A delay estimation approach to change-point detection

Mamadou Mboup, Cédric Join, Michel Fliess

A DELAY ESTIMATION APPROACH TO CHANGE-POINT DETECTION

Mamadou Mboup

Projet ALIEN, INRIA Futurs
&
UFR Mathématiques et Informatique (CRIP5),
Université Paris Descartes,
45 rue des Saints-Pères, 75270 Paris cedex 06,
France
mboup@math-info.univ-paris5.fr

Cédric Join

Projet ALIEN, INRIA Futurs
&
CRAN (CNRS, UMR 7039),
Université Henri Poincaré (Nancy I), BP 239,
54506 Vandœuvre-lès-Nancy,
France
Cedric.Join@cran.uhp-nancy.fr

Michel Fliess

Projet ALIEN, INRIA Futurs
&
LIX (CNRS, UMR 7161),
École polytechnique,
91128 Palaiseau,
France
Michel.Fliess@polytechnique.edu

ABSTRACT

The change-point detection problem is cast into a delay estimation. Using a local piecewise polynomial representation and some elementary algebraic manipulations, we give an explicit characterization of a change-point as a solution of a given polynomial equation. A key feature of this polynomial equation is its coefficients being composed by short time window iterated integrals of the noisy signal. The so designed change-point detector shows good robustness to various type of noises.

Index Terms— Jump parameter systems, Signal detection, Delay estimation, Symbol manipulation, Polynomial approximation

1. INTRODUCTION

Given a piecewise regular signal, the purpose of this paper is to detect its discontinuities and estimate their locations. The problem is challenging especially for applications requiring on-line detection: the difficulties are stemming from corrupting noises which are blurring the discontinuities, and the combined need of fast calculations for real-time implementation and of reliable detection. A large amount of literature is devoted to these questions. We refer to the monograph of M. Basseville and V. Nikiforov [1] for a nice tutorial presentation and many application examples.

When a statistical model description of the observation data is available, then it is very common to approach the problem via hypothesis testing. And, the classical cumulative sum statistic [1] (see [2] for other statistics) is essential in most of the algorithms devised within this framework. Furthermore, since the wavelet transform coefficients of a signal inherently reveal the presence of irregularities, the combination wavelet–test statistic has now became classical for change point detection [3].

Model selection is another common approach and therein, the Bayesian theory plays a important rôle. An example using penalized contrast may be found in [4] (see also [5]). As opposed to the test statistic based methods which are sequential, the ones developed in this framework are rather global: the change points are simultaneously estimated. Clearly, this is not suited for on-line implementation. We may also mention other approaches such as the kernel based ones [6], [7]. These approaches are mainly local.

The solution we are presenting here is both sequential and local. It is based on a direct estimation of the points of singularity of the signal's derivatives. Following this spirit, one immediate solution would be to use a signal derivative estimator, as those proposed in [8], to locate the singularities. But, as we will shortly see, an explicit estimation of the derivative is not necessary. Using a local piecewise polynomial representation of the signal, we are able, in section 2, to cast the change-point problem into a delay estimation. The proposed detection algorithm is next presented in section 3. It's principle is to express a change-point as a solution of a polynomial equation, the coefficients of which are composed by short time window iterated integrals of the noisy signal. These integrals lowpass filter the noise as witnessed (testified) by the extensive simulation results presented in section 4. Different types of singularities are considered therein[1]: change-point in the mean, in the local slope, and a jump on the signal's second order derivative.

2. PROBLEM STATEMENT

2.1. Signal description

Consider the following piecewise smooth signal model,

$$x(t) = \sum_{i=1}^{K} H(t - t_{i-1})\, f_i(t - t_{i-1}), \tag{1}$$

[1]The first author would like to acknowledge M. Djafari (LSS, Gif-sur-Yvette, France) who kindly provided him with the test signals.

where $H(\cdot)$ is the Heaviside function and where each $f_i(t)$ is a smooth segment. The only irregularities of x, commonly called *change-points*, occur at times $t_i, i = 1, \dots, K$, where K is unknow. We set $t_0 = 0$. Based on the observation $y(t) = x(t)+n(t)$, where $n(t)$ is an additive noise corruption, we want to detect the change-points and estimate their locations t_i. Let T be given and assume that there is at most one discontinuity point in each interval $I_\tau^T = (\tau - T, \tau)$, $\tau \geqslant T$. In the sequel, we will set

$$x_\tau(t) = x(t + \tau - T), \quad t \in [0, T], \tau \geqslant T,$$

for the restriction of the signal in I_τ^T and we redefine the discontinuity point, say t_τ, relatively to I_τ^T with: $t_\tau = 0$ if $x_\tau(t)$ is smooth and $0 < t_\tau \leqslant T$ otherwise. The problem now reduces to detect the nonsmoothness of $x_\tau(t)$, for each $\tau \geqslant T$. For this task, we consider next a polynomial model for $x_\tau(t)$.

2.2. Local polynomial model

By assumption, each sliding interval I_τ^T contains at most one change point. The corresponding signal $x_\tau(t)$ then admits a representation of the form

$$x_\tau(t) = [1-H(t-t_\tau)]a(t)+H(t-t_\tau)b(t-t_\tau) \quad t \in [0, T], \tag{2}$$

where $a(t)$ and $b(t)$ are smooth and $x_\tau(t) = b(t)$ if $t_\tau = 0$. Here the possible change-point t_τ is viewed as a delay. Our detection method relies on the direct estimation of this delay. To proceed we need a model for $a(t)$ and $b(t)$ above. Now, a polynomial model is found adequate especially when the length of the interval I_τ^T, *viz.* T, is small. So henceforth, $a(t) = \sum_{i=0}^{p} a_i t^i$ and $b(t) = \sum_{j=0}^{q} b_j t^j$ are two polynomials of degree p and q respectively.

Let us express Eq. (2) in the operational domain[2]:

$$\hat{x}_\tau(s) = \sum_{i=0}^{p} \frac{i! a_i}{s^{i+1}} + e^{-t_\tau s} \sum_{j=0}^{q} \frac{j! b_j}{s^{j+1}} \tag{3}$$

By using elementary differential algebraic operations on this expression and translating the result back in the time domain, one can show that t_τ is identifiable with respect to $x(t)$. These operations, which lead to the design of an estimator of t_τ based on the noisy observation $y(t)$, are exposed below.

3. DETECTOR SYNTHESIS AND IMPLEMENTATION

3.1. Change-point estimation

We start considering the simplest model for $x_\tau(t)$, *viz.* a piecewise constant model. We thus have: $a(t) \equiv a_0$ and $b(t) \equiv b_0$ in (2). Eq. (3) then reduces to

$$\hat{x}_\tau(s) = \frac{a_0}{s} + e^{-t_\tau s} \frac{b_0}{s}. \tag{4}$$

In the sequel, we will write $\hat{x}_\tau$ instead of $\hat{x}_\tau(s)$ to ease the notations. Now, we are going to show that t_τ may be expressed as a function of $x_\tau(t)$ only: we will say that t_τ is identifiable with respect to $\hat{x}_\tau$ (see *e.g.* [10] for various types of identifiability). This will stem from the following manipulations. The first step is to eliminate the unknown coefficients a_0 and b_0. For this, multiply both side of (4) by s: $s\hat{x}_\tau - a_0 = e^{-t_\tau s} b_0$. Noting that $e^{-t_\tau s}$ satisfies $\left[\frac{d}{ds} + t_\tau\right] e^{-t_\tau s} = 0$, we have

$$\left[\frac{d}{ds} + t_\tau\right](s\hat{x}_\tau - a_0) = 0.$$

Differentiation with respect to s will eliminate a_0. The delay t_τ is therefore explicitly given by:

$$s\frac{d^2}{ds^2}\hat{x}_\tau + 2\frac{d}{ds}\hat{x}_\tau + t_\tau\left(s\frac{d}{ds}\hat{x}_\tau + \hat{x}_\tau\right) = 0. \tag{5}$$

Recall that, by the very classical rules of operational calculus, derivation with respect to s is the operational analogue of multiplication by $-t$: $\frac{d}{ds}\hat{x}_\tau \bullet\!\!-\!\!\circ -t x_\tau(t)$. And multiplication by s corresponds to time derivation: $s\hat{x}_\tau \bullet\!\!-\!\!\circ \frac{d}{dt}x_\tau(t)$. While the first operation is easy to implement, numerical differentiation is known to be difficult and ill-conditioned. Let ν be a positive integer greater than the highest power of s in (5) and divide both members of (5) by a s^ν. Then only negative powers of s will intervenes in the resulting equation. Replacing the unobserved signal x_τ by its noisy observation counterpart y_τ, we obtain the linear estimator $\widetilde{t_\tau}$ of t_τ:

$$\frac{\hat{y}_\tau''}{s^{\nu-1}} + 2\frac{\hat{y}_\tau'}{s^\nu} + \widetilde{t_\tau}\left\{\frac{\hat{y}_\tau'}{s^{\nu-1}} + \frac{\hat{y}_\tau}{s^\nu}\right\} = 0, \tag{6}$$

where $\hat{y}_\tau'$ stands for $\frac{d}{ds}\hat{y}_\tau$. Recall that division by s^k, $k > 0$ corresponds to k^{th}-iterated time integration:

$$(k-1)! s^{-k}\hat{x}(s) \bullet\!\!-\!\!\circ \int_0^t (t-\alpha)^{k-1} x(\alpha) d\alpha.$$

The estimate of t_τ thus follows upon expressing (6) back in time domain, for $\nu \geqslant 2$:

$$\left[\int_0^T (\nu t - T)(T-t)^{\nu-2} y_\tau(t) dt\right] \widetilde{t_\tau} = \int_0^T \{(\nu+1)t + 2T\}(T-t)^{\nu-2}\, t y_\tau(t) dt \tag{7}$$

If the interval I_τ^T is devoid of a change point, then the right hand side and the bracketed term above must be zero up to the (output) noise level. In this situation, we will have $\widetilde{t_\tau} = 0$. Note that noises are viewed highly as highly fluctuating phenomena (see [11] for the mathematical details). They are attenuated by the iterated time integrals, which are simple examples of low-pass filters.

[2]Algebraic manipulations of delays are easier via the formalism of operational calculus. See [9] for more details.

3.2. Synthesis of change point detector

The generalization of the preceding developments to higher order polynomial models is straightforward. To see this let us multiply both sides of (3) by s^ℓ where $\ell \triangleq \max\{p, q\} + 1$. Then we may write: $s^\ell \hat{x}_\tau = A(s) + e^{-t_\tau s} B(s)$ where $A(s)$ and $B(s)$ are two polynomials in s, with degree at most $\ell - 1$. Let us isolate $B(s)$ as in $B(s) = e^{t_\tau s}(s^\ell \hat{x}_\tau - A(s))$ and annihilate it by $\frac{d^\ell}{ds^\ell}$. We obtain:

$$\sum_{i=0}^{\ell} \left\{ \binom{\ell}{i} \frac{d^{\ell-i}}{ds^{\ell-i}} (s^\ell \hat{x}_\tau - A(s)) \right\} t_\tau^i = 0.$$

Now, the unknown polynomial $A(s)$ has to be eliminated. This is achieved by applying again $\frac{d^\ell}{ds^\ell}$ to the above expression. The result is next divided by s^ν, $\nu > \ell$, to avoid time differentiation. Replacing $\hat{x}_\tau$ by its corresponding noisy observation $\hat{y}_\tau$ and returning to the time domain, we finally obtain, for each time instant τ, a polynomial expression $\mathcal{D}_{\{p\}\{q\}}(\tau, t_\tau)$ of the form:

$$\mathcal{D}_{\{p\}\{q\}}(\tau, t_\tau) = \int_0^T y(\tau - (T - t)) \sum_{i=0}^{\ell} P_i(t)\, t_\tau^i \, dt. \quad (8)$$

Each $P_i(t)$ in (8) is a polynomial depending on the parameters p, q, ν and T and is defined as:

$$P_i(t) = \sum_{j=0}^{\ell} \binom{\ell}{i} \binom{2\ell - i}{j} \frac{(-1)^{2\ell-i-j} \ell!}{(\ell - j)!(j + \nu - \ell - 1)!} (T - t)^{j+\nu-\ell-1} (\tau - (T - t))^{2\ell-i-j}$$

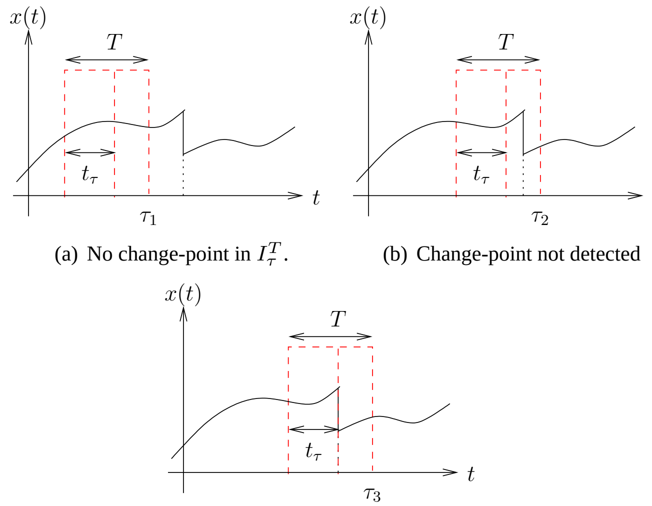
Clearly, this expression does not provide an estimator for t_τ. However, it does provide us with a change-point detector, the implementation of which is described below.

3.3. Implementation

To see how to implement $\mathcal{D}_{\{p\}\{q\}}(\tau, t_\tau)$ as a detector, let us set t_τ to a fixed value, say $t_\tau = rT$ for some $r \in (0,\ 1)$. Consider the following scenarios:

- Case 1: I_τ^T does not contain a change-point. Then, the equality $\mathcal{D}_{\{p\}\{q\}}(\tau, rT) = 0$ holds except for the local mismodelling error and the noise contributions. This situation is depicted in figure 1-(a).
- Case 2: I_τ^T contains a change-point, located at $t_\tau^\star \neq rt$. In this case, $\mathcal{D}_{\{p\}\{q\}}(\tau, rT)$ largely deviates from zero. The selected value for t_τ does not correspond to a root of the polynomial in (8), as shown in figure 1-(b).
- Case 3: I_τ^T contains a change-point at $t_\tau^\star$ which coincides with rT. Then $\mathcal{D}_{\{p\}\{q\}}(\tau, rT)$ must vanishes up to the noise and the local mismodelling errors. Figure 1-(c) illustrates this situation.

Remark 1 *Choosing $t_\tau = 0.5T$ seems to be the best choice especially for the detection of very close change points.*



(a) No change-point in I_τ^T. (b) Change-point not detected

(c) Change-point detected

Fig. 1. Detector

According to these scenarios, a change-point is meant by the passage from case 2 to case 3, *i.e.* when starting beyong a threshold, $|\mathcal{D}_{\{p\}\{q\}}(\tau, t_\tau)|$ falls down and crosses zero. Since the characteristics of the noise are unknown, the selection of the threshold will subsequently be based on some heuristics. Also, the performance of the proposed detection method are investigated in the next section, through Monte Carlo simulation.

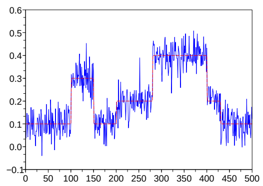
4. SIMULATION RESULTS

The change-point detector in (8) is now numerically implemented using the trapezoidal method, under various signal and noise settings. Very simple local polynomial models are used for all signal settings. Robustness to noise corruption is highlighted using several noise models with different powers: normal, uniform and Perlin noise[3]. Table 1, where the notation $\mathcal{M}_{\{p\}\{q\}}$ refers to (3) with degrees p and q, summarizes the simulation results. For all simulations, the sampling period (without unit) is $T_e = 1$ and the detector is inhibited during $10T_e$ (10 samples) after each detection. For each setting, the distribution of the estimated change-points, computed over 100 Monte Carlo simulations, are shown below. We start with a piecewise constant signal. The original noise-free signal (in dashed line figure 2-(a) and 2-(c)) is composed of 7 different segments, with two very close jumps separated by 30 samples. The signal is corrupted by a white Gaussian noise in figure 2-(a) and by a Perlin noise in 2-(c). The results in figures 2-(b) and (d) show how all the change-points are correctly detected for both noise settings. They illustrate the ability of the method to detect separately very close change-points. In the next experiment, we consider a

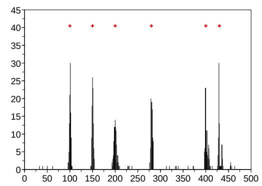
[3] *Perlin's noises* [12] are fractal like and they are quite popular in computer graphics.

Figures reference	Signal model see (3)	T in T_e	Noise type	SNR in dB	Estimated number of segments								
					1	2	3	4	5	6	7	8	≥9
figure 2-(a)-(b)	$\mathcal{M}_{\{0\}\{0\}}$	60	Normal	13	0	0	0	0	0	28	**40**	26	6
figure 2-(b)-(c)	$\mathcal{M}_{\{0\}\{0\}}$	60	Perlin	15	0	0	0	0	4	24	**36**	25	11
figure 3	$\mathcal{M}_{\{1\}\{1\}}$	60	Uniform	16	0	0	0	2	9	25	**27**	24	13
figure 4	$\mathcal{M}_{\{1\}\{0\}}$	30	Normal	10	0	45	**31**	3	6	3	0	0	1

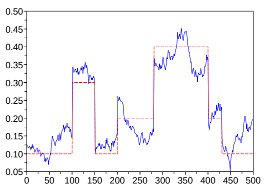
Table 1. Simulation summarized



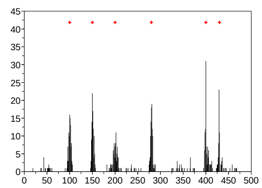
(a) Noisy free signal (- -), signal (–)



(b) Histogram of the detection



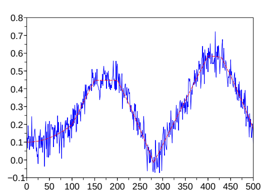
(c) Noisy free signal (- -), signal (–)



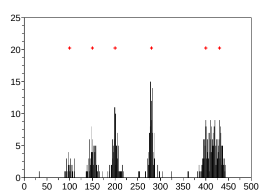
(d) Histogram of the detection

Fig. 2. Piecewise constant signal: changes in the mean

piecewise polynomial signal with degrees 0 to 2, corrupted by a uniform distributed noise. The signal is continuous and it contains 6 change-points. These are of order 1 (discontinuity on the derivative) or order 2 (discontinuity on the second order derivative) and hence more difficult to detect. The obtained results, depicted in Figure 3 remain good. The two close change-points, at $t = 400T_e$ and $t = 430T_e$, are however difficult to separate.



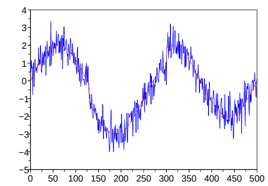
(a) Noisy free signal (- -), signal (–)



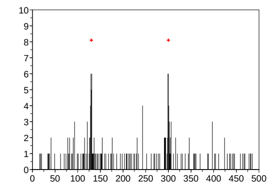
(b) Histogram of the detection

Fig. 3. Piecewise polynomial signal: order 1 and 2.

In view of the preceding experiments, the method presents a good robustness to noise. It also appears that the noise characterisctics do not have a significant effect on the detection. We consider now a piecewise non polynomial signal with 3 segments, through an additive white Gaussian noise (see figure 4-(a)). The results illustrated in figure 4-(b) show that a very simple local model (order 0 or 1) is sufficient. Indeed, for any fixed orders, the local mismodelling error can be made arbitrary small by decreasing T.



(a) Noisy free signal (- -), signal (–)



(b) Histogram of the detection

Fig. 4. Non polynomial signal

5. REFERENCES

[1] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall, 1993.

[2] N. Vaswani, "Additive change detection in nonlinear systems with unknown change parameters," *IEEE Trans. Signal Processing*, vol. 55, no. 3, 2007.

[3] M. Raimondo and N. Tajvidi, "A peaks over threshhold model for change-point detection by wavelets," *Statistica Sinica*, vol. 14, pp. 395–412, 2004.

[4] M. Lavielle, "Using penalized contrasts for change-point problem," *Signal Processing*, vol. 85, pp. 1501–1510, 2005.

[5] T-H. Fan, K-C. Chang, and C-B Lee, "Bayesian estimation of the number of change points in simple linear regression models," *Communications in Statistics–Theory and Methods*, vol. 35, pp. 689–71, 2006.

[6] I. Gijbels, P. Hall, and A. Kneip, "On the estimation of jump points in smooth curves," *Ann. Inst. Statist. Math.*, vol. 51, pp. 231–251, 1999.

[7] F. Désobry, M. Davy, and C. Doncarli, "An online kernel change detection algorithm," *IEEE Trans Signal Processing*, vol. 53, no. 8, pp. 2961–2974, 2005.

[8] M. Mboup, C. Join, and M. Fliess, "A revised look at numerical differentiation with an application to nonlinear feedback control," in *MED'07*, Athens, Greece, 2007.

[9] M. Fliess, C. Join, and M. Mboup, "Algebraic change point detection," *Applicable Algebra in Eng., Communication and Computing*, to appear.

[10] M. Fliess and H. Sira-Ramírez, "An algebraic framework for linear identification," in *ESAIM: COCV*, vol. 9. SMAI, 2003, URL: http://www.emath.fr/cocv.

[11] M. Fliess, "Analyse non standard du bruit," *CRAS, Série 1, Mathématiques*, vol. 342, pp. 797–802, 2006.

[12] K. Perlin, "An image synthesizer," in *SIGGRAPH'85*, 1985, pp. 287–296.