Semiparametric curve alignment and shift density estimation for biological data

T. Trigano, U. Isserles and Y. Ritov

Abstract

Assume that we observe a large number of signals, all of them with identical, although unknown, shape, but with a different random shift. The objective is to estimate the individual time shifts and their distribution. Such an objective appears in several biological applications like neuroscience or ECG signal processing, in which the estimation of the distribution of the elapsed time between repetitive pulses with a possibly low signal-noise ratio, and without a knowledge of the pulse shape is of interest. We suggest an M-estimator leading to a three-stage algorithm: we first split our data set in blocks, then the shift estimation in each block is done by minimizing a cost function based on the periodogram; the estimated shifts are eventually plugged into a standard density estimator. We show that under mild regularity assumptions the density estimate converges weakly to the true shift distribution. The theory is applied both to simulations and to alignment of real ECG signals. The proposed approach outperforms the standard methods for curve alignment and shift density estimation, even in the case of low signal-to-noise ratio, and is robust to numerous perturbations common in ECG signals.

Index Terms

semiparametric methods, density estimation, shift estimation, ECG data processing, nonlinear inverse problems.

I. INTRODUCTION

We investigate in this paper a specific class of stochastic nonlinear inverse problems. We observe a collection of $M + 1$ uniformly sampled signals in a finite interval $[0, T]$

$$y_j(t_i) = s(t_i - \theta_j) + \sigma\varepsilon_j(t_i), \ t_i \in [0, T], \ j = 0 \ldots M \tag{1}$$

where s is an unknown signal, $\{\theta_j, j = 0 \ldots M\}$ are independent real-valued continuous random variables with common probability density function f which represent a shift parameter, and $\varepsilon_0, \ldots, \varepsilon_M$ are independent standard white noise processes with standard deviation σ and independent of $\theta_0, \ldots, \theta_M$. Our aim is to estimate either $\{\theta_j, j = 0 \ldots M\}$, or the shift distribution f. Similar models appear commonly in practice in numerous fields. For instance, a common problem in functional

data analysis (FDA) is to align curves obtained in a series of experiments with varying time shifts, before extracting their common features; we refer to [1] and [2] for an in-depth discussion on the problem of curve alignment in FDA applications. In data mining applications, after splitting the data into different homogeneous clusters, observations of a same cluster may differ. Such variations take into account the variability of individual waveforms inside one given group. In the framework described by (1), the knowledge of the translation parameter θ, and more specifically of its distribution, can be used to determine the inner variability of a given cluster of curves. Several papers (e.g. [3], [4], [5], [6], [7]) focus on this specific model for many different applications in biology or signal processing. Such a problem can also be related to curve alignment problems as in [1], which are typically encountered in medicine (growth curves) and traffic data. Many methods previously introduced rely on a preliminary estimation of s, thus introducing an additional error in the estimation of $\{\theta_j, j = 0 \ldots M\}$. For example, [6] proposed to estimate the shifts by aligning the maxima of the curves, their position being approximated by the zeros of a kernel estimate of the derivative. Similar discussions can be found in recent contributions in the system identification framework, e.g. [8], [9], [10]. In particular, [9] provides a two-stage algorithm which estimates jointly a parametric component and a functional. Since we do not rely on any information or estimation regarding s in this paper, it is of interest to consider it as a nuisance parameter, and the shifts $\{\theta_j, j = 0 \ldots M\}$ (or f) as a parameter of interest, and consider (1) as a semiparametric model as described in [11]. However, if one is interested in s while having estimates of the shifts $\hat{\theta}_1, \ldots, \hat{\theta}_M$, one can easily proceed and use $\hat{s}(t) = M^{-1} \sum_{j=0}^{M} y_j(t + \hat{\theta}_j)$ as an estimate of the signal s.

Our contribution is close to recent shift estimation techniques described in [12] and [13]. Both rely on the fact that the spectral density of one given signal remains invariant by shifting, and therefore, it is well fitted for semiparametric methods when s is unknown. In [12], the problem addressed is the joint estimation of K shifts parameters, when K is a fixed number of curves (unlike what is done in the current paper where $K \to \infty$). This leads to a semiparametric estimation technique similar to the papers of [14], [15]. The advantage of such an estimator is that it is asymptotically efficient, consistent and asymptotically normal. However, when the number of curves to process is important, the method leads to a computationally intensive optimization problem. It is therefore of interest in practical applications to deal with blocks of smaller size which include one identical reference curve, as done in this paper. In [13], the authors estimate the shift probability density function when the number of curves is infinite, but the corresponding alignment procedure is performed one curve after the other, by means of the minimization of a penalized likelihood function. Such an approach makes sense when we have a few curves to compare, but when we dispose of many signals, the shifts parameters may be estimated jointly and more efficiently.

In our main application, we analyze ECG signals. We aim at situations where the heart electrical

activity remains regular enough in the sense that the shape of each cycle remains approximately repetitive, so that after prior segmentation of the ECG recording, the above model still holds. This is the case for heart malfunctions such as sinus or supraventricular tachycardia, as mentioned in [16]. This preliminary segmentation can be done efficiently, for example, by taking segments around the easily identified maxima of the QRS complex, as it can be found in [6], or by means of digital filters as suggested in [17]. It is of interest to estimate $\{\theta_j, j = 0 \ldots M\}$, in (1), since these estimates can be used afterwards for a more accurate estimation of the heart rate distribution. Another measurement often used by cardiologists is the mean ECG signal. A problem encountered in that case is that improperly aligned curves can yield an average on which the characteristics of the ECG cycle are lost. The proposed method leads to a more efficient estimation of the mean cycle by averaging the segments after an alignment according to a well-estimated shift.

The paper is organized as follows. Section II describes the assumptions made and the method to derive the estimators of the shifts and of their distribution. This method is based on the optimization of a cost function, based on the comparison between the power spectrum of the average of blocks of curves and the average of the individual power spectra. Since we consider a large number of curves, we expect that taking the average signal will allow to minimize the cost criterion consistently. We provide in Section III theoretical results on the efficiency of the method and on the weak convergence of the density estimate. In Section IV, we present simulations results, which show that the proposed algorithm performs well for density estimation, and study its performances under different conditions. We also applied the methodology to the alignment of ECG signals, and show that the proposed algorithm outperforms the standard FDA methods. Proofs of the discussed results are presented in the appendix.

II. NONPARAMETRIC ESTIMATION OF THE SHIFT DISTRIBUTION

In this section, we state the main assumptions that will be used in the rest of the paper, and propose an algorithm which leads to an M-estimator of the shifts. Using these estimators, we obtain a plug-in estimate of the shift probability density function.

A. Assumptions

Assume that we observe $M + 1$ sampled noisy signals on a finite time interval $[0, T]$, each one being shifted randomly by θ; a typical signal is expressed by

$$
\begin{aligned}
y_j(t_i) &= s(t_i - \theta_j) + \sigma\varepsilon_j(t_i), \\
t_i &= \frac{(i-1)T}{n}, \quad i = 1 \ldots n, \ j = 0 \ldots M,
\end{aligned}
\tag{2}
$$

where the processes $\{\varepsilon_j, j = 0 \ldots M\}$ are assumed to be standard Gaussian white noises, and the variance σ^2 is assumed to be constant. We also assume that the whole signal is within the sampling frame, which can be formalized by the following assumption:

(H-1) The distribution of θ and the shape s both have bounded non-trivial support, $[0, T_\theta]$ and $[0, T_s]$, respectively, and $T_\theta + T_s < T$.

As pointed out in [18], under this assumption we can consider s as a periodic function with associated period T. Without any loss of generality, we further assume that $T \triangleq 2\pi$ in order to simplify notations. We also assume:

(H-2) $s \in L^2([0, T_s])$ and its derivative $s' \in L^\infty$.

Assumption (H-1) implies that we observe a sequence of identical curves with additive noise, so that the spectral information is the same for all curves. Assumption (H-2) is critical to guarantee the existence of the Energy Spectral Density (ESD) of the studied curve and of the terms appearing in later sections. Note that the boundedness of the derivative is assumed for the sake of convenience (in order to show easily that the discretization error in the later parts can be neglected); the proposed method would also give good results on curves showing discontinuities. We finally make the following assumptions on the random variables appearing in (2):

(H-3) The shifts $\{\theta_j, j = 0 \ldots M\}$ are continuous random variables, independent and identically distributed with common probability density function f which is assumed to be uniformly bounded. We also consider the first shift θ_0 as known, and without loss of generality we fix $\theta_0 \triangleq 0$. Finally, we assume that the variables $\{\varepsilon_j(t_i), j = 0, \ldots, M, i = 1, \ldots, n\}$ are standard normal independent random variables, which are also independent of $\{\theta_j, j = 0 \ldots M\}$.

B. Computation of the shift estimators and of their density

The intuitive idea of the proposed algorithm is as follows. Assume, for the sake of the argument, that $\sigma = 0$; then, when the shifts are known and corrected, the individual signals are equal to their average. Consequently, the average of their ESDs is equal to the ESD of the mean signal. On the other hand, if the shifts are not corrected, then the average signal is a convolution of the original shape with the shift distribution, and hence its ESD is strictly different from the average of the individual ESD's.

Following the method of [13], we propose to plug estimators of $\{\theta_j, j = 1 \ldots M\}$ into an estimate of f. We start by splitting our dataset in N blocks of $K + 1$ curves each, as shown in Figure 1. Observe that y_0 is included in each block, since all the rest of the signals are aligned with it. The motivation to split the dataset into smaller blocks is twofold: it reduces the variance of the estimators of the
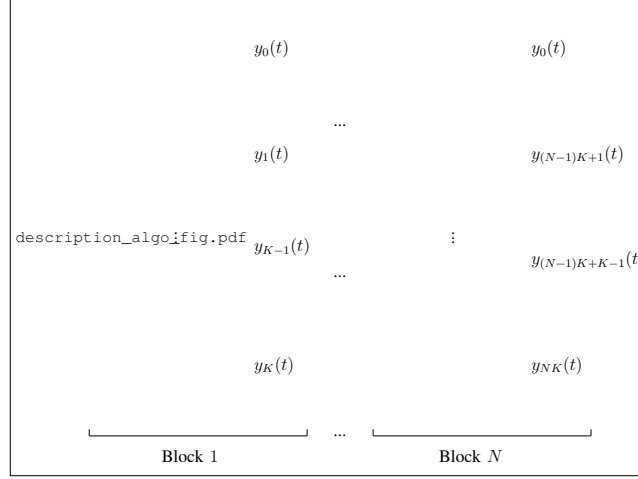
Fig. 1. Split of the curves data set

shifts by estimating them jointly, and also provides smooth functions for the optimization procedure detailed in this section. The first step is therefore to estimate the vectors of shifts $\{\boldsymbol{\theta}_m, m = 1 \ldots N\}$, where for all integer m, $\boldsymbol{\theta}_m \triangleq (\theta_{(m-1)K+1}, \ldots, \theta_{mK})$.

The estimation of $\boldsymbol{\theta}_m$ is achieved by minimizing a cost function. For any continuous-time and 2π-periodic signal y, we denote by S_y its energy spectral density, that is for all ω:

$$S_y(\omega) \triangleq \left| \frac{1}{2\pi} \int_0^{2\pi} y(t) \mathrm{e}^{-i\omega t} \, dt \right|^2 . \tag{3}$$

This quantity is of interest, since it remains invariant by shifting. For each integer $m = 1 \ldots N$, we define the mean of K signals translated by some correction terms $\boldsymbol{\alpha}_m \triangleq (\alpha_{(m-1)K+1}, \ldots, \alpha_{mK})$:

$$\bar{y}_m(t; \boldsymbol{\alpha}_m) \tag{4}$$

$$\triangleq \frac{1}{K + \lambda} \left(\lambda y_0(t) + \sum_{l=(m-1)K+1}^{mK} y_l(t + \alpha_l) \right),$$

where $\lambda \triangleq \lambda(K)$ is a positive number which depends on K, and is introduced in order to give more importance to the reference signal y_0. For any $m = 1, \ldots, N$ we now consider:

$$\frac{1}{M+1} \sum_{l=0}^M S_{y_l}(\omega) - S_{\bar{y}_m(\cdot; \boldsymbol{\alpha}_m)}(\omega) . \tag{5}$$

The function described in (5) represents the difference between the mean of the ESDs and the ESD of the average signal of the m-th block. Since the observed signals are sampled, the integral of $S_y(\omega)$ will be in practice approximated by its Riemann sum, that is

$$\hat{S}_y(k) = \left| \frac{1}{n} \sum_{m=1}^n y(t_m) \mathrm{e}^{-2i\pi mk/n} \right|^2, \quad k \in \mathcal{K} ,$$

where $\mathcal{K} = \{-\frac{n-1}{2}, -\frac{n-3}{2}, \ldots, \frac{n-1}{2}\}$ (note that k in the latter is not necessarily an integer). Let the sequence $C_m(\boldsymbol{\alpha}_m) \triangleq \{C_m(k, \boldsymbol{\alpha}_m) : k \in \mathcal{K}\}$ be defined by

$$C_m(k, \boldsymbol{\alpha}_m) \triangleq \frac{1}{M+1} \sum_{l=0}^{M} \hat{S}_{y_l}(k) - \hat{S}_{\bar{y}_m(\cdot;\boldsymbol{\alpha}_m)}(k) , \tag{6}$$

and let $\{\nu_k, k \in \mathcal{K}\}$ be a sequence of nonnegative numbers such that $\nu_{-k} = \nu_k$ and $\sum_k k^2 \nu_k < \infty$ when n tends to infinity. The proposed M-estimator of $\boldsymbol{\theta}_m$ is denoted by $\hat{\boldsymbol{\theta}}_m$ and is given by

$$\hat{\boldsymbol{\theta}}_m \triangleq \underset{\boldsymbol{\alpha}_m \in [0;2\pi]^K}{\text{Arg min}} \|C_m(\boldsymbol{\alpha}_m)\|_\nu^2, \tag{7}$$

where $\|C_m(\boldsymbol{\alpha}_m)\|_\nu^2 = \sum_{k \in \mathcal{K}} \nu_k |C_m(k, \boldsymbol{\alpha}_m)|^2$.

Remark 2.1: As aforementioned, all the blocks have one curve y_0 in common. We impose this constraint in order to address the problem of identifiability. Without this precaution, replacing $\boldsymbol{\alpha}_m$ by $\boldsymbol{\alpha}_m + (c, c, \ldots, c)$, $c \in \mathbb{R}$ and the signal $s(\cdot)$ by $s(\cdot - c)$ in the m-th block would let (6) invariant. Adding the curve y_0 in each block as a reference allows to estimate the shifts with respect to a same common reference.

The estimator of the probability density function f, denoted by $\hat{f}_{M,h}$, is then computed by plugging the estimated values of the shifts in a known density estimator, such as the regular kernel density estimator [19], that is for all real x in $[0; 2\pi]$:

$$\hat{f}_{M,h}(x) = \frac{1}{(M+1)h} \sum_{m=0}^{M} \psi \left(\frac{x - \hat{\theta}_m}{h} \right) , \tag{8}$$

where the kernel ψ is a nonnegative function integrating to 1 with a bounded derivative and h the classical bandwidth parameter of the kernel. In this paper we provide a proof of weak convergence of the empirical distribution function of $\{\hat{\theta}_j, \ j = 1 \ldots M\}$ under some mild conditions. More specifically, we shall get from Theorem 3.2 that $\hat{f}_{M,h}(x)$ converges pointwise to $f(x)$ when both $M \to \infty$ and $n \to \infty$.

III. THEORETICAL ASPECTS

We provide in this section theoretical results on the estimators described in (7) and (8). We denote by $c_s(k)$ the discrete Fourier transform (DFT) of s:

$$c_s(k) \triangleq \frac{1}{n} \sum_{m=1}^{n} s(t_m) e^{-2i\pi mk/n} , k \in \mathcal{K} ,$$

and by $f_{k,l}$ the DFT of y_l:

$$f_{k,l} \triangleq \frac{1}{n} \sum_{m=1}^{n} y_l(t_m) e^{-2i\pi mk/n} , k \in \mathcal{K}, \ l = 0 \ldots M .$$

Let $\theta_l = \bar{\theta}_l + \epsilon_l$ where $\bar{\theta}_l \in \{t_1, \ldots, t_n\}$ is the sampling point whose value is the closest to the actual shift θ_l, and ϵ_l denotes the discretization error. Observe that since the signal is uniformly sampled

on $[0, 2\pi]$, we may write that $|\epsilon_l| < \pi/n$. Using this notation, relation (2) becomes in the Fourier domain for all $k \in \mathcal{K}$ and $l = 0 \dots M$:

$$
\begin{aligned}
f_{k,l} &= \frac{1}{n} \sum_{m=1}^{n} s(t_m - \theta_l) \mathrm{e}^{-2\mathrm{i}\pi mk/n} \\
&\quad + \frac{\sigma}{\sqrt{n}} \left(V_{k,l} + \mathrm{i} W_{k,l} \right) \\
&= \mathrm{e}^{-\mathrm{i}k\bar{\theta}_l} \frac{1}{n} \sum_{m=1}^{n} s(t_m - \epsilon_l) \mathrm{e}^{-2\mathrm{i}\pi mk/n} \\
&\quad + \frac{\sigma}{\sqrt{n}} \left(V_{k,l} + \mathrm{i} W_{k,l} \right) \\
&= \mathrm{e}^{-\mathrm{i}k\theta_l} c_s(k) + \mathrm{O}(kn^{-1}) + \frac{\sigma}{\sqrt{n}} \left(V_{k,l} + \mathrm{i} W_{k,l} \right),
\end{aligned}
\tag{9}
$$

where the middle equality is obtained by the shift property of the DFT and the last equality stems from Taylor-Lagrange inequality due to (H-2). By the white noise assumption (H-3), the sequences $\{V_{k,l}, \ k \in \mathcal{K}\}$ and $\{W_{k,l}, \ k \in \mathcal{K}\}$ in (9) are independent and identically distributed with same standard multivariate normal distribution $\mathcal{N}_n(0, I_n)$. The $\mathrm{O}(kn^{-1})$ term is a result of the sampling operation and is purely deterministic; since it is assumed that $\sum_k k^2 \nu_k < \infty$, the contribution of this deterministic error to the cost function shall be no more than $\mathrm{O}(n^{-1})$, and will further on be neglected since it is not going to induce shift estimation errors greater than the length of a single bin (i.e. n^{-1}), while it will be shown that the statistical estimation error is $\mathrm{O}_{\mathbb{P}}(n^{-1/2})$.

Note that Assumption (H-3) might be considered too strong, and we indeed state it for simplicity. It can be weakened to include more general random variables $\varepsilon_j(t_1), \dots, \varepsilon_j(t_n)$, as long as the Central Limit Theorem can be applied in (9). The $\sqrt{n}$ term appearing in this equation should be then understood as the normalization constant of the mean error in the k-th tap. In particular the homogeneity of the noise distribution is not needed, and some dependency may be permitted as long as the sequence remains with some mixing property. The process should essentially be such that for any $0 < a < b < T$, $\max_j \mathrm{Var}((b-a) \sum_{a < t_i < b} \varepsilon_j(t_i) \leq c_n \sigma^2(b-a)$, where $c_n \to 0$ and $\lim_{\delta \to 0} \sigma^2(\delta) = 0$. One important situation in which the error terms are not independent and identically distributed is when an adaptive sampling strategy is adopted, such that the acquisition of the observations is concentrated around interesting points of the signal. This discussion is beyond the scope of this paper.

A. Heuristic argument and asymptotic expansion

Before detailing the complete derivation of the estimate properties, we give in this section a heuristic argument which shows the consistency and of θ_1 in a simple case. We assume for simplicity that $M = K \gg n \to \infty$ and that all the shifts $\{\theta_j, \ j = 0 \dots M\}$ are equal to zero, so that α_1 represents only the error made during alignment, and we only have one block to process. We also assume that

the signal s is an odd function, so that $c_s(k) = ic_k$ is a non-zero imaginary number, and there is no reason to align the curves with respect to y_0. We assume, without any loss of generality, that $\sum_{m=0}^{K} \alpha_m = 0$. Define for all integers k and l : $\tilde{V}_{k,l} \triangleq \sigma n^{-1/2} V_{k,l}$ and $\tilde{W}_{k,l} \triangleq \sigma n^{-1/2} W_{k,l}$, and define the random variables $R_{k,l}$ and $\beta_{k,l}$ so that we can write $\tilde{V}_{k,l} + i\tilde{W}_{k,l} = R_{k,l} e^{i\beta_{k,l}}$. Observe now that since

$$C_1(k, \boldsymbol{\alpha}_1) = \frac{1}{K+1} \sum_{l=0}^{K} |f_{k,l}|^2 - \left| \frac{1}{K+1} \sum_{l=0}^{K} e^{i\alpha_l k} f_{k,l} \right|^2 , \tag{10}$$

then $C_1(k, \boldsymbol{\alpha}_1) \geq 0$ due to Cauchy-Schwarz inequality. Since K tends to infinity, the mean energy spectral density, the first term on the right-hand-side (RHS) of (10) is approximately $c_k^2 + O_{\mathbb{P}}(K^{-1/2})$, so that:

$$
\begin{aligned}
C_1(k, \boldsymbol{\alpha}_1) &= c_k^2 + O_{\mathbb{P}}(K^{-1/2}) \\
&\quad - \left| \frac{1}{K+1} \sum_{l=0}^{K} e^{i\alpha_l k} (\tilde{V}_{k,l} + i(c_k + \tilde{W}_{k,l})) \right|^2 \\
&= c_k^2 + O_{\mathbb{P}}(K^{-1/2}) \\
&\quad - \frac{1}{(K+1)^2} \sum_{l=0}^{K} \sum_{m=0}^{K} \Big(c_k^2 \cos((\alpha_l - \alpha_m)k) \\
&\quad - 2c_k R_{k,m} \sin((\alpha_l - \alpha_m)k - \beta_{k,m}) \\
&\quad + R_{k,l} R_{k,m} \cos((\alpha_l - \alpha_m)k + \beta_{k,l} - \beta_{k,m}) \Big) .
\end{aligned}
$$

Expanding the harmonic functions up to $o_{\mathbb{P}}(n^{-1})$, assuming that the $\alpha_m = O_{\mathbb{P}}(n^{-1/2})$ (which shall be later proved in Theorem 3.1), and noting that $R_{k,m} = O_{\mathbb{P}}(n^{-1/2})$, we get for any fixed k that

$$
\begin{aligned}
&- \frac{1}{(K+1)^2} \sum_{l=0}^{K} \sum_{m=0}^{K} R_{k,l} R_{k,m} \cos((\alpha_l - \alpha_m)k + \beta_{k,l} - \beta_{k,m}) \\
&= - \frac{1}{(K+1)^2} \sum_{l=0}^{K} \sum_{m=0}^{K} R_{k,l} R_{k,m} \cos(\beta_{k,l} - \beta_{k,m}) \\
&\quad + \frac{1}{(K+1)^2} \sum_{l=0}^{K} \sum_{m=0}^{K} R_{k,l} R_{k,m} \sin(\beta_{k,l} - \beta_{k,m})(\alpha_l - \alpha_m)k + o_{\mathbb{P}}(n^{-1}) \\
&= - \frac{1}{(K+1)^2} \sum_{l=0}^{K} \sum_{m=0}^{K} R_{k,l} R_{k,m} \cos(\beta_{k,l} - \beta_{k,m}) + o_{\mathbb{P}}(n^{-1}) .
\end{aligned}
$$

Since we assumed that $\sum_{m=0}^{K} \alpha_m = 0$, we obtain:

$$\frac{1}{(K+1)^2} \sum_{l=0}^{K} \sum_{m=0}^{K} 2c_k R_{k,m} \sin((\alpha_l - \alpha_m)k - \beta_{k,m})$$

$$= -\frac{2c_k}{K+1} \sum_{m=0}^{K} R_{k,m} \sin(\beta_{k,m})$$

$$+ \frac{2kc_k}{(K+1)^2} \sum_{m=0}^{K} R_{k,m} \cos(\beta_{k,m}) \sum_{l=0}^{K} (\alpha_l - \alpha_m) + o_{\mathbb{P}}(n^{-1})$$

$$= -\frac{2c_k}{K+1} \sum_{m=0}^{K} R_{k,m} \sin(\beta_{k,m})$$

$$- \frac{2kc_k}{(K+1)^2} \sum_{m=0}^{K} R_{k,m} \cos(\beta_{k,m}) \alpha_m + o_{\mathbb{P}}(n^{-1}) \,.$$

Using the same assumption, we get that

$$-\frac{1}{(K+1)^2} \sum_{l=0}^{K} \sum_{m=0}^{K} c_k^2 \cos((\alpha_l - \alpha_m)k)$$

$$= -c_k^2 + \frac{k^2 c_k^2}{2(K+1)^2} \sum_{l=0}^{K} \sum_{m=0}^{K} (\alpha_l - \alpha_m)^2 + o_{\mathbb{P}}(n^{-1})$$

$$= -c_k^2 + \frac{k^2 c_k^2}{K+1} \sum_{l=0}^{K} \alpha_l^2 + o_{\mathbb{P}}(n^{-1}) \,.$$

Hence, the Taylor expansion of the cost function is equal to

$$C_1(k, \boldsymbol{\alpha}_1)$$

$$= -\frac{1}{(K+1)^2} \sum_{l=0}^{K} \sum_{m=0}^{K} R_{k,l} R_{k,m} \cos(\beta_{k,l} - \beta_{k,m})$$

$$- \frac{2kc_k}{K+1} \sum_{m=0}^{K} R_{k,m} \cos(\beta_{k,m}) \alpha_m - \frac{2c_k}{K+1} \sum_{m=0}^{K} R_{k,m} \sin(\beta_{k,m})$$

$$+ \frac{k^2 c_k^2}{K+1} \sum_{m=0}^{K} \alpha_m^2 + O_{\mathbb{P}}(K^{-1/2}) + o_{\mathbb{P}}(n^{-1}) \,,$$

which is minimized by taking $\hat{\theta}_m = R_{k,m} \cos(\beta_{k,m})/kc_k + o_{\mathbb{P}}(n^{-1/2}) + O_{\mathbb{P}}(K^{-1/2})$. More generally, when the different bands are weighted, we obtain by differentiation

$$\hat{\theta}_m = \frac{\sum_{k \in \mathcal{K}} \nu_k k c_k^3 R_{k,m} \cos(\beta_{k,m})}{\sum_{k \in \mathcal{K}} \nu_k k^2 c_k^4} + o_{\mathbb{P}}(n^{-1/2}) + O_{\mathbb{P}}(K^{-1/2})$$

which establishes the asymptotic expansion (up to the first order) and the asymptotic normality of the estimate when both n and K tend to infinity.

B. Computation of the cost function C_m

The cost function C_m associated with block m can be written as follows:

$$\|C_m(\boldsymbol{\alpha}_m)\|_\nu^2 = \sum_{k\in\mathcal{K}} \nu_k \left(A_M(k) - B_m(k,\boldsymbol{\theta}_m)\right)^2$$

$$+ \sum_{k\in\mathcal{K}} \nu_k \left(B_m(k,\boldsymbol{\theta}_m) - B_m(k,\boldsymbol{\alpha}_m)\right)^2 \tag{11}$$

$$+ 2\sum_{k\in\mathcal{K}} \nu_k \left(B_m(k,\boldsymbol{\theta}_m) - B_m(k,\boldsymbol{\alpha}_m)\right)$$

$$\times \left(A_M(k) - B_m(k,\boldsymbol{\theta}_m)\right),$$

where $A_M(k)$ and $B_m(k,\boldsymbol{\alpha}_m)$ are the first and second terms of the RHS of (6), both taken at point k. We focus on the expansion of the terms associated with $\|C_1(\boldsymbol{\alpha}_1)\|_\nu^2$, since all other cost functions may be expanded in a similar manner up to a change of index. We detail the expansion of $A_M(k)$ and $B_1(k,\boldsymbol{\alpha}_1)$, since $B_1(k,\boldsymbol{\theta}_1)$ can be easily obtained from the latter term.

Recall that $A_M(k) = \frac{1}{M+1}\sum_{l=0}^{M} |f_{k,l}|^2$; we get that

$$A_M(k) = \frac{1}{M+1}\sum_{l=0}^{M} \left| e^{-ik\theta_l}c_s(k) + \frac{\sigma}{\sqrt{n}}(V_{k,l} + iW_{k,l}) \right|^2$$

$$= \frac{1}{M+1}\sum_{l=0}^{M} \left\{ |c_s(k)|^2 + \frac{\sigma^2}{n}(V_{k,l}^2 + W_{k,l}^2) \right.$$

$$\left. + \frac{2\sigma}{\sqrt{n}}V_{k,l}\mathrm{Re}(e^{-ik\theta_l}c_s(k)) + \frac{2\sigma}{\sqrt{n}}W_{k,l}\mathrm{Im}(e^{-ik\theta_l}c_s(k)) \right\}$$

Due to the equalities $\mathrm{Re}(e^{-k\theta_l}c_s(k)) = \cos(k\theta_l)\mathrm{Re}(c_s(k)) + \sin(k\theta_l)\mathrm{Im}(c_s(k))$ and $\mathrm{Im}(e^{-k\theta_l}c_s(k)) = \cos(k\theta_l)\mathrm{Im}(c_s(k)) - \sin(k\theta_l)\mathrm{Re}(c_s(k))$, it follows that

$$A_M(k) = |c_s(k)|^2 + \frac{\sigma^2}{n(M+1)}\sum_{l=0}^{M}(V_{k,l}^2 + W_{k,l}^2) \tag{12}$$

$$+ \frac{2\sigma\mathrm{Re}(c_s(k))}{\sqrt{n}(M+1)}\sum_{l=0}^{M}(V_{k,l}\cos(k\theta_l) - W_{k,l}\sin(k\theta_l))$$

$$+ \frac{2\sigma\mathrm{Im}(c_s(k))}{\sqrt{n}(M+1)}\sum_{l=0}^{M}(V_{k,l}\sin(k\theta_l) + W_{k,l}\cos(k\theta_l))$$

Remark 3.1: By Assumption (H-2) and the law of large numbers the last two terms of (12) converge almost surely to 0 as M tends to infinity. Moreover, the sum of the second term has a χ^2 distribution with $2(M+1)$ degrees of freedom. Thus, the term $A_M(k)$ tends to $|c_s(k)|^2 + 4n^{-1}\sigma^2$ as $M\to\infty$, and therefore to $S_s(k)$ as both M and n tend to infinity.

The first curve of each block is the reference curve, which is considered to be invariant and thus has

a known associated shift, so that $\alpha_0 = \theta_0 = \hat{\theta}_0 = 0$. It stems from (4) and (9) that

$$
B_1(k, \boldsymbol{\alpha}_1) = \left| \frac{1}{\lambda + K} \left[\lambda(c_s(k) + \frac{\sigma}{\sqrt{n}}(V_{k,0} + \mathrm{i}W_{k,0})) \right. \right.
$$

$$
\left. \left. + \sum_{l=1}^{K} \left(\mathrm{e}^{\mathrm{i}k(\alpha_l - \theta_l)} c_s(k) + \frac{\sigma}{\sqrt{n}} \mathrm{e}^{\mathrm{i}k\alpha_l}(V_{k,l} + \mathrm{i}W_{k,l}) \right) \right] \right|^2 ,
$$

thus, if we define λ_m, $m = 0 \ldots K$, such that $\lambda_0 \triangleq \lambda$ and $\lambda_m \triangleq 1$ otherwise:

$$
B_1(k, \boldsymbol{\alpha}_1) = \frac{1}{(K + \lambda)^2}
$$

$$
\times \left(\sum_{l=0}^{K} \lambda_l \left(\mathrm{e}^{\mathrm{i}k(\alpha_l - \theta_l)} c_s(k) + \frac{\sigma}{\sqrt{n}} \mathrm{e}^{\mathrm{i}k\alpha_l}(V_{k,l} + \mathrm{i}W_{k,l}) \right) \right)
$$

$$
\times \left(\sum_{m=0}^{K} \lambda_m \left(\mathrm{e}^{\mathrm{i}k(\theta_m - \alpha_m)} c_s^*(k) + \frac{\sigma}{\sqrt{n}} \mathrm{e}^{-\mathrm{i}k\alpha_m}(V_{k,m} - \mathrm{i}W_{k,m}) \right) \right) ,
$$

and expanding the latter yields

$$
B_1(k, \boldsymbol{\alpha}_1) = \frac{|c_s(k)|^2}{(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \mathrm{e}^{\mathrm{i}k(\alpha_l - \theta_l - \alpha_m + \theta_m)}
$$

$$
+ \frac{\sigma^2}{n(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \{ \mathrm{e}^{\mathrm{i}k(\alpha_l - \alpha_m)} \times \tag{13}
$$

$$
[V_{k,l}V_{k,m} + W_{k,l}W_{k,m} + \mathrm{i}(V_{k,l}W_{k,m} - W_{k,l}V_{k,m})] \}
$$

$$
+ \frac{\sigma c_s(k)}{\sqrt{n}(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \mathrm{e}^{\mathrm{i}k(\alpha_l - \theta_l - \alpha_m)}(V_{k,m} - \mathrm{i}W_{k,m})
$$

$$
+ \frac{\sigma c_s^*(k)}{\sqrt{n}(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \mathrm{e}^{\mathrm{i}k(\theta_m + \alpha_l - \alpha_m)}(V_{k,l} + \mathrm{i}W_{k,l}) .
$$

The functional $\|C_1(\boldsymbol{\alpha}_1)\|_\nu^2$ can be split into a stochastic part which depends on the random variables $\{V_{k,l},\ k = -\frac{n-1}{2} \ldots \frac{n-1}{2}\}$ and $\{W_{k,l},\ k = -\frac{n-1}{2} \ldots \frac{n-1}{2}\}$, and a noise-free part which does not depends on them, and is further on denoted by $D_1(\boldsymbol{\alpha}_1)$. Observe that the first sum in (13) is equal to $|c_s(k)|^2$ when taking $\boldsymbol{\alpha}_1 = \boldsymbol{\theta}_1$; consequently, all the terms stemming from the first and the third sum in (11) depend on the random variables $\{V_{k,l},\ k = -\frac{n-1}{2} \ldots \frac{n-1}{2}\}$ and $\{W_{k,l},\ k = -\frac{n-1}{2} \ldots \frac{n-1}{2}\}$, and are not part of the functional $D_1(\boldsymbol{\alpha}_1)$. This term is equal to:

$$
D_1(\boldsymbol{\alpha}_1) \tag{14}
$$

$$
= \sum_{k \in \mathcal{K}} \nu_k |c_s(k)|^4 \left| \left| \frac{1}{K + \lambda} \sum_{m=0}^{K} \lambda_m \mathrm{e}^{\mathrm{i}k(\alpha_m - \theta_m)} \right|^2 - 1 \right|^2
$$

Details of the calculations are given in Appendix -A. Note that due to (14), D_1 has a unique global minimum which is attained when $\alpha_m = \theta_m$, for all $m = 1 \ldots, K$, that is the actual shift value. We show in Proposition 3.1 that $\|C_1(\boldsymbol{\alpha}_1)\|_\nu^2 - D_1(\boldsymbol{\alpha}_1)$ is negligible when both n and K tend to infinity,

under mild assumptions on λ, so that the proposed cost function behaves asymptotically like $D_1(\boldsymbol{\alpha}_1)$.

Proposition 3.1: Assume that $K \to \infty$, $n \to \infty$, $\lambda \to \infty$, and $\lambda/K \to 0$. Denote the noise-free part associated with $B_1(k, \boldsymbol{\theta}_1) - B_1(k, \boldsymbol{\alpha}_1)$ by $\Delta(k, \boldsymbol{\alpha}_1)$, that is

$$\Delta(k, \boldsymbol{\alpha}_1) \triangleq |c_s(k)|^2 \left| \left| \frac{1}{K + \lambda} \sum_{m=0}^{K} \lambda_m e^{ik(\alpha_m - \theta_m)} \right|^2 - 1 \right|,$$

And denote the noise part by $R(k, \alpha_1) \triangleq B_1(k, \boldsymbol{\theta}_1) - B_1(k, \boldsymbol{\alpha}_1) - \Delta(k, \boldsymbol{\alpha}_1)$. Then:

$$\sum_{k \in \mathcal{K}} \nu_k \big(A_M(k) - B_1(k, \boldsymbol{\theta}_1) \big)^2 = O_{\mathbb{P}} \left(\frac{1}{n^2} \right) + O_{\mathbb{P}} \left(\frac{1}{nK} \right)$$

$$\sum_{k \in \mathcal{K}} \nu_k R(k; \boldsymbol{\alpha}_1)^2 = O_{\mathbb{P}} \left(\frac{1}{n^2} \right)$$

$$+ O_{\mathbb{P}} \Big(\frac{1}{n} \Big) \inf_c \frac{1}{K + \lambda} \sum_{m=0}^{K} \lambda_m (\alpha_m - \theta_m - c)^2 \tag{15}$$

$$\|C_1(\boldsymbol{\alpha}_1)\|_\nu^2 = \sum_{k \in \mathcal{K}} \nu_k \Delta(k, \boldsymbol{\alpha}_1)^2 + O_{\mathbb{P}} \left(\frac{1}{nK} \right) + O_{\mathbb{P}} \left(\frac{1}{n^2} \right)$$

$$+ \left(O_{\mathbb{P}} \Big(\frac{1}{n} \Big) + O_{\mathbb{P}} \Big(\frac{1}{\sqrt{nK}} \Big) \right) \inf_c \frac{1}{K + \lambda} \sum_{m=0}^{K} \lambda_m (\alpha_m - \theta_m - c)^2$$

$$+ O_{\mathbb{P}} \Big(\frac{1}{\sqrt{n}} \Big) \left[\inf_c \frac{1}{K + \lambda} \sum_{m=0}^{K} \lambda_m (\alpha_m - \theta_m - c)^2 \right]^{3/2}.$$

where the $O_{\mathbb{P}}$ hold uniformly in $\boldsymbol{\alpha}_1$.

Proof: See Appendix -B. ∎

Since $\hat{\boldsymbol{\theta}}_1$ is the minimizer of $C_1(\boldsymbol{\alpha}_1)$ and $D_1(\boldsymbol{\theta}_1) = 0$, we get by means of Proposition 3.1 that

$$D_1(\hat{\boldsymbol{\theta}}_1) = \|C_1(\hat{\boldsymbol{\theta}}_1)\|_\nu^2 + \big(D_1(\hat{\boldsymbol{\theta}}_1) - \|C_1(\hat{\boldsymbol{\theta}}_1)\|_\nu^2 \big)$$

$$\leq \|C_1(\boldsymbol{\theta}_1)\|_\nu^2 + \big(D_1(\hat{\boldsymbol{\theta}}_1) - \|C_1(\hat{\boldsymbol{\theta}}_1)\|_\nu^2 \big)$$

$$= D_1(\boldsymbol{\theta}_1) + \big(D_1(\hat{\boldsymbol{\theta}}_1) - \|C_1(\hat{\boldsymbol{\theta}}_1)\|_\nu^2 \big)$$

$$\qquad - \big(D_1(\boldsymbol{\theta}_1) - \|C_1(\boldsymbol{\theta}_1)\|_\nu^2 \big)$$

$$= \big(D_1(\hat{\boldsymbol{\theta}}_1) - \|C_1(\hat{\boldsymbol{\theta}}_1)\|_\nu^2 \big)$$

$$\qquad - \big(D_1(\boldsymbol{\theta}_1) - \|C_1(\boldsymbol{\theta}_1)\|_\nu^2 \big) \tag{16}$$

$$= O_{\mathbb{P}} \left(\frac{1}{n^2} \right) + O_{\mathbb{P}} \left(\frac{1}{nK} \right)$$

$$+ \left(O_{\mathbb{P}} \Big(\frac{1}{n} \Big) + O_{\mathbb{P}} \Big(\frac{1}{\sqrt{nK}} \Big) \right) \inf_c \frac{1}{K + \lambda} \sum_{m=0}^{K} \lambda_m (\hat{\theta}_m - \theta_m - c)^2$$

$$+ O_{\mathbb{P}} \Big(\frac{1}{\sqrt{n}} \Big) \left[\inf_c \frac{1}{K + \lambda} \sum_{m=0}^{K} \lambda_m (\hat{\theta}_m - \theta_m - c)^2 \right]^{3/2},$$

thus showing that $D_1(\hat{\boldsymbol{\theta}}_1)$ is close to zero as both n and K tend to infinity. The main result is using the fact that the only minimizer of D_1 is the true vector of shifts.

C. Theoretical properties of the shift estimation algorithm

The following result gives information on the number of curves well aligned in a given block, and holds for each term in the sum of Equation (14).

Proposition 3.2: Let $\eta \to 0$ as $n, K \to \infty$, $\lambda \geq 1$ and let δ be a real positive number. Suppose that $\alpha_1, \ldots, \alpha_m$ is any sequence such that:

$$\left| \frac{1}{(K+\lambda)} \sum_{m=0}^{K} \lambda_m e^{ik(\theta_m - \alpha_m)} \right| > 1 - \eta \,,$$

for some $k \in \mathcal{K}$. Then there exist two positive constants γ_0 and K_0, such that for $K \geq K_0$, there is a constant c such that the number of curves whose alignment error $\alpha_m - \theta_m - c$ is bigger than η^δ, is bounded by $\gamma_0(K+\lambda)\eta^{1-2\delta}$. Moreover,

$$\sum_{m=1}^{K} (\theta_m - \alpha_m - c)^2 \leq \frac{(K+\lambda)\eta}{\gamma_0 k^2}. \tag{17}$$

Proof: See Appendix -C. ∎

Note that the latter proposition is of interest only when $0 < \delta < 1/2$, since $\delta > 1/2$ would yield a large upper bound. Proposition 3.2 has the following motivation: when the number of curves in each block is large enough, the noise contribution to the criterion will be small, and $\hat{\boldsymbol{\theta}}_1$ will be such that the condition of the proposition holds. Hence, we can conclude that most curves will tend to align. However, they may not align with the reference curve y_0. Consequently, the weighting factor λ is introduced in order to "force" all the curves in a block to align with respect to y_0, as stated in the following proposition:

Proposition 3.3: Assume that λ is an integer, and that $\eta^{1-2\delta} \leq \lambda/(\gamma_0(K+\lambda))$, where γ_0 is the positive constant appearing in the previous proposition. Then, under the assumption of Proposition 3.2, we get that $|c| < \eta^\delta$.

Proof: See Appendix -D. ∎

In other words, when λ is chosen such that $\lambda \to \infty$ and $\lambda/K \to 0$ as $K \to \infty$, the estimate will be close to the actual shifts. We now state the main theorem:

Theorem 3.1: Under Assumptions (H-1)–(H-3), if $K \to \infty$, $n \to \infty$, $\lambda = \lambda(K) \to \infty$, $n^{1/4}\lambda/K \to 0$, and n/K is bounded, then for all $\delta \in (0, 1/2)$, there exists $\gamma > 0$, such that with probability converging to 1

$$\frac{1}{K+\lambda} \sum_{m=0}^{K} \mathbf{1}(|\hat{\theta}_m - \theta_m| > 2n^{-\delta}) \leq \gamma n^{-(1-2\delta)}. \tag{18}$$

Proof: In the following, $\gamma_1, \gamma_2, \ldots$ denote positive constants such that the corresponding inequalities hold. The proof of this theorem is deduced from (16) and Propositions 3.2 and 3.3.

Define

$$A^2 \triangleq \inf_c \frac{1}{K + \lambda} \sum_{m=0}^{K} (\hat{\theta}_m - \theta_m - c)^2.$$

By (14) and (16) we can use Proposition 3.2 with

$$\eta = \frac{\gamma_1}{n} + \frac{\gamma_1}{\sqrt{n}} A + \frac{\gamma_1}{n^{1/4}} A^{3/2}, \tag{19}$$

Since θ_m and $\hat{\theta}_m$ are bounded, we obtain that $\eta = o_{\mathbb{P}}(1)$. Equation (17) in Proposition 3.2 yields

$$A^2 \leq \frac{\gamma_1}{n} + \frac{\gamma_2}{\sqrt{n}} A + \frac{\gamma_3}{n^{1/4}} A^{3/2}$$

Define $B \triangleq \sqrt{n} A$, so that the latter becomes $B^2 - \gamma_1 - \gamma_2 B - \gamma_3 B^{3/2} \leq 0$. A continuity argument yields the boundedness of B, thus:

$$\inf_c \frac{1}{K + \lambda} \sum_{m=0}^{K} (\hat{\theta}_m - \theta_m - c)^2 \leq \frac{\gamma_4}{n}, \tag{20}$$

which shows that $\eta \leq \gamma_5/n$. On the other hand, by (14), (16), and Proposition 3.3 we conclude that with probability converging to 1:

$$\frac{1}{K + \lambda} \sum_{m=0}^{K} \mathbf{1}(|\hat{\theta}_m - \theta_m| > 2\eta^\delta) \leq \gamma_6 \eta^{1-2\delta}, \tag{21}$$

and due to (20), (21) still holds when replacing η by γ_5/n, thus proving (18) and the theorem. In particular, letting δ be as close to $1/2$ as needed shows that the estimator $\hat{\boldsymbol{\theta}}_1$ tends to $\boldsymbol{\theta}_1$ with the standard rate of convergence $n^{-1/2}$. ∎

D. Weak convergence of the density estimator

Due to the previous results, it is now possible to give a theoretical result about the plug-in estimate of the distribution of θ. As suggested in (8), an estimate of the probability density function f can be obtained by plugging the approximated values of the shifts into a known density estimate. We provide here a result on the weak convergence of the empirical estimator.

Theorem 3.2: Let g be a continuous function with a bounded derivative. Under the assumptions of Theorem 3.1, we get almost surely when $M \to \infty, n \to \infty$ that

$$\frac{1}{M + 1} \sum_{m=0}^{M} g(\hat{\theta}_m) \longrightarrow \mathbb{E}[g(\theta)]. \tag{22}$$

Proof of theorem 3.2 can be sketched as follows: due to the Law of Large Numbers, it is equivalent to show that:

$$\frac{1}{M + 1} \sum_{m=0}^{M} (g(\hat{\theta}_m) - g(\theta_m))$$

converges almost surely to 0. Since g has a bounded derivative, we can write that the absolute value of the latter term is bounded by

$$\frac{\sup_x |g'(x)|}{M+1} \sum_{m=0}^{M} |\hat{\theta}_m - \theta_m|.$$

Consequently, due to Theorem 3.1, there exists a constant C such that with probability:

$$\frac{1}{M+1} \sum_{m=0}^{M} (g(\hat{\theta}_m) - g(\theta_m)) \leq C \left(\frac{1}{n^\delta} + \frac{1}{n^{(1-2\delta)}} \right),$$

which completes the proof. More particularly, taking $g(\cdot) = h^{-1} \psi \left(\frac{\cdot - x}{h} \right)$, where $h^2 \min\{n^\delta, n^{1-\delta}\} \to \infty$ we get that (8) tends to $\mathbb{E}\left[h^{-1} \psi \left(\frac{\theta - x}{h} \right) \right]$, thus showing pointwise consistency, that is

$$\hat{f}_{M,h}(x) \longrightarrow f(x) \text{ as } M \to \infty, h \to 0, Mh \to \infty$$

for any continuity point x of f.

Remark 3.2: If n remains bounded as $K \to \infty$, then the parameters θ_m cannot be estimated without an error, and the observed distribution of $\{\hat{\theta}_m\}$ would be a convolution of the distribution of $\{\theta_m\}$ with the estimation error. If n is large enough, the latter distribution is approximately normal with variance which is $O_{\mathbb{P}}(\sigma^2/n)$.

Remark 3.3: The discussion was under the assumption that the $\theta_1, \ldots, \theta_M$ have a continuous distribution with a smooth density. If this would not be the case, then the estimated density will be approximately equal to a smoothed version of the distribution.
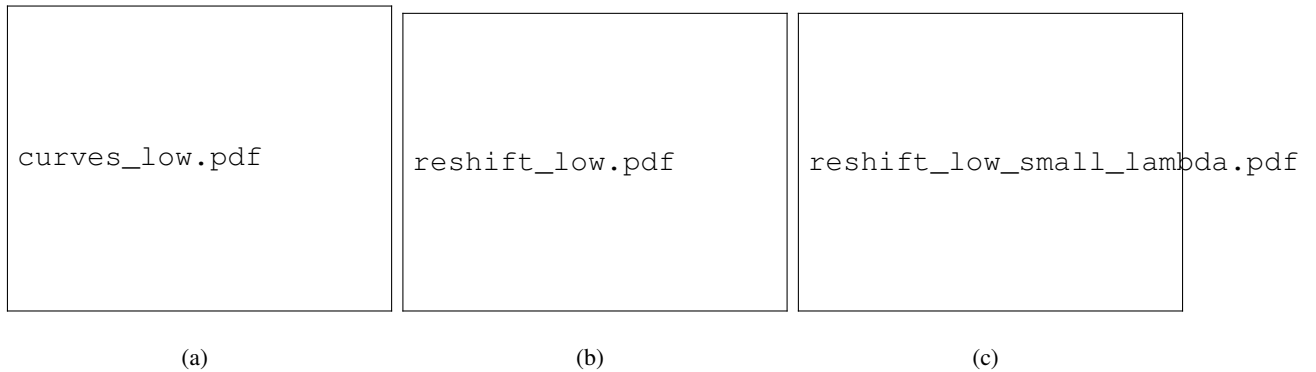
IV. APPLICATIONS

Fig. 2. Results for K=200 and $\sigma^2 = 0.1$; (a) two curves before alignment. (b) comparison between estimated against actual values (blue dots) of the shifts for $\lambda = 50$: good estimates must be close to the identity line (red curve). (c) comparison between estimated and actual values of the shifts for $\lambda = 10$.

We present in this section results based on simulations and real data. Since we provide a generic method suitable for most biological signals, we focus in our simulations on a neuroscience model, while our real datasets stem from the ECG framework. In the latter case, we compare our method

to the one described in [1] which is often used by practitioners, that is a measure of fit based on the squared distance between the average pulse and the shifted pulses leading to a standard Least Square Estimate of the shifts. We present in our simulations results for several values of K. However, a method for choosing automatically the parameter K has been suggested in [20]: since the term $A_M(k)$ can be built iteratively and converges when the number of curves tends to infinity due to Remark 3.1, K can be chosen such that

$$K \triangleq \min \left\{ L \; ; \; \sum_{k \in \mathcal{K}} \nu_k \left(A_M(k) - \frac{1}{L+1} \sum_{l=0}^{L} |f_{k,l}|^2 \right)^2 \leq \epsilon \right\} ,$$

where ϵ is a precision threshold fixed by the user. It is however obvious that the optimal choice of K should depend on the functional properties of the signal s, which are unknown in a semiparametric framework.

A. Simulations results

Using simulations we can study the influence of the parameters K and λ empirically by providing the Mean Integrated Squared Error (MISE) for different values of K and σ^2. We use a fixed number of blocks $N = 20$. The weighting parameter is chosen as $\lambda = [K^\beta]$, where $0 < \beta < 1$. Choosing β close to 1 enables us to align the curves of a given block with respect to the reference curve.

1) Experimental protocol: Simulated data are created according to the discrete model (2), and we compute the estimators for different values of the parameters K, λ and σ^2. For each curve, we sample 512 points equally spaced on the interval $[0; 2\pi]$. We make the experiment with s computed according to the standard Hodgkin-Huxley model for a neural response. The shifts are drawn from a uniform distribution $\mathcal{U}(120\pi/256, 325\pi/256)$, and $\theta_0 = \pi$. The sequence $\{\nu_k, \ k \in \mathcal{K}\}$ is taken such that $\nu_k = 1$ for $k = -\frac{150}{2} \ldots \frac{150}{2}$ and $\nu_k = 0$ otherwise. Though this choice is not optimal, it provides sufficiently good results on the present simulations to illustrate our purpose. Details on the problem of the choice of the tapering sequence $\{\nu_k, \ k \in \mathcal{K}\}$ may be found in [14].

2) Results: We present in Figure 2 results obtained using the alignment procedure, in the case of high noise level ($\sigma^2 = 0.1$). We also compare our estimations with those obtained with an existing method, namely curve alignment according to the comparison between each curve to the mean curve [1]. Results using landmark alignment are displayed in Figure 4. We observe that the efficiency of this approach is less than our estimate achieves with $\lambda = 50$, Figure 2-(b), but is better than the estimate with $\lambda = 10$, Figure 2-(c). An example of density estimation is displayed in Figure 3, using a Gaussian kernel. It should be noticed, however, that h is a free parameter which may exhibit a strong influence on the resulting estimate. A too small value of h leads to over-fitting, whereas taking large values of h leads to hide the multimodality of f, if any. Choosing a data-driven bandwidth selection of h is thus far from trivial, and out of the scope of this paper: we refer to [21],

[22] for an in-depth description of the existing procedures. The bandwidth h is chosen by Silverman's "rule-of-thumb" [23]. We retrieve the uniform distribution of θ. Table I shows the estimated MISE for different values of K and σ^2, with $\lambda = [K^{0.9}]$ and $N = 100$ blocks. The first given number is the value for our estimate, while the second is for the estimator of [1]. Note the dominance of the proposed estimator in all cases, in particular for the more noisy situations.

Fig. 3. Probability density estimation for $N = 20$, $K = 200$ and $\sigma^2 = 0.1$.

B. Results on real data

We now compare the estimated average aligned signal of the two methods applied to ECG signals. The data was obtained from the Hadassah Ein-Karem hospital.

1) Experimental protocol: In order to obtain a series of heart cycles, we first make a preliminary segmentation using the method of [6], namely alignment according to the local maxima of the heart cycle. We then apply our method, and compare it to the alignment obtained by comparing the mean curve to a shifted curve one at a time. We took in this example $K = 30$ and $\lambda = K^{0.75}$.

2) Results: The results are presented in Figure 5. Comparison of Figures 5(c) and 5(d) shows that the proposed method outperforms the standard one. Moreover, when computing the average of the reshifted heart cycle, we observe that our method allows to separate more efficiently the different

Fig. 4.   Shift estimation using Least Square Estimate (see [1]) for one block.

σ^2	K=10	K=20	K=30	K=50	K=100
0	0.0305	0.0228	0.0198	0.0153	0.0106
	0.0306	0.0234	0.0199	0.0156	0.0109
10^{-4}	0.0312	0.0218	0.0183	0.0156	0.0121
	0.0325	0.0232	0.0212	0.0183	0.0158
10^{-2}	0.0296	0.0218	0.0172	0.0143	0.0120
	0.0306	0.0232	0.0192	0.0172	0.0143
1	0.0326	0.0274	0.0248	0.0255	0.0288
	0.0547	0.0806	0.0514	0.0553	0.0741

TABLE I

THE MISE OF THE TWO DENSITY ESTIMATES.

parts of the heart cycle; indeed, the separation between the P-wave, the QRS-complex and the T-wave are much more visible, as it can be seen by comparing the average signals obtained in Figure 5(a) and Figure 5(b).

C. Influence of ECG perturbations on the proposed algorithm

As we saw, the model fits reasonably well the data we have at hand, and in fact perform better than the competing algorithm. The ideal model may not fit other data sets in which the shape of the
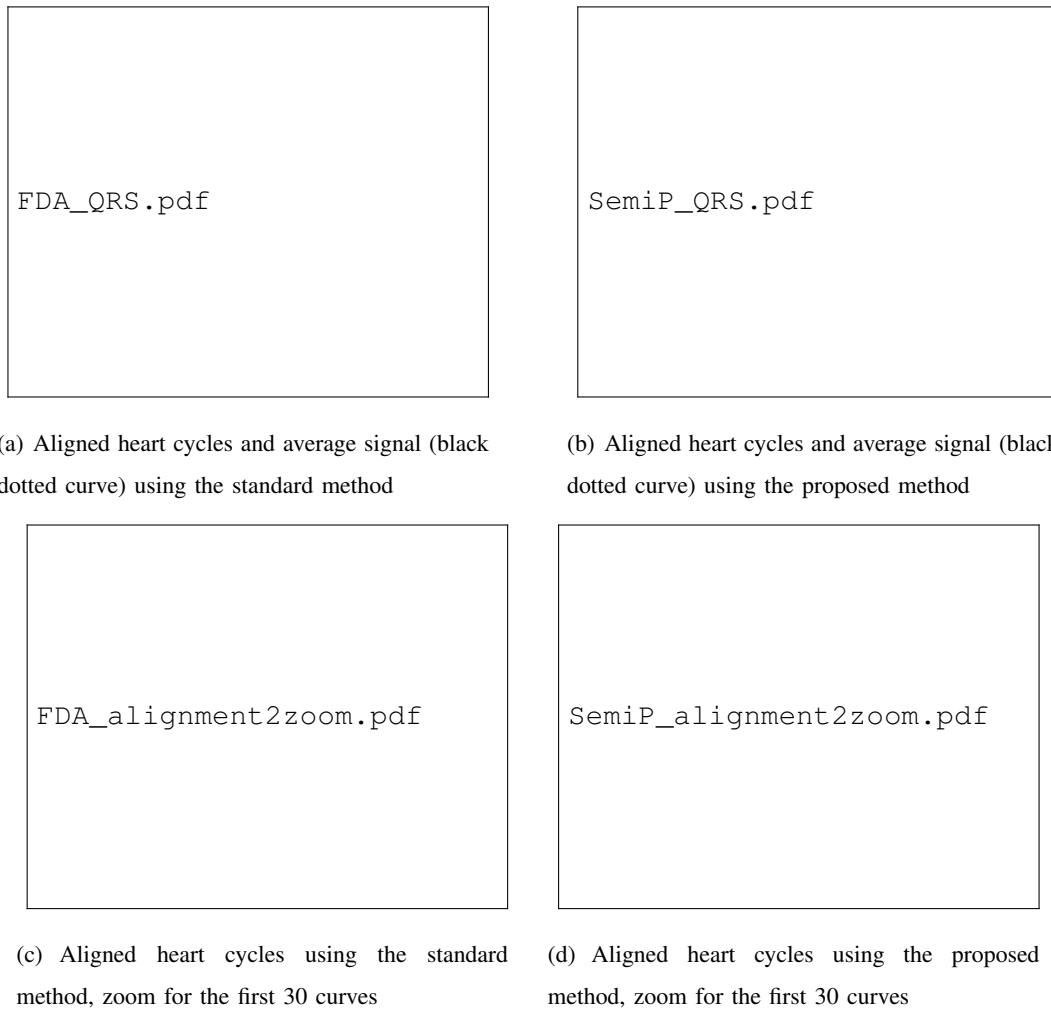
(a) Aligned heart cycles and average signal (black dotted curve) using the standard method

(b) Aligned heart cycles and average signal (black dotted curve) using the proposed method

(c) Aligned heart cycles using the standard method, zoom for the first 30 curves

(d) Aligned heart cycles using the proposed method, zoom for the first 30 curves

Fig. 5.   Comparison between the state-of-the-art and the proposed method for the alignment of heart cycles (arbitrary units). A semiparametric approach appears more appealing to align cycles according to their starting point, and allows to separate more efficiently to P-wave, the QRS complex and the T-wave.

heart pulse changes, or additional perturbations occur. Although no estimation procedure can operate under any possible distortion of the data, we now show that our procedure is quite robust against the main type of potential distortions. The main type of perturbations related to the processing of ECG data are of four kinds (cf. [24]):

- the baseline wandering effect, which can be modeled by the addition of a very low-frequency curve.
- 50 or 60 Hz power-line interference, corresponding to the addition of an amplitude and frequency varying sinusoid.
- Electromyogram (EMG), which is an electric signal caused by the muscle motion during effort test.
- Motion artifact, which comes from the variation of electrode-skin contact impedance produced

by electrode movement during effort test.

To keep the discussion within the scope of the paper, we chose to focus on two perturbations, namely the baseline wander effect and the power-line interference effect. We present in Figure 6 the effect of baseline wander on the proposed algorithm. This effect was simulated by the addition of a low-frequency sine to the ECG measurements. We took here $N = 100, K = 100, \lambda = K^{0.9}$.
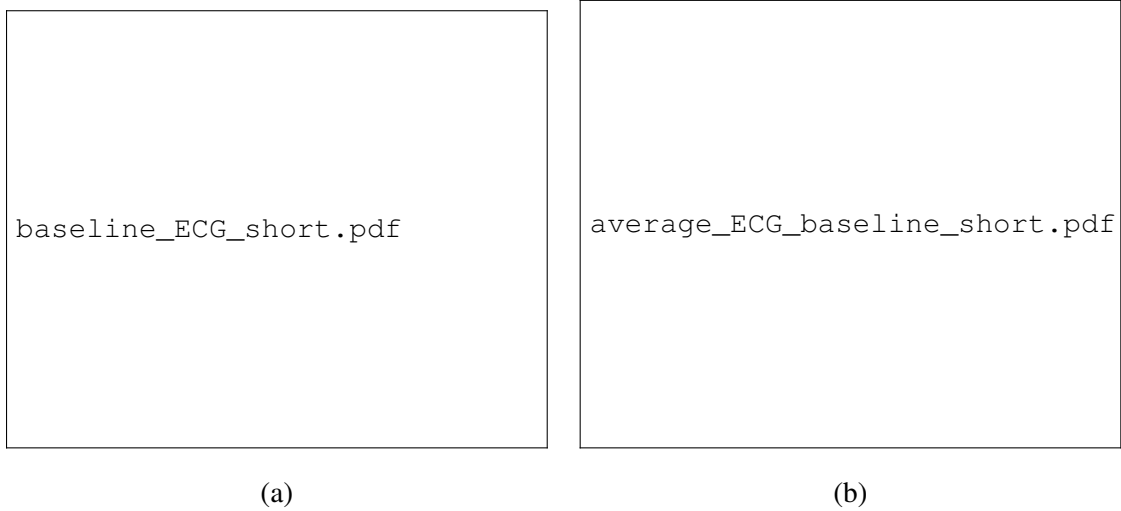

(a)

(b)

Fig. 6.   Effect of the baseline wander phenomenon over the proposed curve alignment method: distorted signal (a), and aligned pulses with the average ECG pulse obtained for one block (b)

We observe that the proposed curve alignment algorithm is robust regarding this kind of perturbations, since we observe well-aligning curves and very little change on the average pulse shape compared to the one obtained without this perturbation. This can be interpreted as follows: since the baseline is in this situation a zero-mean process, the averaging which is done while computing the cost function naturally tends to cancel the baseline. However, we remark that the baseline wander phenomenon can cripple the preliminary segmentation, if the amplitude of the baseline is too high. This problem can be easily circumvented by means of a baseline reduction prefiltering, such as proposed in [24], [25], [26].

We now consider the problem of power-line interference. In order to artificially simulate the original signal with a simulation of the power-line interference, we used the model described in [27], that is we add to the ECG signal the following discrete perturbation:

$$y[n] = (A_0 + \xi_A[n]) \sin\left(\frac{2\pi(f_0 + \xi_f[n])}{f_s} n\right) \ ,$$

where A_0 is the average amplitude of the interference, f_0 its frequency, f_s the sampling frequency of the signal and $\xi_A[n], \xi_f[n]$ are white Gaussian processes used to illustrate possible changes of the amplitude and frequency of the interference. The results of the curve alignment procedure are presented in Figure 7, for a similar choice of N, K and λ.

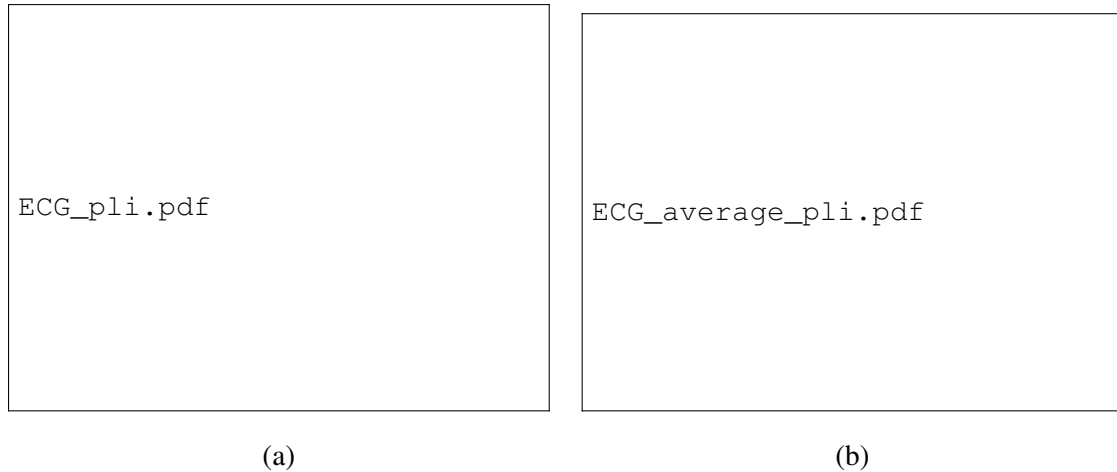(a)   (b)

Fig. 7.   Effect of the power-line interference phenomenon over the proposed curve alignment method: distorted signal (a), and aligned pulses with the average ECG pulse obtained for one block (b)

As shown in Figure 7, the proposed algorithm is robust for this kind of distortion, as we retrieve about the same average signal after alignment of the curves. It shall be noted, once again, that this kind of perturbation can interfere with the segmentation procedure, and that for interferences with high amplitude, a prefiltering step as described in [28], [29], [30] could be applied. Both results illustrate the robustness of semiparametric methods for curve alignment, when compared to standard FDA analysis. We now apply the proposed algorithm to a real ECG signal displayed in Figure 8, which is distorted by power-line interference and baseline wander. After a preliminary segmentation, we get the individual pulses displayed in Figure 9. The aligned curves and the obtained average signal are presented in Figure 10. It can be noted that the proposed method still performs well and is robust to aformentioned perturbations. The average signal obtained by the proposed algorithm is therefore more representative.

D. Discussion

Figures 2(b) and 2(c) are a good illustration of Proposition 3.2. Figure 2(c) shows that when λ is too small, the curves are well aligned within the blocks, but blocks have different constant shifts. Taking a larger λ addresses this problem, as it can be seen in Figure 2(b). Our proposed method uses all the available information and not only the information contained in the neighborhood of the landmarks. The advantage of our method is evident with noisy curves, when locating the maximum of each curve is very difficult.

Not surprisingly, the number of curves in each block K may be low if the noise variance remains very small (first column of Table I), the limiting case $K = 2$ consisting in aligning the curves individually. Theoretically, K should be taken as large as possible. However, this comes with a price, the largest the K the more difficult is the optimization problem.

signal_ecg_real_perturbations.pdf

Fig. 8. ECG signal with real baseline wander and power-line interference (partial)

real_baseline_pwli_before.pdf
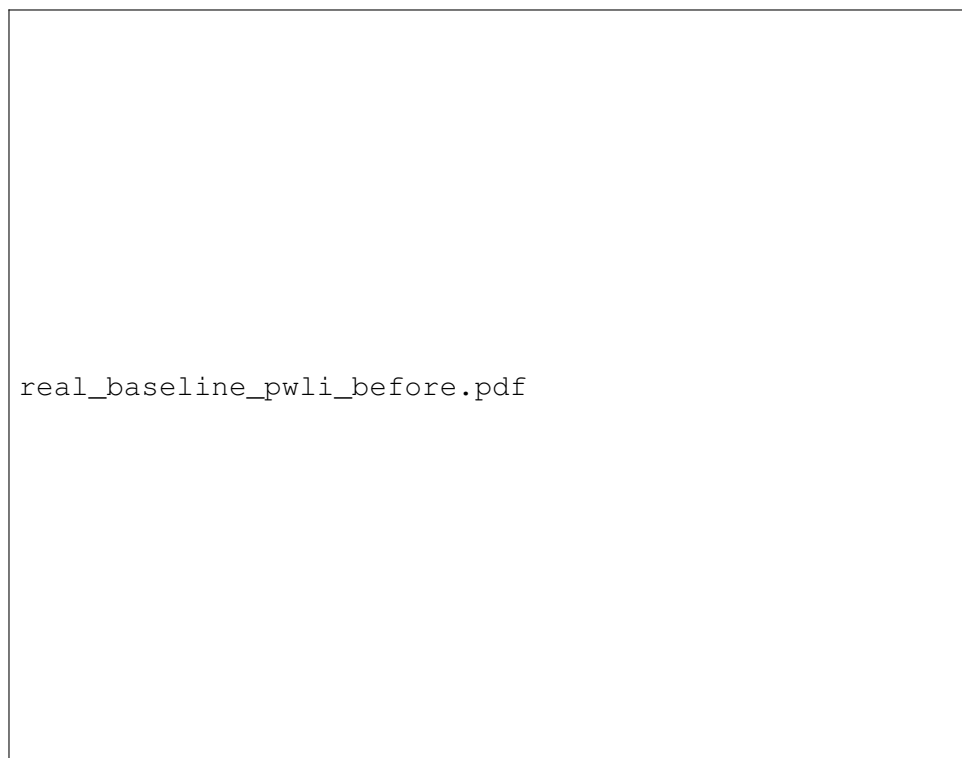
Fig. 9. Obtained curves before the curve alignment procedure and associated average signal (dotted).

real_baseline_pwli_after.pdf

Fig. 10.   Aligned curves by means of the proposed method, and average curve (dotted).

V. CONCLUSION

We proposed in this paper a method for curve alignment and density estimation of the shifts, based on an M-estimation procedure on a functional of the energy spectrum density. The proposed estimator, deduced from blocks of signals of size K, showed good performances in simulations, even when the noise variance is high. On real ECG data, the proposed method outperforms the functional data analysis method, thus leading to a more meaningful average signal, which is of interest for the study of some cardiac arrhythmia. Investigations of the associated kernel estimates, with emphasis on rates of convergence, should appear in a future contribution.

VI. ACKNOWLEDGMENTS

A. Computation of the noise-free part

If the curves are perfectly aligned, that is if $\boldsymbol{\alpha}_1 = \boldsymbol{\theta}_1$, equation (13) becomes

$$
\begin{aligned}
B_1(k, \boldsymbol{\theta}_1) = &\ \frac{|c_s(k)|^2}{(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \\
&+ \frac{\sigma^2}{n(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \{ \mathrm{e}^{\mathrm{i}k(\theta_l - \theta_m)} \times \\
& [V_{k,l} V_{k,m} + W_{k,l} W_{k,m} + \mathrm{i}(V_{k,l} W_{k,m} - W_{k,l} V_{k,m})] \} \\
&+ \frac{\sigma c_s(k)}{\sqrt{n}(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \mathrm{e}^{-\mathrm{i}k\theta_m} (V_{k,m} - \mathrm{i}W_{k,m}) \\
&+ \frac{\sigma c_s^*(k)}{\sqrt{n}(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \mathrm{e}^{\mathrm{i}k\theta_l} (V_{k,l} + \mathrm{i}W_{k,l})
\end{aligned}
\tag{23}
$$

Equation (13) can also be expanded, in order to find an equation close to (12). We find after some calculations that

$$
\begin{aligned}
&B_1(k, \boldsymbol{\theta}_1) \\
&= |c_s(k)|^2 + \frac{\sigma^2}{n(\lambda + K)^2} \sum_{l=0}^{K} \lambda_l^2 (V_{k,l}^2 + W_{k,l}^2) \\
&+ \frac{2\lambda\sigma^2}{n(\lambda + K)^2} \mathrm{Re}\{ \sum_{l=1}^{K} \mathrm{e}^{\mathrm{i}k\theta_l} [V_{k,l} V_{k,0} + W_{k,l} W_{k,0} \\
&\qquad\qquad + \mathrm{i}(V_{k,l} W_{k,0} - W_{k,l} V_{k,0})] \} \\
&+ \frac{2\sigma^2}{n(\lambda + K)^2} \mathrm{Re}\{ \sum_{1 \le l < m \le K} \mathrm{e}^{\mathrm{i}k\theta_l} [V_{k,l} V_{k,m} + W_{k,l} W_{k,m} \\
&\qquad\qquad + \mathrm{i}(V_{k,l} W_{k,m} - W_{k,l} V_{k,m})] \} \\
&+ \frac{2\sigma \mathrm{Re}(c_s(k))}{\sqrt{n}(\lambda + K)} \sum_{l=0}^{K} \lambda_l (V_{k,l} \cos(k\theta_l) - W_{k,l} \sin(k\theta_l)) \\
&- \frac{2\sigma \mathrm{Im}(c_s(k))}{\sqrt{n}(\lambda + K)} \sum_{l=0}^{K} \lambda_l (V_{k,l} \sin(k\theta_l) + W_{k,l} \cos(k\theta_l))
\end{aligned}
\tag{24}
$$

Collecting equations (12), (13) and (24), we can check easily that the only noise-free part comes from the second sum in (11), and is equal to $D_1(\boldsymbol{\alpha}_1)$.

B. Proof of Proposition 3.1

Using Equations (12) and (24), we get that for all k the deterministic part of $A_M(k) - B_1(k, \boldsymbol{\theta}_1)$ vanishes, leading to

$$
A_M(k) - B_1(k, \boldsymbol{\theta}_1) = \frac{\sigma^2}{(M+1)n} \sum_{l=0}^{M} (V_{k,l}^2 + W_{k,l}^2)
$$

$$
- \frac{\sigma^2}{n(\lambda+K)^2} \sum_{l=0}^{K} \lambda_l^2 (V_{k,l}^2 + W_{k,l}^2)
$$

$$
- \frac{2\lambda\sigma^2}{n(\lambda+K)^2} \mathrm{Re}\{\sum_{l=1}^{K} \mathrm{e}^{ik\theta_l}[V_{k,l}V_{k,0} + W_{k,l}W_{k,0}
$$

$$
+ i(V_{k,l}W_{k,0} - W_{k,l}V_{k,0})]\}
$$

$$
- \frac{2\sigma^2}{n(\lambda+K)^2} \mathrm{Re}\{\sum_{1 \leq l < m \leq K} \mathrm{e}^{ik(\theta_l - \theta_m)}[V_{k,l}V_{k,m} + W_{k,l}W_{k,m}
$$

$$
+ i(V_{k,l}W_{k,m} - W_{k,l}V_{k,m})]\}
$$

$$
+ \frac{2\sigma\mathrm{Re}(c_s(k))}{(M+1)\sqrt{n}} \sum_{l=0}^{M} (V_{k,l}\cos(k\theta_l) - W_{k,l}\sin(k\theta_l))
$$

$$
+ \frac{2\sigma\mathrm{Im}(c_s(k))}{(M+1)\sqrt{n}} \sum_{l=0}^{M} (V_{k,l}\sin(k\theta_l) + W_{k,l}\cos(k\theta_l))
$$

$$
- \frac{2\sigma\mathrm{Re}(c_s(k))}{\sqrt{n}(\lambda+K)} \sum_{l=0}^{K} \lambda_l(V_{k,l}\cos(k\theta_l) - W_{k,l}\sin(k\theta_l))
$$

$$
+ \frac{2\sigma\mathrm{Im}(c_s(k))}{\sqrt{n}(\lambda+K)} \sum_{l=0}^{K} \lambda_l(V_{k,l}\sin(k\theta_l) + W_{k,l}\cos(k\theta_l)).
$$

Observe that the first two terms are a linear combination of independent Gamma distributed random variables, with second moments $\frac{\sigma^4(2M+2)(2M+4)}{(M+1)^2 n^2}$ and $\frac{4\sigma^4(\lambda^4 + K + (\lambda^2 + K)^2)}{(K+\lambda)^4 n^2}$, respectively. All the remaining sums in the latter are of i.i.d. random variables with mean zero, and all have sub-Gaussian tails. Consequently, there is a constant D, independent of k, such that when $K \to \infty$, $\lambda \to \infty$ and $\lambda/K \to 0$:

$$
\|A_M(k) - B_1(k, \boldsymbol{\theta}_1)\|_{\mu_2} \leq D \left(\frac{\sigma^2}{n} + \frac{\sigma^2}{nK} + \frac{\sigma|c_s(k)|}{\sqrt{nK}} \right)
$$

where for any random variable X, $\|X\|_{\mu_2} = \sqrt{\mathbb{E}(X^2)}$. Hereafter, D is the same constant, large enough to keep all the inequalities valid. From the latter inequality, we get that:

$$
\sum_{k \in \mathcal{K}} \nu_k \big(A_M(k) - B_1(k, \boldsymbol{\theta}_1) \big)^2 = \mathrm{O}_{\mathbb{P}}\left(\frac{1}{n^2} \right) + \mathrm{O}_{\mathbb{P}}\left(\frac{1}{nK} \right).
$$

We now study the term $R(k, \boldsymbol{\alpha}_1)$, that is the part of $B_1(k, \boldsymbol{\theta}_1) - B_1(k, \boldsymbol{\alpha}_1)$ which depends on the random variables V and W, using their expression in (13) and (23). We get that $R(k, \boldsymbol{\alpha}_1) =$

$I + II + III$, where

$$I \triangleq \frac{\sigma^2}{n(\lambda + K)^2} \left| \sum_{l=0}^{K} \lambda_l \mathrm{e}^{\mathrm{i}k\alpha_l} (V_{k,l} + \mathrm{i}W_{k,l}) \right|^2$$

$$II \triangleq -\frac{\sigma^2}{n(\lambda + K)^2} \left| \sum_{l=0}^{K} \lambda_l \mathrm{e}^{\mathrm{i}k\theta_l} (V_{k,l} + \mathrm{i}W_{k,l}) \right|^2$$

and

$$III \triangleq 2\mathrm{Re}\{ \frac{c_s(k)\sigma}{\sqrt{n}(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \times$$

$$(\mathrm{e}^{\mathrm{i}k(\alpha_l - \theta_l - \alpha_m)} - \mathrm{e}^{-\mathrm{i}k\theta_m})(V_{k,m} - \mathrm{i}W_{k,m})\}$$

According to Cauchy-Schwarz inequality for Hermitian products:

$$I \leq \frac{\sigma^2}{n} \left(\frac{1}{K + \lambda} \sum_{l=0}^{K} \lambda_l \mathrm{e}^{\mathrm{i}k\alpha_l} \mathrm{e}^{-\mathrm{i}k\alpha_l} \right) \left(\frac{1}{K + \lambda} \sum_{l=0}^{K} \lambda_l (V_{k,l}^2 + W_{k,l}^2) \right)$$

Since the first term in parentheses is equal to 1 and the second is bounded in probability by Markov's inequality, we get that $I = O_{\mathbb{P}}(1/n)$, and we get that $II = O_{\mathbb{P}}(1/n)$ using a similar argument. Finally, let us define $U_{k,m} \triangleq \mathrm{e}^{-ik\theta_n}(V_{k,m} - \mathrm{i}W_{k,m})$; we get for any real number c that

$$|III|$$

$$\leq \frac{2|c_s(k)|\sigma}{\sqrt{n}(\lambda + K)^2} \left| \sum_{l=0}^{K} \sum_{m=0}^{K} \lambda_l \lambda_m \left(\mathrm{e}^{\mathrm{i}k(\alpha_l - \theta_l - c - \alpha_m + \theta_m + c)} - 1 \right) U_{k,m} \right|$$

$$+ \frac{2|c_s(k)|\sigma}{\sqrt{n}(\lambda + K)} \left| \sum_{m=0}^{K} \lambda_m U_{k,m} \right| .$$

Since the random variables $\{U_{k,m}, m = 0 \ldots K\}$ are independent and identically distributed with mean zero and finite variances, we get that

$$\sum_{m=0}^{K} \lambda_m U_{k,m} = O_{\mathbb{P}}(K^{1/2}). \tag{25}$$

Consequently, using (25) and Cauchy-Schwarz inequality, we obtain:

$$|III|$$

$$\leq \frac{2|c_s(k)|\sigma}{\sqrt{n}(\lambda + K)^2} \left| \sum_{l=0}^{K} \sum_{m=0}^{K} \lambda_l \lambda_m \mathrm{e}^{\mathrm{i}k(\alpha_l - \theta_l - c - \alpha_m + \theta_m + c)} U_{k,m} \right|$$
$$+ \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{nK}}\right)$$

$$\leq \frac{2|c_s(k)|\sigma}{\sqrt{n}(\lambda + K)} \left| \sum_{m=0}^{K} \lambda_m \mathrm{e}^{-\mathrm{i}k(\alpha_m - \theta_m - c)} U_{k,m} \right| + \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{nK}}\right)$$

$$\leq \frac{|c_s(k)|\sigma}{\sqrt{n}(\lambda + K)} \left| \sum_{m=0}^{K} \lambda_m (\mathrm{e}^{-\mathrm{i}k(\alpha_m - \theta_m - c)} - 1) U_{k,m} \right|$$
$$+ \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{nK}}\right)$$

$$\leq \frac{|c_s(k)|\sigma}{\sqrt{n}} \left(\frac{1}{K + \lambda} \sum_{m=0}^{K} \lambda_m |\mathrm{e}^{-\mathrm{i}k(\alpha_m - \theta_m - c)} - 1|^2 \right)^{1/2}$$
$$\times \left(\frac{1}{K + \lambda} \sum_{m=0}^{K} \lambda_m |U_{k,m}|^2 \right)^{1/2} + \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{nK}}\right)$$

$$= \mathrm{O}_{\mathbb{P}}(1)\frac{|c_s(k)|\sigma}{\sqrt{n}} \left(\frac{1}{K + \lambda} \sum_{m=0}^{K} \lambda_m (\alpha_m - \theta_m - c)^2 \right)^{1/2}$$
$$+ \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{nK}}\right).$$

Recall that $\sum_{\nu \in \mathcal{K}} \nu_k$ is bounded. Equation (15) is obtained if all the bounds above are collected, and Assumption (H-2) is used. Finally, since

$$\left| \|C_1(\boldsymbol{\alpha}_1)\|_\nu^2 - \sum_{k \in \mathcal{K}} \nu_k \Delta(k, \boldsymbol{\alpha}_1)^2 \right|$$

$$= \left| \sum_{k \in \mathcal{K}} \nu_k (A_m(k) - B_1(k, \boldsymbol{\theta}_1) + R(k, \boldsymbol{\alpha}_1) + \Delta(k, \boldsymbol{\alpha}_1))^2 - \sum_{k \in \mathcal{K}} \nu_k \Delta(k, \boldsymbol{\alpha}_1)^2 \right|$$

$$\leq 2 \sum_{k \in \mathcal{K}} \nu_k (A_m(k) - B_1(k, \boldsymbol{\theta}_1))^2 + 2 \sum_{k \in \mathcal{K}} \nu_k R(k, \boldsymbol{\alpha}_1)^2$$

$$+ 2 \left| \sum_{k \in \mathcal{K}} \nu_k (A_m(k) - B_1(k, \boldsymbol{\theta}_1) + R(k, \boldsymbol{\alpha}_1)) \Delta(k, \boldsymbol{\alpha}_1) \right|$$

For any number c, a Taylor expansion up to the second order yields:

$$\Delta(k, \boldsymbol{\alpha}_1) = |c_s(k)|^2 \left| \left| \frac{1}{K+\lambda} \sum_{m=0}^{K} \lambda_m \mathrm{e}^{\mathrm{i}k(\alpha_m - \theta_m - c)} \right|^2 - 1 \right|$$

$$\leq 2|c_s(k)|^2 \left| \frac{1}{K+\lambda} \sum_{m=0}^{K} \lambda_m \mathrm{e}^{\mathrm{i}k(\alpha_m - \theta_m - c)} - 1 \right|$$

$$\leq Dk^2 |c_s(k)|^2 \left| \frac{1}{K+\lambda} \left(\sum_{m=0}^{K} \lambda_m \mathrm{i}(\alpha_m - \theta_m - c) - \frac{1}{2} \lambda_m (\alpha_m - \theta_m - c)^2 \mathrm{e}^{\mathrm{i}k(\alpha_m - \theta_m - c)\xi} \right) \right|$$

$$\leq Dk^2 |c_s(k)|^2 \left[\left| \frac{1}{K+\lambda} \sum_{m=0}^{K} \lambda_m (\alpha_m - \theta_m - c) \right| + \left| \frac{1}{2} \frac{1}{K+\lambda} \sum_{m=0}^{K} \lambda_m (\alpha_m - \theta_m - c)^2 \right| \right]$$

Taking $c = \frac{1}{K+\lambda} \sum_{m=0}^{K} \lambda_m (\alpha_m - \theta_m)$ will cancel the RHS first order term and give the minimum of the RHS second order term; since $\sum_{k \in \mathcal{K}} \nu_k k^2 < \infty$ we get eventually that

$$\sum_{k \in \mathcal{K}} \nu_k \Delta(k, \boldsymbol{\alpha}_1) \leq D \inf_c \frac{1}{K+\lambda} \sum_{m=0}^{K} \lambda_m (\alpha_m - \theta_m - c)^2 \,,$$

thus we get the last term of (15).

C. Proof of Proposition 3.2

Observe that there exists γ_0 in $(0, 1)$ such that, for all x in $[-\pi, \pi]$, we have $\cos x \leq 1 - \gamma_0 x^2$. Since we have

$$\left| \frac{1}{(K+\lambda)} \sum_{0 \leq m \leq K} \lambda_l \exp\left(\mathrm{i}k(\theta_m - \alpha_m)\right) \right| \leq 1 \,,$$

then there exists, according to the assumption, two constants $K_0 \geq 0$ and c such that, for $K \geq K_0$ and every k, we have

$$\mathrm{Re}\left(\frac{\mathrm{e}^{-\mathrm{i}c}}{(K+\lambda)} \sum_{m=0}^{K} \lambda_m \exp\left(\mathrm{i}k(\theta_m - \alpha_m)\right) \right)$$

$$\geq 1 - \eta \,. \quad (26)$$

Hence:

$$1 - \eta \leq \frac{1}{K+\lambda} \sum_{m=0}^{K} \lambda_m \cos\left(k(\theta_m - \alpha_m - c)\right)$$

$$\leq \frac{1}{K+\lambda} \sum_{m=0}^{K} \lambda_m \left(1 - \gamma_0 k^2 (\theta_m - \alpha_m - c)^2\right),$$

and (17) follows. Denote by N the number of curves in the block whose alignment error is "far" from c (up to a 2π factor):

$$N \triangleq \sum_{m=0}^{K} \mathbf{1}\left\{ |\theta_m - \alpha_m - c| \geq \eta^\delta \right\} \,,$$

and assume, for simplicity, that the N last curves are the misaligned curves. Equation (26) implies

$$
\begin{aligned}
1 - \eta &\leq \frac{1}{K+\lambda} \sum_{m=0}^{K-N-1} \lambda_m \cos(k(\theta_m - \alpha_m - c)) \\
&\quad + \frac{1}{K+\lambda} \sum_{m=K-N}^{K} \lambda_m \cos(k(\theta_m - \alpha_m - c)) \\
&\leq \frac{K+\lambda-N}{K+\lambda} + \frac{N}{K}(1 - \gamma_0 k^{2\delta} \eta^{2\delta}) \\
&= 1 - \frac{N}{K+\lambda} \gamma_0 k^{2\delta} \eta^{2\delta} \; .
\end{aligned}
\tag{27}
$$

Equation (27) leads to

$$
N \leq \frac{K+\lambda}{\gamma_0 k^{2\delta}} \eta^{1-2\delta} \; ,
$$

which completes the proof.

D. Proof of Proposition 3.3

Assume that $|c| > \eta^\delta$; since λ is assumed to be an integer, we can see this weighting parameter as the artificial addition of $\lambda - 1$ reference curves. Since $\alpha_0 = \theta_0 \overset{\Delta}{=} 0$, in that case, $|\theta_0 - \alpha_0 - c| > \eta^\delta$, thus giving

$$
\frac{N}{K+\lambda} > \frac{\lambda}{K+\lambda} \geq \gamma \eta^{1-2\delta} \; ,
$$

which would contradict Proposition 3.2. Therefore, we get that $|c| \leq \eta^\delta$.

REFERENCES

[1] B. W. Silveman and J. Ramsay, *Functional Data Analysis*, 2nd ed. Springer Series in Statistics, 2005.

[2] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, 1st ed. Springer Series in Statistics, 2006.

[3] J. O. Ramsay, "Estimating Smooth Monotone Functions," *Journal of the Royal Statistical Society Series B*, vol. 60, no. 2, pp. 365–375, 1998.

[4] J. O. Ramsay and X. Li, "Curve Registration," *Journal of the Royal Statistical Society Series B*, vol. 60, no. 2, pp. 351–363, 1998.

[5] B. Ronn, "Nonparametric Maximum Likelihood Estimation for Shifted Curves," *Journal of the Royal Statistical Society Series B*, vol. 63, no. 2, pp. 243–259, 2001.

[6] T. Gasser and A. Kneip, "Searching for Structure in Curve Sample," *Journal of the Amerian Statistical Association*, vol. 90, no. 432, pp. 1179–1188, 1995.

[7] A. Kneip and T. Gasser, "Statistical Tools to Analyze Data Representing a Sample of Curves," *Annals of Statistics*, vol. 20, no. 3, pp. 1266–1305, 1992.

[8] M. Pawlak, Z. Hasiewicz, and P. Wachel, "On nonparametric identification of wiener systems," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 482–492, 2007.

[9] W. Greblicki and M. Pawlak, *Nonparametric System Identification*. Cambridge University Press, Jun. 2008.

[10] Z. Hasiewicz and G. Mzyk, "Hammerstein system identification by non-parametric instrumental variables," *International Journal of Control*, vol. 82, no. 3, p. 440, 2009.

[11] P. Bickel, C. Klaassen, Y. Ritov, and J. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, 1st ed. Springer, May 1998.

[12] M. Vimond, "Efficient estimation for a subclass of shape invariant models," *The Annals of Statistics*, vol. 38, no. 3, pp. 1885–1912, 2010.

[13] I. Castillo and J. Loubes, "Estimation of the distribution of random shifts deformation," *Mathematical Methods of Statistics*, vol. 18, no. 1, pp. 21–42, Mar. 2009.

[14] F. Gamboa, J.-M. Loubes, and E. Maza, "Semiparametric Estimation of Shifts Between Curves," *Electronic Journal of Statistics*, vol. 1, pp. 616–640, 2007.

[15] M. Lavielle and C. Levy-Leduc, "Semiparametric Estimation of the Frequency of Unknown Periodic Functions and its Application to Laser Vibrometry Signals," *IEEE Transactions in Signal Processing*, vol. 53, no. 7, pp. 2306– 2314, 2005.

[16] A. C. Guyton and J. E. Hall, *Textbook of Medical Physiology*, 9th ed. W. H. Saunders, 1996.

[17] J. Pan and W. Tomkins, "A Real Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230–236, 1985.

[18] Y. Ritov, "Estimating a Signal with Noisy Nuisance Parameters," *Biometrika*, vol. 76, no. 1, pp. 31–37, 1989.

[19] A. DasGupta, *Asymptotic Theory of Statistics and Probability*, 1st ed. Springer Text in Statistics, Mar. 2008.

[20] T. Trigano, U. Isserles, and Y. Ritov, "Semiparametric Shift Estimation for Alignment of ECG Data," in *Proceedings of the EUSIPCO Signal Processing Conference*, 2008.

[21] A. Berlinet and L. Devroye, "A comparison of kernel density estimate," *Publ. Inst. Statist. Univ. Paris*, vol. 38, pp. 3–59, 1994.

[22] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 401–407, 1996.

[23] B. Silverman, *Density Estimation for Statistics and Data Analysis*, 1st ed. Chapman and Hall/CRC, Apr. 1986.

[24] O. Sayadi and M. B. Shamsollahi, "Multiadaptive Bionic Wavelet Transform: Application to ECG Denoising and Baseline Wandering Reduction," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–11, 2007.

[25] M. A. Mneimneh, E. E. Yaz, M. T. Johnson, and R. J. Povinelli, "An Adaptive Kalman Filter for Removing Baseline Wandering in ECG Signals," *Computers in Cardiology*, vol. 33, pp. 253–256, 2006.

[26] B. Mozaffary and M. A. Tinati, "ECG Baseline Wander Elimination using Wavelet Packets," in *Proceedings of World Academy of Science, Engineering and Technology*, vol. 3, 2005.

[27] L. D. Avendano-Valencia, L. E. Avendano, J. M. Ferrero, and G. Castellanos-Dominguez, "Improvement of an Extended Kalman Filter Power Line Interference Suppressor for ECG Signals," *Computers in Cardiology*, vol. 34, pp. 553–556, 2007.

[28] I. Christov, "Dynamic Powerline Interference Subtraction from Biosignals," *Journal of Medical Engineering and Technology*, vol. 24, no. 4, pp. 169–172, 2000.

[29] C. Levkov, G. Mihov, R. Ivanov, I. Daskalov, I. Christov, and I. Dotsinsky, "Removal of Power-Line Interference from the ECG: a Review of the Subtraction Procedure," *BioMedical Engineering OnLine*, vol. 4, no. 50, pp. 1–18, 2005.

[30] A. K. Ziarani and A. Konrad, "A Nonlinear Adaptive Method of Elimination of Power Line Interference in ECG Signals," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 6, pp. 540–547, 2002.