The Annals of Statistics

SEMIPARAMETRIC ESTIMATION OF SHIFTS BETWEEN CURVES

By Fabrice Gamboa

Université Toulouse 3

and

By Jean-Michel Loubes

CNRS and Université Montpellier 2

and

By Elie Maza

Université Toulouse 3

We observe a large number of functions differing from each other only by a translation parameter. While the main pattern is unknown, we propose to estimate the shift parameters using M-estimators. Fourier series enable to transform this statistical problem into a semi-parametric framework. We study the convergence of the estimator and provide its asymptotic behavior. Moreover, we use the method in the applied case of velocity curve forecasting.

1. Introduction. A main issue in data mining is the feature extraction of a large set of curves. Indeed, classification methods enable to split the data into different homogeneous groups, each representing a specific mass behavior. But, within one group, the observations differ slightly the one from another. Such variations take into account the variability of the individuals inside one group. More precisely, there is a mean pattern such that, each observation curve is warped from this archetype by a warping function, see for examples [25] or [15].

In this work, we focus on a particular case where the individuals usually experience similar events, which are explained by a common pattern, but the starting time of the event occurs sooner or later. Classification methods, like repeated measures ANOVA or Principal Component Analysis of curves,

Received 1 January 1; revised 1 January 1; accepted 1 January 1.

AMS 2000 subject classifications: Primary 60G17; secondary 62G07

Keywords and phrases: Semiparametric estimation, Warping model, Empirical process, Fourier transform, M-estimation

see for instance [24] or [26], ignore this type of variability. Hence, computing a representative curve for each group, severely distorts the analysis of the data. Indeed, the average curve (usually the mean or the median) oversmooths the studied phenomenon, and is not a good description of reality.

In our work, we restrict ourselves to the case where all the curves can be deduced the one from another by a shift parameter. Hence, we consider the following model: for $j = 1, \dots, J$ and $i = 1, \dots, n_j$, we observe

$$Y_{ij} = f\left(t_{ij} - \theta_j^*\right) + \sigma\varepsilon_{ij}, \tag{1.1}$$

where J stands for the number of curves $\mathcal{C}_j$ while n_j is the number of observations for the j-th individual. Values t_{ij} are the observation times, which are assumed to be known. The unobserved warping effects θ_j^* for $j = 1, \dots, J$ are shift parameters which translate the unknown function f. We also choose $t_{ij} = t_i$ and $n_j = n$, which means that all curves are observed at the same times with the same occurrence. The errors ε_{ij} for $(i, j) \in \{1, \dots, n\} \times \{1, \dots, J\}$ are i.i.d. with distribution $\mathcal{N}(0, 1)$. Moreover, we assume in the following that $\sigma = 1$ (see Remark 3.2). We aim at finding a good representative of the feature f.

A more general problem has been tackled in the litterature and some work has been done to find a representative of a large sample of close enough functions f_j, $j = 1, \dots, J$, see for examples [25] or [15]. Indeed, in a general case, if we observe y_{ij}, $j = 1, \dots, J$, $i = 1, \dots, n_j$, such that

$$Y_{ij} = f_j(t_{ij}) + \varepsilon_{ij}, \tag{1.2}$$

where, ε_{ij}, $j = 1, \dots, J$, $i = 1, \dots, n_j$, are i.i.d. random variables, representing the observation noise. Such functions f_j, $j = 1, \dots, J$, are close from each other in the sense that there exists an unknown archetype f and unknown warping functions h_j, $j = 1, \dots, J$, such that, for all $j = 1, \dots, J$,

$$\forall t \in [0, T],\ f_j(t) = f \circ h_j(t).$$

Examples of such data might be growth curves, longitudinal data in medicine, speech signals, traffic data or expenditure curves for some goods in the econometric domain. Our main motivation in this paper is the analysis of the vehicle speed evolution on a motorway. The data are curves, describing the evolution, on observation cells, of the daily vehicle speed. After performing classification procedures (see for instance [19] for a complete study), we obtain clusters of functions where in each subgroup appears a typical common behavior. Indeed, all the curves can be deduced one from another by a shift parameter.

This kind of issue led several statisticians to apply transformation to functions in order to get rid of the shifts and to align the curves. If a parametric model would be available a priori, the analysis would be made easier. But, if the data are numerous, there is not generally enough knowledge to build such a model. Thus, they turn into a non parametric framework. When the pattern is known, the problem turns to align a noisy observation with a fixed feature. Piccioni, Scarlatti and Trouvé in [21], Kneip, Li, MacGibbon and Ramsay in [16], or Ramsay and Li in [22] proposed curve registration methods. Their main idea is to align each curve on a target curve f_0, which means finding, for all $j \in \{1, \dots, J\}$, the warping function h_j minimizing

$$F_\lambda(f_0, f_j; h_j) = \int \|f_j \circ h_j(t) - f_0(t)\|^2 \, \mathrm{d}t + \lambda \int w_j^2(t)\mathrm{d}t,$$

where h_j belongs to a particular smooth monotone family defined by the solution of the differential equation $D^2 h_j = w_j D h_j$. Hence, w_j is simply $D^2 h_j / D h_j$, the relative curvature of h_j. Thus, penalizing w_j yields both smoothness and monotonicity of h_j (see [22] for more details). The main drawback of such methods is that they assume that the archetype f_0 is known, which is a reasonable assumption in pattern recognition, but which is unrealistic when the observed phenomenon is not well known as in our study. Alternatively, in a non parametric point of view, the pattern is replaced by its estimate. In this case, the issue is a matter of synchronizing sample curves. Wang and Gasser in [30] and [31] use kernel estimators, as Boularan, Ferré and Vieu in [2] or Núñez-Antón, Rodríguez-Póo and Vieu in [20]. In another work, Gasser and Kneip, in [7], align the curves by aligning the local extrema of the functions, which are estimated as zeroes of the non parametric estimate of the derivative. In all cases, the issue of estimating the shifts is blurred by the estimation of the curves, which leads to non parametric rates of convergence.

Hence, it seems natural to study the problem in a semiparametric framework: the shifts are the parameters of interest to be estimated, while the pattern stands for an unknown nuisance functional parameter. A very general semiparametric regression model called *Self-modelling regression (SEMOR)* has been considered in [14] and [13]. In these papers the model is $f_j(\cdot) = f(\cdot, \theta_j^*)$, $j \in \{1, \dots, J\}$, and a general backfitting algorithm is studied. Roughly speaking, after initializing an estimate of f by a first guess (using for example a kernel method), this algorithm is based on two recusive steps. In the first step, the estimation of $\theta_j^*, j = 1, \ldots, J$, is performed using a least squares criterion. In the second step, the estimate of f is updated using also a least squares method. In [14] a general convergence theorem is proved. In [13] a complete study, including the asymptotic normality of the estimates, is performed for the *Shape-invariant model (SIM)* introduced in [17]. See also [18], [11] and [12] for related works. Actually, the model studied in our paper (regression model (1.1)) falls in the SIM frame, so that, the methods studied in [14] and [13] may be applied. Nevertheless, the estimation procedure developed here is new, structurally simpler and computatonaly easier to implement than the complicated backfitting algorithms.

The difficulty of the work is that the estimation must not rely on the pattern, even if the quantities are deeply linked. That is the reason why we will use an M-estimator built on the Fourier series of the data. Under identifiability assumptions, we provide a consistent method to estimate at the parametric rate of convergence the shifts θ_j^*, $j = 1, \ldots, J$, when f is unknown, and we show that fluctuations of the estimates are asymptotically Gaussian. Further, our estimation method leads to a fast algorithm to align shifted curves without any prior assumption on the feature, due to semiparametric techniques. We point out that this study can be linked first with the study of Golubev in [10], dealing with the semiparametric efficiency in the estimation of shifts in a continuous observation scheme, and also with the study of Gassiat and Lévy-Leduc in [8], dealing with the estimation of the periodicity of a signal. Futher, the mixed effects model (1.1) with random shifts is studied in [3] (see also [4]).

The present paper falls into six parts. Section 2 is devoted to the definition of the model and to the description of the estimation method. In Section 3, we provide asymptotic properties of the estimators. As a matter of fact, we show that the estimators are convergent and asymptotically Gaussian. The estimating method is effectively performed in Section 4, on some simulated data, and then used to analyze road traffic data. We compare our results to another existing method. The technical lemmas and the proofs are gathered in Section 5 and Section 6.

2. Semiparametric estimation of shifts.

2.1. *Model.* For every curve $\mathcal{C}_j$, $j = 1, \ldots, J$, we get n observations y_{ij}, $i = 1, \ldots, n$, measured at equispaced times $t_i = \frac{i-1}{n} T \in [0, T[$, with

$T \in \mathbb{R}_+^*$. We model these observations in the following way:

$$Y_{ij} = f(t_i - \theta_j^*) + \varepsilon_{ij},\ j = 1, \ldots, J,\ i = 1, \ldots, n, \tag{2.1}$$

where, $f : \mathbb{R} \to \mathbb{R}$ is an unknown T-periodic function, $\theta^* = (\theta_1^*, \ldots, \theta_J^*) \in \mathbb{R}^J$ is an unknown shift parameter, θ_j^* is the shift of the j-th curve, and, for all $j \in \{1, \ldots, J\}$, ε_{ij}, $i = 1, \ldots, n$, is a Gaussian white noise, with variance 1. For sake of simplicity, we get an unitary variance, but all our results are still valid for a general variance.
Our aim is to estimate the translation factors θ_j^*, $j = 1, \ldots, J$, without the knowledge of the pattern f. Due to the special structure of the model, Fourier analysis is well suited to conduct such a study. Indeed, the Fourier basis diagonalizes any translation. Then, using a Discrete Fourier Transform (see [1] for more details), we may transform the model (2.1) into the following one (supposing n is odd):

$$d_{jl} = e^{-il\alpha_j^*} c_l(f) + w_{jl},\ j = 1, \ldots, J,\ l = -(n-1)/2, \ldots, (n-1)/2, \tag{2.2}$$

where, $c_l(f) = \frac{1}{n} \sum_{m=1}^n f(t_m) e^{-i2\pi \frac{ml}{n}}$, $l = -(n-1)/2, \ldots, (n-1)/2$, are the discrete Fourier coefficients and $\alpha_j^* = \frac{2\pi}{T} \theta_j^* \in \mathbb{R}$, $j = 1, \ldots, J$, are the phase factors, and, for all $j \in \{1, \ldots, J\}$, w_{jl}, $l = -(n-1)/2, \ldots, (n-1)/2$, is a complex Gaussian white noise, with complex variance $1/n$, and with independent real and imaginary parts. As previously, our goal is to estimate the phase factors α_j^*, $j = 1, \ldots, J$, without the knowledge of the Fourier coefficients of function f. Stricto sensu, the discrete Fourier coefficients are not the real Fourier coefficients of the functions, but the bias induced is similar to the bias induced by any discretization in regression, which vanishes under some regularity assumptions, as shown in [6].

We point out that we are facing a semiparametric model. As a matter of fact, we aim at estimating the parameter $\alpha^* = (\alpha_1^*, \cdots, \alpha_J^*)$ which depends on an unknown nuisance functional parameter $(c_l(f))_{l \in \mathbb{Z}}$, the Fourier coefficients of the unknown function f.

2.2. *Identifiability.* We notice that the model (2.2) is not identifiable for all translation parameters. Indeed, replacing α^* by

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_J \end{pmatrix} = \begin{pmatrix} \alpha_1^* \\ \vdots \\ \alpha_J^* \end{pmatrix} + c \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + 2\pi \begin{pmatrix} k_1 \\ \vdots \\ k_J \end{pmatrix},\ c \in \mathbb{R},\ \begin{pmatrix} k_1 \\ \vdots \\ k_J \end{pmatrix} \in \mathbb{Z}^J, \tag{2.3}$$

and replacing $f(\cdot)$ by $f(\cdot - c)$, let invariant the equation (2.2). So, in order to ensure identifiability of the model, we restrict the parameter space A:

$$\begin{aligned} &\text{i)} \quad A \text{ is compact,} \\ &\text{ii)} \quad \alpha^* \in A, \\ &\text{iii)} \quad \text{if } \alpha \in A \text{ and (2.3) holds for } \alpha, \text{ then } \alpha = \alpha^*. \end{aligned} \tag{2.4}$$

In this article, we will mainly consider, in the fluctuations theorem (Theorem 3.1), the parameter set $A_1 = \left\{\alpha \in [-\pi, \pi[^J : \alpha_1 = 0\right\}$. Hence, constant c defined in (2.3) is equal to 0, and $\alpha = \alpha^*$. Our fluctuations theorem can easily be transposed to other choices of parameter spaces, for example $A_2 = \left\{\alpha \in [-\pi, \pi[^J : \sum_{j=1}^J \alpha_j = 0 \text{ and } \alpha_1 \in \left[0, \frac{2\pi}{J}\right[\right\}$. The condition $\sum_{j=1}^J \alpha_j = 0$ implies in (2.3) that $c = -\frac{2\pi}{J}\sum_{j=1}^J k_j$. So that, with equation (2.3), we can write that

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_J \end{pmatrix} = \begin{pmatrix} \alpha_1^* \\ \vdots \\ \alpha_J^* \end{pmatrix} + \frac{2\pi}{J} \begin{pmatrix} (J-1) & -1 & \cdots & -1 \\ -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ -1 & \cdots & -1 & (J-1) \end{pmatrix} \begin{pmatrix} k_1 \\ \vdots \\ k_J \end{pmatrix}.$$

Hence, we get J different solutions in $[-\pi, \pi[^J \subset \mathbb{R}^J$, and a unique solution with the additional condition $\alpha_1 \in \left[0, \frac{2\pi}{J}\right[$.

2.3. *Estimation.* Since we want to estimate the shifts without prior knowledge of the function f, we will consider a semiparametric method, relying on an M-estimation procedure. Hence, the functional parameter is a nuisance parameter that does not play a role in the rate of converge of the estimates of the parameters, regardless of smoothness conditions for f.

For this, define, for any $\alpha = (\alpha_1, \ldots, \alpha_J) \in A$, the rephased coefficients

$$\tilde{c}_{jl}(\alpha) = e^{il\alpha_j} d_{jl},\ j = 1, \ldots, J,\ l = -(n-1)/2, \ldots, (n-1)/2,$$

and the mean of these rephased Fourier coefficients

$$\hat{c}_l(\alpha) = \frac{1}{J}\sum_{j=1}^J \tilde{c}_{jl}(\alpha),\ l = -(n-1)/2, \ldots, (n-1)/2.$$

We have that $\tilde{c}_{jl}(\alpha^*) = c_l(f) + e^{il\alpha_j^*} w_{jl}$, for all $j \in \{1, \ldots, J\}$, and

$$\hat{c}_l(\alpha^*) = c_l(f) + \frac{1}{J}\sum_{j=1}^J e^{il\alpha_j^*} w_{jl}.$$

Hence, $|\tilde{c}_{jl}(\alpha) - \hat{c}_l(\alpha)|^2$ should be small when α is close to α^*. As a consequence, considering a sequence $(\delta_l)_{l\in\mathbb{Z}}$ such that $\sum_{l\in\mathbb{Z}} \delta_l^2 < +\infty$, we define the following criterion function:

$$M_n(\alpha) = \frac{1}{J}\sum_{j=1}^{J}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \left|\tilde{c}_{jl}(\alpha) - \hat{c}_l(\alpha)\right|^2. \tag{2.5}$$

This random function is positive. Furthermore, its minimum value should be reached close to the true parameter α^*. We assume also that for all sufficiently large n, $\sum_{l=-(n-1)/2}^{(n-1)/2} \delta_l^2 |c_l(f)|^2 > 0$, which is achieved as soon as $\delta_l^2 > 0$ for all $l \in \mathbb{Z}$. Then, the following theorem provides the consistency of the M-estimator, defined by

$$\hat{\alpha}_n = \arg\min_{\alpha\in A} M_n(\alpha).$$

THEOREM 2.1. *Under the following assumptions on f and on the weight sequence $(\delta_l)_{l\in\mathbb{Z}}$:*

$$\sum_{l\in\mathbb{Z}} |c_l(f)|^2 < +\infty, \tag{2.6}$$

$$\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 < +\infty, \tag{2.7}$$

we have that $\hat{\alpha}_n \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} \alpha^*$.

We point out that we only assume that f is square integrable and yet are able to build estimates of the shifts. The computation of the estimator is quick since only a Fast Fourier Transform algorithm and a minimization algorithm of a quadratic functional are needed.

PROOF 2.2 (Proof of Theorem 2.1). *The proof of this theorem follows the classical guidelines of the convergence of M-estimators (see for example [28] or [9]). Indeed, the contrast is split into two parts, a determinist and a random one. Then, it suffices to show that the following conditions hold for the criterion function to ensure consistency of $\hat{\alpha}_n$.*

i) Convergence to a contrast function:

$$M_n(\alpha) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} K(\alpha),\ \alpha \in A, \tag{2.8}$$

where $K(\cdot)$ has a unique minimum at α^*.

ii) *Set the modulus of uniformly continuity W, defined by $W(n,\eta) = \sup_{\|\alpha-\beta\|\leq\eta} |M_n(\alpha) - M_n(\beta)|$. There exists two sequences $(\eta_k)_{k\in\mathbb{N}}$ and $(\epsilon_k)_{k\in\mathbb{N}}$, decreasing to zero, such that for a large enough k, we have*

$$\lim_{n\to+\infty} \mathbf{P}_{\alpha^*}\left(W(n,\eta_k) > \epsilon_k\right) = 0. \tag{2.9}$$

These two conditions are fulfilled, as it is proved in Section 6. Notice that we have chosen to privilegiate the uniform convergence of the modulus of continuity of the contrast and not the uniform convergence of the criterion itself. Nevertheless, the two proofs use the same kind of arguments, i.e proving the uniform convergence of empirical processes of the following form: $\frac{1}{n}\sum_{l=-(n-1)/2}^{(n+1)/2} \delta_l^2 \sum_{j=1}^{J} \cos(l\alpha_j)\xi_{jl}^x \xi_{kl}^y$, where ξ_{jl}^x and ξ_{kl}^y are independent centered Gaussian variables.

■

3. Asymptotic normality. In this section, we prove that the estimator built in the previous section is asymptotically Gaussian, and we give its asymptotic covariance matrix. In general, the asymptotic covariance matrix hardly depends on the geometric structure of the parameter space A. So, for sake of simplicity, we study the asymptotic normality for the parameter space A_1. Hence, the parameter space has dimension $J-1$, and we rewrite this as $\tilde{A}_1 = [-\pi, \pi[^{J-1}$, and, any element in $\tilde{A}_1$ as $\alpha = (\alpha_2, \ldots, \alpha_J)$. Also, for sake of simplicity, in this section and in the proofs of Theorem 3.1, we will write $M_n(\alpha)$ instead of $M_n(0, \alpha_2, \ldots, \alpha_J)$. So, we consider any estimator defined by

$$\hat{\alpha}_n = \arg\min_{\alpha\in\tilde{A}_1} M_n(\alpha).$$

THEOREM 3.1. *Under the assumptions of Theorem 2.1, and the following additional assumption on the weight sequence $(\delta_l)_{l\in\mathbb{Z}}$:*

$$\sum_{l\in\mathbb{Z}} \delta_l^4 l^4 < +\infty, \tag{3.1}$$

we get that

$$\sqrt{n}(\hat{\alpha}_n - \alpha^*) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}_{J-1}(0, \Gamma), \tag{3.2}$$

with

$$\Gamma = \frac{\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2}{\left(\sum_{l\in\mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2\right)^2} \left(I_{J-1} + U_{J-1}\right),$$

where, I_{J-1} is the identity matrix of dimension $J-1$, and U_{J-1} is the square matrix of dimension $J-1$ whose all entries are equal to one.

REMARK 3.2. *If the white noise in the model* (2.1) *has a variance equal to* σ^2*, then the limit distribution in the previous theorem has a covariance matrix equal to* $\sigma^2\Gamma$.

REMARK 3.3. *In Theorem 2.1 and Theorem 3.1, we only assume that* f *lies in* L^2*. That is, the natural class of functions in the semiparametric model* (2.1) *is quite general.*

PROOF 3.4 (Proof of Theorem 3.1). *Recall that the* M*-estimator is defined as the minimum of the criterion function* $M_n(\alpha)$*. Hence, we get*

$$\nabla M_n(\hat{\alpha}_n) = 0,$$

where ∇ *is the gradient operator. A second order decomposition leads to: there exists* $\bar{\alpha}_n$ *in a neighborhood of* α^* *such that*

$$\sqrt{n}\left(\hat{\alpha}_n - \alpha^*\right) = -\left[\nabla^2 M_n\left(\bar{\alpha}_n\right)\right]^{-1} \sqrt{n}\nabla M_n\left(\alpha^*\right), \tag{3.3}$$

where ∇^2 *is the Hessian operator. Now, using the two asymptotic results from Proposition 5.1 and from Proposition 5.2, we get*

$$\sqrt{n}\nabla M_n(\alpha^*) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}_{J-1}(0, \Gamma_0),$$
$$\left[\nabla^2 M_n\left(\bar{\alpha}_n\right)\right]^{-1} \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} V,$$

where V *is a non negative symmetric matrix of dimension* $J-1$*. Hence, if we set* $\Gamma = V'\Gamma_0 V$*, the result of Theorem 3.1 follows easily.*

■

REMARK 3.5. *The extra terms* $(\delta_l)_{l\in\mathbb{Z}}$ *used in the definition* (2.5) *smooth the criterion function* $M_n(\alpha)$*. Indeed, without this term, i.e under the choice* $\delta_l = 1$*, the random part of the criterion function does not converge towards a determinist part but to a random process, preventing any estimation. The weights enable to get rid of this part, smoothing the contrast to zero. We illustrate this purpose on Section 4 by comparing a weighted criterion with a non weighted one (see Figure 3). Moreover, this result will be highlighted in the proof of Theorem 3.1.*

REMARK 3.6. *The problem of choosing the weights* $(\delta_l)_{l\in\mathbb{Z}}$ *in the definition of the criterion function is important. Since we work with* L^2 *functions, the assumption* (3.1) *is satisfied as soon as* $|\delta_l| = O\left(|l|^{-5/4-\nu}\right)$, $\forall \nu > 0$.

In the simulations, we have taken $\delta_l = 1/|l|^{1.3}$. *This choice guarantees consistency and good numerical results. But, when looking at the asymptotic variance, we can see that there is a trade-off which leads to a lower bound for the smoothing sequence, the smaller the weights, the larger the variance. Since the function* f *is unknown and so the sequence does not depend on the Fourier coefficients, hence the optimal choice for* $(\delta_l)_{l\in\mathbb{Z}}$ *should be given by semiparametric efficiency. Using Cauchy-Schwarz's inequality, we get that*

$$\frac{\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2}{\left(\sum_{l\in\mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2\right)^2} \geq \left(\sum_{l\in\mathbb{Z}} l^2 |c_l(f)|^2\right)^{-1}.$$

This case, corresponding to the least favorable case in the semiparametric efficiency framework, is obtained for the optimal choice of coefficients $\delta_l = 1$. *If an asymptotic fluctuation results would hold, we would obtain:*

$$\sqrt{n}\sqrt{\sum_{l\in\mathbb{Z}} l^2 |c_l(f)|^2}\,(\hat{\alpha}_n - \alpha^*) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}_{J-1}\left(0, I_{J-1} + U_{J-1}\right).$$

Nevertheless, the choice of the weight sequence $\delta_l = 1$ *yields to a non convergent estimator. Non optimality as regards asymptotic efficiency is the price to pay both to deal with a discretized version of the regression model and to handle simultaneous estimation for all the unknown functions. Maybe, a different way of estimation could get rid of this drawback. Yet, another choice could have been done to smooth the contrast by restricting the number of Fourier coefficients, as it is done in [10] for example. Some links could also be established between the estimator we consider and a Bayesian penalized maximum likelihood estimator, where the weights* $(\delta_l)_{l\in\mathbb{Z}}$ *stand for a particular choice of a prior over the unknown function* f. *This Bayesian point of view is tackled in [5]. However, the optimal choice of the smoothing parameter to obtain efficiency is a difficult issue in the semiparametric framework (i.e when the weights are not allowed to depend on the Fourier coefficients of the functions) which is tackled in a forthcoming work, [29].*

REMARK 3.7. *Throughout all the work, we assume that the observation noise in the model* (2.1) *is Gaussian. Nevertheless, we could get rid of this assumption with moment conditions for the errors.*

4. Applications and simulations. In this section, we present some numerical applications of the method. The first one gives results on simulated data. The second one is based on an experiment on human fingers force. The last one is carried out with traffic data.

The optimization algorithm used in any resolution is based on a Krylov method (the conjugate gradient method). Indeed, minimizing an L^2 criterion function with a conjugate gradient algorithm yields a reduced step number, and hence, a small complexity.

Simulated data

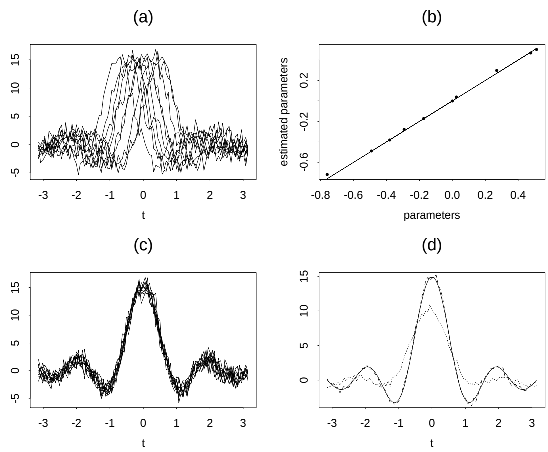

Figure 1: Estimation results with the M-estimation methodology.

Simulated data are carried out as follows:

$$y_{ij} = f(t_i - \theta_j^*) + \varepsilon_{ij},\ j = 1, \dots, J,\ i = 1, \dots, n,$$

with the following choice of parameters: $J = 10$; $n = 100$; values $t_i = -\pi + \frac{i-1}{n}2\pi$, $i = 1, \dots, n$, are equally spaced points on $[-\pi, \pi[$; $f(t) = 15\sin(4t)/(4t)$; $(\theta_2^*, \dots, \theta_J^*)$ are simulated with a uniform law on $[-\pi/4, \pi/4]$ and $\theta_1^* = 0$; for all $j \in \{1, \dots, J\}$, for all $i \in \{1, \dots, n\}$, values ε_{ij} are simulated from a Gaussian law with mean 0 and standard deviation 1. Results are given on Figure 1. The target function f is considered as a 2π-periodic function ($T = 2\pi$), hence $\alpha^* = \theta^*$. The function f is plotted by a solid line in figure 1(d). The figure 1(a) shows simulated data y_{ij}, $j = 1, \dots, J$, $i = 1, \dots, n$. The mean curve of these data is given on Figure 1(d) by the dotted line. We can see that the mean function is representativeness of data. Indeed, the amplitude of higher optimum is reduced and smallest ones disappeared. Figure 1(c) shows shifted curves. The mean function of these shifted curves is given on Figure 1(d) by dashed line. Figure 1(b) plots α_j^* on abscissa axis against $\hat{\alpha}_j$ on ordinate axis, $j = 1, \dots, J$. Estimations are very close to truth parameters. Comparison between mean curves, before and after the shift estimation, is straightforward.

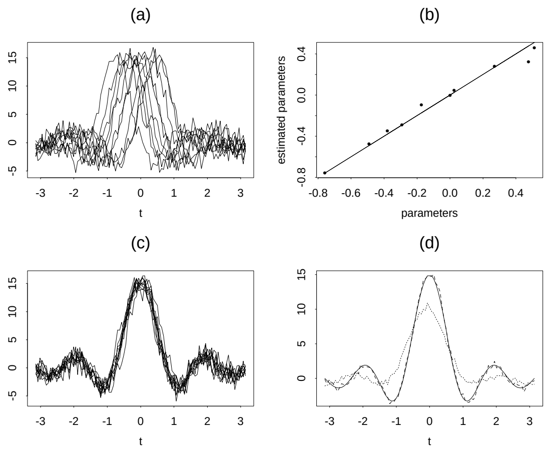

Figure 2: Estimation results with the landmark methodology.

We now compare our estimations with those obtained with an existing method: curve registration by landmarks. This method aims at aligning curves by, first, estimating landmarks of curves (here, the maximum) and by, secondly, aligning these landmarks. For more details on this procedure, see [7]. In Figure 2, we show the results on our simulated data. These results are not so good as those we obtain with our method. That can be explain by the fact that we need first to estimate each curve maximum by a non parametric method which leads to estimation errors. Moreover, our method uses all information given by the data, not only by landmarks.

In order to illustrate the importance of the weight sequence, we compare now the obtained criterion function with and without $(\delta_l)_{l\in\mathbb{Z}}$. For this purpose, simulated data sets are carried out, with $J = 2$, $\theta_1^* = 0$ and $\theta_2^* = \pi/3$. Figure 3 shows the obtained results. The first column of this figure presents these simulated data sets, with respectively, $\sigma = 1$ in Figure 3(a,1), $\sigma = 2$ in Figure 3(a,2) and $\sigma = 3$ in Figure 3(a,3). The second column presents the weighted criterion functions $M_n(\cdot)$ associated respectively with (a,1), (a,2) and (a,3). The third column presents, respectively, the associated non weighted criterion functions. We easely see that without our weight sequence $(\delta_l)_{l\in\mathbb{Z}}$, the criterion function is random, with a variance proportional to the noise variance σ^2. We also see that even with an important noise variance,

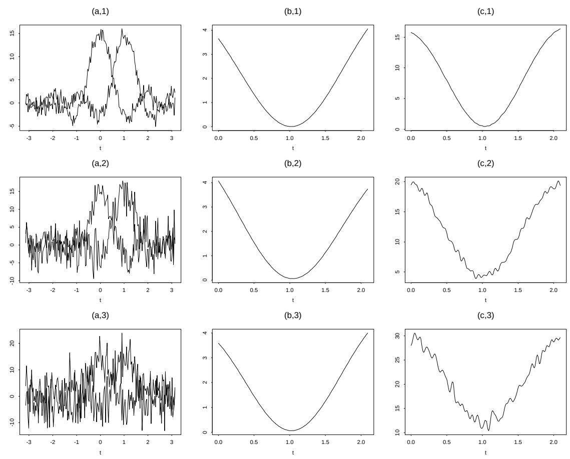

Figure 3: Criterion functions, with and without the weight sequence $(\delta_l)_{l\in\mathbb{Z}}$.

as in Figure 3(a,3), our criterion function is smooth, with a unique minimum arround $\pi/3$.

Pinch force data

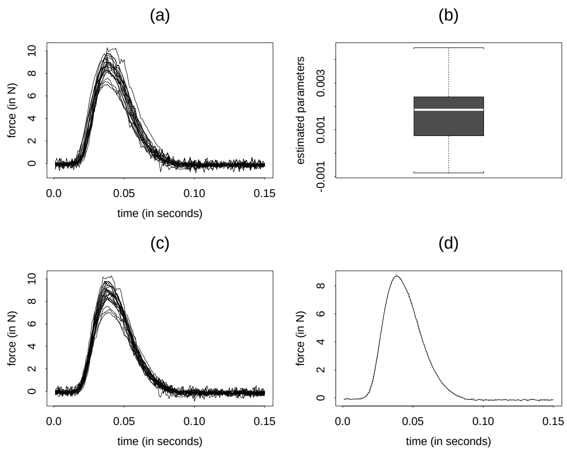

Figure 4: Shift estimation results on the pinch force data set.

Data presented here are extracted from an experiment described in [23], and studied in [22] with a Curve Registration methodology. Data represent the force exerted by the thumb and forefinger on a force meter during 20 brief pinches. These 20 force measurements having arbitrary beginning, Ramsay and Li in [22] begin their study by a landmark alignment of curve maxima (with single shifts). These aligned data are shown in figure 4(a).

Our propose is to study these data with the shift estimation methodology. Shift estimations and shifted curves are respectively shown in Figure 4(b) and Figure 4(c). In Figure 4(b), we only show a boxplot of the estimated parameters because, obviously, we do not know the real parameters. We note that shift parameters are almost all close to zero, between -10^{-3} and 310^{3}. That means that, in this case, landmark alignment shift quite well the data. Hence, in Figure 4(d), the mean curves of shifted curve (solid line) and of primary curves (dotted line) are almost the same ones.

Application to road traffic forecasting

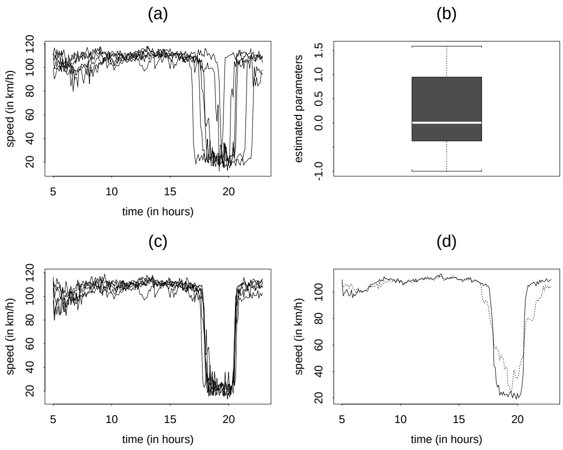

Figure 5: Shift estimation results on a particular traffic data set.

Most of the Parisian road traffic network is equipped with a traffic road measurement infrastructure. The main elements of this infrastructure are counting stations. These sensors are situated approximately every 500 meters on main trunk roads (motorways and speedways principally). Every counting station measures, daily, the average speed of vehicle flow on 6 minutes periods. We consider measurements from 5 AM to 11 PM, hence, the length of the daily measurement is 180. We note y_{ij} the speed measurement of day $j \in \{1, \ldots, J\}$ and of period $i \in \{1, \ldots, n\}$, with $n = 180$.

Our purpose is to improve, with shift estimation, an existing forecasting methodology. This forecasting methodology is described in [19]. This method is based on a classification method. We dispose of a sample of J speed curves and we want to summarize it by a small number N of standard profiles, representatives of each cluster.

Consider several clusters of J speed curves. Indeed, we note frequently that many subgroups are composed by curves describing the same behavior. For example, we observe a speed curve subgroup with a same traffic jam or speed reduction, but with a different start time of the phenomenon for each curve. Thus, Figure 5(a) represent a particular cluster on a particular counting station. Figure 5(b) is a boxplot of the estimated shifts. Shifted curves are plotted on Figure 5(c). So, in this homogeneous cluster, where only a shift phenomenon appear, the mean curves in figure 5(d) of shifted curves (solid line) and of primary curves (dotted line) aren't the same. The shift estimated mean is clearly more representative of individual pattern.

5. Technical Lemmas. The two following propositions, Proposition 5.1 and Proposition 5.2, are used in the proof of asymptotic normality (Theorem 3.1). Their proofs are postponed to the appendix.

PROPOSITION 5.1. *Assume that the assumptions of Theorem 2.1 are fulfilled. Then*

$$\sqrt{n}\nabla M_n(\alpha^*) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}_{J-1}(0,\Gamma), \tag{5.1}$$

where the variance matrix is $\Gamma = \frac{4}{J^2}\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2 \left(I_{J-1} - \frac{1}{J}U_{J-1}\right)$.

PROPOSITION 5.2. *Assume that the assumptions of Theorem 2.1 are fulfilled. Further, assume that* (3.1) *holds. Then, for any sequence* $(\bar{\alpha}_n)_{n\in\mathbb{N}}$ *with* $\|\bar{\alpha}_n - \alpha^*\| \leq \|\hat{\alpha}_n - \alpha^*\|$, *we have*

$$\nabla^2 M_n(\bar{\alpha}_n) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} \frac{2}{J^2}\sum_{l\in\mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2 \left(JI_{J-1} - U_{J-1}\right). \tag{5.2}$$

6. Appendix. Let z be a complex number and $\bar{z}$ its conjugate. We write $\mathfrak{Re}(z) = \frac{1}{2}(z+\bar{z})$ (the real part of z) and $\mathfrak{Im}(z) = \frac{1}{2i}(z-\bar{z})$ (the imaginary part of z).

PROOF 6.1 (Proof of Condition (2.8)). *Consider the following notation: for all* $j = 1,\ldots,J$, *for all* $l = -(n-1)/2,\ldots,(n-1)/2$, $w_{jl} = \frac{1}{\sqrt{n}}\xi_{jl} = \frac{1}{\sqrt{n}}\left(\xi_{jl}^x + i\xi_{jl}^y\right)$. *Here,* $\left(\xi_{jl}^x\right)$ *and* $\left(\xi_{jl}^y\right)$ *are independent Gaussian sequences,* *with law $\mathcal{N}_n(0, I_n)$. Also, set*

$$\forall l = -(n-1)/2, \ldots, (n-1)/2,\ c_l(f) = |c_l(f)|e^{i\theta_l},\ \text{with } \theta_l \in [0, 2\pi[.$$

Hence, the criterion function $M_n(\alpha)$ can be written as follows:

$$M_n(\alpha) = \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)|^2 - \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \left| \frac{1}{J} \sum_{j=1}^{J} e^{il(\alpha_j - \alpha_j^*)} \delta_l c_l(f) \right|^2 \tag{6.1}$$

$$+ \frac{J-1}{nJ^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \sum_{j=1}^{J} \left({\xi_{jl}^x}^2 + {\xi_{jl}^y}^2 \right) \tag{6.2}$$

$$- \frac{2}{nJ^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \sum_{j=1}^{J} \sum_{k>j} \Big[\cos(l(\alpha_j - \alpha_k))(\xi_{jl}^x \xi_{kl}^x + \xi_{jl}^y \xi_{kl}^y) +$$

$$\sin(l(\alpha_j - \alpha_k))(\xi_{jl}^x \xi_{kl}^y - \xi_{jl}^y \xi_{kl}^x) \Big] \tag{6.3}$$

$$+ \frac{2(J-1)}{\sqrt{n}J^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)| \sum_{j=1}^{J} \Big[\cos(\theta_l - l\alpha_j^*) \xi_{jl}^x + \sin(\theta_l - l\alpha_j^*) \xi_{jl}^y \Big] \tag{6.4}$$

$$- \frac{2}{\sqrt{n}J^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)| \sum_{j=1}^{J} \sum_{k \neq j} \Big[\cos(l(\alpha_j - \alpha_j^* - \alpha_k) + \theta_l) \xi_{kl}^x +$$

$$\sin(l(\alpha_j - \alpha_j^* - \alpha_k) + \theta_l) \xi_{kl}^y \Big]. \tag{6.5}$$

So we have split the criterion function into five terms: (6.1), (6.2), (6.3), (6.4) *and* (6.5). *We aim, in this proof, at giving their asymptotic behaviour while next proof is devoted to the study of the uniform convergence of their increments.*

The term (6.1) *is a deterministic one. Let ψ the T-periodic function defined by the Fourier serie $\psi(t) = \sum_{l \in \mathbb{Z}} \delta_l e^{2\pi i l t/T}$. The function ψ is well defined in $L^2([0,T])$ since $(\delta_l)_{l \in \mathbb{Z}} \in l^2(\mathbb{Z})$. Hence, using Parseval theorem, we have that*

$$(6.1) \xrightarrow[n \to +\infty]{} \int_0^T |(\psi * f)(t)|^2 \frac{dt}{T} - \int_0^T \left| \frac{1}{J} \sum_{j=1}^{J} (\psi * f)(t + \theta_j - \theta_j^*) \right|^2 \frac{dt}{T}.$$

The term (6.2) *is a pure noise term:*

$$(6.2) = \frac{J-1}{nJ^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \sum_{j=1}^{J} \left({\xi_{jl}^x}^2 + {\xi_{kl}^y}^2 \right) = \frac{J-1}{nJ^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \chi_{2J,l}^2,$$

where $\chi_{2J,l}^2$, $l = -(n-1)/2, \dots, (n-1)/2$, *is a sequence of i.i.d. chisquared random variables with* $2J$ *degrees of freedom. So,* (6.2) *will vanish asymptotically in probability. Indeed,* $\mathbb{E}((6.2)) = \frac{2(J-1)}{nJ} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \xrightarrow[n\to+\infty]{} 0$ *and* $\mathrm{Var}((6.2)) = \frac{4(J-1)^2}{n^2 J^3} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^4 \xrightarrow[n\to+\infty]{} 0$.

The term (6.3) *is also a pure noise term composed of terms of the type*

$$U_n = \frac{1}{n} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \sum_{j=1}^{J} \sum_{k>j} \cos\left(l(\alpha_j - \alpha_k)\right) \xi_{jl}^x \xi_{kl}^y,$$

with $\mathbb{E}(U_n) = 0$ *and*

$$\mathrm{Var}(U_n) = \frac{1}{n^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^4 \sum_{j=1}^{J} \sum_{k>j} \cos^2\left(l(\alpha_j - \alpha_k)\right) \le \frac{J^2 - J}{2n^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^4 \xrightarrow[n\to+\infty]{} 0.$$

So that (6.3) $\xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} 0$.

The terms (6.4) *and* (6.5) *have the same asymptotic behavior as*

$$V_n = \frac{1}{\sqrt{n}} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)| W_l,$$

with $W_l = \sum_{j=1}^{J} \sum_{k \neq j} \cos(l(\alpha_j - \alpha_j^* - \alpha_k) + \theta_l) \xi_{kl}^x$, *where, as before, the random sequences* $\left(\xi_{jl}^x\right)$ *are i.i.d. and following a Gaussian law* $\mathcal{N}_n(0, I_n)$. *Since* $(c_l(f))_{l\in\mathbb{Z}} \in l^2(\mathbb{Z})$, *we have that*

$$\sum_{l\in\mathbb{Z}} \delta_l^4 |c_l(f)|^2 \|W_l\|_2^2 < +\infty.$$

As a result, the random variables $\sqrt{n}V_n$ *converges in* $\mathbf{P}_{\alpha^*}$*-probability. Hence,* $V_n \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} 0$. *So, we can conclude the convergences of* (6.4) *and* (6.5)*:*

$$(6.4) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} 0 \text{ and } (6.5) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} 0.$$

In conclusion, we have the criterion convergence: $M_n(\alpha) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} K(\alpha)$, *with*

$$K(\alpha) = \int_0^T |(\psi * f)(t)|^2 \frac{\mathrm{d}t}{T} - \int_0^T \left| \frac{1}{J} \sum_{j=1}^J (\psi * f)\left(t + \theta_j - \theta_j^*\right) \right|^2 \frac{\mathrm{d}t}{T}.$$

Moreover, Cauchy-Schwartz inequality yields that

$$\int_0^T \left| \frac{1}{J} \sum_{j=1}^J (\psi * f)(t + \theta_j - \theta_j^*) \right|^2 \frac{\mathrm{d}t}{T}$$
$$\leq \int_0^T \frac{1}{J} \sum_{j=1}^J \left|(\psi * f)(t + \theta_j - \theta_j^*)\right|^2 \frac{\mathrm{d}t}{T} = \int_0^T |(\psi * f)(t)|^2 \frac{\mathrm{d}t}{T},$$

hence, $K(\cdot) \geq 0$, *and the minimum value is reached for*

$$\int_0^T |(\psi * f)(t)|^2 \frac{\mathrm{d}t}{T} = \int_0^T \left| \frac{1}{J} \sum_{j=1}^J (\psi * f)(t + \theta_j - \theta_j^*) \right|^2 \frac{\mathrm{d}t}{T},$$

which is equivalent, using Parseval theorem to

$$\sum_{l\in\mathbb{Z}} |\delta_l c_l(f)|^2 = \sum_{l\in\mathbb{Z}} \left| \frac{1}{J} \sum_{j=1}^J \delta_l c_l(f) e^{il(\alpha_j - \alpha_j^*)} \right|^2. \tag{6.6}$$

So, we have that

$$(6.6) \iff \forall l \in \mathbb{Z}, \left| \frac{1}{J} \sum_{j=1}^J e^{il(\alpha_j - \alpha_j^*)} \right|^2 = 1 \iff \forall j = 1, \ldots, J, \alpha_j = \alpha_j^* + c[2\pi], c \in \mathbb{R}.$$

In a matrix way, we get the equation (2.3), *i.e*

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_J \end{pmatrix} = \begin{pmatrix} \alpha_1^* \\ \vdots \\ \alpha_J^* \end{pmatrix} + c \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + 2\pi \begin{pmatrix} k_1 \\ \vdots \\ k_J \end{pmatrix}, \; c \in \mathbb{R}, \begin{pmatrix} k_1 \\ \vdots \\ k_J \end{pmatrix} \in \mathbb{Z}^J.$$

Hence, since $\alpha \in A$ *and* A *is defined by* (2.4), *we have shown that* $\alpha_j = \alpha_j^*$ *for all* $j = 1, \ldots, J$. *Since* $\alpha \mapsto K(\alpha)$ *achieves its unique minimum for* $\alpha = \alpha^*$, *the condition* (2.8) *is fulfilled.*

■

PROOF 6.2 (Proof of Condition (2.9)). *Set $\alpha = (\alpha_1, \ldots, \alpha_J)$ and $\beta = (\beta_1, \ldots, \beta_J)$ two translation parameters in A. Our aim is to prove that the convergence of the criterion function is uniform, in the sense that, for k large enough and two decreasing to zero sequences $(\eta_k)_{k\in\mathbb{N}}$ and $(\epsilon_k)_{k\in\mathbb{N}}$, we have $\lim_{n\to+\infty} \mathbf{P}\left(\sup_{\|\alpha-\beta\|\leq\eta_k} |M_n(\alpha) - M_n(\beta)| > \epsilon_k\right) = 0$. Using previous expression of the criterion function we write $M_n(\alpha) - M_n(\beta) = A + B + C$, where*

$$A = \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \left| \frac{1}{J} \sum_{j=1}^{J} e^{il(\beta_j - \alpha_j^*)} \delta_l c_l(f) \right|^2 - \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \left| \frac{1}{J} \sum_{j=1}^{J} e^{il(\alpha_j - \alpha_j^*)} \delta_l c_l(f) \right|^2$$

is the deterministic term, and $B + C$ is the stochastic term, with

$$\begin{aligned} B = &- \frac{2}{nJ^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{j=1}^{J} \sum_{k>j} \delta_l^2 \left[\cos(l(\alpha_j - \alpha_k)) - \cos(l(\beta_j - \beta_k))\right] (\xi_{jl}^x \xi_{kl}^x + \xi_{jl}^y \xi_{kl}^y) \\ &- \frac{2}{nJ^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{j=1}^{J} \sum_{k>j} \delta_l^2 \left[\sin(l(\alpha_j - \alpha_k)) - \sin(l(\beta_j - \beta_k))\right] (\xi_{jl}^x \xi_{kl}^y - \xi_{jl}^y \xi_{kl}^x), \end{aligned}$$

$$\begin{aligned} C = &- \frac{2}{\sqrt{n}J^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)| \sum_{j=1}^{J} \sum_{k\neq j} \Big[\cos(\theta_l + l(\alpha_j - \alpha_j^* - \alpha_k)) - \\ &\qquad \cos(\theta_l + l(\beta_j - \alpha_j^* - \beta_k))\Big] \xi_{kl}^x \\ &- \frac{2}{\sqrt{n}J^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)| \sum_{j=1}^{J} \sum_{k\neq j} \Big[\sin(\theta_l + l(\alpha_j - \alpha_j^* - \alpha_k)) - \\ &\qquad \sin(\theta_l + l(\beta_j - \alpha_j^* - \beta_k))\Big] \xi_{kl}^y. \end{aligned}$$

As a result, the stochastic term $B + C$ can be split into two categories of terms, composed, for all $l \in \{-(n-1)/2, \ldots, (n-1)/2\}$, of the centered independent random variables

$$U_l(\alpha, \beta) = \delta_l^2 \sum_{j=1}^{J} \sum_{k>j} \left[\cos(l(\alpha_j - \alpha_k)) - \cos(l(\beta_j - \beta_k))\right] \xi_{jl}^x \xi_{kl}^x,$$

$$V_l(\alpha, \beta) = \delta_l^2 |c_l(f)| \sum_{j=1}^{J} \sum_{k\neq j} \Big[\cos(\theta_l + l(\alpha_j - \alpha_j^* - \alpha_k)) - \cos(\theta_l + l(\beta_j - \alpha_j^* - \beta_k))\Big] \xi_{kl}^x.$$

All the remaining terms in B and C differ from U_l or V_l by the trigonometric functions, but their asymptotic behavior can be deduced in a similar way. Hence, our aim is to bound the probabilities

$$\mathbf{P}\left(\sup_{\|\alpha-\beta\|\leq\eta}\left|\frac{1}{n}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}U_l(\alpha,\beta)\right|>x\right), \tag{6.7}$$

$$\mathbf{P}\left(\sup_{\|\alpha-\beta\|\leq\eta}\left|\frac{1}{\sqrt{n}}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}V_l(\alpha,\beta)\right|>x\right). \tag{6.8}$$

Using the basic inequality $|\cos(p)-\cos(q)|\leq|p-q|$ and after some calculations, we have that

$$\mathrm{Var}\left(\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}U_l(\alpha,\beta)\right)\leq J^2\eta^2\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}\delta_l^4 l^2,$$

$$\mathrm{Var}\left(\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}V_l(\alpha,\beta)\right)\leq J^2\eta^2\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}\delta_l^4 l^2|c_l(f)|^2.$$

Moreover, using exponential bounded moments of random variables U_l and V_l gives, we get from Bernstein's inequality for independent variables, that there exists a positive constant $M<+\infty$, such that

$$\mathbf{P}\left(\left|\frac{1}{n}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}U_l(\alpha,\beta)\right|>x\right)\leq 2\exp\left(-\frac{n^2x^2}{2\left(\mathrm{Var}\left(\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}U_l(\alpha,\beta)\right)+Mnx\right)}\right),$$

$$\mathbf{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}V_l(\alpha,\beta)\right|>x\right)\leq 2\exp\left(-\frac{nx^2}{2\left(\mathrm{Var}\left(\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}V_l(\alpha,\beta)\right)+M\sqrt{n}x\right)}\right).$$

Hence, using (2.6) *and* (2.7)*, we get that both previous probabilities go to zero when n goes to infinity. Now, since α and β lie in a compact set of $\mathbb{R}^J$, with $J<+\infty$ fixed, a standard chaining argument (see for instance [27]) proves that there exist $(\eta_k)_{k\in\mathbb{N}}$ and $(\epsilon_k)_{k\in\mathbb{N}}$ such that for k large enough we* *get*

$$\lim_{n\to+\infty} \mathbf{P}\left(\sup_{\|\alpha-\beta\|\le\eta_k} \left| \frac{1}{n} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} U_l(\alpha,\beta) \right| > \epsilon_k \right) = 0,$$
$$\lim_{n\to+\infty} \mathbf{P}\left(\sup_{\|\alpha-\beta\|\le\eta_k} \left| \frac{1}{\sqrt{n}} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} V_l(\alpha,\beta) \right| > \epsilon_k \right) = 0.$$

The deterministic part A remains to be studied. We can write, after some calculations, that

$$|A| \le \eta \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |l| |c_l(f)|^2. \tag{6.9}$$

But, since (2.6) *and* (2.7) *hold, using Cauchy-Schwartz inequality, we have:* $\|f\|_\delta^2 = \sum_{l\in\mathbb{Z}} \delta_l^2 |l||c_l(f)|^2 \le \left(\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2\right)^{\frac{1}{2}} \left(\sum_{l\in\mathbb{Z}} |c_l(f)|^2\right)^{\frac{1}{2}} < +\infty$. *Hence, as soon as we have chosen two decreasing sequences* $(\eta_k)_{k\in\mathbb{N}}$ *and* $(\epsilon_k)_{k\in\mathbb{N}}$ *such that* $\epsilon_k/\eta_k > \|f\|_\delta^2$, *we have that* $\lim_{n\to+\infty} \mathbf{P}\left[W(n,\eta_k) > \epsilon_k\right] = 0$, *which concludes the proof.*

We point out that choosing the weights $\delta_l \to 0$ enables to prove consistency of the terms (6.4) *and* (6.5)*. As a matter of fact, processes of the type* $Z_n(\alpha) = \frac{1}{\sqrt{n}} \sum_{l=-n}^{n} \cos(l\alpha)\xi_l$ *can be studied using an ergodic theorem. If* $\alpha \notin \mathbb{Q}$, *the dynamical system* $T : X \to X + \alpha$ *over* $\mathbb{T}$ *has an invariant measure* λ *and the only stable sets are* $\mathbb{T}$ *and the null space. Set* $T : [0,T]^J \longrightarrow [0,T]^J$, $(\theta_1,\ldots,\theta_J) \longrightarrow (\theta_1+\alpha_1,\ldots,\theta_J+\alpha_J)$. *Since*

$$\mathrm{Var}(Z_n(\alpha)) \to \kappa \int_0^{2\pi} \cos^2\theta \, \mathrm{d}\theta = \sigma^2,$$
$$\mathrm{Cov}\left(Z_n(\alpha), Z_n(\alpha')\right) \to \kappa \int_{[0,2\pi]^2} \cos\theta_1 \cos\theta_2 \, \mathrm{d}\theta_1 \mathrm{d}\theta_2 = 0,$$

an ergodic theorem for this dynamical system entails that the marginals of the process converge to a white noise,

$$(Z_n(\alpha_1),\ldots,Z_n(\alpha_J)) \xrightarrow{\mathcal{L}} \mathcal{N}_J(0, \sigma^2 Id_J),$$

but the empirical process does not converge. Hence, adding vanishing weights to the empirical loss function enables to get rid of this random part in the limit contrast.

■

PROOF 6.3 (Proof of Proposition 5.1). *After some calculations, we get the following expressions for the first and the second derivatives of the empirical contrast, for all $k \in \{2, \dots, J\}$, for all $m \in \{2, \dots, J\}$:*

$$\frac{\partial M_n}{\partial \alpha_k}(\alpha) = \frac{2}{J} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l \mathfrak{Im}\left(\tilde{c}_{kl}(\alpha)\overline{\hat{c}_l(\alpha)}\right), \tag{6.10}$$

$$\frac{\partial^2 M_n}{\partial \alpha_k^2}(\alpha) = \frac{2}{J^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \mathfrak{Re}\left(\tilde{c}_{kl}(\alpha) \sum_{j\neq k} \overline{\tilde{c}_{jl}(\alpha)}\right), \tag{6.11}$$

$$\forall m \neq k,\ \frac{\partial^2 M_n}{\partial \alpha_k \partial \alpha_m}(\alpha) = -\frac{2}{J^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \mathfrak{Re}\left(\tilde{c}_{kl}(\alpha)\overline{\tilde{c}_{ml}(\alpha)}\right). \tag{6.12}$$

By straightforward calculations, we get that

$$\sqrt{n}\frac{\partial M_n}{\partial \alpha_k}(\alpha^*) = \frac{2}{J} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l \left(|c_l(f)| \left(V_l^k - V_l\right) + W_l^k\right),$$

where, for all $l \in \mathbb{Z}$,

$$W_l^k = \frac{1}{J\sqrt{n}} \sum_{j=1}^{J} \left[\sin(l(\alpha_k^* - \alpha_j^*))(\xi_{kl}^x \xi_{jl}^x + \xi_{kl}^y \xi_{jl}^y) + \cos(l(\alpha_k^* - \alpha_j^*))(\xi_{kl}^y \xi_{jl}^x - \xi_{kl}^x \xi_{jl}^y)\right],$$

$V_l^k = \left(\cos(l\alpha_k^* + \theta_l)\xi_{kl}^x - \sin(l\alpha_k^* + \theta_l)\xi_{kl}^y\right)$ *,and* $V_l = \frac{1}{J}\sum_{j=1}^{J} V_l^j$.
Let, for $l \in \mathbb{Z}$, $Y_l = \left(\xi_{1l}^x\, \xi_{2l}^x \cdots \xi_{Jl}^x\, \xi_{1l}^y\, \xi_{2l}^y \cdots \xi_{Jl}^y\right)'$, and, let f_l^k be the vector of length $2J$, defined by $\left(f_l^k\right)_k = \cos\left(l\alpha_k^* + \theta_l\right)$, $\left(f_l^k\right)_{J+k} = -\sin\left(l\alpha_k^* + \theta_l\right)$, and $\left(f_l^k\right)_i = 0$ if $i \notin \{k, J+k\}$. As a consequence, we get the following expression for V_l^k: $V_l^k = \left\langle f_l^k, Y_l\right\rangle = f_l^{k\prime} Y_l$. In a same way, for $l \in \mathbb{Z}$, let $\bar{B}_l^k$ be the $(2J) \times (2J)$ matrix defined by rows by

$$\left(\bar{B}_l^k\right)_k = (\sin[l(\alpha_k^* - \alpha_1^*)] \cdots \sin[l(\alpha_k^* - \alpha_J^*)]\ -\cos[l(\alpha_k^* - \alpha_1^*)] \cdots -\cos[l(\alpha_k^* - \alpha_J^*)]),$$
$$\left(\bar{B}_l^k\right)_{J+k} = (\cos[l(\alpha_k^* - \alpha_1^*)] \cdots \cos[l(\alpha_k^* - \alpha_J^*)]\ \sin[l(\alpha_k^* - \alpha_1^*)] \cdots \sin[l(\alpha_k^* - \alpha_J^*)]),$$
$$\left(\bar{B}_l^k\right)_i = (0 \cdots 0) \ \text{ if } i \notin \{k, J+k\}.$$

Further, let the symmetric matrix B_l^k be defined by $B_l^k = \frac{\bar{B}_l^k + \left(\bar{B}_l^k\right)'}{2}$. Hence, we may write $W_l^k = \frac{1}{J\sqrt{n}} Y_l' B_l^k Y_l$. Now, we define, for $k = 2, \dots, J$: $\tilde{B}_l^k = \frac{2}{J} B_l^k$, $\tilde{f}_l^k = \frac{2}{J} f_l^k$.

Our aim is to study the asymptotic distribution of the gradient $\sqrt{n}\nabla M_n(\alpha^*)$. *For this purpose, we consider* $u = (u_2, \ldots, u_J)' \in \mathbb{R}^{J-1}$ *and* $t = (t_2, \ldots, t_J)' \in \mathbb{R}^{J-1}$, *and we define the couple of random variables:*

$$(R_n, S_n) = \left(\frac{2}{J} \sum_{k=2}^{J} u_k \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l |c_l(f)| (V_l^k - V_l), \frac{2}{J} \sum_{k=2}^{J} t_k \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l W_l^k \right).$$

Using previous notations, we get

$$R_n = \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l |c_l(f)| \langle g_l(u), Y_l \rangle, \text{ with } g_l(u) = \sum_{k=2}^{J} u_k \left(\tilde{f}_l^k - \frac{1}{J} \sum_{j=1}^{J} \tilde{f}_l^j \right),$$

$$S_n = \frac{1}{J\sqrt{n}} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l Y_l' A_l(t) Y_l, \text{ with } A_l(t) = \sum_{k=2}^{J} t_k \tilde{B}_l^k.$$

The quadratic part is vanishing in probability when n *increases. Indeed, there exists a positive constant* $C(t)$, *depending on* t, *such that* $\|A_l(t)\| \leq C(t)$, *and, this leads to an upper bound for the quadratic term:*

$$\mathbb{E}(|S_n|) \leq \frac{C(t)}{J\sqrt{n}} \mathbb{E}(\|Y_1\|) \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |l|.$$

Moreover, since (3.1) *holds and using Cauchy-Schwartz inequality, we have:*

$$\sum_{l \in \mathbb{Z}^*} \delta_l^2 |l| \leq \left(\sum_{l \in \mathbb{Z}^*} \delta_l^4 l^4 \right)^{\frac{1}{2}} \left(\sum_{l \in \mathbb{Z}^*} \frac{1}{l^2} \right)^{\frac{1}{2}} < +\infty.$$

Hence, $\mathbb{E}(|S_n|) \xrightarrow[n \to +\infty]{} 0$. *So that* $S_n \xrightarrow[n \to +\infty]{\mathbf{P}_{\alpha^*}} 0$. *For the last term, we have that* $\langle g_l(u), Y_l \rangle \sim \mathcal{N}\left(0, \|g_l(u)\|_2^2\right)$, *where* $\|g_l(u)\|_2^2 = \frac{4}{J^2} u' \left(I_{J-1} - \frac{1}{J} U_{J-1} \right) u$, *with,* I_{J-1} *the* $J-1$ *identity matrix, and* U_{J-1} *the* $J-1 \times J-1$ *matrix which all entries are equal to 1. Independence of variables* $(Y_l)_{l \in \mathbb{Z}}$ *yields that*

$$R_n \xrightarrow[n \to +\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{4}{J^2} \sum_{l \in \mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2 u' \left(I_{J-1} - \frac{1}{J} U_{J-1} \right) u \right).$$

Hence, we have proved that

$$\sqrt{n}\nabla M_n(\alpha^*) \xrightarrow[n \to +\infty]{\mathcal{D}} \mathcal{N}_{J-1}\left(0, \frac{4}{J^2} \sum_{l \in \mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2 \left(I_{J-1} - \frac{1}{J} U_{J-1} \right) \right).$$

∎

PROOF 6.4 (Proof of Proposition 5.2). *First, we pay attention to the non diagonal terms of the matrix of the second derivatives. For $m \neq k$, we get after some calculations:*

$$
\begin{aligned}
-\frac{J^2}{2}\frac{\partial^2 M_n}{\partial\alpha_k\partial\alpha_m}(\alpha) &= \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \mathfrak{Re}\left(\tilde{c}_{kl}(\alpha)\overline{\tilde{c}_{ml}(\alpha)}\right) \\
&= \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)|^2 \cos\left(l[\alpha_k - \alpha_k^* + \alpha_m^* - \alpha_m]\right) && (6.13) \\
&+ \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)| \left(\cos[l(\alpha_k - \alpha_k^* - \alpha_m) + \theta_l] w_{ml}^x + \right. \\
&\qquad \left. \sin[l(\alpha_k - \alpha_k^* - \alpha_m) + \theta_l] w_{ml}^y\right) && (6.14) \\
&+ \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)| \left(\cos[l(\alpha_m - \alpha_m^* - \alpha_k) + \theta_l] w_{kl}^x + \right. \\
&\qquad \left. \sin[l(\alpha_m - \alpha_m^* - \alpha_k) + \theta_l] w_{kl}^y\right) && (6.15) \\
&+ \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \left[\cos(l[\alpha_k - \alpha_m])(w_{kl}^x w_{ml}^x + w_{kl}^y w_{ml}^y) - \right. \\
&\qquad \left. \sin(l[\alpha_k - \alpha_m])(w_{kl}^y w_{ml}^x - w_{kl}^x w_{ml}^y)\right]. && (6.16)
\end{aligned}
$$

We now study the asymptotic behaviour of each term separately. Indeed, the second derivatives are taken at a point $\bar{\alpha}_n$ which converges to α^*: $\bar{\alpha}_n$ is in the neighborhood of α^* with radius $\|\alpha^* - \hat{\alpha}_n\|$. Hence, we need conditions to claim uniform convergence of $\nabla^2 M_n(\cdot)$.

In a similar way as in the proof of Condition (2.9)*, as we assume that*

(3.1) *holds and using Cauchy-Schwartz inequality, we have that*

$$\sum_{l\in\mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2 \leq \left(\sum_{l\in\mathbb{Z}} \delta_l^4 l^4 |c_l(f)|^2\right)^{\frac{1}{2}} \left(\sum_{l\in\mathbb{Z}} |c_l(f)|^2\right)^{\frac{1}{2}} < +\infty,$$

hence, the deterministic term (6.13) *converges, uniformly in the variable* α.

Since for all $k \in \{1, \dots, J\}$, *the random variables* w_{kl}^x *and* w_{kl}^y *follow a Gaussian law* $\mathcal{N}(0, 1/n)$, *we consider the independent variables* ξ_{kl}^x *and* ξ_{kl}^y *such that* $w_{kl}^x = \frac{1}{\sqrt{n}}\xi_{kl}^x$ *and* $w_{kl}^y = \frac{1}{\sqrt{n}}\xi_{kl}^y$. *For the two second terms* (6.14) *and* (6.15), *we write*

$$(6.14) = \frac{1}{\sqrt{n}} \sum \delta_l^2 l^2 |c_l(f)| \left(\cos[l(\alpha_k - \alpha_k^* - \alpha_m) + \theta_l]\xi_{ml}^x + \sin[l(\alpha_k - \alpha_k^* - \alpha_m) + \theta_l]\xi_{ml}^y\right),$$

$$(6.15) = \frac{1}{\sqrt{n}} \sum \delta_l^2 l^2 |c_l(f)| \left(\cos[l(\alpha_m - \alpha_m^* - \alpha_k) + \theta_l]\xi_{kl}^x + \sin[l(\alpha_m - \alpha_m^* - \alpha_k) + \theta_l]\xi_{kl}^y\right).$$

Using the assumptions (2.6) *and* (3.1), *there exists a random variable* Y *such that*

$$Y_n = \sqrt{n}(6.15) \xrightarrow[n\to+\infty]{\|\cdot\|_2} Y.$$

So that $\frac{1}{\sqrt{n}}Y_n$ *converges in probability to* 0. *Hence, in a similar way as in the proof of Condition* (2.9), *a standard chaining argument proves that, for all* $\lambda > 0$, *we have* $\mathbf{P}\left(\sup_{\alpha\in\tilde{A}_1} (6.15) > \lambda\right) \xrightarrow[n\to+\infty]{} 0$, *which leads to:* $\sup_{\alpha\in\tilde{A}_1} (6.15) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} 0$. *In the same way, we may also conclude that* (6.14) *and further* (6.16) *both converge uniformly to* 0. *The diagonal terms can be* *written as follows:*

$$\frac{J^2}{2}\frac{\partial^2 M_n}{\partial \alpha_k^2}(\alpha) = \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \mathfrak{Re}\left(\tilde{c}_{kl}\sum_{j\neq k}\overline{\tilde{c}_{jl}}\right)$$

(6.17)

$$= \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)|^2 \sum_{j\neq k}\cos\left(l[\alpha_k-\alpha_k^*+\alpha_j^*-\alpha_j]\right)$$

$$+ \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)| \sum_{j\neq k}\Big(\cos[l(\alpha_k-\alpha_k^*-\alpha_j)+\theta_l]w_{jl}^x +$$

(6.18)

$$\sin\left[l(\alpha_k-\alpha_k^*-\alpha_j)+\theta_l\right]w_{jl}^y\Big)$$

$$+ \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)| \sum_{j\neq k}\Big(\cos[l(\alpha_j-\alpha_j^*-\alpha_k)+\theta_l]w_{kl}^x +$$

(6.19)

$$\sin\Big[l(\alpha_j-\alpha_j^*-\alpha_k)+\theta_l\Big]w_{kl}^y\Big)$$

$$+ \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \sum_{j\neq k}\Big[\cos(l[\alpha_k-\alpha_j])(w_{kl}^x w_{jl}^x + w_{kl}^y w_{jl}^y) -$$

(6.20)

$$\sin(l[\alpha_k-\alpha_j])(w_{kl}^y w_{jl}^x - w_{kl}^x w_{jl}^y)\Big].$$

Using similar arguments as for the previous terms, we can see that, under the same assumptions we get that all the terms (6.17), (6.18), (6.19) *and* (6.20) *converges uniformly, and we get*

$$\frac{\partial^2 M_n}{\partial \alpha_k^2}(\bar{\alpha}_n) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} \frac{2(J-1)}{J^2}\sum_{l\in\mathbb{Z}}\delta_l^2 l^2 |c_l(f)|^2.$$

As a result, gathering the two previous results leads to the following asymptotic behavior:

$$\nabla^2 M_n(\bar{\alpha}_n) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} \frac{2}{J^2}\sum_{l\in\mathbb{Z}}\delta_l^2 l^2 |c_l(f)|^2\left(JI_{J-1}-U_{J-1}\right),$$

which proves the result. Moreover, this matrix is invertible. As a result, we have that

$$\left[\nabla^2 M_n\left(\bar{\alpha}_n\right)\right]^{-1} \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} \frac{J}{2\sum_{l\in\mathbb{Z}}\delta_l^2 l^2 |c_l(f)|^2}\left(I_{J-1}+U_{J-1}\right).$$

■

REFERENCES

[1] R. J. Beerends, H. G. ter Morsche, J. C. van den Berg, and E. M. van de Vrie. *Fourier and Laplace transforms*. Cambridge University Press, Cambridge, 2003. Translated from the 1992 Dutch edition by Beerends.

[2] J. Boularan, L. Ferré, and P. Vieu. Growth curves: a two-stage nonparametric approach. *J. Statist. Plann. Inference*, 38(3):327–350, 1994.

[3] I. Castillo and J-M. Loubes. Semiparametric estimation of random shifts. *Prépublications de l'université Paris Sud 31-2005*, 2005.

[4] D. Chafai and J-M. Loubes. Nonparametric maximum likelihood estimator for a certain class of inverse problems. *Prépublications de l'université Paris Sud 13-2004*, 2004.

[5] A.S. Dalalyan, G.K. Golubev, and A.B. Tsybakov. Bayesian maximum likelihood and semiparametric second order efficiency. Prépublication numéro 842 du laboratoire de probabilités et modèles aléatoires, Université Paris 6 et Paris 7.

[6] D.L. Donoho and I.M. Johnstone. Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields*, 99(2):277–303, 1994.

[7] T. Gasser and A. Kneip. Searching for structure in curve samples. *J. Amer. Statist. Assoc.*, 90:1179–1188, 1995.

[8] E. Gassiat and C. Lévy-Leduc. Efficient semi-parametric estimation of the periods in a superposition of periodic functions with unknown shape. preprint.

[9] V. Genon-Catalot and D. Picard. *Éléments de statistique asymptotique*, volume 11 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Paris, 1993.

[10] G. Golubev. Estimation of the period of a signal with an unknown form against a white. *Problemy Peredachi Informatsii*, 24:38–52, 1988.

[11] W. Härdle and J. S. Marron. Semiparametric comparison of regression curves. *Ann. Statist.*, 18(1):63–89, 1990.

[12] C. Ke and Y. Wang. Semiparametric nonlinear mixed-effects models and their applications. *J. Amer. Statist. Assoc.*, 96(456):1272–1298, 2001. With comments and a rejoinder by the authors.

[13] A. Kneip and J. Engel. Model estimation in nonlinear regression under shape invariance. *Ann. Statist.*, 23(2):551–570, 1995.

[14] A. Kneip and T. Gasser. Convergence and consistency results for self-modeling nonlinear regression. *Ann. Statist.*, 16(1):82–112, 1988.

[15] A. Kneip and T. Gasser. Statistical tools to analyze data representing a sample of curves. *Ann. Statist.*, 20(3):1266–1305, 1992.

[16] A. Kneip, X. Li, K.B. MacGibbon, and J.O. Ramsay. Curve registration by local regression. *Canad. J. Statist.*, 28(1):19–29, 2000.

[17] W.H. Lawton, E.A. Sylvestre, and M.S. Maggio. Self-modeling regression. *Technometrics.*, 14:513–532, 1972.

[18] M. Lindstrom. Self modeling with random scale and shift parameters and a free-knot spline shape function. *Statistics in Medicine*, 14:2009–2021, 1995.
[19] J-M. Loubes, E. Maza, and M. Lavielle. Road trafficking description and short term travel time forecasting, with a classification method. *Preprint*, 2002.
[20] V. Núñez-Antón, J.M. Rodríguez-Póo, and P. Vieu. Longitudinal data with nonstationary errors: a nonparametric three-stage approach. *Test*, 8(1):201–231, 1999.
[21] M. Piccioni, S. Scarlatti, and A. Trouvé. A variational problem arising from speech recognition. *SIAM J. Appl. Math.*, 58(3):753–771 (electronic), 1998.
[22] J.O. Ramsay and X. Li. Curve registration. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(2):351–363, 1998.
[23] J.O. Ramsay, X. Wang, and R. Flanagan. A functional data analysis of the pinch force of human fingers. *Appl. Statist.*, 44(1):17–30, 1995.
[24] C. Rao. Some statistical methods for the comparison of growth curves. *Biometrics*, 14:1–17, 1958.
[25] J.A. Rice and B.W. Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B*, 53(1):233–243, 1991.
[26] B.W. Silverman. Incorporating parametric effects into functional principal components analysis. *J. Roy. Statist. Soc. Ser. B*, 57(4):673–689, 1995.
[27] S.A. van de Geer. *Applications of empirical process theory.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2000.
[28] A.W. van der Vaart. *Asymptotic statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.
[29] M. Vimont. Asymtotic efficiency of shifts estimators with M-estimators. *manuscript*, 2005.
[30] K. Wang and T. Gasser. Alignment of curves by dynamic time warping. *Ann. Statist.*, 25(3):1251–1276, 1997.
[31] K. Wang and T. Gasser. Synchronizing sample curves nonparametrically. *Ann. Statist.*, 27(2):439–460, 1999.

Equipe de probabilités et statistique
Département de mathématiques
UMR 5149, Université Montpellier 2
Place Eugène Bataillon
34000 Montpellier
France
e-mail: Jean-Michel.Loubes@math.univ-montp2.fr
url: www.math.u-psud.fr/~loubes