11

Semiparametric curve alignment and shift density estimation: ECG data processing revisited

T. Trigano[1], U. Isserles[3], T. Montagu[2] and Y. Ritov[3]
[1]*Shamoon College of Engineering, Ashdod Campus, Department of Electrical Engineering, 77141, Ashdod, Israel*
[2]*CEA-LIST, Laboratory of Stochastic Processes and Spectra, CEA-Saclay, 91191 Gif-sur-Yvette, France*
[3]*Hebrew University of Jerusalem, Department of Statistics, Mount Scopus, Israel*

Abstract
We address in this contribution a problem stemming from functional data analysis. Assuming that we dispose of a large number of shifted recorded curves with identical shape, the objective is to estimate the time shifts as well as their distribution. Such an objective appears in several biological applications, for example in ECG signal processing. We are interested in the estimation of the distribution of elapsed durations between repetitive pulses, but wish to estimate it with a possibly low signal-to-noise ratio, or without any knowledge of the pulse shape. This problem is solved within a semiparametric framework, that is without any knowledge of the shape. We suggest an M-estimator leading to two different algorithms whose main steps are as follows: we split our dataset in blocks, on which the estimation of the shifts is done by minimizing a cost criterion, based on a functional of the periodogram. The estimated shifts are then plugged into a standard density estimator. Some theoretical insights are presented, which show that under mild assumptions the alignment can be done efficiently. Results are presented on simulations, as well as on real data for the alignment of ECG signals, and these algorithms are compared to the methods used by practitioners in this framework. It is shown in the results that the presented method outperforms the standard ones, thus leading to a more accurate estimation of the average heart pulse and of the distribution of elapsed times between heart pulses, even in the case of low Signal-to- Noise Ratio (SNR).

1. Introduction

1.1 Description of the problem

Due to the improvements of electronic apparatus and registration systems, it is more and more common place to collect sets of curves or other functional observations. Such registration is often followed by a processing operation, since they tend to represent the same repeated phenomenon. In this contribution the problem of curve registration and alignment from a semiparametric point of view is addressed. More specifically, we assume that we dispose of

M registered curves, each of which being described by the model given in Equation (1.1):

$$y_m(t) = s(t - \theta_m) + \sigma\varepsilon_m(t),\ m = 0 \dots M \tag{1.1}$$

where s is a curve of interest, θ_m is an unknown shift parameter distributed according to some probability density function f_θ, σ is a real positive number and ε_m is a standard Gaussian white noise process. We are interested in the estimation of the curve s and of the shifts $\{\theta_m, m = 0 \dots M\}$ or, when the number of curves M is large, in the estimation of the shift distribution f_θ. Such a problem appears frequently in functional data analysis (FDA) applications, and we refer to Silveman & Ramsay (2005) and Ferraty & Vieu (2006) for examples and case studies related to this issue. In the framework described by the latter equation, 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 (see Ramsay (1998), Ramsay & Li (1998), Ronn (2001), Gasser & Kneip (1995),Kneip & Gasser (1992)) focus on this specific model for many different applications in signal processing for biology. For example, in neuroscience, neurons emit randomly electrical pulses which are recorded by an electrode. Biologists, in many applications, are interested in the estimation of the inter-spike interval, that is, either the estimation of the durations of elapsed time between two electrical pulses, or the estimation of its distribution. As stated in Johnson (1996), it is interesting to model the observed electrical signal as the sample path of a renewal process. We can find in recent contributions (see Pouzat et al. (2004) and Delescluse & Pouzat (2006)) the usefulness of the ISI for spike sorting. In those applications, it is often easy to segment roughly the signal such that we retain only one pulse into each segment, however the realignment of the obtained curves are mainly based in either alignment of the main structural information of the curves (e.g. the zeros, as in Gasser & Kneip (1995); see Kneip & Gasser (1992) for a description of available tools to characterize curves structural information), either in the knowledge of the shape of a standard electrical pulse, as in Ramsay (1998) or Ramsay & Li (1998) (in that case, the problem is often called *template matching*, see Lewicki (1998) and references therein). However, both approaches are sensitive to the level of noise, and some recordings are sometimes too noisy to authorize a satisfactory realignment of the curves. We are therefore interested in finding a method of estimation robust enough in relation to the noise level.

1.2 The curve alignment and estimation from a statistical point of view

The problem of the estimation of the shift parameter θ has been investigated in numerous statistical publications, and this according to two different approaches. The main contributions is this topic focus on the case of a finite number of curves, and provide asymptotic results when the number of sample points tend to infinity. For example, Dalalyan et al. (2006) studied the case of two shifted curves, and proposed a penalized Maximum Likelihood Estimator approach, whereas Gamboa et al. (2007) suggested a semiparametric joint estimation procedure in the case of J curves (J being a fixed number). Some functional data analysis techniques have also been described in Ramsay (1998) and Ramsay & Li (1998), and the authors generally assume that the shift can be expressed as a warping function which has to be estimated. The methods described in Gasser & Kneip (1995) and Kneip & Gasser (1992) are based on template matching procedures; for example, the latter suggested to estimate the sets of the local maxima of s, and to align the different curves accordingly. It shall be noticed that template matching approaches give indeed good results when the curve s is regular enough and the noise variance σ is small; however, they fail when the common shape s shows higher variability or in the case of low SNR. In the case of a finite number of curves, Lavielle & Levy-Leduc

(2005) suggested a semiparametric approach for the estimation of the period of a laser signal, thus following the lead of Ritov (1989).
Another way of looking at the same model has also been proposed: instead of fixing the number of observed curves, it is interesting to make this number tend to infinity and to look at the obtained asymptotics. The first paper dealing with the estimation with a large number of curves can be found in Ritov (1989), and has received a larger attention in the last years. For example, Castillo (2006) and Castillo & Loubes (2007) propose to relate to the nonlinear inverse problem methods and derive estimates based on the works of Dalalyan et al. (2006), whereas Ronn (2001) suggested a nonparametric maximum likelihood estimator approach. More recently, Bigot et al. (2008) and Bigot & Gadat (2008) investigated the estimation of the shape s for an identical model, and suggested a wavelet approach which leads to a near-optimal (in the minimax sense) estimator of the shape. However, their assumption is that the shift distribution f_θ is fully known, thus enabling a deconvolution step in order to compute a nonparametric estimator of s.

1.3 Curve alignment for ECG data

In this contribution we focus more specifically on the analysis of ECG signals. In recordings of the heart electrical activity, at each cycle of contraction and release of the heart muscle, we get a characteristic P-wave, which depicts the depolarization of the atria, followed by a QRS-complex stemming from the depolarization of the ventricles and a T-wave corresponding to the repolarization of the heart muscle. We refer to (Guyton & Hall, 1996, Chapter 12) for an in-depth description of the heart cycle. A typical ECG signal is shown in Figure 1.
Different positions of the electrodes, transient conditions of the heart, as well as some malfunctions and several perturbations (baseline wander, power-line interference, additional electromyographic signal)Fotiadis et al. (2006); Sarnmo & Laguna (2006), can alter the shape of the signal. We aim to situations in which 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 our recording, the above model still holds. This preliminary segmentation can be done, for example, by taking segments around the easily identified maxima of the QRS-complex, as it can be found in Gasser & Kneip (1995). It is therefore of interest to estimate the shift parameters θ_m in (1.1). These estimates can be used afterwards for a more accurate estimation of the heart rate distribution. In normal cases, such estimation can be done accurately by using the common FDA method (e.g. by using only the above prior segmentations). However, when the activity of the heart is more irregular, a more precise alignment can be helpful. This happens for example in cases of cardiac arrhythmia, whose identification can be easier if the heart cycles are accurately aligned. Among interpretations deduced from ECG data, some are based on the so called "signal averaged ECG" (SAECG). SAECG is routinely used in clinics for late potential detection, various heart diseases and arrhythmia detections, as mentioned in Cain et al. (1996), and more specifically for ventricular tachycardia and late potential detection (see e.g. Nava et al. (2000), Rodriguez et al. (2000)). Analysis of the SAECG signal is usually performed by using standard wavelet decomposition Englund et al. (1998). As mentioned in the cited contributions, SAECG is simply a signal averaging technique used to improve the signal-to-noise ratio, since clinicians assume that the ECG waveform is invariant and that the background noise is uncorrelated. Moreover a timing reference (i.e. a landmark) is set at the peak R of the QRS complex (since it is easily detectable) allowing averaging without distortion. This method is to relate to Kneip & Gasser (1992), where the authors chose several landmarks instead of one. We argue that this is actually a crucial point since jitter of

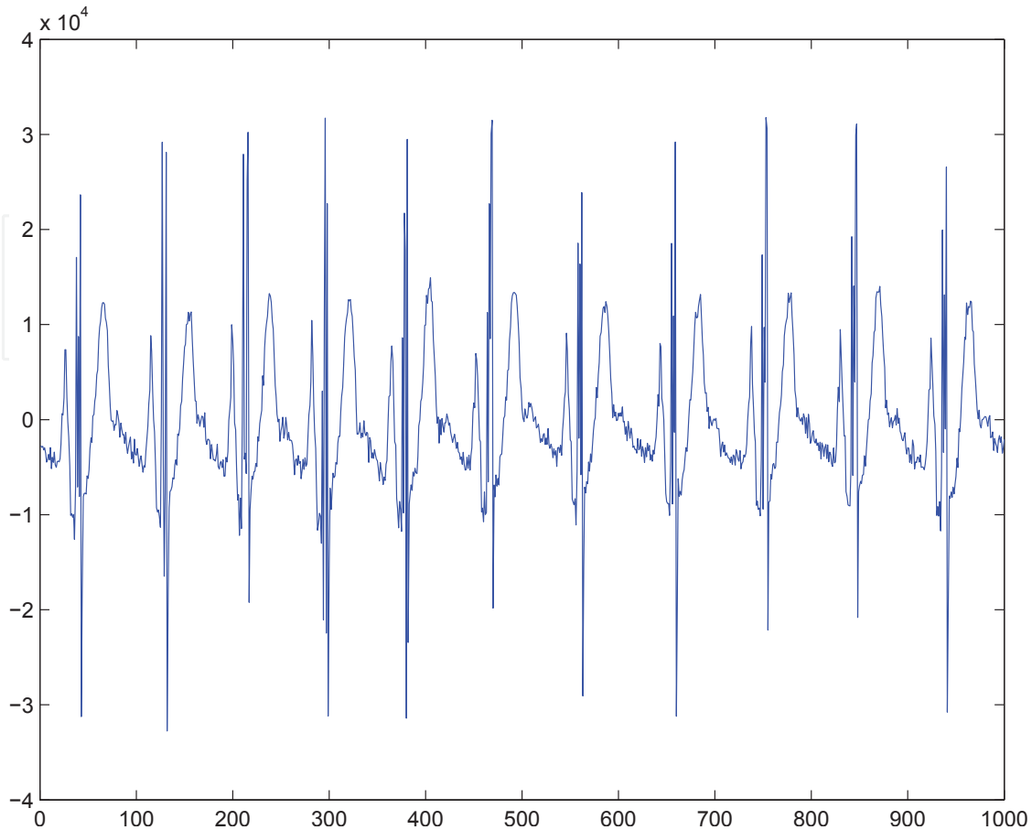


Fig. 1. Typical ECG signal of an healthy subject (arbitrary units).

this temporal reference would distort the resulting SAECG. The proposed method leads to an estimation of the mean cycle by averaging the segments after an alignment according to an estimated θ_j. Additional benefits for a more proper alignment can be found in many other measurements done by cardiologists.

1.4 Chapter organization

This chapter is organized as followed: Section 2 describes the derivation of the shift estimators and the density estimate. Roughly speaking, the estimation of the shifts is based on an M-estimation procedure with a cost function connected to the spectral density of the signal, and we suggest a plug-in estimation for the shift density; This leads to two different algorithms for curve alignment which are described in the end of the section. Some theoretical aspects are described in Section 3. Eventually, Section 4 is dedicated to results on simulations (in order to compare the performances of the two shifts algorithms) and real datasets, and discusses the influence of the standard perturbations which appear in ECG signal processing in the described algorithms.

2. Methodology

We present in this section the main assumptions used in the rest of the chapter and the derivation of the proposed estimator.

2.1 Main assumptions

We assume that after getting an electrical signal such as the one described in Figure 1, a preliminary segmentation is done so that we relate only to the model described in the introduction. We can therefore assume that we have observed sampled noisy curves on a finite time interval $[0, T]$. Each of these curves has being shifted randomly by some random variable θ. A typical curve is expressed in 1.1 and the number of curves is denoted by M. Assuming that the preliminary segmentation has been done efficiently and that each segmented curve has a whole single repetition of the signal, the following assumptions can be done:

(H-1) The distribution of the shift θ (denoted by f_θ) and the shape s both have finite support, respectively $[0, T_\theta]$ and $[0, T_s]$. We also assume that $s \neq 0$ on $[0, T_s]$ and that the derivative s' is bounded.

(H-2) $T_\theta + T_s < T$; that is, we assume that the shape is fully observed in the all the recorded curves.

As mentioned in Ritov (1989), this is equivalent to consider observations on a circle and setting $T = 2\pi$. Therefore, T will be chosen equal to 2π without any loss of generality. We also assume that there is no dependency between the shifts and the additional noise, that is:

(H-3) The random variables $\{\theta_l,\ l = 0 \dots M\}$ are independent and identically distributed, and are in addition independent from $\{\varepsilon_l,\ l = 0 \dots M\}$.

2.2 Description of the shift estimation procedure

In this section, we present a method for the semiparametric curve alignment. This method can be used as a first step for a nonparametric estimation of the shift density, by following the methodology described in Castillo (2006): first provide an estimate for the shifts, using in their example the methodology of Dalalyan et al. (2006), and then plug the obtained values into a standard kernel estimate. We propose an M-estimator to retrieve the shifts, in which the shape information is considered as a nuisance parameter and the shifts are estimated jointly. A similar approach appeared for example in Vimond (2008) leading to another estimator with good asymptotic properties. In another contribution, Gamboa et al. (2007) proposed a semiparametric method for the shifts, with applications to traffic forecasting. This M-estimate, based on a criterion function related to Fourier coefficients, has been shown to be consistent and asymptotically normal as the number of Fourier coefficients increases. However, in this contribution, we chose to focus on the asymptotic properties as the number of curves increases (of course, we also assume that the number of sampling points is large enough, since the M-estimation might lead to inconsistent results otherwise). This approach is explained by the fact that we have in practice little control on the sampling frequency, whereas obtaining a larger dataset of curves is easier.

Following the method of Castillo (2006), we propose to plug M estimates of shifts into a kernel estimate. Consequently, we need to estimate the sequence $\{\theta_l,\ l = 0 \dots M\}$. One important difference, compared to the previously cited works, is that we choose to estimate blocks of parameters jointly instead of one at a time. We therefore split our dataset of curves in N blocks of $K + 1$ curves each, as indicated in Figure 2. Observe that the curve y_0 is included in each block, since we wish to align each curve accordingly to y_0 ; consequently, it shall be assumed in the rest of the chapter that $\theta_0 = 0$. The interest of splitting the dataset of curves into blocks is double: it reduces the variance of the estimators of the shifts by estimating them jointly, and also provides smooth cost functions for the optimization procedure detailed in this section. Indeed, since the recorded signals are based on the same curve s, the average of the

Power Spectral Densities is close to the Power Spectral Density of the average curve provided the shifts are known and have been corrected. This idea will be the cornerstone of the M-estimation procedure proposed in this contribution. We thus estimate jointly the sequence of

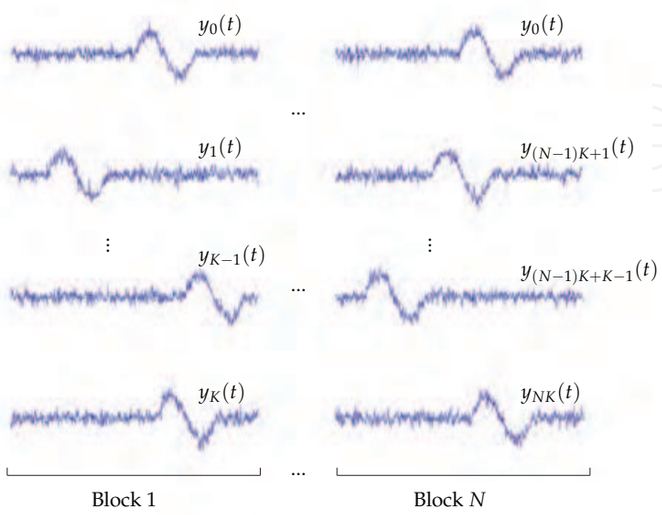


Fig. 2. Split of the curves dataset

vectors $\{\boldsymbol{\theta_m},\ m = 1 \dots N\}$, where for all integer m

$$\boldsymbol{\theta_m} \stackrel{\Delta}{=} (\theta_{(m-1)K+1}, \dots, \theta_{mK}). \tag{2.1}$$

The estimation of $\{\boldsymbol{\theta_m},\ m = 1 \dots N\}$ is achieved by minimizing N cost functions, which are now detailed. Let us denote by S_y the squared modulus of the Fourier Transform of a given continuous curve y, that is for all ω:

$$S_y(\omega) \stackrel{\Delta}{=} \left| \int_0^{2\pi} y(t) \mathrm{e}^{-i\omega t}\, dt \right|^2 .$$

This quantity is of interest, since it remains invariant by shifting. For each integer $m = 1 \dots N$, we define the weighted mean of K curves translated by some correction terms $\mathbf{ff_m} \stackrel{\Delta}{=} (\alpha_{(m-1)K+1}, \dots, \alpha_{mK})$:

$$\bar{y}_m(t; \boldsymbol{\alpha_m}) \stackrel{\Delta}{=} \frac{1}{K + \lambda(K)} \left(\lambda(K) y_0(t) + \sum_{k=(m-1)K+1}^{mK} y_k(t + \alpha_k) \right) . \tag{2.2}$$

where $\lambda(K)$ is a positive number which depends on K, and is introduced in order to give more importance to the reference curve y_0. In the rest of the paper, we shall write λ instead of $\lambda(K)$ in order to avoid cumbersome notations . We now consider the following function:

$$\frac{1}{M+1} \sum_{k=0}^{M} S_{y_k} - S_{\bar{y}_m} . \tag{2.3}$$

The function described in (2.3) represents the difference between the mean of the Power Spectral Densities and the Power Spectral Density of the mean curve. Observe that (2.3) tends to a constant if the curves used in (2.2) are well aligned, that is when $\boldsymbol{\alpha_m}$ is close to $\boldsymbol{\theta_m}$. Since the observed curves are sampled, we will approximate the integral of S_y by its Riemann sum, that is we shall use

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

as an estimator of S_y; that is, the Discrete Fourier Transform (DFT) of a sampled curve $\{y(t_m),\ m = 1 \dots n\}$ will be used in practice instead of the actual Fourier transform of a curve y. The following cost criterion is then introduced and should be minimized in order to reshift all curves into the n-th block:

$$C_m(\boldsymbol{\alpha_m}) \stackrel{\Delta}{=} \frac{1}{M+1} \sum_{k=0}^{M} \hat{S}_{y_k} - \hat{S}_{\bar{y}_m(\cdot;\boldsymbol{\alpha_m})} \,. \tag{2.4}$$

The M-estimator of $\boldsymbol{\theta_m}$, denoted by $\hat{\boldsymbol{\theta}}_{\boldsymbol{m}}$, is therefore given by

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{m}} \stackrel{\Delta}{=} \mathrm{Arg}\min_{\boldsymbol{\alpha}_m \in [0;2\pi]^K} \| C_m(\boldsymbol{\alpha_m}) \|_2^2 \tag{2.5}$$

Remark 1. *It can be noticed that all blocks of $K+1$ curves have one curve y_0 in common. We chose to build the blocks of curves as described in order to address the problem of identifiability. Without this precaution, replacing the solution of (2.5) by $\hat{\boldsymbol{\theta}} + c + 2k\pi$, $k \in \mathbb{Z}$ and s by $s(\cdot - c)$ would let the cost criterion invariant. Adding curve y_0 as a referential allows to estimate $\boldsymbol{\theta} - \theta_0$, thus avoiding the non-identifiability of the model.*

In order to define the criterion function, we chose to split the set of observed curves in N blocks of $K+1$ curves. Indeed, this is not useful if the spectral information is fully known. However, since we observe noisy curves, and since we did not assume any knowledge on the spectral information, the functions S_y have to be estimated. A well known nonparametric estimator is the periodogram, which has been extensively studied (see e.g. Chonavel (2000) and references therein). This estimator is known to be asymptotically unbiased, but its variance does not tend to 0 in the general case; moreover, the pointwise estimation leads to uncorrelated estimated values, therefore the periodogram provides an estimate of the power spectral density of a process with many irregularities, regardless of the regularity of the true power spectrum. A good way to reduce the variance of this estimator is given by the averaged periodogram (or Bartlett's method), based on the mean of several periodogram estimators, thus the necessity of splitting the dataset. We refer to Chonavel (2000) for a more detailed description of this method.

It is interesting to compare the cost function introduced in (2.5) to the estimator introduced in Gamboa et al. (2007). In their contribution, they introduce additional weights to smooth the constrast function, for a fixed number of curves J. More precisely, they propose to estimate $\{\hat{\theta}_j,\ j = 1 \dots J\}$ the minimum of the following criterion function:

$$M_n(\alpha_1, \dots, \alpha_J) = \frac{1}{J} \sum_{j=1}^{J} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \left| \mathrm{e}^{\mathrm{i}\alpha_j l} d_{jl} - \frac{1}{J} \sum_{m=1}^{J} \mathrm{e}^{\mathrm{i}\alpha_m l} d_{ml} \right|^2 , \tag{2.6}$$

where $\{\delta_l,\ l \in \mathbb{Z}\}$ is real sequence such that $\sum_l \delta_l^2 < \infty$ and $\sum_l \delta_l^4 < \infty$, and d_{jl} is the l-th discrete Fourier coefficient associated to the j-th curve. This is to relate to Welch's method to

reduce the variance of the periodogram, and makes sense since the number of curves in Gamboa et al. (2007) is assumed to be fixed. The asymptotics is then provided as the number of samples per curve tends to infinity. Similarly, the method proposed in Castillo (2006) leads to similar smoothing, since the shifting step is done by re-aligning one curve at a time, without using the fact that in the case of shift density estimation, the number of curves is assumed to be big. We argue that the method of averaged periodogram may be preferred for simplicity to weighted periodogram for variance reduction, since we only have one parameter to tune.

2.3 Description of the algorithmic procedures

2.3.1 Curve alignment with respect to a reference curve

Following the derivation of the M-estimator, the first algorithm for curve alignment is pretty straightforward: the dataset is split in blocks which include the reference curve and an optimization procedure (e.g. a conjugate gradient descent) is then performed on each block. The reference curve is by default the first one. The standard plug-in estimate is then used to estimate the density. The procedure is summarized in Algorithm 1.

Algorithm 1 Alignment procedure w.r.t. a reference curve **(A1)**

INPUTS: input curves $y_0, \dots, y_M$ where y_0 is the reference curve and parameters K, λ
OUTPUTS: Shift estimators - $\{\hat{\theta}_j, j = 1 \dots M\}$

Compute the average of the curves periodograms $\bar{S} = \dfrac{1}{M+1} \displaystyle\sum_{j=0}^{M} \hat{S}_{y_j}$.

Split the curve dataset into N blocks of $K+1$ curves, each of them including y_0.
for $m = 1 \dots N$ **do**

$$\bar{y}_m(t; \boldsymbol{\alpha_m}) \stackrel{\Delta}{=} \frac{1}{K+\lambda} \left(\lambda y_0(t) + \sum_{j=(m-1)K+1}^{mK} y_j(t + \alpha_k) \right)$$

Define $C_m(\boldsymbol{\alpha_m}) \stackrel{\Delta}{=} \bar{S} - \hat{S}_{\bar{y}_m(\cdot;\boldsymbol{\alpha_m})}$

Compute $\hat{\boldsymbol{\theta}}_{\boldsymbol{m}} = \operatorname{Arg\,min}_{\boldsymbol{\alpha} \in [0;2\pi]^K} \|C_m(\boldsymbol{\alpha_m})\|_2^2$, where $\hat{\boldsymbol{\theta}}_{\boldsymbol{m}} \stackrel{\Delta}{=} \{\hat{\theta}_{(m-1)K+1}, \dots, \hat{\theta}_{mK}\}$
end for
Return $\{\hat{\theta}_j, j = 1 \dots M\}$

Algorithm **(A1)** is somehow problematic, since it involves the choice of a reference curve y_0 and giving it more importance.This may lead to a bad estimation of the shifts between curves, in the case of very noisy curves (when σ is for example of the same order of magnitude than the signal), since the noise is also magnified in the process. In order to address this issue, it is possible to use another algorithm described in 2.3.2.

2.3.2 A two-stage algorithm for curve alignment and shift density estimation

The two stage algorithm (denoted by **(A2)** in the rest of the chapter) can be described as follows: the dataset is split into N blocks of K curves, and the reference curve y_0 is not included in the blocks. The algorithm primarily performs alignment of the curves within each block. An average of the aligned curves of each block is calculated after this first step. The set of average curves is eventually aligned by similar means. The final shift estimator of each curve is the sum of the shift estimator of the curve among the curves of its block and the shift of the

averaged curve of that block; thus with the two stage algorithm the reference curve is unnecessary. This is the main advantage of a two-stage algorithm, since choosing a reference curve with possibly low SNR leads to significant alignment errors.
The two-stage procedure **(A2)** is summarized in Algorithm 2. In the first stage, estimation of the vector of shifts as in (2.1) is done for each block separately. Since we discard the reference curve y_0, the average of curves in blocks $m \in \{1, \dots, N\}$ translated by some correction terms $\mathbf{ff_m} \stackrel{\Delta}{=} (\alpha_{(m-1)K+1}, \dots, \alpha_{mK})$, is equal to

$$\bar{y}_m(t; \boldsymbol{\alpha_m}) \stackrel{\Delta}{=} \frac{1}{K} \left(\sum_{k=(m-1)K+1}^{mK} y_k(t + \alpha_k) \right) .$$

A cost function is then defined as it was for algorithm **(A1)**:

$$C_m(\boldsymbol{\alpha_m}) \stackrel{\Delta}{=} \frac{1}{M+1} \sum_{k=0}^{M} \hat{S}_{y_k} - \hat{S}_{\bar{y}_m(\cdot;\boldsymbol{\alpha_m})} , \tag{2.7}$$

and the first stage ends by the computation of the M-estimator which minimizes the cost function defined in (2.7). At this stage we obtain the shift estimates $\{\tilde{\boldsymbol{\theta}}_{\boldsymbol{m}}, m = 1 \dots N\}$. At the second stage an average of the aligned curves within each block is calculated, and we align the averaged curves. That is, the average curve is

$$\bar{y}_m(t) \stackrel{\Delta}{=} \frac{1}{K} \left(\sum_{k=(m-1)K+1}^{mK} y_k(t + \tilde{\theta}_k) \right) ,$$

and for a given correction term $\beta \stackrel{\Delta}{=} (\beta_1, \dots, \beta_N)$ the mean of translated average curves is

$$\bar{y}(t; \beta) \stackrel{\Delta}{=} \frac{1}{N+1} \left(\sum_{k=1}^{N} \bar{y}_k(t + \beta_k) \right) .$$

The second cost function for alignment between averaged curves is therefore

$$C(\beta) \stackrel{\Delta}{=} \frac{1}{M+1} \sum_{k=0}^{M} \hat{S}_{y_k} - \hat{S}_{\bar{y}(\cdot;\beta)}$$

An M-estimator is now calculated for the shifts among blocks

$$\vartheta \stackrel{\Delta}{=} \mathrm{Arg\,min}_{\boldsymbol{\beta} \in [0;2\pi]^N} \| C(\beta) \|_2^2 .$$

Finally the estimator for the shift of y_l is the sum of the estimator obtained at the first stage of **(A2)** for this curve and the estimator obtained at the second stage for the averaged curve of the block y_l belongs to, that is:

$$(\hat{\boldsymbol{\theta}}_{\boldsymbol{m}})_k = \vartheta_m + (\tilde{\boldsymbol{\theta}}_{\boldsymbol{m}})_k , m = 1 \dots N, k = 1 \dots K.$$

Algorithm 2 Two-stage alignment procedure**(A2)**

INPUTS: input curves $y_1, \ldots, y_M$ and size of each block K
OUTPUTS: Shift estimators - $\{\hat{\theta}_j, j = 1 \ldots M\}$

Split the curve dataset into N blocks of K curves.

Compute the average of the curves periodograms $\bar{S} = \frac{1}{M} \sum_{j=1}^{M} \hat{S}_{y_j}$

for $m = 1 \ldots N$ **do**

Define $\bar{y}_m(t; \boldsymbol{\alpha_m}) \stackrel{\Delta}{=} \frac{1}{K} \left(\sum_{j=(m-1)K+1}^{mK} y_j(t + \alpha_j) \right)$

Define $C_m(\boldsymbol{\alpha_m}) \stackrel{\Delta}{=} \bar{S} - \hat{S}_{\bar{y}_m(\cdot;\boldsymbol{\alpha_m})}$

Compute $\tilde{\boldsymbol{\theta}}_{\boldsymbol{m}} = \operatorname{Arg\,min}_{\boldsymbol{\alpha} \in [0;2\pi]^K} \|C_m(\boldsymbol{\alpha_m})\|_2^2$, where $\tilde{\boldsymbol{\theta}}_{\boldsymbol{m}} = \{\tilde{\theta}_{(m-1)K+1}, \ldots, \tilde{\theta}_{mK}\}$

Compute $\bar{y}_m(t) \stackrel{\Delta}{=} \frac{1}{K} \left(\sum_{k=(m-1)K+1}^{mK} y_k(t + \tilde{\theta}_k) \right)$

end for

Define $\bar{y}(t; \beta) \stackrel{\Delta}{=} \frac{1}{N+1} \left(\sum_{j=1}^{N} \bar{y}_j(t + \beta_k) \right)$

Define $C(\beta) \stackrel{\Delta}{=} \bar{S} - \hat{S}_{\bar{y}(\cdot;\beta)}$

Compute $\hat{\vartheta} = \operatorname{Arg\,min}_{\boldsymbol{\alpha} \in [0;2\pi]^K} \|C(\beta)\|_2^2$

for $m = 1 \ldots N$ **do**

for $i = 1 \ldots K$ **do**

$(\hat{\boldsymbol{\theta}}_{\boldsymbol{m}})_i = \hat{\vartheta}_m + (\tilde{\boldsymbol{\theta}}_{\boldsymbol{m}})_i$

end for

end for

Return $\hat{\boldsymbol{\theta}}_{\boldsymbol{1}}, \ldots, \hat{\boldsymbol{\theta}}_{\boldsymbol{N}}$.

3. Theoretical properties

We provide in this section some theoretical results about the cost functions obtained in the latter section. In order to prove the efficiency of the method, it is indeed necessary to check that all cost functions have global minima at the actual shifts. We also have to verify that, provided the number of curves is large enough, the cost functions are smooth enough so that the optimization can be done efficiently. The presented results hold for the algorithm **(A1)**, though they can be easily extended to each stage of **(A2)**.

3.1 Expansion of a cost function

Recall that the total number of curves is $M = NK + 1$, where N is the number of blocks and K is the number of curves in each block. The first curve y_0 is a common reference curve for all blocks. We denote by $c_s(k)$ the discrete Fourier transform (DFT) of s taken at point k,

$$c_s(k) \stackrel{\Delta}{=} \frac{1}{n} \sum_{m=1}^{n} s(t_m) \mathrm{e}^{-\frac{2\mathrm{i}\pi mk}{n}},$$

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

$$f_{k,l} \triangleq \frac{1}{n}\sum_{m=1}^{n} y_l(t_m)\mathrm{e}^{-\frac{2\mathrm{i}\pi mk}{n}}\,.$$

Using this notation, relation (1.1) becomes in the Fourier domain for all $k = -\frac{n-1}{2}\ldots\frac{n-1}{2}$ and $l = 0\ldots M$:

$$\begin{aligned} f_{k,l} &= \mathrm{e}^{-\mathrm{i}k\theta_l}\frac{1}{n}\sum_{m=1}^{n} s(t_m+\epsilon_l)\mathrm{e}^{-\frac{2\mathrm{i}\pi mk}{n}} + \frac{\sigma}{\sqrt{n}}\left(V_{k,l}+\mathrm{i}W_{k,l}\right) \\ &= \mathrm{e}^{-\mathrm{i}k\theta_l}c_s(k) + O(\|s'\|_\infty n^{-1}) + \frac{\sigma}{\sqrt{n}}\left(V_{k,l}+\mathrm{i}W_{k,l}\right), \end{aligned} \tag{3.1}$$

where in the latter equation ϵ_l a constant such that $|\epsilon_l| < \pi n^{-1}$, and s' is the first derivative of s which we assumed to be bounded. This error term which results from the sampling operation is purely deterministic, and is further on neglected since it is not expected to induce shift estimation errors greater than the length of a single bin (i.e. n^{-1}). The sequences $\left\{V_{k,l},\ k = -\frac{n-1}{2}\ldots\frac{n-1}{2}\right\}$ and $\left\{W_{k,l},\ k = -\frac{n-1}{2}\ldots\frac{n-1}{2}\right\}$ are independent and identically distributed with same standard multivariate normal distribution $\mathcal{N}_n(0, I_n)$. We now compute the cost function C_m associated with block m:

$$\begin{aligned} \|C_m(\boldsymbol{\alpha_m})\|_2^2 &= \sum_{k=0}^{n-1}\left(A_M(k) - B_m(k,\boldsymbol{\theta_m})\right)^2 \\ &+ \sum_{k=0}^{n-1}\left(B_m(k,\boldsymbol{\theta_m}) - B_m(k,\boldsymbol{\alpha_m})\right)^2 \\ &+ 2\sum_{k=0}^{n-1}\left(B_m(k,\boldsymbol{\theta_m}) - B_m(k,\boldsymbol{\alpha_m})\right)\left(A_M(k) - B_m(k,\boldsymbol{\theta_m})\right), \end{aligned} \tag{3.2}$$

where $A_M(k)$ is the first term of the right hand side of (2.4) and $B_m(k,\boldsymbol{\alpha_m})$ is the second term of the right hand side of (2.4), both taken at point k. Each term of the latter equation is expanded separately. We get that

$$\begin{aligned} A_M(k) &= |c_s(k)|^2 + \frac{\sigma^2}{(M+1)n}\sum_{l=0}^{M}\left(V_{k,l}^2 + W_{k,l}^2\right) \\ &+ \frac{2\sigma\mathrm{Re}(c_s(k))}{(M+1)\sqrt{n}}\sum_{l=0}^{M}(V_{k,l}\cos(k\theta_l) - W_{k,l}\sin(k\theta_l)) \\ &- \frac{2\sigma\mathrm{Im}(c_s(k))}{(M+1)\sqrt{n}}\sum_{l=0}^{M}(V_{k,l}\sin(k\theta_l) + W_{k,l}\cos(k\theta_l))\,. \end{aligned} \tag{3.3}$$

The last two terms of (3.3) converge almost surely to 0 as M tends to infinity, according to Assumption (H-3) and the law of large numbers. Moreover, the sum of the second term is distributed according to a χ^2 distribution with $M+1$ degrees of freedom. Thus, the term $A_M(k)$ tends to $|c_s(k)|^2 + 2n^{-1}\sigma^2$ as $M \to \infty$. We now focus on the expansion of the terms associated to $\|C_1(\boldsymbol{\alpha_1})\|_2^2$, since all other cost functions may be expanded in a similar manner up to a

change of index. The first curve of each block is the reference curve, which is considered to be invariant and thus has a known associated shift $\alpha_0 = \theta_0 = 0$. We obtain

$$B_1(k, \boldsymbol{\alpha_1}) = \left| \frac{1}{\lambda + K} \left[\lambda (c_s(k) + \frac{\sigma}{\sqrt{n}} (V_{k,0} + \mathrm{i} W_{k,0})) \right. \right. \\ \left. \left. + \sum_{l=1}^{K} \left(\mathrm{e}^{\mathrm{i}k(\alpha_l - \theta_l)} c_s(k) + \frac{\sigma}{\sqrt{n}} \mathrm{e}^{\mathrm{i}k\alpha_l} (V_{k,l} + \mathrm{i} W_{k,l}) \right) \right] \right|^2 ,$$

thus, if we define the sequence $\{\lambda_m, m = 0 \dots K\}$ such that $\lambda_0 \triangleq \lambda$ and $\lambda_m \triangleq 1$ otherwise:

$$\begin{aligned} & B_1(k, \boldsymbol{\alpha_1}) \\ & = \frac{|c_s(k)|^2}{(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \mathrm{e}^{\mathrm{i}k(\alpha_l - \theta_l - \alpha_m + \theta_m)} \\ & + \frac{\sigma^2}{n(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \{ \mathrm{e}^{\mathrm{i}k(\alpha_l - \alpha_m)} \times \\ & \quad [V_{k,l} V_{k,m} + W_{k,l} W_{k,m} + \mathrm{i}(V_{k,l} W_{k,m} - W_{k,l} V_{k,m})] \} \\ & + \frac{\sigma c_s(k)}{\sqrt{n}(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \mathrm{e}^{\mathrm{i}(\alpha_l - \theta_l - \alpha_m)} (V_{k,m} - \mathrm{i} W_{k,m}) \\ & + \frac{\sigma c_s^*(k)}{\sqrt{n}(\lambda + K)^2} \sum_{l,m=0}^{K} \lambda_l \lambda_m \mathrm{e}^{\mathrm{i}k(\theta_m + \alpha_l - \alpha_m)} (V_{k,l} + \mathrm{i} W_{k,l}) \end{aligned} \tag{3.4}$$

We now can study the behavior of the cost function, as the number of curves tend to infinity. The functional $\|C_1(\boldsymbol{\alpha_1})\|_2^2$ can be split into a noise-free part, that is a term without random variables V or W, and a random noisy part.

3.2 Decomposition of the cost function into a noise-free part and a noisy part

Recall that the noise-free part of $\|C_1(\boldsymbol{\alpha_1})\|_2^2$ neither depends on $\left\{ V_{k,l},\ k = -\frac{n-1}{2} \dots \frac{n-1}{2} \right\}$ nor $\left\{ W_{k,l},\ k = -\frac{n-1}{2} \dots \frac{n-1}{2} \right\}$), and is denoted further by $D_1(\boldsymbol{\alpha_1})$. This term is equal to:

$$D_1(\boldsymbol{\alpha_1}) = \sum_{k=0}^{n-1} |c_s(k)|^4 \left| \left| \frac{1}{K + \lambda} \sum_{m=0}^{K} \lambda_m \mathrm{e}^{\mathrm{i}k(\alpha_m - \theta_m)} \right|^2 - 1 \right|^2 \tag{3.5}$$

Details of the calculations can be seen in Trigano et al. (2009). Note that due to (3.5), D_1 has a unique global minimum which is attained when $\alpha_m = \theta_m$, for all $m = 1 \dots K$, that is the actual shift value. We now provide two quantitative results about algorithm **(A1)**. Their demonstrations can be also found in Trigano et al. (2009).

Proposition 1. *Let $\{\eta(K, \lambda),\ K \geq 0\}$ be a sequence such that $\eta(K, \lambda) \to 0$ as $K \to +\infty$ for all λ, and let α be a real positive number. Assume that for all $k = 0 \dots n-1$:*

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

then there exists two positive constants γ, and K_0 such that, for $K \geq K_0$, there is a constant $c(K,\lambda)$ such that the number of curves whose alignment error with respect to $\theta - c(K,\lambda)$ is bigger than $\eta(K,\lambda)^\alpha$, denoted by $\#\{m = 1 \dots K \ : \ |\alpha_m - \theta_m - c(K,\lambda)| > \eta(K,\lambda)^\alpha\}$ is bounded as follows:

$$\#\{m \ : \ |\alpha_m - \theta_m - c(K,\lambda)| > \eta(K,\lambda)^\alpha\} \leq \gamma(K+\lambda)\eta(K,\lambda)^{1-2\alpha}$$

Proposition 1 can be intuitively interpreted as follows: provided the optimization procedure is effective enough, if the number of curves in each block is large enough, most curves will tend to align, but not necessarily with respect to 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, and the following proposition holds:

Proposition 2. *Assume that λ is an integer, and that*

$$\gamma\eta(K,\lambda)^{1-2\alpha} \leq \frac{\lambda}{K+\lambda}.$$

Then, under the assumption of Proposition 1, we get that $|c(K,\lambda)| < \eta(K,\lambda)^\alpha$

In other words, when choosing $\lambda(K)$ such that

$$\lambda(K) \to +\infty, \ \frac{\lambda(K)}{K} \to 0 \text{ as } K \to \infty,$$

it is possible to obtain estimates very close to the actual shifts. In order to check that the optimization procedure can indeed be done effectively, we need the noisy part of the cost function to be small under the same conditions.
We now study the noisy part of $\|C_1(\boldsymbol{\alpha_1})\|_2^2$. Recall that due to Equation (3.2), the noisy part of $\|C_1(\boldsymbol{\alpha_1})\|_2^2$ stems from terms of the form $A_M(k) - B_1(k,\boldsymbol{\theta_1})$ and $B_1(k,\boldsymbol{\theta_1}) - B_1(k,\boldsymbol{\alpha_1})$. It is then possible to show the following proposition:

Proposition 3. *Assume that $K \to \infty$, $\lambda \to \infty$ and that $\lambda/K \to 0$, and let ε be any positive number. Let us denote by $R(k)$ the noisy part associated to $B_1(k,\boldsymbol{\theta_1}) - B_1(k,\boldsymbol{\alpha_1})$; we get under these assumptions that*

$$A_M(k) - B_1(k,\boldsymbol{\theta_1}) = \frac{2\sigma^2}{n} + \mathrm{O}_{\mathbb{P}}(n^{-1}K^{-1/2})$$

and

$$R(k) = \mathrm{O}_{\mathbb{P}}(n^{-1}K^{-1/2+\varepsilon}) + \mathrm{O}_{\mathbb{P}}(n^{-1})$$

Proposition 3 is of importance, since it shows that under the assumptions that each block contains a large number of curves, and that the weighting factor λ is important but negligible with respect to K, the cost function $\|C_1(\boldsymbol{\alpha_1})\|^2$ reduces to $D_1(\boldsymbol{\alpha_1})$ plus a constant term, which means that the minimum of $\|C_1(\boldsymbol{\alpha_1})\|^2$ is with high probability close to the minimum of $D_1(\boldsymbol{\alpha_1})$. In practice, a typical choice of the weighting parameter would be $\lambda = K^\beta$, with $0 < \beta < 1$, and verifies all the assumptions of the previous propositions.

4. Applications

We present in this section the results obtained by the presented algorithms, both for simulated and real data. In our simulations, the shape s is created according to a model used in neuroscience, namely the Hodgkin-Huxley model (see Johnson (1996) and references therein), and measure the performances of the shift alignment procedures by using the MISE (Mean Integrated Squared Error) criterion; more specifically, the shifts $\{\theta_m, m = 0 \dots M\}$ are drawn accordingly to a known probability distribution, then estimated by means of the previously described algorithms and, as a measure of performance, the MISE of the obtained plug-in density estimates proposed in the end of Section 2 is computed.
We first investigate the choice of the tuning parameters of the algorithm **(A1)**, that the number of curves in each block K and the weighting parameter λ, and discuss the influence of these values in the obtained density estimator, for different choices of σ^2. We then investigate the performances of the algorithms **(A1)** and **(A2)**, and how they compare to two standard methods:

— The alignment method with respect to the local extrema, as proposed in Gasser & Kneip (1995) (we denote this algorithm by **(A3)**, and

— a measure of fit based on the squared distance between the average pulse and the shifted pulses leading, which is often used by practitioners and is described in Silveman & Ramsay (2005) (this algorithm is denoted by **(A4)**.

It shall be seen that from the point of view of the shift density estimation, the algorithm **(A1)** outperforms the others.
We then apply the presented algorithms for the computation of the SAECG ; the real ECG data have been taken from the MIT-BIH database, and we present the obtained results for three different types of ECG recordings:

— recordings stemming from a normal heart

— recordings from a patient suffering of cardiac arrhythmia

— recordings of noise-stress tests, that is with a lower SNR.

4.1 Results on simulated data

4.1.1 Experimental protocol

Simulated data are created accordingly to the discrete model 1.1 and we compute the estimators for different values of the parameters K, λ and σ^2. For each curve, we sample in order to get 512 points equally spaced on the interval $[0; 2\pi]$. We make the experiment with s simulated according to the Hodgkin-Huxley model of a neural response. The shifts are drawn accordingly to a normal distribution $\mathcal{N}(0, 32^2)$, and $\theta_0 = 0$.

4.1.2 Results

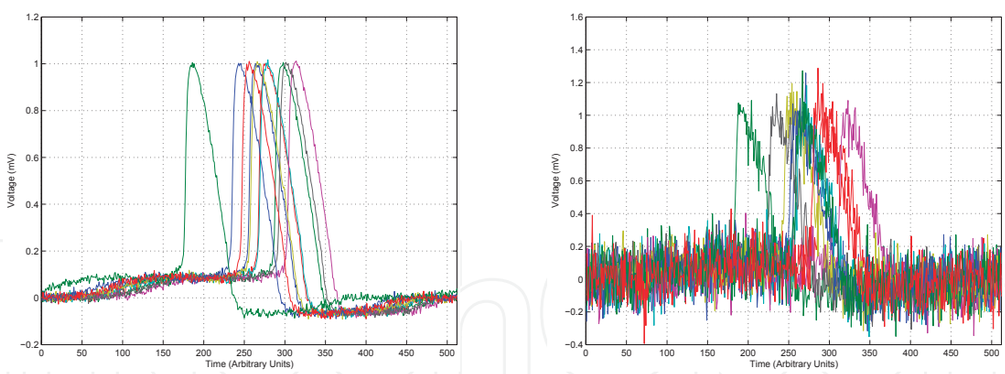
We study the influence of the parameter K and λ empirically by providing the MISE of the plug-in density estimates for different values of K, λ and σ^2, with $N = 100$. In all these experiments, the value of the weighting parameter λ is chosen accordingly to K, that is $\lambda = K^{0.9}$. The numerical values of the MISE are provided for all algorithms in Table 1.
It shall be observed that the algorithm **(A1)** outperforms the other algorithms in most of the cases, except the case of extremely low SNR. We now present graphical results obtained by means of **(A1)**, **(A2)**, **(A3)** and **(A4)**, in the case of high SNR, which can be obtained for example

		K=10	K=20	K=30	K=50	K=100
$\sigma^2 = 0$	**(A1)**	0.0305	0.0228	0.0198	0.0153	0.0106
	(A2)	0.0407	0.0357	0.0372	0.0372	0.0375
	(A3)	0.0306	0.0234	0.0199	0.0156	0.0109
	(A4)	0.0316	0.0248	0.0227	0.0174	0.0136
$\sigma^2 = 10^{-4}$	**(A1)**	0.0312	0.0218	0.0183	0.0156	0.0121
	(A2)	0.0399	0.0383	0.0362	0.0364	0.0364
	(A3)	0.0325	0.0232	0.0212	0.0183	0.0158
	(A4)	0.0322	0.0219	0.0192	0.0168	0.0126
$\sigma^2 = 10^{-2}$	**(A1)**	0.0296	0.0218	0.0172	0.0143	0.0120
	(A2)	0.0410	0.0383	0.0384	0.0371	0.0355
	(A3)	0.0306	0.0232	0.0192	0.0172	0.0143
	(A4)	0.0303	0.0219	0.0182	0.0155	0.0125
$\sigma^2 = 1$	**(A1)**	0.0326	0.0274	0.0248	0.0255	0.0288
	(A2)	0.0460	0.0407	0.0374	0.0381	0.0395
	(A3)	0.0547	0.0806	0.0514	0.0553	0.0741
	(A4)	0.0510	0.0450	0.0414	0.0393	0.0370

Table 1. MISE of the density estimates obtained for different values of the noise variance σ^2 and number of curves in each block K, for a fixed value of $\lambda = K^{0.9}$

.

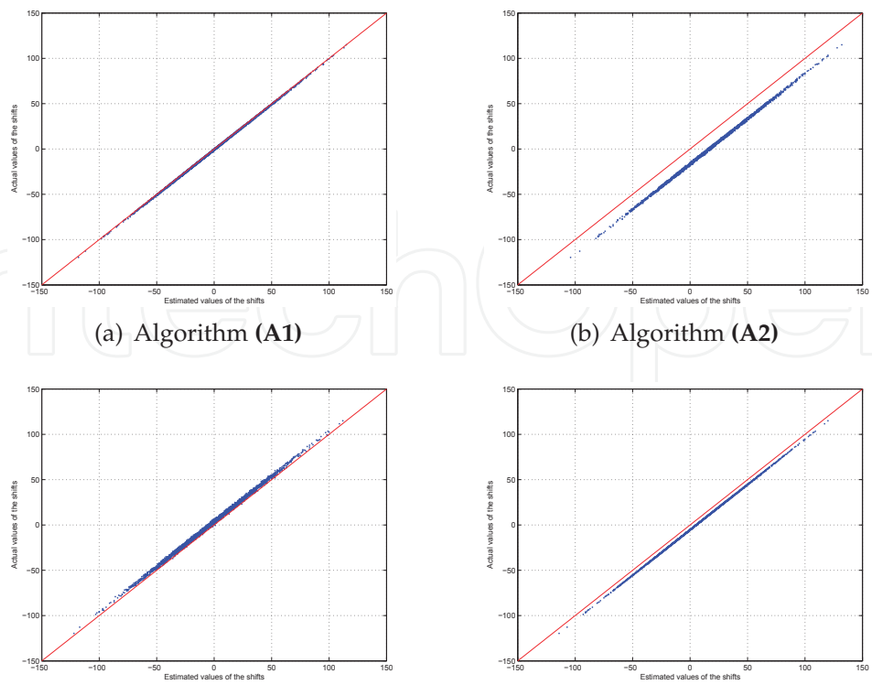
by choosing $\sigma = 0.01$, and in the case of low SNR, which can be obtained by fixing $\sigma = 0.1$. Two typical curves are presented in both cases in figure (3).



(a) High SNR ($\sigma^2 = 10^{-6}$) (b) Low SNR ($\sigma^2 = 10^{-2}$)

Fig. 3. Ten typical curves for different SNR values.

Results are presented for the four algorithms and for these specific choices of σ in figure (4) and figure(5).
Eventually, a graphical comparison of the actual shift distribution and its plug-in density estimate is presented in figure (6), and the results with a weighting parameter λ too small is given in figure (7).



(a) Algorithm **(A1)** (b) Algorithm **(A2)**

(c) Algorithm **(A3)** (d) Algorithm **(A4)**

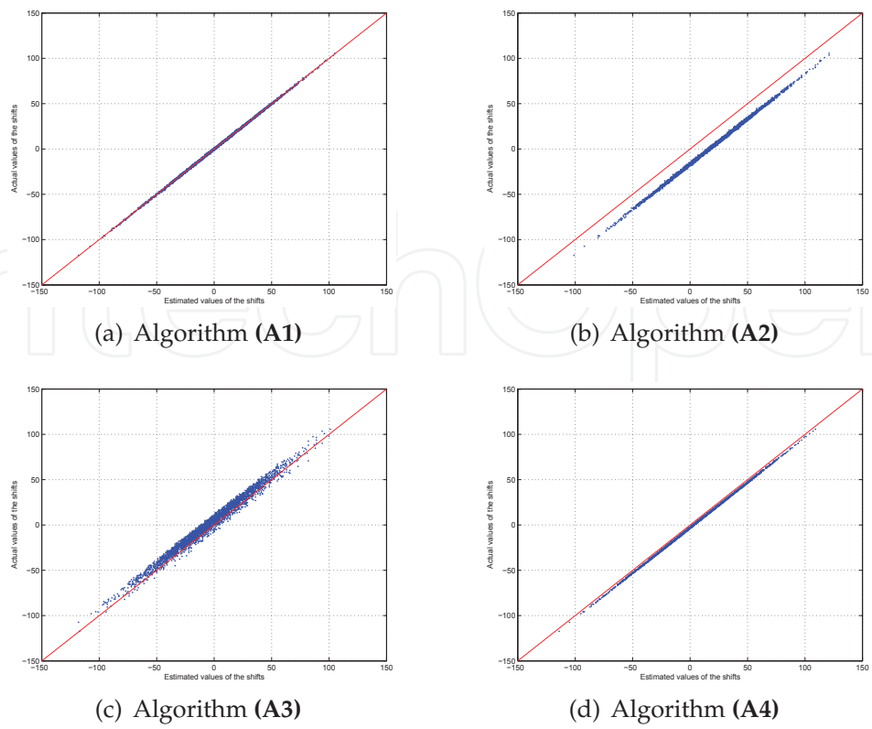
Fig. 4. Results of the curve alignment procedures for $K = 30$, $N = 100$, $\lambda = K^{0.9}$, $\sigma^2 = 10^{-6}$.

4.1.3 Discussion

The graph obtained in figure (7) well illustrates Proposition 1. In this graph, we observe that in each separate block the curves are well aligned, since for each block plotting the actual values of the shifts versus their estimated values gives lines with slope 1. However, they do not align with respect to the location of the reference curve, due to a weighting parameter λ too small. Taking a larger λ allows to address this problem, as it can be seen in figure (4(a)) and figure (5(a)). From the obtained results of Table 1, it can be observed that the algorithm **(A1)** outperforms the others, from the point of view of density estimation. This algorithm is also robust to the noise level σ, as it can be seen from figure (4(a)), figure (5(a)) and the results displayed in Trigano et al. (2008).

Not surprisingly, the algorithm **(A2)** is well suited for alignment (e.g. in order to compute the average signal), but is less adapted for density estimation, since the curves do not align accordingly to the reference curve, as stated in Proposition 1. This means that, provided the expectation of f_θ is known, the algorithm **(A2)** would give results close to **(A1)**. The two-step algorithm should be preferred for example in the case of very low SNR, where each curve is very noisy so that there is no good choice available for a reference curve y_0. Indeed, when comparing the results of **(A1)** and **(A2)** for $\sigma^2 = 0.01$ and $\sigma^2 = 1$, we can see that the MISE degradation is less important for **(A2)** than for **(A1)**.

The method described in Gasser & Kneip (1995), from which the algorithm **(A3)** is derived, is based on the features of the curves s, and is less effective when the noise level is important. This can be observed from figure (5(d)) and the results given in Table 1 for the highest values of σ^2. This is predictable, since any method which relies on a preliminary smoothing of



(a) Algorithm **(A1)** (b) Algorithm **(A2)**

(c) Algorithm **(A3)** (d) Algorithm **(A4)**

Fig. 5. Results of the curve alignment procedures for $K = 30$, $N = 100$, $\lambda = K^{0.9}$, $\sigma^2 = 10^{-2}$.

the curves and an alignment with respect to the maxima of the curve would lead to a more important error than a semiparametric approach if the SNR is low. However, **(A3)** competes relatively well with **(A1)** in the case of high SNR, and may be preferred for simplicity in that case. Nevertheless, since **(A1)** performs an efficient alignment both for high and low values of σ^2, since this parameter introduces only a constant term in the cost functions C_n, which can be omitted in the optimization procedure. We may argue that the main advantage of **(A1)** lies in its generality, and that it should be used for example when the SNR is unknown.
Eventually, it can be observed that **(A4)**, which relates to FDA methods described in Silveman & Ramsay (2005), is relatively efficient, provided the total number of curves is large enough. The MISE results as well as figures (4(c)) and (5(c)) indicate that this method is well fitted for curve alignment. However, a degradation of the performances can be noticed for the lowest values of K. This makes intuitively sense, since a small value of K indicates that only a few curves are used to compute the average signal at each iteration of **(A4)**. Therefore, the alignment w.r.t. the average signal yields a larger MISE. This appears e.g. in figure (4(c)), where a slight bias between the actual shifts and their estimators can be observed. Consequently, the algorithm **(A4)** gives results quite similar to **(A1)**, but requires one block of $N \times K$ curves to compare well to **(A1)**, which leads to significantly longer computational times.
From the theoretical point of view, the good performances of **(A1-2)** with respect to **(A3-4)** can be explained by the study of another M-estimate proposed in Gamboa et al. (2007) for curve alignment, which gives further insight in the comparison with the state-of-the-art method. Indeed, (Gamboa et al., 2007, Theorem 2.1) shows that a statistically consistent alignment can be obtained only when filtering the curves and aligning the low-frequency information.

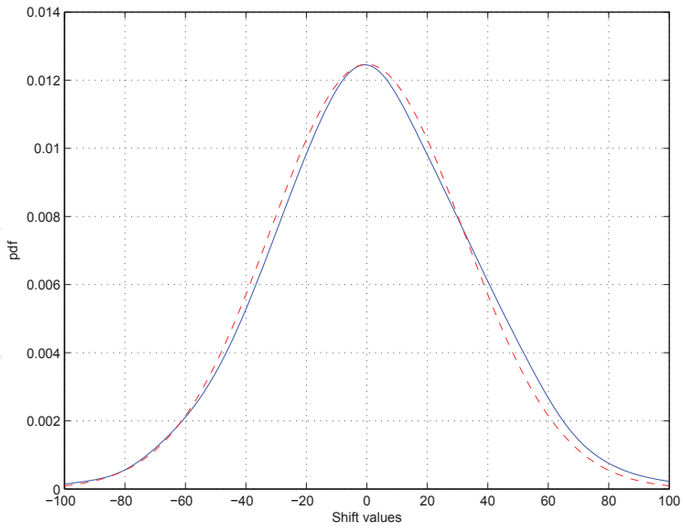


Fig. 6. Comparison between the true shift density (dotted red) and its plug-in estimate (blue).

Therefore, an approach based on the spectral information is more susceptible to achieve good alignment by comparison to the method of Silveman & Ramsay (2005).

4.2 Results on real ECG data

We now wish to compare our method to the state-of-the-art for the alignment of heart cycles, in order to estimate the average signal. We provide the study of the signal presented in figure (1), which was obtained from the MIT-BIH database, and is a recorded signal stemming from both a healthy (figure (8) and (10)) and arrhythmic heart (figure (12)).

4.2.1 Experimental protocol

In order to obtain a series of heart cycles, we first make a preliminary segmentation using the method of Gasser & Kneip (1995), 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 $\lambda = 10$ for 3 iterations.

4.2.2 Results

We present on the following figures results obtained by algorithms **(A1-4)** on three different data sets from the MIT-BIH database. Figure (8) show results obtained on (very) noisy ECG. At first sight, all four methods seem to perform equally well regarding P and T waves. However, zoomed QRS patterns on figure (9) show that alignment procedures **(A1)** and **(A2)** give better results. Indeed, the black mean curve is smoother (jitter is reduced) than the one obtained with the two other methods, reflecting that alignment is achieved in a better way. On the contrary, mean curves obtained by algorithms **(A3)** and **(A4)** are less satisfactory since they are less temporally concentrated ; Moreover, the average curves obtained in that case show several local maxima, which is far from the standard shape of the QRS complex (indeed, we are expecting to find one single mode in this part of the signal, as given by the algorithms **(A1)** and **(A2)**.

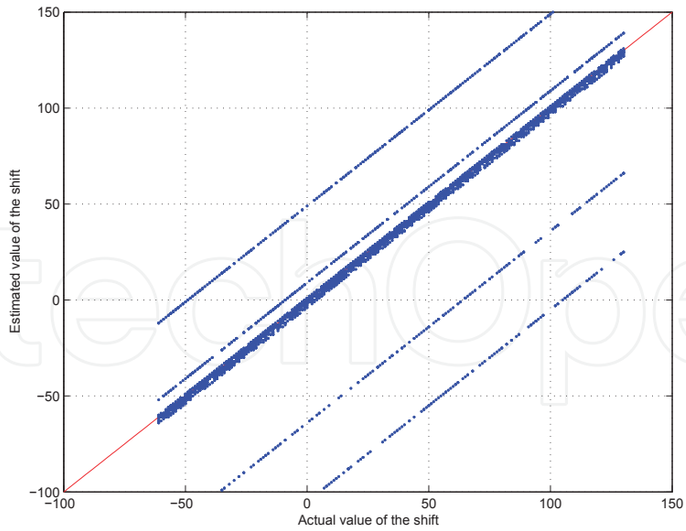


Fig. 7. Results of the shift estimation **(A1)** for $K = 200$, $N = 30$, $\lambda = 10$, $\sigma^2 = 10^{-1}$.

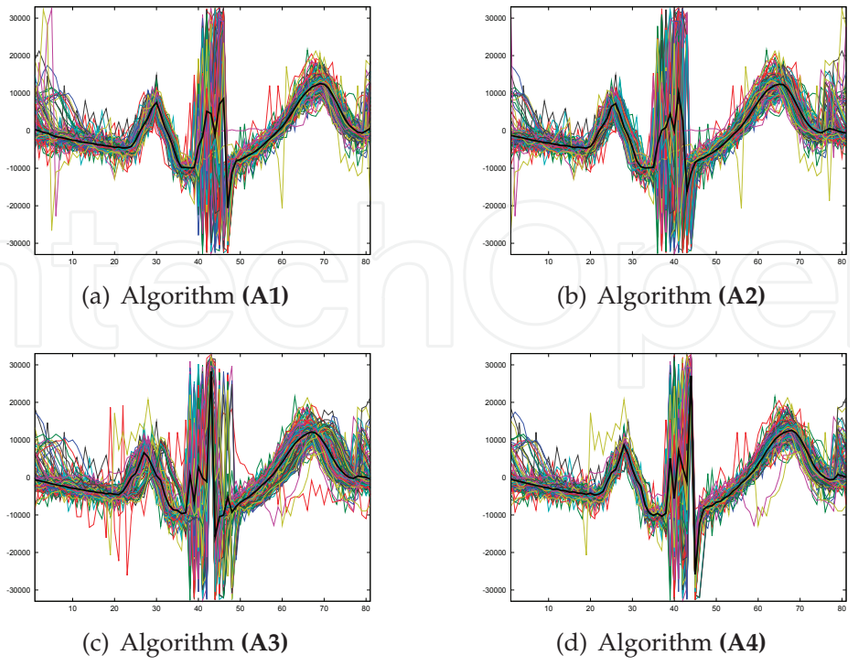
Figure (10) show results obtained with actual power-line interference and baseline shifting. On this data set, algorithm **(A2)** is outperformed by the 3 others as revealed by figure (11(b)). In this case, **(A3)** and **(A4)** perform well since maxima are. Note that **(A1)** mean curve is slightly smoother than **(A3)** and **(A4)** ones.
Figure (12) and (13) show alignment of an arrhythmic data set. These data are typically encountered in routine clinical use and contain ECG with very various waveforms and artifacts. (**A3**-**A4**) clearly achieve alignment according to maxima. Jitter is high in this case, since the shape of the curve can significantly vary from one pulse to another. Consequently, the mean curve does not reflect this variability. Such behavior from (**A3**-**A4**) could be expected, since one underlying assumption is that the curve shape s is made of peaked common noise-free patterns. By contrast, (**A1**-**A2**) achieve better alignment since the resulting jitter is comparatively reduced and the corresponding mean curve smoother. The shape variability is intrinsically taken into account by the way of periodogram coefficients.

4.3 Discussion

Results presented in the previous paragraph show clearly different behaviors of alignment procedure. Algorithms **(A1)** and **(A2)** fully take into account all frequencies composing curves including the different noise contributions, such as low frequencies representing baseline variations or more localized ones such as power-line interference. By contrast, **(A3)** and **(A4)** are based on curve maxima meaning only high frequencies. When these maxima are perturbed by noise or distorted by intrinsic curve change (such as arrhythmic ECG), the corresponding mean curve does not reflects such perturbations. We believe that it makes our periodogram-based alignment methods more robust to the classical ECG perturbations.

5. Conclusion

In this contribution several methods of curve alignment for repeated events were introduced and investigated; this led to plug-in estimate of the density of elapsed times between events. Their performances were presented both on simulations and real ECG, and compared to the

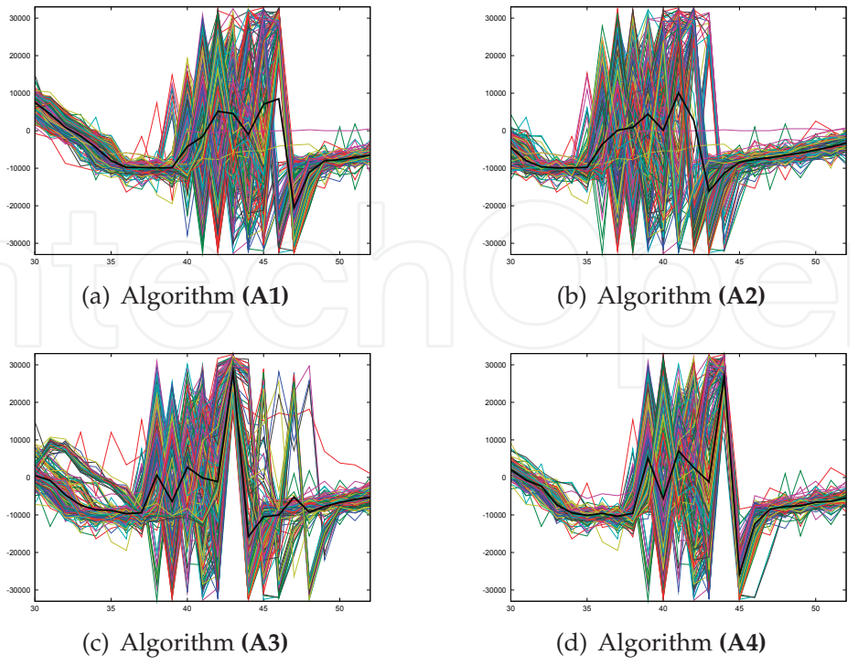


(a) Algorithm **(A1)** (b) Algorithm **(A2)**

(c) Algorithm **(A3)** (d) Algorithm **(A4)**

Fig. 8. Results of the curve alignment procedures on real ECG for $K = 30$, $N = 10$, $\lambda = K^{0.9}$.

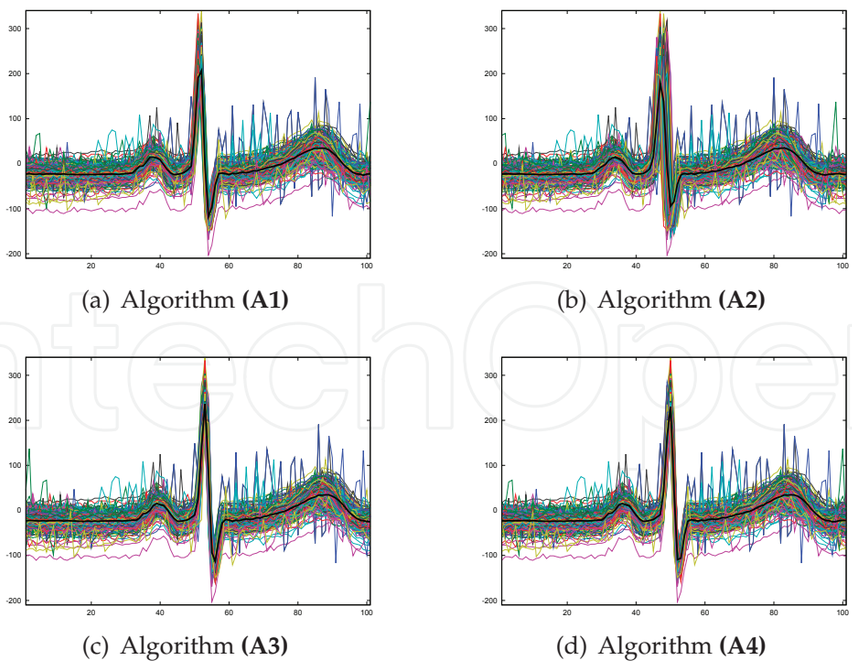
well-known shift correction methods based on the alignment with respect to the local extrema (which is the standard method used in the ECG signal processing framework). It is shown that the algorithms which are based on a semiparametric approach outperform the methods based on an FDA approach. The suggested algorithms provides excellent results, whether the SNR is high or low. On real ECG data, the proposed algorithms gives good results for the computations of the SAECG, and seem according to the experiments extremely robust to distortions such as power-line interference, baseline wander or variability of the pulse shapes. Further results on the rates of convergences of the density estimator and the applicability of the method for the detection of heart diseases should appear in future contributions.



(a) Algorithm **(A1)** (b) Algorithm **(A2)**

(c) Algorithm **(A3)** (d) Algorithm **(A4)**

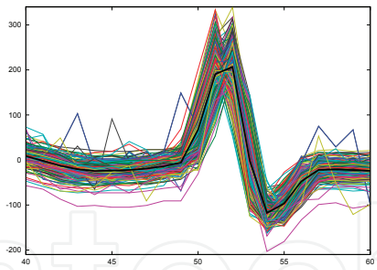
Fig. 9. Zoom on figure (8) on each aligned QRS region.



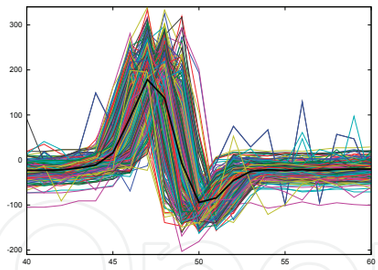
(a) Algorithm **(A1)** (b) Algorithm **(A2)**

(c) Algorithm **(A3)** (d) Algorithm **(A4)**

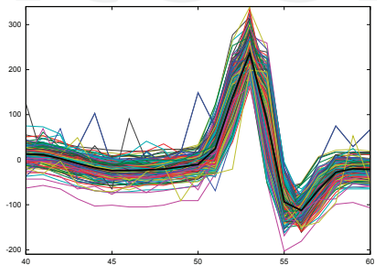
Fig. 10. Results of the curve alignment procedures on real ECG for $K = 30$, $N = 10$, $\lambda = K^{0.9}$.



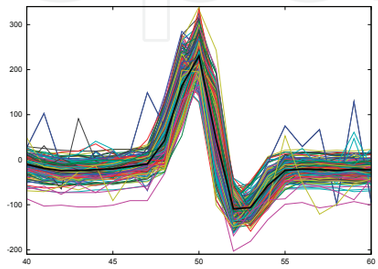
(a) Algorithm **(A1)**



(b) Algorithm **(A2)**

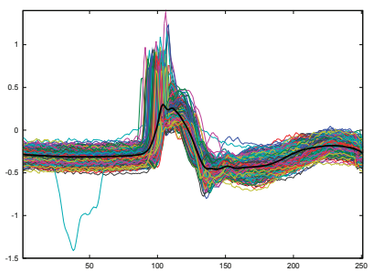


(c) Algorithm **(A3)**

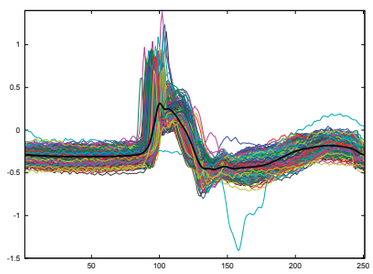


(d) Algorithm **(A4)**

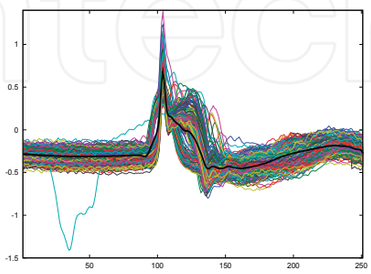
Fig. 11. Zoom of figure (10) on each aligned QRS region.



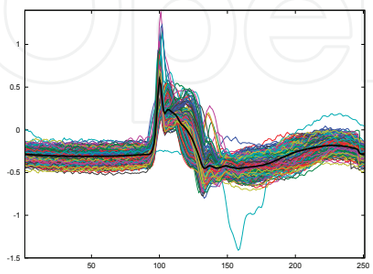
(a) Algorithm **(A1)**



(b) Algorithm **(A2)**

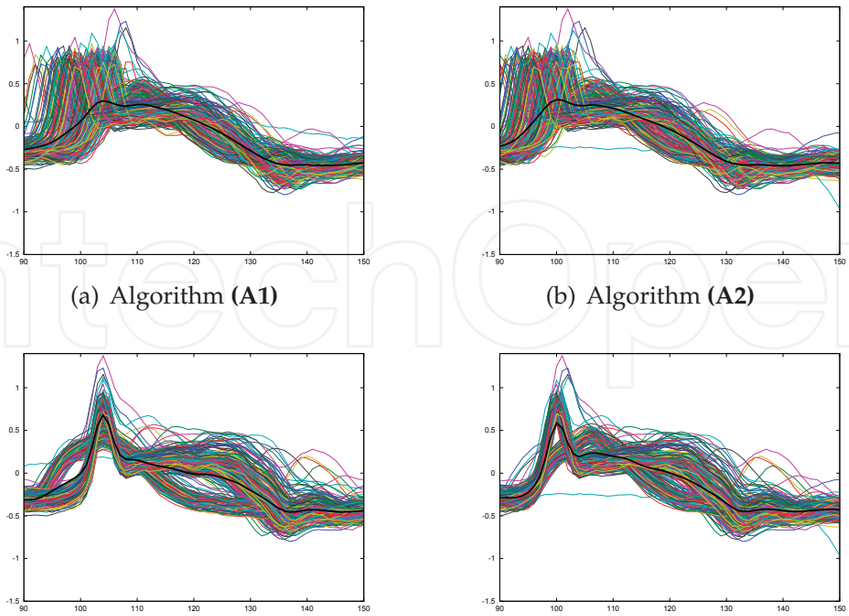


(c) Algorithm **(A3)**



(d) Algorithm **(A4)**

Fig. 12. Results of the curve alignment procedures on real arrhythmic ECG for $K = 30$, $N = 10$, $\lambda = K^{0.9}$.



(a) Algorithm **(A1)** (b) Algorithm **(A2)**

(c) Algorithm **(A3)** (d) Algorithm **(A4)**

Fig. 13. Zoom on figure (12) on each aligned QRS region.

6. References

Bigot, J. & Gadat, S. (2008). An Inverse Problem Point of View for Adaptive Estimation in a Shifted Curves Model, *Submitted* .

Bigot, J., Loubes, J.-M. & Vimond, M. (2008). Semiparametric Eestimation of Shifts on Compact Lie Groups for Image Registration, *Submitted* .

Cain, M., Anderso, J., Arnsdorf, M., Mason, J., Scheinman, M. & Waldo, A. (1996). Signal-Averaged Electrocardiography, *Journal of American College of Cardiology* **27**(1): 238–249.

Castillo, I. (2006). *Estimation semi-paramétrique à l'ordre 2 et applications*, PhD thesis, Université Paris XI.

Castillo, I. & Loubes, J.-M. (2007). Estimation of the Law of Random Shifts Deformation, *submitted to Pattern Analysis, Statistical Modelling and Computational Learning* .

Chonavel, T. (2000). *Statistical Signal Processing*, Springer.

Dalalyan, A. S., Golubev, G. K. & Tsybakov, A. B. (2006). Penalized maximum likelihood and semiparametric second-order efficiency, *Ann. Statist.* **34**(1): 169–201.

Delescluse, M. & Pouzat, C. (2006). Efficient Spike Sorting of Multi-State Neurons Using Inter-Spike Intervals Information, *Technical report*, CNRS UMR 8118.

Englund, A., Hnatkova, K., Kulakowski, P., Elliot, P., McKenna, W. & Malik, M. (1998). Wavelet decomposition analysis of the signal averaged electrocardiogram used for risk stratification of patients with hypertrophic cardiomyopathy, *European Heart Journal* **19**: 1383–1390.

Ferraty, F. & Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*, Springer series in Statistics, 1 edn, Springer.

Fotiadis, D., Likas, A., Michalis, L. & Papaloukas, C. (2006). *Encyclopedia of Biomedical Engineering*, Wiley, chapter Electrocardiogram (ECG): (Automated Diagnosis).

Gamboa, F., Loubes, J.-M. & Maza, E. (2007). Semiparametric Estimation of Shifts Between Curves, *Elec. Journ. of Statist.* **1**: 616–640.

Gasser, T. & Kneip, A. (1995). Searching for structure in curve sample, *J. Am. Statist. Ass.* **90**(432): 1179–1188.

Guyton, A. & Hall, J. E. (1996). *Textbook of Medical Physiology*, 9 edn, W. H. Saunders.

Johnson, D. H. (1996). Point process models of single-neuron discharges, *Journal of Computational Neuroscience* **3**(4): 275–299.

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

Lavielle, M. & Levy-Leduc, C. (2005). Semiparametric estimation of the frequency of unknown periodic functions and its application to laser vibrometry signals, *IEEE Trans. Signal Processing* **53**(7): 2306– 2314.

Lewicki, M. S. (1998). A Review of Methods for Spike Sorting: the Detection and Classification of Neural Action Potentials., *Network: Computation in Neural Systems* **9**(4): R53–R78.

Nava, A., Folino, F., Bauce, B., Turrini, P., Buja, G., Daliento, L. & Thiene, G. (2000). Signal-averaged electrocardiogram in patients with arrythmogenic right ventricular cardiomyopathy and ventricular arrythmias, *European Heart Journal* **21**: 58–65.

Pouzat, C., Delescluse, M., Viot, P. & Diebolt, J. (2004). Improved Spike-Sorting by Modeling Firing Statistics and Burst-Dependent Spike Amplitude Attenuation: a Markov Chain Monte Carlo Approach, *Journal of Neurophysiology* **91**: 2910–2928.

Ramsay, J. O. (1998). Estimating smooth monotone functions, *J. R. Statist. Soc. B* **60**(2): 365–375.

Ramsay, J. O. & Li, X. (1998). Curve registration, *J. R. Statist. Soc. B* **60**(2): 351–363.

Ritov, Y. (1989). Estimating a Signal with Noisy Nuisance Parameters, *Biometrika* **76**(1): 31–37.

Rodriguez, B., Jané, R. & Brooks, D. (2000). Ventricular Tachycardia Risk Detection using Wavelet Decompostion of the Signal Averaged ecg, *Computers in Cardiology* **27**: 731–734.

Ronn, B. (2001). Nonparametric Maximum Likelihood Estimation for Shifted Curves, *J. R. Statist. Soc. B* **63**(2): 243–259.

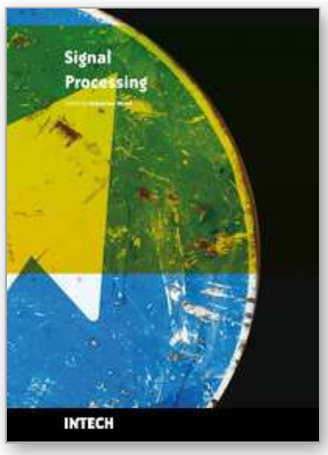
Sarnmo, L. & Laguna, P. (2006). *Encyclopedia of Biomedical Engineering*, Wiley, chapter Electrocardiogram (ECG)Signal Processing.

Silveman, B. W. & Ramsay, J. (2005). *Functional Data Analysis*, Springer Series in Statistics, 2 edn, Springer.

Trigano, T., Isserles, U. & Ritov, Y. (2008). Semiparametric shift estimation for alignment of ecg data, *Proceedings of the EUSIPCO Signal Processing Conference*.

Trigano, T., Isserles, U. & Ritov, Y. (2009). Semiparametric curve alignment and shift density estimation with applications to ecg data, *in revision for IEEE Transactions in Signal Processing* **arXiV preprint**: 0807.1271v3.

Vimond, M. (2008). Efficient Estimation for Homothetic Shifted Regression Models, *To appear in Annals of Statistics* .



Signal Processing

Edited by Sebastian Miron

ISBN 978-953-7619-91-6
Hard cover, 528 pages
Publisher InTech
Published online 01, March, 2010
Published in print edition March, 2010

This book intends to provide highlights of the current research in signal processing area and to offer a snapshot of the recent advances in this field. This work is mainly destined to researchers in the signal processing related areas but it is also accessible to anyone with a scientific background desiring to have an up-to-date overview of this domain. The twenty-five chapters present methodological advances and recent applications of signal processing algorithms in various domains as telecommunications, array processing, biology, cryptography, image and speech processing. The methodologies illustrated in this book, such as sparse signal recovery, are hot topics in the signal processing community at this moment. The editor would like to thank all the authors for their excellent contributions in different areas of signal processing and hopes that this book will be of valuable help to the readers.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen