

# SEMIPARAMETRIC SHIFT ESTIMATION FOR ALIGNMENT OF ECG DATA

Thomas Trigano<sup>†,‡</sup>, Uri Isserles<sup>†</sup> and Ya'acov Ritov<sup>†</sup>

<sup>†</sup> Hebrew University of Jerusalem  
Department of Statistics  
Mount Scopus, 91905, Jerusalem, Israel

<sup>‡</sup> Shamoon College of Engineering  
Department of Electrical Engineering  
77245, Ashdod, Israel

phone: +972-25883312, fax: +972-25883549, email: trigano,urisser,yaacov@mscc.huji.ac.il

## ABSTRACT

We address the problem of curve alignment with a semi-parametric framework, that is without any knowledge of the shape. This study stems from a biological issue, in which we are interested in the estimation of the average heart cycle signal, but wish to estimate it without any knowledge of the pulse shape, which may differ from one patient to another. We suggest in this paper an estimator based on a smoothed functional of the periodogram. Results show better performances than the standard method, as well as its robustness to the noise level.

## 1. INTRODUCTION

The aim of this work is to address a specific curve alignment problem. We pay attention to the issue of estimating either a set of shift parameters  $\{\theta_j, j = 1 \dots M\}$ , or its distribution, when this sample is not observed directly but through its image by an unknown operator  $s$ . More precisely we observe a collection of curves

$$y_{k,l} = s(t_k - \theta_l) + \sigma n_{k,l}, k = 0 \dots m-1, l = 0 \dots M, \quad (1)$$

where the  $n_{k,l}$  are independent standard normal random variables, independent of  $\theta$ ,  $\{t_k, k = 0 \dots m-1\}$  being equally spaced observation times and  $M$  denotes the total number of curves. Such problems appear commonly in practice, for instance in functional data analysis (FDA). In this framework, a common problem is to align curves obtained in a series of experiments before extracting their common features; we refer to [1] for an in-depth discussion on the problem of curve alignment in functional data analysis applications.

In this contribution we focus 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 [2, 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, as well as some malfunctions of the heart, can alter the shape of the signal. We aim at situations where the heart electrical activity is cyclic enough, 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 [3].

It is therefore of interest to estimate the  $\theta_j$  in (1). These estimates can be used afterwards for a more accurate estimation of the heart rate distribution. In regular cases, such estimation can be done accurately by using the common FDA

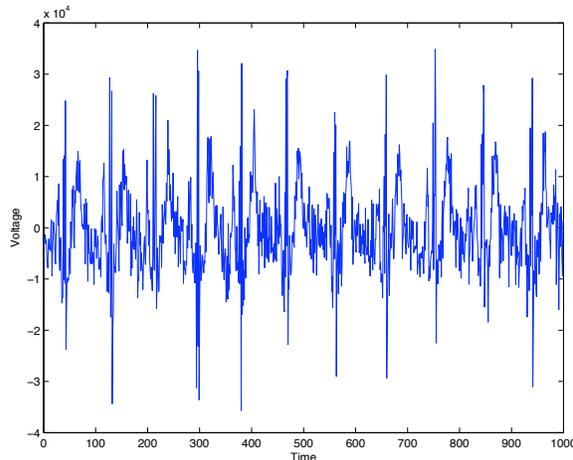


Figure 1: ECG signal from a healthy subject (arbitrary units). This signal is recorded with additional measurement noise, so that FDA alignment methods cannot give good results.

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 arrhythmias, whose identification can be easier if the heart cycles are accurately aligned.

Another measurement often used by cardiologists is the mean ECG signal. A problem encountered in that case is that slightly improperly aligned signals can yield an average on which the characteristics of the heart cycle are lost. The proposed method leads to an estimation of the mean cycle by averaging the segments after an alignment according to an estimated  $\theta_j$ . Additional benefits for a more proper alignment can be found in many other measurements done by cardiologists.

The paper is organized as follows. Section 2 describes the assumptions made and the method to derive the estimator of the shifts, and proposes an estimate for their distribution. Some theoretical results are presented about the settings related to this method. In Section 3, we present results on simulations and real datasets and compare them to the standard method described in [1]. We also comment more precisely the role of the parameters used in our estimator.

## 2. SEMIPARAMETRIC ESTIMATION OF THE SHIFT DISTRIBUTION

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 [4]: first provide an estimate for the shifts, 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 [5] leading to another estimator with good asymptotic properties.

### 2.1 Assumptions

We assume that we always observe the full noisy curve in each sample, which can be formalized by the following assumptions:

1. The distribution of  $\theta$  and the shape  $s$  both have finite support, respectively  $[0, T_\theta]$  and  $[0, T_s]$ .
2.  $T_\theta + T_s < T$ .
3.  $s \in L^2([0, T_s])$ .

Assumptions 1 and 2 imply that we observe a sequence of similar shifted curves with additional noise (that is, that no curve is truncated), so that the spectral information is the same for all recordings. This makes sense for some neuroscience applications, e.g. when estimating the Inter-Spike Interval distribution between neuronal pulses. For the case of ECG data, those assumptions are less realistic, since the QRS-complex of a heart cycle has indeed some observable variability. However, it shall be noted that the low-frequency information remains approximately invariant, thus Assumptions 1 and 2 hold in that sense. Assumption 3 is a standard regularity assumption made on the studied signal. We also denote by  $f$  the probability density function associated to the random variable  $\theta$ .

### 2.2 Description of the shift estimation procedure

We first estimate the sequence  $\{\theta_j, j = 0 \dots M\}$  using an M-estimation procedure, that is by minimizing a cost criterion. In order to define this criterion, we choose to split the set of observed curves in  $N$  blocks of  $K + 1$  curves as represented in Figure 2, and to estimate jointly the sequence of vectors  $\{\Theta_n, n = 1 \dots N\}$ , where for all  $n$

$$\Theta_n \stackrel{\text{def}}{=} (\theta_{(n-1)K+1}, \dots, \theta_{nK}). \quad (2)$$

One important difference, compared to the previously cited works, is that we choose to estimate jointly blocks of parameters instead of one at a time. The estimation of  $\{\theta_n, n = 1 \dots N\}$  is done by minimizing a cost function, which is now described.

Let us denote by  $S_y$  the squared modulus of the Discrete Fourier Transform of a given signal  $y$ . We define, for all  $n = 1 \dots N$ , the mean of curves shifted by some correction terms  $(\alpha_{(n-1)K+1}, \dots, \alpha_{nK})$ :

$$\bar{y}_n(t; \alpha_{(n-1)K+1}, \dots, \alpha_{nK}) \stackrel{\text{def}}{=} \frac{1}{K + \lambda} \left( \lambda y_0(t) + \sum_{l=(n-1)K+1}^{nK} y_l(t + \alpha_l) \right), \quad (3)$$

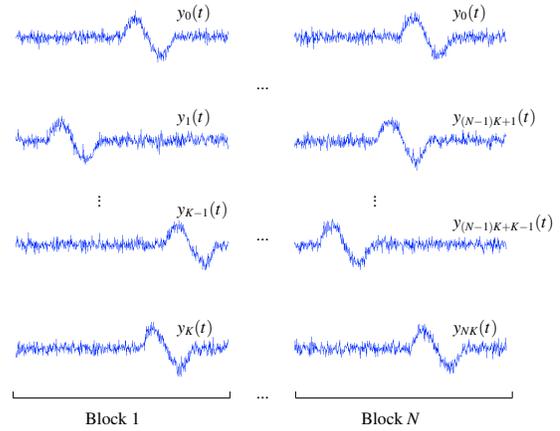


Figure 2: Description of the dataset split into blocks of curves

where  $\lambda$  is a tuning parameter in  $[0; K]$ , and  $y_0$  is a reference curve which is common to all blocks, and is introduced in order to avoid problems of identifiability. We introduce the following cost criterion to be minimized in order to align all curves into the  $n$ -th block:

$$C_n(\alpha_{(n-1)K+1}, \dots, \alpha_{nK}) \stackrel{\text{def}}{=} \left\| \frac{1}{M+1} \sum_{l=0}^M S_{y_l} - S_{\bar{y}_n(\cdot; \alpha_{(n-1)K+1}, \dots, \alpha_{nK})} \right\|_2^2. \quad (4)$$

Note that  $C_n$  attains its minimum if the curves in a block are well aligned. The M-estimator of  $\Theta_n$  is therefore given by

$$\hat{\Theta}_n \stackrel{\text{def}}{=} \underset{x_j \in [0; T]}{\text{Arg min}} C_n(x_1, \dots, x_K). \quad (5)$$

When we are interested in the distribution of the shifts rather than their values, the estimator of  $f$ , denoted by  $\hat{f}$  is naturally computed by plugging the estimated values of the shifts in a regular kernel density estimator, that is for all real  $x$  in  $[0; T]$ :

$$\hat{f}(x) = \frac{1}{Mh} \sum_{l=1}^M \Psi \left( \frac{x - \hat{\theta}_l}{h} \right), \quad (6)$$

where  $\Psi$  is a kernel function integrating to 1 and  $h$  the classical tuning parameter of the kernel.

### 2.3 Properties and settings of the shift estimator

#### 2.3.1 Choice of the number of curves $K$

Since we observe noisy curves and did not assume any knowledge on the spectral information in (4), it has to be estimated first. A well known nonparametric estimator is the periodogram, which has been extensively studied. However, the periodogram of a single curve is too noisy, regardless of the actual regularity of the true spectral information. A good way to reduce the variance of this estimator is given by the averaged periodogram, based on the mean of several periodogram estimators, thus the necessity of splitting the

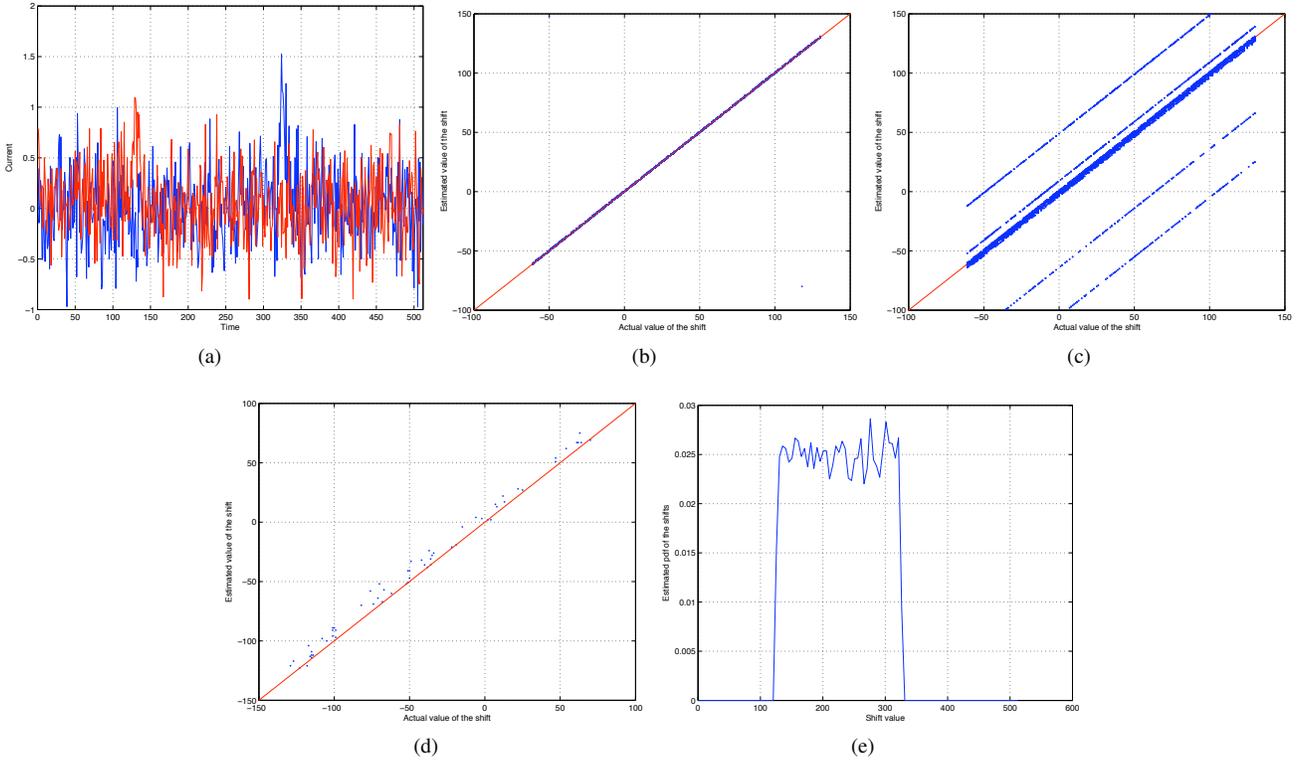


Figure 3: Results for  $K=200$  and  $\sigma^2 = 0.1$ . (a) Two curves before alignment. (b) Comparison between estimated against actual values (blue dots) of the shifts for  $\lambda = 50$ : good estimates must be close of the identity line (red curve). (c) Comparison between estimated and actual values of the shifts for  $\lambda = 10$ . (d) Shift estimation using Least Square Estimate (see [1]) for one block. (e) Probability density estimation for  $N = 20$ ,  $K = 200$  and  $\sigma^2 = 0.1$ .

dataset. This allows to get a “smooth” function  $C_n$  whose optimization leads to an estimate close to the actual values of the shifts. Thus, in the case of noisy curves, the parameter  $K$  should be chosen big enough to make the noise term vanish. It can be shown in [6] that the cost function of a given block  $C_n$  can be decomposed in  $C_n = D_n + R_n$ , where

$$D_n(\alpha_{(n-1)K+1}, \dots, \alpha_{nK}) \stackrel{\text{def}}{=} \sum_{k=0}^{m-1} \left| \frac{1}{K+\lambda} \left( \lambda + \sum_{l=1}^K e^{\frac{2ik\pi(\alpha_l - \theta_l)}{\gamma}} \right) \right|^2 - 1 \quad (7)$$

is the deterministic part of  $C_n$ , whose minimum is attained if and only if  $\alpha_l = \theta_l$  for all  $l = (n-1)K+1 \dots nK$ . Moreover, we get the following result on the random part of the cost function:

**Proposition 2.1** Denote by  $R_n$  the random part associated to the cost function  $C_n$ . Thus, as  $K$  tends to infinity, we get for all  $\beta \in (0, 1)$  that

$$R_n = \frac{2\sigma^2}{m} + O_{\mathbb{P}}(K^{-1+\beta}).$$

Proposition 2.1 indicates that, provided  $K$  is large enough, the cost function  $C_n$  is close to  $D_n + \sigma^2/m$ , thus validating the optimization procedure.

It can also 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 (5) by  $\hat{\theta} + c + 2k\pi$ ,  $k \in \mathbb{Z}$  and  $s$  by  $s(\cdot - c)$  would let the cost criterion invariant. Adding curve  $y_0$  as a referential allows to estimate  $\theta - \theta_0$ , thus avoiding the unidentifiability of the model.

### 2.3.2 Choice of the weight parameter $\lambda$

We now briefly discuss the choice of the tuning parameter  $\lambda$ . In the estimator (5),  $\lambda$  is chosen to give more importance to the curve  $y_0$ , thus forcing the other curves to align accordingly. Indeed, if we take  $\lambda = 1$  (that is, all curves have the same weights), the following proposition holds:

**Proposition 2.2** Let  $\{\eta(K), K \geq 0\}$  be a sequence such that  $\eta(K) \rightarrow 0$  as  $K \rightarrow +\infty$ . Assume that

$$\left| \frac{1}{(K+1)} \left( 1 + \sum_{1 \leq l \leq K} \exp\left(\frac{2i\pi k}{T}(\theta_l - \alpha_l)\right) \right) \right| > 1 - \eta(K),$$

then there exists three constants  $c$ ,  $\gamma$  and  $K_0$  such that, for  $K \geq K_0$ , the number of curves whose alignment error is far from  $c$ , denoted by  $\#\{k : |c - \alpha_k + \theta_k| > \eta(K)^\alpha\}$  is bounded as follows:

$$\#\{k : |c - \alpha_k - \theta_k| > \eta(K)^\alpha\} \leq \frac{1}{\gamma} K \eta(K)^{1-2\alpha}.$$

$K \backslash \sigma^2$	$10^{-6}$	$10^{-2}$	$10^{-1}$
2	$4.45 \cdot 10^{-3}$	$5.00 \cdot 10^{-3}$	$5.15 \cdot 10^{-3}$
20	$5.03 \cdot 10^{-3}$	$4.95 \cdot 10^{-3}$	$5.09 \cdot 10^{-3}$
50	$4.95 \cdot 10^{-3}$	$5.24 \cdot 10^{-3}$	$4.92 \cdot 10^{-3}$
100	$5.01 \cdot 10^{-3}$	$4.90 \cdot 10^{-3}$	$5.15 \cdot 10^{-3}$
200	$4.80 \cdot 10^{-3}$	$5.85 \cdot 10^{-3}$	$5.64 \cdot 10^{-3}$

Table 1: Estimated MISE values for different block sizes  $K$  and noise variances  $\sigma^2$ .

Proofs of Propositions 2.1 and 2.2 can be found in [6]. Proposition 2.2 illustrates that, as the number of curves increases in each block, the curves all tend in practice to align around the same value,  $c$ , which is not necessarily equal to 0. Thus, giving more weight to the reference curve  $y_0$  allows to align the curves accordingly to  $y_0$ . Previous considerations implies the existence of a trade-off for the tuning parameters  $K$  and  $\lambda$ : indeed, Propositions 2.2 and 2.1 indicates that  $K$  should be chosen big in order to well align the curves, but this goes with larger computational time for the optimization procedure. A practical method is therefore to choose  $\lambda$  from the beginning, and make  $K$  increase until the average of  $K$  periodograms does not vary much, that is

$$\left\| \frac{1}{K+1+\lambda} \left( \lambda^2 S_{y_0} + \sum_{1 \leq k \leq K+1} S_{y_k} \right) - \frac{1}{K+\lambda} \left( \lambda^2 S_{y_0} + \sum_{1 \leq k \leq K} S_{y_k} \right) \right\| \leq \varepsilon, \quad (8)$$

where  $\varepsilon$  is a threshold chosen by the user. Such a method guarantees in practice that the cost function to optimize is smooth enough for  $K$  being as small as possible.

### 3. APPLICATIONS

We present in this section results based on simulations and on real ECG data. 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.

In the case of simulations, we study the influence of the parameter  $K$  and  $\lambda$  empirically by providing the Mean Squared Integrated Error (MISE) error for different values of  $K$ ,  $\lambda$  and  $\sigma^2$ , with  $N = 20$ .

#### 3.1 Results on simulations

##### 3.1.1 Experimental protocol

Simulated data are created accordingly to the discrete model 1, and we compute the estimators for different values of the parameters  $K$ ,  $\lambda$  and  $\sigma^2$ . 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 uniform distribution  $\mathcal{U}(120\pi/256, 325\pi/256)$ , and  $\theta_0 = \pi$ .

##### 3.1.2 Results

We present in Figure 3 results obtained in 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 for landmark alignment are displayed in Figure 3(d). We observe that this shift estimation procedure is less efficient. An example of density estimation is displayed in Figure 3(e), using a uniform kernel. We retrieve the uniform distribution of  $\theta$ . Table 1 shows the estimated Mean Integrated Squared Error (MISE), with different values of  $K$  and  $\sigma^2$  and  $\lambda = K$ .

##### 3.1.3 Discussion

Figure 3(c) is a good illustration of Proposition 2.2. In this graph, we observe that in each block the curves are well aligned, since we get for each block that the estimated shifts are distributed according to a line with slope 1, but that they do not align with respect to the location of the reference curve, due to a weighting parameter  $\lambda$  too small. Taking a larger  $\lambda$  allows to address this problem, as it may be seen in Figure 3(b). We thus observe that if  $K$  and  $\lambda$  are well chosen, the shift estimation procedure performs well even if the noise level is high. The method from [1] is less performant when the noise level is too high. Indeed, since the average can be very flat in the case of low SNR, this leads to estimation errors possibly important. On the other hand, the averages of periodograms remain relatively robust to the noise level in all cases, since the noise variance introduces only a constant term, which can be omitted in the optimization procedure. We observe on Table 1 that the parameter  $K$  must be chosen carefully: indeed, if  $K$  is large, the noise terms vanishes, thus making the cost functions  $C_n$  regular enough, but make the optimization problem more difficult to solve. These two considerations induce to find a trade-off in practice.

From the theoretical point of view, the study of another M-estimate proposed in [7] for curve alignment gives further insight in the comparison with the state-of-the-art method. Indeed, [7, Theorem 2.1] shows that a statistically consistent alignment can be obtained only when filtering the curves and aligning the low-frequency information. Therefore, an approach based on the spectral information is more susceptible to achieve good alignment by comparison to the method of [1].

#### 3.2 Results on real 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 Hadassah Ein-Karem hospital, and is a recorded signal stemming from a healthy heart.

##### 3.2.1 Experimental protocol

In order to obtain a series of heart cycles, we first make a preliminary segmentation using the method of [3], 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  $\lambda = 10$ .

##### 3.2.2 Results

Results on real data are presented in Figure 4. It can be noticed that the semiparametric method outperforms the stan-

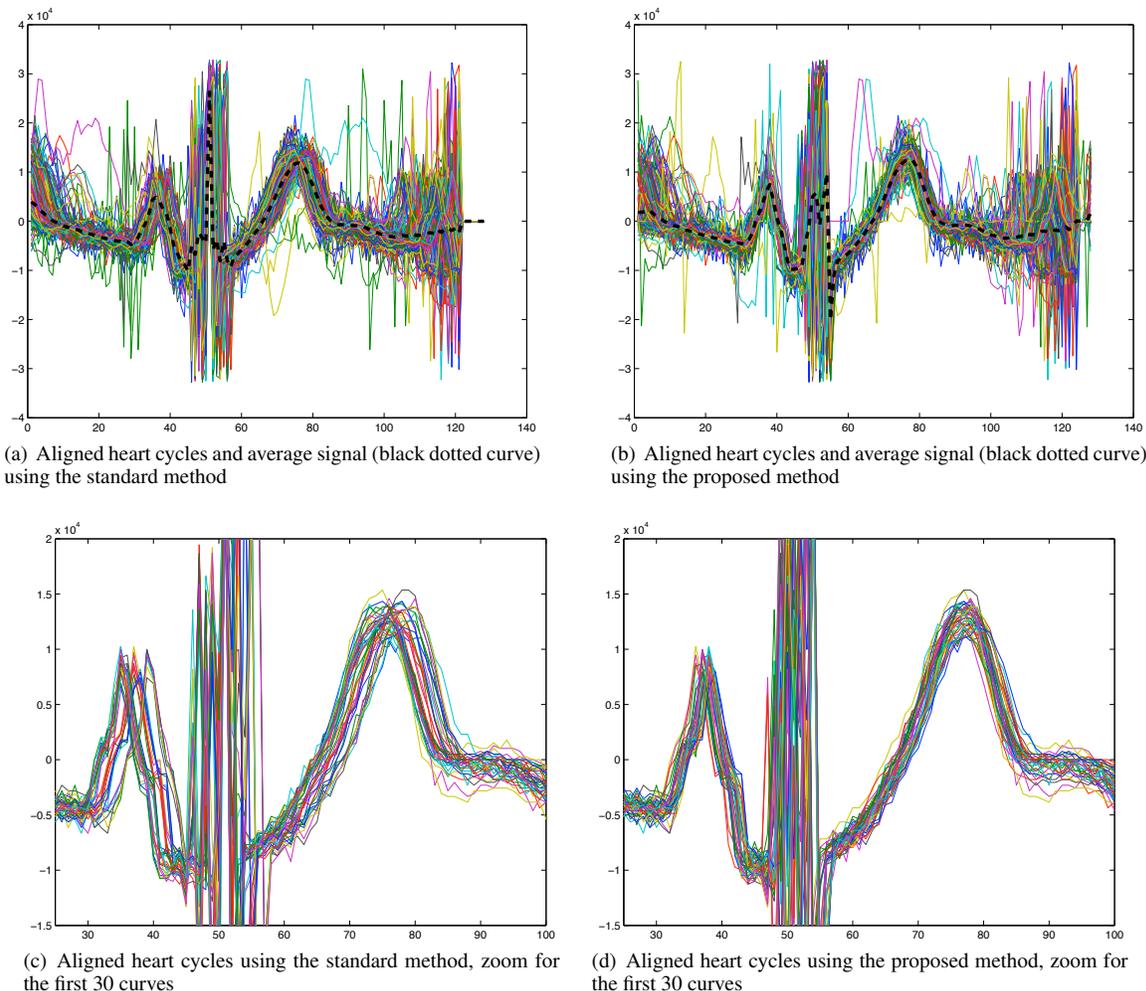


Figure 4: Comparison between the state-of-the-art and the proposed method for the alignment of heart cycles (arbitrary units). A semiparametric approach appears more appealing to align cycles correctly, and allows to separate more efficiently the P-wave, the QRS complex and the T-wave on the average heart cycle.

standard method, by comparing Figures 4(d) and 4(c). Moreover, when computing the average of the reshifted heart cycle, we observe that our method allows to separate more efficiently the different 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 Figure 4(a) and Figure 4(b). This makes the obtained average heart cycle more relevant, since small variations on the position of the QRS complex, which varies quickly, can yield an average where the shape information on the QRS-complex is lost.

#### 4. CONCLUSION

We proposed in this paper a method for curve alignment and density estimation which showed good performances on simulations, even when the noise variance is high. On real ECG data, the proposed method outperforms the standard functional data analysis method, thus leading to a more significant average signal. The theoretical study of the proposed estimator, in terms of consistency and rates of convergence, will appear in a future contribution.

#### REFERENCES

- [1] B. W. Silverman and J. Ramsay. *Functional Data Analysis*. Springer Series in Statistics. Springer, 2 edition, 2005.
- [2] A. .C. Guyton and J. E. Hall. *Textbook of Medical Physiology*. W. H. Saunders, 9 edition, 1996.
- [3] T. Gasser and A. Kneip. Searching for structure in curve sample. *J. Am. Statist. Ass.*, 90(432):1179–1188, 1995.
- [4] I. Castillo. *Estimation semi-paramétrique à l'ordre 2 et applications*. PhD thesis, Université Paris XI, 2006.
- [5] M. Vimond. Efficient Estimation for Homothetic Shifted Regression Models. *To appear in Annals of Statistics*, 2008.
- [6] U. Isserles, Y. Ritov, and T. Trigano. Semiparametric density estimation of shifts between curves. *Submitted*, 2008.
- [7] F. Gamboa, J.-M. Loubes, and E. Maza. Semiparametric Estimation of Shifts Between Curves. *Elec. Journ. of Statist.*, 1:616–640, 2007.