]\left(t - \frac{2j-2k-1}{\gamma}\right)\right| \leqslant A\exp(-B\gamma^{-\frac{1}{\kappa}})$$

and

$$\sup_{|t|\leqslant c, 0\leqslant j\leqslant 2k+1} \left| \mathcal{F}[f_\kappa]'\left(t - \frac{2j-2k-1}{\gamma}\right)\right| \leqslant A\exp(-B\gamma^{-\frac{1}{\kappa}}).$$

Thus,

$$\sup_{|t|\leqslant c} |m_k(t)| \leqslant A\exp(-B\gamma^{-\frac{1}{\kappa}}), \tag{37}$$

and

$$\sup_{|t|\leqslant c} |m_k'(t)| \leqslant A\exp(-B\gamma^{-\frac{1}{\kappa}}). \tag{38}$$

For all $\eta \in \{0,1\}^2$, and $t \in [-c, c]$,

$$\begin{aligned}\left(\prod_{j=1}^{2} \partial_{t_j}^{\eta_j}\right)(\mathcal{F}[\tilde{T}_0] - \mathcal{F}[\tilde{T}_1])(t) &= \prod_{j=1}^{2} \partial_{t_j}^{\eta_j}\left[2i\sum_{k=0}^{\infty} \frac{(-1)^k t_2^{2k+1}\gamma^{2k+1}}{(2k+1)!} m_k(t_1)\right] \\ &= 2i\eta_2 \sum_{k=0}^{\infty} \frac{(-1)^k t_2^{2k}\gamma^{2k+1}}{(2k)!} \partial_{t_1}^{\eta_1} m_k(t_1) + 2i(1-\eta_2)\sum_{k=0}^{\infty} \frac{(-1)^k t_2^{2k+1}\gamma^{2k+1}}{(2k+1)!} \partial_{t_1}^{\eta_1} m_k(t_1),\end{aligned}$$

so that

$$\begin{aligned}&\left|\left(\prod_{j=1}^{2} \partial_{t_j}^{\eta_j}\right)(\mathcal{F}[\tilde{T}_0] - \mathcal{F}[\tilde{T}_1])(t)\right| \\ &\leqslant 2\sum_{k=0}^{\infty} \frac{|t_2|^{2k}\gamma^{2k+1}}{(2k)!} |\partial_{t_1}^{\eta_1} m_k(t_1)| + 2\sum_{k=0}^{\infty} \frac{|t_2|^{2k+1}\gamma^{2k+1}}{(2k+1)!} |\partial_{t_1}^{\eta_1} m_k(t_1)|.\end{aligned}$$

By (37) and (38), there exists a constant $C > 0$ which depends only on d and A such that

$$\left(\prod_{j=1}^{2} \partial_{t_j}^{\eta_j}\right)(\mathcal{F}[\tilde{S}_0] - \mathcal{F}[\tilde{S}_1])(t) \leqslant C\exp(-B\gamma^{-\frac{1}{\kappa}}) \sup_{|t_2|\leqslant c}\left(\gamma\cosh(|t_2|\gamma) + \sinh(|t_2|\gamma)\right).$$

For γ small enough, there exists a constant $C_1 > 0$ which depends only on d, $A > 0$ and $C > 0$ such that

$$\left|\left(\prod_{j=1}^{2} \partial_{t_j}^{\eta_j}\right)(\mathcal{F}[\tilde{T}_0] - \mathcal{F}[\tilde{T}_1])(t)\right| \leqslant C\exp(-B\gamma^{-\frac{1}{\kappa}}).$$

Finally, using (35), there exist constants $C > 0$ and $B > 0$ which depend only on d such that

$$\int_{\mathbb{R}^D} |d(G_0(\kappa) * Q)(x) - d(G_1(\kappa) * Q)(x)| \leqslant C \exp(-B\gamma^{-\frac{1}{\kappa}}).$$

Taking $\gamma = c_\gamma (\log n)^{-\kappa}$ with $c_\gamma \leqslant B_1^\kappa$ shows that there exists $C > 0$ such that

$$\int_{\mathbb{R}^D} |d(G_0(\kappa) * Q)(x) - d(G_1(\kappa) * Q)(x)| \leqslant \frac{C}{n}$$

Let us now consider the case $\kappa = 1$**.** For γ that satisfies $\gamma \leqslant \frac{1}{2c}$, by Lemma 4, for all $\delta \in (0,1)$, there exist two constants $A > 0$, $B > 0$ independent of γ and k such that

$$\sup_{|t|\leqslant c, 0\leqslant j \leqslant 2k+1} \left| \mathcal{F}[f_1] \left(t - \frac{2j-2k-1}{\gamma} \right) \right| \leqslant A \exp(-B\gamma^{-\delta}),$$

and

$$\sup_{|t|\leqslant c, 0\leqslant j \leqslant 2k+1} \left| \mathcal{F}[f_1]' \left(t - \frac{2j-2k-1}{\gamma} \right) \right| \leqslant A \exp(-B\gamma^{-\delta}).$$

Thus, there exists constants $A > 0$ and $B > 0$ independent of γ and k such that

$$\sup_{|t|\leqslant c} |m_k(t)| \leqslant A \exp(-B\gamma^{-\delta}), \tag{39}$$

and

$$\sup_{|t|\leqslant c} |m_k'(t)| \leqslant A \exp(-B\gamma^{-\delta}). \tag{40}$$

Doing the same computation as in the case $\kappa \in (1/2, 1)$ shows that for all $\eta \in \{0,1\}^2$ and $t \in [-c, c]$,

$$\begin{aligned} &\left| \left(\prod_{j=1}^{2} \partial_{t_j}^{\eta_j} \right) (\mathcal{F}[\tilde{T}_0] - \mathcal{F}[\tilde{T}_1])(t) \right| \\ &\leqslant 2 \sum_{k=0}^{\infty} \frac{|t_2|^{2k} \gamma^{2k+1}}{(2k)!} |\partial_{t_1}^{\eta_1} m_k(t_1)| + 2 \sum_{k=0}^{\infty} \frac{|t_2|^{2k+1} \gamma^{2k+1}}{(2k+1)!} |\partial_{t_1}^{\eta_1} m_k(t_1)|. \end{aligned}$$

By (39) and (40), there exist constants $C > 0$ and $B > 0$ which depend only on d such that

$$\left(\prod_{j=1}^{2} \partial_{t_j}^{\eta_j} \right) (\mathcal{F}[\tilde{T}_0] - \mathcal{F}[\tilde{T}_1])(t) \leqslant C \exp(-B\gamma^{-\delta}) \sup_{|t_2|\leqslant c} \Big(\gamma \cosh(|t_2|\gamma) + \sinh(|t_2|\gamma) \Big).$$

For γ small enough, there exists a constant $C > 0$ which depends only on d such that

$$\left| \left(\prod_{j=1}^{2} \partial_{t_j}^{\eta_j} \right) (\mathcal{F}[\tilde{T}_0] - \mathcal{F}[\tilde{T}_1])(t) \right| \leqslant C \exp(-B\gamma^{-\delta}).$$

Finally, using (35), there exists a constant $C > 0$ which depends only on d such that

$$\int_{\mathbb{R}^D} |d(G_0(\kappa) * Q)(x) - d(G_1(\kappa) * Q)(x)| \leqslant C \exp(-B\gamma^{-\delta}).$$

Taking $\gamma = c_\gamma (\log n)^{-\frac{1}{\delta}}$ with $c_\gamma \leqslant B^{\frac{1}{\delta+1}}$ shows that there exists $C > 0$ such that

$$\int_{\mathbb{R}^D} |d(G_0(\kappa) * Q)(x) - d(G_1(\kappa) * Q)(x)| \leqslant \frac{C}{n}.$$

6.18 Proof of Theorem 6

Fix $\kappa_0 \in (1/2, 1]$, $S > 0$, $a > 0$ $d \leqslant D$, $\nu \in (0, \nu_{\text{est}}]$, $c(\nu) > 0$ $E > 0$. Using the end of the proof of Theorem 4, there exists n_0 and c' such that for all $\kappa \in [\kappa_0, 1]$, all $G \in St_{\mathcal{K}}(a, d) \cap \mathcal{L}(\kappa, S, \mathcal{H})$ and all $\mathbb{Q} \in \mathcal{Q}^{(d)}(\nu, c(\nu), E)$, with $(G * \mathbb{Q})^{\otimes n}$-probability at least $1 - \frac{2}{n}$, (28) holds. Let us now choose $c_\sigma = c'$ and consider the event where (28) holds. By the triangular inequality, for any $\kappa \in [\kappa_0, 1]$,

$$\begin{aligned} H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}}_{\widehat{\kappa}_n}) &\leqslant H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}}_{\kappa}) + H_{\mathcal{K}}(\widehat{\mathcal{M}}_{\kappa}, \widehat{\mathcal{M}}_{\widehat{\kappa}_n}) \\ &\leqslant \sigma_n(\kappa) + H_{\mathcal{K}}(\widehat{\mathcal{M}}_{\kappa}, \widehat{\mathcal{M}}_{\widehat{\kappa}_n}). \end{aligned}$$

Now, using the definition of $B_n(\cdot)$, if $\kappa \leqslant \widehat{\kappa}_n$, then

$$H_{\mathcal{K}}(\widehat{\mathcal{M}}_{\kappa}, \widehat{\mathcal{M}}_{\widehat{\kappa}_n}) \leqslant B_n(\widehat{\kappa}_n) + \sigma_n(\kappa)$$

while if $\kappa \geqslant \widehat{\kappa}_n$, then

$$H_{\mathcal{K}}(\widehat{\mathcal{M}}_{\kappa}, \widehat{\mathcal{M}}_{\widehat{\kappa}_n}) \leqslant B_n(\kappa) + \sigma_n(\widehat{\kappa}_n)$$

so that in all cases,

$$\begin{aligned} H_{\mathcal{K}}(\widehat{\mathcal{M}}_{\kappa}, \widehat{\mathcal{M}}_{\widehat{\kappa}_n}) &\leqslant B_n(\widehat{\kappa}_n) + \sigma_n(\kappa) + B_n(\kappa) + \sigma_n(\widehat{\kappa}_n) \\ &\leqslant 2B_n(\kappa) + 2\sigma_n(\kappa) \end{aligned}$$

using the definition of $\widehat{\kappa}_n$, and therefore

$$H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}}_{\widehat{\kappa}_n}) \leqslant 2B_n(\kappa) + 3\sigma_n(\kappa).$$

By the triangular inequality and the definition of $B_n(\cdot)$,

$$\begin{aligned} B_n(\kappa) &\leqslant 0 \vee \sup_{\kappa' \in [\kappa_0, \kappa]} \left\{ H_{\mathcal{K}}(\widehat{\mathcal{M}}_{\kappa}, \mathcal{M}_G) + H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}}_{\kappa'}) - \sigma_n(\kappa') \right\} \\ &\leqslant H_{\mathcal{K}}(\widehat{\mathcal{M}}_{\kappa}, \mathcal{M}_G) + 0 \vee \sup_{\kappa' \in [\kappa_0, \kappa]} \left\{ H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}}_{\kappa'}) - \sigma_n(\kappa') \right\} \\ &\leqslant \sigma_n(\kappa). \end{aligned}$$

Thus, for all $\kappa \in [\kappa_0, 1]$, all $G \in St_{\mathcal{K}}(a, d) \cap \mathcal{L}(\kappa, S, \mathcal{H})$ and all $\mathbb{Q} \in \mathcal{Q}^{(d)}(\nu, c(\nu), E)$, with $(G * \mathbb{Q})^{\otimes n}$-probability at least $1 - \frac{2}{n}$,

$$H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}}_{\widehat{\kappa}_n}) \leqslant 5\sigma_n(\kappa),$$

and using the fact that $H_{\mathcal{K}}(\mathcal{M}_G, \widehat{\mathcal{M}}_{\widehat{\kappa}}) \leqslant \sup_{x, x' \in \mathcal{K}} d(x, x')$ on the event of probability at most $2/n$ where this doesn't hold, Theorem 6 follows.

6.19 Proof of Theorem 7

We shall need two technical lemmas. The following one is easily proved following the arguments at the end of the proof of Theorem 4.

Lemma 12. *Let G be a probability measure with compact support $\mathcal{M}_G$. Assume $G \in St_{\mathcal{M}_G}(a, d, r_0)$ for some constants $a > 0$, $d > 0$ and $r_0 > 0$. Recall that $\Gamma_n := \Gamma_{n,1} = \|\hat{g}_n - \bar{g}\|_\infty$. Then*

(1) For any $C_1 > 0$ and $c > 0$, there exists $h_0 > 0$ such that if $h_n \leqslant h_0$, on the event where

$$C_1 + \Gamma_n < \lambda_n < \; ac_A^d d_A (\frac{1}{h_n})^{D-d} - \Gamma_n,$$

it holds

$$\mathcal{M}_G \subset \widehat{\mathcal{M}} \subset (\mathcal{M}_G)_c.$$

(2) *For m_n, h_n, λ_n chosen as in Theorem 4, for all $C_1 \in (0, ac_A^d d_A)$ and $\delta' > 0$, there exists $C > 0$ and $n_0 \geqslant 0$ such that for all $n \geqslant n_0$, with probability at least $1 - 2\exp(-n^{1/2-\delta'})$,*

$$\Gamma_n^2 \leqslant Ce^{-m_n} \qquad \text{and} \qquad C_1 + \Gamma_n < \lambda_n < \; ac_A^d d_A(\frac{1}{h_n})^{D-d} - \Gamma_n.$$

and in particular, for all $c > 0$ and $\delta' > 0$, there exists $n_0' \geqslant 0$ such that for all $n \geqslant n_0'$, with probability at least $1 - 2\exp(-n^{1/2-\delta'})$,

$$\mathcal{M}_G \subset \widehat{\mathcal{M}} \subset (\mathcal{M}_G)_c.$$

In particular, since $R_n \longrightarrow +\infty$ and $\mathcal{M}_G$ is compact, up to increasing n_0, on this event,

$$\mathcal{M}_G \subset \widehat{\mathcal{M}} \cap \bar{B}(0, R_n) \subset (\mathcal{M}_G)_c.$$

In the rest of the proof of the Theorem, we lighten the notation $\widehat{\mathcal{M}} \cap \bar{B}(0, R_n)$ into $\widehat{\mathcal{M}}$ (equivalently, we redefine the estimator $\hat{\mathcal{M}}$ as the intersection of the estimator of Section 3.2 with the closed euclidean ball of radius R_n).

Lemma 13. *Let G be a probability measure with compact support $\mathcal{M}_G$. Assume $G \in St_{\mathcal{M}_G}(a, d, r_0)$ for some constants $a > 0$, $d > 0$ and $r_0 > 0$. Then for any $\alpha > 0$ and $c > 0$, there exists $C(\alpha, c) > 0$ such that, on the event where*

$$\mathcal{M}_G \subset \widehat{\mathcal{M}} \subset (\mathcal{M}_G)_c,$$

it holds

$$\|\bar{g}\|_{L_1(\mathbb{R}^D \setminus (\widehat{\mathcal{M}})_c)} \leqslant C(\alpha, c) h_n^\alpha \quad \text{and} \quad \int_{\mathbb{R}^D \setminus (\widehat{\mathcal{M}})_c} \|x\|^2 |\bar{g}(x)| dx \leqslant C(\alpha, c) h_n^\alpha.$$

Proof. By definition,

$$\|\bar{g}\|_{L_1(\mathbb{R}^D \setminus (\widehat{\mathcal{M}})_c)} = \frac{1}{h_n^D} \int_{x \in \mathcal{M}_G} \int_{y \in \mathbb{R}^D \setminus (\widehat{\mathcal{M}})_c} \psi_A \left(\frac{\|y - x\|_2}{h_n} \right) dy dG(x).$$

By (4), for any $A > 0$, there exists $C > 0$ such that for any $x \in \mathcal{M}_G$ and $y \in \mathbb{R}^D \setminus (\widehat{\mathcal{M}})_c$,

$$\psi_A \left(\frac{\|y - x\|_2}{h_n} \right) \leqslant C \exp\left(-\beta_A \frac{\|y - x\|_2^{A/(A+1)}}{h_n^{A/(A+1)}} \right) \leqslant C \exp\left(-\beta_A \frac{d(y, \mathcal{M}_G)^{A/(A+1)}}{h_n^{A/(A+1)}} \right).$$

Since $\mathcal{M}_G \subset \widehat{\mathcal{M}}$, for all $y \in \mathbb{R}^D \setminus (\widehat{\mathcal{M}})_c$, $d(y, \mathcal{M}_G) \geqslant c$, so for any $\alpha > 0$, there exists a constant $\tilde{C} > 0$ such that

$$C \exp\left(-\beta_A \frac{d(y, \mathcal{M}_G)^{A/(A+1)}}{h_n^{A/(A+1)}} \right) \leqslant \tilde{C} \frac{h_n^{D+\alpha}}{d(y, \mathcal{M}_G)^{D+\alpha}}.$$

Moreover, since $\mathcal{M}_G$ is compact, $\text{Diam}(\mathcal{M}_G)$ is finite, so that on the event where $\mathcal{M}_G \subset \widehat{\mathcal{M}}$,

$$\begin{aligned} \int_{\mathbb{R}^D \setminus (\widehat{\mathcal{M}})_c} \frac{1}{d(y, \mathcal{M}_G)^{D+\alpha}} dy &\leqslant \int_{\mathbb{R}^D \setminus (\mathcal{M}_G)_c} \frac{1}{d(y, \mathcal{M}_G)^{D+\alpha}} dy < \infty. \\ &\leqslant \int_{\mathbb{R}^D} \left(\frac{1}{c \vee (\|y\| - \text{Diam}(\mathcal{M}_G)/2)} \right)^{D+\alpha} dy < \infty. \end{aligned}$$

Therefore, for all $c > 0$ and $\alpha > 0$, there exists C depending on A, D, c, α and $\text{Diam}(\mathcal{M}_G)$ such that

$$\|\bar{g}\|_{L_1(\mathbb{R}^D \setminus (\widehat{\mathcal{M}})_c)} \leqslant C h_n^\alpha.$$

The proof that the same holds for $\int_{\mathbb{R}^D \setminus (\widehat{\mathcal{M}})_c} \|x\|^2 \bar{g}(x) dx$ is similar. ☐

Let $G \in St_{\mathcal{M}_G}(a, d, r_0)$ be such that if $X \sim G$, then $\Phi_X \in \mathcal{H} \cap \Upsilon_{1,S}$. We use a bias-variance decomposition of $W_2(G, \widehat{P}_{n,\eta})$ as

$$W_2(G, \widehat{P}_{n,\eta}) \leqslant W_2(G, P_{\psi_{A,h}}) + W_2(P_{\psi_{A,h}}, \widehat{P}_{n,\eta}).$$

The proof is done is several steps :

(1) We first show that there exists $C > 0$ depending only on A and D such that the bias satisfies

$$W_2(G, P_{\psi_{A,h_n}}) \leqslant Ch_n.$$

(2) We prove that for any $\alpha \geqslant 1$, on the event where

$$\mathcal{M}_G \subset \widehat{\mathcal{M}} \subset (\mathcal{M}_G)_c,$$

there exists $C' > 0$ such that

$$W_2(P_{\psi_{A,h_n}}, \widehat{P}_{n,\eta}) \leqslant C'(h_n^\alpha + \Gamma_n).$$

(3) We show that the choice of the parameters m_n, h_n and λ_n gives the result.

Proof of (1) Let Y_ψ be a random variable with density ψ_{A,h_n} and independent of X, so that the distribution of $X + Y_\psi$ is $P_{\psi_{A,h_n}}$. Then, by definition of W_2,

$$W_2^2(G, P_{\psi_{A,h_n}}) \leqslant \mathbb{E}(\|X + Y_\psi - X\|_2^2) = \mathbb{E}(\|Y_\psi\|_2^2) = h_n^2 \int_{\mathbb{R}^D} \|u\|^2 \psi_{A,1}(u) du.$$

Proof of (2) If ν and μ are probability measures on $\mathbb{R}^D$ having respective densities f and g with respect to the Lebesgue measure, Lemma 1 in [8] ensures that

$$W_2^2(\nu, \mu) \leqslant 2 \min_{a \in \mathbb{R}^D} \int_{\mathbb{R}^D} \|x - a\|^2 |f(x) - g(x)| dx. \tag{41}$$

This entails

$$\begin{aligned} W_2^2(P_{\psi_{A,h_n}}, \widehat{P}_{n,\eta}) &\leqslant 2 \min_{a \in \mathbb{R}^D} \int_{\mathbb{R}^D} \|x - a\|^2 |\bar{g}(x) - c_n \widehat{g}_n^+(x) 1|_{(\widehat{\mathcal{M}})_\eta})(x)| dx \\ &\leqslant 2 \int_{(\widehat{\mathcal{M}})_\eta} \|x\|^2 |\bar{g}(x) - c_n \widehat{g}_n^+(x)| dx + 2 \int_{\mathbb{R}^D \setminus (\widehat{\mathcal{M}})_\eta} \|x\|^2 \bar{g}(x) dx. \end{aligned} \tag{42}$$

For S compact subset of $\mathbb{R}^D$, write $M_S = \sup_{x \in S} \|x\|^2$ and $\mathrm{Vol}(S)$ for the Lebesgue measure of S. Then

$$\begin{aligned} \int_{(\widehat{\mathcal{M}})_\eta} \|x\|^2 |\bar{g}(x) - c_n \widehat{g}_n^+(x)| dx &\leqslant M_{(\widehat{\mathcal{M}})_\eta} \int_{(\widehat{\mathcal{M}})_\eta} |\widehat{g}_n^+(x) - \bar{g}(x)| dx + 2M_{(\widehat{\mathcal{M}})_\eta} \frac{|c_n - 1|}{c_n} \\ &\leqslant M_{(\widehat{\mathcal{M}})_\eta} \mathrm{Vol}((\widehat{\mathcal{M}})_\eta) \Gamma_n + 2M_{(\widehat{\mathcal{M}})_\eta} \frac{|c_n - 1|}{c_n}. \end{aligned}$$

We also have

$$\begin{aligned} \frac{|c_n - 1|}{c_n} = \left| \frac{1}{c_n} - 1 \right| &= \left| \int_{(\widehat{\mathcal{M}})_\eta} (\widehat{g}_n^+(y) - \bar{g}(y)) dy - \int_{\mathbb{R}^D \setminus (\widehat{\mathcal{M}})_\eta} \bar{g}(y) dy \right| \\ &\leqslant \|\widehat{g}_n - \bar{g}\|_{L_1((\widehat{\mathcal{M}})_\eta)} + \|\bar{g}\|_{L_1(\mathbb{R}^D \setminus (\widehat{\mathcal{M}})_\eta)}. \end{aligned}$$

Using Hölder's inequality,

$$\|\widehat{g}_n - \bar{g}\|_{L_1((\widehat{\mathcal{M}})_\eta)} \leqslant \mathrm{Vol}((\widehat{\mathcal{M}})_\eta) \Gamma_n.$$

By Lemma 13, for any $\alpha > 0$, there exist C such that

$$\int_{(\widehat{\mathcal{M}})_\eta} \|x\|^2 |\bar{g}(x) - c_n \widehat{g}_n^+(x)| dx \leqslant 4M_{(\widehat{\mathcal{M}})_\eta} \mathrm{Vol}((\widehat{\mathcal{M}})_\eta)\Gamma_n + 2M_{(\widehat{\mathcal{M}})_\eta} C h_n^\alpha. \quad (43)$$

For any $c > 0$, when $\widehat{\mathcal{M}} \subset (\mathcal{M}_G)_c$, one has $(\widehat{\mathcal{M}})_\eta \subset (\mathcal{M}_G)_{\eta+c}$. This inclusion entails $M_{(\widehat{\mathcal{M}})_\eta} \leqslant M_{(\mathcal{M}_G)_{\eta+c}}$ and $\mathrm{Vol}((\widehat{\mathcal{M}})_\eta) \leqslant \mathrm{Vol}((\mathcal{M}_G)_{\eta+c})$. Therefore, for any $c > 0$,

$$\int_{(\widehat{\mathcal{M}})_\eta} \|x\|^2 |\bar{g}(x) - c_n \widehat{g}_n^+(x)| dx \leqslant 4M_{(\mathcal{M}_G)_{\eta+c}} \mathrm{Vol}((\mathcal{M}_G)_{\eta+c})\Gamma_n + 2M_{(\mathcal{M}_G)_{\eta+c}} C h_n^\alpha. \quad (44)$$

Again by Lemma 13, on the event where $\mathcal{M}_G \subset \widehat{\mathcal{M}} \subset (\mathcal{M}_G)_\eta$,

$$\int_{\mathbb{R}^D \setminus (\widehat{\mathcal{M}})_\eta} \|x\|^2 |\bar{g}(x)| dx \leqslant C' h_n^\alpha. \quad (45)$$

Finally, using (42), (44) and (45), for any $\alpha \geqslant 1$, there exists $C > 0$ such that

$$W_2^2(P_{\psi_{A,h_n}}, \widehat{P}_{n,\eta}) \leqslant C(h_n^\alpha + \Gamma_n).$$

Proof of (3) Using **(1)** and **(2)**, for sequences h_n, m_n and λ_n satisfying the assumptions of Theorem 7, on the event where $\mathcal{M}_G \subset \widehat{\mathcal{M}} \subset (\mathcal{M}_G)_\eta$, for any $\alpha \geqslant 2$, there exists $C > 0$ such that

$$W_2(G, \widehat{P}_{n,\eta}) \leqslant C(h_n + \sqrt{h_n^\alpha + \Gamma_n}) \leqslant 2C(h_n + \sqrt{\Gamma_n}).$$

We may assume $h_n \leqslant 1$ for all n without loss of generality. As stated in Lemma 12, for any $\delta' > 0$, there exist C' and n_0 such that for all $n \geqslant n_0$, with probability at least $1 - 2\exp(-n^{1/2-\delta'})$, $\sqrt{\Gamma_n} \leqslant C' e^{-m_n/4}$ and $\mathcal{M}_G \subset \widehat{\mathcal{M}} \subset (\mathcal{M}_G)_\eta$, and therefore

$$W_2(G, \widehat{P}_{n,\eta^\star}) \leqslant C m_n^{-1}$$

on this event, up to changing the constant C.

On the event of probability at most $2\exp(-n^{1/2-\delta'})$ where this doesn't hold, since the support of $\widehat{P}_{n,\eta^\star}$ is a subset of $\bar{B}(0, R_n)$, $W_2(G, \widehat{P}_{n,\eta^\star}) \leqslant 2R_n$.

Therefore, taking $\delta' < \delta$ where δ is as defined in the statement of the Theorem, there exists $C > 0$ such that for $n \geqslant n_0$,

$$\mathbb{E}_{(G*\mathbb{Q})^{\otimes n}}[W_2(G, \widehat{P}_{n,\eta^\star})] \leqslant C m_n^{-1},$$

which concludes the proof.

6.20 Proof of Theorem 8

Let $\widehat{P}_n$ be an estimator of G. According to [31],

$$\sup_{\substack{G \in St_{\mathcal{K}}(a,d,r_0) \cap \mathcal{L}(1,S,\mathcal{H}) \\ \mathbb{Q} \in \mathcal{Q}^{(D)}(\nu, c(\nu), E)}} \mathbb{E}_{(G*\mathbb{Q})^{\otimes n}}[W_p(G, \widehat{P}_n)] \geqslant \frac{1}{2} W_p(G_0(\kappa), G_1(\kappa))(1 - \|G_0(\kappa)*Q) - G_1(\kappa)*Q\|_1)^n.$$

We have shown in Theorem 5 that there exists a constant $C > 0$ such that

$$\|G_0(\kappa) * Q - G_1(\kappa) * Q\|_{TV} \leqslant \frac{C}{n},$$

taking γ of the form $c\log(n)^{-1-\delta}$ for any $\delta > 0$ and c small enough, which implies that the minimax risk is lower bounded by $W_p(G_0(\kappa), G_1(\kappa))$. We show that there exists a constant $c > 0$ and $n_0 > 0$ such that for $n \geqslant n_0$

$$W_p(G_0(\kappa), G_1(\kappa)) \geqslant c\gamma.$$

Let $\mathcal{U}_\gamma$ be the set of $u \in \mathbb{R}$ such that $|\cos(\frac{u}{\gamma})| \geqslant 1/2$, that is $\mathcal{U}_\gamma = \bigcup_{k\in\mathbb{Z}}[k\pi\gamma - \frac{\pi\gamma}{3}, k\pi\gamma + \frac{\pi\gamma}{3}]$. For each $k \in \mathbb{Z}$, let $I_{k,\gamma} := [k\pi\gamma - \frac{\pi\gamma}{2}, k\pi\gamma + \frac{\pi\gamma}{2}]$. Let us also define, for any two sets A and B of $\mathbb{R}^d$, $d(A,B) = \inf_{x\in A, y\in B} \|x-y\|_2$. We first show that

$$d(M_0(\gamma) \cap (\mathcal{U}_\gamma \times \mathbb{R}^{D-1}), M_1(\gamma)) \geqslant \gamma(\frac{\alpha}{4\sqrt{2}} \wedge \frac{\pi}{6}).$$

Let $x \in M_0(\gamma) \cap (\mathcal{U}_\gamma \times \mathbb{R}^{D-1})$ and $y \in M_1(\gamma)$. There exists $k \in \mathbb{Z}$ such that $x \in M_0(\gamma) \cap ((\mathcal{U}_\gamma \cap I_{k,\gamma}) \times \mathbb{R}^{D-1})$. If $y \in I_{k,\gamma} \times \mathbb{R}^{D-1}$ (that is, if the first coordinate of x and y are in the same interval $I_{k,\gamma}$), then

$$\|x-y\|_2 \geqslant d(M_0(\gamma) \cap ((\mathcal{U}_\gamma \cap I_{k,\gamma}) \times \mathbb{R}^{D-1}), M_1(\gamma) \cap (I_{k,\gamma} \times \mathbb{R}^{D-1})).$$

All points of $M_0(\gamma)$ are of the form $(\alpha u, \alpha u + \frac{\alpha}{2}\gamma\cos(\frac{u}{\gamma}), 0, \dots 0)^\top$ and the distance between $(\alpha u, \alpha u + \frac{\alpha}{2}\gamma\cos(\frac{u}{\gamma}), 0, \dots, 0)^\top$ and the diagonal defined by $\mathcal{D}_\alpha := \{(\alpha u, \alpha u, 0, \dots, 0)^\top : u \in \mathbb{R}\}$ is $\frac{\alpha}{4\sqrt{2}}\gamma|\cos(\frac{u}{\gamma})|$. Since the sets $M_0(\gamma) \cap ((\mathcal{U}_\gamma \cap I_{k,\gamma}) \times \mathbb{R}^{D-1})$ and $M_1(\gamma) \cap (I_{k,\gamma} \times \mathbb{R}^{D-1})$ are on opposite sides of the diagonal $\mathcal{D}_\alpha$,

$$\begin{aligned} d(M_0(\gamma) \cap ((\mathcal{U}_\gamma \cap I_{k,\gamma}) \times \mathbb{R}^{D-1}), M_1(\gamma) \cap (I_{k,\gamma} \times \mathbb{R}^{D-1})) &\geqslant d(M_0(\gamma) \cap ((\mathcal{U}_\gamma \cap I_{k,\gamma}) \times \mathbb{R}^{D-1}), \mathcal{D}_\alpha) \\ &= \frac{\alpha}{4\sqrt{2}}\gamma, \end{aligned}$$

so that $\|x-y\|_2 \geqslant \frac{\alpha}{4\sqrt{2}}\gamma$. If now $y \notin I_{k,\gamma} \times \mathbb{R}^{D-1}$,

$$\begin{aligned} d(M_0(\gamma) \cap ((\mathcal{U}_\gamma \cap I_{k,\gamma}) \times \mathbb{R}^{D-1}), M_1(\gamma) \cap ((\mathbb{R} \setminus I_{k,\gamma}) \times \mathbb{R}^{D-1})) &\geqslant d(\mathcal{U}_\gamma \cap I_{k,\gamma}, \mathbb{R} \setminus I_{k,\gamma}) \\ &= \frac{\pi\gamma}{6}, \end{aligned}$$

so that $\|x-y\|_2 \geqslant \frac{\pi\gamma}{6}$, and thus $d(M_0(\gamma) \cap (\mathcal{U}_\gamma \times \mathbb{R}^{D-1}), M_1(\gamma)) \geqslant \gamma(\frac{\alpha}{4\sqrt{2}} \wedge \frac{\pi}{6})$.

Now, let us show that $W_p(G_0(1), G_1(1)) \geqslant \gamma(\frac{\alpha}{8\sqrt{2}} \wedge \frac{\pi}{12})$. Let π be a transport plan between $G_0(1)$ and $G_1(1)$, then

$$\begin{aligned} \int_{M_0(\gamma)\times M_1(\gamma)} \|x-y\|_2^p d\pi(x,y) &\geqslant \int_{M_0(\gamma)\cap(\mathcal{U}_\gamma\times\mathbb{R}^{D-1})\times M_1(\gamma)} \|x-y\|_2^p d\pi(x,y) \\ &\geqslant d(M_0(\gamma) \cap (\mathcal{U}_\gamma \times \mathbb{R}^{D-1}), M_1(\gamma))^p \ \pi(M_0(\gamma) \cap (\mathcal{U}_\gamma \times \mathbb{R}^{D-1}) \times M_1(\gamma)) \\ &= d(M_0(\gamma) \cap (\mathcal{U}_\gamma \times \mathbb{R}^{D-1}), M_1(\gamma))^p \ G_0(1)(M_0(\gamma) \cap (\mathcal{U}_\gamma \times \mathbb{R}^{D-1})) \\ &= d(M_0(\gamma) \cap (\mathcal{U}_\gamma \times \mathbb{R}^{D-1}), M_1(\gamma))^p \ \mathbb{P}[U(1) \in \mathcal{U}_\gamma] \end{aligned}$$

since $G_1(1)$ has support $M_1(\gamma)$ and by definition of $G_0(1)$. Therefore, by taking the infimum on all transport plans between $G_0(1)$ and $G_1(1)$,

$$W_p(G_0(1), G_1(1)) \geqslant \gamma(\frac{\alpha}{4\sqrt{2}} \wedge \frac{\pi}{6})\mathbb{P}[U(1) \in \mathcal{U}_\gamma]^{1/p}.$$

$U(1)$ admit a density f_1 with respect to Lebesgue measure that is supported on $[-1,1]$ and continuous. Let us write ω one of its modulus of continuity. We have

$$\begin{aligned}
\mathbb{P}[U(1) \in \mathcal{U}_\gamma] &= \int_{[-1,1]} f_1(x) 1|_{\mathcal{U}_\gamma}(x) dx \\
&= \sum_{k \in [-1/(\pi\gamma), 1/(\pi\gamma)]} \int_{[k\pi\gamma - \frac{\pi\gamma}{3}, k\pi\gamma + \frac{\pi\gamma}{3}]} f_1(x) dx \\
&\leqslant \sum_{k \in [-1/(\pi\gamma), 1/(\pi\gamma)]} \left(\int_{[k\pi\gamma - \frac{\pi\gamma}{3}, k\pi\gamma + \frac{\pi\gamma}{3}]} f_1(k\pi\gamma) dx + \frac{2\pi\gamma}{3} \omega(\pi\gamma/3) \right) \\
&\leqslant \sum_{k \in [-1/(\pi\gamma), 1/(\pi\gamma)]} \frac{2}{3} \int_{[k\pi\gamma - \frac{\pi\gamma}{2}, k\pi\gamma + \frac{\pi\gamma}{2}]} f_1(k\pi\gamma) dx + \frac{3}{\pi\gamma} \frac{2\pi\gamma}{3} \omega(\pi\gamma/3) \\
&\leqslant \sum_{k \in [-1/(\pi\gamma), 1/(\pi\gamma)]} \frac{2}{3} \int_{[k\pi\gamma - \frac{\pi\gamma}{2}, k\pi\gamma + \frac{\pi\gamma}{2}]} f_1(x) dx + \frac{3}{\pi\gamma} \left(\frac{2\pi\gamma}{3} \omega(\pi\gamma/3) + \pi\gamma\omega(\pi\gamma/2) \right) \\
&\leqslant \frac{2}{3} \int_{[-1,1]} f_1(x) dx + 3 \left(\frac{2}{3} \omega(\pi\gamma/3) + \omega(\pi\gamma/2) \right) \\
&\underset{\gamma \to 0}{\longrightarrow} \frac{2}{3} \int_{[-1,1]} f_1(x) dx = \frac{2}{3}.
\end{aligned}$$

Therefore, there exists n_0 such that for all $n \geqslant n_0$, $W_p(G_0(1), G_1(1)) \geqslant \gamma(\frac{\alpha}{8\sqrt{2}} \wedge \frac{\pi}{12})$.

References

[1] Eddie Aamari, Catherine Aaron, and Clément Levrard. *Minimax Boundary Estimation and Estimation with Boundary*. 2021. DOI: 10.48550/ARXIV.2108.03135. URL: https://arxiv.org/abs/2108.03135.

[2] Eddie Aamari and Clément Levrard. "Nonasymptotic rates for manifold, tangent space and curvature estimation". In: *Ann. Statist.* 47.1 (2019), pp. 177–204. ISSN: 0090-5364. DOI: 10.1214/18-AOS1685. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1214/18-AOS1685.

[3] Yariv Aizenbud and Barak Sober. *Non-Parametric Estimation of Manifolds from Noisy Data*. 2021. DOI: 10.48550/ARXIV.2105.04754. URL: https://arxiv.org/abs/2105.04754.

[4] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 2000, pp. xviii+434. ISBN: 0-19-850245-1.

[5] Clément Berenfeld, Paul Rosa, and Judith Rousseau. *Estimating a density near an unknown manifold: a Bayesian nonparametric approach*. 2022. arXiv: 2205.15717 [math.ST].

[6] Jean-Daniel Boissonnat, Frédéric Chazal, and Mariette Yvinec. *Geometric and topological inference*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2018, pp. xii+233. ISBN: 978-1-108-41089-2; 978-1-108-41939-0. DOI: 10.1017/9781108297806. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1017/9781108297806.

[7] Victor-Emmanuel Brunel, Jason M. Klusowski, and Dana Yang. "Estimation of convex supports from noisy measurements". In: *Bernoulli* 27.2 (2021), pp. 772–793. ISSN: 1350-7265. DOI: 10.3150/20-bej1229. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.3150/20-bej1229.

[8] Claire Caillerie et al. "Deconvolution for the Wasserstein metric and geometric inference". In: *Electron. J. Stat.* 5 (2011), pp. 1394–1423. DOI: 10.1214/11-EJS646. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1214/11-EJS646.

[9] Jérémie Capitao-Miniconi and Élisabeth Gassiat. "Deconvolution of spherical data corrupted with unknown noise". In: *Electron. J. Stat.* (to appear).

[10] Frédéric Chazal, David Cohen-Steiner, and Quentin Mérigot. "Geometric inference for probability measures". In: *Found. Comput. Math.* 11.6 (2011), pp. 733–751. ISSN: 1615-3375. DOI: 10.1007/s10208-011-9098-0. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1007/s10208-011-9098-0.

[11] Antonio Cuevas and Ricardo Fraiman. "A plug-in approach to support estimation". In: *Ann. Statist.* 25.6 (1997), pp. 2300–2312. ISSN: 0090-5364. DOI: 10.1214/aos/1030741073. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1214/aos/1030741073.

[12] Jérôme Dedecker, Aurélie Fischer, and Bertrand Michel. "Improved rates for Wasserstein deconvolution with ordinary smooth error in dimension one". In: *Electron. J. Stat.* 9.1 (2015), pp. 234–265. DOI: 10.1214/15-EJS997. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1214/15-EJS997.

[13] Jérôme Dedecker and Bertrand Michel. "Minimax rates of convergence for Wasserstein deconvolution with supersmooth errors in any dimension". In: *J. Multivariate Anal.* 122 (2013), pp. 278–291. ISSN: 0047-259X. DOI: 10.1016/j.jmva.2013.08.009. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1016/j.jmva.2013.08.009.

[14] Sylvain Delattre and Aurélie Fischer. *Estimation via length-constrained generalized empirical principal curves under small noise.* 2019. DOI: 10.48550/ARXIV.1911.06728. URL: https://arxiv.org/abs/1911.06728.

[15] Luc Devroye and Gary L. Wise. "Detection of abnormal behavior via nonparametric estimation of the support". In: *SIAM J. Appl. Math.* 38.3 (1980), pp. 480–488. ISSN: 0036-1399. DOI: 10.1137/0138038. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1137/0138038.

[16] Vincent Divol. "Measure estimation on manifolds: an optimal transport approach". In: *Probab. Theory Related Fields* 183.1-2 (2022), pp. 581–647. ISSN: 0178-8051. DOI: 10.1007/s00440-022-01118-z. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1007/s00440-022-01118-z.

[17] Herbert Federer. "Curvature measures". In: *Trans. Amer. Math. Soc.* 93 (1959), pp. 418–491. ISSN: 0002-9947. DOI: 10.2307/1993504. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.2307/1993504.

[18] Nicolás García Trillos, Daniel Sanz-Alonso, and Ruiyi Yang. "Local regularization of noisy point clouds: improved global geometric estimates and data analysis". In: *J. Mach. Learn. Res.* 20 (2019), Paper No. 136, 37. ISSN: 1532-4435.

[19] Élisabeth Gassiat, Sylvain Le Corff, and Luc Lehéricy. "Deconvolution with unknown noise distribution is possible for multivariate signals". In: *Ann. Statist.* 50.1 (2022), pp. 303–323.

[20] Élisabeth Gassiat, Sylvain Le Corff, and Luc Lehéricy. "Supplementary to "Deconvolution with unknown noise distribution is possible for multivariate signals"". In: *Ann. Statist.* (2022). URL: https://doi.org/10.1214/21-AOS2106SUPP.

[21] Christopher R. Genovese et al. "Manifold estimation and singular deconvolution under Hausdorff loss". In: *Ann. Statist.* 40.2 (2012), pp. 941–963. ISSN: 0090-5364. DOI: 10.1214/12-AOS994. URL: https://doi-org.revues.math.u-psud.fr/10.1214/12-AOS994.

[22] Alexander Goldenshluger and Oleg Lepski. "Universal pointwise selection rule in multivariate function estimation". In: *Bernoulli* 14.4 (2008), pp. 1150–1190. ISSN: 1350-7265. DOI: 10.3150/08-BEJ144. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.3150/08-BEJ144.

[23] Robert C. Gunning and Hugo Rossi. *Analytic functions of several complex variables.* Reprint of the 1965 original. AMS Chelsea Publishing, Providence, RI, 2009, pp. xiv+318. ISBN: 978-0-8218-2165-7. DOI: 10.1090/chel/368.

[24] J. Marcinkiewicz. “Sur une propriété de la loi de Gauß”. In: *Math. Z.* 44.1 (1939), pp. 612–618. ISSN: 0025-5874. DOI: 10.1007/BF01210677. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1007/BF01210677.

[25] Alexander Meister. “Estimating the support of multivariate densities under measurement error”. In: *J. Multivariate Anal.* 97.8 (2006), pp. 1702–1717. ISSN: 0047-259X. DOI: 10.1016/j.jmva.2005.04.004. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1016/j.jmva.2005.04.004.

[26] G. W. Morgan. “A Note on Fourier Transforms”. In: *J. London Math. Soc.* 9.3 (1934), pp. 187–192. ISSN: 0024-6107. DOI: 10.1112/jlms/s1-9.3.187. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1112/jlms/s1-9.3.187.

[27] Jonathan Niles-Weed and Quentin Berthet. “Minimax estimation of smooth densities in Wasserstein distance”. In: *Ann. Statist.* 50.3 (2022), pp. 1519–1540. ISSN: 0090-5364. DOI: 10.1214/21-aos2161. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1214/21-aos2161.

[28] Nikita Puchkin and Vladimir Spokoiny. “Structure-adaptive manifold estimation”. In: *J. Mach. Learn. Res.* 23 (2022), Paper No. [40], 62. ISSN: 1532-4435.

[29] T. J. Sweeting. “On a converse to Scheffé’s theorem”. In: *Ann. Statist.* 14.3 (1986), pp. 1252–1256. ISSN: 0090-5364. DOI: 10.1214/aos/1176350065. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1214/aos/1176350065.

[30] T. Tlas. “Bump functions with monotone Fourier transforms satisfying decay bounds”. In: *J. Approx. Theory* 278 (2022), Paper No. 105742, 6. ISSN: 0021-9045. DOI: 10.1016/j.jat.2022.105742. URL: https://doi-org.ezproxy.universite-paris-saclay.fr/10.1016/j.jat.2022.105742.

[31] Bin Yu. “Assouad, Fano, and Le Cam”. In: *Festschrift for Lucien Le Cam*. Springer, New York, 1997, pp. 423–435.