

Detection of multiple changes in mean by sparse parameter estimation*

Jiří Neubauer^a, Vítězslav Veselý^b

^aUniversity of Defence
Kounicova 65, 61200 Brno, Czech Republic
jiri.neubauer@unob.cz

^bMasaryk University
Lipová 41a, 60200 Brno, Czech Republic
vesely@econ.muni.cz

Received: 13 February 2012 / **Revised:** 21 February 2013 / **Published online:** 18 April 2013

Abstract. The contribution is focused on detection of multiple changes in the mean in a one-dimensional stochastic process by sparse parameter estimation from an overparametrized model. The authors' approach to change point detection differs entirely from standard statistical techniques. A stochastic process residing in a bounded interval with changes in the mean is estimated using dictionary (a family of functions, the so-called atoms, which are overcomplete in the sense of being nearly linearly dependent) and consisting of Heaviside functions. Among all possible representations of the process we want to find a sparse one utilizing a significantly reduced number of atoms. This problem can be solved by ℓ_1 -minimization. The basis pursuit algorithm is used to get sparse parameter estimates. In this contribution the authors calculate empirical probability of successful change point detection as a function depending on the number of change points and the level of standard deviation of additive white noise of the stochastic process. The empirical probability was computed by simulations where locations of change points were chosen randomly from uniform distribution. The authors' approach is compared with LASSO algorithm, ℓ_1 trend filtering and selected statistical methods. Such probability decreases with increasing number of change points and/or standard deviation of white noise. The proposed method was applied on the time series of nuclear magnetic response during the drilling of a well.

Keywords: multiple change point detection, sparse parameter estimation, basis pursuit denoising, LASSO, ℓ_1 trend filtering.

1 Introduction

Detection of changes in the mean helps us to discover time points at which some intrinsic properties of a stochastic process change. This covers a broad range of real-world problems and has been actively discussed. The standard statistical approach to this problem

*This research was supported by the grant GAČR P402/10/P209 and FEM Development Projects Economic Laboratory.

is described, for example, in [1]. The article is focused on detection of multiple changes in the mean in a one-dimensional stochastic process using sparse parameter estimation from an overparametrized model. Sparse modeling is a quickly developing area on the intersection of statistics, machine-learning and signal processing. It can be effectively used in the case of large data sets fitting when standard numerical methods are unstable. Several methods of large-scale data analysis can be found, for example, in [2]. Detection of changes has originally arisen in the context of quality control. Nowadays, we can find a wide range of fields where change point problem is applied, such as epidemiology, medicine (rhythm analysis), ecology, signal processing etc. When analyzing measurements coming from real problems often some events or anomalies are manifested by abrupt changes which may not be quite apparent in their early stage, detection of hearing loss in audiological frequencies [3] being an example out of many illustrating such a situation. Detection of changes in processes need not involve only changes in the mean, but also changes in the variance or other characteristics. For example, the article [4] deals with detection of changed segment in a binomial sequence.

The authors' approach to change point detection is quite different from the standard statistical techniques. The statistical approaches usually use for change point estimation the maximum likelihood method or non-parametric methods, such as methods based on M and R estimates (see [1,5]). In the case of uncorrelated white noise in the model and one change in the mean, the authors Neubauer and Veselý [6] showed that there is no significant difference among these methods. The proposed method of the change point detection is based on entirely different methods, built on a completely different approach. It is a very versatile approach to signal modeling, assuming that the signal is sparse; therefore, it can be described by a small number of suitable functions. A stochastic process residing in a bounded interval with changes in the mean is estimated using dictionary (a family of functions, the so-called atoms) and consisting of Heaviside functions. Among all the possible representations of the process we want to find a sparse one utilizing a significantly reduced number of atoms. This problem can be solved by ℓ_1 -minimization (see [8,9]) or by some of the greedy algorithms which are compared to the ℓ_1 -minimization in [10]. The basis pursuit algorithm is used to get sparse parameter estimates. Basic properties of this approach to one change point detection were studied in [6] by simulations and included a performance comparative study with standard statistical procedures as well. The impact of serial correlation of noise on change point detection is described in [7]. The contribution [11] deals primarily with detection of two change points. According to these results, the basis pursuit approach proposes an alternative technique of the change point detection. In this article the authors calculate empirical probability of successful change point detection as a function depending on the number of change points and the level of standard deviation of additive white noise of the stochastic process. The empirical probability was computed by simulations where locations of change points were chosen randomly from uniform distribution. As an alternative to the basis pursuit algorithm, LASSO method [12], ℓ_1 trend filtering [13] and selected statistical method were taken to judge the proposed method. The last part of the article demonstrates how the aforementioned methods can automate the change point detection in the time series of nuclear magnetic response during the drilling of a well.

2 Basis pursuit and change point detection

In this paragraph we briefly describe the method based on basis pursuit algorithm (BPA) for the detection of the change point in the sample path $\{y_t\}$ in one-dimensional stochastic process $\{Y_t\}$. We assume a deterministic functional model on a bounded interval \mathcal{I} described by the dictionary $G = \{G_j\}_{j \in J}$ with atoms $G_j \in L^2(\mathcal{I})$ and with additive white noise e on a suitable finite discrete mesh $\mathcal{T} \subset \mathcal{I}$:

$$Y_t = x_t + e_t, \quad t \in \mathcal{T},$$

where $x \in \text{sp}(\{G_j\}_{j \in J})$, $e_t \sim WN(0, \sigma^2)$, $t \in \mathcal{T}$, $\sigma > 0$, and J is a big finite indexing set. Smoothed function $\hat{x} = \sum_{j \in J} \hat{\xi}_j G_j =: \mathbf{G}\hat{\xi}$ minimizes on \mathcal{T} ℓ_1 -penalized optimality measure $(1/2)\|y - \mathbf{G}\xi\|_2^2$ as follows:

$$\hat{\xi} = \arg \min_{\xi \in \ell^2(J)} \frac{1}{2} \|y - \mathbf{G}\xi\|_2^2 + \lambda \|\xi\|_1, \quad \|\xi\|_1 := \sum_{j \in J} \|G_j\|_2 |\xi_j|,$$

where $\lambda = \sigma \sqrt{2 \ln(\text{card}J)}$ is a smoothing parameter chosen according to the soft-thresholding rule commonly used in wavelet theory. This choice is natural because one can prove that with any orthonormal basis $G = \{G_j\}_{j \in J}$ the shrinkage via soft-thresholding produces the same smoothing result \hat{x} (see [9]). Such approaches are also known as basis pursuit denoising (BPDN).

Solution of this minimization problem with λ close to zero may not be sparse enough: we are searching small $F \subset J$ such that $\hat{x} \approx \sum_{j \in F} \hat{\xi}_j G_j$ is a good approximation. The procedure of BPDN is described in [6].

We build our dictionary from heaviside-shaped atoms on $L^2(\mathbb{R})$ derived from a fixed “mother function” via shifting and scaling following the analogy with the construction of wavelet bases.

We construct an oversized shift-scale dictionary $G = \{G_{a,b}\}_{a \in \mathcal{A}, b \in \mathcal{B}}$ derived from the “mother function” by varying the shift parameter a and the scale (width) parameter b between values from big finite sets $\mathcal{A} \subset \mathbb{R}$ and $\mathcal{B} \subset \mathbb{R}^+$, respectively ($J = \mathcal{A} \times \mathcal{B}$), on a bounded interval $\mathcal{I} \subset \mathbb{R}$ spanning the space $H = \text{sp}(\{G_{a,b}\}_{a \in \mathcal{A}, b \in \mathcal{B}})$, where

$$G_{a,b}(t) = \begin{cases} 1 & \text{for } t - a > b/2, \\ 2(t - a)/b & \text{for } |t - a| \leq b/2, \ b > 0, \\ 0 & \text{for } t = a, \ b = 0, \\ -1 & \text{otherwise.} \end{cases}$$

In the simulations below $\mathcal{I} = [0, 1]$, $\mathcal{T} = \{t/T\}$ (typically with mesh size $T = 100$), $\mathcal{A} = \{t/T\}_{t=t_0}^{T-t_0}$ (t_0 is a boundary trimming, $t_0 = 5$ was used in the simulations) and scale b fixed to zero ($\mathcal{B} = \{0\}$). Clearly the atoms of such Heaviside dictionary are normalized on \mathcal{I} , i.e. $\|G_{a,0}\|_2 = 1$. Some examples of Heaviside functions are displayed in Fig. 1.

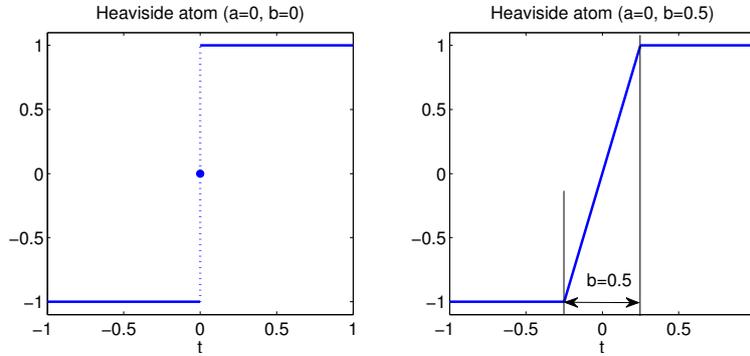


Fig. 1. Heaviside atoms with parameters $a = 0$, $b = 0$ and $a = 0$, $b = 0.5$.

Neubauer and Veselý [6] proposed the method of change point detection if there is just one change point in a one-dimensional stochastic process (or in its sample path). We briefly describe this method. We would like to find a change point in a stochastic process

$$Y_t = \begin{cases} \mu + \epsilon_t & \text{for } t = 1, 2, \dots, c, \\ \mu + \delta + \epsilon_t & \text{for } t = c + 1, \dots, T, \end{cases} \quad (1)$$

where $\mu, \delta \neq 0$, $t_0 \leq c < T - t_0$ are unknown parameters and ϵ_t are independent identically distributed random variables with zero mean and variance σ^2 . The parameter c indicates the change point in the process. Using the basis pursuit algorithm we obtain some significant atoms, we calculate the correlation between the significant atoms and the analyzed process. As a BPDN estimator of the change point c we take the shift parameter of the significant atom matching one of three alternative rules: the atom is the most significant one, it exhibits the highest correlation with the data or it produces the highest value of geometric mean of the correlation coefficient and its contribution to the total ℓ_1 norm. In the case of one change point, the estimator based on the highest correlation performed slightly better.

Now let us assume the model with p change points

$$Y_t = \begin{cases} \mu + \epsilon_t & \text{for } t = 1, 2, \dots, c_1, \\ \mu + \delta_1 + \epsilon_t & \text{for } t = c_1 + 1, \dots, c_2, \\ \dots & \\ \mu + \delta_p + \epsilon_t & \text{for } t = c_p + 1, \dots, T, \end{cases} \quad (2)$$

where $\mu, \delta_1, \dots, \delta_p \neq 0$, $t_0 \leq c_1 < \dots < c_p < T - t_0$ are unknown parameters and ϵ_t are independent identically distributed random variables with zero mean and variance σ^2 .

We use the method of change point estimation described above for detection of p change points c_1, \dots, c_p in model (2). Instead of finding only one significant atom we identify either p most significant atoms, or p significant atoms exhibiting the highest correlation, or p significant atoms having the highest value of geometric mean of the

correlation coefficient and the contribution to the total ℓ_1 norm. The shift parameters of such atoms determine *direct* BPDN estimators for the change points c_1, \dots, c_p . Another possibility is to apply the procedure of one change point detection p -times in sequence (*sequential* BPDN estimator). In the first step we identify one change point in the process Y_t , then we subtract the contribution of such a significant atom from the process (by linear regression)

$$\begin{aligned} Y_t &= \gamma G_{0, \hat{c}_1} + e_t, \\ Y'_t &= Y_t - \hat{\gamma} G_{0, \hat{c}_1}, \end{aligned}$$

and we apply the same method to the new process Y'_t . This sequence is repeated until we get p change point estimates. The shift parameters of selected atoms are again identifiers of the change points c_1, \dots, c_p .

Regarding the theoretical background, we consider three well-known criteria.

The so-called *spark* of a matrix (dictionary) G is the minimum number of columns of G which are linearly dependent. It can be shown (see [14]) that if a solution x of our inverse problem is found which satisfies $\|x\|_0 < \text{spark}(G)/2$ then x is the sparsest and unique solution to this problem. However, in general, establishing spark is an NP-hard problem so in practice this is intractable. Nevertheless, due to a clear structure of our dictionary G , we know that its spark is full, i.e. 101 in our case. Thus, if a solution x would be found with $\|x\|_0 \leq 45$ then we have the true solution. This actually motivates to exploit the sparsity-seeking approach. However, one must be aware of two major facts: First, this result only holds in case without noise and the noisy case is not solved in the literature yet. Second, the statement does not tell anything about how the solution should be obtained!

Another measure of goodness of the dictionary is the mutual coherence (see [8, 14]) which is defined pairwise as $\mu(G) = \max_{i \neq j} g_i^T g_j$ for ℓ_2 -normalized columns. This value can be determined computationally – in our case it is 0.99 which means that the atoms are very close to each other. The relationship between μ and spark is known and thanks to that there exists a theorem saying that if a solution x is found satisfying $\|x\|_0 < (1/2)(1 + 1/\mu(G))$ then it is the unique one. But in our case the value of the right side only 1.0051 which means that this would be true only if the sparsity of the solution would be 1 (without noise, again). In our case of multiple change-points this is contradictory, so the coherence does not help much here.

The third property of a dictionary is the RIP (restricted isometry property, see [15]) which is again not possible to compute in polynomial time, but can be used to specify a condition under which the solution can be found via ℓ_1 -minimization (which we use). We found numerically that the RIP constant δ_k which could be found in a combinatorial way even for such low values as $k = 2$ and $k = 3$ is in the order of 10^1 ; RIP requires $\delta_k < 1$. However this is not a surprise, since the only matrices which are known to obey the RIP are certain classes of random matrices, see [8, 15].

To conclude, no one of the state-of-the art quantities measuring the goodness of the dictionary really helps, so we rely on the simulation results (which is in fact not uncommon in the sparse-land publications area), see [16, 17].

3 The LASSO

The LASSO (least absolute selection and shrinkage operator) is a shrinkage and selection method for linear regression. It minimizes the sum of squared errors, with a bound on the sum of the absolute values of the coefficients. Using the usual linear regression notation the LASSO estimate is defined in [12] as follows:

$$\hat{\beta}^{\text{LASSO}} = \arg \min_{\beta} \sum_{i=1}^T \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 \quad \text{subject to} \quad \sum_{j=1}^p |\beta_j| \leq t.$$

Rewriting this in terms of operators and language of atoms and dictionaries, we obtain a denoising approach similar to basis pursuit.

4 ℓ_1 trend filtering

In [13] the authors propose a variation on Hodrick–Prescott filtering, which they call ℓ_1 trend filtering. The trend is estimated as the minimizer of the objective function

$$\frac{1}{2} \sum_{t=1}^T (y_t - x_t)^2 + \chi \sum_{t=2}^{T-1} |x_{t-1} - 2x_t + x_{t+1}|, \quad (3)$$

which can be written in the matrix form as

$$\frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \chi \|\mathbf{D}\mathbf{x}\|_1,$$

where $\chi \geq 0$ is a smoothing parameter, $\mathbf{x} = (x_1, \dots, x_T)' \in \mathbb{R}^T$, $\mathbf{y} = (y_1, \dots, y_n)' \in \mathbb{R}^T$ and $\mathbf{D} \in \mathbb{R}^{(T-2) \times T}$ is the matrix

$$\mathbf{D} = \begin{pmatrix} 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & -2 & 1 \end{pmatrix}.$$

The ℓ_1 trend estimate is piecewise linear in t . The points where the slope of the estimated trend is changed can be interpreted as abrupt changes (change points) in the process. Model (2) describes only changes in the mean. We adapt the proposed method for detection change point of this kind. The argument appearing in the second term of (3), $x_{t-1} - 2x_t + x_{t+1}$, is the second difference of $\{x_t\}$. It is zero when and only when three points x_{t-1} , $2x_t$ and x_{t+1} are on the line. In the case of the piecewise constant trend, we replace the second difference with the first difference $x_t - x_{t-1}$. We will minimize the function

$$\frac{1}{2} \sum_{t=1}^T (y_t - x_t)^2 + \chi \sum_{t=2}^T |x_t - x_{t-1}|, \quad (4)$$

in the matrix form $(1/2)\|\mathbf{y} - \mathbf{x}\|_2^2 + \chi\|\mathbf{D}\mathbf{x}\|_1$, where $\mathbf{D} \in \mathbb{R}^{(T-1) \times T}$ is the matrix

$$\mathbf{D} = \begin{pmatrix} 1 & -1 & & & & \\ & 1 & -1 & & & \\ & & \ddots & \ddots & & \\ & & & 1 & -1 & \\ & & & & 1 & -1 \end{pmatrix}.$$

The estimated trend is piecewise constant. The points where the estimated trend is changed can be taken as estimates of change points.

5 Statistical methods

Consider model (1). Assuming σ^2 given, the unknown parameters c , μ and δ may be estimated by the least-squares method. The least-squares estimators \hat{c} , $\hat{\mu}$ and $\hat{\delta}$ of the parameters c , μ and δ are defined as solutions of the minimization problem

$$\min \left\{ \sum_{t=1}^k (Y_t - \mu)^2 + \sum_{t=k+1}^T (Y_t - \mu - \delta)^2; k \in \{1, \dots, T-1\}, \mu \in \mathbb{R}, \delta \in \mathbb{R} \right\}.$$

In other words, the unknown parameters are estimated in such a way that the sum of squares of residuals is minimal. The estimates of the parameters μ and δ are (see [5] etc.)

$$\hat{\mu} = \bar{Y}_{\hat{c}} \quad \text{and} \quad \hat{\delta} = \bar{Y}_{\hat{c}}^0 - \bar{Y}_{\hat{c}},$$

where \hat{c} is a solution of the maximization problem

$$\hat{c} = \arg \max \left\{ \sqrt{\frac{T}{k(T-k)}} |S_k|; k \in \{1, \dots, T-1\} \right\}, \tag{5}$$

where we denote $S_k = \sum_{t=1}^k (Y_t - \bar{Y}_T)$, $\bar{Y}_T = (1/T) \sum_{t=1}^T Y_t$, $\bar{Y}_{\hat{c}} = (1/\hat{c}) \sum_{t=1}^{\hat{c}} Y_t$ and $\bar{Y}_{\hat{c}}^0 = 1/(T - \hat{c}) \sum_{t=\hat{c}+1}^T Y_t$.

In the case of multiple change points, we can use the described statistical method of one change point detection via the following procedure. At first, find $\hat{c}^{(1)}$ by solving (5). Secondly, divide observations into two groups $Y_1, \dots, Y_{\hat{c}^{(1)}}$ and $Y_{\hat{c}^{(1)}+1}, \dots, Y_T$ and find the estimator in each group. The whole procedure is repeated until a “constant mean is obtained”. This procedure is called “binary segmentation”.

The statistical computing environment R, package “changepoint”, offers several methods for multiple change point detection. For the purpose of comparison, we will employ two basic approaches: detection of multiple change points using binary segmentation for normally distributed data and the method based on cumulative sums with the use of binary segmentation (see [1, 18–20]).

6 Simulation study

For the purpose of performance study of the proposed method of multiple change point detection we use simulations of process (2). We put $T = 100$, $\mu = 0$, the error terms are independent normally distributed with zero mean and the standard deviations $\sigma = 0.1, 0.2, 0.5$ and 1 . Locations of change points were chosen at random uniformly in the interval $[5, 95]$, jumps δ_i ($i = 1, \dots, p$) taking heights ± 1 (+ or $-$ at random with equal probability). The minimum distance between neighboring change points is 5 . For each number of change points out of $p = 1, \dots, 15$ we calculated 500 simulations of process (2). A uniform distribution was used in the simulations to generate random position of change points. In terms of change point detection, the change points are distributed within intervals of equal length with the same probability. Thus, there is no preferred position of change points. If we choose, for example, a normal distribution with the mean in the middle of the interval, position change points are more likely close to this center. As was shown in [6], detection of change points located in the middle of the process is more successful than in the case that this jump is near the beginning or end of the process. From this perspective, the choice of the uniform distribution is the most appropriate.

We applied ten methods to detect change points:

- the direct BPDN estimator based on p most significant atoms;
- the direct BPDN estimator based on p significant atoms exhibiting the highest correlation with the data;
- the direct BPDN estimator based on the highest geometric mean of the correlation coefficient and the contribution to the total ℓ_1 norm;
- the sequential BPDN estimator based on the most significant atom;
- the sequential BPDN estimator based on the significant atom with the highest correlation;
- the sequential BPDN estimator based on the highest geometric mean of correlation coefficient and the contribution to the total ℓ_1 norm;
- the LASSO algorithm;
- the adapted ℓ_1 trend filtering;
- the estimator for normally distributed data using the binary segmentation method;
- the cumulative sums estimator using the binary segmentation method.

We calculate empirical probability of successful change point detection as

$$\text{empirical probability} = \frac{\text{number of successful detections}}{\text{number of simulations}}.$$

The detection was a success provided that all p change points could be correctly identified. The results are summarized by plots in Fig. 2.

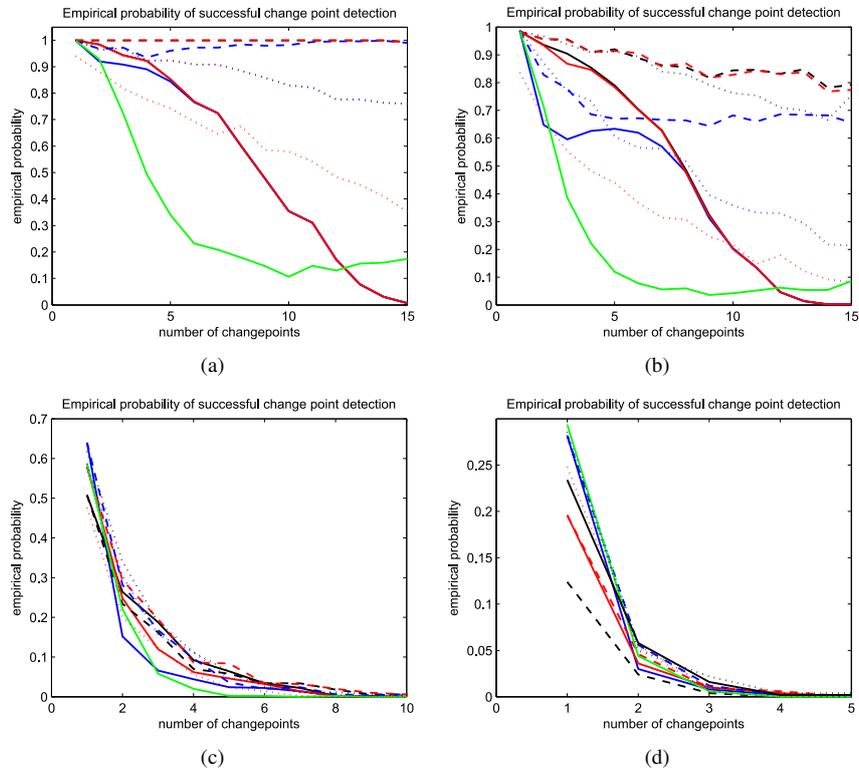


Fig. 2. Empirical probability of successful change point detection for: (a) $\sigma = 0.1$, (b) $\sigma = 0.2$, (c) $\sigma = 0.5$, (d) $\sigma = 1$. BPDN – black solid, BPDN (the highest correlation) – blue solid, BPDN (geometric mean) – red solid, sequential BPDN – black dashed, sequential BPDN (the highest correlation) – blue dashed, sequential BPDN (geometric mean) – red dashed, LASSO – green, adapted ℓ_1 trend filtering – black dotted, normal data with binary segmentation – blue dotted, cumulative sums with binary segmentation – red dotted. (Online version in colour.)

The empirical probability of the sequential methods is generally higher than the probability of the direct detection methods. We assumed that the difference between two neighbouring atoms is at least 5. In the case of sequential methods we exclude all atoms detected in preceding iterations.

Basis pursuit denoising estimates were computed in Matlab utilizing [21] and four-step modification of BPDN (BPA4) from the toolbox [22]. A detailed description of the BPA4 algorithm can be found in [23]. The LASSO estimates were calculated in R using the package “lars”. The algorithm of least angle regression is used to get LASSO estimates [24]. For the purpose of ℓ_1 trend filtering, the Matlab code written by Koh et al. [25] was employed. The smoothing parameter χ was for $\sigma = 0.1$ and 0.2 set to 1, for $\sigma = 0.5$ to 2 and for $\sigma = 1$ to 5. The package “change point” in the statistical environment R was used to obtain the last two estimators.

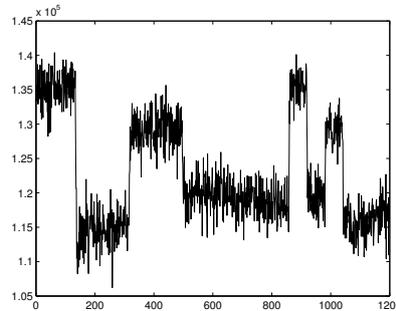


Fig. 3. Nuclear magnetic response during the drilling of a well.

7 Real data example

In this part of the article we will apply all of the aforementioned methods of change point detection to time series comprising measurements of nuclear magnetic response made during the drilling of a well (see [26,27]). The changes should correspond to the transition between the different strata of rock. We will analyze a subset of this dataset containing 1200 measurements (from the index 1551 to 2750). Table 1 contains the results obtained

Table 1. The results of change point detection for time series of nuclear magnetic response during the drilling of a well.

Method	Change points estimates							
BPDN	135	859	1041	919	982	316	498	59
(sorted)	59	135	316	498	859	919	982	1041
BPDN – corr.	134	59	1041	498	920	984	316	856
(sorted)	59	134	316	498	856	920	984	1041
BPDN – geom. mean	135	1041	919	982	59	498	316	859
(sorted)	59	135	316	498	859	919	982	1041
BPDN – seq.	135	859	141	1041	918	316	498	981
(sorted)	135	141	316	498	859	918	981	1041
BPDN – corr seq.	134	316	1041	498	858	919	981	126
(sorted)	126	134	316	498	858	919	981	1041
BPDN – geom. mean seq.	135	1041	316	859	498	920	981	145
(sorted)	135	145	316	498	859	920	981	1041
LASSO	135	1043	317	499	859	921	504	1061
(sorted)	135	317	499	504	859	921	1043	1061
adapted ℓ_1 trend filtering	1041	859	316	134	919	981	498	322
(sorted)	134	316	322	498	859	919	981	1041
Normal data	134	316	496	1042	858	919	981	676
(sorted)	134	316	496	676	858	919	981	1042
CUSUM	135	316	498	1041	858	919	981	676
(sorted)	135	316	498	676	858	919	981	1041

by 10 described methods, where 8 change points were estimated. For each method there is listed a sequence of change point estimates in the order they were detected, then these estimates are sorted in ascending order for clarity. Seven possible position of change points were detected (in the range 134–135, 316–317, 496–499, 919–920, 981–982 and 1041–1042). Only the method based on the LASSO algorithm did not find the change in 981–982. The estimates of the eighth change point whose position is not apparent differs for each method. Similarly to the simulation study, we assumed that the difference between two neighboring change point estimates was at least 5. The smoothing parameter χ in adapted ℓ_1 trend filtering was set to 10^5 .

8 Conclusion

According to the simulation results the basis pursuit approach proposes a reasonable detection method of change points in one-dimensional process. In the case of one change point, the method based on significant atoms with the highest correlation with the process yields better results than the method based on pure ℓ_1 minimization when the standard deviation of the white noise is large. This is not valid for more than one change point which can be drawn from the graphs of the empirical probability of successful change point detection. It is apparent from the graphs that the sequential methods performed better than the direct ones. The LASSO algorithm gives similar results for one or two change points compared to non-sequential approaches when the level of the noise is small. For more than two change points the LASSO algorithm offers lower empirical probability of successful change point detection. For larger standard deviation of the noise and small number of change points the LASSO gives a comparable result as the method based on the correlation coefficient. As we can deduce from the simulations, the adapted ℓ_1 filtering gives almost comparable results with the sequential methods (except the method based on the highest correlation). However, for the standard deviation 0.2 and high number of change points the empirical probability of successful detection of the two proposed methods is higher than for adapted ℓ_1 filtering. If one wants to compare the aforementioned statistical methods, it can be said according to the calculated empirical probability, the estimators using the cumulative sums perform worse than the other one. The sequential methods give, in most cases, better results than the statistical approaches. In the case of large standard deviation and higher number of change points the differences among the proposed methods nearly vanish.

In general, with the increasing number of change points the empirical probability decreases as one can expect. The standard deviation of an additive white noise of the process also affects the probability of successful detection. Higher values of the standard deviation cause lower probability of successful detection.

As to the computational speed it should be noted that the LASSO algorithm, the adapted ℓ_1 trend filtering and the algorithms for statistical estimates are significantly faster than the algorithm BPA4 used for the basis pursuit denoising. In addition to the problem of change point detection, the described four-step procedure using basis pursuit algorithm can solve a number of problems of various types, such as signal processing [9], kernel

smoothing [23], and ARMA models fitting [28]. This 4-step version of basis pursuit algorithm usually gives better results than pure basis pursuit approach. The price for this refinement is the increase in computation time. The aim of the article was not to monitor the speed of algorithms but to focus on the ability of each method to correctly detect potential changes. The dictionary created from Heaviside functions (atoms) can be merged with another set of functions, i.e. Fourier dictionary (see [9]) or Gaussian functions. The resulting dictionary can be applied to smooth the signal (the process) containing abrupt changes (jumps).

The change point detection techniques may be useful in a broad range of real-world problems, for instance in the modeling of economical or environmental time series where jumps can occur.

In general, the scope of sparse modeling based on basis pursuit or other related algorithms [29] goes far beyond the change-point problem.

It is because two factors are crucial for successful modeling: *choice of an adequate model* and a *numerically stable estimation procedure* yielding reliable parameter estimates. Unfortunately the ideas about a choice of an appropriate model are often very vague and produce models where it is hard to balance the requirement on sufficient regularity of the model (as few parameters as possible to guarantee numerical stability) and feasible precision which forces the analyst to increase the number of model components typically leading to overparameterization accompanied with non-uniqueness and numerical instability of solutions. Then the obtained numerical solution (parameter estimates) may strongly depend on the algorithm applied and consequently many of them may accidentally produce instable solutions corrupted by round-off error propagation.

Thus there is a danger that statistical analysis on parameters is partially or fully inadequate because the results of numerical computations may be far from the theoretically exact ones (the random effects are buried in round-off errors and cannot be separated).

For example, the papers [28, 30] may illustrate the above reasoning quite well.

References

1. M. Csörgő, L. Horváth, *Limit Theorems in Change-Point Problem*, Wiley, New York, 1997.
2. G. Dzemyda, L. Sakalauskas, Large-scale data analysis using heuristic methods, *Informatica*, **22**(1):1–10, 2011.
3. A. Janušauskas, V. Marozas, A. Lukoševičius, L. Sörnmo, Detection of hearing loss in audiological frequencies from transient evoked otoacoustic emissions. *Informatica*, **21**(2):191–204, 2010.
4. D. Zuokas, Detecting and locating a changed segment in a binomial sequence: comparison of tests, *Nonlinear Anal. Model. Control*, **10**(2):171–192, 2005.
5. J. Antoch, M. Hušková, D. Jarušková, Change point detection, in *5th ERS IASC Summer School*, IASC, 2000, pp. 1–76.
6. J. Neubauer, V. Veselý, Change point detection by sparse parameter estimation, *Informatica*, **22**(1):149–164, 2011.

7. J. Neubauer, V. Veselý, Impact of serial correlation on change point detection by sparse parameter estimation, in *10th International Conference on Applied Mathematics (APLIMAT 2011)*, Bratislava, 2011, pp. 1249–1254.
8. A.M. Bruckstein, D.L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, *SIAM Rev.*, **51**(1):34–81, 2009.
9. S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, **20**(1):33–61, 1998; *SIAM Rev.*, **43**(1):129–159, 2001.
10. J. Spirik, Algorithms for computing sparse solutions, in *Proceedings of the 17th Conference Student EEICT 2011*, Vol. 3, Novpress s.r.o., Brno, 2011, pp. 123–127.
11. J. Neubauer, V. Veselý, Multiple change point detection by sparse parameter estimation, in *Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics, Paris, France, August 22–27, 2010*, pp. 1453–1459.
12. T. Hastie, R. Tibshirani, J.H. Friedman, *The Elements of Statistical Learning*, Springer, New York, 2001.
13. S.-J. Kim, K. Koh, S. Boyd, D. Gorinevsky, ℓ_1 trend filtering, *SIAM Rev.*, **51**(2):339–360, 2009.
14. D.L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization, in *Proc. Natl. Acad. Sci. USA*, **100**(5):2197–2202, 2003.
15. E.J. Candes, T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory*, **51**:4203–4215, 2005.
16. J.A. Tropp, Average-case analysis of greedy pursuit, in *Proceedings of SPIE*, Vol. 5914, 2005, doi:10.1117/12.618154.
17. D.L. Donoho, J. Tanner, Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing, *Philos. Trans. R. Soc. Lond., Ser. A, Math. Phys. Eng. Sci.*, **367**(1906):4273–4293, 2009.
18. D.V. Hinkley, Inference about the change-point in a sequence of random variables, *Biometrika*, **57**:1–17, 1970.
19. A.J. Scott, M. Knott, a cluster analysis method for grouping means in the analysis of variance, *Biometrics*, **30**(3):507–512, 1974.
20. J. Chen, A.K. Gupta, *Parametric Statistical Change Point Analysis*, Birkhäuser, Boston, 2000.
21. M.A. Saunders, pdsc.m: MATLAB code for minimizing convex separable objective functions subject to $Ax = b, x \geq 0$, 1997–2001, <http://www-leland.stanford.edu/group/SOL/software/pdsc/>.
22. V. Veselý, framebox: MATLAB toolbox for overcomplete modeling and sparse parameter estimation, 2001–2011.
23. J. Zelinka, V. Veselý, I. Horová, Comparative study of two kernel smoothing techniques, in I. Horová (Ed.), *Proceedings of the summer school DATASTAT'2003, Svratka*, Folia Fac. Sci. Nat. Univ. Masaryk. Brun., Math., Vol. 15, Masaryk University, Brno, 2004, pp. 419–436.

24. B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, *Ann. Stat.* **32**(2):407–499, 2004.
25. K. Koh, S.-J. Kim, S. Boyd, Software for ℓ_1 Trend Filtering, ver. May 2008, http://www.stanford.edu/~boyd/l1_tf/.
26. W.J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*, Springer, New York, 1996.
27. R. Garnett, M.A. Osborne, S. Reece, A. Rogers, S.J. Roberts, sequential Bayesian prediction in the presence of changepoints and faults. *Computer J.*, **53**(9):1430–1446, 2010.
28. V. Veselý, J. Tonner, Sparse parameter estimation in overcomplete time series models, *Aust. J. Stat.*, **35**(2–3):371–378, 2006.
29. M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York, Dordrecht, Heidelberg, London, 2010.
30. V. Veselý, J. Tonner, Z. Hrdličková, J. Michálek, M. Kolář, Analysis of PM10 air pollution in Brno based on generalized linear model with strongly rank-deficient design matrix, *Environmetrics*, **20**(6):676–698, 2009.
31. O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, Boston, Basel, Berlin, 2003.
32. T. Hastie, B. Efron, lars: Least Angle Regression, Lasso and Forward Stagewise, R package version 1.1., ver. 2012, <http://CRAN.R-project.org/package=lars>.
33. R. Killick, I.A. Eckley, changepoint: An R package for changepoint analysis, R package version 0.7., ver. 2012, <http://CRAN.R-project.org/package=changepoint>.
34. J. Neubauer, J. Odehnal, Selected methods of economic time series analysis, in *XX International Conference PDMU-2012: Problem of Decision Making under Uncertainties*, University of Defence, Brno, 2012, pp. 151–160.
35. J. Neubauer, J. Odehnal, Detection of changes in time series of economic growth and military expenditure, in *Recent Advances in Energy, Environment, Economic Development*, WSEAS Press, Paris, 2012, pp. 350–355.
36. R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2009, <http://www.R-project.org>.