Bayesian methods for the Shape Invariant Model

Dominique Bontemps[*] and Sébastien Gadat[*]

Institut Mathématiques de Toulouse, Université Paul Sabatier
118 route de Narbonne F-31062 Toulouse Cedex 9 FRANCE
e-mail: dominique.bontemps@math.univ-toulouse.fr
e-mail: sebastien.gadat@math.univ-toulouse.fr

Abstract:
In this paper, we consider the so-called Shape Invariant Model that is used to model a function f^0 submitted to a random translation of law g^0 in a white noise. This model is of interest when the law of the deformations is *unknown*. Our objective is to recover the law of the process $\mathbb{P}_{f^0,g^0}$ as well as f^0 and g^0. To do this, we adopt a Bayesian point of view and find priors on f and g so that the posterior distribution concentrates at a polynomial rate around $\mathbb{P}_{f^0,g^0}$ when n goes to $+\infty$. We then derive results on the identifiability of the SIM, as well as results on the functional objects themselves. We intensively use Bayesian non-parametric tools coupled with mixture models, which may be of independent interest in model selection from a frequentist point of view.

AMS 2000 subject classifications: Primary 62G05, 62F15; secondary 62G20.
Keywords and phrases: Grenander's pattern theory, Shape Invariant Model, Bayesian methods, Convergence rate of posterior distribution, Non-parametric estimation.

1. Introduction

This study deals with the so-called Shape Invariant Model (SIM), that describes a statistical process involving a random geometric deformation of a shape. This type of geometric deformation of a common shape corresponds to a particular case of Grenander's General Pattern Theory ([GM07]). This model may be applicable to a number of fields such as image processing ([AGP91], [PMRC10]) and medicine ([Big13]). It is also used in genetics when dealing with delayed activation curves of genes when drugs are administrated to patients, and in Chip-Seq estimations when translations in protein fixation yield randomly shifted counting processes ([MMW07] and [BGKM13]). It is found in econometrics as well when dealing with Engel curves ([BCK07]) and in landmark registration.

This model has received considerable attention in the statistical community, and most studies in this field focus on the estimation of an unknown functional object using noisy i.i.d. observations. Some studies consider a semi-parametric

approach for the estimation of the deformation parameters (*self-modeling regression framework* used by [KG88]). [Cas12] uses a Bayesian approach to obtain statistical results on the SIM in a semi-parametric setting when the level of noise on observations asymptotically vanishes. Older approaches generally focus on parametric problems (see [GM01] and the discussion therein for an overview). M-estimation is used in [BGL09] and [AKT10] proposes a stochastic algorithm to run estimations in the SIM. A recent study [CD11] used a testing strategy to obtain curve registration. Finally, [BG10] obtained minimax adaptive rates for non-parametric estimations when the law of the randomized translations is known. The SIM could be extended to more general situations of geometrical deformations described through an action of a finite dimensional Lie Group (see [BCG12]). We restricted our work to the case of the one-dimensional Lie group of translation $\mathbb{S}^1$ to warp the functional objects.

This work was inspired by several discussions about the work of [AKT10] concerning the study of the SIM. Our aim was to extend their theoretical parametric Bayesian result to non-parametric settings and to then study the behavior of some posterior distributions. We considered the general case where *both the functional shape and the probability distribution of the deformations are unknown*. This in fact corresponds to the more realistic case. To the best of our knowledge, significant statistical results have yet to be derived in this non-parametric situation.

Our paper describes the evolution of the posterior distribution *when the number of observations grows to* $+\infty$ *with a fixed noise level* σ, *with the aim of obtaining results on the estimation of* f *and* g, *which are assumed to be unknown in our work*. This is considerably different from the study of the asymptotically vanishing noise situation ($\sigma \to 0$) that is studied in [BG12]. This is itself a special feature of the SIM: there is no obvious Le Cam equivalence of experiments (see [LCY00]) between the experiments when $n \mapsto +\infty$ and when $\sigma \mapsto 0$. Very different minimax results occur in [BG10] ($n \longmapsto +\infty$) and in [BG12] ($\sigma \longmapsto 0$). Bayesian non-parametric methods are used with mixture models to obtain the contraction rates of our Bayesian procedures that rely on the important contributions of [BSW99], [GGvdV00], [GvdV01] and [GW00].

We first prove polynomial contraction rates for the posterior distribution of the law of the process generated by the SIM. This is a step toward semi-parametric results that no longer interest us from the point of view of the whole process but, instead, in terms of the common shape or in the law of deformations that underlies it. We prove identifiability, but obtain only logarithmic posterior contraction rates for them. These slower contraction rates are linked to the logarithmic frequentist lower bounds that we also present here.

This paper therefore contributes to the understanding of the SIM under a Bayesian framework. It also illustrates that there is still numerical work to be done to efficiently estimate the common shape or the law of the deformations when both are unknown.

Section 2 presents a description of the SIM, standard notations, and ends with the statements of our main results (contraction rates, identifiability and lower bounds). Section 3 gives a look at some of the challenging issues that

remain to be elucidated. Section 4 provides a metric description of the main probability spaces of the model, and Section 5 presents the proofs of the main result. Finally, Section 6 provides the proof of our semi-parametric results. A few auxiliary results are given in the appendix.

2. Model, notations and main results

2.1. Motivation

We give next a brief description of a possible application that present the main features of the SIM, even though our primary interest is mainly theoretical. Note that although in this study, simulation results are not included, they are nevertheless of major interest for applications. However, development of efficient sampling algorithms is beyond the initial scope of this paper and requires further advances on stochastic algorithms (MCMC and Langevin diffusions).

An important application of the mean shape computation is the estimation of the average cardiac cycle using electrocardiogram (ECG) records. An ECG records the electrical activity of the heart and consists of a succession of similar cycles. An example taken from the MIT-BIH database [GR01] is shown at the top of Fig. 1. The estimation of the mean shape of the successive cycles is useful for the diagnosis of heart disease, and among them, arrhythmia. There are several types of arrhythmia and it is important to determine the variant that is affecting the patient, which can be determined on the basis of an ECG. Unfortunately, segmentation of the ECG records is a non-trivial task, as pointed out by [GK95], and is generally based on feature detections such as maximum, inflexion points, etc. A segmentation of cycles using the crossing level 0 of the ECG [1] is shown in Fig. 1 and it can be observed that this segmentation is far from satisfactory. This phenomenon is even worse when dealing with arrhythmic cycles, which are more irregular than normal cycles and essential for establishing a medical diagnosis. In order to eliminate spacial fluctuations around the mean shape, it is possible to attempt to average the cycles, but as indicated in Fig. 1, the averaging step produces a type of Tichonov regularization. This inverse problem phenomenon occurs as a result of the errors in the alignment procedure, and the law of these warping errors produces a convolution kernel that has a negative impact on the estimation of the mean shape.

The question arises as to whether or not it is possible to produce a sharper warping procedure that perfectly recovers the unknown random deformations between each individual cycle with an asymptotically vanishing warping error. Indeed, as shown by Theorem 5 of [BG10], this last search is useless for smooth signals when increasing the number of observations ($n \longrightarrow +\infty$), and is only possible when the resolution of the recorded signals is increasing (the level of the noise should asymptotically vanish). Even worse, statistical estimation of the true shape with any warping estimator fails and cannot be consistent for
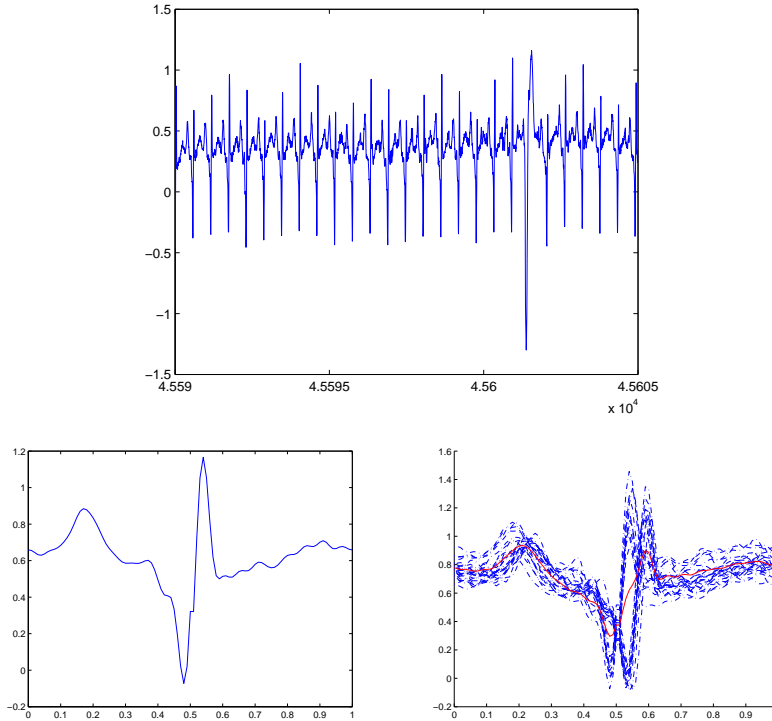
FIG 1. *Top: Original time series "chfdb chf13 45590" of the ECG dataset; Bottom left: One individual cycle; Bottom right: whole cycles (in dashed blue) and the average cycle (in red).*

regular shapes when the number of samples is growing to $+\infty$.

In the following section, we therefore attempt to obtain an estimation of the common shapes with a learning procedure of the deformation law. We are interested in the asymptotic case where the resolution level is kept fixed, since it is not possible to imagine reducing the noise level to 0 in many practical problems (*i.e.*, below the precision of the recording machine), although it may be possible to multiply the observations (for example, by increasing the time of ECG recordings).

2.2. Statistical settings

Shape Invariant Model   We recall here the definition of the Shape Invariant Model. Let $f^0 \in \mathcal{F}$ be a subset of smooth functions and consider a probability measure $g^0 \in \mathfrak{M}([0,1])$ (this last set stands for the set of probability measures on

---

[1]Indeed, the original data is translated by $+$ 1.5 millivolts for the sake of convenience and the electrical activity is always non negative!

$[0, 1]$). We observe n realizations of noisy and randomly shifted complex valued curves $Y_1, \ldots, Y_n$ resulting from the following white noise model:

$$\forall x \in [0, 1] \quad \forall j = 1 \ldots n \qquad dY_j(x) := f^0(x - \tau_j)dx + \sigma dW_j(x). \qquad (2.1)$$

Here, f^0 is the *mean* pattern of the curves $Y_1, \ldots, Y_n$. The random shifts $(\tau_j)_{j=1\ldots n}$ are sampled independently according to the probability measure g^0, and $(W_j)_{j=1\ldots n}$ are independent complex standard Brownian motions on $[0, 1]$, which models the presence of noise in the observations. The noise level σ is kept fixed in our study and is set to 1 for the sake of simplicity. Note that in our white noise model, σ can be directly obtained from the quadratic variation of each individual curve, and is therefore known. Complex valued curves are considered here to simplify notations but our results can be adapted to the real-valued case. A complex standard Brownian motion W_t on $[0, 1]$ is such that W_1 is a standard complex Gaussian random variable whose distribution is designated by $\mathcal{N}_{\mathbb{C}}(0, 1)$. A standard complex Gaussian random variable has independent real and imaginary parts distributed according to $\mathcal{N}_{\mathbb{R}}(0, \frac{1}{2})$.

We aim to describe the behavior of posterior distributions given a sample $(Y_1, \ldots, Y_n)$ when $n \to +\infty$. We intensively use "$\lesssim$", which refers to an inequality up to a multiplicative absolute constant, and $a \sim b$, which means that $a/b \longrightarrow 1$.

2.3. Notations and Definitions

Bayesian framework We briefly recall the Bayesian formalism following the presentation of [GGvdV00] but readers familiar with it can skip this paragraph.

The functional objects f^0 and g^0 that we are looking for belong to $\mathcal{F} \otimes \mathfrak{M}([0, 1])$ and for any couple $(f, g) \in \mathcal{F} \otimes \mathfrak{M}([0, 1])$. Equation (2.1) describes the law of one continuous curve. Its law is designated as $\mathbb{P}_{f,g}$ and possesses a density $p_{f,g}$ with respect to the Wiener measure on the sample space. $\mathbb{P}_{f^0,g^0}$ belongs to a set $\mathcal{P}$ of probability measures over the sample space. This set $\mathcal{P}$ is the set of all possible measures described by Equation (2.1) when (f, g) varies over $\mathcal{F} \otimes \mathfrak{M}([0, 1])$. Given a prior distribution Π_n on $\mathcal{P}$, Bayesian procedures are based on the posterior:

$$\Pi_n \left(B|Y_1, \ldots, Y_n\right) = \frac{\int_B \prod_{j=1}^n p(Y_j)d\Pi_n(p)}{\int_{\mathcal{P}} \prod_{j=1}^n p(Y_j)d\Pi_n(p)},$$

which is a random measure on $\mathcal{P}$ that depends on the observations $Y_1, \ldots, Y_n$. Bayesian estimators can then be obtained using the mode, the mean or the median of the posterior distribution (the approach, for example, adopted in [AKT10]).

The posterior distribution is then said to be *consistent* if it concentrates on arbitrarily small neighborhoods of $\mathbb{P}_{f^0,g^0}$ in $\mathcal{P}$ with a probability tending to 1 when n grows to $+\infty$. If d is a distance on probability measures (*e.g.*,

the Hellinger distance d_H or the Total Variation d_{TV}), a frequentist rate ϵ_n corresponds to:

$$\Pi_n \left(\mathbb{P}_{f,g} : d(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f,g}) \geq M\epsilon_n | Y_1, \dots Y_n \right) \longrightarrow 0 \quad \mathbb{P}_{f^0,g^0} \, a.s. \quad \text{as} \quad n \longrightarrow +\infty.$$

A *frequentist* property of such a posterior distribution describes the contraction rate $(\epsilon_n)_{n \geq 0}$ of such neighborhoods while still capturing most of the posterior mass. We refer to [GGvdV00] (especially their Theorem 2.1) for a complete discussion on the posterior concentration rates, as well as the links with classical non-parametric frequentist benchmarks (see [IH81]). Using Equation (2.1), we attempt to tackle such a Bayesian property.

Fourier analysis We assume that f^0 belongs to $\mathcal{F} \subset L^2_{\mathbb{C}}([0,1])$, endowed with the Euclidean norm $\|h\|^2 := \int_0^1 |h(s)|^2 ds$. Each element $h \in L^2_{\mathbb{C}}([0,1])$ may be extended to a 1-periodic function on $\mathbb{R}$. f^0 is assumed to be 1-periodic, and $\mathfrak{i}$ will stand for the complex number so that $\mathfrak{i}^2 = -1$. Fourier coefficients of h are:

$$\theta_\ell(h) := \int_0^1 e^{-\mathfrak{i}2\pi\ell t} h(t) dt. \tag{2.2}$$

We often use the parameterization of any element of $h \in L^2_{\mathbb{C}}([0,1])$ through its Fourier expansion and will use the shortcut $(\theta_\ell)_{\ell \in \mathbb{Z}}$ instead of $(\theta_\ell(h))_{\ell \in \mathbb{Z}}$. We are interested in the SIM for smooth functions and balls of Sobolev spaces with a smoothness parameter s, when $s \geq 1$:

$$\mathcal{H}_s(A) := \left\{ f \in L^2_{\mathbb{C}}([0,1]) \, : \, \sum_{\ell \in \mathbb{Z}} (1 + |\ell|^{2s}) |\theta_\ell(f)|^2 < A^2 \right\}, \|\theta\|_{\mathcal{H}_s} := \sqrt{\sum_{\ell \in \mathbb{Z}} |\ell|^{2s} |\theta_\ell|^2}.$$

The notation $\mathcal{H}_s$ refers to the whole Sobolev subspace ($A = +\infty$). Finally, given any integer ℓ, we define the "thresholded" elements of $\mathcal{H}_s$ as:

$$\mathcal{H}^\ell := \left\{ f \in L^2_{\mathbb{C}}([0,1]) \quad | \quad \forall |k| > \ell \quad \theta_k(f) = 0 \right\}.$$

Mixture models and probability distributions Equation (2.1) can be written in the Fourier domain as:

$$\forall \ell \in \mathbb{Z} \quad \forall j \in \{1 \dots n\} \qquad \theta_\ell(Y_j) = \theta_\ell^0 e^{-\mathfrak{i}2\pi\ell\tau_j} + \xi_{\ell,j},$$

where $\theta^0 := (\theta_\ell(f^0))_{\ell \in \mathbb{Z}}$ denotes the true unknown Fourier coefficients of f^0, and $\theta_\ell(Y_j)$ those of the observed signals. The white noise model implies that $\xi_{\ell,j} \sim_{i.i.d.} \mathcal{N}_{\mathbb{C}}(0,1)$ for any ℓ, j. The density of the complex Gaussian random variables is designated as:

$$\gamma_\mu(z) := \pi^{-p} e^{-\|z-\mu\|^2}, \forall z \in \mathbb{C}^p,$$

and will be referred to as γ when $\mu = 0$ and $p = 1$. For any $\ell \in \mathbb{Z}$, Equation (2.1) implies that $\theta_\ell(Y) \sim \int_0^1 \gamma_{\theta_\ell^0 e^{-\mathfrak{i}2\pi\ell\varphi}}(\cdot) dg(\varphi)$, which is a mixture law of Gaussian

random variables. We use, for any phase $\varphi \in [0,1]$ and any $\theta \in \ell^2(\mathbb{Z})$, the notation:

$$\forall \ell \in \mathbb{Z} \qquad (\theta \bullet \varphi)_\ell := \theta_\ell e^{-\mathrm{i}2\pi\ell\varphi}.$$

This corresponds to a rotation of each coefficient θ_ℓ with an angle $2\pi\ell\varphi$. Hence

$$\theta(Y) \sim \mathbb{P}_{\theta^0,g} := \int_0^1 \gamma_{\theta^0 \bullet \varphi}(.)dg(\varphi).$$

From one frequency to another, the rotations used in $\theta(Y)$ *are not independent*, and the coefficients $(\theta_\ell(Y))_{\ell \in \mathbb{Z}}$ are highly correlated. When θ has a length k, $p_{\theta,g}$ will be the density w.r.t. the Lebesgue measure on $\mathbb{C}^k$ of the law $\mathbb{P}_{\theta,g}$:

$$\forall z \in \mathbb{C}^k \qquad p_{\theta,g}(z) := \int_0^1 \gamma(z - \theta \bullet \varphi)dg(\varphi).$$

We will use standard objects (see [vdVW96]) such as the Hellinger (d_H) or Total Variation (d_{TV}) distances between probability measures, as well as covering numbers of metric spaces such as $D(\epsilon, \mathcal{P}, d)$, bracketing numbers $N_{[]}(\epsilon, \mathcal{P}, d)$, etc.

2.4. Bayesian priors in the randomly shifted curves model

We detail here the prior Π_n on $\mathcal{P}$, defined on $\mathcal{H}_s \otimes \mathfrak{M}([0,1])$. Equation (2.1) will convert it to a prior on $\mathcal{P}$. We choose f and g independently as follows.

Prior on f The prior on f is adapted from [RR12] and is defined on $\mathcal{H}_s$ by:

$$\pi := \sum_{\ell \geq 1} \lambda(\ell)\pi_\ell.$$

Given any integer ℓ, we randomly switch on the Fourier frequencies in $[\![-\ell, +\ell]\!]$ with probability $\lambda(\ell)$, and π_ℓ is a distribution where $\pi_\ell := \otimes_{k \in \mathbb{Z}} \pi_\ell^k$ with:

$$\forall k \in \mathbb{Z} \qquad \pi_\ell^k = 1_{|k|>\ell}\delta_0 + 1_{|k|\leq\ell}\mathcal{N}_{\mathbb{C}}(0, \xi_n^2).$$

The selected frequencies are chosen with a distribution λ on $\mathbb{N}^\star$, which satisfies:

$$\exists \rho \in (1,2) \quad \exists c_1 \geq c_2 > 0 \quad \forall \ell \in \mathbb{N}^\star \qquad e^{-c_1\ell^2 \log^\rho \ell} \lesssim \lambda(\ell) \lesssim e^{-c_2\ell^2 \log^\rho \ell}.$$

The prior π depends on the variance of the Gaussian laws ξ_n used to sample the Fourier coefficients. We use a sample size-dependent variance (see Remark 5.1):

$$\xi_n^2 := n^{-\mu_s}(\log n)^{-\zeta}, \tag{2.3}$$

where μ_s and ζ may depend on s (non-adaptive prior) or not (adaptive prior).

Prior on g We propose in the sequel two priors on g.

Dirichlet Process prior As pointed out above, the model can be seen as a Gaussian mixture model, and natural priors on g may be built according to a Dirichlet Process (see [GvdV01]). Given any absolutely continuous measure α w.r.t. the Lebesgue measure on $[0,1]$ with a positive continuous density, the Dirichlet Process DP_α generates a random probability measure $g \in \mathfrak{M}([0,1])$. For any finite partition $(A_1, \ldots, A_k)$ of $[0,1]$, the vector $(g(A_1), \ldots, g(A_k))$ on the simplex has a Dirichlet distribution $Dir(\alpha(A_1), \ldots, \alpha(A_k))$ (see details in [Fer73]).

Non-adaptive Gaussian Process Prior We may use another prior to obtain smoothness results on g, and then extend our result further than a simple contraction on *laws*. Smoothness is barely compatible with D.P., and even a kernel convolution with D.P. seems problematic in our situation. To get around this difficulty, we propose the use of Gaussian Processes. In this construction, we assume that $g^0 \in \mathcal{H}_\nu(A)$, *where $\nu \geq 1/2$ and A are both known* and define the corresponding subspace:

$$\mathfrak{M}_\nu([0,1])(A) := \mathfrak{M}([0,1]) \cap \mathcal{H}_\nu(A).$$

We define the integer $k_\nu := \lfloor \nu - 1/2 \rfloor$ to be the largest integer smaller than $\nu - 1/2$ and follow the strategy in Section 4 of [vdVvZ08a], but the additional difficulty is to guarantee the periodicity of g. Its construction is inspired from that of the Brownian bridge $(B_t)_{t \in [0,1]}$, which is obtained as $B_t = W_t - tW_1$ (where W is a standard Brownian motion). For any continuous function f on $[0,1]$, we define:

$$J(f) : t \longmapsto \int_0^t f(s)ds - t \int_0^1 f(u)du$$

and the recursive compositions $J_k = J_{k-1} \circ J$. In order to fit our prior to the values of derivatives of g on the extremal points $\{0,1\}$, we introduce $(\psi_j)_{j=1\ldots k_\nu}$:

$$\forall t \in [0,1] \qquad \psi_j(t) := \sin(2\pi jt) + \cos(2\pi jt).$$

Our prior is now built as follows. We first independently sample a real Brownian bridge $(B_\tau)_{\tau \in [0,1]}$ and $(Z_1, \ldots Z_{k_\nu}) \sim_{i.i.d.} \mathcal{N}_\mathbb{R}(0,1)$. We next compute:

$$\forall \tau \in [0,1] \qquad w_\tau := J_{k_\nu}(B)(\tau) + \sum_{i=1}^{k_\nu} Z_i \psi_i(\tau). \tag{2.4}$$

Given $(w_\tau, \tau \in [0;1])$ generated by Equation (2.4), we build p_w through:

$$\forall \tau \in [0;1] \qquad p_w(\tau) := \frac{e^{w_\tau}}{\int_0^1 e^{w_\tau} d\tau}. \tag{2.5}$$

Hence, this prior based on GP yields a prior on densities on $[0;1]$ and p_w inherits the smoothness k_ν of the Gaussian process $\tau \mapsto w_\tau$. We now restrict this prior to the Sobolev balls of radius $2A$ to obtain a prior $q_{\nu,A}$ on $\mathfrak{M}_\nu([0,1])(2A)$.

In the sequel, we designate Π_n^{DP} (resp. Π_n^{GP}) for the posterior distribution built from the prior obtained with the Dirichlet Process (resp. the Gaussian Process).

2.5. Main results

2.5.1. Contraction rates

Dirichlet Process Prior

Theorem 2.1. *Assume that $f^0 \in \mathcal{H}_s$ with $s \geq 1/2$, then $\mu_s = 1/(2s+2)$ and $\zeta = 0$ in the definition of ξ_n yield a non adaptive prior: for any $g^0 \in \mathfrak{M}([0,1])$ and for a sufficiently large M,*

$$\Pi_n^{DP} \left\{ \mathbb{P}_{f,g} : d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \geq M\epsilon_n | Y_1, \ldots Y_n \right\} \longrightarrow 0 \, \mathbb{P}_{f^0,g^0} \, a.s. \, as \, n \longrightarrow +\infty.$$

Moreover, the contraction rate ϵ_n is given by $\epsilon_n = n^{-s/(2s+2)} \log n$.

The values $\mu = 1/3$ (independent on s) and $\zeta = 3/2$ yield the following for a sufficiently large M:

$$\Pi_n^{DP} \left\{ \mathbb{P}_{f,g} : d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \geq M\epsilon_n | Y_1, \ldots Y_n \right\} \longrightarrow 0 \, \mathbb{P}_{f^0,g^0} \, a.s. \, as \, n \longrightarrow +\infty.$$

where $\epsilon_n = \begin{cases} n^{-s/(2s+2)} \log n & if \quad s \in [1/2, 2], \\ n^{-1/3} \log n & if \quad s \geq 2. \end{cases}$

This result describes the posterior concentration around some Hellinger neighborhoods of $\mathbb{P}_{f^0,g^0}$ within a polynomial rate.

Our prior when $\mu = 1/3$ is adaptive to the regularity s as soon as $s \in [1/2, 2]$, where $\xi_n^2 = n^{-1/3}(\log n)^{-3/2}$. In this case, the convergence rate is $n^{-s/(2s+2)}$ up to a log term. When $s \geq 2$, the above choice for ξ_n^2 is no longer optimal. However, taking $\mu < 1/3$ yields an adaptive prior on the range $s \in [\frac{1}{2\mu} - 1; \frac{1}{\mu} - 1]$. With this last prior, the concentration is no longer guaranteed when $s < \frac{1}{2\mu} - 1$, and for $s > \frac{1}{\mu} - 1$ the rate is stuck to $n^{-(1-\mu)/2}$.

Finally, the non-adaptive prior based on $\xi_n^2 = n^{-1/(2s+2)}$ recovers the rate $-s/(2s+2)$ for any s larger than $1/2$. To the best of our knowledge, the minimax frequentist rate is unknown when both f^0 and g^0 are unknown. Remind that in the standard regression model (without shift) on $\mathcal{H}_s$, the minimax rate of estimation for f is $n^{-s/(2s+1)}$. The rate we obtain in Theorem 2.1 is slightly damaged owing to our upper bound of the statistical complexity of the model coming from the estimation of g, see Proposition 4.2

Gaussian Process Prior

Theorem 2.2. *Assume that $f^0 \in \mathcal{H}_s$ with $s \geq 1/2$ and $g^0 \in \mathfrak{M}_\nu([0,1])(A)$ with $\nu \geq 1/2$, then:*

$$\Pi_n^{GP} \left\{ \mathbb{P}_{f,g} : d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \geq M\epsilon_n | Y_1, \ldots Y_n \right\} \longrightarrow 0 \, \mathbb{P}_{f^0,g^0} \, a.s. \, as \, n \longrightarrow +\infty.$$

for a sufficiently large M, with $\epsilon_n = n^{-\left[\frac{\nu(s \wedge 1)}{2\nu(s \wedge 1)+1} \wedge \frac{s}{2s+2}\right]} \log(n)^\kappa$, for an explicit κ.

We still obtain a polynomial contraction rate here, but the smoothness of g has a great impact on the rate $(\epsilon_n)_{n \geq 0}$. This bound is damaged when ν is close to $1/2$, which is its lower limiting value, and we then obtain a contraction rate of $n^{-1/4}$ if $s \geq 1$. In a sense, Theorem 2.2 is weaker than Theorem 2.1, but it will make it possible to extend our mathematical analysis further since it guarantees that the marginal on the g coordinate of the posterior distribution Π_n^{GP} is supported by smooth densities.

2.5.2. Identifiability

We then attempt to derive results on the objects $f \in \mathcal{H}_s$ and $g \in \mathfrak{M}_\nu([0,1])$ themselves. Hence, a crucial step is to guarantee the identifiability of the model, which is illustrated below. Without any constraint on the first Fourier coefficient, it can be easily checked that identifiability fails. We then restrict our study to:

$$\mathcal{F}_s := \left\{ f \in L_{\mathbb{C}}^2([0,1]) \quad | \quad \theta_1(f) > 0 \quad \text{and} \quad \sum_{\ell \in \mathbb{Z}} (1 + |\ell|^{2s})|\theta_\ell(f)|^2 < +\infty \right\},$$

and the prior used on $\mathcal{F}_s$ is the one induced by π, up to the restriction that the first Fourier coefficient is positive. We constrain $g \in \mathfrak{M}([0,1])^\star$ (subsequently converted to an inverse problem assumption):

$$\mathfrak{M}([0,1])^\star := \{ g \in \mathfrak{M}([0,1]) \quad | \quad \forall k \in \mathbb{Z} \quad \theta_k(g) \neq 0 \}.$$

Theorem 2.3. *The SIM is identifiable as soon as $(f^0, g^0) \in \mathcal{F}_s \times \mathfrak{M}([0,1])^\star$:*

$$\mathcal{I} \quad : \quad (f^0, g^0) \in \mathcal{F}_s \times \mathfrak{M}([0,1])^\star \longmapsto \mathbb{P}_{f^0,g^0} \qquad \text{is injective.}$$

Indeed, the necessary assumption to yield identifiability is slightly more general than the restriction to $\mathfrak{M}([0,1])^\star$. If we define $\mathfrak{M}([0,1])^+$ as follows:

$$\mathfrak{M}([0,1])^+ := \{ g \in \mathfrak{M}([0,1]) \quad | \quad \forall k \in \mathbb{Z}, \exists j_k \in \mathbb{Z} : \theta_{k.j_k}(g) \neq 0 \}.$$

We will see in the proof of Theorem 2.3 that the set $\mathcal{F}_s \times \mathfrak{M}([0,1])^+$ is the minimal set for the identifiability of the model.

2.5.3. Contraction rate on functional objects

Theorem 2.4. *(i) Assume $f^0 \in \mathcal{F}_s$ with $s \geq 1/2$, and $g^0 \in \mathfrak{M}_\nu([0,1])$ with $\nu > 1$:*

$$\Pi_n^{GP} \left\{ g : \|g - g^0\| \geq M\mu_n | Y_1, \dots Y_n \right\} \longrightarrow 0 \, \mathbb{P}_{f^0,g^0} \, a.s. \, as \, n \longrightarrow +\infty$$

for a sufficiently large M, with a contraction rate of $\mu_n = (\log n)^{-\nu}$.

(ii) Assume that $g^0 \in \mathfrak{M}_\nu([0,1])$ satisfies the inverse problem assumption:

$$\exists c > 0 \quad \exists \beta > \nu + \tfrac{1}{2} \quad \forall k \in \mathbb{Z} \qquad |\theta_k(g^0)| \geq ck^{-\beta}$$

then we also have:

$$\Pi_n^{GP} \left\{ f : \|f - f^0\| \geq M\tilde{\mu}_n | Y_1, \ldots Y_n \right\} \longrightarrow 0 \, \mathbb{P}_{f^0, g^0} \, a.s. \, as \, n \longrightarrow +\infty$$

where the contraction rate $\tilde{\mu}_n$ is $\tilde{\mu}_n = (\log n)^{-\frac{2s\nu}{2s+2\beta+1}}$.

The optimality of this result is, to the best of our knowledge, an open problem. Note that when β increases, the rate on f is seriously impacted, which is a common feature of statistical inverse problems. This result is consistent with the fact that estimating f when g is known is an inverse problem (see [BG10]). Our next result shows that it is impossible to obtain frequentist convergence rates better than a power of $\log n$, even if our lower bound does not exactly match the upper bound obtained in the previous result.

2.5.4. Link with heteroscedastic deconvolution with unknown variance

Our problem seems strongly related to standard deconvolution with unknown variance setting. For instance, the first Fourier coefficients are

$$\theta_1(Y_j) = \theta_1 e^{-i2\pi\tau_j} + \epsilon_{1,j}, \forall j \in \{1 \ldots n\}$$

and up to a division by θ_1, it can also be parametrised as

$$\tilde{\theta}_1(Y_j) = e^{-i2\pi\tau_j} + \frac{\epsilon_{1,j}}{\theta_1}, \forall j \in \{1 \ldots n\}, \tag{2.6}$$

which is similar to the problem

$$Y_i = X_i + \epsilon_i, \forall i \in \{1 \ldots n\} \qquad (X_i)_{i=1\ldots n} \sim g, \tag{2.7}$$

studied for instance by [Mat02] where $\epsilon \sim \mathcal{N}_\mathbb{C}(0, \sigma^2)$ whose variance σ^2 is unknown. As pointed in [Mat02] (see also the more recent work [BM05] where similar situations are extensively detailed), such a particular setting is unfavourable for statistical estimation since convergence rates are generally of log order.

Results obtained in [Mat02] and [BM05] are obtained using the van Trees inequality, which is a Bayesian Cramer-Rao bound (see for instance [GL95] for further details). However, Proposition 6.1 gives a polynomial rate for the posterior contraction around θ_1^0 and this rate, at the first glance, seems contradictory with the results given by [Mat02]. Indeed, [Mat02] considers some lower bounds in a larger class than the estimation problem of θ_1 written as (2.6): from a minimax point of view, the supremum over all hypotheses is taken in a somewhat larger set than ours. Moreover, if one considers (2.6), the density of $e^{-i2\pi\tau_j}$ is supported by $\mathbb{S}^1$ instead of the whole complex plane which would be a natural

extension of (2.7). Hence, g is a singular measure with respect to the noise measure and the ability of going beyond the log rate is due to the degeneracy nature of our problem according to the Gaussian complex noise on θ_1. It is an important structural information which is not available when one considers general problems such as (2.7). A new proof is thus needed for the SIM.

2.5.5. Frequentist lower bounds

Theorem 2.5. *Assume that $(f^0, g^0) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0,1])$, then a sufficiently small c exists so that the minimax rate over $\mathcal{F}_s \times \mathfrak{M}_\nu([0,1])$ satisfies:*

$$\liminf_{n \longrightarrow +\infty} (\log n)^{2s+2} \inf_{\hat{f} \in \mathcal{F}_s} \sup_{(f,g) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0,1])} \|\hat{f} - f\|^2 \geq c,$$

and

$$\liminf_{n \longrightarrow +\infty} (\log n)^{2\nu+1} \inf_{\hat{g} \in \mathfrak{M}_\nu([0,1])} \sup_{(f,g) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0,1])} \|\hat{g} - g\|^2 \geq c.$$

This result is far from being contradictory with the polynomial rate obtained in Theorems 2.1 and 2.2. A series of remarks can be made. First, Theorems 2.1 and 2.2 provide contraction rates on the probability distributions in $\mathcal{P}$ and not on the functional space $\mathcal{F}_s$. Second, the link between (f^0, g^0) and $\mathbb{P}_{f^0, g^0}$ relies on the identifiability, and the lower bound is derived with a set of functions $(f_i, g_i)_i$, which are very hard to identify through $\mathcal{I} : (f, g) \mapsto \mathbb{P}_{f,g}$. On this set of functions, the injection is very "flat" and the two-by-two differences of $\mathcal{I}(f^i, g^i)$ are small, so that the pairs of functions (f^i, g^i) become very hard to distinguish. It should be noted that a typical feature of this set is that the convolution products are very similar each others: $f^i \star g^i \simeq f^j \star g^j, i \neq j$. Typically, hard estimations occur when faced with a deconvolution statistical problem with an unknown operator. It is shown in [BG10] that in the SIM, when $n \longrightarrow +\infty$, it is impossible to recover the unknown true shifts. The abrupt degradation between the polynomial rates on probability laws in $\mathcal{P}$ and the logarithmic rates on functional objects in $\mathcal{F}_s \times \mathfrak{M}_\nu([0,1])$ also occurs because of this.

3. Discussion

In this paper, we exhibit a suitable prior that makes it possible to obtain a contraction rate of the posterior distribution near the true underlying distribution $\mathbb{P}_{f^0, g^0}$. Moreover, this rate is polynomial with the number n of observations, even if our SIM is an inverse problem with an unknown translation operator that depends on g. From a technical point of view, the milestones of such results are the tight link between the white noise model and the Fourier expansion, as well as the smoothness of Gaussian law that makes it possible to obtain an efficient covering strategy. Up to a non-restrictive condition, we also obtain a large identifiability class, but the contraction of the posterior is dramatically damaged in this class since we then obtain a logarithm rate instead of a polynomial one. This last point cannot be much improved by using the standard L^2 distance to

measure the neighborhoods of f^0, as pointed out by our last lower bound. Note that we do not obtain exactly the same rates for our lower and upper reconstruction bounds. This may be due to the rough inequality $|\psi_a(\varphi)| \geq \frac{|\psi_a(\varphi)|^2}{\|\psi_a\|_\infty}$ used to obtain Equation (6.1), and may be the reason why we do not obtain optimal rates. Indeed, the degradation of the contraction rate occurs when attempting to invert the identifiability map $\mathcal{I} : (f, g) \mapsto \mathbb{P}_{f,g}$. This difficulty should be understood as a novel consequence of the impossibility to exactly recover the random shift parameters when only n grows to $+\infty$. This phenomenon is highlighted in several papers, including [BG10] or [BGKM13]. However, it may be possible to obtain a polynomial rate using a more appropriate distance adapted to our problem of randomly shifted curves, for instance

$$d_{\text{Fréchet}}(f_1, f_2) := \inf_{\tau \in [0,1]} \|f_1(. - \tau) - f_2(.)\|.$$

We plan to tackle this problem in a future study. The important requirement in this case is to find some relationships between the neighborhoods of $\mathbb{P}_{f^0,g^0}$ and the neighborhoods of f^0 according to the distance $d_{\text{Fréchet}}$. Another interesting study would consider the SIM with a noise level σ depending on n in the Bayesian framework. This asymptotic setting is linked to the work of [BG12] in which their J curves are sampled at the n points of a discrete design in $[0, 1]$. Finally, an open issue that still remains to be elucidated concerns the search for a stochastic algorithm to approach the posterior distribution in our SIM.

4. Metric description of the model

We follow the roadmap of Theorem 2.1 of [GGvdV00] by the introduction of a suitable sieve $\mathcal{P}_{\ell_\epsilon, w_\epsilon}$ defined in Section 4.1.2 (see also [SW01] and [Zha00]), and find optimal calibrations of ϵ, ℓ_ϵ and w_ϵ with respect to n. We then need to find a lower bound of the prior mass around a type of Kullback-Leibler neighborhood of $\mathbb{P}_{f^0,g^0} \in \mathcal{P}$. These sets are defined as

$$\mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0,g^0}, d_{KL}) = \left\{ \mathbb{P}_{f,g} \in \mathcal{P} | d_{KL}(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f,g}) \leq \epsilon_n^2, V(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f,g}) \leq \epsilon_n^2 \right\}.$$

We will indeed consider Hellinger neighborhoods instead of Kullback-Leibler ones (a link is given in Section A). In Section 4.2, we exhibit suitable Hellinger neighborhoods.

4.1. Entropy estimates

4.1.1. Covering numbers for fixed f

When f belongs to the Sobolev space $\mathcal{H}_s$, for $s \geq 1/2$, described by the infinite vector of Fourier coefficients $(\theta_k)_{k \in \mathbb{Z}}$, the set of all possible Gaussian measures is denoted $\mathcal{A}_\theta := \{\gamma_{\theta \bullet \varphi}, \varphi \in [0, 1]\}$. We establish the preliminary result.

Proposition 4.1. *For any sequence $f \in \mathcal{H}_s$ for $s \geq 1/2$, with Fourier coefficients θ:*

i) $d_{TV}(\mathbb{P}_{f,\delta_{\phi_1}}, \mathbb{P}_{f,\delta_{\phi_2}}) \lesssim |\phi_2 - \phi_1|^{s \wedge 1} \|f\|_{\mathcal{H}_{s \wedge 1}}.$
ii) $\log N(\epsilon, \mathcal{A}_\theta, d_H) \lesssim \log \|\theta\|_{\mathcal{H}_{1/2}} + \log \frac{1}{\epsilon}.$

Proof. Remark that for any dimension p and any couple of points $(z_1, z_2) \in \mathbb{C}^p$, if $\|z_1 - z_2\|$ is the Euclidean distance in $\mathbb{C}^p$, then one has

$$d_{TV}(\gamma_{z_1}, \gamma_{z_2}) = \frac{1}{2}\|\gamma_{z_1} - \gamma_{z_2}\|_{L_1} = \left[2\Phi\left(\frac{\|z_1 - z_2\|}{2}\right) - 1\right] \leq \frac{\|z_1 - z_2\|}{\sqrt{2\pi}},$$

where Φ is the c.d.f. of a real standard Gaussian variable. Now, we compute

$$\begin{aligned}
d_{TV}(\mathbb{P}_{f,\delta_{\phi_1}}, \mathbb{P}_{f,\delta_{\phi_2}}) = d_{TV}(\mathbb{P}_{\gamma_\theta}, \mathbb{P}_{\gamma_{\theta \bullet (\phi_2 - \phi_1)}}) &\lesssim \|\theta - \theta \bullet (\phi_2 - \phi_1)\| \\
&\lesssim \sqrt{\sum_{k \in \mathbb{Z}} |\theta_k|^2 |1 - e^{i2\pi k(\phi_2 - \phi_1)}|^2} \\
&\leq \sqrt{\sum_{k \in \mathbb{Z}} 2^{(1-2s)\vee(-1)}\pi |k(\phi_2 - \phi_1)|^{2(s \wedge 1)}|\theta_k|^2} \\
&\lesssim |\phi_2 - \phi_1|^{s \wedge 1} \|f\|_{\mathcal{H}_{s \wedge 1}}
\end{aligned}$$

Now, an ϵ-covering of $\mathcal{A}_\theta$ is simply obtained by covering the tore with intervals of radius $(C\|f\|_{\mathcal{H}_{1/2}}/\epsilon)^2$, which leads to the announced metric entropy. $\quad\square$

We next consider a continuous mixture for g, which is the natural case. Define

$$\mathcal{P}_f := \{\mathbb{P}_{f,g} \mid g \in \mathfrak{M}([0,1])\}.$$

Now, we will only consider functions f with null Fourier coefficients of order higher than ℓ_ϵ and will omit the dependence on ϵ with the notation ℓ. It would be tempting to use Lemma 3.1 and 3.2 of [GvdV01] to bound the metric entropy of $\mathcal{P}_f$. However, as pointed by [MM11], it leads to a too weak result since the upper bound of $N(\epsilon, \mathcal{P}_f, d_H)$ will have a strong dependency on ℓ. It is then necessary to adapt the proof of [GvdV01] to obtain a sufficiently sharp upper bound of the entropy of $\mathcal{P}_f$ (with respect to d_{TV} which is easier to handle here). These results are still true with d_H using $\frac{d_H^2}{2} \leq d_{TV} \leq d_H$.

Proposition 4.2. *If $f \in \mathcal{H}^\ell$ satisfies $\|\theta\|^2 \lesssim 2\ell + 1$ and $\epsilon > 0$ with $\log \frac{1}{\epsilon} \lesssim \ell$:*

$$\log N(\epsilon, \mathcal{P}_f, d_{TV}) \lesssim \ell^2 \left(\log \frac{1}{\epsilon} + \log \|\theta\|_{\mathcal{H}_{1/2}}\right).$$

If furthermore $w \lesssim \sqrt{2\ell + 1}$ then

$$\sup_{f \in \mathcal{H}^\ell : \|\theta(f)\| \leq w} \log N(\epsilon, \mathcal{P}_f, d_{TV}) \lesssim \ell^2 \left(\log \frac{1}{\epsilon} + \log \ell\right).$$

The second inequality opens the way for the case of unknown f since we express the dependency on f and ℓ. The rate $s/(2s+2)$ obtained in Theorem 2.1 comes from this upper bound of the order ℓ^2 (to be compared to ℓ when no shift occurs in the standard regression model). We currently do not know whether smaller covering numbers can be obtained in the SIM.

Proof. To build an ϵ-covering of $\mathcal{P}_f$, we now

- approximate any mixture g by a finite one $\tilde{g}$ such that

$$d_{TV}(\mathbb{P}_{\theta,g}, \mathbb{P}_{\theta,\tilde{g}}) \leq \epsilon/2,$$

with a number of points uniformly bounded (only depends on f and ϵ);
- use Lemma 2 of [GW00] to obtain an upper bound of the metric entropy of finite mixtures from Proposition 4.1.

We first fix the notation $p = 2\ell + 1$ which is the dimension of the multivariate mixture. For any $R > 0$, denote $\mathcal{E}_R$ the centered ball of $\mathbb{C}^p$ of radius R. For sake of simplicity, we will sometimes omit the dependence of p with ϵ. There exists an absolute constant a such that $\|\theta\| \leq w \leq a\sqrt{p}$. Now

$$d_{TV}(\mathbb{P}_{\theta,g}, \mathbb{P}_{\theta,\tilde{g}}) \leq \frac{1}{2} \underbrace{\int_{\mathcal{E}_R^c} |d\mathbb{P}_{\theta,g} - d\mathbb{P}_{\theta,\tilde{g}}|(z)}_{:=(A)} + \frac{1}{2} \underbrace{\int_{\mathcal{E}_R} |d\mathbb{P}_{\theta,g} - d\mathbb{P}_{\theta,\tilde{g}}|(z)}_{:=(B)}.$$

Let ν be a measure on $[0,1]$ that dominates both g and $\tilde{g}$.

Term (A) We will pick R such that (A) is smaller than $\epsilon/2$, first set $R^2 > (1+a)^2 p \geq a^{-2}(1+a)^2 \|\theta\|^2$ and with this choice, $\|z - \theta \bullet \varphi\| > \|z\|/(1+a)$ if $z \in \mathcal{E}_R^c$. This simply implies that,

$$\begin{aligned}
(A) &\leq \pi^{-p} \int_{\mathcal{E}_R^c} \int_0^1 e^{-\frac{\|z\|^2}{(1+a)^2}} \left| \frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right| d\nu(\varphi) dz \\
&\leq 2(1+a)^{2p} \, \mathbb{P}\left(\chi_{2p}^2 \geq \frac{2R^2}{(1+a)^2} \right).
\end{aligned}$$

From the concentration of χ_p^2 statistics given by Lemma 1 of [IL06]:

$$\forall k \geq 1, \forall c > 0 \quad \mathbb{P}\left(\chi_k^2 \geq (1+c)k \right) \leq \frac{1}{c\sqrt{2\pi}} e^{-\frac{k}{2}[c - \log(1+c)] - \frac{1}{2}\log k}. \tag{4.1}$$

Therefore, writing $R^2 = (1+a)^2(1+c)p$ for $c > 0$, one gets

$$(A) \leq \frac{1}{c\sqrt{\pi}} e^{-p[c - \log(1+c) - 2\log(1+a)] - \frac{1}{2}\log p}$$

and this term is smaller than $\epsilon/2$ if we pick c large enough, since $\log \frac{1}{\epsilon} \lesssim p$.

Term (B) We follow the strategy of [GvdV01] which exploits the smoothness of Gaussian laws and exhibit suitable discrete mixtures. We traduce their matching moment conditions to the multivariate setting by exploiting Fourier analysis. Remark that

$$\forall k \in \mathbb{N} \quad \forall y \in \mathbb{R}_+ \qquad \underbrace{\left| e^{-y} - \sum_{j=0}^{k-1} \frac{(-y)^j}{j!} \right|}_{:=R_k(y)} \leq \frac{|y|^k}{k!} \leq \frac{(e|y|)^k}{k^k}. \qquad (4.2)$$

Thus, for all $z \in \mathcal{E}_R$, we have

$$\mathbb{P}_{\theta,g}(z) - \mathbb{P}_{\theta,\tilde{g}}(z) = \pi^{-p} \int_0^1 e^{-\|z - \theta \bullet \varphi\|^2} \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi)$$

$$= \pi^{-p} \sum_{j=0}^{k-1} \frac{(-1)^j}{j!} \int_0^1 \|z - \theta \bullet \varphi\|^{2j} \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi)$$

$$+ \pi^{-p} \int_0^1 R_k \left(\|z - \theta \bullet \varphi\|^2 \right) \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi).$$

We now decompose $\theta = (\theta_{-\ell}, \ldots, \theta_\ell)$ and $z = (z_{-\ell}, \ldots, z_\ell)$ using polar coordinates: $\theta_m = \rho_m^{(1)} e^{i\alpha_m}$ and $z_m = \rho_m^{(2)} e^{i\beta_m}$ for $|m| \leq \ell$. This leads to $\|z - \theta \bullet \varphi\|^2 = \|z\|^2 + \|\theta\|^2 - 2 \sum_{m=-\ell}^{\ell} \rho_m^{(1)} \rho_m^{(2)} \cos(\beta_m - \alpha_m - m\varphi)$. For any integer $j \leq k$, we deduce that

$$\|z - \theta \bullet \varphi\|^{2j} = C_j(z, \theta) + \sum_{r=1}^{j} \sum_{m=-\ell}^{\ell} a_{r,m}(z, \theta) \left[\cos(\beta_m - \alpha_m - m\varphi) \right]^r,$$

where $(a(r, m))_{r=1\ldots k, m=-\ell\ldots\ell}$ only depends on z and θ. Using Euler's identity,

$$\|z - \theta \bullet \varphi\|^{2j} = C_j(z, \theta) + \sum_{r=-j\ell}^{j\ell} b_r(z, \theta) e^{ir\varphi},$$

where b stands for a complex vector obtained through the Binomial formula.

$$\forall z \in \mathcal{E}_R : (\mathbb{P}_{\theta,g} - \mathbb{P}_{\theta,\tilde{g}})(z)$$

$$= \pi^{-p} \sum_{j=0}^{k-1} \frac{(-1)^j}{j!} \int_0^1 \left[C_j(z, \theta) + \sum_{r=-j\ell}^{j\ell} b_r(z, \theta) e^{ir\varphi} \right] \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi)$$

$$+ \pi^{-p} \int_0^1 R_k \left(\|z - \theta \bullet \varphi\|^2 \right) \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi)$$

$$= \pi^{-p} \sum_{j=0}^{k-1} \frac{(-1)^j}{j!} \left[C_j(z, \theta) c_0(g - \tilde{g}) + \sum_{r=-j\ell}^{j\ell} b_r(z, \theta) c_r(g - \tilde{g}) \right]$$

$$+ \pi^{-p} \int_0^1 R_k \left(\|z - \theta \bullet \varphi\|^2 \right) \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi).$$

Caratheodory's theorem shows that one can find $\tilde{g}$ with a finite support of size $2(k-1)\ell + 1 \sim 2k\ell$ such that $c_r(g) = c_r(\tilde{g}$ for all $r \in [-(k-1)\ell, (k-1)\ell]$.

For such finite mixture law $\tilde{g}$, we obtain $\forall z \in \mathbb{C}^p$,

$$\mathbb{P}_{\theta,g}(z) - \mathbb{P}_{\theta,\tilde{g}}(z) = \pi^{-p} \int_0^1 R_k \left(\|z - \theta \bullet \varphi\|^2 \right) \left[\frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi),$$

and of course $(B) \leq 2\pi^{-p} \sup_{z \in \mathcal{E}_R, \varphi \in (0,1)} R_k \left(\|z - \theta \bullet \varphi\|^2 \right) Vol(\mathcal{E}_R)$. According to the choice $R = (1+a)\sqrt{(1+c)p}$ which implies that $\|z - \theta \bullet \varphi\| \leq (1 + 2a)\sqrt{(1+c)p}$, and using the volume of $\mathcal{E}_R$ and Stirling's formula, we obtain

$$(B) \lesssim \pi^{-p} \frac{\left(e(1+2a)^2(1+c)p \right)^k}{k^k} \frac{\pi^p[(1+a)^2(1+c)p]^p}{p!} \lesssim C_1^p C_2^k e^{-k\log(k) + k\log(p)}$$

where we used in the last equation $p^p/p! \leq C^p$. If we define k in (4.2) such that $k \sim b\ell$ for a sufficiently large b, we then obtain for a universal C:

$$(B) = \int_{\mathcal{E}_R} |d\mathbb{P}_{\theta,g} - d\mathbb{P}_{\theta,\tilde{g}}| (z) \lesssim e^{\ell(C - b\log(b))}.$$

In order to bound (B) by $\epsilon/2$, we consider $k_\epsilon \sim b\ell_\epsilon$ for b large enough, since $\log \frac{1}{\epsilon} \lesssim \ell_\epsilon$, we have found $\tilde{g}$ with a discrete support of cardinal $s_\epsilon \sim 2b\ell_\epsilon^2$ points, with s_ϵ not depending on g, such that $d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{f,\tilde{g}}) \leq \epsilon/2$.

Now, the first inequality in Proposition 4.2 comes from Lemma 2 of [GW00]. The second inequality is now deduced *uniformly* from the first one using $\|\theta\|_{\mathcal{H}_1} \leq \ell\|\theta\|$ when $f \in \mathcal{H}^\ell$. $\qquad\square$

4.1.2. General case

We describe the picture when f varies, which is our main objective. We assume that $f \in \mathcal{H}_s$ and define a sieve which depends on a cut-off ℓ and a size w:

$$\mathcal{P}_{\ell,w} := \left\{ \mathbb{P}_{f,g} \mid f \in \mathcal{H}^\ell, \|\theta(f)\| \leq w, g \in \mathfrak{M}([0,1]) \right\}.$$

Theorem 4.1. *For $\epsilon > 0$ small enough and $(\ell_\epsilon, w_\epsilon)$ s.t. $\log \frac{1}{\epsilon} \lesssim \ell_\epsilon$ and $w_\epsilon \lesssim \sqrt{\ell_\epsilon}$:*

$$\log N(\epsilon, \mathcal{P}_{\ell_\epsilon,w_\epsilon}, d_{TV}) \lesssim \ell_\epsilon^2 \left(\log \frac{1}{\epsilon} + \log \ell_\epsilon \right).$$

Remark 4.1. *i) If we focus on the role of w_ϵ, we see in the proof of Proposition 4.2 that a w_ϵ of smaller order than $\sqrt{\ell_\epsilon}$ would not decrease the obtained entropy number. In the same time, a larger radius w_ϵ entails a damaged entropy number of order $\ell_\epsilon(\ell_\epsilon \vee w_\epsilon^2) \left(\log \frac{1}{\epsilon} + \log \ell_\epsilon \right)$. At the end, this would also damage the contraction rate of $\mathbb{P}_{f,g}$.*
ii) Our model is a special case of Gaussian mixture models, nevertheless it may be generalized to other cases within a growing dimension setting.
iii) We will use a higher choice of ℓ_ϵ than $\log \frac{1}{\epsilon}$, this will be fixed in Section 5.1.
iv) Note that we can derive the same upper bound for the metric entropy using the Hellinger distance.

The proof of Theorem 4.1 is based on two simple results. The first one is the Girsanov formula (see *e.g.* Appendix of [BG10]). In our framework, we obtain

$$\frac{d\mathbb{P}_{f,g}}{d\mathbb{P}_{f^0,g^0}}(Y) = \frac{\int_0^1 exp\left(2\Re e\langle f^{-\alpha_1}, dY\rangle - \|f^{-\alpha_1}\|^2\right) dg(\alpha_1)}{\int_0^1 exp\left(2\Re e\langle f^{0,-\alpha_2}, dY\rangle - \|f^{0,-\alpha_2}\|^2\right) dg^0(\alpha_2)}, \tag{4.3}$$

for any measurable trajectory Y. The second result is given as follows.

Lemma 4.1. *Let* $(f, \tilde{f}) \in L^2_{\mathbb{C}}([0,1])^2$, *and* $g \in \mathfrak{M}([0,1])$, *then*

$$d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{\tilde{f},g}) \leq \frac{\|f - \tilde{f}\|}{\sqrt{2}}.$$

Proof. $\mathbb{P}_{f,g}$ is a mixture model: $\mathbb{P}_{f,g} = \int_0^1 \mathbb{P}_{f,\delta_\alpha} dg(\alpha)$. Thus

$$d_{TV}\left(\mathbb{P}_{f,g}, \mathbb{P}_{\tilde{f},g}\right) = \left\|\int_0^1 \left(\mathbb{P}_{f,\delta_\alpha} - \mathbb{P}_{\tilde{f},\delta_\alpha}\right) dg(\alpha)\right\|_{TV}$$

$$\leq \int_0^1 \left\|\mathbb{P}_{f,\delta_\alpha} - \mathbb{P}_{\tilde{f},\delta_\alpha}\right\|_{TV} dg(\alpha),$$

and the last term is bounded by $\left\|\mathbb{P}_{f,\delta_0} - \mathbb{P}_{\tilde{f},\delta_0}\right\|_{TV} \leq d_H\left(\mathbb{P}_{f,\delta_0}, \mathbb{P}_{\tilde{f},\delta_0}\right)$ Assume now $Y \sim \mathbb{P}_{f,\delta_0}$, hence from (2.1) $dY = f(x)dx + dW$, with W is a complex standard Brownian motion. If we denote U a random variable $\mathcal{N}_{\mathbb{C}}(0,1)$, standard argument using Girsanov's formula yields

$$d_H^2\left(\mathbb{P}_{f,\delta_0}, \mathbb{P}_{\tilde{f},\delta_0}\right)/2$$

$$= 1 - \mathbb{E}_{f,\delta_0}\sqrt{\frac{d\mathbb{P}_{\tilde{f},\delta_0}}{d\mathbb{P}_{f,\delta_0}}(Y)} = 1 - \mathbb{E}_{f,\delta_0}\sqrt{\exp\left(2\Re e\langle \tilde{f} - f, dW\rangle - \|\tilde{f} - f\|^2\right)}$$

$$= \left(1 - \exp\left(\frac{-\|\tilde{f} - f\|^2}{2}\right)\mathbb{E}_U\left[\exp\left(\|\tilde{f} - f\|\Re e(U)\right)\right]\right) \leq \frac{\|\tilde{f} - f\|^2}{4}.$$

$\square$

Proof of Theorem 4.1. We build a ϵ-covering of $\mathcal{P}_{\ell,w}$ with $\epsilon/2$-coverings for f and g. First, let $\mathbb{P}_{f,g}$ and $\mathbb{P}_{\tilde{f},\tilde{g}}$ two elements of $\mathcal{P}_{\ell,w}$ and use the triangle inequality

$$d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{\tilde{f},\tilde{g}}) \leq d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{\tilde{f},g}) + d_{TV}(\mathbb{P}_{\tilde{f},g}, \mathbb{P}_{\tilde{f},\tilde{g}}).$$

We will look for a covering method that will use the inequality above and a tensorial argument, it requires to bound both terms. The second term is handled uniformly in $\tilde{f}$ by Proposition 4.2. We handle the first one with Lemma 4.1, deducing $\epsilon/2$-coverings of $\mathbb{P}_{f,g}$ for fixed g from $\epsilon/\sqrt{2}$-coverings of f for the $\|.\|$:

$$\log N\left(\epsilon/\sqrt{2}, \left\{f \in \mathcal{H}^{\ell_\epsilon}, \|\theta(f)\| \leq w_\epsilon\right\}, \|\cdot\|\right) \lesssim \ell_\epsilon \log\frac{w_\epsilon}{\epsilon} = o\left(\ell_\epsilon^2 \log\frac{1}{\epsilon}\right).$$

Therefore only the entropy obtained through Proposition 4.2 matters. $\square$

4.2. Hellinger neighborhoods

We describe Hellinger neighborhoods of $\mathbb{P}_{f^0,g^0}$ in terms of (f, g) and later turn them into Kullback-Leibler ones to compute their prior mass. For sake of simplicity, $\mathbb{E}_0 F(Y)$ will refer to the expectation of $F(Y)$ when $Y \sim \mathbb{P}_{f^0,g^0}$. For a cut-off ℓ_n, we denote $f_{\ell_n}^0 \in \mathcal{H}^{\ell_n}$ the truncation of f^0 at frequency ℓ_n, we get

$$d_H(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f,g}) \leq \overbrace{d_H(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f_{\ell_n}^0,g^0})}^{(E_1)} + \overbrace{d_H(\mathbb{P}_{f_{\ell_n}^0,g^0}, \mathbb{P}_{f_{\ell_n}^0,g})}^{(E_2)} + \overbrace{d_H(\mathbb{P}_{f_{\ell_n}^0,g}, \mathbb{P}_{f,g})}^{(E_3)}.$$

Next, we provide sharp upper bounds on (E_1), (E_2), (E_3) to find a suitable lower bound of the prior mass of Hellinger neighborhoods.

Upper bound of (E_1)   Since $d_H^2 \leq d_{KL}$, the Girsanov formula (4.3) yields

$$(E_1) \leq \sqrt{d_{KL}(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f_{\ell_n}^0,g^0})}$$
$$= \left(\mathbb{E}_0 \left[-\log \frac{\int_0^1 \exp\left(2\Re\langle f_{\ell_n}^{0,-\alpha}, dY \rangle - \|f_{\ell_n}^0\|^2\right) dg^0(\alpha)}{\int_0^1 \exp\left(2\Re\langle f^{0,-\alpha}, dY \rangle - \|f^0\|^2\right) dg^0(\alpha)} \right] \right)^{1/2} := (\tilde{E}_1)$$

We now obtain the upper bound of (E_1) according to the next proposition.

Proposition 4.3. *Assume that $Y \sim \mathbb{P}_{f^0,g^0}$ and $f^0 \in \mathcal{H}_s$, then*

$$(E_1) \leq (\tilde{E}_1) \leq \sqrt{2}\|f^0 - f_{\ell_n}^0\| \leq \sqrt{2}\|f^0\|_{\mathcal{H}_s} \ell_n^{-s}.$$

Proof. Denote Y a random variable sampled from $\mathbb{P}_{f^0,g^0}$. For any function F of the trajectory Y, we will denote $\mathbb{E}_\beta F(Y)$ the expectation of $F(Y)$ up to the condition that the shift is equal to β, and of course one has

$$\mathbb{E}_0[F(Y)] = \int_0^1 \mathbb{E}_\beta[F(Y)] dg^0(\beta).$$

For each possible value of $\beta \in [0, 1]$, we define

$$D_\beta(\alpha) := \exp\left(2\Re\langle f_{\ell_n}^{0,-\alpha}, f^{0,-\beta} \rangle + 2\Re\langle f_{\ell_n}^{0,-\alpha}, dW \rangle - \|f_{\ell_n}^0\|^2\right),$$
$$X_\beta(\alpha) := \exp\left(2\Re\langle (f^0 - f_{\ell_n}^0)^{-\alpha}, f^{0,-\beta} \rangle \right.$$
$$\left. + 2\Re\langle (f^0 - f_{\ell_n}^0)^{-\alpha}, dW \rangle - \|f^0 - f_{\ell_n}^0\|^2\right).$$

We can now split the randomness of the Brownian motion into two parts: the first one is spanned by the Fourier frequencies from $-\ell_n$ to ℓ_n and the second part is its orthogonal (in L^2): $W = W_1 + W_2$. Of course, W_1 and W_2 are independent.

Moreover, $\langle f_{\ell_n}^{0,-\alpha}, dW \rangle = \langle f_{\ell_n}^{0,-\alpha}, dW_1 \rangle$ and $\langle (f^0 - f_{\ell_n}^0)^{-\alpha}, dW \rangle = \langle (f^0 - f_{\ell_n}^0)^{-\alpha}, dW_2 \rangle$. For any fixed β, $D_\beta(\alpha)$ is measurable with respect to the filtration associated to W_1, and $X_\beta(\alpha)$ is independent of W_1. Jensen's inequality implies

$$
\begin{aligned}
(\tilde{E}_1)^2 &= \mathbb{E}\left[\log \frac{\int_0^1 D_\beta(\alpha) X_\beta(\alpha) dg^0(\alpha)}{\int_0^1 D_\beta(\alpha) dg^0(\alpha)} \right] \\
&\leq \log \int_0^1 \mathbb{E}_\beta^{W_2}\left[\mathbb{E}_\beta^{W_1}\left[\left. \frac{\int_0^1 D_\beta(\alpha) X_\beta(\alpha) dg^0(\alpha)}{\int_0^1 D_\beta(\alpha) dg^0(\alpha)} \right| W_2 \right] \right] dg^0(\beta) \\
&\leq \log \int_0^1 \mathbb{E}_\beta^{W_2}\left[X_\beta(\alpha) \mathbb{E}_\beta^{W_1}\left[\left. \frac{\int_0^1 D_\beta(\alpha) dg^0(\alpha)}{\int_0^1 D_\beta(\alpha) dg^0(\alpha)} \right| W_2 \right] \right] dg^0(\beta) \\
&\leq \log \int_0^1 \left(\sup_\alpha \mathbb{E}_\beta^{W_2}\left[X_\beta(\alpha) \right] \right) dg^0(\beta).
\end{aligned}
$$

The notation $\mathbb{E}_\beta^{W_1} F(Y)$ (resp. $\mathbb{E}_\beta^{W_2} F(Y)$) used above refers to the expectation of $F(Y)$ with respect to W_1 (resp. with respect to W_2) with a fixed β. Now, remark that $X_\beta(\alpha) \sim \exp\left(2\Re\langle (f^0 - f_{\ell_n}^0)^{-\alpha}, f^{0,-\beta} \rangle + U \right)$, where $U \sim \mathcal{N}_\mathbb{R}\left(-\|f^0 - f_{\ell_n}^0\|^2, 2\|f^0 - f_{\ell_n}^0\|^2 \right)$, and $\mathbb{E}\left[e^U \right] = 1$. Hence

$$
\begin{aligned}
(\tilde{E}_1)^2 &\leq \log \int_0^1 \sup_\alpha \exp\left(2\Re\langle (f^0 - f_{\ell_n}^0)^{-\alpha}, f^{0,-\beta} \rangle \right) dg^0(\beta) \\
&\leq \log \sup_{\alpha,\beta} \exp\left(2\Re\langle (f^0 - f_\ell^0)^{-\alpha}, f^{0,-\beta} \rangle \right)
\end{aligned}
$$

We can now switch log and sup since log is increasing, and we obtain

$$
(\tilde{E}_1) \leq \sqrt{ 2 \sup_{\alpha,\beta} \Re\langle (f^0 - f_{\ell_n}^0)^{-\alpha}, f^{0,-\beta} \rangle }.
$$

We decompose $f^{0,-\beta} = f_{\ell_n}^{0,-\beta} + f^{0,-\beta} - f_{\ell_n}^{0,-\beta}$ and Cauchy-Schwarz's inequality yields $(\tilde{E}_1) \leq \sqrt{2}\|f^0 - f_{\ell_n}^0\|$. Note that untill now we did not use the hypothesis $f^0 \in \mathcal{H}_s$. It is only needed to get the last inequality in Proposition 4.3. $\qquad \square$

Upper bound of (E_3) We are interested in $d_H(\mathbb{P}_{f_{\ell_n}^0,g}, \mathbb{P}_{f,g})$ when $f_{\ell_n}^0$ is close to f, ℓ_n grows up to $+\infty$ (with the same mixture law on $[0,1]$). The important fact will be its exclusive dependence with respect to $\|f_{\ell_n}^0 - f\|$. This upper bound is immediate from Lemma 4.1:

$$
\forall f \in \mathcal{H}^{\ell_n}, \forall g \in \mathfrak{M}([0,1]) : d_H(\mathbb{P}_{f_{\ell_n}^0,g}, \mathbb{P}_{f,g}) \leq 2^{1/4}\sqrt{\|f - f_{\ell_n}^0\|}.
$$

Upper bound for (E_2) We can use a discrete mixture with η-separated support points as follows.

Proposition 4.4. *Assume that $f^0 \in \mathcal{H}_s$ for $s \geq 1/2$, $g^0 \in \mathfrak{M}([0,1])$, and $\log \frac{1}{\epsilon_n} \lesssim \ell_n$. For any $\eta_n \leq \epsilon_n^4$, there exists a discrete distribution $\tilde{g}$ whose support $(\psi_j)_{j=1\ldots J_n}$ is of size $J_n \lesssim \ell_n^2$, η_n-separated, and such that*

$$d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}) \lesssim \left(1 + \|f^0\|_{\mathcal{H}_{1/2}}^{1/2}\right) \epsilon_n.$$

Furthermore, for any $g \in \mathfrak{M}([0,1])$,

$$d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, g}) \lesssim \left(1 + \|f^0\|_{\mathcal{H}_{1/2}}^{1/2}\right) \epsilon_n$$
$$+ \sqrt{\|f_{\ell_n}^0\|_{\mathcal{H}_{1/2}} \sqrt{\eta_n} + \sum_{j=1}^{J_n} |\breve{g}([\varphi_j - \eta_n/2, \varphi_j + \eta_n/2]) - \tilde{g}(\varphi_j)|}.$$

Proof. The construction used in the proof of Proposition 4.2 provides a mixture $\tilde{\tilde{g}}$ such that $\tilde{\tilde{g}}$ is supported by $\tilde{J}_n := C\ell_n^2$ points (denoted $(\varphi_j)_{j=1\ldots\tilde{J}_n}$) so that $d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, \tilde{\tilde{g}}}) \leq \epsilon_n$. Therefore $\tilde{\tilde{g}} = \sum_{j=1}^{\tilde{J}_n} w_j \delta_{\varphi_j}$. As pointed by [GvdV01], one can slightly modify $\tilde{\tilde{g}}$ so that the support points are separated enough as follows. First, denote $(\psi_j)_{j=1\ldots J_n}$ the subset of $(\varphi_j)_{j=1\ldots\tilde{J}_n}$ which is η_n-separated with a maximal number of elements. Hence, $J_n \leq \tilde{J}_n$ and up to a permutation, one can divide $(\varphi_j)_{j=1\ldots\tilde{J}_n}$ in two parts: $(\varphi_j)_{j=1\ldots\tilde{J}_n} = (\psi_j)_{j=1\ldots J_n} \cup (\varphi_j)_{j=J_n+1\ldots\tilde{J}_n}$. For any $i \in \{J_n + 1, \ldots, \tilde{J}_n\}$, we define $\psi_{j(i)}$ as the closest point of $(\psi_j)_{j=1\ldots J_n}$, the new discrete mixture law is then given by

$$\tilde{g} = \sum_{j=1}^{J_n} \underbrace{\left(w_j + \sum_{i > J_n | j(i) = j} w_i\right)}_{:=\tilde{w}_j} \delta_{\psi_j}.$$

Of course, $\tilde{g}$ has a η_n-separated support. Moreover, we have

$$d_{TV}\left(\mathbb{P}_{f_{\ell_n}^0, \tilde{g}}, \mathbb{P}_{f_{\ell_n}^0, \tilde{\tilde{g}}}\right) = d_{TV}\left(\sum_{i=1}^{\tilde{J}_n} w_j \gamma_{\theta \bullet \varphi_j}, \sum_{i=1}^{J_n} \tilde{w}_i \gamma_{\theta \bullet \psi_i}\right)$$
$$\leq \sum_{j=1}^{J_n} \sum_{i > J_n | j(i) = j} w_i d_{TV}(\gamma_{\theta \bullet \varphi_i}, \gamma_{\theta \bullet \psi_j})$$
$$\lesssim \sum_{j=1}^{J_n} \sum_{i > J_n | j(i) = j} w_i \|\theta\|_{\mathcal{H}_{1/2}} \sqrt{\eta_n} \leq \|\theta\|_{\mathcal{H}_{1/2}} \sqrt{\eta_n}$$

where we successively used convexity and Proposition 4.1. Now the relations between Hellinger and Total Variation distances yield

$$d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}) \leq \epsilon_n + d_H(\mathbb{P}_{f_{\ell_n}^0, \tilde{g}}, \mathbb{P}_{f_{\ell_n}^0, \tilde{\tilde{g}}}) \lesssim \left(1 + \|\theta\|_{\mathcal{H}_{1/2}}^{1/2}\right) \epsilon_n.$$

Now, let g be in $\mathfrak{M}([0,1])$, then an adaptation of Lemma 5.1 of [GvdV01] leads to

$$d_H^2(\mathbb{P}_{f_{\ell_n}^0,\tilde{g}}, \mathbb{P}_{f_{\ell_n}^0,g}) \lesssim \|f_{\ell_n}^0\|_{\mathcal{H}_{1/2}} \sqrt{\eta} + \sum_{j=1}^J |g([\varphi_j - \eta/2, \varphi_j + \eta/2]) - \tilde{g}(\varphi_j)|.$$

which permits to conclude. $\qquad\square$

Description of neighborhoods We can now gather (E_1, E_2, E_3) as follows.

Proposition 4.5. *Assume that $f^0 \in \mathcal{H}_s$ for $s \geq 1/2$ and $g^0 \in \mathfrak{M}([0,1])$ and consider $\ell_n \sim \epsilon_n^{-1/s}$, $\eta_n := \epsilon_n^4$, and $\tilde{g}$ the mixture get in Proposition 4.4. Define*

$$\mathcal{G}_{\epsilon_n} := \left\{ g \in \mathfrak{M}([0,1]) : \sum_{j=1}^{J_n} |g(\psi_j - \eta_n/2, \psi_j + \eta_n/2) - \tilde{g}(\psi_j)| \leq \epsilon_n^2 \right\},$$

$$\mathcal{F}_{\epsilon_n} := \left\{ f \in \mathcal{H}^{\ell_n} : \|f - f_{\ell_n}^0\| \leq \epsilon_n^2 \right\}.$$

There exists a constant C_0 depending only on f^0 such that

$$\forall g \in \mathcal{G}_{\epsilon_n}, \quad \forall f \in \mathcal{F}_{\epsilon_n} : \qquad d_H\left(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f,g}\right) \leq C_0 \epsilon_n.$$

5. Posterior contraction rates

5.1. Proof of Theorem 2.1

We first study the minoration of the Kullback-Leibler neighborhoods.

Proposition 5.1. *Assume that $f^0 \in \mathcal{H}_s$ for $s \geq 1$ and $g^0 \in \mathfrak{M}([0,1])$. For any sequence $(\epsilon_n)_{n \in \mathbb{N}}$ which converges to 0 as $n \to +\infty$, one has*

$$\Pi_n^{DP}\left(\mathbb{P}_{f,g} \in \mathcal{P} : d_{KL}(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \leq \epsilon_n^2, V(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \leq \epsilon_n^2\right) \geq h_n,$$

where c is a constant such that

$$h_n := e^{-(c+o(1))\left[\epsilon_n^{-2/s}(\log(1/\epsilon_n))^{\rho+2/s} \vee \xi_n^{-2}\right]}.$$

We could obtain a suitable lower bound as soon as $\lambda(\ell_n) \geq e^{-c\ell_n^2 \log \ell_n}$ and any λ possessing an heavier tail would also fit. But, we also need to majorize $\Pi_n^{DP}(\mathcal{P} \setminus \mathcal{P}_n)$, see below.

Proof. We have seen in the proof of Proposition A.1 that M_δ^2 is uniformly bounded with respect to $\|f\|$ and $\|f^0\|$ for a suitable choice of δ. We restrict our study to the elements f such that $\|f\| \leq 2\|f^0\|$. We know from Proposition A.1 that as soon as $\tilde{\epsilon}_n \log \frac{1}{\tilde{\epsilon}_n} \leq c\epsilon_n$ with c small enough, $\mathcal{V}_{\tilde{\epsilon}_n}(\mathbb{P}_{f^0,g^0}, d_H) := \left\{\mathbb{P}_{f,g} \in \mathcal{P} | d_H(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f,g}) \leq \tilde{\epsilon}_n, \|f\| \leq 2\|f^0\|\right\} \subset \mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0,g^0}, d_{KL})$. This last condition on $\tilde{\epsilon}_n$ is true as soon as

$$\tilde{\epsilon}_n := \tilde{c}\epsilon_n \left(\log \frac{1}{\epsilon_n}\right)^{-1} \tag{5.1}$$

with $\tilde{c}$ small enough. Now, Proposition 4.5 permits to describe a subset of $\mathcal{V}_{\tilde{\epsilon}_n}(\mathbb{P}_{f^0,g^0}, d_H)$, with subsets $\mathcal{F}_{\tilde{\epsilon}_n}$ and $\mathcal{G}_{\tilde{\epsilon}_n}$ for f and g. Choose $\ell_n := \tilde{\epsilon}_n^{-1/s}$.

We first bound the prior mass on $\mathcal{G}_{\tilde{\epsilon}_n}$. This follows from the bound given by Lemma 6.1 of [GGvdV00] for DP. Note that $J_n \lesssim \ell_n^2 = \tilde{\epsilon}_n^{-2/s} \leq \tilde{\epsilon}_n^{-4}$ and there exists an absolute constant $a \in (0, 1]$ such that the condition $J_n \leq 2(a\tilde{\epsilon}_n)^{-4}$ is fulfilled. Hence one can find constants C and c such that for n large enough

$$\Pi_n\left(\mathcal{G}_{\tilde{\epsilon}_n}\right) \geq \Pi_n\left(\mathcal{G}_{a\tilde{\epsilon}_n}\right) \geq Ce^{-cJ_n \log \frac{1}{\tilde{\epsilon}_n^4}} \geq Ce^{-c\ell_n^2 \log \frac{1}{\tilde{\epsilon}_n}}. \tag{5.2}$$

We next consider the prior mass on $\mathcal{F}_{\tilde{\epsilon}_n}$. Remark that when n is large enough, any element of $\mathcal{F}_{\tilde{\epsilon}_n}$ satisfies $\|f\| \leq 2\|f^0\|$ and the additional condition on $\|f\|$ in the definition of $\mathcal{V}_{\tilde{\epsilon}_n}(\mathbb{P}_{f^0,g^0}, d_H)$ is instantaneously fulfilled. Of course:

$$\Pi_n\left(\mathcal{F}_{\tilde{\epsilon}_n}\right) \geq \lambda(\ell_n) \times \pi_{\ell_n}\left(B\left(\theta^0_{\ell_n}, \tilde{\epsilon}_n^2\right)\right).$$

From our assumption on the prior λ, we have $\lambda(\ell_n) \geq e^{-c\ell_n^2 \log^\rho \ell_n}$, and the value of the volume of the $(4\ell_n + 2)$-dimensional Euclidean ball of radius $\tilde{\epsilon}_n^2$ implies

$$\Pi_n\left(\mathcal{F}_{\tilde{\epsilon}_n}\right) \geq e^{-c\ell_n^2 \log^\rho \ell_n} \inf_{u \in B(0,\tilde{\epsilon}_n^2)} \left(\frac{e^{-\|\theta^0 + u\|^2/\xi_n^2}}{\pi^{2\ell_n+1}\xi_n^{2(2\ell_n+1)}}\right) \left(\tilde{\epsilon}_n^2\right)^{4\ell_n+2} \frac{\pi^{2\ell_n+1}}{\Gamma(2\ell_n + 2)}.$$

For n large enough we get

$$\begin{aligned}
\Pi_n\left(\mathcal{F}_{\tilde{\epsilon}_n}\right) &\geq \exp - \left[c\ell_n^2 \log^\rho \ell_n + \xi_n^{-2}\right. \\
&\quad \left. + (2\ell_n + 1)\left(\log \ell_n + 4\log(1/\tilde{\epsilon}_n) - \log \xi_n^{-2} + \mathcal{O}(1)\right)\right] \\
&\geq \exp\left[-(c + o(1))\left[\ell_n^2 \log^\rho \ell_n \vee \xi_n^{-2}\right]\right]
\end{aligned} \tag{5.3}$$

Gathering (5.2), (5.3), $\ell_n = \tilde{\epsilon}_n^{-1/s}$ and (5.1) lead (for c small enough) to

$$\begin{aligned}
\Pi_n\left(\mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0,g^0}, d_{KL})\right) &\geq \Pi_n\left(\mathcal{F}_{\tilde{\epsilon}_n}\right)\Pi_n\left(\mathcal{G}_{\tilde{\epsilon}_n}\right) \\
&\geq \exp\left[-(c + o(1))\left[\ell_n^2 \log^\rho \ell_n \vee \xi_n^{-2}\right]\right] \\
&\geq \exp\left[-(c + o(1))\left[\epsilon_n^{-2/s}(\log(1/\epsilon_n))^{\rho+2/s} \vee \xi_n^{-2}\right]\right].
\end{aligned}$$

$\square$

Proposition 5.2. *For any sequences $k_n \mapsto +\infty$ and $\epsilon_n \mapsto 0$ as $n \mapsto +\infty$, define $w_n^2 = 4k_n + 2$, then there exists a constant c such that*

$$\Pi_n^{DP}\left(\mathcal{P} \setminus \mathcal{P}_{k_n,w_n}\right) \leq e^{-c[k_n^2 \log^\rho(k_n) \wedge k_n \xi_n^{-2}]},$$

and $$\log D\left(\epsilon_n, \mathcal{P}_{k_n,w_n}, d_H\right) \lesssim k_n^2 \left[\log k_n + \log \frac{1}{\epsilon_n}\right].$$

Proof. The upper bound on the packing number comes directly from Theorem 4.1 since we set $w_n = \sqrt{2k_n + 1}$. To control the prior mass outside the sieve, remark that owing to the construction of our prior, we have

$$\Pi_n \left(\mathcal{P} \setminus \mathcal{P}_{k_n, w_n} \right) \leq \sum_{|k| \geq k_n} \lambda(k) + \Pr\left(\sum_{|k| \leq k_n} |\theta_k|^2 \geq w_n^2 \right), \qquad (5.4)$$

where each θ_k for $-k_n \leq k \leq k_n$ follows a centered Gaussian law of variance ξ_n^2. Now, there exists some constants c and C such that for sufficiently large n: $\sum_{|k| \geq k_n} \lambda(k) \leq C\lambda(k_n) \leq e^{-ck_n^2 \log^\rho(k_n)}$. Regarding now the second term of the upper bound in (5.4), we use (4.1) to get

$$\Pr\left(\sum_{|k| \leq k_n} |\theta_k|^2 \geq w_n^2 \right) = \Pr\left(\sum_{|k| \leq k_n} \left| \frac{\theta_k}{\xi_n} \right|^2 \xi_n^2 \geq w_n^2 \right)$$

$$\leq \mathbb{P}\left(\chi^2_{2k_n+1} \geq 2(2k_n+1)\xi_n^{-2} \right) \qquad (5.5)$$

$$\leq \frac{1}{(\xi_n^{-2} - 1)\sqrt{\pi}} e^{-(2k_n+1)[\xi_n^{-2} - 1 - \log \xi_n^{-2}] - \log(2k_n+1)/2}.$$

The value of ξ_n yields $\Pi_n \left(\mathcal{P} \setminus \mathcal{P}_{k_n, w_n} \right) \leq e^{-c[k_n^2 \log^\rho(k_n) \wedge k_n \xi_n^{-2}]}$. $\qquad \square$

Remark 5.1. *Note that the limited size of w_n and the entropy of the sieve (see Proposition 4.2) prevent the use of a prior independent on the sample size. Indeed, we need in (5.5) a small enough ξ_n to avoid a too large weight outside of the sieve $\mathcal{P}_{\ell_n, w_n}$. Note also that a variance $\xi_{k,n}$ that decays with the frequency k would not bypass this limitation.*

We are now able to conclude the proof of the posterior consistency.

Proof of Theorem 2.1. Take $\epsilon_n := n^{-\alpha}(\log n)^\kappa$ and $k_n := n^\beta (\log n)^\gamma$. From the definition (2.3), we have $\xi_n^{-2} = n^{\mu_s}(\log n)^\zeta$, and we look for admissible values of α, β, κ, γ, μ_s, and ζ to apply Theorem 2.1 of [GGvdV00]. To derive a suitable lower bound of the prior of $\mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0, g^0}, d_{KL})$, Proposition 5.1 yields

$$\epsilon_n^{-2/s} \left(\log \frac{1}{\epsilon_n} \right)^{\rho + 2/s} \vee n^{\mu_s}(\log n)^\zeta \ll n\epsilon_n^2 = n^{1-2\alpha}(\log n)^{2\kappa}.$$

This is true as soon as $\epsilon_n = n^{-\alpha}(\log n)^\kappa$ satisfies $\alpha \leq \frac{s}{2s+2}$ and $\kappa > \frac{\rho s + 2}{2s+2}$. We also obtain the condition on μ_s: $\mu_s \leq 1 - 2\alpha$, and if $\mu_s = 1 - 2\alpha$ then $\zeta < 2\kappa$.

Now, Proposition (5.2) shows that the sieve is not too large provided that

$$k_n^2 \left[\log k_n + \log \frac{1}{\epsilon_n} \right] \lesssim n\epsilon_n^2 = n^{1-2\alpha}(\log n)^{2\kappa}. \qquad (5.6)$$

This condition is satisfied when $2\beta \leq 1-2\alpha$ and $2\gamma+1 \leq 2\kappa$. At last, Proposition (5.2) shows that the prior of the complementary of the Sieve is not too large if

$$k_n^2 \log^\rho k_n \wedge k_n n^{\mu_s} \gtrsim n\epsilon_n^2$$

and we deduce from (5.6) that $2\beta = 1 - 2\alpha$ and $-\rho/2 + \kappa \leq \gamma \leq -1/2 + \kappa$. Moreover, we also see that $\beta + \mu_s \geq 1 - 2\alpha$, hence $\mu_s \geq 1/2 - \alpha$, and if $\mu_s = 1/2 - \alpha$ then $\gamma + \zeta \geq 2\kappa$; the former condition on μ_s yields $\mu_s \geq 1/2 - \alpha \geq \frac{1}{2s+2}$.

To sum up this part, we have to balance the following inequalities

$$\frac{1}{2s+2} \leq 1/2 - \alpha \leq \mu_s \leq 1 - 2\alpha \text{ and if } \mu_s = 1 - 2\alpha \text{ then } \zeta < 2\kappa.$$

Now, we split the proof according to the adaptive or non adaptive case.

Adaptive prior We first set μ independent of s and equal to $1/3$. For any $s \in [1/2, 2]$, we see that $\alpha(s) = s/(2s+2)$ is the largest value of α and $\alpha(s) = 1/3 < s/(2s+2)$ as soon as $s > 2$. The corresponding value of β is $1/(2s+2)$ when $s \in [1/2, 2]$ and $\beta = 1/6$ otherwise. Any choice of $\zeta \in [3/2, 2)$ permits to deal with the conditions on ζ. The other values of γ and κ may be determined by choosing $\rho \in (1, 2), \kappa = 1$ and $\gamma = 1/2$.

Non adaptive prior This case is simpler, for instance we can fix

$$\mu_s = \beta = \frac{1}{2s+2}, \alpha = \frac{s}{2s+2}, \rho = 2, \kappa = 1, \gamma = \frac{1}{2}, \zeta = 0.$$

$\square$

5.2. Proof of Theorem 2.2

The entropy bounds of the model are still valid for the proof of Theorem 2.2. However, the way we previously described the closeness of $\mathbb{P}_{f,g}$ and $\mathbb{P}_{f,\tilde{g}}$ is not satisfactory with a smooth GP prior: Lemma 6.1 of [GGvdV00] designed for DP is not yet convenient for GP and we must derive another neighborhood structure, easily tractable with GP. The next Proposition, of independent interest for mixture models, does the job.

Consider the inverse functions of the distribution functions defined by

$$\forall u \in [0, 1], \quad G^{-1}(u) = \inf\{t \in [0, 1] : g([0, t]) > u\},$$

and recall that the Wasserstein (or Kantorovich) distance W_t, for $t > 0$, is given by

$$W_t(g, \tilde{g}) := \left(\int_0^1 \left| G^{-1}(u) - \tilde{G}^{-1}(u) \right|^t du \right)^{1/t}.$$

Proposition 5.3. *Consider* $f \in \mathcal{H}_s$ *for* $s \geq 1/2$, *and let* $(g, \tilde{g}) \in \mathfrak{M}([0, 1])$, *then*

$$d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{f,\tilde{g}}) \lesssim \|f\|_{\mathcal{H}_{s\wedge 1}} W_{s\wedge 1}(g, \tilde{g})^{s\wedge 1} \leq \|f\|_{\mathcal{H}_{s\wedge 1}} W_1(g, \tilde{g})^{s\wedge 1}$$
$$\leq \|f\|_{\mathcal{H}_{s\wedge 1}} d_{TV}(g, \tilde{g})^{s\wedge 1} \lesssim \|f\|_{\mathcal{H}_{s\wedge 1}} \|g - \tilde{g}\|^{s\wedge 1}.$$

Proof of Proposition 5.3. The convexity of d_{TV}, and Lemma 4.1 yields

$$
\begin{aligned}
d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{f,\tilde{g}}) &= \left\| \int_0^1 \mathbb{P}_{f,\delta_\alpha} dg(\alpha) - \int_0^1 \mathbb{P}_{f,\delta_\alpha} d\tilde{g}(\alpha) \right\|_{TV} \\
&= \left\| \int_0^1 \left(\mathbb{P}_{f,\delta_{G^{-1}(u)}} - \mathbb{P}_{f,\delta_{\tilde{G}^{-1}(u)}} \right) du \right\|_{TV} \\
&\lesssim \|f\|_{\mathcal{H}_{s\wedge 1}} \int_0^1 \left| G^{-1}(u) - \tilde{G}^{-1}(u) \right|^{s\wedge 1} du.
\end{aligned}
$$

by Proposition 4.1. Now, the last inequalities are classical, see for instance [GS02, Theorem 4]. □

Proof of Theorem 2.2. We mimic the proof of Theorem 2.1

Complementary of the sieve   First, we consider the following sieve over $\mathcal{P}$:

$$
\tilde{\mathcal{P}}_{k_n,w_n} := \left\{ \mathbb{P}_{f,g} : (f,g) \in \mathcal{H}^{k_n} \times \mathfrak{M}_\nu([0,1])(2A), \|f\| \le w_n \right\},
$$

where k_n is a sequence such that $k_n \longmapsto +\infty$ as $n \longmapsto +\infty$, and $w_n^2 = 4k_n + 2$. Our sieve is included in the set of all mixture laws (without any restriction on the smoothness of g), and since the marginal of the prior on the f coordinate is the same for the DP and the GP priors, we can apply Proposition 5.2 to get

$$
\Pi_n^{GP} \left(\mathcal{P} \setminus \tilde{\mathcal{P}}_{k_n,w_n} \right) \le e^{-c[k_n^2 \log^\rho(k_n) \wedge k_n \xi_n^{-2}]}.
$$

Entropy estimates   Since $\mathfrak{M}_\nu([0,1])(2A) \subset \mathfrak{M}([0,1])$, our sieve is included in the one of in Theorem 2.1 and we also deduce that for any sequence $\epsilon_n \longrightarrow 0$:

$$
\log N \left(\epsilon_n, \tilde{\mathcal{P}}_{k_n,w_n}, d_H \right) \lesssim k_n^2 \left[\log k_n + \log \frac{1}{\epsilon_n} \right].
$$

Lower bound of $\Pi_n^{GP} \left(\mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0,g^0}, d_{KL}) \right)$ Define $\tilde{\epsilon}_n = c\epsilon_n \left(\log \frac{1}{\epsilon_n} \right)^{-1}, \ell_n \sim \tilde{\epsilon}_n^{-1/s}$, as well as the sets $\mathcal{F}_\epsilon := \left\{ f \in \mathcal{H}^{\ell_n} : \|f - f_{\ell_n}^0\| \le \epsilon^2 \right\}$, which was used before, and $\tilde{\mathcal{G}}_\epsilon := \left\{ g \in \mathfrak{M}_\nu([0,1])(2A) : d_{TV}(g, g^0) \le \epsilon \right\}$. Lemma 4.1, Proposition A.1 and 5.3 implies that if $f \in \mathcal{F}_{\tilde{\epsilon}_n}$ and $g \in \tilde{\mathcal{G}}_{\frac{1}{\tilde{\epsilon}_n^{s\wedge 1}}}$, then $\mathbb{P}_{f,g} \in \mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0,g^0}, d_{KL})$. Proposition 5.1 yields: $\Pi_n^{GP}(\mathcal{F}_{\tilde{\epsilon}_n}) \ge e^{-(c+o(1))\left[\epsilon_n^{-2/s}(\log(1/\epsilon_n))^{\rho+2/s} \vee \xi_n^{-2} \right]}$. According to Theorem B.1 (given in the appendix), and since $k_\nu + 1/2 \le \nu$:

$$
\Pi_n^{GP} \left(\tilde{\mathcal{G}}_{\frac{1}{\tilde{\epsilon}_n^{s\wedge 1}}} \right) \ge e^{-(c+o(1))\tilde{\epsilon}_n^{-\frac{1}{(s\wedge 1)(k_\nu+1/2)}}} \ge e^{-(c+o(1))\tilde{\epsilon}_n^{-\frac{1}{(s\wedge 1)\nu}}}.
$$

Contraction Rate We now find a suitable choice of k_n and ϵ_n, *i.e.*

$$\Pi_n^{GP}\left(\tilde{\mathcal{G}}_{\tilde{\epsilon}_n}\right)\Pi_n\left(\mathcal{F}_{\tilde{\epsilon}_n}\right) \geq e^{-Cn\epsilon_n^2}, \quad \Pi_n^{GP}\left(\mathcal{P}\setminus\mathcal{P}_{k_n,w_n}\right) \leq e^{-(C+4)n\epsilon_n^2},$$

$$\log D\left(\epsilon_n, \tilde{\mathcal{P}}_{k_n,w_n}, d_H\right) \lesssim n\epsilon_n^2.$$

Following the arguments of Section 5.1, we can find $\gamma > 0$ and $\kappa > 0$ such that

$$\epsilon_n := n^{-\left[\frac{\nu(s\wedge 1)}{2\nu(s\wedge 1)+1}\wedge\frac{s}{2s+2}\right]}\log(n)^{\kappa}, \qquad k_n = n^{\frac{1}{2}-\left[\frac{\nu(s\wedge 1)}{2\nu(s\wedge 1)+1}\wedge\frac{s}{2s+2}\right]}\log(n)^{\gamma}.$$

$\square$

Remark 5.2. *Note that we have chosen to deal with a non-parametric structure on g for the sake of generality. Now, if g is assumed unknown and to belong to a parametric space of dimension d, it is reasonnable to expect a better convergence rate. More precisely, one can use Lemma 4.1 and Proposition 5.3 to obtain an entropy number of order $\epsilon^{-1/s}\log(\epsilon^{-1})$ and then deduce a standard non parametric contraction rate on $\mathbb{P}_{f,g}$ of order $n^{-s/(2s+1)}$.*

6. Semi-parametric results for Π_n^{GP}

In the SIM, an important issue is the identifiability with respect to the unknown curve f and mixture law g. We provide a generic identifiability condition and then deduce from Theorem 2.2 a contraction rate around the true f^0 and g^0.

6.1. Identifiability of the model

In previous works, identifiability generally depends on a restriction on the support of g. For instance, [BG10] assume g centered and compactly supported in $[-1/4, 1/4]$ (shapes were defined on $[-1/2; 1/2]$ instead of $[0, 1]$ in our paper) although f is supposed to have a non vanishing first Fourier coefficient ($\theta_1(f) \neq 0$). The same kind of conditions are also assumed in [BG12].

If the condition on the first harmonic on f is imperative to obtain identifiability of g, the restriction on its support size seems artificial and we detail in the sequel how one can avoid it. Recall that for any curve Y, $\theta_1(Y) = \theta_1^0 e^{-i2\pi\tau} + \xi$. Up to a change of variable, we can always modify g in $\tilde{g}$ such that $\theta_1^0 \in \mathbb{R}_+$, for instance fix $\tilde{g}(\varphi) = g(\varphi + \alpha)$ where α is the complex argument of θ_1^0. Consequently, w.l.o.g. we study identifiability of the SIM when f belongs to $\mathcal{F}_s$.

Proof of Theorem 2.3. We use three hierarchical steps. First, we prove that if $\mathbb{P}_{f,g} = \mathbb{P}_{\tilde{f},\tilde{g}} \Longrightarrow \theta_1(f) = \theta_1(\tilde{f})$. Then we deduce that $g = \tilde{g}$ and at last we obtain the identifiability for all other Fourier coefficients of f.

Note that as soon as $\nu > 1/2$, g and $\tilde{g}$ admit densities w.r.t. the Lebesgue measure on $[0, 1]$. In the sequel we use the same notation g for the density of g.

Point 1: Identifiability on $\theta_0(f)$ and $\theta_1(f)$ We denote $\mathbb{P}_{f,g}^k$ the marginal law of $\mathbb{P}_{f,g}$ on the kth Fourier coefficient. We have the following implications

$$d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{\tilde{f},\tilde{g}}) = 0 \implies \left(\mathbb{P}_{f,g} = \mathbb{P}_{\tilde{f},\tilde{g}}\right) \implies \forall k \in \mathbb{Z} : d_{TV}\left(\mathbb{P}_{f,g}^k, \mathbb{P}_{\tilde{f},\tilde{g}}^k\right) = 0.$$

We immediately obtain that $\theta_0(f) = \theta_0(\tilde{f})$ since $\theta_0(f)$ (resp. $\theta_0(\tilde{f})$) represents the mean of the distribution $\mathbb{P}_{f,g}^0$ (resp. $\mathbb{P}_{\tilde{f},\tilde{g}}^0$). But the distribution $\mathbb{P}_{f,g}^0$ does not bring any information on the measure g, and is no longer helpful for identifiability. Concerning now the first Fourier coefficient, remark that

$$d_{TV}\left(\mathbb{P}_{f,g}^1, \mathbb{P}_{\tilde{f},\tilde{g}}^1\right)$$
$$= \frac{1}{2\pi} \int_{\mathbb{C}} \left| \int_0^1 e^{-|\theta_1 e^{i2\pi\alpha} - z|^2} g(\alpha)d\alpha - \int_0^1 e^{-|\tilde{\theta}_1 e^{i2\pi\alpha} - z|^2} \tilde{g}(\alpha)d\alpha \right| dz.$$

Assume that $\tilde{\theta}_1 \neq \theta_1$, w.l.o.g. $\tilde{\theta}_1 > \theta_1 > 0$ and consider the disk $D_{\mathbb{C}}\left(0, \frac{\tilde{\theta}_1 - \theta_1}{2}\right)$, we then get for any z in $D_{\mathbb{C}}\left(0, \frac{\tilde{\theta}_1 - \theta_1}{2}\right)$ and any α: $|\theta_1 e^{i2\pi\alpha} - z| < \frac{\tilde{\theta}_1 + \theta_1}{2}$ and $|\tilde{\theta}_1 e^{i2\pi\alpha} - z| > \frac{\tilde{\theta}_1 + \theta_1}{2}$. Hence, we get $\int_0^1 e^{-|\theta_1 e^{i2\pi\alpha} - z|^2} g(\alpha)d\alpha > e^{-\frac{|\tilde{\theta}_1 + \theta_1|^2}{4}}$ and of course $\int_0^1 e^{-|\tilde{\theta}_1 e^{i2\pi\alpha} - z|^2} \tilde{g}(\alpha)d\alpha < e^{-\frac{|\tilde{\theta}_1 + \theta_1|^2}{4}}$. We obtain

$$d_{TV}\left(\mathbb{P}_{f,g}^1, \mathbb{P}_{\tilde{f},\tilde{g}}^1\right) \geq \frac{1}{2\pi} \int_{D_{\mathbb{C}}\left(0, \frac{\tilde{\theta}_1 - \theta_1}{2}\right)} \left| \int_0^1 e^{-|\theta_1 e^{i2\pi\alpha} - z|^2} g(\alpha)d\alpha \right.$$
$$\left. - \int_0^1 e^{-|\tilde{\theta}_1 e^{i2\pi\alpha} - z|^2} \tilde{g}(\alpha)d\alpha \right| dz > 0.$$

In the opposite, $d_{TV}(\mathbb{P}_{f,g}^1, \mathbb{P}_{\tilde{f},\tilde{g}}^1) = 0$ implies that $\theta_1 = \tilde{\theta}_1$ since $(f, \tilde{f}) \in \mathcal{F}_s(A)$.

Point 2: Identifiability on g We still assume that $d_{TV}(\mathbb{P}_{f,g}^1, \mathbb{P}_{\tilde{f},\tilde{g}}^1) = 0$, so that $\theta_1 = \tilde{\theta}_1$ and we want to infer that $g = \tilde{g}$. Using a polar change of variables

$$d_{TV}\left(\mathbb{P}_{f,g}^1, \mathbb{P}_{\tilde{f},\tilde{g}}^1\right)$$
$$= \frac{1}{2\pi} \int_{\mathbb{C}} e^{-[\theta_1^2 + |z|^2]} \left| \int_0^1 e^{2\Re(z\theta_1 e^{i2\pi\alpha})} (g(\alpha) - \tilde{g}(\alpha)d\alpha \right| dz$$
$$= \frac{1}{4\pi^2} \int_0^{+\infty} \rho e^{-[\theta_1^2 + \rho^2]} \int_0^{2\pi} \left| \int_0^{2\pi} e^{2\rho\theta_1 \cos(u-\varphi)} (g - \tilde{g})(u/2\pi)du \right| d\varphi d\rho$$
$$= \frac{1}{4\pi^2} \int_0^{+\infty} \rho e^{-[\theta_1^2 + \rho^2]} \int_0^{2\pi} |\psi_{2\rho\theta_1}(\varphi)| \, d\varphi d\rho.$$

In the expression above, we denote $h = g - \tilde{g}$ and $\psi_a(\varphi)$ is defined as

$$\psi_a(\varphi) = \int_0^{2\pi} e^{a\cos(u)} h\left(\frac{u+\varphi}{2\pi}\right) du.$$

Of course, $\|\psi_a\|_\infty \le 4\pi e^a$, and we roughly bound $|\psi_a(\varphi)| \ge \frac{|\psi_a(\varphi)|^2}{4\pi e^a}$. Hence,

$$d_{TV}(\mathbb{P}^1_{f,g}, \mathbb{P}^1_{\tilde{f},\tilde{g}}) \ge \frac{1}{8\pi^2} \int_0^{+\infty} \rho e^{-(\theta_1^2 + \rho^2 + 2\theta_1\rho)} \|\psi_{2\rho\theta_1}\|^2 d\rho. \qquad (6.1)$$

Since $\nu > 1$, h and ψ_a may be expanded in Fourier series since $h \in \mathcal{L}^2([0,1])$:

$$h(x) = \sum_{n\in\mathbb{Z}} c_n(h)e^{i2\pi nx}, \qquad \psi_a(\varphi) = \sum_{n\in\mathbb{Z}} c_n(h)\left[\int_0^{2\pi} e^{a\cos(u)}e^{inu}du\right]e^{i2\pi n\varphi}.$$

Thus, the L^2 norm of ψ_a is given by

$$\|\psi_a\|^2 = \sum_{n\in\mathbb{Z}} |c_n(h)|^2 \left|\int_0^{2\pi} e^{a\cos(u)}e^{inu}du\right|^2. \qquad (6.2)$$

Now, denote the first and second kind of Tchebychev polynomials $(T_n)_{n\in\mathbb{Z}}$ and $(U_n)_{n\in\mathbb{Z}}$ satisfying $T_n(\cos\theta) = \cos(n\theta)$ and $(\sin\theta)U_n(\cos\theta) = \sin(n\theta)$, we get

$$\int_0^{2\pi} e^{a\cos(u)}e^{inu}du$$

$$= \int_0^{2\pi} e^{a\cos(u)}\left[T_n(\cos u) + i(\sin u)U_n(\cos u)\right]du$$

$$= \int_0^{2\pi} \sum_{k\ge 0} \frac{a^k(\cos u)^k}{k!}\left[T_n(\cos u) + i(\sin u)\sum_{j=0}^n \beta_j(\cos u)^j\right]du$$

where we have used the analytic expression of U_n given by

$$U_n(\cos u) = \sum_{j=0}^{E((n-1)/2)} (-1)^j C_n^{2j+1}(\cos u)^{n-2j-1}(1 - \cos^2 u)^j.$$

Hence, we obtain for $a \in \mathbb{R}$:

$$\int_0^{2\pi} e^{a\cos(u)}e^{inu}du$$

$$= \int_0^{2\pi} \sum_{k\ge 0} \frac{a^k(\cos u)^k}{k!}T_n(\cos u)du + i\sum_{k\ge 0}\sum_{j=0}^n \beta_j\frac{a^k}{k!}\int_0^{2\pi} \sin u(\cos u)^{k+j}du$$

$$= \int_0^{2\pi} \sum_{k\ge 0} \frac{a^k(\cos u)^k}{k!}T_n(\cos u)du = \int_0^{2\pi} e^{a\cos(u)}\cos(nu)du.$$

Denote $A_n(a) := \int_0^{2\pi} e^{a\cos(u)}\cos(nu)du$, Equation (6.2) yields

$$\|\psi_a\|^2 = \sum_{n\in\mathbb{Z}} |c_n(h)|^2 A_n(a)^2. \qquad (6.3)$$

A_n is analytic and is not the null function, otherwise all its derivative would also vanish but remark that $(\cos u)^n = T_n(\cos u)$ and several derivations yield

$$
\begin{aligned}
A_n^{(n)}(0) \\
= \quad & \frac{d^{(n)}}{da^{(n)}} \left[\sum_{k=0}^{\infty} \frac{a^k}{k!} \int_0^{2\pi} (\cos u)^k \cos(nu)\, du \right] (0) = \int_0^{2\pi} (\cos u)^n T_n(\cos u)\, du \\
= \quad & \int_0^{2\pi} \left[\sum_{k=0}^{n-1} \alpha_k T_k(\cos u) + 2^{1-n} T_n(\cos u) \right] T_n(\cos u)\, du = 2^{1-n}\pi > 0.
\end{aligned}
$$

Note that in the meantime, we also obtain that $A_n^{(j)}(0) = 0, \forall j < n$, so that

$$
A_n(a) \sim_{a \mapsto 0} \frac{2^{1-n}\pi}{n!} a^n. \tag{6.4}
$$

We can conclude the proof of the identifiability of g using (6.3) in (6.1) to obtain

$$
d_{TV}(\mathbb{P}^1_{f,g}, \mathbb{P}^1_{\tilde{f},\tilde{g}}) \geq \frac{1}{8\pi^2} \sum_{n \in \mathbb{Z}} |c_n(h)|^2 \underbrace{\left(\int_0^{+\infty} \rho e^{-[\theta_1 + \rho]^2} A_n(2\rho\theta_1)^2 d\rho \right)}_{:=I_n(\theta_1)}. \tag{6.5}
$$

From (6.4), each integral $I_n(\theta_1) \neq 0, \forall n \in \mathbb{Z}$ and we then conclude that:

$$
d_{TV}(\mathbb{P}^1_{f,g}, \mathbb{P}^1_{\tilde{f},\tilde{g}}) \iff g = \tilde{g} \qquad \text{and} \qquad \theta_1 = \tilde{\theta}_1.
$$

Point 3: Identifiability on f   We prove now that $\mathbb{P}_{f,g} = \mathbb{P}_{\tilde{f},\tilde{g}} \implies f = \tilde{f}$. Since it implies $g = \tilde{g}$, it remains to show that $\theta_k(f) = \theta_k(\tilde{f}), \forall k \in \mathbb{Z} \setminus \{0,1\}$. A similar argument as the one of Point 1 yields

$$
d_{TV}(\mathbb{P}^k_{f,g}, \mathbb{P}^k_{\tilde{f},\tilde{g}}) = 0 \implies |\theta_k| = |\tilde{\theta}_k|.
$$

But we cannot directly conclude here (no restriction is assumed on the phase of Fourier coefficients $\theta_k(f), k \in \mathbb{Z} \setminus \{0,1\}$). Write $\tilde{\theta}_k = \theta_k e^{i\varphi}$, and $g = \tilde{g}$ implies

$$
d_{TV}(\mathbb{P}^k_{f,g}, \mathbb{P}^k_{\tilde{f},g}) = \frac{1}{2\pi} \int_{\mathbb{C}} \left| \underbrace{\int_0^{2\pi} e^{-|z - \theta_k e^{-ik\alpha}|^2} - e^{-|z - \theta_k e^{i(\varphi - k\alpha)}|^2} g(\alpha) d\alpha}_{:=F(z)} \right| dz.
$$

If one writes $z = x + iy$, F is differentiable w.r.t. x and y and $F(0) = 0$. A computation of $\nabla F(0)$ shows that $\nabla F(0)$ is the vector (written in $\mathbb{C}$):

$$
\nabla F(0) = \theta_k e^{-|\theta_k|^2} c_k(g)[1 - e^{i\varphi}].
$$

Since $g \in \mathfrak{M}([0,1])^\star$, this last term is non vanishing except if $\theta_k = 0$ (which implies $\tilde{\theta}_k = 0 = \theta_k$) or if $\varphi \equiv 0(2\pi)$. In both cases, $F'(0) = 0 \iff \tilde{\theta}_k = \theta_k$. Thus, $\theta_k \neq \tilde{\theta}_k \implies \nabla F(0) \neq 0$ and we may find r small enough such that $|F|(z) > 0$ when $z \in B(0,r) \setminus \{0\}$. This ends the proof of identifiability.

Point 4: Minimal assumption on g We now briefly show that $\mathfrak{M}([0,1])^+$ is minimal for the identifiability of the model. Indeed, consider any integer k and a couple of functions $(f, \tilde{f})$ whose Fourier coefficients are $\theta_1(f) = \theta_1(\tilde{f}) = 1$ and $\theta_k(f) = \theta_k(\tilde{f})e^{-i\varphi} \neq 0$. By construction, for any $g \in \mathfrak{M}([0,1])^+$, $\mathbb{P}^1_{f,g} = \mathbb{P}^1_{\tilde{f},g}$. Following the definition of F above, when f and $\tilde{f}$ have only two frequencies switched on, it is thus necessary and sufficient to consider

$$ d_{TV}(\mathbb{P}^k_{f,g}, \mathbb{P}^k_{\tilde{f},g}) = \frac{1}{2\pi} \int_{\mathbb{C}} |F(z)| dz. $$

As an holomorphic function, F is null *iff* all its derivatives at 0 vanish. We have already seen that $F(0) = 0$ and $F'(0) = \theta_k e^{-|\theta_k|^2} c_k(g)[1 - e^{i\varphi}]$. A simple induction then shows that

$$ \forall j \leq m, F^{(j)}(0) = 0 \Longrightarrow \forall j \leq m, c_{kj}(g) = 0. $$

Thus, if $g \in \mathfrak{M}([0,1])^+$, there exists m such that $F^{(m)}(0) \neq 0$ and then F cannot be null, unless $\varphi = 0$. Hence, we see in this framework that

$$ g \in \mathfrak{M}([0,1])^+ \text{ and } \mathbb{P}_{f,g} = \mathbb{P}_{\tilde{f},g} \Longrightarrow f = \tilde{f}. $$

Conversely, if $g \notin \mathfrak{M}([0,1])^+$, there exists $k \in \mathbb{Z}$ such that $c_{jk}(g) = 0$ for all integers $j \in \mathbb{Z}$ but one cannot distinguish $\mathbb{P}_{f,g}$ from $\mathbb{P}_{\tilde{f},g}$ as soon as

$$ |\theta_k(f)| = |\theta_k(\tilde{f})|. $$

In particular, in this case the model is not identifiable. $\qquad\square$

The main difficulty was $d_{TV}(\mathbb{P}^1_{f,g}, \mathbb{P}^1_{\tilde{f},\tilde{g}}) \Longrightarrow g = \tilde{g}$ and we will use this to obtain a contraction rate for (f, g) around (f^0, g^0). The main inequality is (6.5) where $h = g - \tilde{g}$. and we first use it to study a contraction rate around g^0.

6.2. *Contraction rate of the posterior distribution around f^0 and g^0*

At this stage, we assume that $(f, g) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0,1])$, with $s \geq 1$ and $\nu > 1$. **Sketch of proof:** the next proof will be splitted in three parts. First we show that the joint contraction on $\mathbb{P}_{f,g}$ around $\mathbb{P}_{f^0,g^0}$ implies a contraction property of the first Fourier coefficient of f around $\theta_1(f^0)$. We use this last property to obtain the contraction of g around g^0. At last, we use the whole marginals of the complete process to derive contraction rates for the other Fourier coefficients of f around those of f^0.

6.2.1. *Contraction rate on the first Fourier coefficient*

Proposition 6.1. *If $(f, g) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0,1])$, then for a sufficiently large M:*

$$ \Pi^{GP}_n \left(\theta_1(f) \in B\left(\theta^0_1, M\epsilon^{1/3}_n\right)^c \Big| Y_1, \ldots, Y_n \right) \longrightarrow 0 \quad \mathbb{P}_{f^0,g^0} \text{ a.s.} \quad as \quad n \to +\infty. $$

Moreover, the contraction rate is $n^{-1/3 \times [\nu(s \wedge 1)/(2\nu(s \wedge 1)+1) \wedge s/(2s+2)]}(\log n)^{1/3}$.

Proof. In this proof, we will omit the dependence on f in the notation $\theta_1(f)$, which will be shortenned as θ_1. From the beginning of the proof of Theorem 2.3, one can show that for any θ_1 such that $0 < \eta < |\theta_1 - \theta_1^0| < \theta_1^0/2$, and for any $g \in \mathfrak{M}_\nu([0,1])$:

$$d_{TV}\left(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}\right) \geq d_{TV}\left(\mathbb{P}_{f,g}^1, \mathbb{P}_{f^0,g^0}^1\right),$$

owing to the variational definition of the Total Variation distance. Then

$$d_{TV}\left(\mathbb{P}_{f,g}^1, \mathbb{P}_{f^0,g^0}^1\right)$$
$$\geq \frac{1}{2\pi}\int_{B\left(0,\frac{|\theta_1-\theta_1^0|}{4}\right)}\left|\int_0^1 g(\alpha)e^{-|z-\theta_1 e^{i2\pi\varphi}|^2} - g^0(\alpha)e^{-|z-\theta_1^0 e^{i2\pi\varphi}|^2}d\varphi\right|dz$$
$$\geq \frac{\eta^2}{32}\left|e^{-(3\theta_1^0+\theta_1)^2/16} - e^{-(3\theta_1+\theta_1^0)^2/16}\right| \geq C(\theta_1^0)\eta^3,$$

for a small enough $C(\theta_1^0)$. Simple inclusions and Pinsker's inequality yields

$$\{\theta_1 \in B(0,\eta)^c\} \subset \{\theta_1 | d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \geq C(\theta_1^0)\eta^3\}$$
$$\subset \{\theta_1 | d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \geq C(\theta_1^0)\eta^3\}.$$

The proof is now achieved according to Theorem 2.2. $\qquad\square$

6.2.2. *Posterior contraction rate around g^0*

Proof of Theorem 2.4, i). Using Theorem 2.2 and for M large enough

$$\Pi_n^{GP}\left\{\mathbb{P}_{f,g} \quad s.t.\, d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \geq M\epsilon_n | Y_1, \dots Y_n\right\} \longrightarrow 0 \quad \text{as} \quad n \to +\infty.$$

Since $d_{TV}(\mathbb{P}_{\theta_1,g}^1, \mathbb{P}_{\theta_1^0,g^0}^1) \leq d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \leq d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0})$, we then get

$$\Pi_n^{GP}\left\{\mathbb{P}_{f,g} \quad s.t.\, d_{TV}(\mathbb{P}_{\theta_1,g}^1, \mathbb{P}_{\theta_1^0,g^0}^1) \geq M\epsilon_n | Y_1, \dots Y_n\right\} \longrightarrow 0 \quad \text{as} \quad n \to +\infty. \tag{6.6}$$

For any $g \in \mathfrak{M}_\nu([0,1])$, the triangular inequality yields

$$d_{TV}\left(\mathbb{P}_{\theta_1^0,g}^1, \mathbb{P}_{\theta_1,g}^1\right) + d_{TV}\left(\mathbb{P}_{\theta_1,g}^1, \mathbb{P}_{\theta_1^0,g^0}^1\right) \geq d_{TV}\left(\mathbb{P}_{\theta_1^0,g}^1, \mathbb{P}_{\theta_1^0,g^0}^1\right). \tag{6.7}$$

Let $\tilde{f}$ s.t. $\tilde{\theta} = \theta(\tilde{f})$ with $\tilde{\theta}_1 = \theta_1(f)$, and for any $k \in \mathbb{Z}\backslash\{1\}$, $\tilde{\theta}_k = \theta_k(f^0)$. Then Lemma 4.1 yields $d_{TV}\left(\mathbb{P}_{\theta_1^0,g}^1, \mathbb{P}_{\theta_1,g}^1\right) = d_{TV}\left(\mathbb{P}_{\tilde{f},g}^1, \mathbb{P}_{f^0,g}^1\right) \leq \frac{\|\tilde{f}-f^0\|}{\sqrt{2}} = \frac{|\theta_1-\theta_1^0|}{\sqrt{2}}$. Therefore, as $n \to +\infty$:

$$\Pi_n^{GP}\left(\mathbb{P}_{f,g}\, s.t.\, d_{TV}\left(\mathbb{P}_{\theta_1^0,g}^1, \mathbb{P}_{\theta_1,g}^1\right) \leq \frac{M}{\sqrt{2}}\epsilon_n^{1/3}\middle| Y_1, \dots, Y_n\right)$$
$$\geq \Pi_n^{GP}\left(\mathbb{P}_{f,g}\, s.t.\, |\theta_1 - \theta_1^0| \leq M\epsilon_n^{1/3}\middle| Y_1, \dots, Y_n\right) \longrightarrow 1 \tag{6.8}$$

In conclusion, (6.6),(6.7) and (6.8) shows that for M large enough:

$$\Pi_n^{GP}\left(\mathbb{P}_{f,g}\ s.t.\ d_{TV}\left(\mathbb{P}_{\theta_1^0,g}^1,\mathbb{P}_{\theta_1^0,g^0}^1\right) \le M\epsilon_n^{1/3}\,\bigg|\,Y_1,\dots,Y_n\right) \longrightarrow 1 \quad \text{as} \quad n \to +\infty.$$

We now use (6.5) applied with $\theta_1 = \theta_1^0$ to obtain our consistency rate:

$$d_{TV}(\mathbb{P}_{\theta_1^0,g}^1,\mathbb{P}_{\theta_1^0,g^0}^1) \ge \frac{1}{8\pi^2}\sum_{n\in\mathbb{Z}}|c_n(g-g^0)|^2\int_0^{+\infty}\rho e^{-(\rho+\theta_1^0)^2}A_n(2\rho\theta_1^0)^2 d\rho. \quad (6.9)$$

Now, equivalents given by Lemma C.1 (see Appendix C) shows that the main part of the integral above corresponds to $a \in [0, c\sqrt{n}]$. One can find κ such that

$$\int_0^{+\infty}\rho e^{-(\rho+\theta_1^0)^2}A_n(2\rho\theta_1^0)^2 d\rho$$

$$\ge \int_0^{\frac{\sqrt{n}}{2\theta_1^0}}\frac{4\pi^2\rho^{2n+1}\{\theta_1^0\}^{2n}}{n!^2}e^{-(\rho+\theta_1^0)^2}\left(1-\kappa\frac{[2\rho\theta_1^0]}{n}\right)^2 d\rho$$

$$\ge \left(1-\frac{\kappa}{\sqrt{n}}\right)^2\frac{4\pi^2\{\theta_1^0\}^{2n}}{n!^2}e^{-\left(\theta_1^0+\frac{\sqrt{n}}{2\theta_1^0}\right)^2}\int_0^{\frac{\sqrt{n}}{2\theta_1^0}}\rho^{2n+1}d\rho.$$

Now, we can apply the Stirling formula to obtain:

$$\frac{4\pi^2\{\theta_1^0\}^{2n}}{n!^2}e^{-\left(\theta_1^0+\frac{\sqrt{n}}{2\theta_1^0}\right)^2}\int_0^{\frac{\sqrt{n}}{2\theta_1^0}}\rho^{2n+1}d\rho$$

$$\sim \frac{4\pi^2\{\theta_1^0\}^{2n}}{(n/e)^{2n}2\pi n}e^{-\left(\theta_1^0+\frac{\sqrt{n}}{2\theta_1^0}\right)^2}\frac{\left(\sqrt{n}/(2\theta_1^0)\right)^{2n+2}}{2n+2}$$

$$\sim \frac{2\pi}{n(2n+2)}e^{-2n\log\left[\frac{n}{e\theta_1^0}\right]-\left(\theta_1^0+\frac{\sqrt{n}}{2\theta_1^0}\right)^2+(n+1)\log\left[\frac{n}{4\{\theta_1^0\}^2}\right]},$$

which is lower bounded by $C(\theta_1^0)e^{-n\log(n)}$. Such lower bound in (6.9) yields

$$d_{TV}(\mathbb{P}_{\theta_1^0,g}^1,\mathbb{P}_{\theta_1^0,g^0}^1) \ge \tilde{c}\sum_{k\in\mathbb{Z}}|c_k(g-g^0)|^2 e^{-ck\log k},$$

for $(c,\tilde{c})$ sufficiently small. We now end the proof of the Theorem: choose a frequency cut-off k_n that depends on n and remark that

$$\forall g \in \mathfrak{M}_\nu([0,1]) \quad \|g-g^0\|^2 = \sum_{|\ell|\le k_n}|c_\ell(g-g^0)|^2 + \sum_{|\ell|>k_n}|c_\ell(g-g^0)|^2$$

$$\lesssim e^{ck_n\log k_n}\sum_{|\ell|\le k_n}|c_\ell(g-g^0)|^2 e^{-c\ell\log\ell} + k_n^{-2\nu}$$

$$\lesssim e^{ck_n\log k_n}d_{TV}(\mathbb{P}_{\theta_1^0,g}^1,\mathbb{P}_{\theta_1^0,g^0}^1) + k_n^{-2\nu}.$$

Equation (6.8) implies that the above term is lower than $e^{ck_n \log k_n} \epsilon_n^{1/3} + k_n^{-2\nu}$ up to a multiplicative constant, with probability close to 1 as n goes to $+\infty$. The optimal choice for k_n yields $[k_n + 2\nu] \log k_n = \frac{1}{3} \log \frac{1}{\epsilon_n}$, which ensures that

$$\Pi_n \left\{ g \, s.t. \, \|g - g^0\|^2 \le M \log(n)^{-2\nu} | Y_1, \dots Y_n \right\} \longrightarrow 1 \, \text{as} \, n \to +\infty.$$

$\square$

In the last lines, we use the knowledge of ν as well as the radius A of the Sobolev ellipsoid $\mathfrak{M}_\nu([0,1])(2A)$ to build a suitable threshold k_n. But we cannot control easily the posterior weights on $\mathfrak{M}_\nu([0,1])(2A)$ from the posterior around $\mathbb{P}_{f^0, g^0}$: that's why it is difficult to conclude with an adaptive prior.

6.2.3. Posterior contraction rate around f^0

We then obtain a result for the neighborhoods of f^0. Proposition 6.1 leads to a polynomial order on θ_1. It is far from being the case for the other frequencies.

Proof of Theorem 2.4, ii). The proof of $ii)$ is inspired from the one of $i)$.

Point 1: Triangular inequality For any $f \in \mathcal{F}_s$, we have for any $k \in \mathbb{Z}$:

$$d_{TV}(\mathbb{P}^k_{f,g^0}, \mathbb{P}^k_{f^0,g^0}) \le d_{TV}(\mathbb{P}^k_{f,g^0}, \mathbb{P}^k_{f,g}) + d_{TV}(\mathbb{P}^k_{f^0,g^0}, \mathbb{P}^k_{f,g}),$$

and if $\epsilon_n \ll \log(n)^{-\nu}$, we have as $n \longrightarrow +\infty$:

$$\Pi_n^{GP} \left(\forall k \in \mathbb{Z} \quad d_{TV}(\mathbb{P}^k_{f,g}, \mathbb{P}^k_{f^0,g^0}) < M\epsilon_n \Big| Y_1, \dots, Y_n \right) \longrightarrow 1 \qquad (6.10)$$

Point 2: $\Pi_n^{GP} \left(\sup_{k \in \mathbb{Z}} d_{TV}(\mathbb{P}^k_{f,g^0}, \mathbb{P}^k_{f,g}) < M \log(n)^{-\nu} \Big| Y_1, \dots, Y_n \right) \to 1$ We can use the Cauchy-Schwarz inequality as follows:

$$d_{TV}(\mathbb{P}^k_{f,g^0}, \mathbb{P}^k_{f,g}) = \frac{1}{2\pi} \int_{\mathbb{C}} \left| \int_0^{2\pi} e^{-|z - \theta_k e^{ik\varphi}|^2} [g(\varphi) - g^0(\varphi)] d\varphi \right| dz$$

$$\le \frac{\|g - g^0\|}{2\pi} \int_{\mathbb{C}} \left[\int_0^{2\pi} e^{-2|z - \theta_k e^{ik\varphi}|^2} d\varphi \right]^{1/2} dz$$

Young's inequality yields $|z - \theta_k e^{ik\varphi}|^2 = |z|^2 + |\theta_k|^2 - 2\Re \left(\bar{z} \theta_k e^{ik\varphi} \right) \ge \frac{|z|^2}{2} - |\theta_k|^2$:

$$d_{TV}(\mathbb{P}^k_{f,g^0}, \mathbb{P}^k_{f,g}) \le \frac{\|g - g^0\|}{2\pi} \int_{\mathbb{C}} \left(e^{-|z|^2 + 2|\theta_k|^2} \right)^{1/2} dz \le \|g - g^0\| \frac{e^{|\theta_k|^2}}{2}. \quad (6.11)$$

We denote $\mathcal{A}_n = \{(f,g) : \exists k \in \mathbb{Z} \quad d_{TV}(\mathbb{P}^k_{f,g^0}, \mathbb{P}^k_{f,g}) \ge M \log(n)^{-\nu}\}$ and the two sets $\mathcal{B} = \{f : \forall k \in \mathbb{Z} \quad |\theta_k| \le |\theta_k^0| + M \log(n)^{-\nu}\}$ and $\mathcal{C} = \{f : \forall k \in \mathbb{Z} \quad |\theta_k^0| \le |\theta_k| + M \log(n)^{-\nu}\}$. Let k and θ_k such that $|\theta_k| > |\theta_k^0| + M \log(n)^{-\nu}$, then

$$d_{TV}(\mathbb{P}^k_{f^0,g^0}, \mathbb{P}^k_{f,g}) = \frac{1}{2\pi} \int_{\mathbb{C}} \left| \int_0^{2\pi} \left[e^{-|z - \theta_k e^{ik\varphi}|^2} g(\varphi) - e^{-|z - \theta_k^0 e^{ik\varphi}|^2} g^0(\varphi) \right] d\varphi \right| dz.$$

For any $z \in B_n := B\left(0, \frac{M \log(n)^{-\nu}}{3}\right)$, one has for any $\varphi \in [0, 2\pi]$

$$|z - \theta_k^0 e^{ik\varphi}| \leq \frac{M \log(n)^{-\nu}}{3} + |\theta_k^0| \leq 2\frac{M \log(n)^{-\nu}}{3} + |\theta_k^0|$$

$$\leq |\theta_k| - \frac{M \log(n)^{-\nu}}{3} \leq |z - \theta_k e^{ik\varphi}|.$$

Hence if $|\theta_k| \geq |\theta_k^0| + M \log(n)^{-\nu}$, one has

$$d_{TV}(\mathbb{P}_{f^0,g^0}^k, \mathbb{P}_{f,g}^k)$$

$$\geq \frac{1}{2\pi} \int_{B_n} \left| \int_0^{2\pi} \left[e^{-|z - \theta_k e^{ik\varphi}|^2} g(\varphi) - e^{-|z - \theta_k^0 e^{ik\varphi}|^2} g^0(\varphi) \right] d\varphi \right| dz$$

$$\geq \frac{1}{2\pi} \int_{B_n} e^{-[|\theta_k^0| + M\log(n)^{-\nu}/3]^2} - e^{-[|\theta_k^0| + 2M\log(n)^{-\nu}/3]^2} dz$$

$$\geq c|\theta_k^0|^2 e^{-|\theta_k^0|^2} \log(n)^{-3\nu},$$

for a sufficiently small absolute constant $c > 0$. Since $(\theta_k^0)_{k \in \mathbb{Z}}$ is bounded, for n large enough, we know that $\|\theta^0\|^2 e^{-\inf_k |\theta_k^0|^2} \log(n)^{-3\nu} \gg \epsilon_n$ and (6.10) implies

$$\Pi_n^{GP}\left(\mathcal{B}^c | Y_1, \ldots, Y_n\right) \longrightarrow 0 \quad \text{as} \quad n \longrightarrow +\infty. \tag{6.12}$$

A similar argument yields

$$\Pi_n^{GP}\left(\mathcal{C}^c | Y_1, \ldots, Y_n\right) \longrightarrow 0 \quad \text{as} \quad n \longrightarrow +\infty.$$

Gathering now (6.12) and (6.11), we get for a sufficiently large M

$$\Pi_n^{GP}\left(\mathcal{A}_n | Y_1, \ldots, Y_n\right) = \Pi_n^{GP}\left(\mathcal{A}_n \cap \mathcal{B} \cap \mathcal{C} | Y_1, \ldots, Y_n\right)$$

$$+ \Pi_n^{GP}\left(\mathcal{A}_n \cap (\mathcal{B} \cap \mathcal{C})^c | Y_1, \ldots, Y_n\right)$$

$$\leq \Pi_n^{GP}\left(\|g - g^0\| \geq M e^{-\left(1 + \sup_k |\theta_k^0|^2\right)} \log(n)^{-\nu}\right)$$

$$+ \Pi_n^{GP}\left(\mathcal{B}^c | Y_1, \ldots, Y_n\right) + \Pi_n^{GP}\left(\mathcal{C}^c | Y_1, \ldots, Y_n\right)$$

Now, Theorem 2.4 $i)$ leads, as $n \longrightarrow +\infty$, to

$$\Pi_n^{GP}\left(\sup_{k \in \mathbb{Z}} d_{TV}(\mathbb{P}_{f,g^0}^k, \mathbb{P}_{f,g}^k) < M \log(n)^{-\nu} \middle| Y_1, \ldots, Y_n\right) \longrightarrow 1. \tag{6.13}$$

Point 3: Contraction of θ_k near θ_k^0 Arguments of Point 2 shows that

$$\Pi_n^{GP}\left(f : \forall k \in \mathbb{Z} \quad \left||\theta_k| - |\theta_k^0|\right| < M \log(n)^{-\nu}\right) \longrightarrow 1 \quad \text{as} \quad n \longrightarrow +\infty.$$

We now study the situation when $\left||\theta_k| - |\theta_k^0|\right| < M \log(n)^{-\nu}$, and we can write $\theta_k = \theta_k^0 e^{i\varphi} + \xi_n$ where ξ_n is a complex number such that $|\xi_n| \leq M \log(n)^{-\nu}$.

$$d_{TV}(\mathbb{P}_{f,g^0}^k, \mathbb{P}_{f^0,g^0}^k) = \frac{1}{2\pi} \int_{\mathbb{C}} \left| \underbrace{\int_0^{2\pi} \left[e^{-|z - \theta_k e^{ik\alpha}|^2} - e^{-|z - \theta_k^0 e^{ik\alpha}|^2} \right] g^0(\alpha) d\alpha}_{:=F(z)} \right| dz$$

Indeed, $F(0) = o(1)$ and a Taylor expansion near 0 yields that

$$
\begin{aligned}
F(z) = 2e^{-|\theta_k^0|^2} \int_0^{2\pi} & \left[1 + \Re\mathrm{e} \left(z\bar{\theta}_k e^{-ik\alpha} \right) - (1 + \Re\mathrm{e} \left(z\bar{\theta}_k^0 e^{-ik\alpha} \right) \right] g^0(\alpha) d\alpha \\
& + o(|z|) + \mathcal{O}(|\xi_n|).
\end{aligned}
$$

If we write $\theta_k = \theta_k^0 e^{\mathrm{i}\varphi} + \mathcal{O}(\log(n)^{-\nu})$, for η small enough such that $|z| \leq \eta$:

$$
|F(z)| \geq c e^{-|\theta_k^0|^2} \left| \sin(\varphi/2) \Re\mathrm{e} \left(z\mathrm{i}e^{\mathrm{i}\varphi/2}\bar{\theta}_k^0 c_{-k}(g^0) \right) \right| + \mathcal{O}(\log(n)^{-\nu})
$$

Now, denote $\bar{u} = \frac{\mathrm{i}e^{\mathrm{i}\varphi/2}\bar{\theta}_k^0 c_{-k}(g^0)}{|\theta_k^0| \times |c_{-k}(g^0)|}$ which is a complex number of norm 1, and let $v = \bar{u}e^{\mathrm{i}\pi/2}$. The vector v is orthogonal to $\bar{u}$ and z may be decomposed as

$$
z = a\bar{u} + bv.
$$

We then choose $|b| < |a|/2$ and denotes $\mathcal{R}_a$ the area where z is living. For $a < \eta$ small enough, there exists an absolute constant c independent of k such that

$$
\begin{aligned}
d_{TV}(\mathbb{P}_{f,g^0}^k, \mathbb{P}_{f^0,g^0}^k) \geq \int_{\mathcal{R}_a} |F(z)| & \geq c\eta^3 e^{-|\theta_k^0|^2} |\sin(\varphi/2)||\bar{\theta}_k^0||c_{-k}(g^0)| \\
& + \mathcal{O}\left(\log(n)^{-\nu} \right).
\end{aligned}
$$

Since $|\theta_k - \theta_k^0| = 2|\sin(\varphi/2)||\theta_k^0| + \mathcal{O}\left(\log(n)^{-\nu} \right)$, we deduce that :

$$
d_{TV}(\mathbb{P}_{f,g^0}^k, \mathbb{P}_{f^0,g^0}^k) \geq c\eta^3 e^{-|\theta_k^0|^2} |c_{-k}(g^0)||\theta_k - \theta_k^0| + \mathcal{O}\left(\log(n)^{-\nu} \right). \qquad (6.14)
$$

We conclude using (6.13) and (6.14) that for a sufficiently large M, as $n \longrightarrow +\infty$:

$$
\Pi_n \left(f : \sup_{k \in \mathbb{Z}} \left| (\theta_k - \theta_k^0) c_{-k}(g^0) \right| < M \log(n)^{-\nu} \middle| Y_1, \ldots, Y_n \right) \longrightarrow 1 \qquad (6.15)
$$

Point 4: Contraction on f^0   We now produce a standard thresholding proof:

$$
\begin{aligned}
\|f - f^0\|^2 & \\
= & \sum_{|\ell| > k_n} |\theta_\ell - \theta_\ell^0|^2 + \sum_{|\ell| \leq k_n} |\theta_\ell - \theta_\ell^0|^2 \lesssim k_n^{-2s} + \sum_{|\ell| \leq k_n} \frac{|\theta_\ell - \theta_\ell^0|^2 |c_{-\ell}(g^0)|^2}{|c_{-\ell}(g^0)|^2} \\
\lesssim & \ k_n^{-2s} + k_n^{2\beta} \sum_{|\ell| \leq k_n} |\theta_\ell - \theta_\ell^0|^2 |c_{-\ell}(g^0)|^2 \\
\lesssim & \ k_n^{-2s} + k_n^{2\beta+1} \sup_{|\ell| \leq k_n} |\theta_\ell - \theta_\ell^0|^2 |c_{-\ell}(g^0)|^2
\end{aligned}
$$

Hence, the frequency cut-off $k_n = (\log n)^{\frac{2\nu}{2\beta+2s+1}}$ and (6.15) implies

$$
\Pi_n \left(f : \|f - f^0\|^2 \leq M (\log n)^{-4s\nu/(2s+2\beta+1)} \middle| Y_1, \ldots, Y_n \right) \longrightarrow 1
$$

as $n \longrightarrow +\infty$ for M large enough. This last result is the desired inequality. $\qquad \square$

Remark 6.1. *The lower bound obtained on $d_{TV}(\mathbb{P}^k_{f,g^0}, \mathbb{P}^k_{f^0,g^0})$ is important to understand how one can build an appropriate net of functions $(f_j, g_j) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0,1])$ hard to distinguish with the L^2 distance. When $|\theta_k| \neq |\theta^0_k|$, it is quite easy to distinguish the two hypotheses. On the contrary, when the modulus are the same, the behaviour of the Fourier coefficients of g^0 becomes important. This is a clue to exhibit a "difficult" net.*

6.3. Lower bound from a frequentist point of view

We have chosen to use the Fano Lemma (see [IH81] for instance) instead of Le Cam's method, since we will only be able to find some *discrete* (instead of convex) set of pairs (f_j, g_j) in $\mathcal{F}_s \times \mathfrak{M}_\nu([0,1])$ closed according to the Total Variation distance.

Proof of Theorem 2.5. We are looking for a set $(f_j, g_j)_{j=1...p_n}$ such that each $\mathbb{P}_{f_j,g_j}$ are closed together with rather different functional parameters f_j or g_j. Reading carefully the Bayesian contraction rate is informative to build p_n hypotheses which are difficult to distinguish. First, we know that since each f_j should belong to $\mathcal{F}_s$, we must impose for any f_j that $\theta_1(f_j) > 0$. Proposition 6.1 shows that two laws $\mathbb{P}_{f_j,g_j}$ and $\mathbb{P}_{f_{j'},g_{j'}}$ are statistically very different as soon as $\theta_1(f_j) \neq \theta_1(f_{j'})$. Then we build our net using a common value for $\theta_1(f_j)$:

$$\forall j \in \{1 \ldots p_n\} \qquad \theta_1(f_j) = 1.$$

Point 1: Net of functions $(f_j)_{j=1...p_n}$ We choose the following construction

$$\forall j \in \{1 \ldots p_n\} \qquad f_j(x) = e^{\mathrm{i}2\pi x} + p_n^{-s} e^{\mathrm{i}2\frac{(j-1)}{p_n}\pi} e^{\mathrm{i}2\pi p_n x}. \qquad (6.16)$$

The number of elements p_n will be next adjusted and will grow to $+\infty$. Our construction naturally satisfies that each $f_j \in \mathcal{F}_s$ since the modulus of the p_n-th Fourier coefficient is of size p_n^{-s}. At last, we have: $\forall (j, j') \in \{1 \ldots p_n\}^2, j \neq j'$:

$$\|f_j - f_{j'}\|^2 \geq p_n^{-2s} \times \left| e^{\mathrm{i}2\pi/p_n} - 1 \right|^2 \geq 4p_n^{-2s} \sin^2(\pi/p_n) \sim_{n \mapsto +\infty} 4\pi^2 p_n^{-2s-2}.$$

Point 2: Net of measures $(g_j)_{j=1...p_n}$ The cornerstone of the lower bound is how to adjust the measures of the random shifts to make the distributions $\mathbb{P}_{f_j,g_j}, j = 1 \ldots p_n$, as close as possible. First, remark that we will still use Total Variation instead of the entropy between laws since Kullback Leibler distance is difficult to handle with mixtures, we use the chain of inequalities: $\forall j \neq j'$

$$d_{TV}\left(\mathbb{P}_{f_j,g_j}, \mathbb{P}_{f_{j'},g_{j'}}\right) \leq \eta \quad \Rightarrow \quad d_H\left(\mathbb{P}_{f_j,g_j}, \mathbb{P}_{f_{j'},g_{j'}}\right) \leq \sqrt{2\eta}$$
$$\Rightarrow \quad d_{KL}\left(\mathbb{P}_{f_j,g_j}, \mathbb{P}_{f_{j'},g_{j'}}\right) \lesssim \sqrt{\eta} \log \frac{1}{\eta}.$$

Hence, from the tensorisation of the entropy, we must find a net such that $d_{TV}\left(\mathbb{P}_{f_j,g_j}, \mathbb{P}_{f_{j'},g_{j'}}\right) \leq \eta_n$ with $-\sqrt{\eta_n} \log \eta_n = \mathcal{O}(1/n)$ to obtain a tractable

application of the Fano Lemma. It imposes to find mixture laws such that $d_{TV}\left(\mathbb{P}_{f_j,g_j}, \mathbb{P}_{f_{j'},g_{j'}}\right) \lesssim \frac{1}{(n\log n)^2}$, it is sufficient to build $(g_j)_{j=1\ldots p_n}$ satisfying

$$\forall j \in \{1\ldots p_n\} \qquad d_{TV}\left(\mathbb{P}_{f_j,g_j}, \mathbb{P}_{f_1,g_1}\right) \lesssim \frac{1}{(n\log n)^2}. \qquad (6.17)$$

For sake of convenience, we replace p_n by p. In a similar way, $\theta_p^j = \theta_p(f_j)$ given by $\theta_p^j = e^{i2\pi\alpha_j}\theta_p^1$ where $\alpha_j = \frac{j-1}{p_n}$. From (6.16), we have

$$d_{TV}\left(\mathbb{P}_{f_j,g_j}, \mathbb{P}_{f_1,g_1}\right) = \frac{1}{2\pi^2}\int_{\mathbb{C}\times\mathbb{C}}\left|\int_0^1 e^{-|z_1-e^{i2\pi\varphi}|^2-|z_2-e^{i2\pi p\varphi}\theta_p^1|^2}g_1(\varphi)d\varphi\right.$$
$$\left. - \int_0^1 e^{-|z_1-e^{i2\pi\varphi}|^2-|z_2-e^{i2\pi p\varphi}\theta_p^j|^2}g_j(\varphi)d\varphi\right|dz_1dz_2$$

Now, we use the smoothness of Gaussian densities, denote F defined on $\mathbb{R}^4$ by

$$F(x_1,y_1,x_2,y_2) := \int_0^1 \left(e^{-\|z-\theta^1\bullet\varphi\|^2}g_1(\varphi) - e^{-\|z-\theta^j\bullet\varphi\|^2}g_j(\varphi)\right)d\varphi,$$

where $z = (x_1+iy_1, x_2+iy_2)$ and $\theta^j \bullet \varphi = (e^{i2\pi\varphi}, \theta_j^p e^{i2\pi p\varphi})$.

To control F, we adapt the proof of Proposition 4.2. Only the sketch of the proof is given here. We use a truncation in $\mathcal{R}_{R_n} := B_{\mathbb{R}^2}(0, R_n)^2$. Outside $\mathcal{R}_{R_n}$, we use the key inequality given by (4.2). Inside $\mathcal{R}_{R_n}$ we need to satisfy some constraints on the Fourier coefficients. Since here the only non null Fourier coefficients are of order 1 and p, we have finally to ensure that

$$\forall m,l \leq d \quad \forall(s,\tilde{s}) \in \{-1;+1\}^2 \qquad c_{sm+\tilde{s}\ell p}(g_j)e^{\tilde{s}\ell\alpha_j} = c_{sm+\tilde{s}\ell p}(g_1). \qquad (6.18)$$

Hence, the maximum size of d is $d = p/4$. We have from (6.18) and (4.2):

$$d_{TV}\left(\mathbb{P}_{f_j,g_j}, \mathbb{P}_{f_1,g_1}\right) = \frac{1}{2\pi^2}\int_{\mathcal{R}_{R_n}}|F(x_1,y_1,x_2,y_2)|dx_1dy_1dx_2dy_2$$
$$+ \frac{1}{2\pi^2}\int_{\mathcal{R}_{R_n}^c}|F(x_1,y_1,x_2,y_2)|dx_1dy_1dx_2dy_2$$
$$\lesssim e^{-R_n^2/2} + \left(\frac{(eR_n)^{p/4}}{(p/4)^{p/4}}\right)^4 \lesssim e^{-R_n^2/2} + \frac{(eR_n)^p}{(p/4)^p},$$

where the last point is deduced from inequality (4.2). We choose now R_n such as $R_n := 3\sqrt{\log n}$ to obtain that $e^{-R_n^2/2} \ll (n\log n)^{-2}$ as required in condition (6.17). Now, we control the last term of the last inequality: the Stirling formula yields

$$\frac{(eR_n)^p}{(p/4)^p} \lesssim e^{p\log(3\sqrt{\log n})-p\log p/4}.$$

If one chooses $p_n = \kappa \log n$ with $\kappa > 12$, we then obtain that

$$d_{TV}\left(\mathbb{P}_{f_j,g_j}, \mathbb{P}_{f_1,g_1}\right) \lesssim e^{-Cp_n\log p_n} \lesssim (n\log n)^{-2}.$$

Such a choice of R_n and p_n ensures that (6.17) is fulfilled.

Point 3: Fano's Lemma application   We have to make sure that our conditions (6.18) for the Fourier coefficients of the g_j's lead to valid densities. Take for instance, for some $\beta > \nu + 1/2$,

$$a = \frac{A}{\left(2 \sum_{k \geq 1} k^{-2\beta+2\nu}\right)^{1/2}} \wedge \frac{1}{\left(2 \sum_{k \geq 1} k^{-2\beta}\right)^{1/2}}.$$

Then take $c_0(g_j) := 1$, and $\forall k \in \mathbb{Z}^\star$, $c_k(g_1) := a|k|^{-\beta}$. This ensures that

$$\sum_{k \in \mathbb{Z}^\star} |c_k(g_j)| \leq 1,$$

and therefore all g_j remains nonnegative. Note that the densities g_j fulfill the condition of Theorem 2.4; the lower bounds below are also valid in this slightly smaller model. We then conclude our proof: we aim to apply the Fano Lemma (see *e.g.* [IH81]) with $\alpha_n = p_n^{-2s-2}$ and $\beta_n = \mathcal{O}(1)$ for the net $(f_j)_{j=1...p_n}$:

$$\liminf_{n \to +\infty} (\log n)^{2s+2} \inf_{\hat{f} \in \mathcal{F}_s} \sup_{(f,g) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0,1])} \|\hat{f} - f\|^2 \geq c.$$

Our construction implies also that each g_j are rather different each others since one has for instance, $c_p(g_j)e^{i\alpha_j} = c_p(g_1) = c_p(g_{j'})e^{i\alpha_{j'}}$. Thus

$$\forall j \neq j' \qquad \|g_j - g_{j'}\|_2^2 \geq |c_p(g_j) - c_p(g_{j'})|^2 = p^{-2\nu} |e^{\alpha_j} - e^{\alpha_{j'}}|^2 \geq cp^{-2\nu-2}.$$

Applying the Fano Lemma to $(g_j)_{j=1...p_n}$ implies

$$\liminf_{n \to +\infty} (\log n)^{2\nu+2} \inf_{\hat{g} \in \mathcal{F}_s} \sup_{(f,g) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0,1])} \|\hat{g} - g\|^2 \geq c.$$

$\square$

Appendix A: Link between Kullback-Leibler and Hellinger neighborhoods

We relate Hellinger neighborhoods and Kullback-Leibler ones following [WS95].

Proposition A.1. *i*) *For any* $\mathbb{P}_{f^0,g^0} \in \mathcal{P}$, *and for any* $f \in \mathcal{H}_s$ *such that* $\|f\| \leq 2\|f^0\|$, *and any* $g \in \mathfrak{M}([0,1])$, *define* $q = \frac{d\mathbb{P}_{f^0,g^0}}{d\mathbb{P}_{f,g}}$ *and*

$$M_\delta^2 := \int_{q \geq e^{1/\delta}} q^{\delta+1} d\nu, \qquad \forall \delta \geq 0. \tag{A.1}$$

There exists $\delta \in (0,1]$ *such that* M_δ^2 *is uniformly bounded with respect to* f.
ii) *There exists a constant* C *such that for* ϵ *small enough, one has*

$$\mathcal{V}_\epsilon(\mathbb{P}_{f^0,g^0}, d_H) \subset \mathcal{V}_{C\epsilon \log(\epsilon^{-1})}(\mathbb{P}_{f^0,g^0}, d_{KL}).$$

In our model, Hellinger neighborhoods are almost Kullback-Leibler ones (up to a log term) since a sufficiently large moment exists for q ($q \log q/q^{1+\delta}$ tends to 0 when q tends to $+\infty$, and a second order expansion of $q \log q - q + 1$ around 1 yields a term similar to $[\sqrt{q} - 1]^2$).

Proof. This proposition uses a corollary of the Rice formula (see *e.g.* [AW09]) stated in Lemma A.1. We use Girsanov's formula (4.3):

$$
\frac{d\mathbb{P}_{f^0,g^0}}{d\mathbb{P}_{f,g}}(Y) = \exp\left(\|f\|^2 - \|f^0\|^2\right) \frac{\int_0^1 e^{2\Re e\langle f^{0,-\alpha_1}, f^{0,-\tau}+dW\rangle} dg^0(\alpha_1)}{\int_0^1 e^{2\Re e\langle f^{-\alpha_2}, f^{0,-\tau}+dW\rangle} dg(\alpha_2)}
$$

$$
\leq \quad e^{\|f\|^2 - \|f^0\|^2} \exp\left(2 \sup_{\alpha_1,\alpha_2} \Re e\langle f^{0,-\alpha_1} - f^{-\alpha_2}, f^{0,-\tau}\rangle\right) \leq e^{(\|f\|+\|f^0\|)^2} e^{Z_1+Z_2}
$$

where the last line is obtained using Cauchy-Schwarz's inequality and :

$$
Z_1 := 2 \sup_{\alpha_1} \Re e\langle f^{0,-\alpha_1}, dW\rangle = 2 \sup_{\alpha_1} \Re e \int_0^1 \overline{f^0}(s - \alpha_1) dW_s,
$$

$$
Z_2 := 2 \sup_{\alpha_2} \Re e\langle -f^{-\alpha_2}, dW\rangle = 2 \sup_{\alpha_2} \Re e \int_0^1 -\overline{f}(s - \alpha_2) dW_s.
$$

We now set $\delta \in (0,1]$ and we define the trajectories $\mathcal{E}_\delta$ as

$$
\mathcal{E}_\delta := \left\{ Y = f^{0,-\tau} + W \quad | \quad \frac{d\mathbb{P}_{f^0,g^0}}{d\mathbb{P}_{f,g}}(Y) \geq e^{1/\delta} \right\}.
$$

Hence, following the definition of M_δ^2 of (A.1), we have

$$
M_\delta^2 = \mathbb{E}_{Y \sim \mathbb{P}_{f^0,g^0}} \left[\left(\frac{d\mathbb{P}_{f^0,g^0}}{d\mathbb{P}_{f,g}}(Y) \right)^\delta \mathbf{1}_{Y \in \mathcal{E}_\delta} \right].
$$

For δ small enough, ($\delta \leq \frac{1}{2(\|f\|+\|f^0\|)^2}$):

$$
\begin{aligned}
M_\delta^2 &\leq e^{\delta(\|f\|+\|f^0\|)^2} \mathbb{E} e^{\delta(Z_1+Z_2)} \mathbf{1}_{Z_1+Z_2 \geq \frac{1}{\delta} - (\|f\|+\|f^0\|)^2} \\
&\leq e^{\delta(\|f\|+\|f^0\|)^2} \mathbb{E} e^{\delta(Z_1+Z_2)} \mathbf{1}_{Z_1+Z_2 \geq \frac{1}{2\delta}} \\
&\leq e^{\delta(\|f\|+\|f^0\|)^2} \mathbb{E} e^{\delta(Z_1+Z_2)} \mathbf{1}_{e^{\delta(Z_1+Z_2)} \geq \sqrt{e}}.
\end{aligned}
$$

Integrating by parts the last expectation, the use of Lemma A.1 yields

$$
M_\delta^2 \leq e^{\delta(\|f\|+\|f^0\|)^2} \int_{\sqrt{e}}^{+\infty} \mathbb{P}\left(e^{\delta(Z_1+Z_2)} > u\right) du
$$

$$
= e^{\delta(\|f\|+\|f^0\|)^2} \int_{\sqrt{e}}^{+\infty} \left[\mathbb{P}\left(\frac{Z_1}{2} \geq \frac{\log u}{4\delta}\right) + \mathbb{P}\left(\frac{Z_2}{2} \geq \frac{\log u}{4\delta}\right) \right] du
$$

$$
M_\delta^2 \leq C(f^0, f) e^{\delta(\|f\|+\|f^0\|)^2} \int_{\sqrt{e}}^{+\infty} \left[e^{-\frac{\log^2(u)}{16\delta^2\|f^0\|^2}} + e^{-\frac{\log^2(u)}{16\delta^2\|f\|^2}} \right] du. \qquad (A.2)
$$

Now, we can choose δ non negative and small enough such that $M_\delta^2 < \infty$ since for $u \geq \sqrt{e}$, we have $e^{-\frac{\log^2(u)}{16\delta^2\|f^0\|^2}} \leq e^{-\frac{\log(u)}{32\delta^2\|f^0\|^2}} = u^{-1/32\delta^2\|f^0\|^2}$, which is an integrable function as soon as $\delta^2 < \frac{1}{32\|f^0\|^2}$. The same result holds replacing f^0 by f and M_δ^2 is uniformly bounded if $\|f\| \leq 2\|f^0\|$. $\qquad\square$

We now show that the technical inequality used in (A.2) is satisfied.

Lemma A.1. *Let W a complex standard Brownian motion and u a complex 1-periodic map of $\mathcal{H}_s$, of class $\mathcal{C}^2$. Then when $t/\|u\| \longrightarrow +\infty$, we have*

$$\mathbb{P}\left(\sup_\alpha \Re\langle u^{-\alpha}, dW \rangle > t\right) \lesssim \frac{\|u'\|}{2\pi\|u\|} \exp\left(\frac{-t^2}{\|u\|^2}\right).$$

In particular, if $u \in \mathcal{H}^\ell$, we have $\mathbb{P}\left(\sup_\alpha \Re\langle u^{-\alpha}, dW \rangle > t\right) \lesssim \frac{\ell}{2\pi} \exp\left(\frac{-t^2}{\|u\|^2}\right).$

Proof. We define the following Gaussian centered process,

$$\forall \alpha \in [0,1] \qquad X(\alpha) := \frac{\sqrt{2}\Re\left(\int_0^1 \overline{u}(s-\alpha)dW_s\right)}{\|u\|},$$

whose covariance function is given by $\Gamma(t) = \mathbb{E}[X(0)X(t)]$. Obviously, one has $\Gamma(0) = 1$ and Cauchy-Schwarz's inequality implies that $\Gamma(s) \leq \Gamma(0)$. Moreover, since Γ is $\mathcal{C}^1([0,1])$, we deduce that $\Gamma'(0) = 0$ and simple computation yields

$$\Gamma"(0) = \frac{\Re\left(\int_0^1 u'(s)u"(s)ds\right)}{\|u\|^2} = -\frac{\|u'\|^2}{\|u\|^2}.$$

Rice's formula (see exercice 4.2, chapter 4 of [AW09]) yields that

$$\mathbb{P}\left(\sup_\alpha X(\alpha) > t\right) \sim \frac{\|u'\|}{2\pi\|u\|}e^{-t^2/2} \text{ as } t \longrightarrow +\infty.$$

The second point is a simple consequence of inequality $\|u'\| \leq \ell\|u\|$. $\qquad\square$

Appendix B: Small ball probability for integrated Brownian bridge

Recall that p_v defined by (2.5) refers to the probability distribution which is proportionnal to e^v. We detail here a lower bound of the prior weight around g^0. As a log density model, it is enough to find a lower bound of the weight around w^0 if one writes $g^0 \propto e^{w^0}$ according to Lemma 3.1 of [vdVvZ08a]. Recall the notation of the prior weight $q_{\nu,A}(\mathcal{G}_\epsilon)$, where $\mathcal{G}_\epsilon$ were previously defined as:

$$\mathcal{G}_\epsilon := \left\{g \in \mathfrak{M}_\nu([0,1])(2A) : d_{TV}(g,g^0) \leq \epsilon\right\}.$$

Theorem B.1. *There exists c such that for ϵ small enough*

$$q_{\nu,A}(\mathcal{G}_\epsilon) \geq ce^{-\epsilon^{-\frac{1}{k_\nu+1/2}}}.$$

Proof. **Structure of the prior** We denote $w_0 := \log g^0$, which is a k_ν-differentiable function of $[0, 1]$, extended to a 1-periodic element of $\mathcal{C}^{k_\nu}(\mathbb{R})$. We define $\tilde{q}$ the prior defined by (2.4) on this class of periodic functions (and omit the dependence on ν and A for sake of simplicity). The prior $q_{\nu,A}$ is then derived from $\tilde{q}$ through (2.5). We can remark that our situation looks similar to the one described in paragraph 4.1 of [vdVvZ08a] for integrated brownian motion. The log-density w_0 is approximated by a "Brownian bridge started at random":

$$w = J_{k_\nu}(B) + \sum_{i=1}^{k_\nu} Z_i \psi_i,$$

where B is a real Brownian bridge between 0 and 1. We suppose B built as $B_t = W_t - tW_1$ on the basis of a Brownian motion W on $[0, 1]$. The key point of the operateur J is that $J_{k_\nu}(B)(0) = J_{k_\nu}(B)(1) = 0$ and $J_k(f)' = J_{k-1}(f) - \int_0^1 J_{k-1}(f)$. Hence, an induction argument yields $J_{k_\nu}(B)^{(j)}(0) = J_{k_\nu}(B)^{(j)}(1)$, whenever $j \in \{1, \ldots, k_\nu\}$. Hence, $J_{k_\nu}(B)$ and its first k_ν derivatives are 1-periodic. Since functions ψ_i are also 1-periodic and $\mathcal{C}^\infty(\mathbb{R})$, our prior $\tilde{q}$ generates admissible functions of $[0, 1]$ to approximate w_0. We will denote this set of admissible trajectories $\mathcal{C}_1^{k_\nu}$ (1-periodic functions which are k_ν times differentiable).

Transformation of the Brownian bridge Denote $\mathbb{B}_1$ the separable Banach space of Brownian bridge trajectories and $\mathbb{B}_2 = \mathbb{R}^{k_\nu+1}$. The map T : $(B, Z_0, \ldots, Z_{k_\nu}) \longmapsto J_{k_\nu}(B) + \sum_{i=0}^{k_\nu} Z_i \psi_i$ is *injective* from the Banach space $\mathbb{B} = \mathbb{B}_1 \times \mathbb{B}_2$ to the set $\underline{\mathbb{B}} := T(\mathbb{B})$. More precisely, by induction each map $J_k(B)$ may be decomposed as

$$\forall t \in [0, 1] \qquad J_k(B)(t) = I_k(W)(t) + \sum_{i=1}^{k+1} c_{i,k}(W)t^i, \tag{B.1}$$

where $c_{i,k}(W)$ are explicit linear functionals that depend on W_1 and on the collection $\left(\int_0^1 (1-t)^{k-j} W_t dt \right)_{1 \le j \le k}$ (and not on t), and I_k is the operator used in [vdVvZ08a] defined as $I_1(f) = \int_0^t f$ and $I_k = I_1 \circ I_{k-1}$ for $k \ge 2$. Hence,

$$\forall t \in [0, 1], \quad T(B, Z_1, \ldots, Z_{k_\nu})^{(k)}(t)$$
$$= W_t + c_{k,k}(W)k! + c_{k+1,k}(W)(k+1)!t + \sum_{i=0}^{k_\nu} Z_i \psi_i^{(k)}(t) \tag{B.2}$$

According to the Karhunen-Loeve expansion of Brownian bridge, and since each $\psi_i^{(k)}$ possesses a non vanishing cosinus term: $t \mapsto \cos(2\pi it)$, we then deduce that

$$T(B^1, Z_1^1, \ldots, Z_{k_\nu}^1) = T(B^2, Z_1^2, \ldots, Z_{k_\nu}^2)$$

implies that $Z_i^1 = Z_i^2$ for $i \in \{0, \ldots, k_\nu\}$, and next that $W^1 = W^2$ and $B^1 = B^2$.

Thus, it is possible to apply Lemma 7.1 of [vdVvZ08b] to deduce that the RKHS associated to the Gaussian process (2.4) in $\underline{\mathbb{B}}$ is $\underline{\mathbb{H}} := T\mathbb{H}$ where $\mathbb{H}$ is the

RKHS derived in the space $\mathbb{B} = \mathbb{B}_1 \times \mathbb{B}_2$. Moreover, the map T is an isometry from $\mathbb{H}$ to $\underline{\mathbb{H}}$ for the RKHS-norms. At last, $\mathbb{B}_1$ and $\mathbb{B}_2$ are independent and

$$\mathbb{H} := \left\{ (f, z) \in AC([0,1]) \times \mathbb{R}^{k_\nu + 1} : f(0) = f(1) = 0, \int_0^1 f'^2 < \infty \right\},$$

where $AC([0,1])$ is the set of absolutely continuous functions on $[0,1]$, $\mathbb{H}$ is endowed with the inner product: $\langle (f_1, z^1), (f_2, z^2) \rangle_{\mathbb{H}} := \int_0^1 f_1' f_2' + \langle z^1, z^2 \rangle_{\mathbb{R}^{k_\nu + 1}}$.

Extremal derivatives We study the process $b := \sum_{i=0}^{k_\nu} Z_i \psi_i$ and are looking for realizations of $(Z_i)_i$ that suitably matches arbitrarily values $w_0^{(j)}(0) = w_0^{(j)}(1)$. We can check that for any integer p: $\psi_k^{(2p)}(t) = (-1)^p (2\pi k)^{2p} \psi_k(t)$, and $\psi_k^{(2p+1)}(t) = (-1)^p (2\pi k)^{2p+1} [-\sin(2\pi k t) + \cos(2\pi k t)]$. Hence, the matching of $w_0^{(j)}(0)$ by $b^{(j)}(0)$ is quantified by

$$w_0^{(j)}(0) - b^{(j)}(0) = w_0^{(j)}(0) - \sum_{k=0}^{k_\nu} (-1)^{\lfloor j/2 \rfloor} (2\pi k)^j Z_k.$$

If one denotes $\alpha_k := 2\pi k$, the vector of derivatives as $d_0 := (w_0^{(j)}(0))_{j=0\ldots k_\nu}$, $Z = (Z_0, \ldots, Z_{k_\nu + 1})$ and the squared matrix of size $(k_\nu + 1) \times (k_\nu + 1)$:

$$A_0 := \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \alpha_1 & \alpha_2 & \ldots & \alpha_{k_\nu} \\ -\alpha_1^2 & -\alpha_2^2 & \ldots & -\alpha_{k_\nu}^2 \\ -\alpha_1^3 & -\alpha_2^3 & \ldots & -\alpha_{k_\nu}^3 \\ \alpha_1^4 & \alpha_2^4 & \ldots & \alpha_{k_\nu}^4 \\ \vdots & & & \end{pmatrix},$$

then we are looking for values of Z such that $d_0 = A_0 Z$. The matrix A_0 is invertible since it may be linked with the Vandermonde matrix. We can now establish that the support of the prior (adherence of $\underline{\mathbb{B}}$) is exactly $\mathcal{C}_1^{k_\nu}$. Indeed, the support of the transformed Brownian bridge $J_k(B)$ is included in elements of $\mathcal{C}_1^{k_\nu}$ with at the most $k+1$ constraints on the values of their $k_\nu + 1$ first derivatives at the point 0. These constraints are given by the coefficients $(c_{i,k_\nu})_{i=0\ldots k_\nu}$ in (B.2). From the invertibility of the matrix A_0, it is possible to match *any* term $w_0^{(j)}(0), 0 \leq j \leq k_\nu$ with the additional process b [see vdVvZ08b, section 10].

Small ball probability estimates We now turn into the core of the proof of the Theorem. Since the Total Variation distance is bounded from above by the Hellinger distance, an immediate application of Lemma 3.1 of [vdVvZ08a] shows that it is sufficient to find a lower bound of the $\tilde{q}(\tilde{\mathcal{G}}_\epsilon)$ where

$$\tilde{\mathcal{G}}_\epsilon := \left\{ w \in \mathcal{C}_1^{k_\nu}([0,1]) : \|w - w_0\|_\infty \leq \epsilon \right\}.$$

Following the argument of [KLL94] on *shifted* Gaussian ball, we have

$$\log\left(\tilde{q}\left(\tilde{\mathcal{G}}_\epsilon\right)\right) \geq - \inf_{h\in\underline{\mathbb{H}}:\|h-w_0\|_\infty\leq\epsilon} \|h\|^2_{\underline{\mathbb{H}}} - \log\tilde{q}\left(\|J_k(B)+b\|_\infty \leq \epsilon\right).$$

From the isometry T from $\mathbb{H}$ to $\underline{\mathbb{H}}$, we can write that the approximation term $\inf_{h\in\underline{\mathbb{H}}:\|h-w_0\|_\infty\leq\epsilon}\|h\|^2_{\underline{\mathbb{H}}}$ is of the same order as the approximation term that we can derive in $\mathbb{H}$, and following arguments of Theorem 4.1 in [vdVvZ08a]:

$$\inf_{h\in\underline{\mathbb{H}}:\|h-w_0\|_\infty\leq\epsilon} \|h\|^2_{\underline{\mathbb{H}}} \lesssim \epsilon^{-\frac{1}{k_\nu+1/2}}.$$

We now obtain a lower bound of the small ball probability of the *centered* Gaussian ball: b and J_{k_ν} are independent Gaussian processes and $\log\left(\frac{1}{\epsilon}\right) \lesssim \log\mathbb{P}\left(\|b\|_\infty \leq \epsilon\right)$. The main difficulty relies on $\phi_0(\epsilon) := \log\mathbb{P}\left(\|J_k(B)\|_\infty \leq \epsilon\right)$. Going back to (B.1), we see that $J_k(B)$ can be decomposed into two noninde-pendent Gaussian processes: $I_k(W)$ and a polynomial $\sum_{i=1}^{k+1} c_{i,k}(W)t^i$ which is a linear functional of W_1 and of the collection $\left(\int_0^1(1-t)^{k-j}W_t dt\right)_{1\leq j\leq k}$. Therefore

$$\log\left(\frac{1}{\epsilon}\right) \lesssim \log\mathbb{P}\left(\|J_k(B) - I_k(W)\|_\infty \leq \epsilon\right).$$

Now, applying Theorems 3.4 and 3.7 of [LS01] yields

$$\log\mathbb{P}\left(\|J_{k_\nu}(B)\|_\infty \leq \epsilon\right) \sim \log\mathbb{P}\left(\|I_{k_\nu}(W)\|_\infty \leq \epsilon\right) \geq -\epsilon^{-\frac{1}{k_\nu+1/2}},$$

which is of the same order as the approximation term. We end the proof gath-ering our lower bound on shifted Gaussian ball and the terms above. $\qquad\square$

Appendix C: Equivalents on Modified Bessel functions

Lemma C.1. *For $n\in\mathbb{Z}$ and $a > 0$, with $A_n(a) := \int_0^{2\pi} e^{a\cos(u)}\cos(nu)du$,*

$$\forall a\in[0,\sqrt{n}] \qquad A_n(a) \sim \frac{2\pi}{n!}\left(\frac{a}{2}\right)^n\left(1+\mathcal{O}\left(\frac{a}{n}\right)\right).$$

Proof. This equivalent is related to the modified Bessel functions (see *e.g.* [AS64] and [LL10]). More precisely, $I_m(a)$ is defined as

$$\forall m\in\mathbb{N}, \forall a > 0 \qquad I_m(a) := \sum_{k\geq 0}\frac{1}{k!(k+m)!}\left(\frac{a}{2}\right)^{2k+m},$$

and we have (see for instance [AS64]) $I_0(a) + 2\sum_{m=1}^{+\infty} I_m(a)\cos(mu) = e^{a\cos u}$. Hence, we easily deduce that $A_n(a) = 2\pi I_n(a)$. For small a, standard results on modified Bessel functions (Equation (9.7.7) of [AS64], p. 378.) yield

$$\forall a\in[0,\sqrt{n}] \qquad I_n(a) \sim \frac{1}{n!}\left(\frac{a}{2}\right)^n\left(1+\mathcal{O}\left(\frac{a}{n}\right)\right). \qquad\qquad \text{(C.1)}$$

$\square$

This integral is strongly related to the density of continuous time random walk: if $B_n(a) = \frac{e^{-a}A_n(a)}{2\pi}$, one has $B_n(0) = 0, \forall n \neq 0$ and $B_0(0) = 1$ and at last

$$B'_n(a) = \frac{B_n(a-1) + B_n(a+1)}{2} - B_n(a).$$

We can recognize here the forward Kolmogorov equation and $B_n(a)$ is the probability that a C.T.R.W. is located at $n \in \mathbb{Z}$ at time a. We then deduce equivalents of $B_n(a)$: from the Brownian approximation of the C.T.R.W. , we have

$$B_n(a) \sim \frac{1}{\sqrt{2\pi a}} e^{-n^2/(2a)}, \forall a \gg n^2. \tag{C.2}$$

Moreover, from [AS64], we know that

$$I_n(a) \sim \frac{e^a}{\sqrt{2\pi a}}, \quad \text{as soon as} \quad a \geq 2n, \tag{C.3}$$

and this equivalent is sharp when a is large (see Eq. (9.7.1) p. 377 of [AS64])

$$\forall a \geq 4n^2 \qquad I_n(a) \geq \frac{1}{2} \times \frac{e^a}{\sqrt{2\pi a}}.$$

(C.3) yields the heuristic equivalent proposed by (C.2): $B_n(a) = e^{-a}I_n(a) \sim \frac{1}{\sqrt{2\pi a}}$, and (C.1) gives a different information for small a. We stopped here our investigations since integral of (C.1) is much more larger than integral of (C.3).

For $a \in [\sqrt{n}, 2n]$, we do not have found any satisfactory equivalent on modified Bessel functions. Formula of [AS64] is still tractable but do not lead to "uniform enough" formula (we need to integrate this equivalent).

Acknowledgements

References

[AGP91] Y. Amit, U. Grenander, and M. Piccioni. Structural Image Restoration through Deformable Templates. *JASA*, 86(414):376–387, 1991.

[AKT10] S. Allassonnière, E. Kuhn, and A. Trouvé. Construction of Bayesian deformable models via a stochastic approximation algorithm: a convergence study. *Bernoulli*, 16(3):641–678, 2010.

[AS64] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. 1964.

[AW09] Jean-Marc Azaïs and Mario Wschebor. *Level sets and extrema of random processes and fields*. Wiley & Sons Inc., Hoboken, NJ, 2009.

[BCG12] J. Bigot, C. Christophe, and S. Gadat. Random action of compact Lie groups and minimax estimation of a mean pattern. *IEEE, Transactions on Information Theory*, 58(6):3509–3520, 2012.

[BCK07] R. Blundell, X. Chen, and D. Kristense. Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves. *Econometrica*, 75(6):1613–1669, 2007.

[BG10] J. Bigot and S. Gadat. A deconvolution approach to estimation of a common shape in a shifted curves model. *Ann. Statist.*, 38(4):2422–2464, 2010.

[BG12] J. Bigot and X. Gendre. Minimax properties of Fréchet means of discretely sampled curves. *Preprint*, 2012.

[BGKM13] J. Bigot, S. Gadat, T. Klein, and C. Marteau. Intensity estimation of non-homogeneous Poisson processes from shifted trajectories. *Electron. J. Stat.*, 7:881–931, 2013.

[BGL09] J. Bigot, S. Gadat, and J.M. Loubes. Statistical M-Estimation and Consistency in large deformable models for Image Warping. *Journal of Mathematical Imaging and Vision*, 34(3):270–290, 2009.

[Big13] J. Bigot. Fréchet means of curves for signal averaging and application to ECG data analysis. *Ann. of Applied Stat.*, 2013.

[BM05] C. Butucea and C. Matias. Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, 11(2):309–340, 2005.

[BSW99] A. Barron, M. J. Schervish, and L. Wasserman. The consistency of posterior distributions in nonparametric problems. *Annals of Statistics*, 27:536–561, 1999.

[Cas12] I. Castillo. A semi-parametric Bernstein-von Mises theorem for Gaussian process priors. *Prob. Th. and Rel. Fields*, 1:53–99, 2012.

[CD11] O. Collier and A Dalalyan. Curve registration by nonparametric goodness-of-fit testing. *Preprint*, 2011.

[Fer73] T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1:209–230, 1973.

[GGvdV00] S. Ghosal, J. K. Ghosh, and A. W. van der Vaart. Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531, 2000.

[GK95] T. Gasser and A Kneip. The impact of the MIT-BIH Arrhythmia Database. *Journal of the American Statistical Association*, 432(90):1179–1188, 1995.

[GL95] R. D. Gill and B. Y. Levit. Applications of the Van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, 1(1-2):59–79, 1995.

[GM01] C. Glasbey and K. Mardia. A penalized likelihood approach to image warping. *J. of the Royal Stat. Soc. (B)*, 63(3):465–514, 2001.

[GM07] U. Grenander and M. Miller. *Pattern Theory: From Representation*

to Inference. Oxford Univ. Press, Oxford, 2007.

[GR01] Moody G. and Mark R. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng in Med and Biol*, 3(20):45–50, 2001.

[GS02] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.

[GvdV01] S. Ghosal and A. W. van der Vaart. Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.*, 29(5):1233–1263, 2001.

[GW00] C. R. Genovese and L. Wasserman. Rates of convergence for the Gaussian mixture sieve. *Ann. Statist.*, 28(4):1105–1127, 2000.

[IH81] I. A. Ibragimov and R. Z. Has'minskii. *Statistical estimation*, volume 16 of *Applications of Mathematics*. Springer-Verlag, New York, 1981. Asymptotic theory.

[IL06] T. Inglot and T. Ledwina. Asymptotic optimality of new adaptive test in regression model. *Ann. Inst. H. Poincaré Probab. Statist.*, 42(5):579–590, 2006.

[KG88] A. Kneip and T. Gasser. Convergence and consistency results for self-modelling regression. *Ann. Statist.*, 16:82–112, 1988.

[KLL94] J. Kuelbs, W. V. Li, and W. Linde. The Gaussian measure of shifted balls. *Probab. Theory Related Fields*, 98(2):143–162, 1994.

[LCY00] L. Le Cam and G. Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer Verlag, New-York, 2000.

[LL10] G. F. Lawler and V. Limic. *Random walk: a modern introduction*, volume 123. Cambridge University Press, Cambridge, 2010.

[LS01] W. Li and Q.-M. Shao. Gaussian processes: inequalities, small ball probabilities and applications. In *Stochastic processes: theory and methods*, pages 533–597. North-Holland, Amsterdam, 2001.

[Mat02] C. Matias. Semiparametric deconvolution with unknown noise variance. *ESAIM Probab. Statist.*, 6:271–292 (electronic), 2002. New directions in time series analysis (Luminy, 2001).

[MM11] C. Maugis and B. Michel. A non asymptotic penalized criterion for Gaussian mixture model selection. *ESAIM Probab. Statist.*, 15:41–68 (electronic), 2011.

[MMW07] J. Mortazavi, R. Myers, and B. Wold. Genome-wide mapping of in vivo protein- dna interactions. *Science*, 316:1497–1502, 2007.

[PMRC10] W. Park, D. R. Madden, D. N. Rockmore, and G. S. Chirikjian. Deblurring of class-averaged images in single-particle electron microscopy. *Inverse Problems*, 26(3):035002, 29, 2010.

[RR12] V. Rivoirard and J. Rousseau. Posterior concentration rates for infinite dimensional exponential families. *Bayesian Analysis*, 7(2):311–334, 2012.

[SW01] Xiaotong Shen and Larry Wasserman. Rates of convergence of posterior distributions. *Ann. Statist.*, 29(3):687–714, 2001.

[vdVvZ08a] A. W. van der Vaart and J. H. van Zanten. Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.*, 36(3):1435–1463, 2008.

[vdVvZ08b] A. W. van der Vaart and J. H. van Zanten. Reproducing kernel Hilbert spaces of Gaussian priors. In *Pushing the limits of contemporary statistics: contributions in honor of J. K. Ghosh*, volume 3, pages 200–222. Inst. Math. Statist., Beachwood, OH, 2008.

[vdVW96] A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.

[WS95] W.H. Wong and X. Shen. Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.*, 23(2):339–362, 1995.

[Zha00] Linda H. Zhao. Bayesian aspects of some nonparametric problems. *Ann. Statist.*, 28(2):532–552, 2000.