



**HAL**  
open science

# Bayesian posterior consistency in the functional randomly shifted curves model

Dominique Bontemps, Sébastien Gadat

► **To cite this version:**

Dominique Bontemps, Sébastien Gadat. Bayesian posterior consistency in the functional randomly shifted curves model. 2013. hal-00768169v2

**HAL Id: hal-00768169**

**<https://ut3-toulouseinp.hal.science/hal-00768169v2>**

Submitted on 12 Mar 2013

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

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

# BAYESIAN POSTERIOR CONSISTENCY AND CONTRACTION RATES IN THE SHAPE INVARIANT MODEL

BY DOMINIQUE BONTEMPS\* AND SÉBASTIEN GADAT\*

*Institut Mathématiques de Toulouse, Université Paul Sabatier*

In this paper, we consider the so-called Shape Invariant Model which stands for the estimation of a function  $f^0$  submitted to a random translation of law  $g^0$  in a white noise model. We are interested in such a model when the law of the deformations is *unknown*. We aim to recover the law of the process  $\mathbb{P}_{f^0, g^0}$  as well as  $f^0$  and  $g^0$ .

In this perspective, we adopt a Bayesian point of view and find prior on  $f$  and  $g$  such that the posterior distribution concentrates around  $\mathbb{P}_{f^0, g^0}$  at a polynomial rate when  $n$  goes to  $+\infty$ . We obtain a logarithmic posterior contraction rate for the shape  $f^0$  and the distribution  $g^0$ . We also derive logarithmic lower bounds for the estimation of  $f^0$  and  $g^0$  in a frequentist paradigm.

**1. Introduction.** We are interested in this work in the so-called Shape Invariant Model (SIM). Such model aims to describe a statistical process which involves a deformation of a functional shape according to some randomized geometric variability. Such geometric deformation of a common unknown shape may be well-suited in various and numerous fields, like image processing (see for instance [AGP91] or [PMRC10]). It corresponds to a particular case of the general Grenander's theory of shapes (see [GM07] for a detailed introduction on this topic). This kind of model is also useful in medicine: the recent work of [Big11] deals with the differentiation between normal and arrhythmic cycles in electrocardiogram. It appears in genetics if one deals with some delayed activation curves of genes when drugs are administrated to patients, or in Chip-Seq estimation when translations in protein fixation yield randomly shifted counting processes (see for instance [MMW07] and [BGKM12]). It also occurs in econometric for the analysis of Engel curves [BCK07], in landmark registration [Big06]. . .

Such a model has received a large interest in the statistical community as pointed by the large amount of references on this subject. Some works

---

\*The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under references ANR-JCJC-SIMI1 DEMOS and ANR Bandhits.

*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

consider a semi-parametric approach for the estimation (*self-modeling regression framework* used by [KG88] and [BGV09]). In [Cas12], the author applies some Bayesian techniques to obtain also statistical results on SIM in a semi-parametric setting when the level of noise on observations asymptotically vanishes. Older approaches use parametric settings (see [GM01] and the discussion therein for an overview) and study the so-called Fréchet mean of pattern. Standard  $M$ -estimation or Bayesian methods are exploited in [BGL09] or [AAT07] and same authors develop in [AKT10] a nice stochastic algorithm to run estimation in such a model. Some recent works follow some testing strategies to obtain curve registration [CD11], [Col12]. At last, note that [BG10] obtains some minimax adaptive results for non-parametric estimations in the Shape Invariant Model when one knows the law of the randomized translations.

All these works are interested in the statistical process of deformation of the "mean common shape" and generally aim to recover this unknown functional object according to noisy i.i.d. observations. Moreover, the Shape Invariant Model is considered as a standard benchmark for statistical methods which aim to compute estimations in some more general deformable models. Of course, the SIM could be extended to some more general situations of geometrical deformations described through an action of a finite dimensional Lie Group (see [BCG12] for a precise non parametric description). We have decided to restrict our work here to the simplest case of the one dimensional Lie group of translation  $\mathbb{S}^1$  to warp the functional objects.

This work has been inspired by several discussions with Alain Trouvé about the work [AKT10] for the study of the Shape Invariant Model. We aim to extend their parametric Bayesian framework to the non-parametric setting and then study the behaviour of some posterior distributions. Hence, the motivation of the paper is mainly theoretical: we want to describe the asymptotic evolution of the posterior probability distributions when data are coming from the SIM. Of course, we need to build suitable prior which yield nice contraction rate for this posterior distribution. We have decided to consider the general case where both the functional shape and the probability distribution of the deformations are unknown. Indeed, it corresponds to the more realistic case. From the best of our knowledge, no sharp statistical results have been derived yet in this non-parametric situation.

Our work will describe the evolution of the posterior distribution when the number of observations grows to  $+\infty$  with a fixed noise level  $\sigma$ . It is an important difference with the study of the asymptotically vanishing noise situation ( $\sigma \rightarrow 0$ ). It is itself a special feature of the Shape Invariant Model: there is no obvious Le Cam equivalence of experiments (see [LCY00]) for the

SIM between the experiments when  $n \rightarrow +\infty$  and when  $\sigma \mapsto 0$ . It is illustrated by the very different minimax results obtained in [BG10] ( $n \rightarrow +\infty$ ) and in [BG12] ( $\sigma \mapsto 0$ ). We will use in the sequel quite standard Bayesian non parametric methods to obtain the frequentist consistency and some contraction rates of the Bayesian procedures. Such tools rely on some important contributions of [BSW99] and [GGvdV00] for the posterior behaviour in general situations, as well as Bayesian properties on mixture models stated in [GvdV01] and [GW00] or prior distribution on smooth densities through Gaussian processes given by [vdVvZ08a].

The paper is organised as follows. Section 2 presents a sharp description of the Shape Invariant Model (shortened as SIM in the sequel), as well as standard elements on Bayesian and Fourier analysis. It also provides some notations for mixture models. It ends with the statements of our main results. Section 3 presents the proof of the posterior contraction around the true law on functional curves, which is our first main result. Section 4 provides some general identifiability results and up to these identifiability conditions, shows the posterior contraction on the functional objects themselves. At last, this section also establishes a lower bound result of reconstruction in a frequentist paradigm. We end the paper with numerous challenging issues.

We gather in the appendix sections some technical points: the metric description of the Shape Invariant Model embedded in a special randomized curves space and the calibration of suitable priors for the SIM.

## 2. Model, notations and main results.

### 2.1. Statistical settings.

*Shape Invariant Model.* We recall here the random Shape Invariant Model. We assume  $f^0$  to be a function which belongs to a subset  $\mathcal{F}$  of smooth functions. We also consider a probability measure  $g^0$  which is an element of the set  $\mathfrak{M}([0, 1])$ . This last set stands for the set of probability measures on  $[0, 1]$ . We observe  $n$  realizations of noisy and randomly shifted complex valued curves  $Y_1, \dots, Y_n$  coming from the following white noise model

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

Here,  $f^0$  is the *mean* pattern of the curves  $Y_1, \dots, Y_n$  although the random shifts  $(\tau_j)_{j=1 \dots n}$  are sampled independently according to the probability measure  $g^0$ . Moreover,  $(W_j)_{j=1 \dots n}$  are independent complex standard Brownian motions on  $[0, 1]$  and model the presence of noise in the observations, the noise level is kept fixed in our study and is set to 1 for sake of simplicity.

In the sequel,  $f^{-\tau}$  will denote the pattern  $f$  shifted by  $\tau$ , that is to say the function  $x \mapsto f(x - \tau)$ . Complex valued curves are considered here for the simplicity of notations. However all our results can be adapted to the simpler case where all curves  $Y_j$ 's are real valued. 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 denoted by  $\mathcal{N}_{\mathbb{C}}(0, 1)$ ; a standard complex Gaussian random variable have independent real and imaginary parts with a real centered Gaussian distribution of variance  $1/2$ .

This work will address the question of the behaviour of some posterior distributions on  $\mathcal{F} \otimes \mathfrak{M}([0, 1])$  given some functional  $n$ -sample  $(Y_1, \dots, Y_n)$ . Since our work will be mainly asymptotic with  $n \rightarrow +\infty$ , we intensively use some standard notation such as " $\lesssim$ " which refers to an inequality up to a multiplicative absolute constant. In the meantime,  $a \sim b$  stands for  $a/b \rightarrow 1$ .

*Bayesian framework.* Since most of statistical works on the SIM are frequentists, we have decided to briefly recall here the Bayesian formalism following the presentation of [GGvdV00]. Familiar readers can thus omit this paragraph.

Functional objects  $f^0$  and  $g^0$  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 denoted  $\mathbb{P}_{f,g}$  and possesses a density  $p_{f,g}$  with respect to the Wiener measure on the sample space. Since  $f^0$  and  $g^0$  are unknown,  $\mathbb{P}_{f^0,g^0}$  is also unavailable but belongs to a set  $\mathcal{P}$  of probability measure over the sample space. This set  $\mathcal{P}$  is the set of all possible measures described by (2.1) when  $(f, g)$  varies into  $\mathcal{F} \otimes \mathfrak{M}([0, 1])$ .

Given some prior distribution  $\Pi_n$  on  $\mathcal{P}$  (generally defined through a prior on  $\mathcal{F} \otimes \mathfrak{M}([0, 1])$ ), Bayesian procedures are generally built using the posterior distribution defined by

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

which is a random measure on  $\mathcal{P}$  that depends on the observations  $Y_1, \dots, Y_n$ . For instance, Bayesian estimators can be obtained using the mode, the mean or the median of the posterior distribution. This is exactly the approach adopted by [AKT10] which is mainly dedicated to compute such a posterior mean in a parametric setting with a stochastic EM algorithm.

The posterior distribution is then said *consistent* if it concentrates to arbitrarily small neighbourhoods of  $\mathbb{P}_{f^0,g^0}$  in  $\mathcal{P}$  with a probability tending to 1 when  $n$  grows to  $+\infty$ . One *frequentist* property of such a posterior distribution describes the contraction rate of such neighbourhoods meanwhile

still capturing most of posterior mass. According to equation (2.1), we thus tackle such a Bayesian consistency and compute such convergence rates in the frequentist paradigm. Of course, these properties will highly depend on the metric structure of the sets  $\mathcal{P}$  and  $\mathcal{F}$ .

*Functional setting and Fourier analysis.* Without loss of generality, the function  $f^0$  is assumed to be periodic with period 1 and to belong to a subset  $\mathcal{F}$  of  $L^2_{\mathbb{C}}([0, 1])$ , the space of squared integrable functions on  $[0, 1]$  endowed with the euclidean norm  $\|h\| := \int_0^1 |h(s)|^2 ds$ . Moreover, each element  $h \in L^2_{\mathbb{C}}([0, 1])$  may naturally be extended to a periodic function on  $\mathbb{R}$  of period 1. Since we will intensively use some Fourier analysis in the sequel, let us first recall some notations:  $i$  will stand for the complex number such that  $i^2 = -1$ . The Fourier coefficients of  $h$  are denoted

$$(2.2) \quad \theta_{\ell}(h) := \int_0^1 e^{-i2\pi\ell t} h(t) dt.$$

All along the paper, we will often use the parametrisation of any element of  $h \in L^2_{\mathbb{C}}([0, 1])$  through its Fourier expansion and will simply use the notation  $(\theta_{\ell})_{\ell \in \mathbb{Z}}$  instead of  $(\theta_{\ell}(h))_{\ell \in \mathbb{Z}}$ .

Our work is dedicated to the analysis of SIM when  $\mathcal{F}$  models smooth functions of  $[0, 1]$ . Hence, natural subspaces of  $L^2_{\mathbb{C}}([0, 1])$  are Sobolev spaces  $\mathcal{H}_s$  with a smoothness parameter  $s$ :

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

A useful set of functions for the identifiability part will also be the following restriction of  $\mathcal{H}_s$ :

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

In the sequel, we aim to find prior on  $\mathcal{P}$  that reaches good frequentist properties, and if possible adaptive with the smoothness parameter  $s$  since this parameter is generally unknown. We will consider only some regular cases when  $s \geq 1$ , the quantity  $\sum_{\ell} \ell^2 |\theta_{\ell}|^2$  is thus bounded and we denote the Sobolev norm

$$\|\theta\|_{\mathcal{H}_1} := \sqrt{\sum_{\ell \in \mathbb{Z}} \ell^2 |\theta_{\ell}|^2}.$$

It will also be useful to consider in some cases Fourier "thresholded" elements of  $\mathcal{H}_s$ . Hence, we set for any integer  $\ell$  (which is the frequency threshold)

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

*Mixture model.* According to equation (2.1), we can write in the Fourier domain that

$$(2.3) \quad \forall \ell \in \mathbb{Z} \quad \forall j \in \{1 \dots n\} \quad \theta_\ell(Y_j) = \theta_\ell^0 e^{-i2\pi j \tau_j} + \xi_{\ell,j},$$

where  $\theta^0 := (\theta_\ell^0)_{\ell \in \mathbb{Z}}$  denotes the true unknown Fourier coefficients of  $f^0$ . Owing to the white noise model, the variables  $(\xi_{\ell,j})_{\ell,j}$  are independent standard (complex) Gaussian random variables:  $\xi_{\ell,j} \sim_{i.i.d.} \mathcal{N}_{\mathbb{C}}(0, 1), \forall \ell, j$ .

For sake of simplicity,  $\gamma$  will refer to  $\gamma(z) := \pi^{-1} e^{-|z|^2}, \forall z \in \mathbb{C}$ , the density of the standard complex Gaussian centered distribution  $\mathcal{N}_{\mathbb{C}}(0, 1)$ , and  $\gamma_\mu(\cdot) := \gamma(\cdot - \mu)$  is the density of the standard complex Gaussian with mean  $\mu$ . We keep also the same notation for  $p$  dimensional complex Gaussian densities  $\gamma(z) := \pi^{-p} e^{-\|z\|^2}, \forall z \in \mathbb{C}^p$ , where  $\|z\|$  is the euclidean  $p$  dimensional norm of the complex vector  $z$ .

For any frequency  $\ell$ , equation (2.1) implies that  $\theta_\ell(Y)$  follows a mixture of complex Gaussian standard variables with mean  $\theta_\ell^0 e^{-i2\pi \ell \varphi}, \varphi \in [0, 1]$ :

$$\theta_\ell(Y) \sim \int_0^1 \gamma_{\theta_\ell^0 e^{-i2\pi \ell \varphi}}(\cdot) dg(\varphi).$$

In the sequel, for any phase  $\varphi \in [0, 1]$  sampled according to any distribution  $g$ , and for any  $\theta \in \ell^2(\mathbb{Z})$ ,  $\theta \bullet \varphi$  will denote the element of  $\ell^2(\mathbb{Z})$  given by

$$\forall \ell \in \mathbb{Z} \quad (\theta \bullet \varphi)_\ell := \theta_\ell e^{-i2\pi \ell \varphi}.$$

When  $\theta$  is a complex vector, for instance  $\theta = (\theta_{-\ell}, \dots, \theta_\ell)$ , we keep the same notation  $\theta \bullet \varphi := (\theta_{-\ell} e^{i2\pi \ell \varphi}, \dots, \theta_0, \theta_1 e^{-i2\pi \varphi}, \dots, \theta_\ell e^{-i2\pi \ell \varphi})$  to refer to the  $2\ell + 1$  dimensional vector. It corresponds to a rotation of each coefficient  $\theta_\ell$  around the origin with an angle  $2\pi \ell \varphi$ . According to this notation, the law of the infinite series (of Fourier coefficients of  $Y$ ) can thus be rewritten as

$$\theta(Y) \sim \int_0^1 \gamma_{\theta \bullet \varphi}(\cdot) dg(\varphi).$$

One should remark the important fact that from one frequency to another, the rotations used to build  $\theta(Y)$  are not independent, which traduces the fact that the coefficients  $(\theta_\ell(Y))_\ell$  are highly correlated.

*2.2. Notations on Mixture models.* Our study will intensively use some classical tools of mixture models, see for instance the papers of [GvdV01] or [GW00]. We thus choose to keep some notations already used in such works.

For any vector  $\theta \in \ell^2_{\mathbb{C}}(\mathbb{Z})$  corresponds a function  $f \in L^2([0, 1])$  according to equation (2.2) and for any measure  $g \in \mathfrak{M}([0, 1])$ ,  $\mathbb{P}_{\theta,g}$  will refer to the

law of the vector of  $\ell^2(\mathbb{Z})$  described by the location mixture of Gaussian variables:

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

This mixture model is of infinite dimension since  $\theta$  belongs to  $\ell^2(\mathbb{Z})$ . Following an obvious notation shortcut,  $\mathbb{P}_{f,g}$  will be its equivalent for the functional law on curves derived from  $\mathbb{P}_{\theta,g}$ . When  $\theta$  is of finite length  $k$ ,  $p_{\theta,g}$  will be the density with respect to the Lebesgue measure on  $\mathbb{C}^k$  of the law  $\mathbb{P}_{\theta,g}$ :

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

We also use standard objects such as the Hellinger distance  $d_H$  between probability measures and the Total Variation distance  $d_{TV}$ , as well as covering numbers of metric spaces such as  $D(\epsilon, \mathcal{P}, d)$ . These objects are precisely described in Appendix A.

*Bayesian frequentist consistency rate.* In our setting,  $d$  is chosen according to one of the metric introduced above ( $d_H$  or  $d_{TV}$ ) on the set

$$\mathcal{P} := \{\mathbb{P}_{f,g} | (f, g) \in \mathcal{H}_s \otimes \mathfrak{M}([0, 1])\}.$$

We can now remind Theorem 2.1 of [GGvdV00] which will be useful for our purpose.

**THEOREM 2.1** (Posterior consistency and convergence rate, [GGvdV00]). *Assume that a sequence  $(\epsilon_n)_n$  with  $\epsilon_n \rightarrow 0$  and  $n\epsilon_n^2 \rightarrow +\infty$ , a constant  $C > 0$ , and a sequence of sets  $\mathcal{P}_n \subset \mathcal{P}$  satisfy*

$$(2.4) \quad \log D(\epsilon_n, \mathcal{P}_n, d) \leq n\epsilon_n^2$$

$$(2.5) \quad \Pi_n(\mathcal{P} \setminus \mathcal{P}_n) \leq e^{-n\epsilon_n^2(C+4)}$$

$$(2.6) \quad \Pi_n(\mathbb{P}_{f,g} \in \mathcal{P} | d_{KL}(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f,g}) \leq \epsilon_n^2, V(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f,g}) \leq \epsilon_n^2) \geq e^{-n\epsilon_n^2 C}.$$

*Then there exists a sufficiently large  $M$  such that*

$$\Pi_n(\mathbb{P}_{f,g} : d(\mathbb{P}_{f^0,g^0}, \mathbb{P}_{f,g}) \geq M\epsilon_n | Y_1, \dots, Y_n) \longrightarrow 0$$

*in  $\mathbb{P}_{f^0,g^0}$  probability as  $n \rightarrow +\infty$ .*



The posterior concentration rate obtained in the above result is  $\epsilon_n$ . The growing set  $\mathcal{P}_n$  is referred to as a Sieve over  $\mathcal{P}$ . Generally, this rate  $\epsilon_n$  can be compared to the classical frequentist benchmark: for instance [GGvdV00] obtained for the Log Spline model a contraction rate  $\epsilon_n = n^{-s/(2s+1)}$  when the unknown underlying density belongs to an Hölder class  $\mathcal{C}^s([0, 1])$ , and this rate is known to be the optimal one (in the sense that it is the minimax one) in the frequentist paradigm over Hölder densities of regularity  $s$  (see [IH81]). Similarly, the recent work of [RR12] considers the situation of density estimation for infinite dimensional exponential families and reaches also contraction rates close or equal to the known optimal frequentist one.

*2.3. Bayesian prior and posterior concentration in the randomly shifted curves model.* We detail here the Bayesian prior  $\Pi_n$  on  $\mathcal{P}$  used to obtain a polynomial concentration rate. Note that such prior will be in our work independent on the unknown smoothness parameter  $s$ . As pointed in the paragraph above, it is sufficient to define some prior on the space  $\mathcal{H}_s \otimes \mathfrak{M}([0, 1])$  since equation (2.1) will then transport this prior to a law  $\Pi_n$  on  $\mathcal{P}$ . The two parameters  $f$  and  $g$  are picked independently at random following the next prior distributions.

*Prior on  $f$ .* The prior on  $f$  is slightly adapted from [RR12]. It is defined on  $\mathcal{H}_s$  through

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

Given any integer  $\ell$ , the idea is to decide to randomly switch on with probability  $\lambda(\ell)$  all the Fourier frequencies from  $-\ell$  to  $+\ell$ . Then,  $\pi_\ell$  is a distribution defined on  $\ell^2(\mathbb{Z})$  such that  $\pi_\ell := \otimes_{k \in \mathbb{Z}} \pi_\ell^k$  and

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

The randomisation of selected frequencies is done using  $\lambda$ , a probability distribution on  $\mathbb{N}^*$  which satisfies for  $\rho \in (1, 2)$ :

$$\exists (c_1, c_2) \in \mathbb{R}_+ \quad \forall \ell \in \mathbb{N}^* \quad e^{-c_1 \ell^2 \log^\rho \ell} \lesssim \lambda(\ell) \lesssim e^{-c_2 \ell^2 \log^\rho \ell}.$$

The prior  $\pi$  depends on the variance of the Gaussian laws  $\xi_n$  used to sample the Fourier coefficients. In the sequel, we use a variance that depends on  $n$  according to

$$(2.7) \quad \xi_n^2 := n^{-\mu_s} (\log n)^{-\zeta},$$

where  $\mu_s$  and  $\zeta$  are parameters that may depend on  $s$  (non adaptive prior) or not (adaptive prior).

*Prior on general probability distribution  $g$ .* As our model does not seem so far from a mixture Gaussian model, a natural prior on  $g$  is built according to a Dirichlet process following the ideas of [GvdV01]. Given any finite base measure  $\alpha$  that has a positive continuous density on  $[0, 1]$  w.r.t. the Lebesgue measure, the Dirichlet process  $D_\alpha$  generates a random probability measure  $g$  on  $[0, 1]$ . For any finite partition  $(A_1, \dots, A_k)$  of  $[0, 1]$ , the probability vector  $(g(A_1), \dots, g(A_k))$  on the  $k$ -dimensional simplex has a Dirichlet distribution  $Dir(\alpha(A_1), \dots, \alpha(A_k))$ . Such process may be built according to the Stick-Breaking construction (see for instance [Fer73]). In the sequel, we refer to  $q_{D,\alpha}$  for this prior based on the Dirichlet process  $D_\alpha$ .

*Prior on smooth probability distributions  $g$ .* In the sequel, we will also consider the special case of *smooth* densities to push our results further. This set is characterised by a regularity parameter  $\nu$  and a radius  $A$ :

$$\mathfrak{M}_\nu([0, 1])(A) := \left\{ g \in \mathfrak{M}([0, 1]) \mid \sum_{k \in \mathbb{Z}} k^{2\nu} |\theta_k(g)|^2 < A^2 \right\},$$

where  $\mathfrak{M}([0, 1])$  is the set of probability on  $[0, 1]$ . At last, we will also need the set

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

The prior on  $\mathfrak{M}_\nu([0, 1])(A)$  will be rather different from the one defined on  $\mathfrak{M}([0, 1])$  above. Such smoothness is not compatible with Dirichlet priors and even kernel convolution with Dirichlet process seems problematic in our situation. Thus, we have chosen to use some prior based on gaussian process. *In this case, we assume for smooth mixture models that we know the smoothness parameter  $\nu$  of  $g^0$ , as well as the radius  $A$  of the Sobolev balls where  $g^0$  is living.*

Given  $\nu \geq 1/2$  and  $A > 0$ , we define the integer  $k_\nu := \lfloor \nu - 1/2 \rfloor$  to be the largest integer smaller than  $\nu - 1/2$ . We follow the strategy of section 4 in [vdVvZ08a] and the important point is that we have to take into account the 1-periodicity of the density  $g$ , as well as its regularity. In this view, we denote  $B$  a Brownian bridge between 0 and 1. The Brownian bridge can be obtained from a Brownian motion trajectory  $W$  using  $B_t = W_t - tW_1$ . Then, For any continuous function  $f$  on  $[0, 1]$ , we define the linear map

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

and all its composition are  $J_k = J_{k-1} \circ J$ . Moreover, in order to adapt our prior to the several derivatives of  $g$  at points 0 and 1, we use the family of

maps  $(\psi_j)_{j=1\dots k_\nu}$  defined as

$$\forall t \in [0, 1] \quad \psi_k(t) := \sin(2\pi kt) + \cos(2\pi kt).$$

Our prior is now built as follows, we first sample a real Brownian bridge  $(B_\tau)_{\tau \in [0,1]}$  and  $Z_1, \dots, Z_{k_\nu}$  independent real standard normal random variables. This enables to generate the Gaussian process

$$(2.8) \quad \forall \tau \in [0, 1] \quad w_\tau := J_{k_\nu}(B)(\tau) + \sum_{i=1}^{k_\nu} Z_i \psi_i(\tau).$$

Given  $(w_\tau, \tau \in [0; 1])$  generated by (2.8), we build  $p_w$  through

$$(2.9) \quad \forall \tau \in [0; 1] \quad p_w(\tau) := \frac{e^{w_\tau}}{\int_0^1 e^{w_\tau} d\tau}.$$

Hence, a prior on Gaussian process yields a prior on densities on  $[0; 1]$  and  $p_w$  inherits of the smoothness  $k_\nu$  of the Gaussian process  $\tau \mapsto w_\tau$ . According to our construction, we now consider the restriction of the prior defined above to the Sobolev balls of radius  $2A$ . This finally defines a prior distribution  $q_{\nu,A}$  on  $\mathfrak{M}_\nu([0, 1])(2A)$ .

*2.4. Main results.* Our two priors on the elements  $g$  enable to define two priors on the model. The first one is  $\pi \otimes q_{D,\alpha}$  for general distributions  $g$  although the second one is  $\pi \otimes q_{\nu,A}$  for smooth mixture distributions. For sake of convenience, we will always use the notation  $\Pi_n$  to refer to one of these two priors and just precise which prior is used in the statements of the theorems. We now give the two results on the randomly SIM.

*2.4.1. Posterior contraction for  $\Pi_n := \pi \otimes q_{D,\alpha}$ .*

**THEOREM 2.2.** *Define the prior distribution  $\Pi_n := \pi \otimes q_{D,\alpha}$  and assume that  $f^0 \in \mathcal{H}_s$  with  $s \geq 1$ , then the values  $\mu_s = 2/(2s + 2)$  and  $\zeta = 0$  in the definition of  $\xi_n$  yield a non adaptive prior such that*

$$\Pi_n \left\{ \mathbb{P}_{f,g} \text{ s.t. } d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \leq M\epsilon_n | Y_1, \dots, Y_n \right\} = 1 + \mathcal{O}_{\mathbb{P}_{f^0,g^0}}(1)$$

when  $n \rightarrow +\infty$ , for a sufficiently large constant  $M$ . Moreover, the contraction rate  $\epsilon_n$  is given by

$$\epsilon_n = n^{-s/(2s+2)} \log n.$$

The values  $\mu = 1/4$  and  $\zeta = 3/2$  yield the contraction rate

$$\Pi_n \left\{ \mathbb{P}_{f,g} \text{ s.t. } d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0,g^0}) \leq M\epsilon_n | Y_1, \dots, Y_n \right\} = 1 + \mathcal{O}_{\mathbb{P}_{f^0,g^0}}(1)$$

for a sufficiently large constant  $M$ , when  $n \rightarrow +\infty$  with

$$\epsilon_n = \begin{cases} n^{-s/(2s+2)} \log n & \text{if } s \in [1, 3] \\ n^{-3/8} \log n & \text{if } s \geq 3. \end{cases}$$

Let us briefly comment this result. It first describes the posterior concentration around some neighbourhood of the true law  $\mathbb{P}_{f^0, g^0}$  within a polynomial rate. Our prior is adaptive with the regularity  $s$  as soon as  $s \in [1, 3]$  setting  $\xi_n^2 = n^{-1/4}(\log n)^{-3/2}$ . For this range of  $s$ , the convergence rate is  $n^{-s/(2s+2)}$  up to a logarithmic term. To the best of our knowledge, the minimax frequentist rate is unknown for the problem on recovering  $\mathbb{P}_{f^0, g^0}$  when both  $f^0$  and  $g^0$  are unknown. An interpretation of such polynomial rate is rather difficult to provide. It may be interpreted as  $-s/(2s + d)$  where  $d$  is the number of dimension to estimate in the model ( $f^0$  and  $g^0$ ). When  $s$  becomes larger than 3, the rate of Theorem (2.2) is "blocked" to  $3/8$  (which corresponds to  $s/(2s + 2)$  when  $s = 3$ ) and does not match with  $s/(2s + 2)$ . This difficulty is mainly due to the important condition  $w_\epsilon^2 \lesssim l_\epsilon$  in Theorem 3.1.

At last, the non adaptive prior based on  $\xi_n^2 = n^{-2/(2s+2)}$  recovers the good rate  $-s/(2s + 2)$  for all  $s$  larger than 1.

2.4.2. *Posterior contraction for  $\Pi_n := \pi \otimes q_{\nu, A}$ .* The second result concerns the special case of smooth densities  $g$  and provide a result for the non adaptive prior based on the knowledge of  $\nu$  and  $A$ .

THEOREM 2.3. *For the prior  $\Pi_n := \pi \otimes q_{\nu, A}$  using  $\xi_n^2 = n^{-1/4}(\log n)^{-3/2}$  in the definition of  $\pi$ , if  $f^0 \in \mathcal{H}_s$  with  $s \geq 1$  and  $g^0 \in \mathfrak{M}_\nu([0, 1])(A)$ , then there exists a sufficiently large  $M$  such that*

$$\Pi_n \left\{ \mathbb{P}_{f, g} \text{ s.t. } d_H(\mathbb{P}_{f, g}, \mathbb{P}_{f^0, g^0}) \leq M\epsilon_n | Y_1, \dots, Y_n \right\} = 1 + \mathcal{O}_{\mathbb{P}_{f^0, g^0}}(1)$$

when  $n \rightarrow +\infty$ , where for an explicit  $\kappa > 0$ :

$$\epsilon_n = n^{-\left[ \frac{\nu}{2\nu+1} \wedge \frac{s}{2s+2} \wedge \frac{3}{8} \right]} \log(n)^\kappa.$$

We still obtain a similar result of polynomial order but note that the smoothness parameter  $\nu$  of  $g^0$  influences now the contraction rate.

2.4.3. *Posterior contraction on functional objects.* The former results establish some results on the law  $\mathbb{P}_{f, g} \in \mathcal{P}$ . It is also possible to derive a second result on the objects  $f \in \mathcal{H}_s$  themselves, provided that we consider smooth classes for mixture distributions and provides a somewhat quite weak result on the posterior convergence towards the true objects  $f^0$  and  $g^0$ .

THEOREM 2.4. For the prior  $\Pi_n := \pi \otimes q_{\nu,A}$  using  $\xi_n^2 = n^{-1/4}(\log n)^{-3/2}$  in the definition of  $\pi$ , the two following results hold.

i) Assume that  $f^0 \in \mathcal{F}_s$  with  $s \geq 1$  and  $g^0 \in \mathfrak{M}_\nu([0,1])(A)$  with  $\nu > 1$ , then there exists a sufficiently large  $M$  such that

$$\Pi_n \{g \text{ s.t. } \|g - g^0\| \leq M\mu_n |Y_1, \dots, Y_n\} = 1 + \mathcal{O}_{\mathbb{P}_{f^0, g^0}}(1)$$

with the contraction rate  $\mu_n = (\log n)^{-\nu}$ .

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

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

then we also have

$$\Pi_n \{f \text{ s.t. } \|f - f^0\| \leq M\tilde{\mu}_n |Y_1, \dots, Y_n\} = 1 + \mathcal{O}_{\mathbb{P}_{f^0, g^0}}(1)$$

when  $n \rightarrow +\infty$ . Moreover, the contraction rate  $\tilde{\mu}_n$  is given by

$$\tilde{\mu}_n = (\log n)^{-\frac{4s\nu}{2s+2\beta+1}}.$$

2.4.4. *Identifiability and lower bounds.* In paragraph 4.3, we will stress the fact that it is indeed impossible to obtain frequentist convergence rates better than some power of  $\log n$  even if our lower bound does not match exactly with the upper bound obtained in the previous result.

THEOREM 2.5. i) The Shape Invariant Model is identifiable as soon as  $(f^0, g^0) \in \mathcal{F}_s \times \mathfrak{M}([0,1])^*$ : the canonical application

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

ii) Assume that  $(f^0, g^0) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0,1])(A)$ , then there exists a sufficiently small  $c$  such that the minimax rate of estimation over  $\mathcal{F}_s \times \mathfrak{M}_\nu([0,1])(A)$  satisfies

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

and

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

This result is far from being contradictory with the polynomial rate obtained in Theorem 2.2. One can make at least three remarks:

- The first result provides a contraction rate on the probability distribution in  $\mathcal{P}$  and not on the functional space  $\mathcal{F}_s$ .
- The link between  $(f^0, g^0)$  and  $\mathbb{P}_{f^0, g^0}$  relies on the identifiability of the model, and the lower bound is derived from a net of functions  $(f_i, g_i)_i$ , which are really hard to identify according to the application  $\mathcal{I} : (f, g) \mapsto \mathbb{P}_{f, g}$ . On this net of functions, the injection is very “flat” and the two by two differences of  $\mathcal{I}(f^i, g^i)$  are as small as possible and thus the pairs of functions  $(f^i, g^i)$  become very hard to distinguish.
- In fact, [BG10] have shown that in the SIM, when  $n \rightarrow +\infty$ , 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  $\mathcal{F}_s \times \mathfrak{M}_\nu([0, 1])(A)$  also occurs owing to such reason. One may argue that such an artefact could be avoided if one chooses a different distance on  $\mathcal{F}_s$ , which may be better suited to our framework, such as

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

We do not have purchased further investigations with this distance on  $\mathcal{F}_s$  but it would certainly be a nice progress to obtain posterior contraction using such a distance. We expect a polynomial rate, but it is clearly an open (and probably hard) task.

**3. Proof of Theorem 2.2 and Theorem 2.3.** We aim to check conditions (2.5) and (2.6) and then apply Theorem 2.1. In this view, we first define in section 3.1 a sieve  $\mathcal{P}_{\ell_\epsilon, w_\epsilon}$ , and our goal is to find some optimal calibration of  $\epsilon$ ,  $\ell_\epsilon$  and  $w_\epsilon$  with respect to  $n$ . We thus need to find a lower bound of the prior mass around some Kullback-Leibler neighbourhood of  $\mathbb{P}_{f^0, g^0} \in \mathcal{P}$ . These sets are defined as

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

This will be done indeed considering Hellinger neighbourhoods instead of Kullback-Leibler ones. A link between these two kinds of neighbourhood is given in section 3.2. In section 3.3, we work with the Hellinger neighbourhoods to exhibit some admissible sizes for  $\epsilon_n$ ,  $\ell_n$  and  $w_n$ . At last, we prove Theorem 2.2 in section 3.4.

In all this section, we delay most technical proofs to the Appendix.

3.1. *Entropy estimates.* We first establish some useful results on the complexity of our model  $\mathbb{P}_{f, g}$  when  $f \in \mathcal{H}_s$  and  $g \in \mathfrak{M}([0, 1])$  in various situations ( $f$  known, unknown, parametric or not).

3.1.1. *Case of known  $f$ .* We first give some useful results when  $f$  is known and belongs to a finite dimensional vector space (the number of active Fourier coefficients is restricted to  $[-\ell, \ell]$  for a given  $\ell$ ). Then  $\ell$  will be allowed to grow with  $n$  and depend on a parameter  $\epsilon$  introduced below. Hence,  $f$  is described by the parameter  $\theta = (\theta_{-\ell}, \dots, \theta_0, \dots, \theta_\ell)$ , and we define the set of all possible Gaussian measures

$$\mathcal{A}_\theta := \{\gamma_{\theta \bullet \varphi}, \varphi \in [0, 1]\}.$$

Following the arguments of [GW00], it is possible to establish the following preliminary result.

PROPOSITION 3.1. *For any sequence  $\theta \in \mathbb{C}^{2\ell+1}$ , one has*

$$N_{\square}(\epsilon, \mathcal{A}_\theta, d_H) \leq \frac{4\pi\sqrt{2(2\ell+1)}\|\theta\|_{\mathcal{H}_1}(1+o(1))}{\epsilon},$$

where  $o(1)$  goes to zero independently on  $\ell$  and  $\theta$  as  $\epsilon \rightarrow 0$ , and

$$\log N(\epsilon, \mathcal{A}_\theta, d_H) \lesssim \log \ell + \log \|\theta\|_{\mathcal{H}_1} + \log \frac{1}{\epsilon}.$$

Assume now that  $g$  possesses a finite number of  $k$  points in its support, one can deduce from the proposition above a simple corollary that exploits the complexity of the simplex of dimension  $k-1$  (see for instance the proof of Lemma 2 in [GW00]).

PROPOSITION 3.2. *Assume that  $f$  is parametric and known ( $\theta \in \mathbb{C}^{2\ell+1}$ ) and define*

$$\mathcal{M}_\theta^k := \left\{ \sum_{i=1}^k g(\varphi_i) \gamma_{\theta \bullet \varphi_i} : \varphi_i \in [0, 1], g(\varphi_i) \geq 0, \forall i \in \llbracket 1, k \rrbracket \text{ and } \sum_{i=1}^k g(\varphi_i) = 1 \right\}$$

for a number of components  $k$  that may depend on  $\epsilon$  (as  $\ell$  does). Then

$$H_{\square}(\epsilon, \mathcal{M}_\theta^k, d_H) \lesssim k \left( \log \ell + \log \|\theta\|_{\mathcal{H}_1} + \log \frac{1}{\epsilon} \right).$$

We then naturally provide a description of the situation when  $f$  is known and parametrized by an infinite sequence  $\theta \in \ell^2(\mathbb{Z})$ . According to the previous computations, and using a truncation argument at frequency  $\ell_\epsilon = \epsilon^{-1/s}$  in the Sobolev space  $\mathcal{H}_s$ , one can show the following result.

COROLLARY 1. Assume  $f \in \mathcal{H}_s$  known for  $s \geq 1$  ( $\theta := \theta(f)$  such that  $\sum_{j \in \mathbb{Z}} |\theta_j|^2 |j|^{2s} < +\infty$ ), using the same set  $\mathcal{A}_\theta$  as in Proposition 3.1 with  $\ell_\epsilon = \epsilon^{-1/s}$ , then

$$H_{\square}(\epsilon, \mathcal{A}_\theta, d_H) \lesssim \frac{s+1}{s} \log \frac{1}{\epsilon} + \log \|\theta\|_{\mathcal{H}_1}.$$

Similarly, one also has

$$H_{\square}(\epsilon, \mathcal{M}_\theta^k, d_H) \lesssim k \left( \frac{s+1}{s} \log \frac{1}{\epsilon} + \log \|\theta\|_{\mathcal{H}_1} \right).$$

The next step is to consider a continuous mixture for  $g$ , which is the more natural case. For  $f$  known, let

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

Once again, we will only consider functions  $f$  with null Fourier coefficients of order higher than  $\ell_\epsilon$ . For sake of simplicity, we will omit the dependence on  $\epsilon$  with the notation  $\ell$ .

It would be quite tempting to use the results of [GvdV01] to bound the bracketing entropy of  $\mathcal{P}_f$ , but indeed as pointed by [MM11] applying directly the bounds obtained in Lemma 3.1 and Lemma 3.2 of [GvdV01] to our setting yields a too weak result: the size of the upper bound on  $H_{\square}(\epsilon, \mathcal{P}_f, d_H)$  will have a too strong dependency on  $\ell$ . By the way, we have to carefully adapt the approach of [GvdV01] to obtain a sufficiently sharp upper bound of the entropy of  $\mathcal{P}_f$ . Such bound is given in the next result, in which we provide a majorization of the entropy with respect to the Total Variation distance which is easier to handle here. Note that all the previous results are still true if we use  $d_H$  instead of  $d_{TV}$  since (A.2) also permits to retrieve entropy bounds for  $d_H$  from entropy bounds for  $d_{TV}$ .

PROPOSITION 3.3. Let  $\epsilon > 0$  and  $s > 0$ , if  $\log \frac{1}{\epsilon} \lesssim \ell$  and  $f \in \mathcal{H}^\ell$  is such that  $\|\theta\|^2 \lesssim 2\ell + 1$ , then

$$\log N(\epsilon, \mathcal{P}_f, d_{TV}) \lesssim \ell^2 \left( \log \frac{1}{\epsilon} + \log \|\theta\|_{\mathcal{H}_1} \right).$$

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  $f$  given below. It is possible since in the first inequality we have carefully expressed the dependency on  $f$  and  $\ell$ .

The method to build an  $\epsilon$ -covering of  $\mathcal{P}_f$  follows two natural steps:

- 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 components of the finite mixture  $\tilde{g}$  uniformly bounded in  $g$  (depending on  $f$  and  $\epsilon$ );

- use Proposition 3.2 for the finite mixture to well approximate  $\mathbb{P}_{\theta,\tilde{g}}$ .

The proof itself is delayed to the Appendix.

3.1.2. *Case of unknown  $f$ .* We now describe the picture when  $f$  is unknown, which is the main objective of this paper. We assume that  $f$  belongs to  $\mathcal{H}_s$ . In order to bound the bracketing entropy, we define a sieve over  $\mathcal{H}_s$  which depends on a frequency cut-off  $\ell$  and a size parameter  $w$ . We then get

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

THEOREM 3.1. *Let be given  $\epsilon > 0$  small enough, and assume that  $\ell_\epsilon$  and  $w_\epsilon$  are such that  $\log \frac{1}{\epsilon} \lesssim \ell_\epsilon$  and  $w_\epsilon \lesssim \sqrt{\ell_\epsilon}$ , then*

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

The proof of Theorem 3.1 is based on two simple results. The first one is the Girsanov formula obtained by [BG10] in appendix A.2.2 (in the case of known  $g$ ): it can be extended to the situation of unknown  $g$  and complex trajectories as in (2.1), which leads to

$$(3.1) \quad \frac{d\mathbb{P}_{f,g}}{d\mathbb{P}_{f^0,g^0}}(Y) = \frac{\int_0^1 \exp(2\Re\langle f^{-\alpha_1}, dY \rangle - \|f^{-\alpha_1}\|^2) dg(\alpha_1)}{\int_0^1 \exp(2\Re\langle f^0, -\alpha_2, dY \rangle - \|f^0, -\alpha_2\|^2) dg^0(\alpha_2)},$$

for any measurable trajectory  $Y$ .

The second result is given in the following lemma.

LEMMA 3.1. *Let  $f$  and  $\tilde{f}$  be any functions in  $L^2_{\mathbb{C}}([0,1])$ ,  $g$  be any shift distribution 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 OF THEOREM 3.1. The idea of the demonstration is to build a  $\epsilon$ -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 remark that by 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 majorization of the first one comes from Lemma 3.1. The second term is handled uniformly in  $\tilde{f}$  by Proposition 3.3.

Now, we build  $\epsilon/2$ -coverings of  $\mathbb{P}_{f,g}$  for fixed  $g$  from an  $\epsilon/\sqrt{2}$ -covering of  $f$  for the  $L^2$ -norm:

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

□

According to inequality (A.2) and since  $\log \frac{1}{\epsilon^2} \lesssim \log \frac{1}{\epsilon}$ , we can easily deduce the next corollary.

COROLLARY 2. *Let be given  $\epsilon > 0$  small enough, and assume  $\log \frac{1}{\epsilon} \lesssim \ell_\epsilon$  and  $w_\epsilon \leq \sqrt{2\ell_\epsilon + 1}$ , then*

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

REMARK 3.1. *i) Even if the model studied here is a very special case of Gaussian mixture models, one may think that such kind of results may help the analysis of more general mixture cases within a growing dimension setting.*

*ii) In our case, we will use a much higher choice of  $\ell_\epsilon$  than  $\log \frac{1}{\epsilon}$ . This choice will be fixed in section 3.4.*

3.2. *Link between Kullback-Leibler and Hellinger neighbourhoods.* We first recall a useful result of Wong & Shen given as Theorem 5 in [WS95]. It enables to handle Hellinger neighbourhoods instead of  $\mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0, g^0}, d_{KL})$ , which is generally easier for mixture models.

THEOREM 3.2 (Wong & Shen). *Let  $\mu$  and  $\nu$  be two measures such that  $\mu$  is a.c. with respect to  $\nu$  with a density  $q = d\mu/d\nu$ . Assume that  $d_H(\mu, \nu)^2 = \int [\sqrt{q} - 1]^2 d\nu \leq \epsilon^2$  and that there exists  $\delta \in (0, 1]$  such that*

$$(3.2) \quad M_\delta^2 := \int_{q \geq \epsilon^{1/\delta}} q^{\delta+1} d\nu < \infty.$$

Then, for  $\epsilon$  small enough, there exists a universal constant  $C$  large enough such that

$$d_{KL}(\mu, \nu) = \int q \log q d\nu \leq C \log(M_\delta) \epsilon^2 \log \frac{1}{\epsilon},$$

and

$$V(\mu, \nu) \leq \int q \log^2 q d\nu \leq C \log(M_\delta)^2 \epsilon^2 \left[ \log \frac{1}{\epsilon} \right]^2.$$

Hence, Hellinger neighbourhoods are almost Kullback-Leibler ones (up to some logarithm terms) provided that a sufficiently large moment exists for  $q$  ( $q \log q$  is killed by  $q^{1+\delta}$  for large values of  $q$  and a second order expansion of  $q \log q - q + 1$  around 1 yields a term similar to  $[\sqrt{q} - 1]^2$ ). Next proposition shows that condition (3.2) is satisfied in our SIM.

**PROPOSITION 3.4.** *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  $q = \frac{d\mathbb{P}_{f^0, g^0}}{d\mathbb{P}_{f, g}}$ . There exists  $\delta \in (0, 1]$  such that the constant defined in equation (3.2)  $M_\delta^2$  is uniformly bounded with respect to  $f$ .*

**3.3. Hellinger neighbourhoods.** Proposition 3.4 will allow to use Theorem 3.2, thus we now aim to find a lower bound on Hellinger neighbourhood of  $\mathbb{P}_{f^0, g^0}$ . Consider a frequency cut-off  $\ell_n$  that will be fixed later. For any  $f \in \mathcal{H}_s^{\ell_n}$  and  $g \in \mathfrak{M}([0, 1])$ , remind that we denote  $\theta := \theta(f)$  as well as  $\theta^0 = \theta(f^0)$ . We define  $f_{\ell_n}^0$  the  $L^2$  projection of  $f^0$  on the subspace  $\mathcal{H}_s^{\ell_n}$ .

For sake of simplicity,  $\mathbb{E}_0 F(Y)$  will refer to the expectation of a function  $F$  of the trajectory  $Y$  when  $Y$  follows  $\mathbb{P}_{f^0, g^0}$ . The triangle inequality applied to the Hellinger distance shows that

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

In the sequel, we will provide sufficiently sharp upper bound on  $(E_1)$ ,  $(E_2)$ ,  $(E_3)$  so that we will be able to find a suitable lower bound of the prior mass of Hellinger neighbourhoods.

*Upper bound of (E<sub>1</sub>).* We first bound (E<sub>1</sub>) using  $d_H^2 \leq d_{KL}$  with the Girsanov formula (3.1)

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

We now obtain the upper bound of (E<sub>1</sub>) according to the next proposition.

PROPOSITION 3.5. *Assume that  $Y \sim \mathbb{P}_{f^0, g^0}$  and  $f^0 \in \mathcal{H}_s$ , then*

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

*Upper bound of (E<sub>3</sub>).* We will be interested in the Hellinger distance when  $f_{\ell_n}^0$  is close to  $f$ , and the dimension  $\ell_n$  grows up to  $+\infty$  (the mixture law on  $[0, 1]$  is the same for the two laws). The important fact will be its exclusive dependence with respect to the  $L^2$  distance between  $f_{\ell_n}^0$  and  $f$ . This upper bound is given in the next proposition, whose proof is immediate from Lemma 3.1 and equation (A.2).

PROPOSITION 3.6. *Assume that  $f \in \mathcal{H}_s^{\ell_n}$  and  $g \in \mathfrak{M}([0, 1])$ , then*

$$d_H(\mathbb{P}_{f_{\ell_n}^0, g}, \mathbb{P}_{f, g}) \leq 2^{1/4} \sqrt{\|f - f_{\ell_n}^0\|}.$$

*Upper bound for (E<sub>2</sub>).* This term is clearly the more difficult to handle. We will obtain a convenient result using some elements obtained in Proposition 3.3. For a given  $\epsilon_n > 0$ ,  $\ell_n, f_{\ell_n}^0 \in \mathcal{H}_s^{\ell_n}$  and  $g^0 \in \mathfrak{M}([0, 1])$ , we know that one may find a mixture model  $\tilde{g}$  such that  $d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}) < \epsilon_n$  and  $\tilde{g}$  has  $C\ell_n^2$  points of support in  $[0, 1]$  as soon as  $\epsilon_n$  is small enough and  $\log \frac{1}{\epsilon_n} \lesssim \ell_n$  (the condition  $\|f_{\ell_n}^0\|^2 \leq 2\ell_n + 1$  is immediate since  $f^0$  does not depend on  $n$ ). The next step is to control the Hellinger distance  $d_H(\mathbb{P}_{f_{\ell_n}^0, g}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}})$  for  $g \in \mathfrak{M}([0, 1])$ , and this can be done thanks to an adaptation in dimension  $2\ell_n + 1$  of Lemma 5.1 of [GvdV01].

LEMMA 3.2. *Let be given  $\tilde{g}$  a discrete mixture law whose support is of cardinal  $J$  whose support points  $(\varphi_j)_{j=1\dots J}$  are such that  $\tilde{g}(\varphi_j) = p_j$  and*

$\eta$ -separated, i.e.  $|\varphi_j - \varphi_i| \geq \eta, \forall i \neq j$ , then  $\forall \tilde{g} \in \mathfrak{M}([0, 1])$

$$d_H^2(\mathbb{P}_{f_{\ell_n}^0, \tilde{g}}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}) \leq \sqrt{\frac{\pi}{2}} \|f_{\ell_n}^0\|_{\mathcal{H}_1} \eta + 2 \sum_{j=1}^J |\tilde{g}([\varphi_j - \eta/2, \varphi_j + \eta/2]) - \tilde{g}(\varphi_j)|.$$

Note that Lemma 3.2 needs a discrete mixture with  $\eta$ -separated support points. The following result permits to obtain such a mixture.

PROPOSITION 3.7. *Assume that and  $f^0 \in \mathcal{H}_s$  for  $s \geq 1$ ,  $g^0 \in \mathfrak{M}([0, 1])$ , and  $\log \frac{1}{\epsilon_n} \lesssim \ell_n$ . For any  $\eta_n \leq \epsilon_n^2$ , there exists a discrete distribution  $\tilde{g}$  with in its support at most  $J_n \lesssim \ell_n^2$  points denoted  $(\psi_j)_{j=1 \dots J_n}$ , such that these points are  $\eta_n$ -separated, and*

$$d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}) \leq \left(1 + (8\pi)^{1/4} \|f_{\ell_n}^0\|_{\mathcal{H}_1}^{1/2}\right) \epsilon_n.$$

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

$$d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, g}) \leq \left(1 + (8\pi)^{1/4} \|f^0\|_{\mathcal{H}_1}^{1/2}\right) \epsilon_n + \sqrt{\sqrt{\frac{\pi}{2}} \|f^0\|_{\mathcal{H}_1} \eta_n + 2 \sum_{j=1}^{J_n} |g(\psi_j - \eta_n/2, \psi_j + \eta_n/2) - \tilde{g}(\psi_j)|}.$$

Latter in the paper we obtain a more general upper bound for  $(E_2)$ , based on the Wasserstein distance. We could use it to retrieve Proposition 3.7, but it also leads to Hellinger neighbourhoods described in terms of the Total Variation distance from  $g$  to  $g^0$ . This last distance is adapted to smooth densities  $g$  but not to the ones considered here, when the prior distribution for  $g$  is a Dirichlet process.

*Description of a Hellinger neighbourhood.* We can now gather the upper bounds of  $(E_1)$ ,  $(E_2)$ , and  $(E_3)$  to get the following result.

PROPOSITION 3.8. *Assume that  $f^0 \in \mathcal{H}_s$  for  $s \geq 1$  and  $g^0 \in \mathfrak{M}([0, 1])$ . Choose the threshold such as  $\epsilon_n^{-1/s} \lesssim \ell_n \lesssim \epsilon_n^{-1/s}$  and  $\eta_n := \epsilon_n^2$ , and consider the finite mixture  $\tilde{g}$  provided by Proposition 3.7. Define*

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

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

Then, there exists a constant  $C_0$  depending only on  $\|f^0\|_{\mathcal{H}_1}$  such that for any  $g \in \mathcal{G}_{\epsilon_n}$  and  $f \in \mathcal{F}_{\epsilon_n}$ ,

$$d_H(\mathbb{P}_{f^0, g^0}, \mathbb{P}_{f, g}) \leq C_0 \epsilon_n.$$

3.4. *Checking the conditions of Theorem 2.1.* We first prove the minoration for the lower bound (2.6), necessary to apply Theorem 2.1.

PROPOSITION 3.9. *Assume that and  $f^0 \in \mathcal{H}_s$  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  $n \rightarrow +\infty$ , and for the prior defined in paragraph 2.3, there exists a constant  $c > 0$  such that*

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

where

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

Proposition 3.9 relies on Theorem 3.2, which permits to use Hellinger neighbourhoods instead of  $\mathcal{V}_{\epsilon_n}(\mathbb{P}_{f^0, g^0}, d_{KL})$ , and on Proposition 3.8, which describes suitable Hellinger neighbourhoods. To control their prior mass, we remind the following useful result appeared as Lemma 6.1 of [GGvdV00]. This enables to find a lower bound of  $\ell_1$ -ball of radius  $r$  under Dirichlet prior.

LEMMA 3.3 ([GGvdV00]). *Let  $r > 0$  and  $(X_1, \dots, X_N)$  be distributed according to the Dirichlet distribution on the  $\ell_1$  simplex of dimension  $N - 1$  with parameters  $(m, \alpha_1, \dots, \alpha_N)$ . Assume that  $\sum_j \alpha_j = m$  and  $A r^b \leq \alpha_j \leq 1$  for some constants  $A$  and  $b$ . Let  $(x_1, \dots, x_N)$  be any points on the  $N$  simplex, there exists  $c$  and  $C$  that only depend on  $A$  and  $b$  such that if  $r \leq 1/N$*

$$\Pr \left( \sum_{j=1}^N |X_j - x_j| \leq 2r \right) \geq C \exp \left( -cN \log \frac{1}{r} \right)$$

In the proof of Proposition 3.9 (delayed to the Appendix), one can see that we could obtain a suitable lower bound as soon as  $\lambda(\ell_n) \geq e^{-c\ell_n^2 \log \ell_n}$  for a constant  $c$ . Of course, a distribution  $\lambda$  with some heavier tail would also suit here. However, such a heavier tail is not suitable for the control of the term  $\Pi_n(\mathcal{P} \setminus \mathcal{P}_n)$  which is detailed in the next proposition.

PROPOSITION 3.10. *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*

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

and

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

We are now able to conclude the proof of the posterior consistency for Theorem 2.2.

PROOF OF THEOREM 2.2. Take  $\epsilon_n := n^{-\alpha}(\log n)^\kappa$  and  $k_n := n^\beta(\log n)^\gamma$ . From our definition (2.7), we have also  $\xi_n^{-2} = n^{\mu_s}(\log n)^\zeta$ , and we look for admissible values of  $\alpha, \beta, \kappa, \gamma, \mu_s$ , and  $\zeta$  in order to satisfy (2.4), (2.5) and (2.6).

Proposition (3.9) imposes that in order to satisfy (2.6), we could check that

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

This is true as soon as  $\epsilon_n$  satisfies

$$\alpha \leq \frac{s}{2s+2} \quad \text{and} \quad \kappa > (\rho s + 2)/(2s + 2).$$

Moreover, we obtain the first condition on  $\mu_s$ :  $\mu_s \leq 1 - 2\alpha$ , and if  $\mu_s = 1 - 2\alpha$  then  $\zeta < 2\kappa$ .

Now, Proposition (3.10) shows that (2.4) is fulfilled provided that

$$(3.3) \quad k_n^2 \left[ \log k_n + \log \frac{1}{\epsilon_n} \right] \lesssim n\epsilon_n^2 = n^{1-2\alpha}(\log n)^{2\kappa}.$$

This condition is satisfied when  $2\beta \leq 1 - 2\alpha$  and  $2\gamma + 1 \leq 2\kappa$ . At last, Proposition (3.10) again ensures that (2.5) is true as soon as

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

and we deduce from (3.3) that

$$2\beta = 1 - 2\alpha \quad \text{and} \quad -\rho/2 + \kappa \leq \gamma \leq -1/2 + \kappa.$$

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  $\mu_s$  yields  $\mu_s \geq 1/2 - \alpha \geq \frac{1}{2s+2}$  (which naturally drives us to set  $\mu_s = 1/4$  (case  $s = 1$ ) for adaptive prior).

We split the proof according to the adaptive or non adaptive case.

*Adaptive prior.* We first set  $\mu$  independent of  $s$  and equal to  $1/4$ . For any  $s \in [1, 3]$ , we see that  $\alpha(s) = s/(2s + 2)$  is the admissible largest value of  $\alpha$  and  $\alpha(s) = 3/8 < s/(2s + 2)$  as soon as  $s > 3$ . The corresponding value of  $\beta$  is  $1/(2s + 2)$  when  $s \in [1, 3]$  and  $\beta = 1/8$  otherwise. Any choice of  $\zeta \in [3/2, 2)$  permits to deal with the conditions on  $\zeta$  that appears when  $s = 1$  or  $s \geq 3$ . The other values of  $\gamma$  and  $\kappa$  may be determined with respect to  $\rho$ . For instance, if we choose  $\rho \in (1, 2)$ , we can take  $\kappa = 1$  and  $\gamma = 1/2$ .

*Non adaptive prior.* The non adaptive case is much more simpler since it is sufficient to fix

$$\mu_s = 1 - 2\alpha = 2/(2s + 2)$$

and  $\zeta = 0$  to obtain suitable calibrations for  $\alpha, \beta, \kappa$  and  $\gamma$ . This achieves the proof.  $\square$

We now slightly discuss on the proof of Theorem 2.3. We follow exactly the same strategy but the neighbourhoods are a little bit modified on the  $g$  coordinate. In this view, we establish a useful bound which concerns the closeness of two laws  $\mathbb{P}_{f,g}$  and  $\mathbb{P}_{f,\tilde{g}}$ , when we keep the same shape  $f \in \mathcal{H}_1$ . In a sense, the next proposition replaces Lemma 3.2, Proposition 3.7 and Proposition 3.8 in the special framework of smooth densities.

Consider the inverse functions of the distribution functions defined by

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

The Wasserstein (or Kantorovich) distance  $W_1$  is given by

$$W_1(g, \tilde{g}) := \int_0^1 \left| G^{-1}(t) - \tilde{G}^{-1}(u) \right| dt.$$

PROPOSITION 3.11. *Consider  $f \in \mathcal{H}_1$ , and let  $g$  and  $\tilde{g}$  be any measures on  $(0, 1]$ . Then*

$$\begin{aligned} d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{f,\tilde{g}}) &\leq \sqrt{2\pi} \|f\|_{\mathcal{H}_1} W_1(g, \tilde{g}) \\ &\leq \sqrt{2\pi} \|f\|_{\mathcal{H}_1} d_{TV}(g, \tilde{g}) \leq \pi \|f\|_{\mathcal{H}_1} \|g - \tilde{g}\|/\sqrt{2}. \end{aligned}$$

PROOF OF THEOREM 2.3. We mimic the proof of Theorem 2.2.

*Complementary of the sieve.* First, we consider the sieve over  $\mathcal{P}$  defined as the set of all possible laws when  $f$  has truncated Fourier coefficients and a restricted  $L^2$  norm:

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



where  $k_n$  is a sequence such that  $k_n \mapsto +\infty$  as  $n \mapsto +\infty$ , and  $w_n^2 = 4k_n + 2$ .

Since our sieve is included in the set of all mixture laws, we can apply Proposition 3.10 and get

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

*Entropy estimates.* Since  $\mathfrak{M}_\nu([0, 1])(2A) \subset \mathfrak{M}([0, 1])$ , our sieve is included in the sieve considered above, we also deduce that for any sequence  $\epsilon \mapsto 0$ :

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

*Lower bound of the prior of Kullback neighbourhoods.* We use the description of Kullback neighbourhoods based on our preliminary results. We define  $\tilde{\epsilon}_n = c\epsilon_n \left( \log \frac{1}{\epsilon_n} \right)^{-1}$ , an integer  $\ell_n$  such that  $\tilde{\epsilon}_n^{-1/s} \lesssim \ell_n \lesssim \tilde{\epsilon}_n^{-1/s}$ , and the sets

$$\mathcal{F}_{\tilde{\epsilon}_n} := \left\{ f \in \mathcal{H}_s^{\ell_n} : \|f - f_{\ell_n}^0\| \leq \tilde{\epsilon}_n^2 \right\},$$

and

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

We deduce from Lemma 3.1, Proposition 3.11 and arguments of Proposition 3.9 that as soon as  $f \in \mathcal{F}_{\tilde{\epsilon}_n}$  and  $g \in \mathcal{G}_{\tilde{\epsilon}_n}$ ,  $\mathbb{P}_{f, g}$  belongs to an  $\epsilon_n$  Kullback neighbourhood of  $\mathbb{P}_{f^0, g^0}$ . From Proposition 3.9, we can use the following lower bound of the prior mass on  $\mathcal{F}_{\tilde{\epsilon}_n}$ :

$$\Pi_n(\mathcal{F}_{\tilde{\epsilon}_n}) \geq e^{-(c+o(1)) \left[ \tilde{\epsilon}_n^{-2/s} (\log(1/\epsilon_n))^{\rho+2/s} \vee \xi_n^{-2} \right]}.$$

According to Theorem C.1 given in the appendix, we know that

$$\Pi_n(\mathcal{G}_{\tilde{\epsilon}_n}) \geq e^{-(c+o(1))\tilde{\epsilon}_n^{-1/(k_\nu+1/2)}} \geq e^{-(c+o(1))\tilde{\epsilon}_n^{-\frac{1}{\nu}}}$$

since  $k_\nu + 1/2 \leq \nu$  (see also [LS01] for a very complete survey on the small ball probability estimation for Gaussian processes).

*Contraction Rate.* We now find a suitable choice of  $k_n$  and  $\epsilon_n$  in order to satisfy Theorem 2.1 of [GGvdV00], *i.e.*

$$\begin{aligned} \Pi_n(\mathcal{G}_{\tilde{\epsilon}_n}) \Pi_n(\mathcal{F}_{\tilde{\epsilon}_n}) &\geq e^{-Cn\epsilon_n^2} \\ \log D(\epsilon_n, \mathcal{P}_{k_n, w_n}, d_H) &\lesssim n\epsilon_n^2 \\ \Pi_n(\mathcal{P} \setminus \mathcal{P}_{k_n, w_n}) &\leq e^{-(C+4)n\epsilon_n^2}. \end{aligned}$$

Following the arguments already developed in Theorem 2.2, we can find  $\gamma > 0$  and  $\kappa > 0$  such that

$$\epsilon_n := n^{-\left[ \frac{\nu}{2\nu+1} \wedge \frac{s}{2s+2} \wedge \frac{3}{8} \right]} \log(n)^\kappa, \quad k_n = n^{\frac{1}{2} - \left[ \frac{\nu}{2\nu+1} \wedge \frac{s}{2s+2} \wedge \frac{3}{8} \right]} \log(n)^\gamma.$$

□

**4. Identifiability and semiparametric results.** In the Shape Invariant Model, an important issue is the identifiability of the model with respect to the unknown curve  $f$  and the unknown mixture law  $g$ . We first discuss on a quite generic identifiability condition for  $\mathbb{P}_{f,g}$ . Then, we deduce from the previous section a contraction rate of the posterior distribution around the true  $f^0$  and  $g^0$ .

4.1. *Identifiability of the model.* In previous works on SIM, the identifiability of the model is generally given according to a restriction on the support of  $g$ . For instance, [BG10] assume the support of  $g$  to be an interval included in  $[-1/4, 1/4]$  (their shapes are defined on  $[-1/2; 1/2]$  instead of  $[0, 1]$  in our paper) and  $g$  is assumed to have 0 mean although  $f$  is supposed to have a non vanishing first Fourier coefficient ( $\theta_1(f) \neq 0$ ). The same kind of condition on the support of  $g$  is 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 such a hypothesis. First, we recall that for any curve  $Y$  sampled from the SIM, the first Fourier coefficient is given by  $\theta_1(Y) = \theta_1^0 e^{-i2\pi\tau} + \xi$  (here  $\theta_1^0 = \theta_1(f^0)$ ). Hence, up to a simple change of variable in  $\tau$ , we can always modify  $g$  in  $\tilde{g}$  such that  $\theta_1^0 \in \mathbb{R}_+$ . It is for instance sufficient to fix  $\tilde{g}(\varphi) = g(\varphi + \alpha)$  where  $\alpha$  is the complex argument of  $\theta_1^0$ . Hence, to impose such an identifiability condition, we have chosen to restrict  $f$  to  $\mathcal{F}_s$ . This condition is not restrictive up to a change of measure for the random variable  $\tau$ . The next result provides an identifiability criterion for the model, both for  $f$  and  $g$ .

**THEOREM 4.1.** *Assume that  $f \in \mathcal{F}_s$  defined above and  $g \in \mathfrak{M}([0, 1])^*$  defined by*

$$\mathfrak{M}([0, 1])^* := \{g \in \mathfrak{M}([0, 1]) \mid \exists(c, C) \in \mathbb{R}_+^* : \forall k \in \mathbb{Z}, c < |k|^\nu |\theta_k(g)| < C\},$$

*where  $\nu > 1$ , then the Shape Invariant Model described by (2.3) is identifiable.*

**PROOF.** The demonstration of Theorem 4.1 is decomposed using three hierarchical steps. First, we prove that if  $\mathbb{P}_{f,g} = \mathbb{P}_{\tilde{f},\tilde{g}}$ , then one has necessarily  $\theta_1(f) = \theta_1(\tilde{f})$ . Then we deduce from this point 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 with respect to the Lebesgue measure on  $[0, 1]$ . In the sequel we use the same notation  $g$  to refer to 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  $k^{\text{th}}$  Fourier coefficient when the curve follows the Shape Invariant Model (2.3). Of course, we have the following implications

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

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 note that the distribution  $\mathbb{P}_{f,g}^0$  does not bring any information on the measure  $g$ , and is not helpful for its identifiability. Concerning now the first Fourier coefficient, we use the notation  $\theta_1 := \theta_1(f)$ ,  $\tilde{\theta}_1 := \theta_1(\tilde{f})$  and remark that

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

Assume now that  $\tilde{\theta}_1 \neq \theta_1$ , without loss of generality  $\tilde{\theta}_1 > \theta_1 > 0$  and consider the disk  $D_{\mathbb{C}} \left( 0, \frac{\tilde{\theta}_1 - \theta_1}{2} \right)$ , we then get  $\forall z \in D_{\mathbb{C}} \left( 0, \frac{\tilde{\theta}_1 - \theta_1}{2} \right), \forall \alpha \in [0, 1]$  :

$$|\theta_1 e^{i2\pi\alpha} - z| < \frac{\tilde{\theta}_1 + \theta_1}{2} \text{ and } |\tilde{\theta}_1 e^{i2\pi\alpha} - z| > \frac{\tilde{\theta}_1 + \theta_1}{2}.$$

Hence, for all  $z \in D_{\mathbb{C}} \left( 0, \frac{\tilde{\theta}_1 - \theta_1}{2} \right)$ , 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 can thus write the following lower bound of the Total Variation<sup>1</sup>.

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

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  $f$  and  $\tilde{f}$  belong to  $\mathcal{F}_s(A)$ .

---

<sup>1</sup>It is indeed possible to write an explicit lower bound which will depend on  $|\theta_1 - \tilde{\theta}_1|^2$ , with a radius smaller than  $\frac{\tilde{\theta}_1 - \theta_1}{2}$ .

*Point 2: Identifiability on  $g$ .* We still assume that  $d_{TV}(\mathbb{P}_{f,g}^1, \mathbb{P}_{\tilde{f},\tilde{g}}^1) = 0$ . We know that  $\theta_1 = \tilde{\theta}_1$  and we want to infer that  $g = \tilde{g}$ . We are going to establish this result using only the first harmonic of the curves. Using a polar change of variables  $z = \rho e^{i\varphi}$ , we can write that

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

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,  $\psi_a$  is upper bounded by  $4\pi e^a$ , and a very rough inequality yields<sup>2</sup>  $|\psi_a(\varphi)| \geq \frac{|\psi_a(\varphi)|^2}{4\pi e^a}$ . Hence,

$$(4.1) \quad d_{TV}(\mathbb{P}_{f,g}^1, \mathbb{P}_{\tilde{f},\tilde{g}}^1) \geq \frac{1}{8\pi^2} \int_0^{+\infty} \rho e^{-(\theta_1^2 + \rho^2 + 2\theta_1\rho)} \|\psi_{2\rho\theta_1}\|^2 d\rho.$$

Using the fact that  $\nu > 1$ ,  $h$  may be expanded in Fourier series since  $h \in \mathcal{L}^2([0, 1])$ :

$$h(x) = \sum_{n \in \mathbb{Z}} c_n(h) e^{i2\pi n x},$$

and we can also obtain the Fourier decomposition of  $\psi_a$ :

$$\psi_a(\varphi) = \sum_{n \in \mathbb{Z}} c_n(h) \int_0^{2\pi} e^{a \cos(u)} e^{inu} du e^{i2\pi n \varphi}.$$

Thus, the  $L^2$  norm of  $\psi_a$  is given by

$$(4.2) \quad \|\psi_a\|^2 = \sum_{n \in \mathbb{Z}} |c_n(h)|^2 \left| \int_0^{2\pi} e^{a \cos(u)} e^{inu} du \right|^2,$$

---

<sup>2</sup>Such an inequality is not very sharp and we can instead use an argument based on the Laplace transform of  $g$  and  $\tilde{g}$ . The main advantage of such inequality is to handle  $L^2$  norms instead of  $L^1$  ones.

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

$$\begin{aligned} \int_0^{2\pi} e^{a \cos(u)} e^{inu} du &= \int_0^{2\pi} e^{a \cos(u)} [T_n(\cos u) + i(\sin u)U_n(\cos u)] du \\ &= \int_0^{2\pi} \sum_{k \geq 0} \frac{a^k (\cos u)^k}{k!} \left[ T_n(\cos u) + i(\sin u) \sum_{j=0}^n \beta_j (\cos u)^j \right] du \end{aligned}$$

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

$$\begin{aligned} \int_0^{2\pi} e^{a \cos(u)} e^{inu} du &= \int_0^{2\pi} \sum_{k \geq 0} \frac{a^k (\cos u)^k}{k!} T_n(\cos u) du \\ &\quad + i \sum_{k \geq 0} \sum_{j=0}^n \beta_j \frac{a^k}{k!} \int_0^{2\pi} \sin u (\cos u)^{k+j} du \\ &= \int_0^{2\pi} \sum_{k \geq 0} \frac{a^k (\cos u)^k}{k!} T_n(\cos u) du \\ &= \int_0^{2\pi} e^{a \cos(u)} \cos(nu) du \in \mathbb{R} \quad \text{if } a \in \mathbb{R}. \end{aligned}$$

We denote  $A_n$  the following (holomorphic) function of the variable  $a$  as

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

and equation (4.2) yields

$$(4.3) \quad \|\psi_a\|^2 = \sum_{n \in \mathbb{Z}} |c_n(h)|^2 A_n(a)^2.$$

Moreover, for each  $n$ ,  $A_n$  is not the null function, otherwise it would be the case for each of its derivative but remark that  $(\cos u)^n$  may be decomposed

in the basis  $(T_k)$  and using successive derivations

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

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

$$(4.4) \quad A_n(a) \underset{a \rightarrow 0}{\sim} \frac{2^{1-n} \pi}{n!} a^n.$$

We can conclude the proof of the identifiability of  $g$  using (4.3) in (4.1) to obtain

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

From (4.4), we can deduce that each integral  $I_n(\theta_1) \neq 0, \forall n \in \mathbb{Z}$  and we then conclude that:

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

*Point 3: Identifiability on  $f$ .* We end the argument and prove that  $\mathbb{P}_{f,g} = \mathbb{P}_{\tilde{f},\tilde{g}}$  implies  $f = \tilde{f}$ . We already know that  $g = \tilde{g}$  and it remains to establish the equality for all the Fourier coefficients whose frequency is different from 0 and 1. By a similar argument as the one used for the identifiability of  $\theta_1$  (Point 1), we can easily show that

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

But we cannot directly conclude here since it is not reasonable to restrict the phase of each others coefficients  $\theta_k(f)$  to a special value (as it is the case for  $\theta_1(f)$  which is positive). We assume that  $\tilde{\theta}_k = \theta_k e^{i\varphi}$ . Since  $g = \tilde{g}$ , we have

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

Now, if one considers  $z = x + iy$ ,  $F$  is differentiable with respect to  $x$  and  $y$  and  $F(0) = 0$ . A simple computation of  $\nabla F(0)$  shows that  $\nabla F(0)$  is the vector (written in the complex form)

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

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

In a sense, the main difficulty of the proof above is the implication of  $d_{TV}(\mathbb{P}_{f,g}^1, \mathbb{P}_{\tilde{f},\tilde{g}}^1) \implies g = \tilde{g}$ . Then, the identifiability follows using a chaining argument  $\theta_1(f) \rightarrow g \rightarrow \theta_k(f), \forall k \notin \{0, 1\}$ . We will see that this part of the proof can also be used to obtain a contraction rate for  $f$  and  $g$  around  $f^0$  and  $g^0$ . We recall here the main inequality used above:  $\forall \theta_1 > 0$  and  $\forall (g, \tilde{g}) \in \mathfrak{M}_\nu([0, 1])(A)$ , the identifiability on  $g$  is traduced by

$$(4.5) \quad d_{TV}(\mathbb{P}_{\theta_1, g}^1, \mathbb{P}_{\theta_1, \tilde{g}}^1) \geq \frac{1}{8\pi^2} \sum_{n \in \mathbb{Z}} |c_n(g - \tilde{g})|^2 \left( \int_0^\infty \rho e^{-(\rho + \theta_1)^2} A_n(2\rho\theta_1)^2 d\rho \right)$$

The aim of the next paragraph is to exploit this inequality to produce a contraction rate of  $g$  around  $g^0$ .

#### 4.2. Contraction rate of the posterior distribution around $f^0$ and $g^0$ .

4.2.1. *Link with deconvolution with unknown variance operator.* We provide in this section an upper bound on the contraction rate of the posterior law around  $f^0$  and  $g^0$ . This question is somewhat natural owing to the identifiability result obtained in the previous section. We thus assume for the rest of the paper that  $f \in \mathcal{F}_s$  and  $g \in \mathfrak{M}_\nu([0, 1])(A)$  for some parameters  $s \geq 1$  and  $\nu > 1$ .

Remark first that our problem written in the Fourier domain seems strongly related to the standard deconvolution with unknown variance setting. For instance, the first observable Fourier coefficients are

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

and up to a division by  $\theta_1$ , it can also be parametrised as

$$(4.6) \quad \tilde{\theta}_1(Y_j) = e^{-i2\pi\tau_j} + \frac{\epsilon_{1,j}}{\theta_1}, \forall j \in \{1 \dots n\},$$

which is very similar to the problem  $Y = X + \epsilon$  studied for instance by [Mat02] where  $\epsilon$  follows a Gaussian law whose variance (here  $1/\theta_1^2$ ) is unknown. As pointed in [Mat02] (see also the more recent work [BM05] where similar situations are extensively detailed), such a particular setting is rather unfavourable for statistical estimation since convergence rates are generally of logarithmic order. Such a phenomenon also occurs in our setting, except for the first Fourier coefficient of  $f$  as pointed in the next proposition.

The roadmap of this paragraph is similar to the proof of Theorem 4.1. We first provide a simple lower bound of  $d_{TV}$  which enables to conclude for the first Fourier coefficient. Then, we still use the first marginal to compute a contraction rate for the posterior distribution on  $g$  around  $g^0$ . At last, we chain all these results to provide a contraction rate for the posterior distribution on  $f$  around  $f^0$ .

4.2.2. *Contraction rate on the first Fourier coefficient.*

PROPOSITION 4.1. *Assume that  $(f, g) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0, 1])(A)$ , then the posterior distribution satisfies*

$$\Pi_n \left( \theta_1 \in B \left( \theta_1^0, M \epsilon_n^{1/3} \right)^c \mid Y_1, \dots, Y_n \right) \mapsto 0$$

in  $\mathbb{P}_{f^0, g^0}$  probability as  $n \rightarrow +\infty$  for a sufficiently large  $M$ . The contraction rate around the true Fourier coefficient is thus at least  $n^{-1/3 \times [\nu/(2\nu+1) \wedge s/(2s+2) \wedge 3/8]} (\log n)^{1/3}$ .

PROOF. The demonstration is quite simple. Remark that using the beginning of the proof of Theorem 4.1, one can show that for any  $\theta_1$  such that  $0 < \eta < |\theta_1 - \theta_1^0| < \theta_1^0/2$ , one can bound, for any  $g \in \mathfrak{M}_\nu([0, 1])(A)$ , the Total Variation distance between  $\mathbb{P}_{f, g}$  and  $\mathbb{P}_{f^0, g^0}$ . Remark that

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

owing to the restriction of  $\mathbb{P}_{f, g}$  to the first Fourier marginal and the variational definition of the Total Variation distance. Then

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



for a suitable small enough constant  $C(\theta_1^0)$ . Now, one can use simple inclusions and Pinsker inequality

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

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

4.2.3. *Posterior contraction rate around  $g^0$ .* We now study the contraction rate of the posterior distribution around the true mixture law  $g^0$ . This result is stated below.

THEOREM 4.2. *Assume  $(f^0, g^0) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0, 1])(A)$ , then*

$$\Pi_n (g : \|g - g^0\|^2 > M \log^{-2\nu}(n) | Y_1, \dots, Y_n) \longrightarrow 0$$

in  $\mathbb{P}_{f^0, g^0}$  probability as  $n \rightarrow +\infty$  for a sufficiently large  $M$ .

PROOF. We first restrict ourselves to the first marginal on Fourier coefficient as before. Using Theorem 2.2, we know that

$$\Pi_n \{ \mathbb{P}_{f,g} : d_H(\mathbb{P}_{f,g}, \mathbb{P}_{f^0, g^0}) \geq M \epsilon_n | Y_1, \dots, Y_n \} \rightarrow 0 \quad \text{as } n \rightarrow +\infty.$$

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

$$(4.7) \quad \Pi_n \left\{ \mathbb{P}_{f,g} : d_{TV}(\mathbb{P}_{\theta_{1,g}^1}, \mathbb{P}_{\theta_{1,g^0}^1}) \geq M \epsilon_n | Y_1, \dots, Y_n \right\} \rightarrow 0 \quad \text{as } n \rightarrow +\infty.$$

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

$$(4.8) \quad d_{TV} \left( \mathbb{P}_{\theta_{1,g}^1}, \mathbb{P}_{\theta_{1,g}^0} \right) + d_{TV} \left( \mathbb{P}_{\theta_{1,g}^1}, \mathbb{P}_{\theta_{1,g^0}^1} \right) \geq d_{TV} \left( \mathbb{P}_{\theta_{1,g}^0}, \mathbb{P}_{\theta_{1,g^0}^1} \right).$$

Now, let  $\tilde{f}$  be defined by  $\theta_1(\tilde{f}) = \theta_1(f)$ , and for any  $k \in \mathbb{Z} \setminus \{1\}$ ,  $\theta_k(\tilde{f}) = \theta_k(f^0)$ . Then Lemma 3.1 yields

$$d_{TV} \left( \mathbb{P}_{\theta_{1,g}^0}, \mathbb{P}_{\theta_{1,g}^1} \right) = d_{TV} \left( \mathbb{P}_{\tilde{f}, g}^1, \mathbb{P}_{f^0, g}^1 \right) \leq \frac{\|\tilde{f} - f^0\|}{\sqrt{2}} = \frac{|\theta_1 - \theta_1^0|}{\sqrt{2}}.$$

Therefore

$$(4.9) \quad \begin{aligned} \Pi_n \left( \mathbb{P}_{f,g} \text{ s.t. } d_{TV} \left( \mathbb{P}_{\theta_{1,g}^0}, \mathbb{P}_{\theta_{1,g}^1} \right) \leq \frac{M}{\sqrt{2}} \epsilon_n^{1/3} \middle| Y_1, \dots, Y_n \right) \\ \geq \Pi_n \left( \mathbb{P}_{f,g} \text{ s.t. } |\theta_1 - \theta_1^0| \leq M \epsilon_n^{1/3} \middle| Y_1, \dots, Y_n \right) \longrightarrow 1 \end{aligned}$$

as  $n \rightarrow +\infty$ . In conclusion, we deduce from (4.7),(4.8) and (4.9) that for  $M$  large enough:

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

We then use equation (4.5) applied with  $\theta_1 = \theta_1^0$  and the last equation to obtain our rate of consistency. Remark that

$$(4.10) \quad d_{TV}(\mathbb{P}_{\theta_1^1, g}^1, \mathbb{P}_{\theta_1^0, g^0}^1) \geq \frac{1}{8\pi^2} \sum_{n \in \mathbb{Z}} |c_n(g - g^0)|^2 \int_0^{+\infty} \rho e^{-(\rho + \theta_1^0)^2} A_n(2\rho\theta_1^0)^2 d\rho,$$

where we have used the definition

$$A_n(a) = \int_0^{2\pi} e^{a \cos(u)} \cos(nu) du.$$

Now, we use equivalents given by Lemma D.1 detailed in Appendix D. We only keep the integral of  $A_n$  for  $a \in [0, c\sqrt{n}]$  since it can be shown that the tail of such integral will yield negligible term. We just use the equivalent given by (D.1). One can find a sufficiently small constant  $\kappa$  such that

$$\begin{aligned} & \int_0^{+\infty} \rho e^{-(\rho + \theta_1^0)^2} A_n(2\rho\theta_1^0)^2 d\rho \\ & \geq \int_0^{\frac{\sqrt{n}}{2\theta_1^0}} \frac{4\pi^2 \rho^{2n+1} \{\theta_1^0\}^{2n}}{n!^2} e^{-(\rho + \theta_1^0)^2} \left( 1 - \kappa \frac{[2\rho\theta_1^0]}{n} \right)^2 d\rho \\ & \geq \left( 1 - \frac{\kappa}{\sqrt{n}} \right)^2 \frac{4\pi^2 \{\theta_1^0\}^{2n}}{n!^2} e^{-(\theta_1^0 + \frac{\sqrt{n}}{2\theta_1^0})^2} \int_0^{\frac{\sqrt{n}}{2\theta_1^0}} \rho^{2n+1} d\rho \end{aligned}$$

Now, we can apply the Stirling formula to obtain:

$$\begin{aligned} & \frac{4\pi^2 \{\theta_1^0\}^{2n}}{n!^2} e^{-(\theta_1^0 + \frac{\sqrt{n}}{2\theta_1^0})^2} \int_0^{\frac{\sqrt{n}}{2\theta_1^0}} \rho^{2n+1} d\rho \\ & \sim \frac{4\pi^2 \{\theta_1^0\}^{2n}}{(n/e)^{2n} 2\pi n} e^{-(\theta_1^0 + \frac{\sqrt{n}}{2\theta_1^0})^2} \frac{(\sqrt{n}/(2\theta_1^0))^{2n+2}}{2n+2} \\ & \sim \frac{2\pi}{n(2n+2)} e^{-2n \log \left[ \frac{n}{e\theta_1^0} \right] - (\theta_1^0 + \frac{\sqrt{n}}{2\theta_1^0})^2 + (n+1) \log \left[ \frac{n}{4\{\theta_1^0\}^2} \right]}. \end{aligned}$$

Hence, this last term is lower bounded by  $C(\theta_1^0)e^{-n \log(n)}$ . As a consequence, we can plug such lower bound in (4.10) to get

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

for  $c$  sufficiently small. We now end the proof of the Theorem: choose a frequency cut-off  $k_n$  that depends on  $n$  and remark that

$$\begin{aligned} \forall g \in \mathfrak{M}_\nu([0, 1])(A) \quad \|g - g^0\|^2 &= \sum_{|\ell| \leq k_n} |c_\ell(g - g^0)|^2 + \sum_{|\ell| > k_n} |c_\ell(g - g^0)|^2 \\ &\lesssim e^{ck_n \log k_n} \sum_{|\ell| \leq k_n} |c_\ell(g - g^0)|^2 e^{-c\ell \log \ell} + k_n^{-2\nu} \\ &\lesssim e^{ck_n \log k_n} d_{TV}(\mathbb{P}_{\theta_1^0, g}^1, \mathbb{P}_{\theta_1^0, g^0}^1) + k_n^{-2\nu}. \end{aligned}$$

We know from Equation (4.9) that the last bound is smaller 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

$$[k_n + 2\nu] \log k_n = \frac{1}{3} \log \frac{1}{\epsilon_n}.$$

This thus ensures that

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

□

4.2.4. *Posterior contraction rate around  $f^0$ .* We then aim to obtain a similar result for the posterior weight on neighbourhoods of  $f^0$ . Even if our results are quite good for the first coefficient  $\theta_1$ , we will see that indeed, this is far from being the case for the rest of its Fourier expansion.

**THEOREM 4.3.** *Assume  $(f^0, g^0) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0, 1])(A)$  and*

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

then

$$\Pi_n \left( f : \|f - f^0\|^2 > M (\log n)^{-2s \times \frac{2\nu}{2s+2\nu+1}} |Y_1, \dots, Y_n\right) \longrightarrow 0$$

in  $\mathbb{P}_{f^0, g^0}$  probability as  $n \rightarrow +\infty$ , for a sufficiently large  $M$ .

**PROOF.** The idea of the proof is very similar to the former used arguments, we aim to study the posterior weight on neighbourhoods of the true Fourier coefficients of  $f^0$ , whose frequency is larger than 1.

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

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

The second term does not exceed  $\epsilon_n \ll \log(n)^{-\nu}$  with a probability tending to 1, more precisely

$$(4.11) \quad \Pi_n \left( \forall k \in \mathbb{Z} \quad d_{TV}(\mathbb{P}_{f,g}^k, \mathbb{P}_{f^0,g^0}^k) < M\epsilon_n \mid Y_1, \dots, Y_n \right) \longrightarrow 1$$

as  $n \rightarrow +\infty$ .

*Point 2:*  $\Pi_n \left( \sup_{k \in \mathbb{Z}} d_{TV}(\mathbb{P}_{f,g^0}^k, \mathbb{P}_{f,g}^k) < M \log(n)^{-\nu} \mid Y_1, \dots, Y_n \right) \rightarrow 1$ . To obtain such a limit, we can use first the Cauchy-Schwarz inequality as follows:

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

Now, the Young inequality implies that for any  $M > 0$ ,

$$|z - \theta_k e^{ik\varphi}|^2 = |z|^2 + |\theta_k|^2 - 2\Re(\bar{z}\theta_k e^{ik\varphi}) \geq |z|^2 \left(1 - \frac{1}{M}\right) - |\theta_k|^2(M-1)$$

and the choice  $M = 2$  yields

$$(4.12) \quad d_{TV}(\mathbb{P}_{f,g^0}^k, \mathbb{P}_{f,g}^k) \leq \frac{\|g - g^0\|}{2\pi} \int_{\mathbb{C}} \left( e^{-|z|^2 + 2|\theta_k|^2} \right)^{1/2} dz \leq \|g - g^0\| \frac{e^{|\theta_k|^2}}{2}.$$

To obtain that the former term is bounded, we first establish that indeed the posterior distribution asymptotically only weights functions  $f$  with bounded Fourier coefficients. We hence denote

$$\mathcal{A}_n = \{(f, g) : \exists k \in \mathbb{Z} \quad d_{TV}(\mathbb{P}_{f,g^0}^k, \mathbb{P}_{f,g}^k) \geq M \log(n)^{-\nu}\}$$

and the two sets

$$\mathcal{B} = \{f : \forall k \in \mathbb{Z} \quad |\theta_k| \leq |\theta_k^0| + M \log(n)^{-\nu}\}$$

and

$$\mathcal{C} = \{f : \forall k \in \mathbb{Z} \quad |\theta_k^0| \leq |\theta_k| + M \log(n)^{-\nu}\}.$$

We first consider an integer  $k$  and  $\theta_k$  such that  $|\theta_k| > |\theta_k^0| + M \log(n)^{-\nu}$ , then

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

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

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

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

$$\begin{aligned} d_{TV}(\mathbb{P}_{f^0, g^0}^k, \mathbb{P}_{f, g}^k) &\geq \frac{1}{2\pi} \int_{B_n} \left| \int_0^{2\pi} \left[ e^{-|z - \theta_k e^{ik\varphi}|^2} g(\varphi) - e^{-|z - \theta_k^0 e^{ik\varphi}|^2} g^0(\varphi) \right] d\varphi \right| dz \\ &\geq \frac{1}{2\pi} \int_{B_n} e^{-[|\theta_k^0| + M \log(n)^{-\nu}/3]^2} - e^{-[|\theta_k^0| + 2M \log(n)^{-\nu}/3]^2} dz \\ &\geq c |\theta_k^0|^2 e^{-|\theta_k^0|^2} \log(n)^{-3\nu}, \end{aligned}$$

for a sufficiently small absolute constant  $c > 0$ . Since the sequence  $(\theta_k^0)_{k \in \mathbb{Z}}$  is bounded, for  $n$  large enough, we know that  $\|\theta^0\|^2 e^{-\inf_k |\theta_k^0|^2} \log(n)^{-3\nu} \gg \epsilon_n$ . We can deduce from (4.11) that

$$(4.13) \quad \Pi_n(\mathcal{B}^c | Y_1, \dots, Y_n) \longrightarrow 0 \quad \text{as } n \longrightarrow +\infty.$$

A similar argument yields

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

Gathering now (4.13) and (4.12), we get for a sufficiently large  $M$

$$\begin{aligned} \Pi_n(\mathcal{A}_n | Y_1, \dots, Y_n) &= \Pi_n(\mathcal{A}_n \cap \mathcal{B} \cap \mathcal{C} | Y_1, \dots, Y_n) \\ &\quad + \Pi_n(\mathcal{A}_n \cap (\mathcal{B} \cap \mathcal{C})^c | Y_1, \dots, Y_n) \\ &\leq \Pi_n\left(\|g - g^0\| \geq M e^{-(1 + \sup_k |\theta_k^0|^2)} \log(n)^{-\nu}\right) \\ &\quad + \Pi_n(\mathcal{B}^c | Y_1, \dots, Y_n) + \Pi_n(\mathcal{C}^c | Y_1, \dots, Y_n) \end{aligned}$$

We can now apply Theorem 4.2 to obtain the desired result:

$$(4.14) \quad \Pi_n\left(\sup_{k \in \mathbb{Z}} d_{TV}(\mathbb{P}_{f, g^0}^k, \mathbb{P}_{f, g}^k) < M \log(n)^{-\nu} \mid Y_1, \dots, Y_n\right) \longrightarrow 1 \quad \text{as } n \longrightarrow +\infty.$$

*Point 3: Contraction of  $\theta_k$  near  $\theta_k^0$ .* From the arguments of Point 2, we see that

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

We now study the situation when  $||\theta_k| - |\theta_k^0|| < M \log(n)^{-\nu}$ , and we can write  $\theta_k = \theta_k^0 e^{i\varphi} + \xi_n$  where  $\xi_n$  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,  $F(0) \simeq 0$  since a Taylor expansion near 0 yields at first order in  $z$  and  $\xi_n$  that

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

If one uses now  $\theta_k = \theta_k^0 e^{i\varphi} + \mathcal{O}(\log(n)^{-\nu})$ , the computation of the integral above yields for  $c < 2$  and  $\eta$  small enough such that  $|z| \leq \eta$ :

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

Now, denote  $\bar{u} = \frac{ie^{i\varphi/2} \bar{\theta}_k^0 c_{-k}(g^0)}{|\theta_k^0| \times |c_{-k}(g^0)|}$  which is a complex number of norm 1, and let  $v = \bar{u} e^{i\pi/2}$ . 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, we obtain that there exists an absolute constant  $c$  independent of  $k$  such that

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

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

$$(4.15) \quad d_{TV}(\mathbb{P}_{f,g^0}^k, \mathbb{P}_{f^0,g^0}^k) \geq c\eta^3 e^{-|\theta_k^0|^2} |c_{-k}(g^0)| |\theta_k - \theta_k^0| + \mathcal{O}(\log(n)^{-\nu}).$$

Thus, we can conclude using (4.14) and (4.15) that there exists a sufficiently large  $M$  such that, as  $n \longrightarrow +\infty$ ,

$$(4.16) \quad \Pi_n \left( f : \sup_{k \in \mathbb{Z}} |(\theta_k - \theta_k^0) c_{-k}(g^0)| < M \log(n)^{-\nu} \middle| Y_1, \dots, Y_n \right) \longrightarrow 1.$$

*Point 4: Contraction on  $f^0$ .* We can now produce a very similar proof to the one used at the end of Theorem 4.2:

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

Hence, (4.16) implies

$$\Pi_n \left( f : \|f - f^0\|^2 \leq k_n^{-2s} + k_n^{2\beta+1} \log(n)^{-2\nu} \mid Y_1, \dots, Y_n \right) \longrightarrow 1 \text{ as } n \longrightarrow +\infty.$$

The optimal choice of the frequency cut-off is  $k_n = (\log n)^{\frac{2\nu}{2\beta+2s+1}}$ , which yields

$$\Pi_n \left( f : \|f - f^0\|^2 \leq M (\log n)^{-4s\nu/(2s+2\beta+1)} \mid Y_1, \dots, Y_n \right) \longrightarrow 1 \text{ as } n \rightarrow +\infty.$$

This last result is the desired inequality.  $\square$

REMARK 4.1. *The lower bound obtained on  $d_{TV}(\mathbb{P}_{f,g^0}^k, \mathbb{P}_{f^0,g^0}^k)$  will be important to understand how one should build an appropriate net of functions  $(f_j, g_j) \in \mathcal{F}_s \times \mathfrak{M}_\nu([0,1])(A)$  hard to distinguish according to the  $L^2$  distance. When  $|\theta_k| \neq |\theta_k^0|$ , it is quite easy to distinguish the two hypotheses but it is far from being the case when their modulus is equal. In such a case, the behaviour of the Fourier coefficients of  $g^0$  becomes important. This is a clue to exhibit an efficient lower bound through the Fano lemma (for instance). This is detailed in the next paragraph.*

4.3. *Lower bound from a frequentist point of view.* We complete now our study of the Shape Invariant Model by a small investigation on how one could obtain some lower bounds in the frequentist paradigm. We could consider several methods. Among them, the first one could be the use of results in the literature, such as the works of [Mat02] or [BM05]. Indeed, in the convolution model with unknown variance

$$(4.17) \quad Y_i = X_i + \epsilon_i, \forall i \in \{1 \dots n\} \quad (X_i)_{i=1 \dots n} \sim g,$$

we already know that one cannot beat some  $\log n$  power for the convergence rate of any estimator of both  $g$  and of the variance of the noise  $\sigma^2$ . Such a nice result is obtained using the so-called van Trees inequality which is a Bayesian Cramer-Rao bound (see for instance [GL95] for further details). However their result cannot be used here: Proposition 4.1 is much more optimistic since we obtain there a polynomial rate for the posterior contraction around  $\theta_1^0$ .

First, note the results given by [Mat02] and Proposition 4.1 are not opposite. Indeed, [Mat02] considers lower bounds in a larger class than the estimation problem of  $\theta_1$  written as (4.6): from a minimax point of view, the supremum over all hypotheses is taken in a somewhat larger set than ours. Moreover, if one considers (4.6), 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 (4.17). Hence,  $g$  is a singular measure with respect to the noise measure: the ability of going beyond the logarithmic convergence rates is certainly due to this degeneracy nature of our problem according to the Gaussian complex noise. It is an important structural information which is not available when one considers general problems such as (4.17).

Following such considerations, we are thus driven to build some nets of hypotheses hard to distinguish between and then apply some classical tools for lower bound results. 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  $\mathcal{F}_s \times \mathfrak{M}_\nu([0, 1])(A)$  closed according to the Total Variation distance. We first recall the version of the Fano Lemma we used.

LEMMA 4.1 (Fano's Lemma). *Let  $r \geq 2$  be an integer and  $\mathcal{M}_r \subset \mathcal{P}$  which contains  $r$  probability distributions indexed by  $j = 1 \dots r$  such that*

$$\forall j \neq j' \quad d(\theta(P_j), \theta(P_{j'})) \geq \alpha_r,$$

and

$$d_{KL}(P_j, P_{j'}) \leq \beta_r.$$

Then, for any estimator  $\hat{\theta}$ , the following lower bound holds

$$\max_j \mathbb{E}_j \left[ d(\hat{\theta}, \theta(P_j)) \right] \geq \frac{\alpha_r}{2} \left( 1 - \frac{\beta_r + \log 2}{\log r} \right).$$

We derive now our lower bounds result.



THEOREM 4.4. *There exists a sufficiently small  $c$  such that the minimax rates of estimation over  $\mathcal{F}_s \times \mathfrak{M}_\nu([0,1])(A)$  satisfy*

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

and

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

PROOF. We will adapt the Fano Lemma to our setting and we are looking for a set  $(f_j, g_j)_{j=1 \dots p_n}$  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  $\theta_1(f_j) > 0$ . From Proposition 4.1, we know that one can easily distinguish two laws  $\mathbb{P}_{f_j, g_j}$  and  $\mathbb{P}_{f_{j'}, g_{j'}}$  as soon as  $\theta_1(f_j) \neq \theta_1(f_{j'})$ . Then we build our net using a common choice for the first Fourier coefficient of each  $f_j$  in our net. For instance, we impose that

$$\forall j \in \{1 \dots p_n\} \quad \theta_1(f_j) = 1.$$

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

$$(4.18) \quad \forall j \in \{1 \dots p_n\} \quad f_j(x) = e^{i2\pi x} + p_n^{-s} e^{i2\frac{(j-1)}{p_n}\pi} e^{i2\pi p_n x}.$$

The number of elements in the net  $p_n$  will be adjusted in the sequel and will grow to  $+\infty$ . Note that our construction naturally satisfies that each net  $(f_j)_{j=1 \dots p_n}$  belongs to  $\mathcal{F}_s$  since the modulus of the  $p_n$ -th Fourier coefficient is of size  $p_n^{-s}$ . At last, we have the following rather trivial inequality:  $\forall (j, j') \in \{1 \dots p_n\}^2$

$$\|f_j - f_{j'}\|^2 \geq p_n^{-2s} \times \left| e^{i2\pi/p_n} - 1 \right|^2 \geq 4p_n^{-2s} \sin^2(\pi/p_n) \sim_{n \rightarrow +\infty} 4\pi^2 p_n^{-2s-2}.$$

*Point 2: Net of measures  $(g_j)_{j=1 \dots p_n}$ .* The core of the lower bound is how to adjust the measures of the random shifts to make the distributions  $\mathbb{P}_{f_j, g_j}$ ,  $j = 1 \dots p_n$ , as close as possible. First, remark that the Fano Lemma 4.1 is formulated with entropy between laws although it is quite difficult to handle when dealing with mixtures. In the sequel, we will choose to still use the Total Variation distance, and then use the chain of inequalities:  $\forall j \neq j'$

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

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 (in which  $P_j = \mathbb{P}_{f_j, g_j}^{\otimes n}$ ). It imposes to find some mixture laws such that  $d_{TV} \left( \mathbb{P}_{f_j, g_j}, \mathbb{P}_{f_{j'}, g_{j'}} \right) \lesssim \frac{1}{(n \log n)^2}$ . From the triangular inequality, it is sufficient to build  $(g_j)_{j=1 \dots p_n}$  satisfying

$$(4.19) \quad \forall j \in \{1 \dots p_n\} \quad d_{TV} \left( \mathbb{P}_{f_j, g_j}, \mathbb{P}_{f_1, g_1} \right) \lesssim \frac{1}{(n \log n)^2}.$$

For sake of convenience, we will omit the dependence of  $p_n$  on  $n$  and simplify the notation to  $p$ . In a similar way,  $\theta_p^j$  will denote the  $p$ -th Fourier coefficient of  $f_j$  given by  $\theta_p^j = e^{i2\pi\alpha_j} \theta_p^1$  where  $\alpha_j = \frac{j-1}{p_n}$ . From our choice of  $(f_j)_{j=1 \dots p_n}$  given by (4.18), we have

$$\begin{aligned} d_{TV} \left( \mathbb{P}_{f_j, g_j}, \mathbb{P}_{f_1, g_1} \right) &= \frac{1}{2\pi^2} \int_{\mathbb{C} \times \mathbb{C}} \left| \int_0^1 e^{-|z_1 - e^{i2\pi\varphi}|^2 - |z_2 - e^{i2\pi p\varphi} \theta_p^1|^2} g_1(\varphi) d\varphi \right. \\ &\quad \left. - \int_0^1 e^{-|z_1 - e^{i2\pi\varphi}|^2 - |z_2 - e^{i2\pi p\varphi} \theta_p^j|^2} g_j(\varphi) d\varphi \right| dz_1 dz_2 \end{aligned}$$

We will use the smoothness of Gaussian densities to obtain a suitable upper bound. Call  $F$  the function 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  $\theta^j \bullet \varphi = (e^{i2\pi\varphi}, \theta_p^j e^{i2\pi p\varphi})$ .

To control  $F$ , we adapt the proof of Proposition 3.3. We use a truncature for  $(x_1, x_2, y_1, y_2) \in \mathcal{R}_{R_n} := B_{\mathbb{R}^2}(0, R_n)^2$ . Outside  $\mathcal{R}_{R_n}$ , we use the key inequality (that comes from a Taylor's expansion):

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

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 to ensure that

$$(4.20) \quad \forall m, l \leq d \quad \forall (s, \tilde{s}) \in \{-1; +1\}^2 \quad c_{sm + \tilde{s}lp}(g_j) e^{\tilde{s}l\alpha_j} = c_{sm + \tilde{s}lp}(g_1).$$

Hence, the maximum size of  $d$  is  $d = p/4$ . We have

$$\begin{aligned} d_{TV}(\mathbb{P}_{f_j, g_j}, \mathbb{P}_{f_1, g_1}) &= \frac{1}{2\pi^2} \int_{\mathcal{R}_{R_n}} |F(x_1, y_1, x_2, y_2)| dx_1 dy_1 dx_2 dy_2 \\ &\quad + \frac{1}{2\pi^2} \int_{\mathcal{R}_{R_n}^c} |F(x_1, y_1, x_2, y_2)| dx_1 dy_1 dx_2 dy_2 \\ &\lesssim e^{-R_n^2/2} + \left( \frac{(eR_n)^{p/4}}{(p/4)^{p/4}} \right)^4 \lesssim e^{-R_n^2/2} + \frac{(eR_n)^p}{(p/4)^p}. \end{aligned}$$

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 (4.19). 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

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

Such a choice of  $R_n$  and  $p_n$  ensures that (4.19) is fulfilled.

We have to make sure that our conditions (4.20) 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  $\forall k \in \mathbb{Z}^*$ ,  $c_k(g_1) := a|k|^{-\beta}$ . This ensures that

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

and therefore all  $g_j$  remains nonnegative.

Note that the densities  $g_j$  fulfill the condition appearing in 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 Lemma 4.1) with  $\alpha_n = p_n^{-2s-2}$  and  $\beta_n = \mathcal{O}(1)$  for the parametrization of  $(f_j)_{j=1 \dots p_n}$ . We then deduce the first lower bound

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

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

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

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

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

This ends the proof of the lower bound. □

**5. Concluding remarks.** In this paper, we exhibit a suitable prior which enable 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 unknown operator of translation which depends on  $g$ . From a technical point of view, the keystones of such results are the tight link between the white noise model and the Fourier expansion as well as the smoothness of Gaussian law which permits to obtain an efficient covering strategy.

Up to non restrictive condition, we can also obtain a large identifiability class but in this class, the contraction of the posterior is dramatically damaged since we then obtain a logarithm rate instead of a polynomial one. This last point cannot be so much improve using the standard  $L^2$  distance to measure the neighbourhoods of  $f^0$  as pointed by our last lower bound. Remark that we do not obtain exactly the same rates for our lower and upper bounds of reconstruction. This may be due to the rough inequality  $|\psi_a(\varphi)| \geq \frac{|\psi_a(\varphi)|^2}{\|\psi_a\|_\infty}$  used to obtain (4.1), it may be the reason why we do not obtain optimal rates.

Indeed, the degradation of the contraction rate occurs when one tries to invert the identifiability map  $\mathcal{I} : (f, g) \mapsto \mathbb{P}_{f, g}$ . Such difficulty should be understood as a novel consequence of the impossibility to exactly recover the random shifts parameters when only  $n$  grows to  $+\infty$ . Such phenomenon is highlighted for instance in several papers such as [BG10] or [BGKM12].

However, it may be possible to obtain a polynomial rate using a more appropriate distance adapted to our problem of randomly shifted curves as pointed in our main results:

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

We plan to tackle this problem in a future work. The important requirement in this view is to find some relations between the neighbourhoods of  $\mathbb{P}_{f^0, g^0}$  and the neighbourhoods of  $f^0$  according to the distance  $d_{Frechet}$ .

An interesting extension would consider the SIM with a noise level  $\sigma$  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  $[0, 1]$ .

At last, an open and challenging question concerns the research of stochastic algorithm to approach the posterior distribution in our non parametric Shape Invariant Model. One may think of an adaptation of the SAEM strategy proposed in [AKT10] even if this approach is at the moment valid only in a parametric setting.

#### APPENDIX A: TOPOLOGY ON PROBABILITY SPACES

*Probability distances.* We study consistency using standard distance over probability measures. If  $P$  and  $Q$  are two probability measures over a set  $X$ , absolutely continuous with respect to a reference measure  $\lambda$ ,  $d_H$  refers to the Hellinger distance defined as

$$d_H(P, Q) := \sqrt{\int_X \left[ \sqrt{\frac{dP}{d\lambda}} - \sqrt{\frac{dQ}{d\lambda}} \right]^2 d\lambda}.$$

Note that  $d_H$  does not depend on the choice of the dominating measure  $\lambda$ , and that the definition can be extended to any finite measures  $P$  and  $Q$  in a straightforward way.

When needed, we use the Total Variation distance between two probability measures  $P$  and  $Q$ . If  $\mathcal{B}$  is the  $\sigma$ -algebra of measurable sets with the reference measure  $\lambda$ , this distance is given by

$$d_{TV}(P, Q) := \sup_{A \in \mathcal{B}} |P(A) - Q(A)| = \frac{1}{2} \int_X \left| \frac{dP}{d\lambda} - \frac{dQ}{d\lambda} \right| d\lambda.$$

At last, we recall the definition of the Kullback-Leibler divergence (entropy) between  $P$  and  $Q$  since it is sometimes be used in the work:

$$d_{KL}(P, Q) := \int_X -\log \frac{dQ}{dP} dP.$$

In the sequel, we shall also use  $V(P, Q)$  defined as a second order moment associated to the Kullback-Leibler divergence

$$V(P, Q) := \int_X \left( \log \frac{dQ}{dP} \right)^2 dP.$$

It may be reminded the classical Pinsker's inequality

$$(A.1) \quad \sqrt{\frac{1}{2}d_{KL}(P, Q)} \geq d_{TV}(P, Q),$$

as well as

$$(A.2) \quad \frac{1}{2} d_H(P, Q)^2 \leq d_{TV}(P, Q) \leq d_H(P, Q).$$

*Model Complexity.* To obtain the posterior consistency and convergence rate, we shall use results given by Theorem 2.1 of [GGvdV00] which is stated below. This theorem exploits the notion of complexity of the studied model, and this complexity is traduced according to packing or covering numbers. For any set of probability measures  $\mathcal{P}$  endowed with a metric  $d$ ,  $D(\epsilon, \mathcal{P}, d)$  refers to the  $\epsilon$ -packing number (the maximum number of points in  $\mathcal{P}$  such that the minimal distance between each pair is larger than  $\epsilon$ ). The  $\epsilon$ -covering number  $N(\epsilon, \mathcal{P}, d)$  is the minimum number of balls of radius  $\epsilon$  needed to cover  $\mathcal{P}$ . These two numbers are linked through the following inequality

$$N(\epsilon, \mathcal{P}, d) \leq D(\epsilon, \mathcal{P}, d) \leq N(\epsilon/2, \mathcal{P}, d).$$

At last, for  $d$  a metric on finite measures, an  $\epsilon$ -bracket is a set of the form

$$[L, U] := \left\{ P \text{ s.t. } \frac{dL}{d\lambda} \leq \frac{dP}{d\lambda} \leq \frac{dU}{d\lambda} \right\},$$

for  $L$  and  $U$  two finite measures such that  $d(L, U) \leq \epsilon$  and  $\lambda$  any dominating measure. The  $\epsilon$ -bracketing number  $N_{[]}(\epsilon, \mathcal{P}, d)$  is the minimal number of  $\epsilon$ -brackets needed to cover  $\mathcal{P}$ . Note that  $N_{[]}(\epsilon, \mathcal{P}, d_H)$  is an upper bound of the  $(\epsilon/2)$ -covering number  $N(\epsilon/2, \mathcal{P}, d_H)$ . The bracketing entropy is then defined by  $H_{[]}(\epsilon, \mathcal{P}, d) := \log N_{[]}(\epsilon, \mathcal{P}, d)$ .

## APPENDIX B: TOOLS FOR THE PROOF OF THEOREM 2.2 AND THEOREM 2.3

### B.1. Entropy estimates.

PROOF OF PROPOSITION 3.1. The proof is similar to Lemma 1 of [GW00], we set  $p = 2\ell + 1$  and for any  $\epsilon > 0$ , we are going to build an explicit bracketing of  $\mathcal{A}_\theta$  and then bound  $N_{[]}(\epsilon, \mathcal{A}_\theta, d_H)$ . For an integer  $K$  which will be chosen in the sequel, we define  $[\varphi_-^i, \varphi_+^i]$  of size  $\Delta_\varphi = 1/K$ , with  $\varphi_-^i = (i-1)\Delta_\varphi$  and  $\varphi_+^i = i\Delta_\varphi$ . For any  $\delta > 0$ , we consider the lower and upper brackets

$$l_i := (1 + \delta)^{-1} \gamma_{\theta \bullet \varphi_-^i, (1+\delta)^{-\alpha} Id} \quad \text{and} \quad u_i := (1 + \delta) \gamma_{\theta \bullet \varphi_+^i, (1+\delta)^\alpha Id}.$$

We are looking for some admissible values of  $\alpha$ ,  $\delta$ , and  $K$  such that the set  $([l_i, u_i])_{i=1\dots K}$  is an  $\epsilon$ -bracket of  $\mathcal{A}_\theta$  for  $d_H$ . Of course, for all  $\varphi \in [\varphi_-^i, \varphi_+^i]$ ,  $l_i \leq \gamma_{\theta \bullet \varphi, Id}(\cdot) \leq u_i$  should hold, but we can check that  $\forall x \in \mathbb{C}$ ,

$$\frac{l_i(x)}{\gamma_{\theta \bullet \varphi, Id}(x)} \leq \frac{1}{1+\delta} \frac{1}{(1+\delta)^{-p\alpha}} e^{\frac{\|\theta \bullet \varphi - \theta \bullet \varphi_-^i\|^2}{1-(1+\delta)^{-\alpha}}} \leq (1+\delta)^{p\alpha-1} e^{\frac{4\pi^2 \Delta_\varphi^2 \|\theta\|_{\mathcal{H}_1}^2}{1-(1+\delta)^{-\alpha}}}.$$

Hence, we must have  $\alpha \leq 1/p$ , and we must also satisfy

$$|\Delta_\varphi|^2 \leq \frac{1-p\alpha}{4\pi^2 \|\theta\|_{\mathcal{H}_1}^2} (1 - (1+\delta)^{-\alpha}) \log(1+\delta) = \frac{\alpha(1-p\alpha)\delta^2}{4\pi^2 \|\theta\|_{\mathcal{H}_1}^2} (1 + o(1)),$$

where  $o(1)$  does not depend on  $p$  and goes to zero as  $\delta \rightarrow 0$  uniformly in  $\alpha$  in any positive neighbourhood of zero. In a same way considering  $\gamma_{\theta \bullet \varphi, Id} u_i^{-1}$ , we obtain

$$\forall x \in \mathbb{C} \quad \frac{\gamma_{\theta \bullet \varphi, Id}(x)}{u_i(x)} \leq (1+\delta)^{\alpha p-1} e^{\frac{4\pi^2 \Delta_\varphi^2 \|\theta\|_{\mathcal{H}_1}^2}{(1+\delta)^{\alpha-1}}},$$

and the same conditions arise. In order to minimize the cardinal of the bracketing,  $\Delta_\varphi$  must be as large as possible, we then maximize  $\alpha(1-p\alpha)$  and choose  $\alpha = (2p)^{-1}$ .

We must now check that  $d_H(l_i, u_i) \leq \epsilon$ . Rapid computations show that

$$d_H(l_i, u_i)^2 = \delta^2 + d_H(\gamma_{\theta \bullet \varphi_-^i, (1+\delta)^{-\alpha} Id}(\cdot), \gamma_{\theta \bullet \varphi_+^i, (1+\delta)^\alpha Id}(\cdot))^2.$$

Using standard formula on Hellinger distance for multivariate gaussian laws, we obtain

$$\begin{aligned} d_H(l_i, u_i)^2 &= \delta^2 + 2 \left[ 1 - \frac{2^p}{((1+\delta)^\alpha + (1+\delta)^{-\alpha})^p} \right] \\ &= \delta^2 + 2 \left[ 1 - \frac{2^p \sqrt{1+\delta}}{(1+(1+\delta)^{1/p})^p} \right]. \end{aligned}$$

One can easily check that, whatever  $p \geq 1$ ,  $(1+(1+\delta)^{1/p})^p \leq 2^p e^{\delta/2}$ , which yields

$$d_H(l_i, u_i)^2 \leq \frac{3}{2} \delta^2 + o(\delta^2) \leq 2\delta^2$$

for  $\delta$  small enough. An admissible choice of  $\delta$  should be  $\delta = \epsilon/\sqrt{2}$ , which insures  $d_H(l_i, u_i) \leq \epsilon$ . We then obtain

$$\Delta_\varphi^2 \leq \frac{\delta^2 + o(\delta^2)}{16\pi^2 p \|\theta\|_{\mathcal{H}_1}^2} = \frac{\epsilon^2 + o(\epsilon^2)}{32\pi^2 p \|\theta\|_{\mathcal{H}_1}^2},$$

where  $o(\epsilon^2)$  does not depend on  $p$ . The number of brackets is now  $K = \Delta_\varphi^{-1}$ , this ends the proof of the proposition.  $\square$

PROOF OF PROPOSITION 3.3. We first fix the notation  $p = 2\ell + 1$  which refers to the dimension of the multivariate mixture. For any  $R > 0$  which will be chosen later,  $\mathcal{E}_R$  is the ball of in  $\mathbb{C}^p$  of radius  $R$ . For sake of simplicity, we will sometimes omit the dependence on  $\epsilon$  with the notation  $p$ . According to the hypotheses in Proposition 3.3, there exists an absolute constant  $a$  such that  $\|\theta\| \leq w \leq a\sqrt{p}$ . We first write

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

Let  $\nu$  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,

$$\forall \varphi \in [0, 1] \quad \forall z \in \mathcal{E}_R^c \quad \|z - \theta \bullet \varphi\| > \|z\|/(1+a).$$

This simply implies that,

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

To deal with we last term we use a concentration of chi-square statistics inequality (see Lemma 1 of [IL06]): for any  $k \geq 1$  and  $c > 0$ ,

$$(B.1) \quad \mathbb{P}(\chi_k^2 \geq (1+c)k) \leq \frac{1}{c\sqrt{2\pi}} e^{-\frac{k}{2}[c - \log(1+c)] - \frac{1}{2} \log k}.$$

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 then consider (B), following the strategy of [GvdV01] which exploits the smoothness of Gaussian densities. We will exhibit a discrete mixture law which will be close to  $\mathbb{P}_{\theta,g}$ , for any given  $g$ . Taylor's expansion theorem yields:

$$(B.2) \quad \forall k \in \mathbb{N} \quad \forall y \in \mathbb{R}_+ \quad \underbrace{\left| e^{-y} - \sum_{j=0}^{k-1} \frac{(-y)^j}{j!} \right|}_{:= R_k(y)} \leq \frac{|y|^k}{k!} \leq \frac{(e|y|)^k}{k^k}.$$



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

$$\begin{aligned} \mathbb{P}_{\theta,g}(z) - \mathbb{P}_{\theta,\tilde{g}}(z) &= \pi^{-p} \int_0^1 e^{-\|z - \theta \bullet \varphi\|^2} \left[ \frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi) \\ &= \pi^{-p} \sum_{j=0}^{k-1} \frac{(-1)^j}{j!} \int_0^1 \|z - \theta \bullet \varphi\|^{2j} \left[ \frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi) \\ &\quad + \pi^{-p} \int_0^1 R_k(\|z - \theta \bullet \varphi\|^2) \left[ \frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi). \end{aligned}$$

We now decompose  $\theta = (\theta_{-\ell}, \dots, \theta_\ell)$  and  $z = (z_{-\ell}, \dots, z_\ell)$  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

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

where  $(a(r, m))_{r=1 \dots k, m=-\ell \dots \ell}$  is a complex matrix which only depends on  $z$  and  $\theta$ . 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 by the Binomial formula and coefficients  $a_{r,m}(z, \theta)$ . Consequently, for all  $z \in \mathcal{E}_R$

$$\begin{aligned} (\mathbb{P}_{\theta,g} - \mathbb{P}_{\theta,\tilde{g}})(z) &= \pi^{-p} \sum_{j=0}^{k-1} \frac{(-1)^j}{j!} \int_0^1 \left[ C_j(z, \theta) \right. \\ &\quad \left. + \sum_{r=-j\ell}^{j\ell} b_r(z, \theta) e^{ir\varphi} \right] \left[ \frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi) \\ &\quad + \pi^{-p} \int_0^1 R_k(\|z - \theta \bullet \varphi\|^2) \left[ \frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi) \\ &= \pi^{-p} \sum_{j=0}^{k-1} \frac{(-1)^j}{j!} \left[ C_j(z, \theta) c_0(g - \tilde{g}) + \sum_{r=-j\ell}^{j\ell} b_r(z, \theta) c_r(g - \tilde{g}) \right] \\ &\quad + \pi^{-p} \int_0^1 R_k(\|z - \theta \bullet \varphi\|^2) \left[ \frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi). \end{aligned}$$

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

$$\forall r \in [-(k-1)\ell, (k-1)\ell] \quad c_r(g) = c_r(\tilde{g}).$$

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

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

and of course

$$\begin{aligned} (B) &\leq \pi^{-p} \int_{\mathcal{E}_R} \left| \int_0^1 R_k(\|z - \theta \bullet \varphi\|^2) \left[ \frac{dg}{d\nu}(\varphi) - \frac{d\tilde{g}}{d\nu}(\varphi) \right] d\nu(\varphi) \right| dz \\ &\leq 2\pi^{-p} \sup_{z \in \mathcal{E}_R, \varphi \in (0,1)} R_k(\|z - \theta \bullet \varphi\|^2) \text{Vol}(\mathcal{E}_R). \end{aligned}$$

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

$$\begin{aligned} (B) &\lesssim \pi^{-p} \frac{(e(1+2a)^2(1+c)p)^k}{k^k} \frac{\pi^p [(1+a)^2(1+c)p]^p}{p!} \\ &\lesssim C_1^p C_2^k e^{-k \log(k) + k \log(p)}, \end{aligned}$$

where we used in the last equation  $p^p/p! \leq C^p$ . If we define the threshold  $k$  in (B.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 thus choose  $k_\epsilon \sim b\ell_\epsilon$  for a sufficiently large absolute constant  $b$ . For such a choice, 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  $s_\epsilon$  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 3.3 comes from Proposition 3.2.

The second inequality in Proposition 3.3 is proved from the first one, using the relation  $\|\theta\|_{\mathcal{H}_1} \leq \ell\|\theta\|$  valid for any  $f \in \mathcal{H}^\ell$ .  $\square$

PROOF OF LEMMA 3.1. We follow a straightforward argument:  $\mathbb{P}_{f,g}$  is a mixture model so

$$\mathbb{P}_{f,g} = \int_0^1 \mathbb{P}_{f,\delta_\alpha} dg(\alpha).$$

Thus

$$\begin{aligned} d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{\tilde{f},g}) &= \left\| \int_0^1 (\mathbb{P}_{f,\delta_\alpha} - \mathbb{P}_{\tilde{f},\delta_\alpha}) dg(\alpha) \right\|_{TV} \\ &\leq \int_0^1 \left\| \mathbb{P}_{f,\delta_\alpha} - \mathbb{P}_{\tilde{f},\delta_\alpha} \right\|_{TV} dg(\alpha) \\ &= \left\| \mathbb{P}_{f,\delta_0} - \mathbb{P}_{\tilde{f},\delta_0} \right\|_{TV} \leq d_H(\mathbb{P}_{f,\delta_0}, \mathbb{P}_{\tilde{f},\delta_0}). \end{aligned}$$

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

$$\begin{aligned} d_H^2(\mathbb{P}_{f,\delta_0}, \mathbb{P}_{\tilde{f},\delta_0}) &= 2 \left( 1 - \mathbb{E}_{f,\delta_0} \sqrt{\frac{d\mathbb{P}_{\tilde{f},\delta_0}}{d\mathbb{P}_{f,\delta_0}}(Y)} \right) \\ &= 2 \left( 1 - \mathbb{E}_{f,\delta_0} \sqrt{\exp(2\Re\langle \tilde{f} - f, dW \rangle - \|\tilde{f} - f\|^2)} \right) \\ &= 2 \left( 1 - \exp\left(\frac{-\|\tilde{f} - f\|^2}{2}\right) \mathbb{E}_U \left[ \exp(\|\tilde{f} - f\|\Re(U)) \right] \right) \\ &= 2 \left( 1 - \exp\left(\frac{-\|\tilde{f} - f\|^2}{4}\right) \right) \leq \frac{\|\tilde{f} - f\|^2}{2}. \end{aligned}$$

□

## B.2. Link between Kullback-Leibler and Hellinger neighbourhoods.

PROOF OF PROPOSITION 3.4. This proposition uses a corollary of Rice's formula (see [AW09] for various applications of such formula), stated in Lemma B.1 and postponed after this proof.

We begin with Girsanov's formula (3.1). Write now  $Y = f^{0,-\tau} + W$  where  $W$  stands for a complex standard Brownian motion independent of the ran-

dom shift  $\tau$  (whose law is  $g^0$ ). The  $L^2$  norm is invariant with any shift thus

$$\begin{aligned} \frac{d\mathbb{P}_{f^0, g^0}}{d\mathbb{P}_{f, g}}(Y) &= \exp(\|f\|^2 - \|f^0\|^2) \frac{\int_0^1 e^{2\Re\langle f^{0, -\alpha_1}, f^{0, -\tau} + dW \rangle} dg^0(\alpha_1)}{\int_0^1 e^{2\Re\langle f^{-\alpha_2}, f^{0, -\tau} + dW \rangle} dg(\alpha_2)} \\ &\leq \exp(\|f\|^2 - \|f^0\|^2) \exp\left(2 \sup_{\alpha_1, \alpha_2} \Re\langle f^{0, -\alpha_1} - f^{-\alpha_2}, f^{0, -\tau} \rangle\right) \\ &\quad \exp\left(2 \sup_{\alpha_1, \alpha_2} \Re\langle f^{0, -\alpha_1} - f^{-\alpha_2}, dW \rangle\right) \\ &\leq e^{(\|f\| + \|f^0\|)^2} e^{Z_1 + Z_2}, \end{aligned}$$

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

$$\begin{aligned} Z_1 &:= 2 \sup_{\alpha_1} \Re\langle f^{0, -\alpha_1}, dW \rangle = 2 \sup_{\alpha_1} \Re \int_0^1 \overline{f^0}(s - \alpha_1) dW_s, \\ Z_2 &:= 2 \sup_{\alpha_2} \Re\langle -f^{-\alpha_2}, dW \rangle = 2 \sup_{\alpha_2} \Re \int_0^1 -\overline{f}(s - \alpha_2) dW_s. \end{aligned}$$

We now set  $\delta \in (0, 1]$  (it will be precisely fixed in the sequel) and we define the trajectories  $\mathcal{E}_\delta$  as

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

Hence, following the definition of  $M_\delta^2$  of (3.2), 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  $\delta$  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 B.1 yields

$$\begin{aligned}
M_\delta^2 &\leq e^{\delta(\|f\|+\|f^0\|)^2} \int_{\sqrt{e}}^{+\infty} \mathbb{P}\left(e^{\delta(Z_1+Z_2)} > u\right) du \\
&= e^{\delta(\|f\|+\|f^0\|)^2} \int_{\sqrt{e}}^{+\infty} \left[ \mathbb{P}\left(\frac{Z_1}{2} \geq \frac{\log u}{4\delta}\right) + \mathbb{P}\left(\frac{Z_2}{2} \geq \frac{\log u}{4\delta}\right) \right] du \\
\text{(B.3)} \quad M_\delta^2 &\leq C(f^0, f) e^{\delta(\|f\|+\|f^0\|)^2} \int_{\sqrt{e}}^{+\infty} \left[ e^{-\frac{\log^2(u)}{16\delta^2\|f^0\|^2}} + e^{-\frac{\log^2(u)}{16\delta^2\|f\|^2}} \right] du.
\end{aligned}$$

Now, we can choose  $\delta$  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  $\delta^2 < \frac{1}{32\|f^0\|^2}$ , and the same holds with  $f$  instead of  $f_0$ . Note that  $M_\delta^2$  is uniformly bounded if  $f$  is picked into a ball centered at 0 with radius  $2\|f^0\|$ .  $\square$

We now show that the technical inequality used in (B.3) is satisfied.

LEMMA B.1. *Let  $W$  a complex standard Brownian motion and  $u$  a complex 1-periodic map of  $\mathcal{H}_s$ . We assume that  $u$  is of class  $\mathcal{C}^2$ . Then when  $t/\|u\| \rightarrow +\infty$ , 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 process

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

$X$  is a Gaussian centered process. Its 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  $\Gamma$  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 for instance exercise 4.2, chapter 4 of [AW09]) then yields that when  $t \rightarrow +\infty$ , we have

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

This ends the proof of the first inequality. Assume furthermore that  $u \in \mathcal{H}^\ell$ , Parseval's equality implies that  $\|u'\| \leq \ell\|u\|$  and we obtain the second inequality.  $\square$

### B.3. Hellinger neighbourhoods.

PROOF OF PROPOSITION 3.5. Recall first that if  $Y$  follows  $\mathbb{P}_{f^0, g^0}$ , one shift  $\beta$  is randomly sampled according to  $g^0$ . Conditionally to this shift  $\beta$ ,  $Y$  is described through a white noise model  $dY(x) = f^0(x - \beta)dx + dW(x)$ . 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  $\beta$ , 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

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

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  $\ell_n$  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  $\beta$ ,  $D_\beta(\alpha)$  is measurable with respect to the filtration associated to  $W_1$ , and  $X_\beta(\alpha)$  is independent of  $W_1$ . We thus obtain

using Jensen's inequality and this filtration property that

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

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  $\beta$ .

Now, one should remark that  $X_\beta(\alpha)$  has the same law as

$$\exp \left( 2\Re \langle (f^0 - f_{\ell_n}^0)^{-\alpha}, f^{0,-\beta} \rangle + U \right),$$

where  $U \sim \mathcal{N}_\mathbb{R}(-\|f^0 - f_{\ell_n}^0\|^2, 2\|f^0 - f_{\ell_n}^0\|^2)$ , and  $\mathbb{E}[e^U] = 1$ . Hence

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

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

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

Again, we can use the orthogonal decomposition  $f^{0,-\beta} = f_{\ell_n}^{0,-\beta} + f^{0,-\beta} - f_{\ell_n}^{0,-\beta}$  and Cauchy-Schwarz's inequality yields  $(\tilde{E}_1) \leq \sqrt{2} \|f^0 - f_{\ell_n}^0\|$ .

Note that until now we did not use the hypothesis  $f^0 \in \mathcal{H}_s$ . It is only needed to get the last inequality in Proposition 3.5.  $\square$

To establish Lemma 3.2, we first remind the following useful result.

**LEMMA B.2.** *For any 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  $\Phi$  stands for the cumulative distribution function of a real standard Gaussian variable.

PROOF OF LEMMA 3.2. Adapting the proof of Lemma 5.1 of [GvdV01],

$$\begin{aligned} \|\mathbb{P}_{f_{\ell_n}^0, \tilde{g}} - \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}\|_{L_1} &\leq \sum_{j=1}^J \int_{\varphi_j - \eta/2}^{\varphi_j + \eta/2} \|\gamma_0(\cdot - \theta \bullet \varphi) - \gamma_0(\cdot - \theta \bullet \varphi_j)\|_{L_1} d\tilde{g}(\varphi) \\ &\quad + 2 \sum_{j=1}^J |\tilde{g}([\varphi_j - \eta/2, \varphi_j + \eta/2]) - p_j|. \end{aligned}$$

Using Lemma B.2 ends the proof.  $\square$

PROOF OF PROPOSITION 3.7. The construction used in the proof of Proposition 3.3 provide a mixture  $\tilde{g}$  such that  $\tilde{g}$  is supported by  $\tilde{J}_n := C\ell_n^2$  points (denoted  $(\varphi_j)_{j=1 \dots \tilde{J}_n}$ ) so that  $d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}) \leq \epsilon_n$ . Therefore  $\tilde{g} = \sum_{j=1}^{\tilde{J}_n} w_j \delta_{\varphi_j}$ . As pointed by [GvdV01], one can slightly modify  $\tilde{g}$  so that the support points are separated enough as follows. First, denote  $(\psi_j)_{j=1 \dots J_n}$  the subset of  $(\varphi_j)_{j=1 \dots \tilde{J}_n}$  which is  $\eta_n$ -separated with a maximal number of elements. Hence,  $J_n \leq \tilde{J}_n$  and up to a permutation, one can divide  $(\varphi_j)_{j=1 \dots \tilde{J}_n}$  in two parts:  $(\varphi_j)_{j=1 \dots \tilde{J}_n} = (\psi_j)_{j=1 \dots J_n} \cup (\varphi_j)_{j=J_n+1 \dots \tilde{J}_n}$ . For any  $i \in \{J_n + 1, \dots, \tilde{J}_n\}$ , we define  $\psi_{j(i)}$  as the closest point of  $(\psi_j)_{j=1 \dots J_n}$ , 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,  $\tilde{g}$  as a support which is  $\eta_n$ -separated. Moreover, we have

$$\begin{aligned} 2d_{TV} \left( \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}} \right) &= \int_{\mathbb{C}^{\ell_n}} \left| \sum_{i=1}^{J_n} \tilde{w}_i \gamma(z - \theta \bullet \psi_i) - \sum_{i=1}^{\tilde{J}_n} w_i \gamma(z - \theta \bullet \varphi_i) \right| dz \\ &= \int_{\mathbb{C}^{\ell_n}} \left| \sum_{j=1}^{J_n} (\tilde{w}_j - w_j) \gamma(z - \theta \bullet \psi_j) - \sum_{i > J_n} w_i \gamma(z - \theta \bullet \varphi_i) \right| dz \\ &= \int_{\mathbb{C}^{\ell_n}} \left| \sum_{j=1}^{J_n} \sum_{i > J_n | j(i)=j} w_i [\gamma(z - \theta \bullet \psi_j) - \gamma(z - \theta \bullet \varphi_i)] \right| dz. \end{aligned}$$



Then, Fubini's theorem yields

$$d_{TV} \left( \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}, \mathbb{P}_{f_{\ell_n}^0, \tilde{\tilde{g}}} \right) \leq \sum_{j=1}^{J_n} \sum_{i > J_n | j(i)=j} w_i d_{TV}(\gamma_{\theta \bullet \varphi_i}, \gamma_{\theta \bullet \psi_j}),$$

and we deduce from Lemma B.2 that

$$d_{TV} \left( \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}, \mathbb{P}_{f_{\ell_n}^0, \tilde{\tilde{g}}} \right) \leq \sqrt{2\pi} \sum_{j=1}^{J_n} \sum_{i > J_n | j(i)=j} w_i \|\theta\|_{\mathcal{H}_1} \eta_n \leq \sqrt{2\pi} \|\theta\|_{\mathcal{H}_1} \eta_n.$$

Now the relations between Hellinger and Total Variation distances (A.2) yield

$$d_H(\mathbb{P}_{f_{\ell_n}^0, g^0}, \mathbb{P}_{f_{\ell_n}^0, \tilde{g}}) \leq \epsilon_n + d_H(\mathbb{P}_{f_{\ell_n}^0, \tilde{g}}, \mathbb{P}_{f_{\ell_n}^0, \tilde{\tilde{g}}}) \leq \left(1 + (8\pi)^{1/4} \|\theta\|_{\mathcal{H}_1}^{1/2}\right) \epsilon_n.$$

Lemma 3.2 permits to conclude.  $\square$

#### B.4. Checking the conditions of Theorem 2.1.

PROOF OF PROPOSITION 3.9. We have seen in the proof of Proposition 3.4 that  $M_\delta^2$  is uniformly bounded with respect to  $\|f\|$  and  $\|f^0\|$  for a suitable choice of  $\delta$ . We restrict our study to the elements  $f$  such that  $\|f\| \leq 2\|f^0\|$ . We know from Proposition 3.4 and Theorem 3.2 that as soon as  $\tilde{\epsilon}_n \log \frac{1}{\epsilon_n} \leq c\epsilon_n$  with  $c$  small enough:

$$\begin{aligned} \mathcal{V}_{\tilde{\epsilon}_n}(\mathbb{P}_{f^0, g^0}, d_H) &:= \left\{ \mathbb{P}_{f, g} \in \mathcal{P} \mid d_H(\mathbb{P}_{f^0, g^0}, \mathbb{P}_{f, g}) \leq \tilde{\epsilon}_n \text{ and } \|f\| \leq 2\|f^0\| \right\} \\ &\subset \mathcal{V}_{\tilde{\epsilon}_n}(\mathbb{P}_{f^0, g^0}, d_{KL}). \end{aligned}$$

This last condition on  $\tilde{\epsilon}_n$  is true as soon as

$$(B.4) \quad \tilde{\epsilon}_n := \tilde{c}\epsilon_n \left( \log \frac{1}{\epsilon_n} \right)^{-1}$$

with  $\tilde{c}$  small enough. Now, Proposition 3.8 permits to describe a subset of  $\mathcal{V}_{\tilde{\epsilon}_n}(\mathbb{P}_{f^0, g^0}, d_H)$ , by the definition of 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 lower bound given by Lemma 3.3. The prior for  $g$  is a Dirichlet process with a finite base measure  $\alpha$  admitting a continuous positive density on  $[0, 1]$ . Since  $\eta_n$  goes to zero, for  $n$  large enough  $\alpha(\psi_j - \eta_n/2, \psi_j + \eta_n/2)$  for any  $j = 1 \dots J_n$ . Note that  $J_n \lesssim \ell_n^2 = \tilde{\epsilon}_n^{-2/s} \leq \tilde{\epsilon}_n^{-2}$ . Thus, there exists an absolute constant

$a \in (0, 1]$  such that the condition  $J_n \leq 2(a\tilde{\epsilon}_n)^{-2}$  is fulfilled, and one can find universal constants  $C$  and  $c$  such that for  $n$  large enough

$$(B.5) \quad \Pi_n(\mathcal{G}_{\tilde{\epsilon}_n}) \geq \Pi_n(\mathcal{G}_{a\tilde{\epsilon}_n}) \geq Ce^{-cJ_n \log \frac{1}{\tilde{\epsilon}_n}} \geq Ce^{-c\ell_n^2 \log \frac{1}{\tilde{\epsilon}_n}}.$$

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. Remark that from the construction of our prior on  $f$ , one has

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

From our assumption on the prior  $\lambda$ , 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

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

For  $n$  large enough we get

$$(B.6) \quad \begin{aligned} \Pi_n(\mathcal{F}_{\tilde{\epsilon}_n}) &\geq \exp - [c\ell_n^2 \log^\rho \ell_n + \xi_n^{-2} \\ &\quad + (2\ell_n + 1) (\log \ell_n + 4 \log(1/\tilde{\epsilon}_n) - \log \xi_n^{-2} + \mathcal{O}(1))] \\ &\geq \exp [-(c + o(1)) [\ell_n^2 \log^\rho \ell_n \vee \xi_n^{-2}]] \end{aligned}$$

Gathering (B.5) and (B.6), the relations  $\ell_n = \tilde{\epsilon}_n^{-1/s}$  and (B.4) lead to

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

for constants  $c > 0$ .  $\square$

**PROOF OF PROPOSITION 3.10.** The upper bound on the packing number comes directly from Theorem 3.1 since we set  $w_n = \sqrt{2k_n + 1}$ .

Now, to control the prior mass outside the sieve, remark first that owing to the construction of our prior, we have

$$(B.7) \quad \Pi_n(\mathcal{P} \setminus \mathcal{P}_{k_n, w_n}) \leq \sum_{|k| \geq k_n} \lambda(k) + \Pr \left( \sum_{|k| \leq k_n} |\theta_k|^2 \geq w_n^2 \right),$$

where each  $\theta_k$  for  $-k_n \leq k \leq k_n$  follows a centered Gaussian law of variance  $\xi_n^2$ . 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 (B.7), we use (B.1) to get

$$\begin{aligned} \Pr \left( \sum_{|k| \leq k_n} |\theta_k|^2 \geq w_n^2 \right) &= \Pr \left( \sum_{|k| \leq k_n} \left| \frac{\theta_k}{\xi_n} \right|^2 \xi_n^2 \geq w_n^2 \right) \\ &\leq \mathbb{P} \left( \chi_{2k_n+1}^2 \geq 2(2k_n+1)\xi_n^{-2} \right) \\ &\leq \frac{1}{(\xi_n^{-2}-1)\sqrt{\pi}} e^{-(2k_n+1)[\xi_n^{-2}-1-\log \xi_n^{-2}]-\log(2k_n+1)/2}. \end{aligned}$$

Now, using the value of  $\xi_n$ , we obtain

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

This concludes the proof of the Proposition. □

### B.5. Bound of $d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{f,\tilde{g}})$ with the Wasserstein metric.

PROOF OF PROPOSITION 3.11. We use a change of variable, the convexity of  $d_{TV}$ , and Lemma 3.1 to get

$$\begin{aligned} d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{f,\tilde{g}}) &= \frac{1}{2} \left\| \int_0^1 \mathbb{P}_{f,\delta_\alpha} dg(\alpha) - \int_0^1 \mathbb{P}_{f,\delta_\alpha} d\tilde{g}(\alpha) \right\|_{L^1} \\ &= \frac{1}{2} \left\| \int_0^1 \left( \mathbb{P}_{f,\delta_{G^{-1}(u)}} - \mathbb{P}_{f,\delta_{\tilde{G}^{-1}(u)}} \right) du \right\|_{L^1} \\ &\leq \int_0^1 d_{TV} \left( \mathbb{P}_{f-G^{-1}(u), \delta_0}, \mathbb{P}_{f-\tilde{G}^{-1}(u), \delta_0} \right) du \\ &\leq \frac{1}{\sqrt{2}} \int_0^1 \left\| f^{-G^{-1}(u)} - f^{-\tilde{G}^{-1}(u)} \right\| du. \end{aligned}$$

Then

$$\begin{aligned} \left\| f^{-G^{-1}(u)} - f^{-\tilde{G}^{-1}(u)} \right\| &= \sqrt{\sum_{k \in \mathbb{Z}} |c_k(f)|^2 \left| e^{-i2\pi k G^{-1}(u)} - e^{-i2\pi k \tilde{G}^{-1}(u)} \right|^2} \\ &\leq 2\pi \left| G^{-1}(u) - \tilde{G}^{-1}(u) \right| \sqrt{\sum_{k \in \mathbb{Z}} k^2 |c_k(f)|^2}. \end{aligned}$$

Therefore we get the first inequality:

$$d_{TV}(\mathbb{P}_{f,g}, \mathbb{P}_{f,\tilde{g}}) \leq \sqrt{2\pi} \|f\|_{\mathcal{H}_1} \int_0^1 \left| G^{-1}(u) - \tilde{G}^{-1}(u) \right| du.$$

Now, the second inequality is a classical result: see for instance [GS02, Theorem 4]. The last inequality is well known too.  $\square$

### APPENDIX C: SMALL BALL PROBABILITY FOR INTEGRATED BROWNIAN BRIDGE

In the sequel, we still use the notation  $p_\nu$  defined by (2.9) to refer to the probability distribution which is proportionnal to  $e^\nu$ . We detail here how one can obtain a lower bound of the prior weight around any element  $g^0$ . Since we deal with a log density model, it will be enough to find a lower bound of the weight around  $w^0$  if one writes  $g^0 \propto e^{w^0}$  according to Lemma C.1 (which is the Lemma 3.1 of [vdVvZ08a]).

LEMMA C.1 ([vdVvZ08a]). *For any real and measurable functions  $v$  and  $w$  of  $[0, 1]$ , the Hellinger distance between  $p_v$  and  $p_w$  is bounded by*

$$d_H(p_v, p_w) \leq \|v - w\|_\infty e^{\|v-w\|_\infty/2}.$$

We now obtain a lower bound of the prior weight on the set  $\mathcal{G}_\epsilon$  previously defined as:

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

This bound is given by the following Theorem.

THEOREM C.1. *The prior  $q_{\nu,A}$  defined by (2.8) and (2.9) satisfies for  $\epsilon$  small enough:*

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

where  $c$  is a constant which does not depend on  $\epsilon$ .

PROOF. The proof is divided in 4 steps.

*Structure of the prior.* We denote  $w_0 := \log g^0$ , which is a  $k_\nu$ -differentiable function of  $[0, 1]$ , that can be extended to a 1-periodic element of  $\mathcal{C}^{k_\nu}(\mathbb{R})$ . We define  $\tilde{q}$  the prior defined by (2.8) on such a class of periodic functions (and omit the dependence on  $\nu$  and  $A$  for sake of simplicity). The prior  $q_{\nu,A}$  is then derived from  $\tilde{q}$  through (2.9). We can remark that our situation looks similar to the one described in paragraph 4.1 of [vdVvZ08a] for integrated

brownian motion. Indeed, the log-density  $w_0$  should be approximated by some "Brownian bridge started at random" using

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

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  $[0, 1]$ . Of course, in the above equation, one can immediately check that  $J_{k_\nu}(B)(0) = J_{k_\nu}(B)(1) = 0$ . Moreover, the relation  $J_k(f)' = J_{k-1}(f) - \int_0^1 J_{k-1}(f)$  and an induction argument yields

$$\forall j \in \{1, \dots, k_\nu\} \quad J_{k_\nu}(B)^{(j)}(0) = J_{k_\nu}(B)^{(j)}(1).$$

Hence,  $J_{k_\nu}(B)$  and its first  $k_\nu$  derivatives are 1-periodic. Of course, the functions  $\psi_i$  are also 1-periodic and  $\mathcal{C}^\infty(\mathbb{R})$  and thus our prior  $\tilde{q}$  generates admissible functions of  $[0, 1]$  to approximate  $w_0$ . We will denote this set of admissible trajectories  $\mathcal{C}_1^{k_\nu}$  to refer to 1-periodic functions which are  $k_\nu$  times differentiable.

*Transformation of the Brownian bridge.* We denote  $\mathbb{B}_1$  the separable Banach space of Brownian bridge trajectories between 0 and 1 and  $\mathbb{B}_2 = \mathbb{R}^{k_\nu+1}$ . It is possible to check that the map

$$T : (B, Z_0, \dots, Z_{k_\nu}) \mapsto J_{k_\nu}(B) + \sum_{i=0}^{k_\nu} Z_i \psi_i$$

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, an recursive argument shows that each map  $J_k(B)$  may be decomposed as

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

where  $c_{i,k}(W)$  are explicit linear functionals that depend on  $W_1$  and on the collection  $(\int_0^1 (1-t)^{k-j} W_t dt)_{1 \leq j \leq k}$  (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  $k \geq 2$ . Hence,

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

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

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

necessarily implies that  $Z_i^1 = Z_i^2$  for  $i \in \{0, \dots, k_\nu\}$ , 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 Reproducing Kernel Hilbert Space (shortened as RKHS in the sequel) associated to the Gaussian process (2.8) in  $\underline{\mathbb{B}}$  is  $\underline{\mathbb{H}} := T\mathbb{H}$  where  $\mathbb{H}$  is the RKHS derived in the simplest 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, since the sets  $\mathbb{B}_1$  and  $\mathbb{B}_2$  are independent, the RKHS  $\mathbb{H}$  may be described as

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

where  $AC([0, 1])$  is the set of absolutely continuous functions on  $[0, 1]$ ,  $\mathbb{H}$  is endowed with the following 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 influence of 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  $w_0^{(j)}(0) = w_0^{(j)}(1)$ . In this view, simple computations yield that for any integer  $p$ :

$$\psi_k^{(2p)}(t) = (-1)^p (2\pi k)^{2p} \psi_k(t),$$

and

$$\psi_k^{(2p+1)}(t) = (-1)^p (2\pi k)^{2p+1} [-\sin(2\pi kt) + \cos(2\pi kt)].$$

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  $d_0 := (w_0^{(j)}(0))_{j=0\dots k_\nu}$ ,  $Z = (Z_0, \dots, Z_{k_\nu+1})$  and the squared matrix of size  $(k_\nu + 1) \times (k_\nu + 1)$ :

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

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 the set of 1-periodic functions  $\mathcal{C}_1^{k_\nu}$  which possesses 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  $(c_{i,k_\nu})_{i=0\dots k_\nu}$  in (C.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 C.1 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

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

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

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

It reminds to obtain a lower bound of the small ball probability of the *centered* Gaussian ball. Note that  $b$  and  $J_{k_\nu}$  are independent Gaussian processes. We have somewhat trivially that  $\log \left( \frac{1}{\epsilon} \right) \lesssim \log \mathbb{P} (\|b\|_\infty \leq \epsilon)$ . Thus,

the main difficulty relies on the lower bound of

$$\phi_0(\epsilon) := \log \mathbb{P} (\|J_k(B)\|_\infty \leq \epsilon).$$

Going back to (C.1), we see that  $J_k(B)$  can be decomposed into two nonindependent 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  $(\int_0^1 (1-t)^{k-j} W_t dt)_{1 \leq j \leq k}$ . Therefore

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

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

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

which is of the same order as the approximation term. Gathering now our lower bound on shifted Gaussian ball and the term above ends the proof of the Theorem.  $\square$

#### APPENDIX D: EQUIVALENTS ON MODIFIED BESSEL FUNCTIONS

LEMMA D.1. *For any  $n \in \mathbb{Z}$  and  $a > 0$ , define*

$$A_n(a) := \int_0^{2\pi} e^{a \cos(u)} \cos(nu) du.$$

*Then, the following equivalent holds:*

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

PROOF. This equivalent is related to the modified Bessel functions (see [AS64] for classical equivalents on Bessel functions and [LL10] for standard results on continuous time random walks). More precisely,  $I_m(a)$  is defined as

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

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  $a$ , it is possible to use standard results on modified Bessel functions. Equation (9.7.7) of [AS64], p. 378. yields

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

□

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

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

Hence,  $B_n(a)$  is the probability of a continuous time random walk to be in place  $n \in \mathbb{Z}$  at time  $a$ . In this way, we get some asymptotic equivalents of  $B_n(a)$  (and so of  $A_n(a)$ ): from the Brownian approximation of the CTRW, we should suspect that for  $a$  large enough

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

Moreover, from [AS64], we know that

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

and this equivalent is sharp when  $a$  is large enough: from equation (9.7.1) p. 377 of [AS64], we know that

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

We remark that (D.3) yields the heuristic equivalent suspected in (D.2):  $B_n(a) = e^{-a} I_n(a) \sim \frac{1}{\sqrt{2\pi a}}$ , although (D.1) provides a quite different information for smaller  $a$ . We do not have purchase more investigation on this asymptotic since we will see that indeed, (D.1) is much more larger than (D.3).

For  $a \in [\sqrt{n}, 2n]$ , we do not have found any satisfactory equivalent of modified Bessel functions. Formula of [AS64] is still tractable but yields some different equivalent which is not "uniform enough" since we need to integrate this equivalent for our bayesian analysis. This is not so important since we can see for our range of application that the most important weight belongs to the smaller values of  $a$ .

## ACKNOWLEDGEMENTS

S. G. is indebted to Jean-Marc Azaïs, Patrick Cattiaux and Laurent Miclo for stimulating discussions related to some technical parts of this work. Authors also thank Jérémie Bigot, Ismaël Castillo, Xavier Gendre, Judith Rousseau and Alain Trouvé for enlightening exchanges.

## REFERENCES

- [AAT07] S. Allasonnière, Y. Amit, and A. Trouvé. Toward a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society (B)*, 69:3–29, 2007.
- [AGP91] Y. Amit, U. Grenander, and M. Piccioni. Structural Image Restoration through Deformable Templates. *JASA*, 86(414):376–387, 1991.
- [AKT10] Stéphanie Allasonnière, Estelle Kuhn, and Alain Trouvé. Construction of Bayesian deformable models via a stochastic approximation algorithm: a convergence study. *Bernoulli*, 16(3):641–678, 2010.
- [AS64] Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.
- [AW09] Jean-Marc Azaïs and Mario Wschebor. *Level sets and extrema of random processes and fields*. John 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] Jérémie Bigot and Sébastien Gadat. A deconvolution approach to estimation of a common shape in a shifted curves model. *Ann. Statist.*, 38(4):2422–2464, 2010.
- [BG12] Jérémie Bigot and Xavier Gendre. Minimax properties of Fréchet means of discretely sampled curves. *Preprint*, 2012.
- [BGKM12] Jérémie Bigot, Sébastien Gadat, Thierry Klein, and Clément Marteau. Intensity estimation of non-homogeneous Poisson processes from shifted trajectories. *Preprint*, 2012.
- [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.
- [BGV09] J. Bigot, F. Gamboa, and M. Vimond. Estimation of translation, rotation and scaling between noisy images using the Fourier Mellin transform. *SIAM Journal on Imaging Sciences*, 2(2):614–645, 2009.
- [Big06] J. Bigot. Landmark-based registration of curves via the continuous wavelet transform. *Journal of Computational and Graphical Statistics*, 15(3):542–564, 2006.
- [Big11] Jérémie Bigot. Fréchet means of curves for signal averaging and application to ECG data analysis. *preprint*, 2011.
- [BM05] Cristina Butucea and Catherine 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] Ismaël Castillo. A semi-parametric Bernstein-von Mises theorem for Gaussian process priors. *Probability Theory and Related Fields*, 1:53–99, 2012.
- [CD11] O. Collier and A. Dalalyan. Curve registration by nonparametric goodness-of-fit testing. *Preprint*, 2011.
- [Col12] Olivier Collier. Minimax hypothesis testing for curve registration. *Electronic Journal of Statistics*, 6:1129–1154, 2012.
- [Fer73] Thomas S. Ferguson. A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1:209–230, 1973.
- [GGvdV00] Subhashis Ghosal, Jayanta K. Ghosh, and Aad W. van der Vaart. Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531, 2000.
- [GL95] Richard D. Gill and Boris Y. Levit. Applications of the Van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, 1(1-2):59–79, 1995.
- [GM01] C. A. Glasbey and K. V. Mardia. A penalized likelihood approach to image warping. *Journal of the Royal Statistical Society (B)*, 63(3):465–514, 2001.
- [GM07] U. Grenander and M. Miller. *Pattern Theory: From Representation to Inference*. Oxford Univ. Press, Oxford, 2007.
- [GS02] Alison L. Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.
- [GvdV01] Subhashis Ghosal and Aad 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] Christopher R. Genovese and Larry 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] Tadeusz Inglot and Teresa 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] James Kuelbs, Wenbo V. Li, and Werner Linde. The Gaussian measure of shifted balls. *Probab. Theory Related Fields*, 98(2):143–162, 1994.
- [LCY00] Lucien Le Cam and Grace Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer series in statistics. Springer Verlag, New-York, 2000.
- [LL10] Gregory F. Lawler and Vlada Limic. *Random walk: a modern introduction*, volume 123 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
- [LS01] W. V. Li and Q.-M. Shao. Gaussian processes: inequalities, small ball probabilities and applications. In *Stochastic processes: theory and methods*, volume 19 of *Handbook of Statist.*, pages 533–597. North-Holland, Amsterdam, 2001.
- [Mat02] Catherine Matias. Semiparametric deconvolution with unknown noise variance. *ESAIM Probab. Statist.*, 6:271–292 (electronic), 2002. New directions in time series analysis (Luminy, 2001).
- [MM11] Cathy Maugis and Bertrand 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] Wooram Park, Dean R. Madden, Daniel N. Rockmore, and Gregory S. Chirikjian. Deblurring of class-averaged images in single-particle electron microscopy. *Inverse Problems*, 26(3):035002, 29, 2010.
- [RR12] Vincent Rivoirard and Judith Rousseau. Posterior concentration rates for infinite dimensional exponential families. *Bayesian Analysis*, 7(2):311–334, 2012.
- [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 Jayanta K. Ghosh*, volume 3 of *Inst. Math. Stat. Collect.*, pages 200–222. Inst. Math. Statist., Beachwood, OH, 2008.
- [WS95] Wing Hung Wong and Xiaotong Shen. Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.*, 23(2):339–362, 1995.

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](mailto:dominique.bontemps@math.univ-toulouse.fr)  
E-MAIL: [sebastien.gadat@math.univ-toulouse.fr](mailto:sebastien.gadat@math.univ-toulouse.fr)