Offset removing in the domain of signal shapes

Hervé Rix[1]

[1] Laboratoire I3S, UMR 6070 UNS-CNRS, Université de Nice-Sophia Antipolis, Les Algorithmes Euclide-B, 2000 Rte des lucioles, BP 121, 06903 Sophia Antipolis cedex.

Abstract-- The recognition of a given shape in a positive signal using the Distribution Function Method (DFM) assumes an affine transform on the abscissa and only a multiplicative coefficient on the ordinates, without any offset. The aim of the paper is to extent DFM when shape equality includes an offset. In fact, this problem is a particular case of signal shape recognition in a sum of two signals whose shapes are known. The first application in mind is the beat to beat extraction of a given electrocardiographic (ECG) wave (e.g. T-wave) added to a random offset. The method for signal shape recognition will be first presented in view of this application. Then a simulation study is done, pointing out the influence of noise on estimations.

Key Words

Signal shape, Distribution Function, Combination of shapes, positive signals.

I. INTRODUCTION

The notion of shape of a signal received several definitions, more or less precise. When the shape is introduced with the idea of clustering a set of signals into a partition of co-sets of equal shape signals, the classical way is to consider a group of functions which, by definition, leave the shape invariant. Shape equality is then a relation of equivalence on the signal set, each class of equivalence representing a unique shape. Dealing with 1D-signals, the group can act both on the left (on the amplitudes) and on the right (on the abscissa) of the signals. In fact two different groups G_1 and G_2 could be chosen for left and right, leading to the following definition: two signals s(t) and v(t), functions of the real variable t, are the same shape, according to G_1 and G_2 if and only if we can write:

$v = u_1 \circ s \circ u_2$, where $u_1 \in G_1$ and $u_2 \in G_2$.

Commonly, the affine group is used in the applications, for the left and right sides. In this case, u_1 and u_2 are two increasing affine functions, so that:

$$v(t) = ks(at+b)+c \quad k>0,\ a>0$$

Often, in applications, the parameter c is removed and assumed to be zero. If this hypothesis can be accepted and the signals being positive on their support, the Distribution Function Method (DFM) [1], has been proven to detect, measure and eventually model subtle differences in shape. In chromatography [2] it makes it possible to reduce a lot the detectability limit of a second component in an apparently pure peak, in comparison with the previous algorithms based on model fitting or deconvolution. In biomedical engineering [3-4], the similarity criterions proposed by the method have been completed by shape averaging techniques using the normalized integrals, for clustering purpose [5-8]. Since the DFM works on signals after removing the base line, the aim of this paper is to extent the method, including a possible offset in shape equality. The idea of looking for signal shape through a normalized integral function, which is subjacent to all these works, needs to deal with positive signals. This is the case when the signals are proportional to probability density functions (pdf) (e.g. chromatographic peaks, distributions of time of flight of photons) of when they are spectra. When this property is not true, we need to work on a positive function of the signals, for example taking the square or the absolute value, or adding an offset. Concerning this last possibility, it can be easily shown that if we add an offset to each signal so that the new minimum is zero, the equality of shape is preserved. The signal shapes are changed, but if they are equal they remain equal. In the following, a new method of signal shape recognition including an eventual offset is presented (section II) and a numerical simulation illustrates its performances in presence of noise added to the observations (section III) before conclusions.

II. MATEARIALS AND METHODS

Let $(s_j(t))_{j=1,\ldots,N}$ be a set of signals and $s_0(t)$ a reference signal. All the supports are assumed to be included in the time interval [0, T], and the signals positive on their support. Following DFM notations, let us define the equality of shape for two signals s(t) and v(t) by the relation:

$$v(t) = \frac{k}{\alpha} s\left(\frac{t-t_0}{\alpha}\right) \qquad k>0,\ \alpha>0 \qquad (1)$$

Now we are looking for a signal $s_j(t)$ which is the same shape as $s_0(t)$, within an offset:

$$s_j(t) = k_j \frac{1}{\alpha_j} s_0(\frac{t-t_j}{\alpha_j}) + c_j \Pi_{[0,T]}(t) \quad (2)$$

$\Pi_{[0,T]}(t) = 1 \; for \quad 0 \le t \le T \; and \; 0 \; elsewhere$.

Going to the integral functions, assuming the signals are positive on their support, we define:

$$S_0^*(t) = \frac{\int_0^t s_0(\tau)d\tau}{\int_0^T s_0(\tau)d\tau} = \frac{\int_0^t s_0(\tau)d\tau}{As_0} = \frac{S_0(t)}{As_0} \quad (3)$$

where As_0 is the area under s_0.

$$D^*(t) = \frac{t}{T}, \quad \text{if } 0 < t < T \quad (4)$$

is the distribution function of an offset on [0, T].

$$S_j(t) = \int_{-\infty}^t s_j(\tau)d\tau \quad , S_j^*(t) = \frac{S_j(t)}{As_j} \quad (5)$$

Assuming equation (2), we have:

$$S_j(t) = k_j \int_{-\infty}^t \frac{1}{\alpha_j} s_0(\frac{\tau - t_j}{\alpha_j})d\tau + c_j \int_{-\infty}^t \Pi_{[0,T]}(\tau)d\tau \quad (6)$$

$$As_j = S_j(+\infty) = k_j As_0 + c_j T \quad (7)$$

$$S_j^*(t) = \frac{k_j S_0(\frac{t-t_j}{\alpha_j}) + c_j D(t)}{k_j As_0 + c_j T} \quad (8)$$

$$S_j^*(t) = \frac{S_0^*(\frac{t-t_j}{\alpha_j}) + \frac{c_j T}{k_j As_0} D^*(t)}{1 + \frac{c_j T}{k_j As_0}} \quad (9)$$

$$S_j^*(t) = \mu S_0^*(\frac{t-t_j}{\alpha_j}) + (1-\mu) D^*(t) \quad (10)$$

$$where \; \mu = \frac{1}{1 + \frac{c_j T}{k_j As_0}} \qquad 0 < \mu < 1 \quad (11)$$

The normalized integral function, which is a distribution function (df), of signal s_j is thus a convex combination of the df of the reference signal composed with an affine function, and the df of the offset. Without offset, $c_j = 0$, $\mu=1$, we retrieve in (10) that sj and s_0 are the same shape, according to definition (1).

Now, from (10) we can write:

$$S_0^*(\frac{t-t_j}{\alpha_j}) = \frac{1}{\mu} S_j^*(t) + (1 - \frac{1}{\mu}) D^*(t) \quad (12)$$

$$S_0^*(\frac{t-t_j}{\alpha_j}) = \beta S_j^*(t) + (1-\beta) D^*(t)$$

where $\beta = \frac{1}{\mu} > 1$ (13)

Application to signal shape recognition:

In that case, the data are the reference signal $s_0(t)$, the time length of the observation window T, and the observed signal $s_j(t)$; the unknown parameters are t_j, α_j, k_j and c_j .

If $s_j(t)$ and $s_0(t)$ are, within an offset, the same shape, then there exits $\beta > 1$, such that the linear combination given by the right hand side of (13) is a df linked to $S_0^*(t)$ by an increasing affine function. In other words, putting:

$$H_\beta(t) = \beta S_j^*(t) + (1-\beta) D^*(t),$$

we have to minimize, in function of β, the shape difference between $H_\beta(t)$ and $S_0^*(t)$. The estimation of β_{min}, corresponding to the shape difference minimum, is done using the Distribution Function Method (DFM) [1]. For example, we need to look for the value of β which gives the minimum departure from the least mean square line fitted on the function:

$$S_0^{*-1}(\beta S_j^*(t) + (1-\beta) D^*(t)).$$

Note that estimating β_{min} gives an estimate of ratio c_j/k_j. The coefficients of the least mean square line give estimations of parameters t_j and α_j.

III. RESULTS

Starting with equation (13), we took a Gaussian shape:

$s_0(t) = (1/\sqrt{2\pi}) \exp(-t^2/2)$ in the observation window $-5 \le t \le +5$ (14)

Fixing the index j = 1, we assume the observed signal $s_1(t)$ is the sum of a Gaussian shape signal and an offset, in the same observation window.

$$s_1(t) = \frac{k_1}{\alpha_1} s_0(\frac{t-t_1}{\alpha_1}) + c_1 \quad -5 \le t \le 5 \quad (15)$$

For given values of the parameter β, let us consider the function:

$$H_\beta(t) = \beta S_1^*(t) + (1-\beta) D^*(t) \quad (16)$$

SNR	a_1	$\hat{a}_1 \pm$ std	$\hat{t}_1 \pm$ std	(c_1/k_1)^
$21dB$	0.8	0.79 ± 0.014	0.50 ± 0.02	0.082
	1.0	0.984 ± 0.018	0.50 ± 0.025	0.0825
	1.2	1.185 ± 0.02	0.50 ± 0.03	0.0825
xxx	xx	xxxxxxxxxxxxx	xxxxxxxxxxx	xxxxxxx
$27dB$	0.8	0.797 ± 0.007	0.501 ± 0.009	0.0805
	1.0	0.997 ± 0.009	0.50 ± 0.01	0.0805
	1.2	1.20 ± 0.01	0.50 ± 0.01	0.0805
xxx	xx	xxxxxxxxxxxxx	xxxxxxxxxxx	xxxxxxx
$33dB$	0.8	0.800 ± 0.003	0.500 ± 0.005	0.080
	1.0	1.000 ± 0.0045	0.500 ± 0.006	0.080
	1.2	1.200 ± 0.005	0.500 ± 0.007	0.080

TABLE 1 Estimation of a_1, t_1 and the ratio c_1/k_1, for 3 values of SNR. Statistics made on 1000 noisy realizations of noise added to the observed signal. The values of c_1 and k_1 are constant and respectively equal to 0.04 and 0.5.

Assuming β belongs to the interval [1, β_0], allowing this function be a distribution function, we are looking for the value of β which minimizes the shape difference with $S_0^*(t)$ given in (3). Interpolating the inverses of the both distribution functions on 100 points equally distributed in [0, 1], we obtain 100 couples (t_i, t_i'). The least mean square line fitting t' in function of t gives a mean residue which is the similarity criterion. Fig.1 shows this criterion called DEL in function of parameter β, without noise. The simulation was done with:

$$k_1 = 0.5; \quad \alpha_1 = 1; \quad t_1 = 0.5; \quad c_1 = 0.04$$

The corresponding value of β, say β^*, making equation (13) is true is β^*= 1.8. In Fig.1, we can check that the value β_{min} which minimizes the similarity criterion DEL is equal to 1.8 too. The criterion being quite zero, all the parameters are well estimated.

Now, adding sequences of zero mean Gaussian noise, using randn from MATLAB, and averaging the estimations on 1000 realizations of the noise, lead to results in TABLE 1, for tree values of the SNR which are in the range of realistic values for ECG waves.

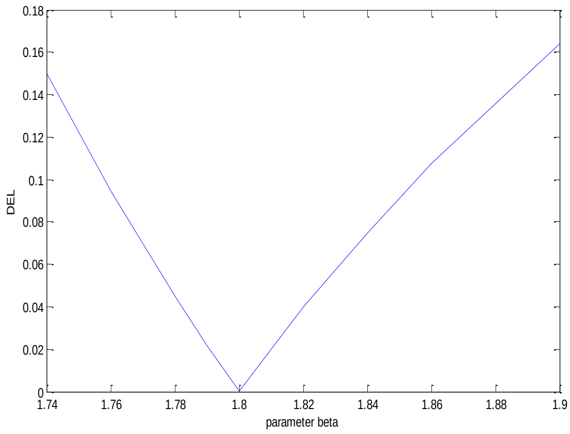

Fig.1: Noise free signal. Shape difference between H_β and S_0^*, in function of β

IV. DISCUSSION – CONCLUSIONS

These simulation results show the proposed method is able to estimate the parameters with accuracy even in presence of noise. Actualy the method can be viewed as an extension of the DFM, when the equality of shape is now defined by the composition with an affine function both on the left and on the right of the signal. But the offset can be replaced by another signal whose shape is known. The separation of two overlapping components is another example giving good results [5].

AKNOWLEDGMENTS

This paper is presented in the framework of contract: ANR 2010 JCJC 0303 01 "PERSIST".

REFERENCES

[1] H. Rix, J.P. Malengé, Detecting small variations in Shape, IEEE *Trans. on Systems, Man and Cybernetics*, Vol. 10, pp. 90-96, 1980.

[2] H. Rix, J.P. Malengé, Detection of an impurity (1%) at low resolution (0.25), *Journal of High Resolution Chromatography and Chromatography Communications*, Vol. 3, pp. 172-176, 1980.

[3] H. Rix, S. Boudaoud, O. Meste, Clustering Signal Shapes: application to P-Waves in ECG, Proc. *2nd European Medical & Biological Engineering Conference*, Vol.1, pp. 364-365, 2002.

[4] S. Boudaoud, H. Rix, J.J. Blanc, J.C., Cornily, O. Meste, Integrated Shape Averaging of the P-wave applied to AF risk detection, *Proc. Computers in Cardiology*, Vol. 30, pp. 125-128, 2003.

[5] H. Rix, Signal Shape Recognition in a Sum of Two Signals, **I3S/RR-2010-12-FR** , pp. 1-14, 2010