



**HAL**  
open science

# Semiparametric inference for mixtures of circular data

Claire Lacour, Thanh Mai Pham Ngoc

► **To cite this version:**

Claire Lacour, Thanh Mai Pham Ngoc. Semiparametric inference for mixtures of circular data. 2021.  
hal-03149434v1

**HAL Id: hal-03149434**

**<https://hal.science/hal-03149434v1>**

Preprint submitted on 12 Mar 2021 (v1), last revised 25 May 2022 (v2)

**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.

# Semiparametric inference for mixtures of circular data

Claire Lacour, Thanh Mai Pham Ngoc

March 12, 2021

## Abstract

We consider  $X_1, \dots, X_n$  a sample of data on the circle  $\mathbb{S}^1$ , whose distribution is a two-component mixture. Denoting  $R$  and  $Q$  two rotations on  $\mathbb{S}^1$ , the density of the  $X_i$ 's is assumed to be  $g(x) = pf(R^{-1}x) + (1-p)f(Q^{-1}x)$ , where  $p \in (0, 1)$  and  $f$  is an unknown density on the circle. In this paper we estimate both the parametric part  $\theta = (p, R, Q)$  and the nonparametric part  $f$ . The specific problems of identifiability on the circle are studied. A consistent estimator of  $\theta$  is introduced and its asymptotic normality is proved. We propose a Fourier-based estimator of  $f$  with a penalized criterion to choose the resolution level. We show that our adaptive estimator is optimal from the oracle and minimax points of view when the density belongs to a Sobolev ball. Our method is illustrated by numerical simulations.

**Keywords:** *Mixture model ; Semiparametric estimation ; Circular data*

## 1 Introduction

Circular data are collected when the topic of interest is a direction or a time of day. These particular data appear in many applications: earth sciences (e.g. wind directions), medicine (e.g. circadian rhythm), ecology (e.g. animal movements), forensics (crime incidence). Different surveys on statistical methods for circular data can be found: Mardia and Jupp (2000), Jammalamadaka and SenGupta (2001), Ley and Verdebout (2017) or more recently Pewsey and García-Portugués (2020). In the present work, we consider a mixture model with two components equal up to a rotation. We observe  $X_1, \dots, X_n$  a sample of data on  $\mathbb{S}^1$  with probability distribution function:

$$g(x) = p_0f(R_0^{-1}x) + (1 - p_0)f(Q_0^{-1}x) = p_0f(x - \alpha_0) + (1 - p_0)f(x - \beta_0). \quad (1.1)$$

In the right hand side we have identified  $f : \mathbb{S}^1 \rightarrow \mathbb{R}$  and its periodized version on  $\mathbb{R}$ . Here  $R_0$  and  $Q_0$  are two unknown rotations of the circle.  $R_0$  is a rotation with angle  $\alpha_0$  and  $Q_0$  is a rotation with angle  $\beta_0$ . The aim is to estimate both  $\theta_0 = (p_0, \alpha_0, \beta_0)$  and the nonparametric part  $f$ . Note that we can rewrite model (1.1) as

$$X_i = Y_i + \varepsilon_i \pmod{2\pi}, \quad i = 1, \dots, n \quad (1.2)$$

where  $Y_i$  has density  $f$  and  $\varepsilon_i$  is a Bernoulli angle, which is equal to  $\alpha_0$  with probability  $p_0$  and  $\beta_0$  otherwise. Accordingly, model (1.1) can be viewed as a circular convolution model with unknown noise operator  $\varepsilon$ . The circular convolution model has been studied by Goldenshluger (2002) in the case of known noise operator whereas Johannes and Schwarz (2013) dealt with unknown error distribution but have at their disposal an independent sample of the noise to estimate this latter. It is worth pointing out that Goldenshluger (2002) and Johannes and Schwarz (2013) made the usual

assumptions on the decay of the Fourier coefficients of the density of  $\varepsilon$ , whereas in model (1.1) the Fourier coefficients are not decreasing.

Mixture models for describing multimodal circular data date back to Pearson (1894) and have been largely used since then. An important case in the literature is the mixture of two von Mises distributions which has been explored in numerous works. Let us cite among others papers by Bartels (1984), Spurr (1981) or Chen et al. (2008). From a practical point of view, algorithms have also been proposed to deal with it, including maximum likelihood algorithms by Jones and James (1969) or a characteristic function based procedure by Spurr and Koutbey (1991). Note that on the unit hypersphere, Banerjee et al. (2005) investigated clustering methods for mixtures of von Mises Fisher distributions. In our framework, we shall not assume any parametric form of the density and hence the model is said to be semiparametric. To the best of our knowledge, this is the first work devoted to the study of the semiparametric mixture model for circular data. This semiparametric model is more complex and intricate than the usual parametric one encountered in the circular literature. In the spherical case, Kim and Koo (2000) studied the general mixture framework for a location parameter but assuming that the nonparametric part  $f$  is known. On the real line, this semiparametric model has been studied by Bordes et al. (2006), Hunter et al. (2007), Butucea and Vandekerkhove (2014) or Gassiat and Rousseau (2016) for dependent latent variables. For the multivariate case, see for instance Hall and Zhou (2003), Hall et al. (2005), Gassiat et al. (2018), Hohmann and Holzmann (2013). When dealing with the specific case of one of the two components being parametric, one refers to work by Ma and Yao (2015) and references therein.

Identifiability questions are at the heart of the theory of mixture models and the circular context is no exception. Thus, our first task is to study the identifiability of the model. From a mathematical point of view, the topology of the circle makes the problem very different from the linear case. In the circular parametric case, Fraser et al. (1981) obtained identifiability results for the von Mises distributions, then extended in Kent (1983) to generalized von Mises distributions while Holzmann et al. (2004) focused on wrapped distributions, basing their analysis on the Fourier coefficients. Here, the Fourier coefficients turn out to be very useful as well but the nonparametric paradigm makes the study quite different and intricate. Our identifiability results are obtained under mild assumptions on the Fourier coefficients. We require that the coefficients are real which can be related to the usual symmetry assumption in mixture models (see for instance Hunter et al. (2007)) and we impose that only the first 4 coefficients do not vanish. Interestingly enough, some not intuitive phenomena appear. A striking case occurs when the angles  $\alpha_0$  and  $\beta_0$  are distant from  $2\pi/3$ , the model is then nonidentifiable which is quite surprising at first sight.

Once the identifiability of the model obtained, we resort to a contrast function in the line of Butucea and Vandekerkhove (2014) to estimate the Euclidian parameter  $\theta_0$ . In that regard, we prove the consistency of our estimator and an asymptotic normality result. Thereafter, for the estimation of the nonparametric part, a penalized empirical risk estimation method is used. The estimator of the density turns to be adaptive (meaning that it does not require the specification of the unknown smoothness parameter), a property which was not reached so far for this semiparametric model even in the linear case. The procedure devised is hence relevant for practical purposes. We prove an oracle inequality and minimax rates are achieved by our estimator for Sobolev regularity classes. Eventually, a numerical section shows the good performances of the whole estimation procedure.

The paper is organized as follows. Section 2 is devoted to the identifiability of the model. Section 3 tackles the estimation of the parameter  $\theta_0$  whereas Section 4 focuses on the estimation of the nonparametric part. Finally Section 5 presents numerical implementations of our procedure. Proofs are gathered in Section 6.

## 2 Identifiability

In this section, to keep the notation as light and clear as possible, we drop the subscript 0 in the parameters. For any function  $g$  and any angle  $\alpha$ , denote  $g_\alpha(x) := g(x - \alpha)$ . For any complex number  $a$ ,  $\bar{a}$  is the complex conjugate of  $a$ . For any integrable function  $\phi : \mathbb{S}^1 \rightarrow \mathbb{R}$ , we denote for any  $l \in \mathbb{Z}$ ,  $\phi^{*l} = \int_{\mathbb{S}^1} \phi(x) e^{-ilx} \frac{dx}{2\pi}$ , the Fourier coefficients.

Let us now study the identifiability of model (1.1). First, it is obvious that if  $p = 0$ ,  $\alpha$  is not identifiable, and if  $p = 1$ ,  $\beta$  is not identifiable. In the same way,  $p$  is not identifiable if  $\alpha = \beta$ . Moreover, as explained in Hunter et al. (2007) for a translation mixture on the real line, the case  $p = 1/2$  has to be avoided. Indeed, denoting  $g$  a density and for instance  $f = \frac{1}{2}g_1 + \frac{1}{2}g_{-1}$  and  $f' = \frac{1}{2}g_2 + \frac{1}{2}g_{-2}$  we have

$$f_1 + f_3 = f'_2 + f'_4.$$

In addition, it is well known that, in such a mixture model,  $(p, \alpha, \beta)$  cannot be distinguished from  $(1 - p, \beta, \alpha)$ : it is the so-called *label switching* problem. So we will assume that  $p \in (0, 1/2)$  (for mixtures on  $\mathbb{R}$  it is assumed alternatively that  $\alpha < \beta$  but ordering angles is less relevant).

Now let us study the specific problems of identifiability on the circle, that do not appear on  $\mathbb{R}$ . First, if  $f$  is the uniform probability, the model is not identifiable, so we have to exclude this case. Another case to exclude is the case of  $\delta$ -periodic functions. Indeed in this case  $f_\alpha = f_{\alpha+\delta}$ . These functions have the property that  $f^{*l} = 0$  for all  $l \notin (2\pi/\delta)\mathbb{Z}$ . So we will require that the Fourier coefficients of  $f$  do not cancel out too much. Here we will assume

$$\text{for all } l \in \{1, 2, 3, 4\}, \quad f^{*l} \neq 0, \text{ and } f^{*l} = \overline{f^{*l}}.$$

This last assumption can be related to the symmetry of  $f$ . Indeed if  $f$  is zero-symmetric then all its Fourier coefficients are real. Symmetry is a usual assumption in this mixture context, to distinguish between the translations of  $f$ : for any  $\delta \in \mathbb{R}$ ,

$$pf(x - \alpha) + (1 - p)f(x - \beta) = pf_\delta(x - \alpha + \delta) + (1 - p)f_\delta(x - \beta + \delta)$$

More precisely, Hunter et al. (2007) show that symmetry is a sufficient and necessary condition for identifiability of the model mixture on  $\mathbb{R}$ . In the circle framework, it is natural to work with Fourier coefficients rather than Fourier transform as on  $\mathbb{R}$ . A lot of circular densities have their Fourier coefficients real, provided that their location parameter is  $\mu = 0$ : for example the Jones-Pewsey density, which includes the cardioid, the wrapped Cauchy density, and the von Mises density. Here we require the assumption only for the first 4 Fourier coefficients of  $f$  (due to our proof), which is milder than symmetry.

Let us now state our identifiability result under these assumptions. Note that Holzmann et al. (2004) have studied the identifiability of this model when  $f$  belongs to a parametric scale-family of densities, but here we face a nonparametric problem concerning  $f$ .

**Theorem 1.** *Assume that  $\theta = (p, \alpha, \beta)$  and  $\theta' = (p', \alpha', \beta')$  belong to*

$$\{(p, \alpha, \beta) \in (0, 1/2) \times \mathbb{S}^1 \times \mathbb{S}^1, \quad \alpha \neq \beta \pmod{2\pi}\}$$

*and that  $f, f'$  belongs to*

$$\left\{ f : \mathbb{S}^1 \rightarrow \mathbb{R} \text{ density such that, for all } l \in \{1, 2, 3, 4\}, f^{*l} \in \mathbb{R} \setminus \{0\} \right\}.$$

*Suppose  $pf_\alpha + (1 - p)f_\beta = p'f_{\alpha'} + (1 - p')f_{\beta'}$ . Then*

1. either  $(p', \alpha', \beta') = (p, \alpha, \beta)$  and  $f' = f$ ,
2. or  $(p', \alpha', \beta') = (p, \alpha + \pi, \beta + \pi)$  and  $f' = f_\pi$
3. or if  $\beta - \alpha = \pi \pmod{2\pi}$  then  $f'$  is a linear combination of  $f$  and  $f_\pi$ , and either  $(\alpha', \beta') = (\alpha, \beta)$ , or  $(\alpha', \beta') = (\beta, \alpha)$ .
4. or if  $\beta - \alpha = \pm 2\pi/3 \pmod{2\pi}$  then  $f'$  is a linear combination of  $f_{\pi/3}, f_{-\pi/3}, f_\pi$  and  $p' = (1 - 2p)/(2 - 3p)$  and
  - (a) if  $\beta - \alpha = 2\pi/3$ ,  $(\alpha', \beta') = (\alpha + \pi, \beta - \pi/3)$ ,
  - (b) if  $\beta - \alpha = -2\pi/3$ ,  $(\alpha', \beta') = (\alpha + \pi, \beta + \pi/3)$

Case 2. arises from a specific feature of circular distributions: if  $f$  is symmetric with respect to 0 then it is symmetric with respect to  $\pi$ . Unlike the real case, a symmetry assumption does not exclude the case  $f'(x) = f(x - \pi)$ . To bypass this we could assume for instance  $f^{*1} > 0$ . Indeed for each  $l \in \mathbb{Z}$ ,  $(f_\pi)^{*l} = f^{*l}(-1)^l$ , so the Fourier coefficients of  $f$  and  $f_\pi$  have opposite sign for any odd  $l$ . With our assumption, we recover among  $f$  and  $f_\pi$  the one with positive first Fourier coefficient, i.e. with positive mean resultant length. Nevertheless our estimation procedure begins with the parametric part so that this assumption concerning only the nonparametric part will not allow us to distinguish  $\alpha$  from  $\alpha + \pi$  in this first parametric estimation step. That is why we rather choose to assume that  $\alpha$  and  $\beta$  belong to  $[0, \pi) \pmod{\pi}$ .

Case 3. concerns bipolar data since  $\alpha$  and  $\beta$  are diametrically opposed (separated by  $\pi$  radians). In this case  $\alpha$  and  $\beta$  are identifiable, but  $p$  and  $f$  not. Indeed, for any density  $f$  and any  $p'$ , we can find  $q$  such that  $f' = qf + (1 - q)f_\pi$  verifies  $pf_\alpha + (1 - p)f_\beta = p'f'_{\alpha'} + (1 - p')f'_{\beta'}$ .

Let us now discuss the case 4., which is the most curious (we shall only comment the first case (a), the other is similar). Let us set

$$f'(x) = (1 - p)f\left(x - \frac{\pi}{3}\right) + (1 - p)f\left(x + \frac{\pi}{3}\right) + (2p - 1)f(x - \pi)$$

as represented in Figure 1. This function is symmetric if  $f$  is symmetric. Then we can write  $f'_{\pi/3}$ :

$$f'\left(x - \frac{\pi}{3}\right) = (1 - p)f\left(x - \frac{2\pi}{3}\right) + (1 - p)f(x) + (2p - 1)f\left(x - \frac{4\pi}{3}\right),$$

as well as  $f'_\pi$ :

$$f'(x - \pi) = (1 - p)f\left(x - \frac{4\pi}{3}\right) + (1 - p)f\left(x - \frac{2\pi}{3}\right) + (2p - 1)f(x).$$

Hence a mixture of  $f'_\pi$  and  $f'_{\pi/3}$  gives a mixture of  $f(x), f(x - \frac{2\pi}{3}), f(x - \frac{4\pi}{3})$ :

$$\begin{aligned} p'f'(x - \pi) + (1 - p')f'\left(x - \frac{\pi}{3}\right) &= [p'(2p - 1) + (1 - p')(1 - p)]f(x) \\ &\quad + [p'(1 - p) + (1 - p')(1 - p)]f\left(x - \frac{2\pi}{3}\right) \\ &\quad + [p'(1 - p) + (1 - p')(2p - 1)]f\left(x - \frac{4\pi}{3}\right) \end{aligned}$$

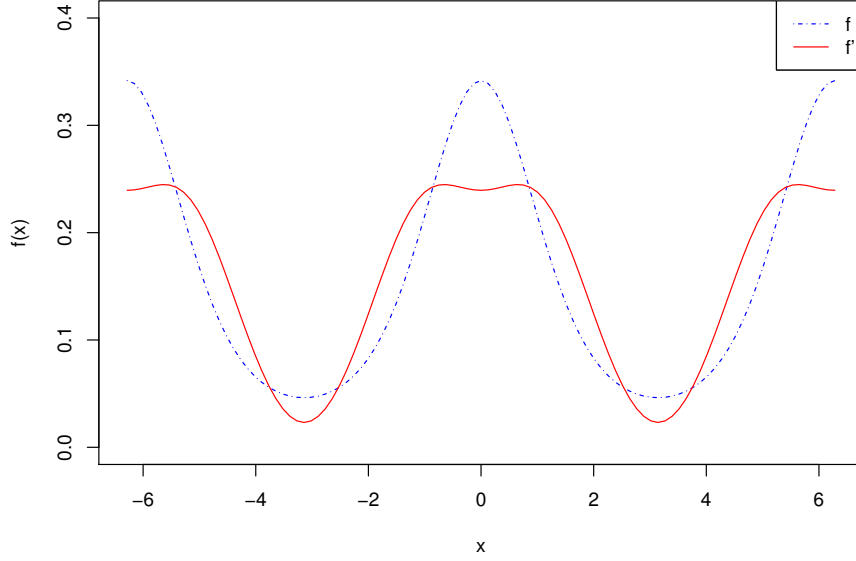


Figure 1: Plot of a circular density  $f$  (dashed blue), and of  $f' = (1-p)f_{\frac{\pi}{3}} + (1-p)f_{-\frac{\pi}{3}} + (2p-1)f_{\pi}$  (solid red). Here  $f$  is the von Mises density with mean 0 and concentration 1.

If now  $p' = (1-2p)/(2-3p)$ , then  $p'(1-p) + (1-p')(2p-1) = 0$  and the third component  $f(x - \frac{4\pi}{3})$  vanishes. Thus

$$p'f'(x - \pi) + (1-p')f'(x - \frac{\pi}{3}) = pf(x) + (1-p)f(x - \frac{2\pi}{3}).$$

In such a particular case, we cannot identify  $\theta$  nor  $f$ . However this happens only when  $\beta - \alpha = \pm 2\pi/3$ . So, to exclude these case, we will now assume  $\beta \neq \alpha \pmod{2\pi/3}$ .

Finally, we shall assume that  $f \in \mathcal{F}$  with some assumptions for  $\mathcal{F}$ :

**Assumption 1.**

$$\mathcal{F} \subset \left\{ f : \mathbb{S}^1 \rightarrow \mathbb{R} \text{ density s.t. for all } l \in \{1, 2, 3, 4\}, \quad f^{*l} \in \mathbb{R} \setminus \{0\} \right\}$$

or

**Assumption 2.**

$$\mathcal{F} \subset \left\{ f : \mathbb{S}^1 \rightarrow \mathbb{R} \text{ density s.t. for all } l \in \{1, 2, 3, 4\}, \quad f^{*l} \in \mathbb{R} \setminus \{0\}, \quad f^{*1} > 0 \right\}$$

and we shall assume that  $\theta \in \Theta$  with some assumptions for  $\Theta$ :

**Assumption 3.**

$$\Theta \subset \left\{ (p, \alpha, \beta) \in \left(0, \frac{1}{2}\right) \times \mathbb{S}^1 \times \mathbb{S}^1, \quad \alpha \neq \beta \pmod{\pi, 2\pi/3} \right\}$$

where  $\alpha \neq \beta \pmod{2\pi/3, \pi}$  means  $\beta - \alpha \notin \{-\frac{2\pi}{3}, 0, \frac{2\pi}{3}, \pi\} + 2\pi\mathbb{Z}$ , or

**Assumption 4.**

$$\Theta \subset \left\{ (p, \alpha, \beta) \in \left(0, \frac{1}{2}\right) \times [0, \pi) \times [0, \pi), \quad \alpha \neq \beta \pmod{2\pi/3} \right\}$$

Note that Assumption 4 implies Assumption 3, and Assumption 2 implies Assumption 1. We can write the following result.

**Corollary 2.** *Under Assumptions 1 and 4, or under Assumptions 2 and 3, model (1.1) is identifiable. Under Assumptions 1 and 3, model (1.1) is identifiable modulo  $\pi$ , that is to say that if  $pf_\alpha + (1-p)f_\beta = p'f_{\alpha'} + (1-p')f_{\beta'}$ , then  $p' = p$  and either  $(\alpha', \beta') = (\alpha, \beta)$  and  $f' = f$ , or  $(\alpha', \beta') = (\alpha + \pi, \beta + \pi)$  and  $f' = f_\pi$ .*

Moreover, the proof of Theorem 1 provides the following statement.

**Lemma 3.** *Under Assumption 3, denoting  $M^l(\theta) := pe^{-i\alpha l} + (1-p)e^{-i\beta l}$ , for all  $\theta, \theta' \in \Theta$ ,*

$$\forall 1 \leq l \leq 4, \Im \left( M^l(\theta') \overline{M^l(\theta)} \right) = 0 \Leftrightarrow \theta' = \theta \text{ or } \theta' = \theta + \pi.$$

where  $\theta' = \theta + \pi$  means  $(p', \alpha', \beta') = (p, \alpha + \pi, \beta + \pi)$ .

### 3 Estimation for the parametric part

Now, let us denote for all  $l \in \mathbb{Z}$

$$M^l(\theta) := pe^{-i\alpha l} + (1-p)e^{-i\beta l}.$$

In model (1.1) the Fourier coefficients of  $g$  satisfy for any  $l$ :

$$g^{*l} = (p_0 e^{-i\alpha_0 l} + (1-p_0) e^{-i\beta_0 l}) f^{*l}.$$

Thus  $g^{*l} = M^l(\theta_0) f^{*l}$  and the previous lemma gives that  $\theta = \theta_0$  (or  $\theta_0 + \pi$ ) if and only if, for each  $l \in \{1, \dots, 4\}$ ,

$$\Im \left( M^l(\theta_0) \overline{M^l(\theta)} \right) = 0 \Leftrightarrow \Im \left( g^{*l} \overline{M^l(\theta)} \right) = 0$$

using that  $f^{*l}$  are non-zero real numbers. This invites us to consider

$$S(\theta) := \sum_{l=-4}^4 \left( \Im \left( g^{*l} \overline{M^l(\theta)} \right) \right)^2 = \sum_{l=-4}^4 \left( \Im \left( g^{*l} \{pe^{i\alpha l} + (1-p)e^{i\beta l}\} \right) \right)^2.$$

Note that  $g^{*0} \overline{M^0(\theta)} = 1/(2\pi)$  and that  $\Im \left( g^{*(-l)} \overline{M^{-l}(\theta)} \right) = \Im \left( \overline{g^{*l}} M^l(\theta) \right) = -\Im \left( g^{*l} \overline{M^l(\theta)} \right)$  so that we can also write

$$S(\theta) = 2 \sum_{l=1}^4 \left( \Im \left( g^{*l} \overline{M^l(\theta)} \right) \right)^2.$$

The empirical counterpart of  $S(\theta)$  is

$$\begin{aligned}
\tilde{S}_n(\theta) &= \sum_{l=-4}^4 \left( \Im \left( \widehat{g^{\star l} M^l(\theta)} \right) \right)^2 \\
&= \sum_{l=-4}^4 \left( \Im \left( \frac{1}{2\pi n} \sum_{k=1}^n e^{-ilX_k} \overline{M^l(\theta)} \right) \right)^2 \\
&= \frac{1}{4\pi^2 n^2} \sum_{l=-4}^4 \sum_{1 \leq k, j \leq n} \Im \left( e^{ilX_k} M^l(\theta) \right) \Im \left( e^{ilX_j} M^l(\theta) \right).
\end{aligned}$$

Next, we consider a slightly modified version of  $\tilde{S}_n(\theta)$  by removing the diagonal terms

$$S_n(\theta) = \frac{1}{4\pi^2 n(n-1)} \sum_{l=-4}^4 \sum_{1 \leq k \neq j \leq n} \Im \left( e^{ilX_k} M^l(\theta) \right) \Im \left( e^{ilX_j} M^l(\theta) \right). \quad (3.1)$$

Let us denote

$$Z_k^l(\theta) := \Im \left( \frac{e^{ilX_k}}{2\pi} M^l(\theta) \right) \quad \text{and} \quad J^l(\theta) := \Im \left( \overline{g^{\star l} M^l(\theta)} \right).$$

Hence

$$S_n(\theta) = \frac{1}{n(n-1)} \sum_{l=-4}^4 \sum_{k \neq j} Z_k^l(\theta) Z_j^l(\theta).$$

Note that we have  $\mathbb{E}(Z_k^l(\theta)) = J^l(\theta)$ , and  $S_n(\theta)$  is an unbiased estimator of  $S(\theta)$ .

Let the estimator of  $\theta_0$  be

$$\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} S_n(\theta). \quad (3.2)$$

For this estimator we can prove the following consistency result.

**Theorem 4.** *Consider  $\Theta$  a compact set included in*

$$\left\{ (p, \alpha, \beta) \in (0, 1/2) \times \mathbb{S}^1 \times \mathbb{S}^1, \quad \alpha \neq \beta \pmod{2\pi/3, \pi} \right\}$$

*and the estimator  $\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} S_n(\theta)$ . We have  $\hat{\theta}_n \rightarrow \theta_0 \pmod{\pi}$  in probability.*

The last convergence means that for all  $\epsilon > 0$ , the probability  $\mathbb{P}(\|\hat{\theta}_n - \theta_0\| \leq \epsilon \text{ or } \|\hat{\theta}_n - \theta_0 - \pi\| \leq \epsilon)$  tends to 1 when  $n$  goes to  $+\infty$ , where  $\|\cdot\|$  denotes the Euclidean norm.

*Proof.*  $\Theta$  is a compact set and  $S$  is continuous. Lemma 12 ensures that  $S_n$  is Lipschitz hence uniformly continuous, and Proposition 13 ensures that for all  $\theta$ ,  $|S_n(\theta) - S(\theta)|$  tends to 0 in probability. Then it is sufficient to apply a classical Lemma to conclude. See the details in Section 6.2  $\square$

From now on, we assume that  $\Theta$  is a compact set included in  $(0, \frac{1}{2}) \times [0, \pi) \times [0, \pi)$ , as in Assumption 4. Then,  $\theta_0 + \pi$  is excluded and under Assumption 4,  $\hat{\theta}_n \rightarrow \theta_0$  in probability. Moreover this estimator is asymptotically normal. We denote  $\dot{\phi}(\theta)$  the gradient of any function  $\phi$  with respect to  $\theta = (p, \alpha, \beta)$ ,  $\ddot{\phi}(\theta)$  the Hessian matrix and for any matrix  $A$ , we denote  $A^\top$  its transpose.



**Theorem 5.** Consider  $\Theta$  a compact set included in

$$\{(p, \alpha, \beta) \in (0, 1/2) \times [0, \pi) \times [0, \pi), \quad \alpha \neq \beta \pmod{2\pi/3}\}$$

and the estimator  $\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} S_n(\theta)$ . Let  $\mathcal{A}$  be the Hessian matrix of  $S$  in  $\theta_0$ :  $\mathcal{A} = \ddot{S}(\theta_0) = 2 \sum_{l=-4}^4 \dot{J}^l(\theta_0) \dot{J}^l(\theta_0)^\top$ . Then, if  $\mathcal{A}$  is invertible,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

where  $\Sigma = \mathcal{A}^{-1}V\mathcal{A}^{-1}$ ,  $V = 4\mathbb{E}(UU^\top)$  and  $U = \sum_{l=-4}^4 \dot{J}^l(\theta_0)Z_1^l(\theta_0)$ .

The proof can be found in Section 6.3.

## 4 Nonparametric part

Let us now estimate the nonparametric part. We shall use the following norm: for any function  $\phi$ , we denote  $\|\phi\|_2 = (\frac{1}{2\pi} \int_{\mathbb{S}^1} \phi^2(x) dx)^{1/2}$ . Recall that for all  $l \in \mathbb{Z}$ ,  $g^{*l} = M^l(\theta_0)f^{*l}$  where  $g$  is the density of the observations  $X_k$  and  $g^{*l}$  its Fourier coefficient. Then  $f^{*l} = g^{*l}/M^l(\theta_0)$ . We can verify that  $M^l(\theta_0) \neq 0$ . Indeed, for any  $\theta \in \Theta$ ,

$$|M^l(\theta)|^2 = p^2 + (1-p)^2 + 2p(1-p) \cos[l(\beta - \alpha)] \geq (1-2p)^2 > 0.$$

Nevertheless this division by  $M^l(\theta_0)$  lead us to impose a new assumption. We assume that there exists  $P \in (0, 1/2)$  such that  $0 < p < P$  for any  $p$ , ie.

**Assumption 5.**  $\Theta$  is a compact set included in

$$\{(p, \alpha, \beta) \in (0, P) \times [0, \pi) \times [0, \pi), \quad \alpha \neq \beta \pmod{2\pi/3}\}$$

Under this assumption,  $|M^l(\theta)|$  is always bounded from below by  $1 - 2P$ . Now, to estimate  $g^{*l} = \int_{\mathbb{S}^1} e^{-ilx} g(x) dx / (2\pi)$ , it is natural to define

$$\widehat{g}^{*l} = \frac{1}{2\pi n} \sum_{k=1}^n e^{-ilX_k}.$$

If  $\hat{\theta} = \hat{\theta}_n$  is the previous estimator of the parametric part, we set the plugin estimator of the Fourier coefficient:

$$\widehat{f}^{*l} = \frac{1}{2\pi n} \sum_{k=1}^n M^l(\hat{\theta})^{-1} e^{-ilX_k}.$$

Finally, for  $L$  a positive integer, set

$$\hat{f}_L(x) = \sum_{l=-L}^L \widehat{f}^{*l} e^{ilx}.$$

To measure the performance of this estimator, we use Parseval equality to write

$$\|f - \hat{f}_L\|_2^2 = \sum_{|l|>L} |f^{*l}|^2 + \sum_{l=-L}^L |f^{*l} - \widehat{f}^{*l}|^2$$

which is the classical bias variance decomposition. Moreover it is possible to prove that the variance term satisfy  $\sum_{l=-L}^L \mathbb{E}|f^{\star l} - \widehat{f}^{\star l}|^2 = O(\frac{2L+1}{n})$  (see Lemma 16 below). To control the bias term we recall the definition of the Sobolev ellipsoid:

$$W(s, R) = \{f : \mathbb{S}^1 \rightarrow \mathbb{R}, \sum_{l \in \mathbb{Z}} |l|^{2s} |f^{\star l}|^2 \leq R^2\}.$$

For such a smooth  $f$ , the risk of estimator  $\hat{f}_L$  is then bounded in the following way:

$$\mathbb{E}\|f - \hat{f}_L\|_2^2 \leq R^2 L^{-2s} + C \frac{2L+1}{n}.$$

It is clear that an optimal value for  $L$  is of order  $n^{1/(2s+1)}$  but this value is unknown. We rather choose a data-driven method to select  $L$ . We introduce a classical minimization of a penalized empirical risk. Set

$$\hat{L} = \operatorname{argmin}_{L \in \mathcal{L}} \left\{ - \sum_{l=-L}^L |\widehat{f}^{\star l}|^2 + \lambda \frac{2L+1}{n} \right\} \quad (4.1)$$

where  $\mathcal{L}$  is a finite set of resolution level, and  $\lambda$  a constant to be specified later. The next theorem states an oracle inequality which highlights the bias variance decomposition of the quadratic risk and justifies our estimation procedure.

**Theorem 6.** *Assume Assumption 5 and Assumption 1. Assume that  $f$  belongs to the Sobolev ellipsoid  $W(s, R)$  with  $s > 1/2$ . Let  $\mathcal{L} = \{1, \dots, \lfloor (n-1)/2 \rfloor\}$  and  $\epsilon > 0$ . If the penalty constant verifies  $\lambda > (1 + \epsilon^{-1})(1 - 2P)^{-2}$  then,*

$$\mathbb{E}\|\hat{f}_{\hat{L}} - f\|_2^2 \leq (1 + 2\epsilon) \mathbb{E} \min_{L \in \mathcal{L}} \left\{ \|\hat{f}_L - f\|_2^2 + 2\lambda \frac{2L+1}{n} \right\} + \frac{C}{n}$$

where  $C$  is a positive constant depending on  $\epsilon, \|f\|_2, P, \theta_0, R, s$ . Moreover

$$\mathbb{E}\|\hat{f}_{\hat{L}} - f\|_2^2 = O\left(n^{-2s/(2s+1)}\right).$$

In consequence our estimator has a quadratic risk in  $n^{-2s/(2s+1)}$  which is known to be the optimal rate of convergence for estimating a density with smoothness  $s$ .

**Remark 1.** *The proof of the oracle inequality stated in Theorem 6 works for any  $\mathcal{L} \subset \mathbb{N}$  with  $\#\mathcal{L}$  not larger than a power of  $n$  and  $\max(\mathcal{L}) \rightarrow \infty$ . The rate of convergence is obtained if  $\mathcal{L}$  contains  $cn^{1/(2s+1)}$ . Note that the penalty only depends on  $P$  which is some safety margin around  $1/2$ , that can be chosen by the statistician. For the practical choice of the penalty, see Section 5.*

Eventually, note that some densities may be *supersmooth*, in the following sense:

$$\sum_{l \in \mathbb{Z}} \exp(2b|l|^r) |f^{\star l}|^2 \leq R^2.$$

In this case, the quadratic bias is bounded by  $R^2 \exp(-2bL^r)$  which gives the following fast rate of convergence:

$$\mathbb{E}\|\hat{f}_{\hat{L}} - f\|_2^2 = O\left(\frac{(\log n)^{1/r}}{n}\right).$$

## 5 Numerical results

All computations are performed with Matlab software and the Optimization Toolbox.

We shall implement our statistical procedure to both estimate the parameter  $\theta_0$  and the density  $f$ . We consider three popular circular densities namely, the von Mises density, the wrapped Cauchy and the wrapped normal densities. We remind their expression (see Ley and Verdebout (2017)). The von Mises density is given by:

$$f_{VM}(x) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x-\mu)},$$

with  $\kappa \geq 0$ ,  $I_0(\kappa)$  the modified Bessel function of the first kind and of order 0. The wrapped Cauchy density has density:

$$f_{WC}(x) = \frac{1}{2\pi} \frac{1 - \gamma^2}{1 + \gamma^2 - 2\gamma \cos(x - \mu)},$$

with  $0 \leq \gamma \leq 1$ . The wrapped normal density expression is:

$$f_{WN}(x) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} e^{-\frac{(x-\mu+2k\pi)^2}{2\sigma^2}},$$

$\sigma > 0$ . For more clarity, we set  $\sigma^2 =: -2\log(\rho)$ . Hence, we have  $0 \leq \rho \leq 1$ .

All these densities are characterized by a concentration parameter  $\kappa$ ,  $\gamma$  or  $\rho$  and a location parameter  $\mu$ . Remind that values  $\kappa = 0$ ,  $\gamma = 0$  and  $\rho = 0$  correspond to the uniform density on the circle. To meet symmetry assumptions of Theorem 1, we consider in the sequel that the location parameter is set to  $\mu = 0$ .

First, let us focus on the parametric part. We set  $\theta_0 = (p_0, \alpha_0, \beta_0) = (\frac{1}{4}, \frac{\pi}{8}, \frac{2\pi}{3})$ . Obtaining the estimate  $\hat{\theta}_n$  of  $\theta_0$  (see (3.2)) requires to solve a nonlinear minimization problem. To this end, we resort to the function *fmincon* of the Matlab Optimization toolbox. The function *fmincon* finds a constrained minimum of a function of several variables. Two parameters are to be specified: the domain over which the minimum is searched and an initial value. We consider the domain  $\{(0, \frac{1}{2}) \times [0, \pi) \times [0, \pi)\}$ . For more stability and to avoid possible local minimums, we perform the procedure over 10 initials values uniformly drawn on  $\{(0, \frac{1}{2}) \times [0, \pi) \times [0, \pi)\}$ . The final estimator  $\hat{\theta}_n$  corresponds to the minimum value of the empirical contrast  $S_n(\theta)$  given in (3.1) over the 10 trials.

Table 1 gathers mean squared errors for our estimation procedure. When analyzing Table 1, one clearly sees that increasing the number of observations improves noticeably the performances. As expected, von Mises densities with smaller concentration parameter are more difficult to estimate. Nonetheless, the overall performances are satisfying. Table 2 displays the performances of the method-of-moments estimation procedure developed by Spurr and Koutbeiy (1991) to handle the problem of estimating the parameters in mixtures of von Mises distributions. In bold, in Table 1, we point out cases for which our procedure outperforms the Spurr and Koutbeiy one. At closer inspection of Tables 1 and 2, for the von Mises density with  $\kappa = 2$  and  $n = 1000$ , the Spurr and Koutbeiy procedure is better. Otherwise, the performances of the two procedures are pretty similar and good. It is worth noticing that the method in Spurr and Koutbeiy (1991) is completely parametric and takes advantage of the knowledge of the distributions. In this regard, our procedure which is semiparametric is competitive with a parametric method.

Figure 2 illustrates the asymptotic normality of our estimator  $\hat{\theta}_n$  stated in Theorem 5.

density	$n = 100$			$n = 1000$		
	$p$	$\alpha$	$\beta$	$p$	$\alpha$	$\beta$
$f_{VM}, \kappa = 2$	<b>0.0121</b>	0.6848	<b>0.1131</b>	<b>0.0017</b>	0.1919	0.0238
$f_{VM}, \kappa = 5$	<b>0.0030</b>	<b>0.0285</b>	0.0049	<b>1.4632e-04</b>	<b>0.0017</b>	<b>4.4861e-04</b>
$f_{VM}, \kappa = 7$	0.0033	0.0133	0.0031	<b>1.6721e-04</b>	<b>0.0013</b>	<b>3.0102e-04</b>
$f_{WC}, \rho = 0.8$	0.0029	0.0124	0.0024	2.0788e-04	8.5435e-04	1.8942e-04
$f_{WN}, \rho = 0.8$	0.0077	0.1679	0.0457	0.0020	0.0238	0.0037

Table 1: Mean squared errors for estimating parameter  $\theta_0$  over 50 Monte Carlo replications.

density	$n = 100$			$n = 1000$		
	$p$	$\alpha$	$\beta$	$p$	$\alpha$	$\beta$
$f_{VM}, \kappa = 2$	0.0938	0.4212	0.1171	0.0062	0.0685	0.0062
$f_{VM}, \kappa = 5$	0.0031	0.0360	0.0049	2.9965e-04	0.0025	6.6273e-04
$f_{VM}, \kappa = 7$	0.0031	0.0084	0.0029	2.4553e-04	0.0014	3.5541e-04

Table 2: Spurr and Koutbeiy procedure: mean squared errors for estimating parameter  $\theta_0$  over 50 Monte Carlo replications.

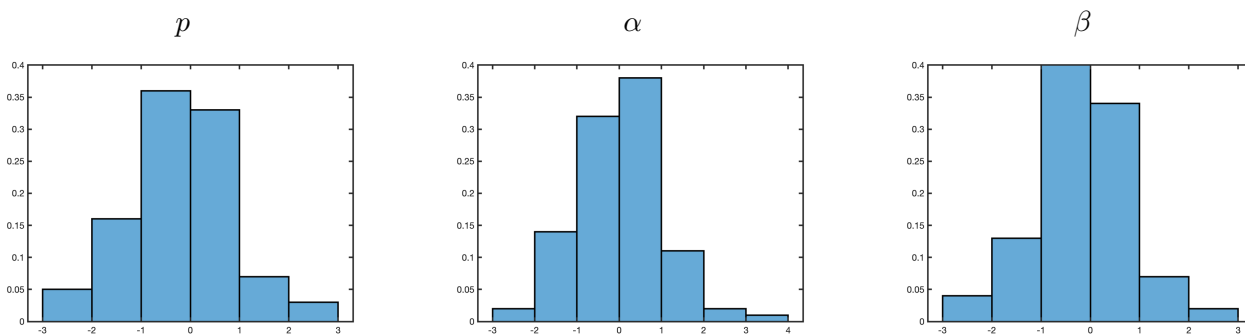


Figure 2: Histograms of the centered and standardized statistics  $\hat{\theta}_n$  for the von Mises density  $f_{VM}$  with  $\kappa = 5$ ,  $n = 1000$  and 100 Monte Carlo replications.

Now, let us turn to the nonparametric estimation part namely the estimation of the density  $f$ . The estimator of  $f$  is given by  $\hat{f}_{\hat{L}}$  (see Theorem 6). It requires the computation of a data-driven resolution level choice  $\hat{L}$  (given in (4.1)) which implies a tuning parameter  $\lambda$ . To select the proper  $\lambda$ , we follow the data-driven slope estimation approach due to Birgé and Massart (see Birgé and Massart (2001) and Birgé and Massart (2007)) and whose an overview in practice is presented in Baudry et al. (2012). To implement the slope heuristics method, one has to plot for  $L = 1$  to  $L_{\max}$ , the couples of points  $(\frac{2L+1}{n}, \sum_{l=-L}^L |\hat{f}^{*l}|^2)$ . For  $L \geq L_0$ , one should observe a linear behaviour (see Figure 3). Then, once the slope is estimated, say  $a$ , by a linear regression method, one eventually

takes  $\hat{\lambda} = 2a$  and the final resolution level is:

$$\hat{L} = \operatorname{argmin}_{L \in \mathcal{L}} \left\{ - \sum_{l=-L}^L |\widehat{f^{\star l}}|^2 + \hat{\lambda} \frac{2L+1}{n} \right\}.$$

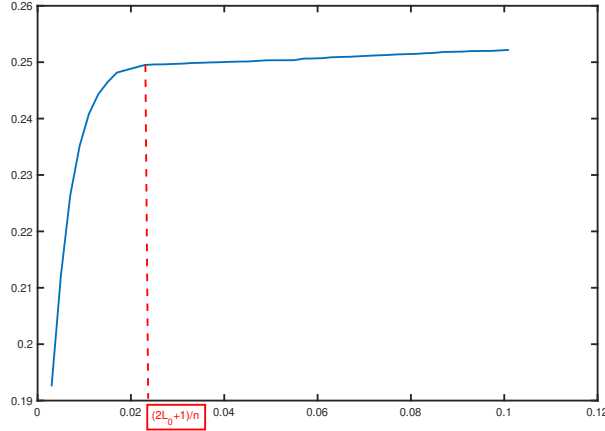


Figure 3: For the wrapped Cauchy density  $f_{WC}$  with  $\gamma = 0.8$  and  $n = 1000$ : plot of couples  $(\frac{2L+1}{n}, \sum_{l=-L}^L |\widehat{f^{\star l}}|^2)$  for  $L = \{1, \dots, 50\}$ .

Finally, Figure 4 shows reconstructions of the density  $f$  and the mixture density  $g$  as well. The estimates are good.

## 6 Proofs

### 6.1 Proof of Theorem 1 (identifiability)

Denote

$$M^l(\theta) := pe^{-i\alpha l} + (1-p)e^{-i\beta l}.$$

Suppose  $pf(x - \alpha) + (1-p)f(x - \beta) = p'f'(x - \alpha') + (1-p')f'(x - \beta')$ . The calculation of the Fourier coefficients gives, for all  $l \in \mathbb{Z}$ ,  $f^{\star l} M^l(\theta) = (f')^{\star l} M^l(\theta')$  which implies

$$f^{\star l} |M^l(\theta)|^2 = (f')^{\star l} M^l(\theta') \overline{M^l(\theta)}.$$

Then, our assumptions on  $f$  and  $f'$  entail

$$M^l(\theta') \overline{M^l(\theta)} \text{ is real } \quad \forall l \in \{1, \dots, 4\}.$$

Let us now study the consequence of this fact. Denote

$$\gamma_1 = \alpha' - \beta, \quad \gamma_2 = \alpha' - \alpha, \quad \gamma_3 = \beta' - \beta, \quad \gamma_4 = \beta' - \alpha$$

the 4 angles. Denote also the associated weights in  $(0, 1)$ :

$$\lambda_1 = p'(1-p), \quad \lambda_2 = p'p, \quad \lambda_3 = (1-p')(1-p), \quad \lambda_4 = p(1-p').$$

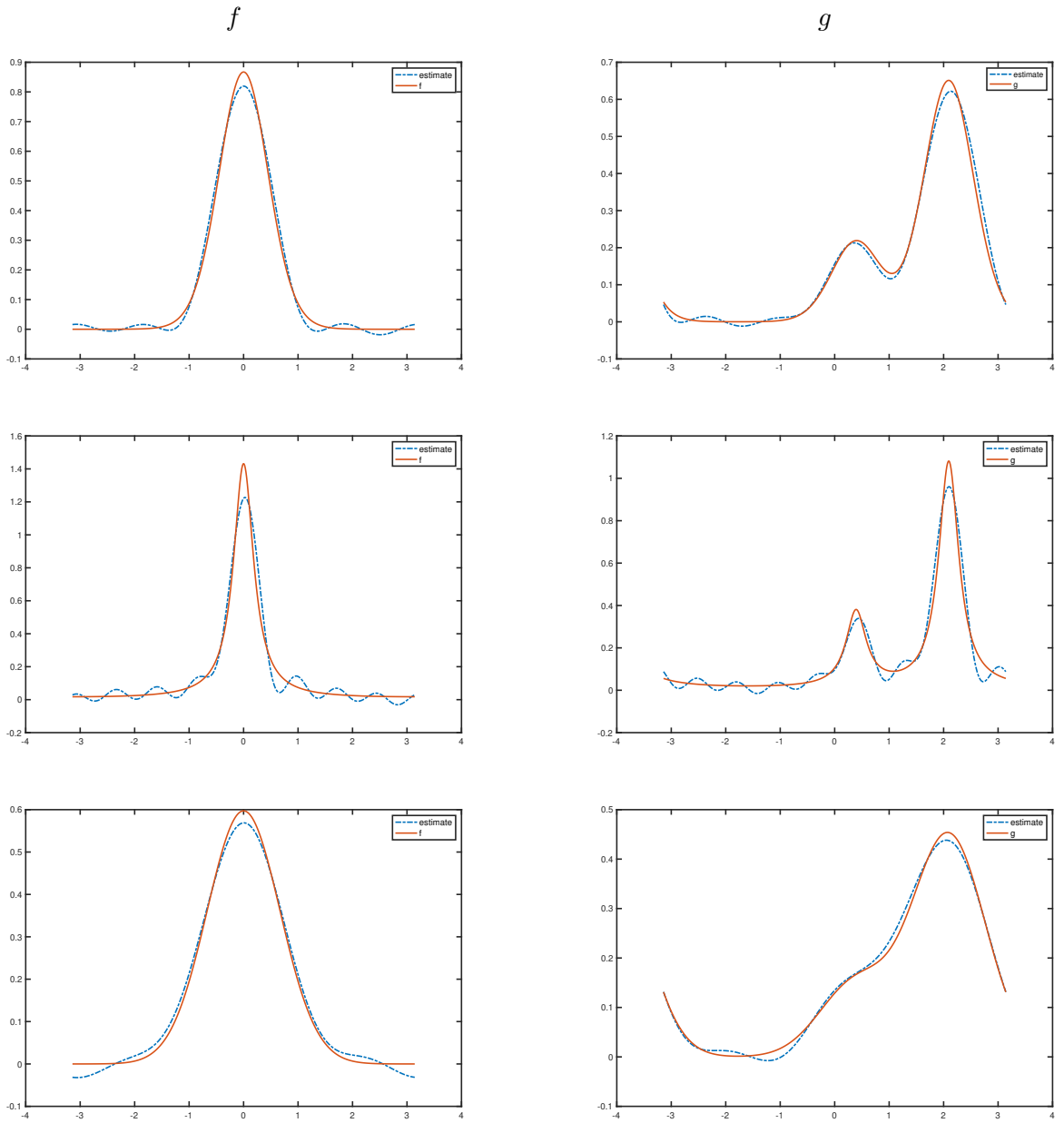


Figure 4: Estimation of the density  $f$  and the mixture density  $g$  for  $n = 1000$ . From top to bottom: the von Mises density with  $\kappa = 5$ , the wrapped Cauchy with  $\gamma = 0.8$  and the wrapped normal density with  $\rho = 0.8$ .

With this notation

$$M^l(\theta') \overline{M^l(\theta)} = \lambda_1 e^{-i\gamma_1 l} + \lambda_2 e^{-i\gamma_2 l} + \lambda_3 e^{-i\gamma_3 l} + \lambda_4 e^{-i\gamma_4 l}.$$

Then  $M^l(\theta')\overline{M^l(\theta)}$  is real if and only if  $\sum_{k=1}^4 \lambda_k \sin(l\gamma_k) = 0$  and we have to solve the equations

$$\forall l = 1, 2, 3, 4, \quad \sum_{k=1}^4 \lambda_k \sin(l\gamma_k) = 0. \quad (6.1)$$

This system of equations is studied in Lemmas 7 and 8 below.

Let us now reason with the representatives of the  $\gamma_k$  in  $(-\pi, \pi]$ . Lemma 8 says that the possible values for the  $\gamma_k$ 's are  $0, \pi, \gamma, -\gamma$ , for some  $\gamma \in (0, \pi)$ . Note that here

$$\gamma_1 - \gamma_2 = \gamma_3 - \gamma_4 = \alpha - \beta \neq 0 \quad \text{and} \quad \gamma_1 - \gamma_3 = \gamma_2 - \gamma_4 = \alpha' - \beta' \neq 0 \quad (6.2)$$

and then the  $\gamma_k$ 's take at least 2 different values: either 4 different values; or  $\gamma_2 = \gamma_3$  and the other distinct; or  $\gamma_1 = \gamma_4$  and the other distinct; or  $\gamma_2 = \gamma_3$  and  $\gamma_1 = \gamma_4$ .

• Let us first study the case where all the  $\gamma_k$ 's are distinct. There are  $4! = 24$  ways of having  $(\gamma_{i_1}, \gamma_{i_2}, \gamma_{i_3}, \gamma_{i_4}) = (-\gamma, 0, \gamma, \pi)$ . But 16 combinations lead to  $p = 1/2$  or  $p' = 1/2$ . For example, if  $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (-\gamma, 0, \gamma, \pi)$  then (6.1) becomes

$$\lambda_1 \sin(-l\gamma) + \lambda_2 \sin(0) + \lambda_3 \sin(l\gamma) + \lambda_4 \sin(l\pi) = 0.$$

Thus  $\lambda_1 = \lambda_3$ , which gives  $p' = 1/2$ . In the same way, there are 4 possibilities giving  $\lambda_1 = \lambda_3$ , 4 possibilities giving  $\lambda_1 = \lambda_2$ , 4 possibilities giving  $\lambda_2 = \lambda_4$ , 4 possibilities giving  $\lambda_3 = \lambda_4$ . All of this is impossible, since  $p \in (0, 1)$  and  $p \neq 1/2$ . In addition, in the 4 cases where  $\gamma_1 = -\gamma_4$ , we obtain via (6.2)  $\gamma_3 = -\gamma_2$  which is impossible if  $\{\gamma_2, \gamma_3\} = \{0, \pi\}$ . Idem if  $\gamma_2 = -\gamma_3$  and  $\{\gamma_1, \gamma_4\} = \{0, \pi\}$ . Thus it is finally impossible that all the  $\gamma_k$ 's are distinct.

• Let us now study the case where the  $\gamma_k$ 's take 3 distinct values ( $\gamma_2 = \gamma_3$  or  $\gamma_1 = \gamma_4$ ) and belong to  $\{0, \pi, \gamma\}$  or  $\{0, \pi, -\gamma\}$ . In the case where  $\gamma_2 = \gamma_3$ , coming back to equation (6.1), we understand that all the rearrangements lead to  $\lambda_4 = 0$  or  $\lambda_1 = 0$  or  $\lambda_2 + \lambda_3 = 0$ , which is impossible. In the same way, if  $\gamma_1 = \gamma_4$ , equation (6.1) leads to  $\lambda_2 = 0$  or  $\lambda_3 = 0$  or  $\lambda_1 + \lambda_4 = 0$ , which is impossible.

• The next case is when the  $\gamma_k$ 's take 3 distinct values and belong to  $\{0, \gamma, -\gamma\}$  or  $\{\pi, \gamma, -\gamma\}$ . If  $\gamma_2 = \gamma_3$ , we can then list the 6 cases:

$\gamma_1$	$\gamma_2 = \gamma_3$	$\gamma_4$	consequence
$-\gamma$	$0/\pi$	$\gamma$	$p = p', \alpha' = \alpha, \beta' = \beta \pmod{\pi}$
$\gamma$	$0/\pi$	$-\gamma$	$p = p', \alpha' = \alpha, \beta' = \beta \pmod{\pi}$
$-\gamma$	$\gamma$	$0/\pi$	$\lambda_1 = \lambda_2 + \lambda_3$
$\gamma$	$-\gamma$	$0/\pi$	$\lambda_1 = \lambda_2 + \lambda_3$
$0/\pi$	$\gamma$	$-\gamma$	$\lambda_4 = \lambda_2 + \lambda_3$
$0/\pi$	$-\gamma$	$\gamma$	$\lambda_4 = \lambda_2 + \lambda_3$

Note that  $\lambda_1 = \lambda_2 + \lambda_3 \Leftrightarrow p'(2 - 3p) = 1 - p$ , which is possible only if  $p < 1/2$  and  $p' > 1/2$  (recall that we suppose  $p < 1/2$  and  $p' < 1/2$ ). In the same way  $\lambda_4 = \lambda_2 + \lambda_3 \Leftrightarrow p'(1 - 3p) = 1 - 2p$ , which is possible only if  $p > 1/2$  and  $p' < 1/2$ .

Finally, if  $\gamma_1 = \gamma_4$ , we have the 6 last cases:

$\gamma_2$	$\gamma_1/\gamma_4$	$\gamma_3$	consequence
$-\gamma$	$0/\pi$	$\gamma$	$p' = 1 - p$
$\gamma$	$0/\pi$	$-\gamma$	$p' = 1 - p,$
$-\gamma$	$\gamma$	$0/\pi$	$p' = \frac{p}{3p-1}$
$\gamma$	$-\gamma$	$0/\pi$	$p' = \frac{p}{3p-1}$
$0/\pi$	$\gamma$	$-\gamma$	$p' = \frac{1-2p}{2-3p}, \beta - \alpha = 2\pi/3, \alpha' = \alpha + \pi, \beta' = \alpha - \pi/3$
$0/\pi$	$-\gamma$	$\gamma$	$p' = \frac{1-2p}{2-3p}, \beta - \alpha = -2\pi/3, \alpha' = \alpha + \pi, \beta' = \alpha + \pi/3$

Note that  $p' = p/(3p-1) > 1/2$  if  $p < 1/2$ .

• The last case occurs when the  $\gamma_k$ 's take 2 distinct values. If the  $\gamma_k$ 's take exactly 2 different values, necessarily

$$\gamma_1 = \gamma_4 \pmod{2\pi} \text{ and } \gamma_2 = \gamma_3 \pmod{2\pi} \Rightarrow 0 = \gamma_1 - \gamma_4 + \gamma_3 - \gamma_2 = 2(\alpha - \beta) \pmod{2\pi}$$

which is possible only if  $\alpha - \beta = \pi \pmod{2\pi}$  (recall that  $\alpha - \beta$  is always assumed  $\neq 0$ ). And in the same way  $\alpha' - \beta' = \pi \pmod{2\pi}$ . Then  $\gamma_1 - \gamma_2 = \alpha - \beta = \pi \pmod{2\pi}$ . Thus the two different values of the  $\gamma_k$  are at a distant of  $\pi$ . The first possibility is that these two values are 0 and  $\pi$ , which corresponds to the first case of Lemma 8. There are two subcases: 1a.  $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (\pi, 0, 0, \pi)$  or 1b.  $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (0, \pi, \pi, 0)$ . In the subcase 1a.  $(\alpha', \beta') = (\alpha, \beta)$ . Equations

$$\begin{cases} pf + (1-p)f_\pi = p'f' + (1-p')f'_\pi \\ pf_\pi + (1-p)f = p'f'_\pi + (1-p')f' \end{cases}$$

entails that  $f'$  is a linear combination of  $f$  and  $f_\pi$ . In the subcase 1b.  $(\alpha', \beta') = (\alpha + \pi, \beta + \pi) = (\beta, \alpha)$ .

The second possibility is that the two distinct values  $\gamma_1 = \gamma_4$  and  $\gamma_2 = \gamma_3$  are not multiples of  $\pi$ , which corresponds to the fourth case of Lemma 8. Then  $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (\gamma_1, -\gamma_1, -\gamma_1, \gamma_1)$  and

$$\gamma_1 - (-\gamma_1) = \gamma_1 - \gamma_2 = \pi \pmod{2\pi}$$

which entails  $\gamma_1 = \pi/2 \pmod{\pi}$ . Equation (6.1) becomes

$$(\lambda_1 - \lambda_2 - \lambda_3 + \lambda_4) \sin(l\pi/2) = 0$$

so that  $\lambda_1 + \lambda_4 = \lambda_2 + \lambda_3$ , which gives

$$p'(1-p) + p(1-p') = p'p + (1-p')(1-p) \Rightarrow p' + p - 2pp' = 1/2 \Rightarrow p' = 1/2$$

which is impossible.

• Let us recap the only possible cases that we have obtained:

- ▷  $p = p', \alpha' = \alpha, \beta' = \beta \pmod{\pi}$ ,
- ▷  $p' = \frac{1-2p}{2-3p}, \beta - \alpha = 2\pi/3, \alpha' = \alpha + \pi, \beta' = \alpha - \pi/3$ ,
- ▷  $p' = \frac{1-2p}{2-3p}, \beta - \alpha = -2\pi/3, \alpha' = \alpha + \pi, \beta' = \alpha + \pi/3$ ,
- ▷  $\alpha - \beta = \pi, (\alpha', \beta') = (\alpha, \beta)$  or  $(\alpha', \beta') = (\beta, \alpha)$ .

This completes the proof of the theorem.

**Lemma 7.** *Let  $\gamma_1, \dots, \gamma_4$  be four reals. Let  $A$  be the matrix  $(\sin(i\gamma_j))_{1 \leq i, j \leq 4}$ . Then*

$$\det A = 64 \prod_{k=1}^4 \sin(\gamma_k) \prod_{1 \leq i < j \leq 4} (\cos(\gamma_i) - \cos(\gamma_j)).$$

*Proof.* From matrix  $A$ , doing line modification  $L_3 \leftarrow L_3 - L_1$ , and  $L_4 \leftarrow L_4 - L_2$ , we obtain (recall that  $\sin(2p) = 2 \sin(p) \cos(p)$  and  $\sin(p) - \sin(q) = 2 \sin(\frac{p-q}{2}) \cos(\frac{p+q}{2})$ )

$$\det A = \begin{vmatrix} \sin(\gamma_1) & \sin(\gamma_2) & \sin(\gamma_3) & \sin(\gamma_4) \\ 2 \sin(\gamma_1) \cos(\gamma_1) & 2 \sin(\gamma_2) \cos(\gamma_2) & 2 \sin(\gamma_3) \cos(\gamma_3) & 2 \sin(\gamma_4) \cos(\gamma_4) \\ 2 \sin(\gamma_1) \cos(2\gamma_1) & 2 \sin(\gamma_2) \cos(2\gamma_2) & 2 \sin(\gamma_3) \cos(2\gamma_3) & 2 \sin(\gamma_4) \cos(2\gamma_4) \\ 2 \sin(\gamma_1) \cos(3\gamma_1) & 2 \sin(\gamma_2) \cos(3\gamma_2) & 2 \sin(\gamma_3) \cos(3\gamma_3) & 2 \sin(\gamma_4) \cos(3\gamma_4) \end{vmatrix}.$$



Using 4-linearity of the determinant:

$$\det A = 8 \left( \prod_{j=1}^4 \sin(\gamma_j) \right) \begin{vmatrix} 1 & 1 & 1 & 1 \\ \cos(\gamma_1) & \cos(\gamma_2) & \cos(\gamma_3) & \cos(\gamma_4) \\ \cos(2\gamma_1) & \cos(2\gamma_2) & \cos(2\gamma_3) & \cos(2\gamma_4) \\ \cos(3\gamma_1) & \cos(3\gamma_2) & \cos(3\gamma_3) & \cos(3\gamma_4) \end{vmatrix}.$$

Now, denote  $x_k = \cos(\gamma_k)$  and remark that  $\cos(i\gamma_k) = T_i(\cos \gamma_k) = T_i(x_k)$  where  $T_i$  is the  $i$ th Chebyshev polynomial:  $T_0 = 1, T_1 = X, T_2 = 2X^2 - 1, T_3 = 4X^3 - 3X$ . We have  $T_2 + T_0 = 2X^2$  and  $T_3 + 3T_1 = 4X^3$ . Then, doing  $L_3 \leftarrow L_3 + L_1$ , and  $L_4 \leftarrow L_4 + 3L_2$ :

$$\det A = 8 \left( \prod_{j=1}^4 \sin(\gamma_j) \right) \begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ 2x_1^2 & 2x_2^2 & 2x_3^2 & 2x_4^2 \\ 4x_1^3 & 4x_2^3 & 4x_3^3 & 4x_4^3 \end{vmatrix} = 64 \left( \prod_{j=1}^4 \sin(\gamma_j) \right) \begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ x_1^2 & x_2^2 & x_3^2 & x_4^2 \\ x_1^3 & x_2^3 & x_3^3 & x_4^3 \end{vmatrix}$$

This is a Vandermonde matrix, hence

$$\det A = 64 \left( \prod_{j=1}^4 \sin(\gamma_j) \right) \prod_{1 \leq i < j \leq 4} (x_i - x_j) = 64 \prod_{k=1}^4 \sin(\gamma_k) \prod_{1 \leq i < j \leq 4} (\cos(\gamma_i) - \cos(\gamma_j)).$$

□

**Lemma 8.** *Let  $\gamma_1, \dots, \gamma_4$  be four reals. Let  $\lambda_1, \dots, \lambda_4 \in \mathbb{R} \setminus \{0\}$  such that*

$$\sum_{k=1}^4 \lambda_k \sin(l\gamma_k) = 0, \quad l = 1, \dots, 4. \quad (6.3)$$

*Then, one of the following cases holds:*

1. *All  $\gamma_k$  are multiples of  $\pi$ .*
2. *Exactly two  $\gamma_k$  are multiples of  $\pi$ :  $\gamma_{i_1} = \gamma_{i_2} = 0 \pmod{\pi}$  and  $\gamma_{i_3} = \pm\gamma_{i_4} \pmod{2\pi}$ .*
3. *Only one  $\gamma_k$  is multiple of  $\pi$ :  $\gamma_{i_1} = 0 \pmod{\pi}$  and  $\gamma_{i_2} = \pm\gamma_{i_3} = \pm\gamma_{i_4} \pmod{2\pi}$ .*
4. *No  $\gamma_k$  is multiple of  $\pi$  and  $\gamma_1 = \pm\gamma_2 = \pm\gamma_3 = \pm\gamma_4 \pmod{2\pi}$ .*

*Proof.* First observe that, since  $\sum_{k=1}^4 \lambda_k \sin(l\gamma_k) = 0$  with  $\lambda \neq 0_{\mathbb{R}^4}$ , necessarily  $\det(A) = 0$  where  $A = (\sin(i\gamma_j))_{1 \leq i, j \leq 4}$ . Using Lemma 7

$$\prod_{k=1}^4 \sin(\gamma_k) \prod_{1 \leq i < j \leq 4} (\cos(\gamma_i) - \cos(\gamma_j)) = 0. \quad (6.4)$$

Now, let us study the various cases that make this quantity vanish.

For the first case, note that if three  $\gamma_k$  are multiples of  $\pi$ :  $\gamma_{i_1} = \gamma_{i_2} = \gamma_{i_3} = 0 \pmod{\pi}$  then equation (6.3) becomes  $\lambda_{i_4} \sin(l\gamma_{i_4}) = 0$  and the last angle is also null modulo  $\pi$ .

In case 2., equation (6.3) entails

$$\lambda_{i_3} \sin(l\gamma_{i_3}) + \lambda_{i_4} \sin(l\gamma_{i_4}) = 0, \quad l = 1, 2$$

with  $\gamma_{i_3} \not\equiv 0 \pmod{\pi}$ ,  $\gamma_{i_4} \not\equiv 0 \pmod{\pi}$ . Then, since  $(\lambda_{i_3}, \lambda_{i_4}) \neq (0, 0)$ ,

$$0 = \begin{vmatrix} \sin(\gamma_{i_3}) & \sin(\gamma_{i_4}) \\ \sin(2\gamma_{i_3}) & \sin(2\gamma_{i_4}) \end{vmatrix} = 2 \sin(\gamma_{i_3}) \sin(\gamma_{i_4}) (\cos(\gamma_{i_4}) - \cos(\gamma_{i_3})).$$

Then  $\cos(\gamma_{i_3}) = \cos(\gamma_{i_4})$ . Either  $\gamma_{i_3} = \gamma_{i_4} \pmod{2\pi}$ , or  $\gamma_{i_3} = -\gamma_{i_4} \pmod{2\pi}$ .

Let us now study case 3. For the sake of simplicity we assume that  $\gamma_4 = 0 \pmod{\pi}$  and  $\gamma_k \not\equiv 0 \pmod{\pi}$  for  $k = 1, 2, 3$ . Equation (6.3) gives

$$\lambda_1 \sin(l\gamma_1) + \lambda_2 \sin(l\gamma_2) + \lambda_3 \sin(l\gamma_3) = 0, \quad l = 1, 2, 3.$$

With the same proof as Lemma 7, we obtain

$$\prod_{k=1}^3 \sin(\gamma_k) \prod_{1 \leq i < j \leq 3} (\cos(\gamma_i) - \cos(\gamma_j)) = 0.$$

Then  $\gamma_1 = \pm\gamma_2 \pmod{2\pi}$  or  $\gamma_1 = \pm\gamma_3 \pmod{2\pi}$  or  $\gamma_2 = \pm\gamma_3 \pmod{2\pi}$ . Moreover, if, for example,  $\gamma_1 = \pm\gamma_2 \pmod{2\pi}$  then

$$(\lambda_1 \pm \lambda_2) \sin(l\gamma_1) + \lambda_3 \sin(l\gamma_3) = 0, \quad l = 1, 2$$

We are reduced to the previous case, then  $\gamma_1 = \pm\gamma_3 \pmod{2\pi}$ .

In the case 4., equation (6.4) becomes  $\prod_{1 \leq i < j \leq 4} (\cos(\gamma_i) - \cos(\gamma_j)) = 0$ , which provides 6 possible equalities. Assume, for example,  $\cos(\gamma_1) - \cos(\gamma_2) = 0$  and consequently  $\gamma_1 = \pm\gamma_2 \pmod{2\pi}$ . Then

$$(\lambda_1 \pm \lambda_2) \sin(l\gamma_1) + \lambda_3 \sin(l\gamma_3) + \lambda_4 \sin(l\gamma_4) = 0, \quad l = 1, 2, 3.$$

Reasoning as in previous case,  $\gamma_1 = \pm\gamma_3 = \pm\gamma_4 \pmod{2\pi}$ . □

## 6.2 Proof of Theorem 4 (consistency)

This proof and the following are inspired from Butucea and Vandekerkhove (2014). Let us denote  $\tilde{\Theta} = (0, 1/2) \times \mathbb{S}^1 \times \mathbb{S}^1$ . Denote  $\dot{\phi}(\theta)$  the gradient of any function  $\phi$  with respect to  $\theta = (p, \alpha, \beta)$ , and  $\ddot{\phi}(\theta)$  the Hessian matrix.

The proof of Theorem 4 relies on some preliminary results, given in the sequel.

**Proposition 9.** *Under Assumption 3 the contrast function  $S$  verifies the following properties:  $S(\theta) \geq 0$ , and  $S(\theta) = 0$  if and only if  $\theta = \theta_0$  or  $\theta = \theta_0 + \pi$ .*

*Proof.* It is clear that  $S(\theta) \geq 0$  and that

$$S(\theta_0) = \sum_{l=-4}^4 \left( \Im \left( g^{*l} \overline{M^l(\theta_0)} \right) \right)^2 = \sum_{l=-4}^4 \left( \Im \left( f^{*l} |M^l(\theta_0)|^2 \right) \right)^2 = 0.$$

By Lemma 3, if  $\theta \neq \theta_0 \pmod{\pi}$ , there exists  $l_1 \in \{1, \dots, 4\}$  such that  $\Im \left( M^{l_1}(\theta_0) \overline{M^{l_1}(\theta)} \right) \neq 0$  so that  $S(\theta) \geq \left( \Im \left( g^{*l_1} \overline{M^{l_1}(\theta)} \right) \right)^2 > 0$ . □

**Lemma 10.** 1. For all  $\theta$  in  $\tilde{\Theta}$ ,  $|M^l(\theta)| \leq 1$ .

2. For all  $1 \leq k \leq n$ , for all  $l$  in  $\mathbb{Z}$ ,

$$\sup_{\theta \in \tilde{\Theta}} |Z_k^l(\theta)| \leq \frac{1}{2\pi}, \quad \sup_{\theta \in \tilde{\Theta}} |J^l(\theta)| \leq \frac{1}{2\pi}.$$

3. For all  $1 \leq k \leq n$ , for all  $l$  in  $\mathbb{Z}$ ,

$$\sup_{\theta \in \tilde{\Theta}} \|\dot{Z}_k^l(\theta)\| \leq \frac{1+|l|}{\sqrt{2\pi}}, \quad \sup_{\theta \in \tilde{\Theta}} \|j^l(\theta)\| \leq \frac{1+|l|}{\sqrt{2\pi}}.$$

where  $\|\cdot\|$  is the Euclidean norm.

4. For all  $1 \leq k \leq n$ , for all  $l$  in  $\mathbb{Z}$ ,

$$\sup_{\theta \in \tilde{\Theta}} \|\ddot{Z}_k^l(\theta)\|_F \leq \frac{|l|+l^2}{\pi}, \quad \sup_{\theta \in \tilde{\Theta}} \|\ddot{j}^l(\theta)\|_F \leq \frac{|l|+l^2}{\pi}.$$

where  $\|\cdot\|_F$  is the Frobenius norm.

*Proof.* Point 1 is straightforward.

2. Let us start with  $Z_k^l(\theta)$ . We recall that  $Z_k^l(\theta) = \Im \left( \frac{e^{ilX_k}}{2\pi} M^l(\theta) \right)$ . Then

$$|Z_k^l(\theta)| \leq \frac{1}{2\pi} |M^l(\theta)| \leq \frac{1}{2\pi}.$$

Furthermore

$$|J^l(\theta)| \leq |g^{*l}| |M^l(\theta)| \leq \frac{1}{2\pi} \int_{\mathbb{S}_1} g \leq \frac{1}{2\pi}.$$

3. We have

$$\dot{Z}_k^l(\theta) = \frac{1}{2\pi} \Im \left( e^{ilX_k} \dot{M}^l(\theta) \right) = \frac{1}{2\pi} \Im \left( e^{ilX_k} \begin{pmatrix} e^{-il\alpha} - e^{-il\beta} \\ -ilpe^{-i\alpha l} \\ -il(1-p)e^{-i\beta l} \end{pmatrix} \right)$$

and

$$j^l(\theta) = \Im \left( \overline{g^{*l}} \dot{M}^l(\theta) \right) = \Im \left( \overline{g^{*l}} \begin{pmatrix} e^{-il\alpha} - e^{-il\beta} \\ -ilpe^{-i\alpha l} \\ -il(1-p)e^{-i\beta l} \end{pmatrix} \right).$$

We get

$$\|j^l(\theta)\| \leq \frac{1}{2\pi} (2 + p^2l^2 + (1-p)^2l^2)^{1/2} \leq \frac{1+|l|}{\sqrt{2\pi}}$$

and we have the same bound for  $\|\dot{Z}_k^l(\theta)\|$ .

4. We have

$$\begin{aligned} \ddot{Z}_k^l(\theta) &= \Im \left( \frac{e^{ilX_k}}{2\pi} \ddot{M}^l(\theta) \right) \\ &= \Im \left( \frac{e^{ilX_k}}{2\pi} \begin{pmatrix} 0 & -ile^{-il\alpha} & ile^{-il\beta} \\ -ile^{-il\alpha} & -l^2pe^{-il\alpha} & 0 \\ ile^{-il\beta} & 0 & -l^2(1-p)e^{-il\beta} \end{pmatrix} \right). \end{aligned}$$

Thus

$$\|\ddot{Z}_k^l\|_F \leq \frac{1}{2\pi} (4l^2 + l^4 p^2 + l^4 (1-p)^2)^{1/2} \leq \frac{|l| + l^2}{\pi}$$

This ends the proof of the lemma.  $\square$

**Lemma 11.** *There exists a numerical positive constant  $C$  such that the following inequalities hold.*

1. For all  $1 \leq k \leq n$ , for all  $l$  in  $\mathbb{Z}$

$$\forall \theta, \theta' \in \tilde{\Theta} \quad \|\dot{Z}_k^l(\theta) - \dot{Z}_k^l(\theta')\| \leq C \|\theta - \theta'\| (1 + |l| + l^2).$$

2. We also have

$$\|\ddot{Z}_k^l(\theta) - \ddot{Z}_k^l(\theta')\|_F \leq C \|\theta - \theta'\| (1 + |l| + l^2 + |l|^3).$$

*Proof.* We use Taylor expansions at first order and then apply same bounding techniques as in Lemma 10.  $\square$

**Lemma 12.** 1. The function  $S$  is Lipschitz continuous over  $\tilde{\Theta}$ .

2. The function  $S_n(\theta)$  is Lipschitz continuous over  $\tilde{\Theta}$ .

3. The function  $\ddot{S}_n(\theta)$  is Lipschitz continuous over  $\tilde{\Theta}$ .

*Proof.* We will write  $C$  for a constant that may change from line to line but is numerical.

Let us start with point 1. We recall that  $S(\theta) = \sum_l J^l(\theta)^2$ . Let  $\theta$  and  $\theta'$  in  $\tilde{\Theta}$ . As  $\tilde{\Theta}$  is a convex set, we get, thanks to the mean value theorem

$$\begin{aligned} |S(\theta) - S(\theta')| &= \left| \sum_{l=-4}^4 J^l(\theta)^2 - J^l(\theta')^2 \right| = \left| 2(\theta - \theta')^\top \sum_{l=-4}^4 J^l(\theta_u) J^l(\theta_u) \right| \\ &\leq C \|\theta - \theta'\| \sum_{l=-4}^4 (1 + |l|) \leq C \|\theta - \theta'\| \end{aligned}$$

with  $\theta_u$  lying between  $\theta$  and  $\theta'$ , and using Lemma 10.

Let us shift to point 2. Due to the mean value theorem, we have

$$\begin{aligned} |S_n(\theta) - S_n(\theta')| &= \left| \frac{1}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^4 \left( Z_k^l(\theta) Z_j^l(\theta) - Z_k^l(\theta') Z_j^l(\theta') \right) \right| \\ &= \left| \frac{1}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^4 \left( (\theta - \theta')^\top \nabla [Z_k^l(\theta) Z_j^l(\theta)] \Big|_{\theta=\theta_u} \right) \right| \\ &= \left| \frac{2(\theta - \theta')^\top}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^4 \dot{Z}_k^l(\theta_u) Z_j^l(\theta_u) \right|, \end{aligned}$$

with  $\theta_u$  lying between  $\theta$  and  $\theta'$ . Then using 1. and 2. of Lemma 10 we get

$$|S_n(\theta) - S_n(\theta')| \leq \frac{C \|\theta - \theta'\|}{n(n-1)} \sum_{i \neq j} \sum_{l=-4}^4 (1 + |l|) \leq C \|\theta - \theta'\|$$

which ends the proof of the second point.

Concerning point 3. we have that

$$\ddot{S}_n(\theta) = \frac{2}{n(n-1)} \sum_{k \neq j} \sum_{l=4}^4 (\ddot{Z}_k^l(\theta) Z_j^l(\theta) + \dot{Z}_k^l(\theta) \dot{Z}_j^l(\theta)^\top).$$

Hence

$$\begin{aligned} \|\ddot{S}_n(\theta) - \ddot{S}_n(\theta')\|_F &\leq \frac{2}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^4 \left( \|\ddot{Z}_k^l(\theta) - \ddot{Z}_k^l(\theta')\|_F \|Z_j^l(\theta)\|_F + \|\ddot{Z}_k^l(\theta')\|_F \|Z_j^l(\theta) - Z_j^l(\theta')\|_F \right. \\ &\quad \left. + \|\dot{Z}_k^l(\theta')\|_F \|\dot{Z}_j^l(\theta) - \dot{Z}_j^l(\theta')^\top\|_F + \|(\dot{Z}_k^l(\theta') - \dot{Z}_k^l(\theta))\dot{Z}_j^l(\theta)^\top\|_F \right) \end{aligned}$$

Using Taylor expansions and Lemma 10 and 11, we get that

$$\|\ddot{S}_n(\theta) - \ddot{S}_n(\theta')\|_F \leq C \|\theta - \theta'\| \sum_{l=-4}^4 (1 + |l| + l^2 + |l|^3).$$

□

**Proposition 13.** *There exist a positive constant  $C$  such that*

$$\sup_{\theta \in \tilde{\Theta}} \mathbb{E}[(S_n(\theta) - S(\theta))^2] \leq \frac{C}{n}.$$

*Proof.* The definitions of  $S_n$  and  $S$  provide

$$S_n(\theta) - S(\theta) = \mathbb{E} \left[ \left( \frac{1}{n(n-1)} \sum_{l=-4}^4 \sum_{k \neq j} (Z_k^l(\theta) Z_j^l(\theta) - J^l(\theta)^2) \right)^2 \right] = T_n + V_n$$

where

$$T_n = \frac{2}{n(n-1)} \sum_{l=-4}^4 \sum_{k < j} (Z_k^l(\theta) - J^l(\theta))(Z_j^l(\theta) - J^l(\theta))$$

and

$$V_n = \frac{2}{n} \sum_{l=-4}^4 \sum_{k=1}^n (Z_k^l(\theta) - J^l(\theta)) J^l(\theta).$$

Note that  $\mathbb{E}(Z_k^l(\theta) - J^l(\theta)) = 0$  which entails  $\mathbb{E}[T_n V_n] = 0$ . Then

$$\mathbb{E}[(S_n(\theta) - S(\theta))^2] = \mathbb{E}[(T_n + V_n)^2] = \mathbb{E}[T_n^2] + \mathbb{E}[V_n^2].$$

Now, since the variables  $\left( \sum_{l=-4}^4 (Z_k^l(\theta) - J^l(\theta))(Z_j^l(\theta) - J^l(\theta)) \right)_{k < j}$  are uncorrelated,

$$\begin{aligned} \mathbb{E}[T_n^2] &= \frac{2}{n(n-1)} \mathbb{E} \left[ \left( \sum_{l=-4}^4 (Z_1^l(\theta) - J^l(\theta))(Z_2^l(\theta) - J^l(\theta)) \right)^2 \right] \\ &\leq \frac{2}{n(n-1)} \mathbb{E} \left[ \left( \sum_{l=-4}^4 \frac{2}{2\pi} \cdot \frac{2}{2\pi} \right)^2 \right] \leq \frac{C}{2n} \end{aligned}$$

using Lemma 10. We focus now on  $V_n$ : in the same way

$$\begin{aligned}\mathbb{E}[V_n^2] &= \frac{4}{n} \mathbb{E} \left[ \left( \sum_{l=-4}^4 (Z_1^l(\theta) - J^l(\theta)) J^l(\theta) \right)^2 \right] \\ &\leq \frac{4}{n} \mathbb{E} \left[ \left( \sum_{l=-4}^4 Z_1^l(\theta) J^l(\theta) \right)^2 \right] \leq \frac{C}{2n},\end{aligned}$$

using Lemma 10 again.  $\square$

Theorem 4 is finally proved using the following lemma, its assumptions being ensured by Proposition 9, Lemma 12 and Proposition 13.

**Lemma 14.** *Assume that  $\Theta$  is a compact set and let  $S : \Theta \rightarrow \mathbb{R}$  be a continuous function. Assume that*

$$S(\theta) = \min_{\Theta} S \Leftrightarrow \theta = \theta_0 \text{ or } \theta = \theta'_0$$

where  $\theta_0, \theta'_0 \in \Theta$ . Let  $S_n : \Theta \rightarrow \mathbb{R}$  be a function which is uniformly continuous and such that for all  $\theta$   $|S_n(\theta) - S(\theta)|$  tends to 0 in probability. Let  $\hat{\theta}_n$  be a point such that  $S_n(\hat{\theta}_n) = \inf_{\Theta} S_n$ . Then  $\hat{\theta}_n \rightarrow \theta_0$  or  $\theta'_0$  in probability.

This is a classical result in the theory of minimum contrast estimators, when  $\theta_0 = \theta'_0$ . We reproduce the proof since it is slightly adapted to the case of two argmins.

*Proof.* Let  $\epsilon > 0$  and  $B$  be the union of the open ball with center  $\theta_0$  and radius  $\epsilon$  and the open ball with center  $\theta'_0$  and radius  $\epsilon$ . Since  $S$  is continuous and  $B^c \subset \Theta$  is a compact set, there exists  $\theta_\epsilon \in B^c$  such that  $S(\theta_\epsilon) = \inf_{B^c} S$ . Using the assumption, since  $\theta_\epsilon \neq \theta_0$ , and  $\theta_\epsilon \neq \theta'_0$

$$\delta := S(\theta_\epsilon) - S(\theta_0) > 0.$$

Since  $S_n$  is uniformly continuous, there exists  $\alpha > 0$  such that

$$\forall \theta, \theta' \quad \|\theta - \theta'\| < \alpha \Rightarrow |S_n(\theta) - S_n(\theta')| \leq \delta/2.$$

Moreover  $B^c$  is a compact set then there exists a finite set  $(\theta_i)$  such that  $B^c \subset \cup_{i=1}^I B(\theta_i, \alpha)$ . Denote  $\Delta_n := \max_{0 \leq i \leq I} |S_n(\theta) - S(\theta)|$ . The assumption ensures that  $\Delta_n$  tends to 0 in probability. Let  $\theta \in B^c$ . There exists  $1 \leq i \leq I$  such that  $\|\theta - \theta_i\| < \alpha$ , and then  $|S_n(\theta) - S_n(\theta_i)| \leq \delta/2$ . Thus

$$\begin{aligned}S_n(\theta) - S_n(\theta_0) &= (S_n(\theta) - S_n(\theta_i)) + (S_n(\theta_i) - S(\theta_i)) + (S(\theta_i) - S(\theta_0)) + (S(\theta_0) - S_n(\theta_0)) \\ &\geq -\delta/2 - \Delta_n + \delta - \Delta_n\end{aligned}$$

using that  $S(\theta_i) - S(\theta_0) \geq S(\theta_\epsilon) - S(\theta_0) = \delta$ . Then

$$\inf_{\theta \in B^c} S_n(\theta) - S_n(\theta_0) \geq \delta/2 - 2\Delta_n.$$

Now, if  $\|\hat{\theta}_n - \theta_0\| \geq \epsilon$  and  $\|\hat{\theta}_n - \theta'_0\| \geq \epsilon$  then  $\hat{\theta}_n \in B^c$  and

$$\inf_{\Theta} S_n = S_n(\hat{\theta}_n) = \inf_{B^c} S_n.$$

In particular  $\inf_{B^c} S_n \leq S_n(\theta_0)$  so that

$$\mathbb{P}(\|\hat{\theta}_n - \theta_0\| \geq \epsilon \text{ and } \|\hat{\theta}_n - \theta'_0\| \geq \epsilon) \leq \mathbb{P}(0 \geq \inf_{B^c} S_n - S_n(\theta_0) \geq \delta/2 - 2\Delta_n) \leq \mathbb{P}(\Delta_n \geq \delta/4) \rightarrow 0$$

since  $\Delta_n$  tends to 0 in probability.  $\square$

### 6.3 Proof of Theorem 5 (asymptotic normality)

The Taylor's theorem and the definition of  $\hat{\theta}_n$  give

$$\dot{S}_n(\hat{\theta}_n) = \dot{S}_n(\theta_0) + \ddot{S}_n(\theta_n^*)(\hat{\theta}_n - \theta_0) = 0,$$

where  $\theta_n^*$  lies in the line segment with extremities  $\theta_0$  and  $\hat{\theta}_n$ . Equivalently we have,

$$\ddot{S}_n(\theta_n^*)(\hat{\theta}_n - \theta_0) = -\dot{S}_n(\theta_0).$$

We recall that

$$S_n(\theta_0) = \frac{1}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^4 Z_k^l(\theta_0) Z_j^l(\theta_0)$$

and

$$\dot{S}_n(\theta_0) = \frac{2}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^4 \dot{Z}_k^l(\theta_0) Z_j^l(\theta_0)$$

and

$$\ddot{S}_n(\theta_0) = \frac{2}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^4 \ddot{Z}_k^l(\theta_0) Z_j^l(\theta_0) + \dot{Z}_k^l(\theta_0) \dot{Z}_j^l(\theta_0)^\top.$$

**Step 1-** Let us prove that

$$\sqrt{n} \dot{S}_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, V).$$

We remind that  $J^l(\theta_0) = 0$ . Hence

$$\mathbb{E}(\dot{S}_n(\theta_0)) = 2 \sum_{l=-4}^4 \dot{J}^l(\theta_0) J^l(\theta_0) = 0.$$

We can break down  $\dot{S}_n(\theta_0)$  in the following way:

$$\begin{aligned} \dot{S}_n(\theta_0) &= \frac{2}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^4 (\dot{Z}_k^l(\theta_0) - j^l(\theta_0) + j^l(\theta_0)) Z_j^l(\theta_0) \\ &= \frac{4}{n(n-1)} \sum_{k < j} \sum_{l=-4}^4 (\dot{Z}_k^l(\theta_0) - j^l(\theta_0)) Z_j^l(\theta_0) + \frac{2}{n} \sum_{j=1}^n \sum_{l=-4}^4 j^l(\theta_0) Z_j^l(\theta_0) \\ &=: A_n + B_n. \end{aligned}$$

Note that  $A_n$  and  $B_n$  are centered variables. Let us show that  $\sqrt{n}A_n = o_P(1)$ . Note that the variables  $W_{jk} := \left( \sum_{l=-4}^4 (\dot{Z}_k^l(\theta_0) - j^l(\theta_0)) Z_j^l(\theta_0) \right)_{k < j}$  are centered and uncorrelated. Then

$$\mathbb{E}(\|A_n\|^2) = \mathbb{E} \left( \left\| \frac{4}{n(n-1)} \sum_{k < j} W_{jk} \right\|^2 \right) = \frac{8}{n(n-1)} \mathbb{E} \|W_{12}\|^2.$$

Using Lemma 10, there exists  $C > 0$  such that

$$\|W_{12}\| \leq \sum_{l=-4}^4 \frac{2(1+|l|)}{\sqrt{2\pi}} \frac{1}{2\pi} \leq C$$

so that  $\mathbb{E}(\|\sqrt{n}A_n\|^2) \leq 8C^2/(n-1)$ . Finally, invoking Markov inequality we have that  $\sqrt{n}A_n = o_P(1)$ . We can write  $\sqrt{n}B_n$  in the following way:

$$\sqrt{n}B_n = \frac{2}{\sqrt{n}} \sum_{k=1}^n U_k(\theta_0),$$

where we set  $U_k(\theta_0) := \sum_{l=-4}^4 \dot{J}^l(\theta_0) Z_k^l(\theta_0)$ . Note that the  $U_k(\theta_0)$ 's are i.i.d and centered. Invoking the central limit theorem, we have that

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n U_k(\theta_0) \xrightarrow{d} \mathcal{N}(0, V/4),$$

where  $V/4$  is the covariance matrix of  $U_1(\theta_0)$ , equal to  $\mathbb{E}(U_1(\theta_0)U_1(\theta_0)^\top)$ .

**Step 2-** Let us prove that  $\ddot{S}_n(\theta_n^*) \xrightarrow{P} \mathcal{A}(\theta_0)$  where  $\mathcal{A}(\theta_0) = 2 \sum_{l=-4}^4 \dot{J}^l(\theta_0) \dot{J}^l(\theta_0)^\top$ . We have

$$\|\ddot{S}_n(\theta_n^*) - \mathcal{A}(\theta_0)\|_F \leq \|\ddot{S}_n(\theta_n^*) - \ddot{S}_n(\theta_0)\|_F + \|\ddot{S}_n(\theta_0) - \mathbb{E}\ddot{S}_n(\theta_0)\|_F + \|\mathbb{E}\ddot{S}_n(\theta_0) - \mathcal{A}(\theta_0)\|_F.$$

We get due to the Lipschitz property of  $\ddot{S}_n$  stated in Lemma 12 that

$$\mathbb{P}\left(\|\ddot{S}_n(\theta_0) - \ddot{S}_n(\theta_n^*)\|_F \geq \varepsilon\right) \leq \mathbb{P}(K\|\theta_n^* - \theta_0\| \geq \varepsilon) \rightarrow 0,$$

because  $\hat{\theta}_n \xrightarrow{P} \theta_0$ . Furthermore, we have

$$\begin{aligned} \mathbb{E}(\ddot{S}_n(\theta_0)) &= \ddot{S}(\theta_0) = 2 \sum_{l=-4}^4 (\ddot{J}^l(\theta_0) \underbrace{\dot{J}^l(\theta_0)}_{=0} + \dot{J}^l(\theta_0) \dot{J}^l(\theta_0)^\top) \\ &= 2 \sum_{l=-4}^4 \dot{J}^l(\theta_0) \dot{J}^l(\theta_0)^\top = \mathcal{A}(\theta_0). \end{aligned}$$

Last, let us focus on the term  $\|\ddot{S}_n(\theta_0) - \mathbb{E}\ddot{S}_n(\theta_0)\|_F$ . We remind that

$$\ddot{S}_n(\theta_0) - \mathbb{E}\ddot{S}_n(\theta_0) = \frac{2}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^4 \left( \ddot{Z}_k^l(\theta_0) Z_j^l(\theta_0) + \dot{Z}_k^l(\theta_0) \dot{Z}_j^l(\theta_0)^\top - \dot{J}^l(\theta_0) \dot{J}^l(\theta_0)^\top \right).$$

From now on, we drop indices  $l$  and  $\theta_0$  to simplify the notation. We center the variables in order to find uncorrelatedness:

$$\ddot{Z}_k Z_j + \dot{Z}_k \dot{Z}_j^\top - \dot{J} \dot{J}^\top = \underbrace{(\ddot{Z}_k - \ddot{J}) Z_j}_A + \underbrace{\ddot{J} Z_j}_B + \underbrace{(\dot{Z}_k - \dot{J})(\dot{Z}_j - \dot{J})^\top}_C + \underbrace{\dot{J}(\dot{Z}_j - \dot{J})^\top}_D + \underbrace{(\dot{Z}_k - \dot{J})(\dot{J})^\top}_E$$



(remind that  $\mathbb{E}(Z_j) = J^l(\theta_0) = 0$ ). Then  $\ddot{S}_n(\theta_0) - \mathbb{E}\ddot{S}_n(\theta_0) = 2 \sum_{l=-4}^4 (A + B + C + D + E)$  where

$$\begin{aligned} A &= \frac{2}{n(n-1)} \sum_{k < j} (\ddot{Z}_k - \ddot{J}) Z_j \\ B &= \frac{1}{n} \sum_{j=1}^n \ddot{J} Z_j \\ C &= \frac{2}{n(n-1)} \sum_{k < j} (\dot{Z}_k - \dot{J})(\dot{Z}_j - \dot{J})^\top \\ D &= \frac{1}{n} \sum_{j=1}^n \dot{J}(\dot{Z}_j - \dot{J})^\top \\ E &= \frac{1}{n} \sum_{k=1}^n (\dot{Z}_k - \dot{J}) \dot{J}^\top = D^\top \end{aligned}$$

Using the weak law of large numbers for uncorrelated centered variables, we obtain that  $\|\ddot{S}_n(\theta_0) - \mathbb{E}\ddot{S}_n(\theta_0)\|_F \xrightarrow{P} 0$  which completes the proof of the Theorem.

#### 6.4 Proof of Theorem 6 (nonparametric estimation)

The proof of the oracle inequality is based on the two following lemmas. The conclusion follows, choosing  $2\gamma = \epsilon/(1 + \epsilon)$  and  $\lambda = \gamma^{-1}\kappa(1 - 2P)^{-2} = 2\kappa(1 + \epsilon^{-1})(1 - 2P)^{-2}$ .

**Lemma 15.** *Let  $\lambda > 0$  and  $\mathcal{L}$  be a finite set of resolution level and define*

$$\hat{L} = \operatorname{argmin}_{L \in \mathcal{L}} \left\{ - \sum_{l=-L}^L |\widehat{f^{*l}}|^2 + \lambda \frac{2L+1}{n} \right\}.$$

Then, for all  $0 < \gamma < 1/2$ ,

$$(1 - 2\gamma) \|\widehat{f}_{\hat{L}} - f\|_2^2 \leq \min_{L \in \mathcal{L}} \left\{ (1 + 2\gamma) \|\widehat{f}_L - f\|_2^2 + 2\lambda \frac{2L+1}{n} \right\} + \frac{1}{\gamma} \sum_{L \in \mathcal{L}} \left( \sup_{t \in B_L} \nu_n^2(t) - \lambda \gamma \frac{2L+1}{n} \right)$$

where  $B_L = \{t \in \mathbb{C}^{\mathbb{Z}}, \sum_{l \in \mathbb{Z}} t_l^2 = 1, t_l = 0 \text{ if } |l| > L\}$   $\nu_n(t) = \sum_{l \in \mathbb{Z}} \overline{t_l} (\widehat{f^{*l}} - f^{*l})$ .

*Proof.* We recall that the dot product  $\langle f, g \rangle$  means  $\frac{1}{2\pi} \int \overline{f(x)} g(x) dx$  and that  $\|\cdot\|_2$  is the associated norm. Usual Fourier analysis gives for any  $L$ :

$$\begin{aligned} \|\widehat{f}_L - f\|_2^2 &= -\|\widehat{f}_L\|_2^2 + 2(\|\widehat{f}_L\|_2^2 - \langle \widehat{f}_L, f \rangle) + \|f\|_2^2 \\ &= - \sum_{l=-L}^L |\widehat{f^{*l}}|^2 + 2 \sum_{l=-L}^L \overline{\widehat{f^{*l}}} (\widehat{f^{*l}} - f^{*l}) + \|f\|_2^2 \\ &= - \sum_{l=-L}^L |\widehat{f^{*l}}|^2 + 2\nu_n(\widehat{f}_L^*) + \|f\|_2^2 \end{aligned}$$

where we denote  $\widehat{f}_L^*$  the sequence in  $\mathbb{C}^{\mathbb{Z}}$  such that  $(\widehat{f}_L^*)_l = \widehat{f}^{*l}$  if  $-L \leq l \leq L$  and 0 otherwise.

Now let  $L$  be an arbitrary resolution level in  $\mathcal{L}$ . Using the definition of  $\widehat{L}$ ,

$$-\sum_{l=-\widehat{L}}^{\widehat{L}} |\widehat{f}^{*l}|^2 + \lambda \frac{2\widehat{L}+1}{n} \leq -\sum_{l=-L}^L |\widehat{f}^{*l}|^2 + \lambda \frac{2L+1}{n}.$$

Thus

$$\|\widehat{f}_{\widehat{L}} - f\|_2^2 - 2\nu_n(\widehat{f}_{\widehat{L}}^*) + \lambda \frac{2\widehat{L}+1}{n} \leq \|\widehat{f}_L - f\|_2^2 - 2\nu_n(\widehat{f}_L^*) + \lambda \frac{2L+1}{n}$$

which leads to

$$\|\widehat{f}_{\widehat{L}} - f\|_2^2 \leq \|\widehat{f}_L - f\|_2^2 + 2\nu_n(\widehat{f}_{\widehat{L}}^* - \widehat{f}_L^*) - \lambda \frac{2\widehat{L}+1}{n} + \lambda \frac{2L+1}{n}.$$

But, denoting  $\|\cdot\|_\ell$  the natural norm of  $\ell^2(\mathbb{C}^{\mathbb{Z}})$

$$\begin{aligned} 2\nu_n(\widehat{f}_{\widehat{L}}^* - \widehat{f}_L^*) &= 2\nu_n \left( \frac{\widehat{f}_{\widehat{L}}^* - \widehat{f}_L^*}{\|\widehat{f}_{\widehat{L}}^* - \widehat{f}_L^*\|_\ell} \right) \|\widehat{f}_{\widehat{L}}^* - \widehat{f}_L^*\|_\ell \leq \gamma \|\widehat{f}_{\widehat{L}}^* - \widehat{f}_L^*\|_\ell^2 + \frac{1}{\gamma} \nu_n^2 \left( \frac{\widehat{f}_{\widehat{L}}^* - \widehat{f}_L^*}{\|\widehat{f}_{\widehat{L}}^* - \widehat{f}_L^*\|_\ell} \right) \\ &\leq 2\gamma (\|\widehat{f}_{\widehat{L}} - f\|_2^2 + \|f - \widehat{f}_L\|_2^2) + \frac{1}{\gamma} \sup_{t \in B_{L \vee \widehat{L}}} \nu_n^2(t) \end{aligned}$$

where  $L \vee \widehat{L} = \max(L, \widehat{L})$ . Thus

$$\begin{aligned} \|\widehat{f}_{\widehat{L}} - f\|_2^2 (1 - 2\gamma) &\leq \|\widehat{f}_L - f\|_2^2 (1 + 2\gamma) + \frac{1}{\gamma} \sup_{t \in B_{L \vee \widehat{L}}} \nu_n^2(t) - \lambda \frac{2\widehat{L}+1}{n} + \lambda \frac{2L+1}{n} \\ &\leq \|\widehat{f}_L - f\|_2^2 (1 + 2\gamma) + \frac{1}{\gamma} \left( \sup_{t \in B_{L \vee \widehat{L}}} \nu_n^2(t) - \lambda \gamma \frac{2\widehat{L}+2L+2}{n} \right) + 2\lambda \frac{2L+1}{n} \\ &\leq \|\widehat{f}_L - f\|_2^2 (1 + 2\gamma) + \frac{1}{\gamma} \left( \sup_{t \in B_{L \vee \widehat{L}}} \nu_n^2(t) - \lambda \gamma \frac{2(L \vee \widehat{L})+1}{n} \right) + 2\lambda \frac{2L+1}{n}. \end{aligned}$$

□

**Lemma 16.** *Assume that  $f$  belongs to the Sobolev ellipsoid  $W(s, R)$  with  $s > 1/2$ . With the notation of Lemma 15, if  $\mathcal{L} = \{1, \dots, L_n\}$  with  $L_n \rightarrow \infty$  but not larger than a power of  $n$ , for all  $\kappa \geq 1/2$ ,*

$$\sum_{L \in \mathcal{L}} \mathbb{E} \left( \sup_{t \in B_L} \nu_n^2(t) - \frac{\kappa}{(1-2P)^2} \frac{2L+1}{n} \right) \leq \frac{C}{n},$$

where  $C$  is a positive constant depending on  $\|f\|_2, P, \theta_0, R, s$ .

*Proof.* Denote  $R^l = \frac{1}{M^l(\widehat{\theta})} - \frac{1}{M^l(\theta_0)}$ . First note that

$$\nu_n(t) = \frac{1}{2\pi n} \sum_{k=1}^n \sum_{l \in \mathbb{Z}} \bar{t}_l \left( \frac{e^{-ilX_k}}{M^l(\widehat{\theta})} - \frac{2\pi g^{*l}}{M^l(\theta_0)} \right) = \nu_{n,1}(t) + \nu_{n,2}(t) + \nu_{n,3}(t)$$

where

$$\begin{aligned}\nu_{n,1}(t) &= \frac{1}{2\pi n} \sum_{k=1}^n \sum_{l \in \mathbb{Z}} \bar{t}_l \left( \frac{e^{-ilX_k} - 2\pi g^{*l}}{M^l(\theta_0)} \right) \\ \nu_{n,2}(t) &= \frac{1}{2\pi n} \sum_{k=1}^n \sum_{l \in \mathbb{Z}} \bar{t}_l \left( e^{-ilX_k} - 2\pi g^{*l} \right) R^l \\ \nu_{n,3}(t) &= \frac{1}{n} \sum_{k=1}^n \sum_{l \in \mathbb{Z}} \bar{t}_l g^{*l} R^l.\end{aligned}$$

Thus, if  $\kappa/3 \geq \kappa_1 + \kappa_2 + \kappa_3$ ,

$$\begin{aligned}\mathbb{E} \left( \sup_{t \in B_L} \nu_n^2(t) - \frac{\kappa}{(1-2P)^2} \frac{2L+1}{n} \right) &\leq 3\mathbb{E} \left( \sup_{t \in B_L} \nu_{n,1}^2(t) - \frac{\kappa_1}{(1-2P)^2} \frac{2L+1}{n} \right) \\ &+ 3\mathbb{E} \left( \sup_{t \in B_L} \nu_{n,2}^2(t) - \frac{\kappa_2}{(1-2P)^2} \frac{2L+1}{n} \right) + 3\mathbb{E} \left( \sup_{t \in B_L} \nu_{n,3}^2(t) - \frac{\kappa_3}{(1-2P)^2} \frac{2L+1}{n} \right).\end{aligned}$$

Control of  $\nu_{n,2}$  Note that  $|R^l| \leq 2/(1-2P)$ , so for  $t \in B_L$ ,

$$\nu_{n,2}^2(t) \leq \left( \frac{2}{1-2P} \sum_{l=-L}^L |\bar{t}_l(\widehat{g}^{*l} - g^{*l})| \right)^2.$$

Then

$$\mathbb{E} \left( \sup_{t \in B_L} \nu_{n,2}^2(t) \right) \leq \left( \frac{2}{1-2P} \right)^2 \sum_{l=-L}^L \mathbb{E} \left| \widehat{g}^{*l} - g^{*l} \right|^2 \leq \frac{1}{\pi^2(1-2P)^2} \frac{2L+1}{n}.$$

Denoting  $\kappa_2 = 1/\pi^2$ , we obtain

$$\mathbb{E} \left( \sup_{t \in B_L} \nu_{n,2}^2(t) - \frac{\kappa_2}{(1-2P)^2} \frac{2L+1}{n} \right) \leq 0.$$

Control of  $\nu_{n,3}$  First note that

$$\left| g^{*l} R^l \right| = \left| f^{*l} \frac{M^l(\theta_0) - M^l(\hat{\theta})}{M^l(\hat{\theta})} \right| \leq \frac{|f^{*l}|}{1-2P} \left| M^l(\theta_0) - M^l(\hat{\theta}) \right|.$$

Thus, using Schwarz inequality

$$\mathbb{E} \left( \sup_{t \in B_L} \nu_{n,3}^2(t) \right) \leq \sum_{l=-L}^L \frac{|f^{*l}|^2}{(1-2P)^2} \mathbb{E} \left| M^l(\theta_0) - M^l(\hat{\theta}) \right|^2.$$

But

$$\begin{aligned}|M^l(\theta_0) - M^l(\hat{\theta})| &\leq |(p_0 - \hat{p})e^{-i\alpha_0 l} + \hat{p}(e^{-i\alpha_0 l} - e^{-i\hat{\alpha} l}) + (1 - p_0 - 1 + \hat{p})e^{-i\beta_0 l} + (1 - \hat{p})(e^{-i\beta_0 l} - e^{-i\hat{\beta} l})| \\ &\leq |p_0 - \hat{p}| + |e^{-i\alpha_0 l} - e^{-i\hat{\alpha} l}| + |p_0 - \hat{p}| + |e^{-i\beta_0 l} - e^{-i\hat{\beta} l}| \\ &\leq 2|p_0 - \hat{p}| + |l||\alpha_0 - \hat{\alpha}| + |l||\beta_0 - \hat{\beta}| \leq 2|l| \|\theta_0 - \hat{\theta}\|_1\end{aligned}$$

(note that it is also true for  $l = 0$  since  $M^0(\theta_0) = M^0(\hat{\theta}) = 1$ ). According to the Theorem 5, there exists a constant  $K(\theta_0) > 0$  such that for all  $n$ ,  $\mathbb{E}(n\|\hat{\theta} - \theta_0\|_1^2) \leq K(\theta_0)$  and then  $\mathbb{E} \left| M^l(\theta_0) - M^l(\hat{\theta}) \right|^2 \leq 4K(\theta_0)l^2/n$ . Finally

$$\mathbb{E} \left( \sup_{t \in B_L} \nu_{n,3}^2(t) \right) \leq \frac{4K(\theta_0)}{(1-2P)^2n} \sum_{l=-L}^L l^2 |f^{*l}|^2 = \frac{4K(\theta_0)}{(1-2P)^2n} \sum_{\substack{l=-L \\ l \neq 0}}^L |l|^{2-2s} |l|^{2s} |f^{*l}|^2.$$

Since in the sum  $|l|^{2-2s} \leq \max(1, L^{2-2s})$  and  $f \in W(s, R)$

$$\begin{aligned} \sum_{L=1}^{L_n} \mathbb{E} \left( \sup_{t \in B_L} \nu_{n,3}^2(t) - \frac{\kappa_3}{(1-2P)^2} \frac{2L+1}{n} \right) &\leq \frac{2}{(1-2P)^2n} \sum_{L=1}^{L_n} (2K(\theta_0)R^2 \max(1, L^{2-2s}) - \kappa_3 L) \\ &\leq \frac{1}{(1-2P)^2n} (K'(\theta_0, R^2, \beta) \max(L_n, L_n^{3-2s}) - \kappa_3 L_n^2) \\ &\leq \frac{1}{(1-2P)^2n} K''(\theta_0, R^2, \beta, \kappa_3) \end{aligned}$$

for any  $\kappa_3 > 0$ , since  $s > 1/2$  and  $L_n = \max(\mathcal{L}) \rightarrow \infty$ .

#### Control of $\nu_{n,1}$

To control  $\nu_{n,1}$ , we need Talagrand inequality.

**Lemma 17.** *Let  $X_1, \dots, X_n$  be i.i.d. random variables, and define  $\nu_n(t) = \frac{1}{n} \sum_{k=1}^n \psi_t(X_k) - \mathbb{E}[\psi_t(X_k)]$ , for  $t$  belonging to a countable class  $\mathcal{B}$  of real-valued measurable functions. Then, for  $\delta > 0$ , there exist three constants  $c_l$ ,  $l = 1, 2, 3$ , such that*

$$\begin{aligned} \mathbb{E} \left[ \left( \sup_{t \in \mathcal{B}} (\nu_n(t))^2 - c(\delta)H^2 \right)_+ \right] &\leq c_1 \left\{ \frac{v}{n} \exp \left( -c_2 \delta \frac{nH^2}{v} \right) \right. \\ &\quad \left. + \frac{M_1^2}{C^2(\delta)n^2} \exp \left( -c_3 C(\delta) \sqrt{\delta} \frac{nH}{M_1} \right) \right\}, \end{aligned} \quad (6.5)$$

with  $C(\delta) = (\sqrt{1+\delta} - 1) \wedge 1$ ,  $c(\delta) = 2(1+2\delta)$  and

$$\sup_{t \in \mathcal{B}} \|\psi_t\|_\infty \leq M_1, \quad \mathbb{E} \left[ \sup_{t \in \mathcal{B}} |\nu_n(\psi_t)| \right] \leq H, \quad \text{and} \quad \sup_{t \in \mathcal{B}} \text{Var}(\psi_t(X_1)) \leq v.$$

Inequality (6.5) is a classical consequence of Talagrand's Inequality given in Klein and Rio (2005): see for example Lemma 5 (page 812) in Lacour (2008). Using density arguments, we can apply it to the unit sphere of a finite dimensional linear space.

Here  $\nu_{n,1}(t) = \frac{1}{n} \sum_{k=1}^n \psi_t(X_k) - \mathbb{E}[\psi_t(X_k)]$  with

$$\psi_t(X) = \frac{1}{2\pi} \sum_{l \in \mathbb{Z}} \bar{t}_l \frac{e^{-ilX}}{M^l(\theta_0)}, \quad \mathbb{E}(\psi_t(X)) = \sum_{l \in \mathbb{Z}} \bar{t}_l \frac{g^{*l}}{M^l(\theta_0)}$$

Let us compute  $M_1, H$  and  $v$ .

- Using Cauchy Schwarz inequality, for  $t \in B_L$ ,

$$|\psi_t(u)|^2 = \left| \frac{1}{2\pi} \sum_{l=-L}^L \bar{t}_l \frac{e^{-ilu}}{M^l(\theta_0)} \right|^2 \leq \frac{1}{4\pi^2} \sum_{l=-L}^L |t_l|^2 \sum_{l=-L}^L \left| \frac{e^{-ilu}}{M^l(\theta_0)} \right|^2 \leq \frac{1}{4\pi^2(1-2p_0)^2} (2L+1),$$

thus  $M_1 = \frac{1}{2\pi(1-2p_0)} \sqrt{2L+1}$ .

- Using Cauchy Schwarz inequality, for  $t \in B_L$ ,

$$\sup_{t \in B_L} \left| \frac{1}{2\pi n} \sum_{k=1}^n \sum_{l \in \mathbb{Z}} \bar{t}_l \left( \frac{e^{-ilX_k}}{M^l(\theta_0)} - \mathbb{E} \left( \frac{e^{-ilX_k}}{M^l(\theta_0)} \right) \right) \right|^2 \leq \sum_{l=-L}^L \left| \frac{1}{2\pi n} \sum_{k=1}^n \left( \frac{e^{-ilX_k}}{M^l(\theta_0)} - \mathbb{E} \left( \frac{e^{-ilX_k}}{M^l(\theta_0)} \right) \right) \right|^2,$$

then

$$\begin{aligned} \mathbb{E} \left( \sup_{t \in B_L} |\nu_{n,1}(\psi_t)|^2 \right) &\leq \sum_{l=-L}^L \text{Var} \left( \frac{1}{2\pi n} \sum_{k=1}^n \frac{e^{-ilX_k}}{M^l(\theta_0)} \right) \leq \sum_{l=-L}^L \frac{1}{4\pi^2 n} \text{Var} \left( \frac{e^{-ilX_1}}{M^l(\theta_0)} \right) \\ &\leq \frac{1}{4\pi^2 n} \sum_{l=-L}^L \mathbb{E} \left| \frac{e^{-ilX_1}}{M^l(\theta_0)} \right|^2 \leq \frac{1}{4\pi^2(1-2p_0)^2} \frac{2L+1}{n}, \end{aligned}$$

thus  $H^2 = \frac{1}{4\pi^2(1-2p_0)^2} \frac{2L+1}{n}$ .

- It remains to control the variance. If  $t \in B_L$

$$\text{Var}(\psi_t(X)) \leq \mathbb{E} \left| \frac{1}{2\pi} \sum_{l=-L}^L \bar{t}_l \frac{e^{-ilX}}{M^l(\theta_0)} \right|^2 = \frac{1}{4\pi^2} \sum_{l,l'} t_l \bar{t}_{l'} \frac{\mathbb{E}(e^{-ilX} \overline{e^{-il'X}})}{M^l(\theta_0) \overline{M^{l'}(\theta_0)}} = \frac{1}{2\pi} \sum_{l,l'} t_l \bar{t}_{l'} \frac{g^{*(l-l')}}{M^l(\theta_0) \overline{M^{l'}(\theta_0)}}$$

Using twice Schwarz inequality

$$\begin{aligned} \text{Var}(\psi_t(X)) &\leq \frac{1}{2\pi} \sqrt{\sum_l \left| \frac{t_l}{M^l(\theta_0)} \right|^2 \sum_{l'} \left| \sum_{l'} \frac{\bar{t}_{l'}}{M^{l'}(\theta_0)} g^{*(l-l')} \right|^2} \\ &\leq \frac{1}{2\pi(1-2p_0)} \sqrt{\sum_l \left| \sum_{l'} \frac{\bar{t}_{l'}}{M^{l'}(\theta_0)} g^{*(l-l')} \right|^2} \\ &\leq \frac{1}{2\pi(1-2p_0)} \sqrt{\sum_l \sum_{l'} \left| \frac{\bar{t}_{l'}}{M^{l'}(\theta_0)} \right|^2 \sum_{l'} |g^{*(l-l')}|^2} \\ &\leq \frac{1}{2\pi(1-2p_0)} \sqrt{\sum_l \frac{1}{|1-2p_0|^2} \sum_{j \in \mathbb{Z}} |g^{*j}|^2} \\ &\leq \frac{\|f\|_2}{2\pi(1-2p_0)^2} \sqrt{2L+1}, \end{aligned}$$

since  $\sum_{j \in \mathbb{Z}} |g^{*j}|^2 \leq \sum_{j \in \mathbb{Z}} |f^{*j}|^2 = \|f\|_2^2$ . Thus  $v = \frac{\|f\|_2}{2\pi(1-2p_0)^2} \sqrt{2L+1}$ .

Inequality (6.5) becomes

$$\begin{aligned} & \mathbb{E} \left[ \left( \sup_{t \in B_L} (\nu_{n,1}(t))^2 - \frac{c(\delta)}{4\pi^2(1-2p_0)^2} \frac{2L+1}{n} \right)_+ \right] \\ & \leq c_1 \left\{ \frac{\|f\|_2 \sqrt{2L+1}}{2\pi(1-2p_0)^2 n} \exp\left(-c_2 \delta \frac{\sqrt{2L+1}}{2\pi\|f\|_2}\right) + \frac{2L+1}{4\pi^2(1-2p_0)^2 C^2(\delta) n^2} \exp\left(-c_3 C(\delta) \sqrt{\delta n}\right) \right\} \\ & \leq \frac{K}{n} \left\{ \sqrt{2L+1} \exp\left(-c\sqrt{2L+1}\right) + \frac{2L+1}{n} \exp\left(-c\sqrt{n}\right) \right\} \end{aligned}$$

with  $K$  and  $c$  positive constants depending on  $\|f\|_2, p_0, c_1, c_2, c_3, \delta$ . This ends the control of  $\nu_{n,1}$  with  $\kappa_1 = \frac{c(\delta)}{4\pi^2}$  since

$$\sum_{L \in \mathcal{L}} \left\{ \sqrt{2L+1} e^{-c\sqrt{2L+1}} + \frac{2L+1}{n} e^{-c\sqrt{n}} \right\} \leq \sum_{L=1}^{\infty} \sqrt{2L+1} e^{-c\sqrt{2L+1}} + \#\mathcal{L} e^{-c\sqrt{n}} = O(1).$$

Choosing  $\kappa_3$  and  $\delta$  sufficiently small, it is sufficient to take

$$\kappa \geq 3\kappa_1 + 3\kappa_2 + 3\kappa_3 = \frac{9}{2\pi^2} + \frac{3\delta}{\pi^2} + \kappa_3 = \frac{1}{2}$$

to obtain the oracle inequality.

Let us derive the rate of convergence. Since  $\nu_n(t) = \sum_{l \in \mathbb{Z}} \widehat{t}_l (\widehat{f}^{*l} - f^{*l})$ ,

$$\sum_{l=-L}^L |\widehat{f}^{*l} - f^{*l}|^2 = \nu_n(\widehat{f}_L^* - f_L^*) \leq \sup_{t \in B_L} \nu_n(t) \|\widehat{f}_L^* - f_L^*\|_\ell$$

where we denote  $f_L^*$  the sequence in  $\mathbb{C}^{\mathbb{Z}}$  such that  $(f_L^*)_l = f^{*l}$  if  $-L \leq l \leq L$  and 0 otherwise. Hence  $\|\widehat{f}_L^* - f_L^*\|_\ell^2 \leq \sup_{t \in B_L} \nu_n(t) \|\widehat{f}_L^* - f_L^*\|_\ell$  so that  $\|\widehat{f}_L^* - f_L^*\|_\ell \leq \sup_{t \in B_L} \nu_n(t)$ . Then, using Lemma 16

$$\mathbb{E} \sum_{l=-L}^L |\widehat{f}^{*l} - f^{*l}|^2 = \mathbb{E} \|\widehat{f}_L^* - f_L^*\|_\ell^2 \leq \frac{\kappa}{(1-2P)^2} \frac{2L+1}{n} + \frac{C}{n} \leq C' \frac{2L+1}{n}.$$

Using Parseval equality,

$$\mathbb{E} \|f - \widehat{f}_L\|_2^2 = \sum_{|l| > L} |f^{*l}|^2 + C' \frac{2L+1}{n} \leq R^2 L^{-2s} + C' \frac{2L+1}{n}.$$

Thus, the oracle inequality gives

$$\mathbb{E} \|\widehat{f}_L - f\|_2^2 \leq (1+2\epsilon) \min_{L \in \mathcal{L}} \left\{ R^2 L^{-2s} + (C' + 2\lambda) \frac{2L+1}{n} \right\} + \frac{C}{n} \leq C'' n^{-2s/(2s+1)}$$

choosing  $L = cn^{1/(2s+1)}$ .

□

## References

- Banerjee, A., Dhillon, I. S., Ghosh, J., and Sra, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions. *J. Mach. Learn. Res.*, 6:1345–1382.
- Bartels, R. (1984). Estimation in a bidirectional mixture of von mises distributions. *Biometrics*, 40(3):1345–1382.
- Baudry, J.-P., Maugis, C., and Michel, B. (2012). Slope heuristics: overview and implementation. *Stat. Comput.*, 22(2):455–470.
- Birgé, L. and Massart, P. (2001). Gaussian model selection. *J. Eur. Math. Soc. (JEMS)*, 3(3):203–268.
- Birgé, L. and Massart, P. (2007). Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields*, 138(1-2):33–73.
- Bordes, L., Mottelet, S., Vandekerkhove, P., et al. (2006). Semiparametric estimation of a two-component mixture model. *The Annals of Statistics*, 34(3):1204–1232.
- Butucea, C. and Vandekerkhove, P. (2014). Semiparametric mixtures of symmetric distributions. *Scandinavian Journal of Statistics*, 41(1):227–239.
- Chen, J., Li, P., and Fu, Y. (2008). Testing homogeneity in a mixture of von Mises distributions with a structural parameter. *Canad. J. Statist.*, 36(1):129–142.
- Fraser, M. D., Hsu, Y. S., and Walker, J. J. (1981). Identifiability of finite mixtures of von Mises distributions. *Ann. Statist.*, 9(5):1130–1131.
- Gassiat, E. and Rousseau, J. (2016). Nonparametric finite translation hidden Markov models and extensions. *Bernoulli*, 22(1):193–212.
- Gassiat, E., Rousseau, J., and Vernet, E. (2018). Efficient semiparametric estimation and model selection for multidimensional mixtures. *Electron. J. Stat.*, 12(1):703–740.
- Goldenshluger, A. (2002). Density deconvolution in the circular structural model. *J. Multivariate Anal.*, 81(2):360–375.
- Hall, P., Neeman, A., Pakyari, R., and Elmore, R. (2005). Nonparametric inference in multivariate mixtures. *Biometrika*, 92(3):667–678.
- Hall, P. and Zhou, X.-H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.*, 31(1):201–224.
- Hohmann, D. and Holzmann, H. (2013). Two-component mixtures with independent coordinates as conditional mixtures: nonparametric identification and estimation. *Electron. J. Stat.*, 7:859–880.
- Holzmann, H., Munk, A., and Stratmann, B. (2004). Identifiability of finite mixtures-with applications to circular distributions. *Sankhyā: The Indian Journal of Statistics*, pages 440–449.
- Hunter, D. R., Wang, S., and Hettmansperger, T. P. (2007). Inference for mixtures of symmetric distributions. *The Annals of Statistics*, pages 224–251.

- Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in circular statistics*, volume 5 of *Series on Multivariate Analysis*. World Scientific Publishing Co., Inc., River Edge, NJ.
- Johannes, J. and Schwarz, M. (2013). Adaptive circular deconvolution by model selection under unknown error distribution. *Bernoulli*, 19(5A):1576–1611.
- Jones, T. A. and James, W. (1969). Analysis of bimodal orientation data. *Math. Geol.*, 1:129–135.
- Kent, J. T. (1983). Identifiability of finite mixtures for directional data. *Ann. Statist.*, 11(3):984–988.
- Kim, P. T. and Koo, J.-Y. (2000). Directional mixture models and optimal estimation of the mixing density. *Canad. J. Statist.*, 28(2):383–398.
- Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077.
- Lacour, C. (2008). Adaptive estimation of the transition density of a particular hidden Markov chain. *Journal of Multivariate Analysis*, 99(5):787–814.
- Ley, C. and Verdebout, T. (2017). *Modern directional statistics*. Chapman & Hall/CRC Interdisciplinary Statistics Series. CRC Press, Boca Raton, FL.
- Ma, Y. and Yao, W. (2015). Flexible estimation of a semiparametric two-component mixture model with one parametric component. *Electron. J. Stat.*, 9(1):444–474.
- Mardia, K. V. and Jupp, P. E. (2000). *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.
- Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Phil. Trans. Royal Soc., A*, 185:71–110.
- Pewsey, A. and García-Portugués, E. (2020). Recent advances in directional statistics. *arXiv:2005.06889*.
- Spurr, B. D. (1981). On estimating the parameters in mixtures of circular normal distributions. *J. Internat. Assoc. Math. Geol.*, 13(2):163–173.
- Spurr, B. D. and Koutbeiy, M. A. (1991). A comparison of various methods for estimating the parameters in mixtures of von Mises distributions. *Comm. Statist. Simulation Comput.*, 20(2-3):725–741.