DATA-ADAPTIVE STRUCTURAL CHANGE-POINT DETECTION VIA ISOLATION

A PREPRINT

Andreas Anastasiou
Department of Mathematics and Statistics
University of Cyprus
anastasiou.andreas@ucy.ac.cy

Sophia Loizidou
Department of Mathematics
University of Luxembourg
sophia.loizidou@uni.lu

ABSTRACT

In this paper, a new data-adaptive method, called DAIS (Data Adaptive ISolation), is introduced for the estimation of the number and the location of change-points in a given data sequence. The proposed method can detect changes in various different signal structures; we focus on the examples of piecewise-constant and continuous, piecewise-linear signals. We highlight, however, that our algorithm can be extended to other frameworks, such as piecewise-quadratic signals. The data-adaptivity of our methodology lies in the fact that, at each step, and for the data under consideration, we search for the most prominent change-point in a targeted neighborhood of the data sequence that contains this change-point with high probability. Using a suitably chosen contrast function, the change-point will then get detected after being isolated in an interval. The isolation feature enhances estimation accuracy, while the data-adaptive nature of DAIS is advantageous regarding, mainly, computational complexity and accuracy. The simulation results presented indicate that DAIS is at least as accurate as state-of-the-art competitors.

Keywords Change-point detection · data-adaptivity · isolation · thresholding criterion

1 Introduction

Change-point detection, also known as data segmentation, is the problem of finding abrupt changes in data when, at least, one of their properties over time changes. It has attracted a lot of interest over the years, mostly due to its importance in time series analysis and the wide range of applications where change-point detection methods are needed. These include genomics (Cao and Biao Wu [2015]), neuroscience (Kovács et al. [2020]), seismic data (Piana Agostinetti and Sgattoni [2021]), astronomy (Fisch et al. [2018]) and finance (Cho and Fryzlewicz [2012]).

There are two directions; sequential (or online) and a posteriori (or offline) change-point detection. We focus on the latter, where the goal is to estimate the number and locations of changes in given data. We work under the model

$$X_t = f_t + \sigma\epsilon_t, \quad t = 1, 2, \ldots, T, \tag{1}$$

where X_t are the observed data and f_t is a one-dimensional deterministic signal with structural changes at certain points. This paper is concerned with piecewise-constant and continuous piecewise-linear signals f_t, meaning that the changes are in the mean or the slope. We highlight that the proposed algorithm can though be extended to other, possibly more complicated, signal structures. Although detecting changes in the mean is a simple case, as noted by Brodsky and Darkhovsky [2000] more complex change-point problems that allow changes in properties other than the mean, can be reduced to the segmentation problem we are studying. This can be done by applying a suitable transformation to the data that reveals the changes as those in the mean of the transformed data. For example, Cho and Fryzlewicz [2015] and Anastasiou et al. [2022] provide methods for the detection of multiple change-points in the second-order (i.e. autocovariance and cross-covariance) structure of possibly high dimensional time series, using Haar wavelets as building blocks. The signal is transformed into a piecewise-constant one with identical change-point locations. We want to emphasise that the novelty of our paper does not come from the structure of changes that the method can be applied to. The importance of our work lies in the data-adaptive nature of the proposed algorithm itself.

The various methods available for solving the problem of change-point detection can be mainly split into two categories, according to whether they are optimization-based methods or they use an appropriately chosen contrast function. The former group includes methods that look for the optimal partition of the data, by performing model selection using a penalization function to avoid overfitting. The latter category's methods do not search for the globally optimal partition of the signal. Instead, the change-point locations are chosen as the most probable location at the given step of the algorithm. Only some of the already existing methods are mentioned in this introduction but comprehensive overviews and more detailed explanations can be found in Truong et al. [2020] and Yu [2020].

Starting with the optimization-based methods for detecting changes in the mean when f_t is piecewise-constant, one of the most common penalty functions is the Schwarz Information criterion (Schwarz [1978]), which was used by Yao [1988] for change-point detection, assuming Gaussian random variables ϵ_t in (1). Under the same assumption, Yao and Au [1989] studies an estimator based on least squares. Relaxing this assumption and allowing instead for more general distributions belonging to the exponential family, Hawkins [2001] introduces a dynamic program which uses maximum likelihood estimates of the location of the change-points. Ninomiya [2005] introduces an AIC-type criterion for change-point models and Ding et al. [2017] shows that while an AIC-like information criterion does not give a strongly consistent selection of the optimal number of change-points, a BIC-like information criterion does. Jackson et al. [2005] employs dynamic programming to guarantee that the exact global optimum, in terms of creating the segments, is found, while Killick et al. [2012] proposes an algorithm, called PELT, whose computational cost is linear in the number of observations. Rigaill [2015] introduces the pDPA algorithm, which includes a pruning step towards complexity reduction. Maidstone et al. [2017] combines the ideas from PELT and pDPA leading to two new algorithms, FPOP and SNIP, which have low computational complexity and, therefore, increased speed.

Focusing now on the category where a contrast function is used, for f_t being piecewise-constant the relevant function is the absolute value of the CUSUM statistic. A method that has received a lot of attention is the Binary Segmentation algorithm, as introduced in Vostrikova [1981]. At each step, the data sequence is split according to the already detected change-points. However, because of checking for a single change-point in intervals that could have any number of change-points, Binary Segmentation has suboptimal accuracy; many methods have been developed with the scope to improve on such drawbacks. One of these methods is proposed by Fryzlewicz [2014], called WBS, which calculates the value of the contrast function on a large number of randomly drawn intervals, which allows detection of change-points in small spacings. Fryzlewicz [2020] proposes WBS2 which draws new intervals each time a detection occurs. A different approach is adopted by Baranowski et al. [2019] with the NOT algorithm. By choosing the narrowest interval that exceeds a pre-defined threshold, the probability that there is exactly one change-point in each interval increases. The ID method in Anastasiou and Fryzlewicz [2022] achieves, first, isolation and then detection of each change-point through an idea based on expanding intervals in a sequential way, starting from both the beginning and the end of the data sequence. The isolation of each change-point maximizes the detection power. Kovács et al. [2020] proposes a seeded binary segmentation algorithm which uses a deterministic construction of search intervals which can be pre-computed. Chu et al. [1995], Eichinger and Kirch [2018] and Levajković and Messer [2023] consider moving sum statistics, while Chan and Chen [2017], through a 'bottom-up' approach, explores reverse segmentation. This involves the creation of a 'solution path' by deleting the change-point with the smallest CUSUM value in the segment determined by its closest left and right neighbors, in order to obtain a hierarchy of nested models.

Methods for detecting more general structural changes can again be split in the same two categories. Focusing on optimization-based methods, Bai and Perron [1998] considers the estimation of linear models based on the least squares principle. Kim et al. [2009] introduces a trend filtering (TF) approach to produce piecewise-linear trend estimates using an ℓ_1 penalty. Fearnhead et al. [2019] presents a locally dynamic approach, called CPOP, that finds the best continuous piecewise-linear fit to the data using penalized least squares and an ℓ_0 penalty. Friedman [1991] proposes the MARS method for flexible regression using splines functions with the degree and knot locations determined by the data. Two methods for optimizing the knot locations and smoothing parameters for least-squares or penalized splines are introduced by Spiriti et al. [2013]. Wiggins [2015] introduces the frequentist information criterion for change-point detection which can detect changes in the mean, slope, standard deviation or serial correlation structure of a signal whose noise can be modeled by Gaussian, Wiener, or Ornstein-Uhlenbeck processes. Sosa-Costa et al. [2018] introduces PLANT, a bottom-up type algorithm that finds the points at which there is potential variation of the slope using a likelihood-based approach, and then recursively merges the adjacent segments. Considering algorithms using a suitably chosen contrast function, the algorithms NOT and ID by Baranowski et al. [2019] and Anastasiou and Fryzlewicz [2022], respectively, can be applied to piecewise-linear signals.

There are some methods that do not fall in either of the categories as described above. Considering piecewise-constant signals, Li et al. [2016] proposes FDRSeg, which controls the false discovery rate in the sense that the number of falsely detected change-points is bounded linearly by the number of true jumps. Fryzlewicz [2018] achieves a multiscale decomposition of the data with respect to an adaptively chosen unbalanced Haar wavelet basis, using a transformation of the data called TGUH transform. Maeng and Fryzlewicz [2023] extends TGUH to piecewise-linear signals.

In this project, we propose a data-adaptive change-point detection method, called DAIS (Data Adaptive ISolation), that attempts to isolate the change-points before detection, by also taking into account the potential true locations of the change-points. This means that the algorithm does not start from a random interval, or the beginning/end of the signal, as most algorithms in the literature, but instead starts checking around a point that there is reason to believe it could be a true change-point. The idea behind the data-adaptive nature of the algorithm arises from the fact that in the case that the unobserved true signal f_t in (1) is piecewise-constant, taking pairwise differences between consecutive time steps will reveal a constant signal with the value 0, with 'spikes' where the change-points occur. This means that the largest 'spike', in absolute value, occurs at the location of the change-point with the largest jump in the sequence. Similarly, in the case that f_t is continuous and piecewise-linear, differencing the signal twice will again reveal a signal with 'spikes' at the locations of the change-points. We try to use this fact in the observed signal, X_t. In the algorithm, we identify the location of this 'spike', which we will be referring to as the 'largest difference' for the rest of the paper, and test around it for possible change-points. The novelty of our work lies in the data-adaptive nature of the algorithm as just described.

The DAIS algorithm uses left- and right-expanding intervals around the location of the largest difference found, denoted $d_{s,e}$, in the interval $[s, e]$ in order to identify the potential change-point that might have caused the spike and so must be close to $d_{s,e}$. This is done in a deterministic way around $d_{s,e}$, expanding once either only to the left or only to the right at each step, in an alternate way. Using expanding intervals, we achieve isolation of the change-points, which is desirable as the detection power of the contrast function is maximized in such cases (see Sections 3.1 and 3.2 for the choice of the contrast function). Due to the alternating sides of the expansions, the location of the largest difference, which we have a reason to believe to be also the location of a change-point, is at the midpoint of the interval being checked (or close to it) after an even number of steps has been performed. This increases the power of the contrast function and so gives an advantage to the detection power of the method. Using an expansion parameter λ_T (more details in Section 4.2) at most $\lceil (e - s + 1)/\lambda_T \rceil + 1$ steps are necessary to check the whole length of the interval around the largest difference. As soon as a change-point is detected, DAIS restarts on two disjoint intervals, one ending at the start-point of the interval where the detection occurred and one starting from the end-point of that same interval. A more detailed explanation of the algorithm can be found in Section 2.

The paper is organized as follows. Section 2 provides details on the methodology of DAIS. We present a simple example followed by a detailed explanation of the algorithm and of the connection between the location of the largest difference and the isolation aspect. In Section 3, we provide theoretical results regarding the consistency of the number of change-points detected and the accuracy of their estimated locations as well as some theoretical results concerning the assumption of Section 2.3. Section 4 includes an explanation of how some parameters of DAIS are selected, some modifications to achieve better performance and a note on computational complexity compared with some competitors. In Section 5, we provide results on simulated data and a comparison to state-of-the-art competitors. In Section 6, two examples using real world data are presented. The first one is about crime data reported daily and the second is weekly data on the Euro to British pound exchange rate. Finally, Section 7 concludes the paper with a summary of the most important findings.

2 Methodology

2.1 Simple example

We begin by presenting a toy example of how DAIS works in practice in the case of f_t in (1) being a piecewise-constant signal. We have a sequence of length $T = 100$, $\sigma = 1$ and a change-point at location $r = 65$ with jump of magnitude 1.5. The data are presented in the left plot of Figure 1 with the true underlying signal plotted in red. For the sake of presentation of this example, the expansion parameter is chosen to be $\lambda_T = 10$. We first calculate the absolute difference of consecutive observations, X_t, for $t = 1, 2, \ldots, T$. As can be seen by the plot on the right of Figure 1, in this example the largest difference occurs at the location of the change-point ($d_{1,100} = 65$). This means that $\text{argmax}_{t=1,2,\ldots,99}|X_{t+1} - X_t| = 65$. As shown in the left plot of Figure 2, DAIS creates right- and left-expanding intervals around this point and so checks, in this order, the intervals $[65, 74]$, $[55, 74]$ and $[55, 84]$, where detection occurs. Note here that the algorithm does not detect the change-point in the first two intervals being checked, most probably either because the magnitude of the jump compared to σ is small, or because the number of observations used is not sufficiently large, or a combination of the two. The change-point is identified at location 65 and then the algorithm restarts in the intervals $[1, 55]$ and $[84, 100]$. As these two intervals contain no change-points in our example, the largest differences are at points where the absolute value of the difference between the noise of two consecutive observations is large. The largest differences are now detected at points 46 and 85 in the two intervals, respectively. The algorithm starts checking the interval $[1, 55]$, with the expansions being performed in a similar way as to the smaller interval $[84, 100]$, on which we focus for simplicity. For the interval $[84, 100]$ the left- and right-expanding subintervals

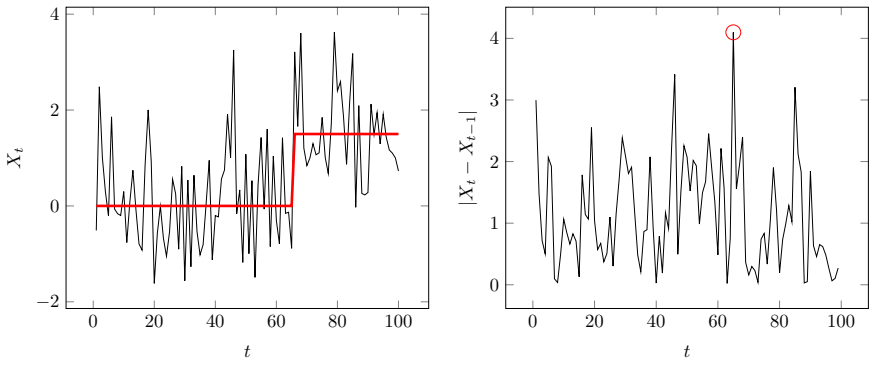

Figure 1: **Left plot:** The data (in black) and the underlying signal (in red) used for the toy example. The change-point is at location 65. **Right plot:** The absolute values of consecutive pairwise differences. The maximum, with location $d_{1,100}$, is marked with a red circle.

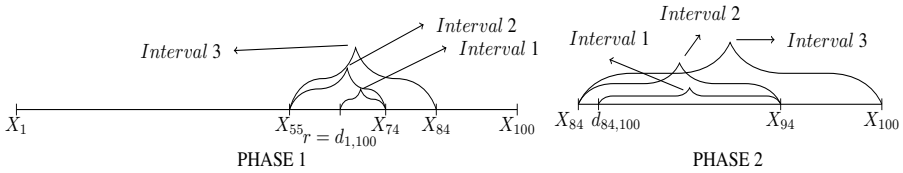

Figure 2: Example with one change-point at $r = 65$ with jump 1.5. **Left plot:** The largest difference is at $d_{1,100} = 65$ and the change-point is detected when the third expanding interval is checked. After detection the next intervals that are checked are $[1, 55]$ and $[84, 100]$. **Right plot:** The left- and right-expanding intervals around the location of the largest difference, $d_{74,100} = 85$, when checking $[84, 100]$. No change-point is detected.

checked are: $[85, 94]$, $[74, 94]$, and $[84, 100]$, as shown in the right plot of Figure 2. Since no change-point is detected in either of the intervals, the algorithm terminates.

2.2 The DAIS algorithm

In the following paragraphs, the method is presented in more generality. For the rest of the paper, we use s, e to denote the start- and the end-point, respectively, of the interval under consideration, while $T \in \mathbb{N}$ denotes the total length of the given signal. $N \in \mathbb{N}_0$ and r_i for $i = 1, 2, \ldots, N$ are the number and locations of the true change-points sorted in increasing order, while $\hat{N}$, $\hat{r}_i$ for $i = 1, 2, \ldots, \hat{N}$ are their estimated values. We denote by $\lambda_T \in \mathbb{N}$ the expansion parameter and $\zeta_T \in \mathbb{R}^+$ the threshold, which is used to decide if a point is a change-point; more details on their choices are given in Section 4.2. In addition, $d_{s,e}$ is the location of the largest difference in the interval $[s, e]$ and is defined as

$$d_{s,e} = \begin{cases} \operatorname{argmax}_{t \in \{s, s+1, \ldots, e-1\}} \{|X_{t+1} - X_t|\}, & f_t \text{ piecewise-constant}, \\ \operatorname{argmax}_{t \in \{s, s+1, \ldots, e-2\}} \{|X_{t+2} - 2X_{t+1} + X_t|\}, & f_t \text{ piecewise-linear}. \end{cases} \tag{2}$$

We denote by $C^b_{s,e}(\boldsymbol{X})$ the contrast function applied to the location b using the values $\{X_s, X_{s+1}, \ldots, X_e\}$ for $b \in [s, e)$. The functions used for the cases of piecewise-constant and continuous, piecewise-linear signals are explained in Sections 3.1 and 3.2, respectively. For $K^l = \lceil \frac{d_{s,e}-s+1}{\lambda_T} \rceil$, $K^r = \lceil \frac{e-d_{s,e}+1}{\lambda_T} \rceil$, $K^{\max} = \max\{K^l, K^r\}$, $K^{\min} = \min\{K^l, K^r\}$ and $K = K^{\max} + K^{\min}$ the start- and end-points of the intervals (c^l_m, c^r_k) in which the algorithm tests for the existence of a change-point, are given by

$$\begin{aligned} c^l_m &= \max\{d_{s,e} - m\lambda_T, s\}, \quad m = 0, 1, \ldots, K^{\max}, \\ c^r_k &= \min\{d_{s,e} + k\lambda_T - 1, e\}, \quad k = 1, 2, \ldots, K^{\max}. \end{aligned}$$

These are collected in the ordered sets, $L^b_{d_{s,e},s,e}$ and $R^b_{d_{s,e},s,e}$, defined in (3), which consist of the left and right end-points, respectively, of the intervals used in DAIS. Since the expansions occur in turn once from the right and once from the left, the points appear twice in each set, up until the moment when one end-point is equal to either s or e. The left- and right-expanding intervals checked are collected in the ordered set I_j in (4). The ordered sets of end-points for

Algorithm 1 DAIS

function DAIS(s, e, λ_T, ζ_T)
if $e - s < 1$ **then**
  STOP
else
  Set $d_{s,e}$ as in (2)
  For $j \in \{1, 2, \ldots, K\}$ let $I_j = [s_j, e_j]$ as in (4)
  $i = 1$
  (Main part)
  $b_i = \text{argmax}_{\text{t}\in[\text{s}_\text{i},\text{e}_\text{i})} \text{C}^{\text{t}}_{\text{s}_\text{i},\text{e}_\text{i}}(\mathbf{X})$
  if $C^{b_i}_{s_i,e_i}(\mathbf{X}) > \zeta_T$ **then**
    add b_i to the list of estimated change-points
    DAIS($s, s_i, \lambda_T, \zeta_T$)
    DAIS($e_i, e, \lambda_T, \zeta_T$)
  else
    $i = i + 1$
    if $i \leq K$ **then**
      Go back to **(Main part)** and repeat
    else
      STOP
    end if
  end if
end if
end function

the expansions are:

$$L^b_{d_{s,e},s,e} = \left\{c^l_0, c^l_1, c^l_1, c^l_2, \ldots, c^l_{K^{\min}-1}, c^l_{K^{\min}-1}, c^l_{K^{\min}}, c^l_{K^{\min}+1}, \ldots, c^l_{K^{\max}}\right\},$$
$$R^b_{d_{s,e},s,e} = \left\{c^r_1, c^r_1, c^r_2, c^r_2, \ldots, c^r_{K^{\min}}, c^r_{K^{\min}}, c^r_{K^{\min}+1}, c^r_{K^{\min}+2}, \ldots, c^r_{K^{\max}}\right\}, \quad (3)$$

while the ordered set of the intervals checked can be written as:

$$I_j = \left[s_j, e_j\right] = \left[L^b_{d_{s,e},s,e}[j], R^b_{d_{s,e},s,e}[j]\right], \text{ for } j \in \{1, 2, \ldots, K\}, \quad (4)$$

where A[i] denotes the i^{th} element of the ordered set A. It is easy to show that $|L^b_{d_{s,e},s,e}| = |R^b_{d_{s,e},s,e}| = K$, where $|A|$ denotes the cardinality of the set A. Note that the largest number of expansions required in order to check the whole length of the interval $[s, e]$ around $d_{s,e}$ is $K = \lceil \frac{e-s+1}{\lambda_T} \rceil + 1$.

Algorithm 1 provides a pseudocode that briefly explains DAIS' workflow. The first step of DAIS is to identify, within the observed data, X_t, $t = 1, 2, \ldots, T$, the location of the largest difference $d_{1,T}$, as defined in (2). In the case that f_t is piecewise-constant, $d_{1,T}$ is the value of t that maximizes the pairwise absolute differences between consecutive observations. In the case that f_t is continuous, piecewise-linear, $d_{1,T}$ is the value of t that maximizes the absolute value of the twice differenced data. Applying the appropriate transformation on the unobserved true signal f_t in (1), will reveal a constant signal at 0, with non-zero values at the locations where the change-points occur, so we expect that the largest difference of X_t will be at a point near the change-point with the jump of the largest magnitude. A proper explanation of what we consider to be 'near' is discussed in Section 2.3 in detail. After calculating $d_{1,T}$, expanding intervals are used to check for possible change-points. The algorithm expands the intervals once from the right and once from the left at every step by an expansion parameter of magnitude λ_T. This is done in order to identify the location of the change-point that is closest to the detected largest difference, which is the reason why the algorithm is data-adaptive. This expansion is performed for at most $\lceil \frac{T}{\lambda_T} \rceil + 1$ steps in total. Upon the detection of a change-point, the algorithm restarts after discarding the data that have already been checked. If the change was detected in the subinterval $[s^*, e^*]$ then DAIS will be reapplied to $[1, s^*]$ and $[e^*, T]$. When checking any interval $[s, e]$, a contrast function is applied to the data, testing for possible change-points at location b, where $b \in \{s, s + 1, \ldots, e - 1\}$. The contrast function is denoted by $C^b_{s,e}(\boldsymbol{X})$ and the maximum value attained in the interval is compared to a threshold ζ_T. If $C^b_{s,e}(\boldsymbol{X})$ exceeds the threshold, we take b to be a change-point. If no change-point is detected, then the algorithm terminates. The choice of the contrast function for the different cases of signals f_t in (1), is explained in Sections 3.1 and 3.2, while the choice of the values for ζ_T and λ_T is explained in Section 4.2.

There are two main advantages to the algorithm due to its data-adaptive nature. Firstly, since DAIS starts checking for potential change-points in an area where there is reason to believe that includes the location of the change-point

with the change of the largest magnitude, the change-point is detected fast. Secondly, when the change-point is at the location of the largest difference, or close to it, due to the way the expansions occur, the change-point lies in the middle of the intervals being checked when an even number of expansions have been performed. This provides advantages in detection power due to the change-point being detected in balanced intervals. Thus, the data-adaptive nature of the algorithm enhances DAIS' speed and accuracy in estimating the locations of the change-points.

2.3 Location of the largest difference and the isolation aspect

As already explained in detail in Section 2, the algorithm calculates the location of the largest difference $d_{s,e}$, as defined in (2), in the interval $[s, e]$ under consideration at each step and checks for change-points around it. In this subsection, we explain the role of the location of the largest difference in the detection of a change-point in an interval where isolation is guaranteed.

As will be explained in the proof of Theorem 1 in Appendix B, the detection of the change-point r_j will certainly occur if both end-points are at least at a distance $\delta_T/2n$ from the true location of the change-point, but it can happen at any smaller interval where the value of the contrast function exceeds the pre-specified threshold ζ_T. The positive number n satisfies $3/2 \leq n \leq \delta_T/2\lambda_T$ where, for $r_0 = 1, r_{N+1} = T$, $\delta_T = \min_{j=1,\ldots,N+1}\{r_j - r_{j-1}\}$ is the smallest distance between two consecutive change-points and λ_T is the expansion parameter. In the worst-case scenario, meaning that the interval in which detection occurs is the biggest possible and therefore the most expansions have been performed, it holds that, firstly, both end-points are at least a distance $\delta_T/2n$ from the change-point and, secondly, at least one of the end-points will be in one of the intervals:

$$I_j^L = \left(r_j - \frac{\delta_T}{n}, r_j - \frac{\delta_T}{2n}\right], \quad I_j^R = \left[r_j + \frac{\delta_T}{2n}, r_j + \frac{\delta_T}{n}\right) \tag{5}$$

for $3/2 \leq n \leq \delta_T/2\lambda_T$. Setting $n = \frac{\delta_T}{2\lambda_T}$ ensures that the detection happens as quickly as possible. Note that in Anastasiou and Fryzlewicz [2022] and Fryzlewicz [2014] $n = 3/2$ while in Baranowski et al. [2019] $n = 3$ when defining similar intervals for the location of the end-points of the interval where detection occurs.

Since the length of the intervals in (5) is $\delta_T/2n$, the condition $2n\lambda_T \leq \delta_T$ is required, so that there is always at least one end-point that lies within I_j^L and I_j^R. Since $\lambda_T \geq 1$, we implicitly require $\delta_T \geq 2n$. Define $\delta_j = r_{j+1} - r_j$ and $\delta_{s,e}^j = \min\{r_{j+1} - d_{s,e}, d_{s,e} - r_j\}$ for $j = 0, 1, \ldots, N$. Without loss of generality, suppose that $d_{s,e} \in [r_J, r_{J+1}]$ for some $J \in \{0, 1, \ldots, N\}$. Isolation is guaranteed whenever the following holds

$$0 \leq \delta_{s,e}^J \leq \frac{\delta_J}{2} - \frac{3\delta_T}{4n}. \tag{6}$$

Note that the lower bound of n ensures that the right hand side of the above equation is non-negative. The bound in (6) both guarantees detection in an interval where the change-point is isolated but also allows for the detection of the neighbouring change-point. This is because the end-point of the interval the algorithm will be applied to after detection of the first change-point (either r_J or r_{J+1}), will have a distance of at least $\delta_T/2n$ from all the undetected change-points; see (29) and (B.26) in the supplementary material for a proof that $\delta_T/2n$ is the minimum distance of the end-points from the change-point that guarantees detection.

It should be mentioned that $\delta_j = \mathcal{O}(\delta_T)$ for any $j = 0, 1, \ldots, N$. Also, since $2n\lambda_T \leq \delta_T$, $n\lambda_T = \mathcal{O}(\delta_T)$ and the upper bound of $\delta_{s,e}^J$ is $\mathcal{O}(\delta_T)$. Now, as is given by assumption (A2) of Section 3.1, it must hold that $\sqrt{\delta_T}\underline{f}_T \geq \underline{C}\sqrt{\log T}$. This means that when $T \to \infty$ then either δ_T or $\underline{f}_T$ increases or both do.

- If δ_T increases then the upper bound of $\delta_{s,e}^J$ increases with the same rate.
- If $\underline{f}_T$ increases, the signal-to-noise ratio becomes larger and so the probability that the largest difference is exactly at the location of the change-point increases.

In either case, the probability that $d_{s,e}$ lies within $\delta_{s,e}^J$ of any change-point increases as T becomes large. For the case of piecewise-linear signals, we can reach the same conclusion using assumption (A3) in Section 3.2.

Considering now any $d_{s,e} \in \{1, 2, \ldots, T\}$ and using (6), it must hold that

$$r_J - \frac{\delta_{J-1}}{2} + \frac{3\delta_T}{4n} \leq d_{s,e} \leq r_J + \frac{\delta_J}{2} - \frac{3\delta_T}{4n} \tag{7}$$

for the change-point r_J that is closest to $d_{s,e}$. Note that the length of the interval where $d_{s,e}$ cannot lie under this assumption is $2(\lceil 3\delta_T/4n \rceil - 1)$ which can be as small as $3\lambda_T - 2$. This means that for appropriately chosen values of n and λ_T, any location is an acceptable location for $d_{s,e}$ besides the midpoint between consecutive change-points.

We highlight that in the specific case of ϵ_t in (1) being i.i.d. random variables following the $\mathcal{N}(0,1)$, a discussion on the calculation of the exact probability that the largest difference is at a location that guarantees detection in an interval where the change-point is isolated, for all change-points, can be found in Section 3.3.

3 Theory

DAIS requires knowledge of the standard deviation, σ, of the data generation mechanism in (1). If σ is unknown, it can be estimated using the Median Absolute Deviation method (MAD), with the estimator defined as $\hat{\sigma} := C \times \text{median}(|\boldsymbol{x} - \text{median}(\boldsymbol{x})|)$ for $\boldsymbol{x} = (x_1, x_2, \dots x_T)$, as proposed by Hampel [1974]. In the case of Gaussian data, for $C = 1.4826$ it is a consistent estimator of the population standard deviation σ as proven by Rousseeuw and Croux [1993]. In this section, we assume that ϵ_t in (1) are i.i.d. random variables following the $\mathcal{N}(0,1)$ distribution. This assumption is commonly used in the literature to prove the theoretical results. We also assume that $\sigma = 1$ in order to make notation easier in the proof of the theorems presented in this section. With this assumption, the model becomes

$$X_t = f_t + \epsilon_t, \quad t = 1, 2, \dots, T. \tag{8}$$

In Sections 3.1 and 3.2 we provide the theoretical results on the consistency of the algorithm for the cases of piecewise-constant and continuous, piecewise-linear signals, respectively. The proofs of Theorems 1 and 2 can be found in Appendix B. In Section 3.3, we return to the discussion of Section 2.3 and present some results in this specific case where ϵ_t in (1) are i.i.d. from the $\mathcal{N}(0,1)$ distribution.

3.1 Piecewise-constant signals

Under piecewise-constancy, $f_t = \mu_j$ for $t = r_{j-1} + 1, \dots, r_j$ and $f_{r_j} \neq f_{r_j+1}$. The statistic used as a contrast function in this case is the absolute value of the CUSUM statistic, which is defined as

$$C^b_{s,e}(\mathbf{X}) = \left|\tilde{X}^b_{s,e}\right| = \left| \sqrt{\frac{e-b}{\ell(b-s+1)}} \sum_{t=s}^{b} X_t - \sqrt{\frac{b-s+1}{\ell(e-b)}} \sum_{t=b+1}^{e} X_t \right|, \tag{9}$$

where $s \leq b < e$ and $\ell = e - s + 1$. The value of $C^b_{s,e}(\mathbf{X})$ is small if b is not a change-point and large otherwise.

To prove the consistency of our method in accurately estimating the number and locations of the estimated change-points, we work under the following assumptions:

(A1) The location of the largest difference $d_{s,e}$ as defined in (2), satisfies $d_{s,e} \in \bigcup_{j=1}^{N} [r_j - \frac{\delta_{j-1}}{2} + \frac{3\delta_T}{4n}, r_j + \frac{\delta_j}{2} - \frac{3\delta_T}{4n}]$ every time the algorithm is applied to an interval $[s, e]$ which contains at least one change-point.

(A2) The minimum distance, δ_T, between two successive change-points and the minimum magnitude of jumps $\underline{f}_T$ are connected by $\sqrt{\delta_T}\underline{f}_T \geq \underline{C}\sqrt{\log T}$ for a large enough constant $\underline{C}$.

The first assumption is a formal statement of what is discussed in Section 2.3. It is not a strict assumption and the exact probability of the consecutive pairwise difference being larger at a change-point compared to a point where no change occurs, in the case where ϵ_t is assumed to be Gaussian, is provided in Section 3.3. Below, we provide the relevant theorem for the consistency of DAIS in accurately estimating the true number and the locations of the change-points in the case of piecewise-constant signals.

Theorem 1. *Let $\{X_t\}_{t=1,2,\dots,T}$ follow model (8), with f_t being a piecewise-constant signal and assume that the random sequence $\{\epsilon_t\}_{t=1,2,\dots,T}$ is independent and identically distributed (i.i.d.) from the normal distribution with mean zero and variance one and also that assumptions (A1) and (A2) hold. Let N and $r_j, j = 1, 2, \dots N$ be the number and location of the change-points, while $\hat{N}$ and $\hat{r}_j, j = 1, 2, \dots \hat{N}$ their estimates, sorted in increasing order. In addition, $\Delta^f_j = \left|f_{r_j+1} - f_{r_j}\right|, j = 1, 2, \dots N$ is the magnitude of each discontinuity in f_t, $\underline{f}_T = \min_j \Delta^f_j$ and $\delta_T = \min_{j=1,2,\dots,N+1} |r_j - r_{j-1}|$, where $r_0 = 0, r_{N+1} = T$. Then there exist positive constants C_1, C_2, C_3, C_4 which do not depend on T such that for $C_1\sqrt{\log T} \leq \zeta_T < C_2\sqrt{\delta_T}\underline{f}_T$ and for sufficiently large T, we obtain*

$$\mathbb{P}\left(\hat{N} = N, \max_{j=1,2,\dots,N}\left(|\hat{r}_j - r_j|\left(\Delta^f_j\right)^2\right) \leq C_3 \log T\right) \geq 1 - \frac{C_4}{T}. \tag{10}$$

The lower bound of the probability in (10) is $1 - \mathcal{O}(1/T)$. From Theorem 1, we can conclude that in order to be able to match the estimated change-points with the true ones, δ_T needs to be larger than $\max_{j=1,2,\ldots,N} |\hat{r}_j - r_j|$, which means that δ_T must be at least of order $\mathcal{O}(\log T)$. Also, assumption (A2) ensures that the rate attained for $\delta_T \underline{f}_T^2$, which characterizes the complexity of the problem, is $\mathcal{O}(\log T)$ and, as Chan and Walther [2013] argues, the lowest possible $\delta_T \underline{f}_T^2$ that allows detection of the change-points is $\mathcal{O}(\log T - \log\log T)$. Therefore, the rate obtained by DAIS is optimal up to a double logarithmic factor.

3.2 Continuous, piecewise-linear signals

For the case of continuous, piecewise-linear signals $f_t = \mu_{j,1} + \mu_{j,2} t$ for $t = r_{j-1}+1, \ldots, r_j$ and $f_{r_j-1} + f_{r_j+1} \neq 2f_{r_j}$ with the additional constraint $\mu_{j,1} + \mu_{j,2} r_j = \mu_{j+1,1} + \mu_{j+1,2} r_j$ so that the signal is continuous. Baranowski et al. [2019] shows that the best choice of contrast function is $C_{s,e}^b(\mathbf{X}) = \left|\langle \boldsymbol{X}, \boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{b}} \rangle\right|$, where the contrast vector $\boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{b}} = (\phi_{s,e}^b(1), \ldots, \phi_{s,e}^b(T))$ is given by

$$\phi_{s,e}^b(t) = \begin{cases} \alpha_{s,e}^b \beta_{s,e}^b \big[(e+2b-3s+2)t - (be+bs-2s^2+2s)\big], & t \in \{s, \ldots, b\}, \\ -\frac{\alpha_{s,e}^b}{\beta_{s,e}^b}\big[(3e-2b-s+2)t - (2e^2+2e-be-bs)\big], & t \in \{b+1, \ldots, e\}, \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

where $\ell = e - s + 1$ and

$$\alpha_{s,e}^b = \left(\frac{6}{\ell(\ell^2-1)(1+(e-b+1)(b-s+1)+(e-b)(b-s))}\right)^{\frac{1}{2}},$$

$$\beta_{s,e}^b = \left(\frac{(e-b+1)(e-b)}{(b-s+1)(b-s)}\right)^{\frac{1}{2}}.$$

Under the Gaussianity assumption, $C_{s,e}^b(\mathbf{X})$ as defined above, is maximized at the same point as the generalized log-likelihood ratio for all possible single change-points within $[s, e)$.

Before presenting the theoretical result for the consistency of the number and location of the estimated change-points in the case of continuous piecewise-linear signals, we require the following assumption, which is equivalent to (A2) in Section 3.1, but now under the current setting of changes in the slope.

(A3) The minimum distance, δ_T, between two successive change-points and the minimum magnitude of jumps $\underline{f}_T = \min_{j=1,2,\ldots,N} |2f_{r_j} - f_{r_j+1} - f_{r_j-1}|$ are connected by $\delta_T^{3/2} \underline{f}_T \geq C^* \sqrt{\log T}$ for a large enough constant C^*.

Theorem 2. *Let $\{X_t\}_{t=1,2,\ldots,T}$ follow model* (8)*, with f_t being a continuous piecewise-linear signal and assume that the random sequence $\{\epsilon_t\}_{t=1,2,\ldots,T}$ is independent and identically distributed (i.i.d.) from the normal distribution with mean zero and variance one and also that (A1) and (A3) hold. Let N and $r_j, j = 1, 2, \ldots N$ be the number and location of the change-points, while $\hat{N}$ and $\hat{r}_j, j = 1, 2, \ldots \hat{N}$ are their estimates, sorted in increasing order. In addition, $\Delta_j^f = \left|2f_{r_j} - f_{r_j+1} - f_{r_j-1}\right|, j = 1, 2, \ldots N$ is the magnitude of each change of slope in f_t, $\underline{f}_T = \min_j \Delta_j^f$ and $\delta_T = \min_{j=1,2,\ldots,N+1} |r_j - r_{j-1}|$, where $r_0 = 0, r_{N+1} = T$. Then there exist positive constants C_1, C_2, C_3, C_4 which do not depend on T such that for $C_1\sqrt{\log T} \leq \zeta_T < C_2 \delta_T^{3/2} \underline{f}_T$ and for sufficiently large T, we obtain*

$$\mathbb{P}\left(\hat{N} = N, \max_{j=1,2,\ldots,N}\left(|\hat{r}_j - r_j| \left(\Delta_j^f\right)^{2/3}\right) \leq C_3 (\log T)^{1/3}\right) \geq 1 - \frac{C_4}{T}. \tag{12}$$

Again, The lower bound of the probability is $1 - \mathcal{O}(1/T)$ and the difficulty of the problem, $\delta_T^{3/2} \underline{f}_T$ is analogous to $\sqrt{\delta_T} \underline{f}_T$ in the case of piecewise-constant signals. The proof of Theorem 2 can be found in the supplementary material.

3.3 Probability of isolation

We now return to the discussion of Section 2.3. Under the assumption of ϵ_t being i.i.d. random variables following the $\mathcal{N}(0, 1)$ distribution, and in the case that f_t is piecewise-constant, we note that in order to calculate the probability that

the largest difference is exactly at the location of a change-point, we need to calculate the following:

$$\mathbb{P}\left(\bigcup_{j=1}^{N}\bigcap_{t=1}^{T-1}|X_{r_j+1}-X_{r_j}|\geq|X_{t+1}-X_t|\right). \tag{13}$$

For $t\notin\{r_j-1,r_j,r_j+1\}, j\in\{1,\ldots,N\}$, we focus on the probability that the largest difference occurs at the location of one of the N change-points. Note that $|X_{r_j+1}-X_{r_j}|$ and $|X_{t+1}-X_t|$ are independent for these values of t. We will work on

$$\mathbb{P}\left(|X_{r_j+1}-X_{r_j}|\geq|X_{t+1}-X_t|\right), \tag{14}$$

which is the probability that the observations at the change-point have a larger absolute difference than a pair of observations in the signal where no change occurs. Using (1), the fact that $\epsilon_t\sim\mathcal{N}(0,1)$ and that $f_{t'+1}-f_{t'}=0$ for $t'\neq r_j, j\in\{1,\ldots,N\}$, (14) can be written as

$$\begin{aligned}\mathbb{P}\left(|X_{r_j+1}-X_{r_j}|\geq|X_{t+1}-X_t|\right)&=\mathbb{P}\left(|f_{r_j+1}-f_{r_j}+\sigma Z_2|\geq|\sigma\tilde{Z}_2|\right)\\&=\mathbb{P}\left(|\alpha_j+Z_2|\geq|\tilde{Z}_2|\right)\end{aligned} \tag{15}$$

where $\alpha_j=(f_{r_j+1}-f_{r_j})/\sigma$ and $Z_2,\tilde{Z}_2\sim\mathcal{N}(0,2)$ are independent. In the above equation, $|\alpha_j+Z_2|$ follows the Folded Normal (FN) distribution with mean α_j and variance 2. Similarly, $|\tilde{Z}_2|$, follows the same distribution with mean 0 and variance 2, which reduces to the simpler case of the Half Normal (HN) distribution with variance 2. So, (15) can be equivalently written as

$$\mathbb{P}(Y_1\geq Y_2) \tag{16}$$

for $Y_1\sim FN(\alpha_j,2)$ and $Y_2\sim HN(0,2)$. Denoting by f_{Y_1} the probability density function of Y_1 and F_{Y_2} the cumulative density function of Y_2, (16) can be written as

$$\int_0^\infty\mathbb{P}(Y_2\leq y)f_{Y_1}(y)dy=\int_0^\infty F_{Y_2}(y)f_{Y_1}(y)dy=\mathbb{E}_{Y_1}\Big[F_{Y_2}(Y_1)\Big]=\mathbb{E}_{Y_1}\Big[\text{erf}\Big(\frac{Y_1}{2}\Big)\Big]$$

where $\text{erf}(z)=\frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\text{dt}$ is the Gauss error function. Using Ng and Geller [1969] (equation (38) on page 10), we can calculate the above in terms of α_j as

$$\mathbb{P}(Y_1\geq Y_2)=\frac{1}{2}\left[1+\text{erf}\Big(\frac{\alpha_j}{2\sqrt{2}}\Big)^2\right].$$

It is easy to see now that this probability goes to 1 as α_j gets larger and in fact $\mathbb{P}(Y_1\geq Y_2)>0.99$ for $\alpha_j=5.15$. However, calculating or bounding from below the probability in (13) is a very difficult task due to the fact that $|X_t-X_{t-1}|$ and $|X_{t-1}-X_{t-2}|$ are dependent. Therefore, we provide results based on simulations.

We ran a simulation study to observe the behavior of the probability of the largest difference being in a neighborhood that guarantees detection in an interval where the change-point is isolated as $\alpha=\min_{j=1,2,\ldots,N}\alpha_j$ increases. By the definition of α, it holds that $\alpha=\underline{f}_T/\sigma$, where $\underline{f}_T=\min_{j=0,1,\ldots,N}\left|f_{r_j+1}-f_{r_j}\right|$. The location of the largest difference is close to the change-point with the largest jump in the interval under consideration, and so in the case of multiple change-points, the probability in question depends on the value of the largest magnitude of change. In these simulations we use the value of $\underline{f}_T$, to obtain the smallest probability of the largest difference being in a location that guarantees detection throughout the whole algorithm, until all change-points have been detected.

For the simulation study, we used one change-point, r, with δ_T fixed at value 20 and set $\sigma=1$. We repeated this 10000 times for various values of the length T. We calculated the Monte Carlo estimates of the probabilities that the location of the largest difference ensured that detection occurred in an interval where the change-point is isolated as T increased. In other words, using (7), we calculated the proportion of times that

$$d_{1,T}\in\Big[\frac{r+1}{2}+\frac{3\delta_T}{4n},\frac{T+r}{2}-\frac{3\delta_T}{4n}\Big],$$

where n is chosen as $n=\frac{\delta_T}{2\lambda_T}$ for $\lambda_T=3$. In order to be consistent with assumption (A2), $\underline{f}_T=c\sqrt{\log T}$, where $c>0$ is a constant. In our simulation study, c is chosen such that α is equal to five when $T=50$. This choice of c led to an early convergence of the relevant proportion; please, see Table 1. Smaller values of c would be acceptable but larger values of the sample size, T, would be required until convergence to one becomes apparent. The results in Table 1 show that the probability of the largest difference occurring at a point that guarantees isolation in the interval where detection occurred increases with the length of the signal. This is expected as $\underline{f}_T=\mathcal{O}(\sqrt{\log(T)})$.

Table 1: Proportion of times that the location of the largest difference was in an interval where detection while the change-point is isolated is guaranteed.

T	α	proportion
50	5.00	0.894
80	5.30	0.907
100	5.43	0.908
150	5.66	0.922
200	5.82	0.923
250	5.95	0.933
500	6.31	0.938
750	6.51	0.941
1000	6.65	0.939
2000	6.97	0.947
5000	7.38	0.953
10000	7.68	0.957
50000	8.32	0.963
100000	8.58	0.966

4 Computational complexity and practicalities

4.1 Computational complexity

We will now explain why DAIS needs to check at most $\lceil \frac{T}{\lambda_T} \rceil + 2N + 1$ intervals before the algorithm is terminated, where N is the total number of change-points of the signal, T is its length and λ_T the chosen expansion parameter. If there are no change-points, the algorithm will check at most $\lceil \frac{T}{\lambda_T} \rceil + 1$ left- and right-expanding intervals around the location of the largest difference, as explained in Section 2. In the case that the signal has change-points, since every time detection occurs the already used observations are discarded and the algorithm is restarted on two disjoint subintervals, the total number of expansions required increases by 2 (1 for each of the newly created subintervals). Therefore, the maximum number of iterations depends on the number of change-points and DAIS needs to check at most $K' = \lceil \frac{T}{\lambda_T} \rceil + 2N + 1$ intervals. Since we take $\lambda_T < \delta_T$, then $K' > \lceil T/\delta_T \rceil + 2N + 1$. Combining this with the definition of δ_T, which implies that $N \leq T/\delta_T$, it can be concluded that the lower bound of K' is $\mathcal{O}(T/\delta_T)$. As explained by Anastasiou and Fryzlewicz [2022], the lower bound for the maximum number of intervals that need to be checked in ID is $\mathcal{O}(T/\delta_T)$ and in WBS and NOT it is $\mathcal{O}(T^2/\delta_T^2)$ up to a logarithmic factor, which means that DAIS is faster than the latter two algorithms and is of the same order as ID.

Now, focusing on the algorithms ID and DAIS, we note that the ideas behind the two methods are similar, since both first isolate the change-point and then detect it. However, the data-adaptive nature of DAIS leads to computational advancements. Using again the definition of δ_T and the equation $\delta_T \geq 2n\lambda_T$, for which the explanation is in Section 2.3, we get

$$N \leq \frac{T}{2n\lambda_T} < \frac{T}{2\lambda_T} - 1.$$

So, $K' < 2T/\lambda_T$ and since ID checks at most $2\lceil \frac{T}{\lambda_T} \rceil$ intervals, it holds that the maximum number of intervals that are required for DAIS to check a signal is always less than the maximum number of intervals that ID requires. For example, when the signal has no change-points, using the same expansion parameter λ_T, DAIS checks almost half as many intervals as ID.

4.2 Parameter selection and Variants

Choice of the threshold ζ_T: In Theorems 1 and 2, the rate of the lower bound of the threshold ζ_T is $\mathcal{O}(\sqrt{\log T})$ and so we use

$$\zeta_T = C\sqrt{\log T}. \tag{17}$$

For the choice of the positive threshold constant C, we ran an extensive simulation study using various signal structures and Gaussian noise. The best behavior occurred for $C = 1.7$ in the case of piecewise-constant signals and $C = 2.1$ in the case of piecewise-linear signals. These are the constants that were used for the simulations in Section 5. Anastasiou and Fryzlewicz [2022] proves in Corollary 1 that, in the case of piecewise-constant signals, as $T \to \infty$, the threshold can be taken to be at most $\sqrt{3\log T}$, meaning that $C \leq \sqrt{3} \approx 1.73$; our choice does not violate this result.

Table 2: Methods used in the simulations

Type of Signal	Method Notation	Reference	R package
Piecewise-constant	ID	Anastasiou and Fryzlewicz [2022]	IDetect
	PELT	Killick et al. [2012]	changepoint
	WBS	Fryzlewicz [2014]	wbs
	WBS2	Fryzlewicz [2020]	breakfast
	NOT	Baranowski et al. [2019]	not
	MOSUM	Meier et al. [2021]	mosum
	MSCP	Levajković and Messer [2023]	mscp
Continuous, piecewise-linear	ID	Anastasiou and Fryzlewicz [2022]	IDetect
	CPOP	Fearnhead et al. [2019]	cpop
	NOT	Baranowski et al. [2019]	not
	MARS	Friedman [1991]	earth
	TF	Kim et al. [2009]	-
	TS	Maeng and Fryzlewicz [2023]	trendsegmentR

Choice of the expanding parameter λ_T: As can be seen by the proofs of Theorems 1 and 2, in a given signal, detection will occur for any $\lambda_T \leq \delta_T/2n$, where $3/2 \leq n \leq \delta_T/2\lambda_T$. Using $\lambda_T = 1$ ensures that each change-point is isolated for as many intervals as possible, but this increases the computational time. In practice, we use $\lambda_T = 3$, which is small enough to have very good accuracy in the detection of the change-points, but not too small which could magnify the computational cost.

Variants: In practice, we use a modification to the algorithm as described in Section 2, in order to make it more accurate. DAIS restarts from the detected change-point instead of the start- and end-points of the interval where the detection occurred. This leads to the improvement of the accuracy of the algorithm without affecting the speed.

5 Simulations

This section compares the performance of DAIS with state-of-the-art competitors. We only focus on algorithms that estimate both the number and the locations of the change-points in the given univariate signal in an offline manner. The competitors used in the simulation study can be found in Table 2, along with the R library from which each function was obtained. For ID and WBS we report the results for both the information criterion and the thresholding (for $C = 1.56$ for ID and $C = \sqrt{2}$ for WBS) stopping rules. The notation is ID_ic, WBS_ic and ID_th, WBS_th, respectively. In order to allow the competitors achieve their best performance, in some cases inputs of the functions as developed in R were adjusted. The competitors ID, WBS2, PELT, MOSUM and MARS were applied using the default values. The value for the maximum number of change-points to be detected was altered in each signal appropriately for WBS and NOT. For MSCP, the value of the minimal window considered is set to be less than the smallest distance between consecutive change-points. The estimation of the standard deviation for CPOP is set to be done using the MAD method and TF is employed based on the implementation found in `https://github.com/hadley/l1tf`. For TS, the algorithm for a continuous piecewise-linear signal with independent observations is used. Code on the simulations, as well as on how to implement DAIS, can be found in `https://github.com/Sophia-Loizidou/DAIS`.

We measure the difference between the number of change-points detected and the true number of change-points ($\hat{N} - N$). As a measure of accuracy of the location of the detected change-points, we give Monte-Carlo estimates of the mean squared error,

$$\text{MSE} = T^{-1} \sum_{t=1}^{T} \mathbb{E}\left(\hat{f}_t - f_t\right)^2,$$

where T is the total length of the sequence. For piecewise-constant signals, $\hat{f}_t$ is the ordinary least square approximation of f_t between two successive change-points while for piecewise-linear signals it is the splines fit. We also report the scaled Hausdorff distance, defined as

$$d_H = n_s^{-1} \max\left\{ \max_j \min_k |r_j - \hat{r}_k|, \max_k \min_j |r_j - \hat{r}_k| \right\}$$

where $n_s = \max_{j=1,\ldots,N} \delta_j$ is the length of the largest segment between successive change-points. The Hausdorff distance, d_H, is given for all signals except for signals that do not contain any change-points; in such cases, d_H is uninformative. Finally, the average time taken for each method to find the change-points is reported. The signals used, can be found in Appendix A and some further simulations results are in the Supplementary Material.

Table 3: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signal (S1). The average MSE and computational time are also given.

	$\hat{N} - N$					
Method	0	1	2	≥ 3	MSE	Time (s)
DAIS	99	0	1	0	1.87×10^{-4}	0.196
ID_th	84	8	7	1	5.26×10^{-4}	0.268
ID_ic	99	1	0	0	1.68×10^{-4}	0.063
WBS_th	22	16	31	31	31.7×10^{-4}	0.079
WBS_ic	99	1	0	0	1.68×10^{-4}	0.079
WBS2	91	5	3	1	2.75×10^{-4}	3.250
PELT	100	0	0	0	1.39×10^{-4}	0.002
NOT	100	0	0	0	1.39×10^{-4}	0.062
MOSUM	94	2	4	0	2.96×10^{-4}	0.028
MSCP	100	0	0	0	1.39×10^{-4}	83.250

Table 3 concerns a signal with no change-points. Most algorithms have very good results with negligible differences in the accuracy in the number of change-points detected, WBS_th with the threshold constant $C = \sqrt{2}$ seems to overestimate. Similar conclusions hold for Table 4, in which the big length of the signal (S2) forces some algorithms, especially MSCP, to be extremely slow. Moreover, comparing DAIS with ID_th which use the same value for the expansion parameter λ_T, it is evident that DAIS requires a smaller number of expansions and thus is faster.

The signal (S3) (Table 5) is a particularly difficult structure for most algorithms because it includes pairs of consecutive change-points that shift the data sequence to opposite directions, leading to masking the existence of the change-points. Detection is more difficult when the change-points are close to each other and their isolation is unlikely or occurs only in a limited number of intervals. The best performing methods are DAIS and MOSUM, having 9% improvement in the number of times the correct number of change-points is detected compared to the other algorithms. The reason why DAIS performs very well is that it will most likely start the expansion around the change-points and so it has an advantage in detecting them, as the power of the contrast function $C^b_{s,e}(\boldsymbol{X})$ is maximized when b is at the midpoint of $[s, e]$. In Tables 6 and 7, the majority of the algorithms perform well; however, WBS_th and MSCP seem to over- and under-estimate, respectively, the number of the change-points. In Table 8, some methods struggle to detect the change-points, which are really close to each other. Based on the simulations, we conclude that overall, DAIS performs very well in various different piecewise-constant signal scenarios, involving the number of change-points, the distance between them, as well as the magnitudes of the changes, when most competitors struggle to accurately detect the changes in some of the cases tested.

Proceeding now with the piecewise-linear framework, the simulations results in Table 9, show that most algorithms perform well, with the exception of MARS and TF. Tables 10, 11, 12 show that in the case of piecewise-linear signals, DAIS performs as well as the algorithms ID_ic and CPOP, and better than other competitors available in the literature. The average computational time for DAIS is very low, and most of the times, significantly lower than the best-performing competitors.

6 Real data

6.1 Crime data

In this section, DAIS is applied to real data and the underlying signal is assumed to be piecewise-constant. The chosen dataset includes daily crime reports and can be found in `https://catalog.data.gov/dataset/crime`. We use daily observations starting from 22/01/2020, up to 31/07/2023. Since the data involve counts of crimes, they are positive integer-valued and thus a transformation is required to bring them closer to Gaussian data with constant variance. This is done using the Anscombe transform, Anscombe [1948], $\alpha : \mathbb{N} \to \mathbb{R}$, with $\alpha(x) = 2\sqrt{x + 3/8}$.

The top plot of Figure 3 is a plot of the true data, before the transformation, along with the estimated underlying signal, $\hat{f}_t$, according to the change-points detected by DAIS, plotted in red. The fits obtained by three other algorithms, ID, MOSUM and NOT, are also shown in the bottom plot of the figure, plotted with blue, yellow and green lines, respectively. All methods obtain different signals on the same data.

Table 4: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signal (S2). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$							
Method	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS	0	99	0	1	0	4.33×10^{-4}	43.10×10^{-4}	0.322
ID_th	0	88	1	10	1	7.77×10^{-4}	35.80×10^{-4}	0.416
ID_ic	0	100	0	0	0	4.25×10^{-4}	1.09×10^{-4}	0.110
WBS_th	0	43	14	24	19	16.00×10^{-4}	186.00×10^{-4}	0.140
WBS_ic	0	100	0	0	0	4.19×10^{-4}	1.06×10^{-4}	0.140
WBS2	0	92	3	4	1	5.46×10^{-4}	17.60×10^{-4}	5.950
PELT	0	100	0	0	0	4.19×10^{-4}	1.06×10^{-4}	0.004
NOT	0	100	0	0	0	4.19×10^{-4}	1.06×10^{-4}	0.103
MOSUM	0	92	2	6	0	5.49×10^{-4}	203.60×10^{-4}	0.049
MSCP	6	93	1	0	0	341.35×10^{-4}	334.74×10^{-4}	276.294

Table 5: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signal (S3). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$								
Method	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS	15	2	79	2	2	0	0.015	0.096	0.004
ID_th	2	1	72	9	13	3	0.015	0.079	0.004
ID_ic	28	0	70	2	0	0	0.016	0.154	0.002
WBS_th	0	0	18	17	19	46	0.028	0.259	0.022
WBS_ic	29	0	70	1	0	0	0.016	0.159	0.022
WBS2	4	3	72	10	6	5	0.015	0.090	0.471
PELT	80	0	20	0	0	0	0.027	0.413	0.001
NOT	30	0	68	2	0	0	0.016	0.164	0.023
MOSUM	10	1	80	5	4	0	0.013	0.083	0.011
MSCP	84	15	1	0	0	0	0.031	0.442	157.422

Table 6: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signal (S4). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$									
Method	≤ -3	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS	0	0	2	95	3	0	0	0.023	0.014	0.002
ID_th	0	0	1	92	7	0	0	0.023	0.014	0.001
ID_ic	0	0	0	87	12	1	0	0.020	0.125	0.002
WBS_th	0	0	0	66	27	6	1	0.025	0.016	0.012
WBS_ic	0	0	0	62	25	7	6	0.025	0.167	0.012
WBS2	0	0	2	92	5	1	0	0.025	0.015	0.035
PELT	0	0	7	93	0	0	0	0.022	0.014	0.001
NOT	0	0	0	93	7	0	0	0.021	0.123	0.047
MOSUM	0	0	2	97	1	0	0	0.018	0.012	0.004
MSCP	100	0	0	0	0	0	0	0.954	0.182	4.830

Table 7: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signals (S5). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$									
Method	≤ -3	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS	0	0	0	96	4	0	0	1.699	0.012	0.002
ID_th	0	0	0	80	14	6	0	1.767	0.018	0.001
ID_ic	0	0	0	90	9	1	0	1.648	0.013	0.002
WBS_th	0	0	0	48	23	11	18	1.797	0.040	0.014
WBS_ic	0	0	0	94	6	0	0	1.485	0.009	0.014
WBS2	0	0	0	94	5	0	1	1.48	0.012	0.106
PELT	0	0	0	100	0	0	0	1.485	0.008	0.001
NOT	0	0	0	96	4	0	0	1.525	0.009	0.021
MOSUM	0	0	0	90	8	2	0	1.513	0.014	0.007
MSCP	47	35	16	2	0	0	0	10.730	0.108	19.422

Table 8: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signal (S6). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$					
Method	-99	$[-98, -11]$	$[-10, 10]$	MSE	d_H	Time (s)
DAIS	0	5	95	0.435	0.011	0.016
ID_th	0	0	100	0.314	0.004	0.009
ID_ic	98	2	0	4.000	0.990	0.033
WBS_th	0	4	96	0.549	0.011	0.025
WBS_ic	100	0	0	4.000	0.990	0.025
WBS2	0	0	100	0.280	0.003	0.339
PELT	2	93	5	2.490	0.251	0.007
NOT	100	0	0	4.000	0.990	0.024
MOSUM	100	0	0	4.000	0.990	0.003
MSCP	0	100	0	3.944	0.093	92.196

Table 9: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signal (S7). The average MSE and computational time are also given.

	$\hat{N} - N$					
Method	0	1	2	≥ 3	MSE	Time (s)
DAIS	100	0	0	0	1.95×10^{-3}	0.011
ID_th	83	10	4	3	2.94×10^{-3}	0.030
ID_ic	100	0	0	0	1.95×10^{-3}	0.007
CPOP	100	0	0	0	1.95×10^{-3}	2.830
NOT	100	0	0	0	1.95×10^{-3}	0.062
MARS	0	100	0	0	5.22×10^{-3}	0.003
TF	7	44	37	12	2.87×10^{-3}	0.419
TS	100	0	0	0	1.95×10^{-3}	0.502

Table 10: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signal (S8). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$							
Method	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS	0	99	1	0	0	0.032	0.099	0.009
ID_th	0	76	20	4	0	0.039	0.153	0.01
ID_ic	0	96	4	0	0	0.030	0.113	0.007
CPOP	0	99	1	0	0	0.013	0.056	5.390
NOT	0	99	1	0	0	0.016	0.063	0.185
MARS	2	9	42	39	8	0.034	0.200	0.006
TF	0	0	0	0	100	0.018	0.451	0.736
TS	1	95	4	0	0	0.095	0.185	0.854

Table 11: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signal (S9). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$									
Method	≤ -15	$(-15, -2]$	-1	0	1	$[2, 15)$	≥ 15	MSE	d_H	Time (s)
DAIS	0	0	1	99	0	0	0	0.262	0.292	0.016
ID_th	0	0	0	69	26	5	0	0.269	0.317	0.013
ID_ic	0	0	0	97	3	0	0	0.201	0.245	0.396
CPOP	0	0	0	97	3	0	0	0.388	0.235	0.445
NOT	100	0	0	0	0	0	0	4.730	99.000	0.075
MARS	100	0	0	0	0	0	0	4.700	98.500	0.004
TF	0	0	0	0	0	0	100	0.205	0.387	0.617
TS	0	36	32	32	0	0	0	0.858	0.579	1.380

Table 12: Distribution of $\hat{N} - N$ of 100 simulated data sequences of (S10). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$									
Method	≤ -15	$(-15, -2]$	-1	0	1	$[2, 15)$	≥ 15	MSE	d_H	Time (s)
DAIS	0	0	0	100	0	0	0	0.038	0.203	0.016
ID_th	0	0	0	86	14	0	0	0.042	0.270	0.011
ID_ic	1	4	17	78	0	0	0	0.106	1.720	0.353
CPOP	0	0	0	98	2	0	0	0.252	0.294	0.072
NOT	100	0	0	0	0	0	0	1.060	119.000	0.048
MARS	100	0	0	0	0	0	0	1.060	118.000	0.003
TF	0	0	0	0	0	3	97	0.190	0.324	0.299
TS	100	0	0	0	0	0	0	1.060	119.000	0.478

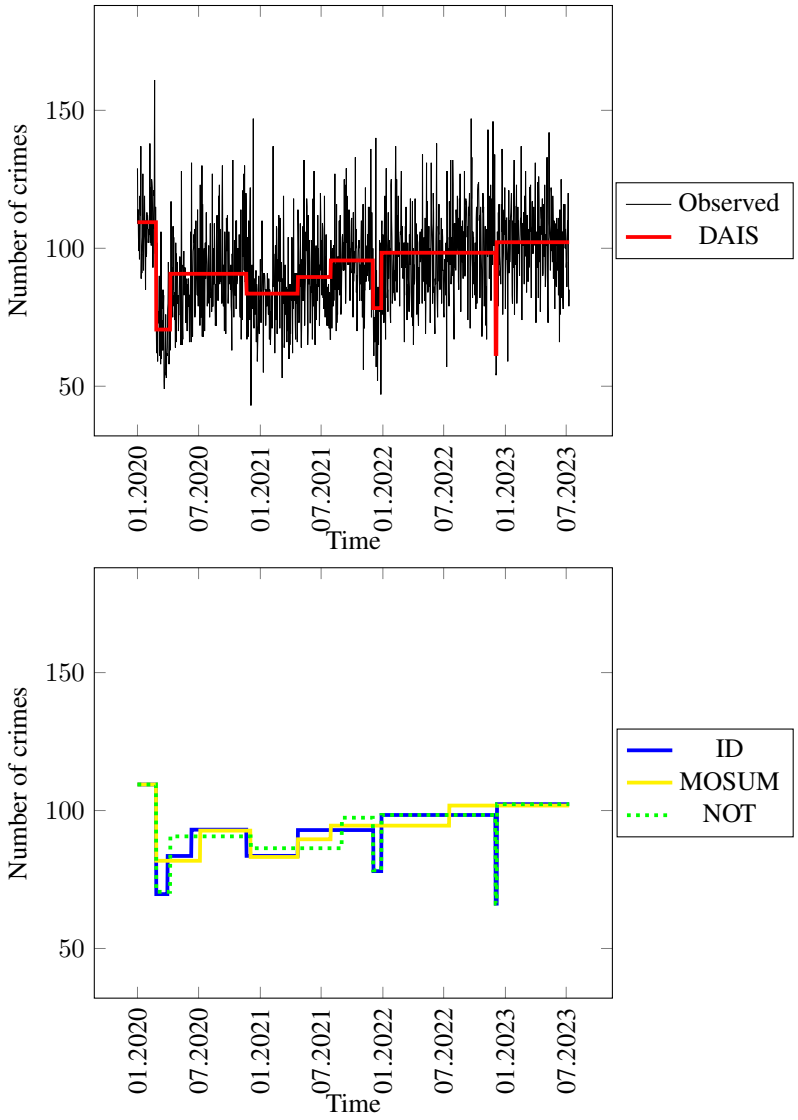

Figure 3: The observed time series data of daily crime reports in Montgomery County and fitted piecewise-constant mean signal obtained by applying DAIS (red), ID (blue), MOSUM (yellow) and NOT (green) to the data.

We now attempt to provide a possible explanation about the change-points estimated by DAIS that express the most important movements in the data. The first change-point occurs on the 17^{th} of March 2020 and corresponds to the first days of positive cases of COVID-19 when the first rules were imposed to limit the spread of the virus among the community in the state of Maryland, where Montgomery County is located. As expected, since people were forced to stay at home, the number of crimes reported dropped significantly. The second change-point is on the 28^{th} of April 2020, when the official authorities started lifting the restrictions and the number of crimes increased. Some small decrease around the 11^{th} of December 2020 can be explained by people's behavioral change around the holidays. The number of crimes returns close to the previous levels, on the 13^{th} of May 2021 and a further increase is observed at the end of the summer, 19^{th} of August. This could possibly be due to people going on holidays. As has also been observed for 2020, there is again a reduced number of reported crimes around the Christmas holidays of 2021, starting from the 23^{th} of December 2021 until the 17^{th} of January 2022. It is notable that the next change-point detected is almost a year later, as there is a sudden decrease on the 23^{rd} of December 2022, with an increase on the 26^{th}.

It is important to note that there is a clear trend as change-points were detected by DAIS on 11/12/2020, 23/12/2021 and 23/12/2022. All three are near the Christmas holidays and a decrease in the daily number of crimes reported is noticed, followed by the trend increasing on the 13/05/2021, 17/01/2022 and 26/12/2023. ID detects the same movements in

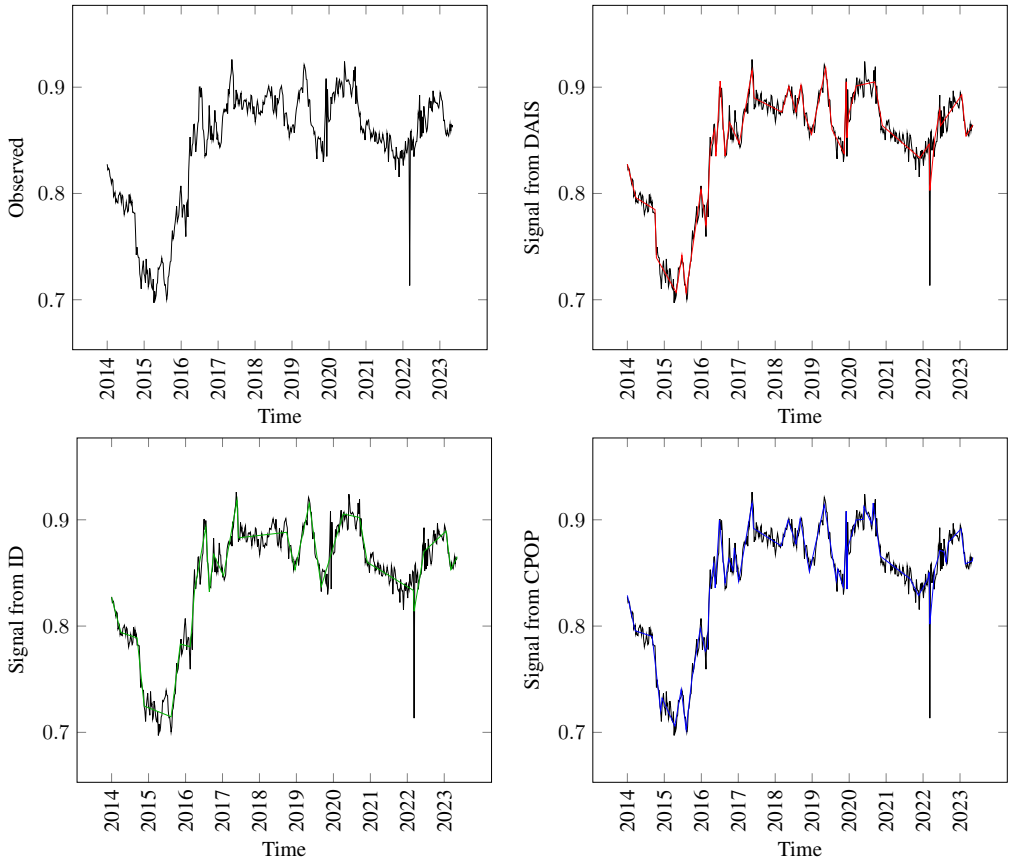

Figure 4: The observed time series data of Euro to British pound sterling exchange rate and fitted piecewise-linear signal obtained by applying DAIS (red), ID (blue) and CPOP (green).

the data, detecting all 6 change-points mentioned above, while NOT only misses the change on 13/05/2021. MOSUM detects a change-point close to Christmas of 2020 only. Interestingly, when the same data were used up to 15/06/2023, ID and NOT were only able to detect the change-points around Christmas 2020 and 2021. Using less data DAIS was able to detect the same change-points that two well-established methods managed to identify only after the addition of more information. It seems that the performance of DAIS is at least as good as that of some of the best-performing methods available in the literature.

6.2 Euro to British pound exchange rate

The DAIS algorithm for changes in the slope can be applied to the Euro to British pound exchange rate. The data are weekly Close prices adjusted for splits for the period between 07/04/2014 and 11/08/2023. The results can be seen in Figure 4. The first plot is the observed signal, while the other plots are the estimated signals from DAIS, ID and CPOP. All four plots have the same range of values on the y-axis so that they are comparable.

All three fits are very similar, with ID detecting the least number of change-points (25) and thus having a smoother signal, while DAIS and CPOP detect 41 and 57, respectively. From the plots it can be seen that around the start of 2016, there is an increase in the value of the exchange rate. This could be explained by the fact that 2016 was the year of the referendum where Brexit was voted by the citizens. It is worth noting that the drop in March of 2020, that is only detected by DAIS, coincides with the day that the Prime Minister Boris Johnson announced the Government would be implementing measures intended to halt the spread of the COVID-19 virus. The big drop around June of 2022 falls around the time when fuel prices soared, according to `https://www.theguardian.com/theguardian/2022/jun/09`.

7 Conclusions

In this project, DAIS is introduced, which is a new data-adaptive method for detecting structural changes in a given data sequence. The first step of the method is to identify the location of the largest difference in the sequence, as defined in (2) for the cases of changes in the mean and changes in the slope, in an effort to start searching from an area around the most apparent change-point in the case that there is one. Expanding intervals around the location of the largest difference are used to achieve isolation of the change-points, which enhances the detection power. The theoretical results regarding the consistency of the number of change-points detected and the accuracy of their estimated locations are given in Section 3. The simulation results presented in Section 5, indicate that DAIS is at least as accurate as the state-of-the-art competitors, with its data-adaptive nature being advantageous regarding the detection of the true change-points in difficult structures, for example when the change-points are really close to each other. DAIS uses intervals around the location of the largest difference and so the change-points are, with high probability, near the midpoint of these intervals, which results in the power of the contrast function being maximized. As explained in Sections 2.3 and 3, for the proof of the theoretical properties of DAIS we require that the location of the largest difference is at a location where detection, while the change-point is isolated in an interval, is guaranteed. However, for the correct choice of the parameters, the points that are problematic are just the midpoints between two consecutive change-points. Finally, in practice, the algorithm performs at least as well as other widely-used algorithms in the literature.

A Simulation Study

The signals used in the simulations of Section 5 are reported below. For the case of piecewise-constant the signals f_t in (1) are the following:

(S1) $justnoise$: sequence of length 6000 with no change-points. The standard deviation is $\sigma = 1$.

(S2) $long\_signal$: sequence of length 11000 with 1 change-point at 5500 with values before and after the change-point 0 and 1.5. The standard deviation is $\sigma = 1$.

(S3) $small\_dist$: sequence of length 1000 with 2 change-points at 485 and 515 with values between change-points 0, 1, 0. The standard deviation is $\sigma = 1$.

(S4) $stairs$: piecewise-constant signal of length 150 with 14 change-points at 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140 with values between change-points 1, 2, 3, . . ., 15. The standard deviation is $\sigma = 0.3$.

(S5) mix: piecewise-constant signal of length 301 with 9 change-points at 11, 21, 41, 61, 91, 121, 161, 201, 251 with values between change-points 7, -7, 6, -6, 5, -5, 4, -4, 3, -3. The standard deviation is $\sigma = 4$.

(S6) $many\_cpts$: piecewise-constant signal of length 700 with 99 change-points at 7, 14, . . ., 693 with values between change-points 0, 4, 0, 4, . . ., 0, 4. The standard deviation is $\sigma = 1$.

The signals used for the case of piecewise-linear f_t are:

(S7) $justnoise\_wave$: piecewise-linear signal without change-points of length 1000. The starting intercept is $f_1 = 0$ and slope $f_2 - f_1 = 1$. The standard deviation is $\sigma = 1$.

(S8) $wave1$: continuous piecewise-linear signal of length 1408 with 7 change-points at 256, 512, 768, 1024, 1152, 1280, 1344 with changes in slope $-1/64$, $2/64$, $-3/64$, $4/64$, $-5/64$, $6/64$, $-7/64$. The starting intercept is $f_1 = 1$ and slope $f_2 - f_1 = 1/256$. The standard deviation is $\sigma = 1$.

(S9) $wave2$: continuous piecewise-linear signal of length 1500 with 99 change-points at 15, 30, . . ., 1485 with changes in slope -1, 1, -1 . . ., -1. The starting intercept is $f_1 = -1/2$ and slope $f_2 - f_1 = 1/40$. The standard deviation is $\sigma = 1$.

(S10) $wave3$: continuous piecewise-linear signal of length 840 with 119 change-points at 7, 14, 21, . . ., 833 with changes in slope -1, 1, -1 . . ., -1. The starting intercept is $f_1 = -1/2$ and slope $f_2 - f_1 = 1/32$. The standard deviation is $\sigma = 0.3$.

B Proof of Theorem 1

Before proving Theorem 1, we introduce some more notation. Similar to (9), we denote the CUSUM of the unobserved signal as $\tilde{f}^b_{s,e} = \sqrt{\frac{e-b}{\ell(b-s+1)}} \sum_{t=s}^{b} f_t - \sqrt{\frac{b-s+1}{\ell(e-b)}} \sum_{t=b+1}^{e} f_t$. For the following proof, the contrast vector

$\boldsymbol{\psi}^{\boldsymbol{b}}_{\boldsymbol{s,e}} = \left(\psi^b_{s,e}(1), \psi^b_{s,e}(2), \ldots, \psi^b_{s,e}(T_1)\right)$ is defined through the contrast function

$$\psi^b_{s,e}(t) = \begin{cases} \sqrt{\frac{e-b}{\ell(b-s+1)}}, & t = s, s+1, \ldots, b, \\ -\sqrt{\frac{b-s+1}{\ell(e-b)}}, & t = b+1, b+2, \ldots, e, \\ 0, & \text{otherwise} \end{cases}, \tag{18}$$

where $s \leq b < e$ and $\ell = e - s + 1$. Notice that for any vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_{T_1})$, we have that $\langle \boldsymbol{v}, \boldsymbol{\psi}^{\boldsymbol{b}}_{\boldsymbol{s,e}} \rangle = \tilde{v}^b_{s,e}$.

For the proof of Theorem 1, we require the following Lemma.

Lemma 1. *Suppose $\boldsymbol{f} = (f_1, f_2, \ldots, f_T)^T$ is a piecewise-constant vector. Pick any interval $[s, e] \subset [1, T]$ such that $[s, e-1]$ contains exactly one change-point r. Let $\rho = |r - b|$, $\Delta^f = |f_{r+1} - f_r|$, $\eta_L = r - s + 1$ and $\eta_R = e - r$. Then,*

$$\left\| \boldsymbol{\psi}^{\boldsymbol{b}}_{\boldsymbol{s,e}} \langle \boldsymbol{f}, \boldsymbol{\psi}^{\boldsymbol{b}}_{\boldsymbol{s,e}} \rangle - \boldsymbol{\psi}^{\boldsymbol{r}}_{\boldsymbol{s,e}} \langle \boldsymbol{f}, \boldsymbol{\psi}^{\boldsymbol{r}}_{\boldsymbol{s,e}} \rangle \right\|_2^2 = \left(\tilde{f}^r_{s,e}\right)^2 - \left(\tilde{f}^b_{s,e}\right)^2.$$

In addition,

1. *for any $r \leq b < e$, $\left(\tilde{f}^r_{s,e}\right)^2 - \left(\tilde{f}^b_{s,e}\right)^2 = (\rho\eta_L/(\rho + \eta_L)) \left(\Delta^f\right)^2$;*
2. *for any $s \leq b < r$, $\left(\tilde{f}^r_{s,e}\right)^2 - \left(\tilde{f}^b_{s,e}\right)^2 = (\rho\eta_R/(\rho + \eta_R)) \left(\Delta^f\right)^2$.*

Proof. See Lemma 4 from Baranowski et al. [2019]. □

The proof of Theorem 1 consists of 6 steps. Step 1 is to show that the observed $|\tilde{X}^b_{s,e}|$ is uniformly close to the unobserved $|\tilde{f}^b_{s,e}|$ for all $1 \leq s \leq b < e \leq T$. This will allow us to extend some results that will be derived for the signal, f_t, to the data sequence, X_t, in which we are interested. In Step 2, we control the distance between $|\tilde{X}^{b_1}_{s,e}| - |\tilde{X}^{b_2}_{s,e}|$ and $|\tilde{f}^{b_1}_{s,e}| - |\tilde{f}^{b_2}_{s,e}|$ for all possible combinations of s, e, b_1, b_2, where $1 \leq s < e \leq T$ and $b_1, b_2 \in [s, e)$. In Step 3, we show that it suffices to restrict the proof to an interval with a single change-point because each change-point will be isolated in an interval where detection will occur with high probability, as discussed in Section 2.3. In this step, we also show that the estimated change-point $\hat{r}_j$ will be close to the actual change-point r_j. Since after detection DAIS restarts in intervals with end- (or start-) point the start- (or end-) point of the interval where the detection occurred, in Step 4 we prove that there is no change-point, besides the detected one, in the intervals that are skipped, with probability 1. In Step 5, we show that the new intervals used after detection allow for the detection of all the remaining change-points. Finally, in Step 6 we show that when there is no change-point in the interval being checked, the algorithm will not have any false detections and will terminate.

Proof. We will prove the more specific result

$$\mathbb{P}\left(\hat{N} = N, \max_{j=1,2,\ldots,N} \left(|\hat{r}_j - r_j| \left(\Delta^f_j\right)^2 \right) \leq C_3 \log T \right) \geq 1 - \frac{1}{6\sqrt{\pi}T}, \tag{19}$$

which implies result (10).

Step 1: Allow us to denote by

$$A_T = \left\{ \max_{s,b,e:1 \leq s \leq b < e \leq T} \left| \tilde{X}^b_{s,e} - \tilde{f}^b_{s,e} \right| \leq \sqrt{8 \log T} \right\}. \tag{20}$$

We will show that $\mathbb{P}(A_T) \geq 1 - 1/(12\sqrt{\pi}T)$. From (8) and (9), simple steps yield $\tilde{X}^b_{s,e} - \tilde{f}^b_{s,e} = \tilde{\epsilon}^b_{s,e}$, where $\tilde{\epsilon}^b_{s,e} \sim \mathcal{N}(0,1)$. Thus, for $Z \sim \mathcal{N}(0,1)$, using the Bonferroni inequality we get that

$$\begin{aligned}\mathbb{P}\Big((A_T)^c\Big) &= \mathbb{P}\left(\max_{s,b,e:1\leq s\leq b<e\leq T}\left|\tilde{X}^b_{s,e} - \tilde{f}^b_{s,e}\right| > \sqrt{8\log T}\right)\\ &\leq \sum_{1\leq s\leq b<e\leq T}\mathbb{P}\left(\left|\tilde{\epsilon}^b_{s,e}\right| > \sqrt{8\log T}\right) \leq \frac{T^3}{6}\mathbb{P}(|Z| > \sqrt{8\log T})\\ &= \frac{T^3}{3}\mathbb{P}\left(Z > \sqrt{8\log T}\right) \leq \frac{T^3}{3}\frac{\phi(\sqrt{8\log T})}{\sqrt{8\log T}} \leq \frac{1}{12\sqrt{\pi}T},\end{aligned}$$

where $\phi(\cdot)$ is the probability density function of the standard normal distribution.

Step 2: For intervals $[s,e)$ that contain only one true change-point r, for $\boldsymbol{\psi^b_{s,e}}$ as defined in (18), we denote by

$$B_T = \left\{\max_{1\leq s\leq b<e<T}\max_{\substack{r_{j-1}<s\leq r_j\\ r_j<e\leq r_{j+1}\\ s\leq b<e}}\frac{\left|\left\langle \boldsymbol{\psi^b_{s,e}}\langle \boldsymbol{f},\boldsymbol{\psi^b_{s,e}}\rangle - \boldsymbol{\psi^r_{s,e}}\langle \boldsymbol{f},\boldsymbol{\psi^r_{s,e}}\rangle, \boldsymbol{\epsilon}\right\rangle\right|}{\|\boldsymbol{\psi^b_{s,e}}\langle \boldsymbol{f},\boldsymbol{\psi^b_{s,e}}\rangle - \boldsymbol{\psi^r_{s,e}}\langle \boldsymbol{f},\boldsymbol{\psi^r_{s,e}}\rangle\|_2} \leq \sqrt{8\log T}\right\}. \tag{21}$$

Because

$$\left|\left\langle \boldsymbol{\psi^b_{s,e}}\langle \boldsymbol{f},\boldsymbol{\psi^b_{s,e}}\rangle - \boldsymbol{\psi^{r_j}_{s,e}}\langle \boldsymbol{f},\boldsymbol{\psi^{r_j}_{s,e}}\rangle\right\rangle\right| \Big/ \left(\|\boldsymbol{\psi^b_{s,e}}\langle \boldsymbol{f},\boldsymbol{\psi^b_{s,e}}\rangle - \boldsymbol{\psi^{r_j}_{s,e}}\langle \boldsymbol{f},\boldsymbol{\psi^{r_j}_{s,e}}\rangle\|_2\right)$$

follows the standard normal distribution, we use a similar approach as in Step 1, to show that $\mathbb{P}\Big((B_T)^c\Big) \leq 1/12\sqrt{\pi}T$. Therefore, Step 1 and Step 2 lead to

$$\mathbb{P}(A_T \cap B_T) \geq 1 - \frac{1}{6\sqrt{\pi}T}.$$

Step 3: This is the main part of our proof. From now on, we assume that A_T and B_T both hold. The constants we use are

$$C_1 = \sqrt{C_3} + \sqrt{8}, C_2 = \frac{1}{\sqrt{4n}} - \frac{2\sqrt{2}}{\underline{C}}, C_3 = 2(2\sqrt{2}+4)^2, \tag{22}$$

where $\underline{C}$ satisfies assumption (A2), $\sqrt{\delta_T}\underline{f}_T \geq \underline{C}\sqrt{\log T}$ and $3/2 \leq n \leq \delta_T/2\lambda_T$ as defined in Section 2.3. For $j \in \{1,2,\ldots,N\}$ define I^L_j and I^R_j as in (5). The location of the largest difference in an interval $[s,e]$, $1 \leq s < e \leq T$, is defined as $d_{s,e} = \text{argmax}_{t\in\{s,s+1,\ldots,e-1\}}\{|X_{t+1} - X_t|\}$ as in (2) and for $K^l = \lceil\frac{d_{s,e}-s+1}{\lambda_T}\rceil$, $K^r = \lceil\frac{e-d_{s,e}+1}{\lambda_T}\rceil$, $K^{\max} = \max\{K^l, K^r\}$, define

$$\begin{aligned}c^l_m &= \max\{d_{s,e} - m\lambda_T, s\}, \quad m = 0,1,\ldots,K^{\max},\\ c^r_k &= \min\{d_{s,e} + k\lambda_T - 1, e\}, \quad k = 1,2,\ldots,K^{\max}.\end{aligned} \tag{23}$$

Since the length of the intervals in (5) is $\delta_T/2n$ and $\lambda_T \leq \delta_T/2n$, for $m,k \in \{0,1,\ldots,K^{\max}\}$ we ensure that there exists at least one m and at least one k such that $c^l_m \in I^L_j$ and $c^r_k \in I^R_j$ for all $j \in \{1,2,\ldots,N\}$ and for all possible locations of the largest difference.

At the beginning of DAIS, $s = 1, e = T$ and the first change-point that will get detected depends on the value of $d_{1,T}$. As already explained in Section 2.3, the largest difference $d_{s,e}$ will be at most at a distance $\delta^j_{1,T}$ from the nearest change-point, where $\delta^j_{1,T} \leq \frac{\delta_j}{2} - \frac{3\delta_T}{4n}$ for $\delta_j = |r_{j+1} - r_j|$. The first point to get detected will be the point that is closest to the largest difference $d_{1,T}$. The interval where the detection of this change-point occurs, cannot contain more than one change-points, as was explained in Section 2.3 and is proved at (24).

Without loss of generality, we suppose that the first change-point to get detected is r_J for $J \in \{1,2,\ldots,N\}$. We will show that there exists an interval $[c^l_m, c^r_k]$, for $m,k \in \{0,1,\ldots,K^{\max}\}$, such that r_J is isolated assuming that $d_{1,T} > r_J$ and the same approach works for $d_{1,T} \leq r_J$. Note that since the closest change-point to the largest difference is r_J, it must hold that $d^J_{1,T} = d_{1,T} - r_J$. We are now considering the largest possible interval that may be checked before detection occurs in the sense that the largest number of expansions will be performed. Detection is guaranteed when checking this interval, but it may occur in any interval smaller than this. Showing that the change-point will be isolated in the largest interval means that it will always be isolated. Since we are considering $d_{1,T} > r_J$, the worst

case scenario occurs for $m = k$, $c^l_m \in I^L_J$ and $c^r_k - r_J > \delta_T/2n$, but it doesn't necessarily hold that $c^r_k \in I^R_J$. So, $r_J - c^l_m \leq \delta_T/n$ and $c^r_k - d_{1,T} + 1 = d_{1,T} - c^l_m$. It follows that

$$\begin{aligned} c^r_k - c^l_m &= c^r_k - d_{1,T} + d_{1,T} - c^l_m = 2(d_{1,T} - c^l_m) - 1 < 2(d_{1,T} - r_J + r_J - c^l_m) \\ &< 2\Big(\delta^J_{1,T} + \frac{\delta_T}{n}\Big) \leq \delta_J + \frac{\delta_T}{2n}. \end{aligned} \tag{24}$$

Now, since $r_J - c^l_m \leq \delta_T/n$, there can be at most one change-point in the interval $[c^l_m, r_J]$. Considering the interval $[r_J, c^r_k]$, using (24) and $r_J - c^l_m > \delta_T/2n$, it holds that $c^r_k - r_J = c^r_k - c^l_m + c^l_m - r_J < \delta_J$, so $c^r_k < r_{J+1}$ and r_J is isolated in $[r_J, c^r_k]$. The two results combined prove that the change-point is isolated in the interval $[c^l_m, c^r_k]$.

We will now show that for $\tilde{b} = \text{argmax}_{c^l_m \leq t < c^r_k} |\tilde{X}^t_{c^l_m, c^r_k}|$, it holds that $\left|\tilde{X}^{\tilde{b}}_{c^l_m, c^r_k}\right| > \zeta_T$. Using (8), we have that

$$\left|\tilde{X}^{\tilde{b}}_{c^l_m, c^r_k}\right| \geq \left|\tilde{X}^{r_J}_{c^l_m, c^r_k}\right| \geq \left|\tilde{f}^{r_J}_{c^l_m, c^r_k}\right| - \sqrt{8 \log T}. \tag{25}$$

But,

$$\begin{aligned} \left|\tilde{f}^{r_J}_{c^l_m, c^r_k}\right| &= \left| \sqrt{\frac{c^r_k - r_J}{(c^r_k - c^l_m + 1)(r_J - c^l_m + 1)}} (r_J - c^l_m + 1) f_{r_J} \right. \qquad (26) \\ &\qquad \left. - \sqrt{\frac{r_J - c^l_m + 1}{(c^r_k - c^l_m + 1)(c^r_k - r_J)}} (c^r_k - r_J) f_{r_J+1} \right| \\ &= \left| \sqrt{\frac{(c^r_k - r_J)(r_J - c^l_m + 1)}{(c^r_k - c^l_m + 1)}} f_{r_J} - \sqrt{\frac{(r_J - c^l_m + 1)(c^r_k - r_J)}{(c^r_k - c^l_m + 1)}} f_{r_J+1} \right| \\ &= \sqrt{\frac{(c^r_k - r_J)(r_J - c^l_m + 1)}{c^r_k - c^l_m + 1}} \Delta^f_J \geq \sqrt{\frac{(c^r_k - r_J)(r_J - c^l_m + 1)}{2 \max\{c^r_k - r_J, r_J - c^l_m + 1\}}} \Delta^f_J \\ &= \sqrt{\frac{\min\{c^r_k - r_J, r_J - c^l_m + 1\}}{2}} \Delta^f_J. \qquad (27) \end{aligned}$$

If we show that

$$\min\{c^r_k - r_J, r_J - c^l_m + 1\} \geq \frac{\delta_T}{2n} \tag{28}$$

then using assumption (A2) and the results in (25) (26), (28), we have that

$$\begin{aligned} \left|\tilde{X}^{\tilde{b}}_{c^l_m, c^r_k}\right| &\geq \sqrt{\frac{\delta_T}{4n}} \Delta^f_J - \sqrt{8 \log T} \geq \sqrt{\frac{\delta_T}{4n}} \underline{f_T} - \sqrt{8 \log T} \\ &= \left(\frac{1}{\sqrt{4n}} - \frac{2\sqrt{2 \log T}}{\sqrt{\delta_T} \underline{f}_T} \right) \sqrt{\delta_T} \underline{f}_T \geq \left(\frac{1}{\sqrt{4n}} - \frac{2\sqrt{2}}{\underline{C}} \right) \sqrt{\delta_T} \underline{f}_T \\ &= C_2 \sqrt{\delta_T} \underline{f}_T > \zeta_T \end{aligned} \tag{29}$$

and thus, the change-point will get detected.

But (28) holds as, for $m = k$, it holds that $c^r_k - d_{1,T} + 1 = d_{1,T} - c^l_m$ and so

- If $d_{1,T} \leq r_J$, then $c^r_k \in I^R_J$ and so $c^r_k - r_J > \delta_T/2n$ and $r_J - c^l_m + 1 > r_J - d_{1,T} + d_{1,T} - c^l_m + 1 \geq c^r_k - r_J > \delta_T/2n$.
- If $d_{1,T} > r_J$, then $c^l_m \in I^L_J$ and so $r_J - c^l_m + 1 > \delta_T/2n$ and $c^r_k - r_J = c^r_k - d_{1,T} + d_{1,T} - r_J > d_{1,T} - c^l_m - 1 \geq r_J - c^l_m \geq \delta_T/2n$.

Therefore, we have proved that there will be an interval of the form $[c^l_{\tilde{m}}, c^r_{\tilde{k}}]$, such that the interval contains r_J and no other change-point and $\max_{c^l_{\tilde{m}} \leq t < c^r_{\tilde{k}}} \left|\tilde{X}^t_{c^l_{\tilde{m}}, c^r_{\tilde{k}}}\right| > \zeta_T$. For $k^* \in \{0, 1, \ldots K^{\max}\}$ and $m^* \in \{k^* - 1, k^*\}$, denote by $c^l_{m^*} \geq c^l_{\tilde{m}}$ and $c^r_{k^*} \leq c^r_{\tilde{k}}$ the first left- and right-expanding points, respectively, that this happens and

let $b_J = \text{argmax}_{c^l_{m^*} \le t < c^r_{k^*}} \left|\tilde{X}^t_{c^l_{m^*}, c^r_{k^*}}\right|$, with $\left|\tilde{X}^{b_J}_{c^l_{m^*}, c^r_{k^*}}\right| > \zeta_T$. Note that b_J cannot be an estimation of $r_j, j \neq J$, as r_J is isolated in the interval where it is detected. Our aim now is to find $\gamma_T > 0$, such that for any $b^* \in \{c^l_{m^*}, c^l_{m^*}+1, \ldots, c^r_{k^*}-1\}$ with $|b^* - r_J|\left(\Delta^f_J\right)^2 > \gamma_T$, we have that

$$\left(\tilde{X}^{r_J}_{c^l_{m^*}, c^r_{k^*}}\right)^2 > \left(\tilde{X}^{b^*}_{c^l_{m^*}, c^r_{k^*}}\right)^2. \tag{30}$$

Proving (30) and using the definition of b_J, we can conclude that $|b_J - r_J|\left(\Delta^f_J\right)^2 \le \gamma_T$. Now, using (8), it can be shown, for $\boldsymbol{\psi}^{\boldsymbol{b}}_{\boldsymbol{s},\boldsymbol{e}}$ as defined in (18), that (30) is equivalent to

$$\begin{aligned}\left(\tilde{f}^{r_J}_{c^l_{m^*}, c^r_{k^*}}\right)^2 - \left(\tilde{f}^{b^*}_{c^l_{m^*}, c^r_{k^*}}\right)^2 > \left(\tilde{\epsilon}^{b^*}_{c^l_{m^*}, c^r_{k^*}}\right)^2 - \left(\tilde{\epsilon}^{r_J}_{c^l_{m^*}, c^r_{k^*}}\right)^2 \\ + 2\left\langle \boldsymbol{\psi}^{b^*}_{c^l_{m^*}, c^r_{k^*}} \langle \boldsymbol{f}, \boldsymbol{\psi}^{b^*}_{c^l_{m^*}, c^r_{k^*}} \rangle - \boldsymbol{\psi}^{r_J}_{c^l_{m^*}, c^r_{k^*}} \langle \boldsymbol{f}, \boldsymbol{\psi}^{r_J}_{c^l_{m^*}, c^r_{k^*}} \rangle, \boldsymbol{\epsilon} \right\rangle.\end{aligned} \tag{31}$$

Without loss of generality, assume that $b^* \in [r_J, c^r_{k^*})$ and a similar approach holds when $b^* \in [c^l_{m^*}, r_J)$. Using Lemma 1, we have that for the left-hand side of (31),

$$\left(\tilde{f}^{r_J}_{c^l_{m^*}, c^r_{k^*}}\right)^2 - \left(\tilde{f}^{b^*}_{c^l_{m^*}, c^r_{k^*}}\right)^2 = \frac{|r_J - b^*|(r_J - c^l_{m^*}+1)}{|r_J - b^*| + (r_J - c^l_{m^*}+1)}\left(\Delta^f_J\right)^2 :\ = \Lambda. \tag{32}$$

Also, for the right-hand side

$$\left(\tilde{\epsilon}^{b^*}_{c^l_{m^*}, c^r_{k^*}}\right)^2 - \left(\tilde{\epsilon}^{r_J}_{c^l_{m^*}, c^r_{k^*}}\right)^2 \le \max_{s,e,b: s \le b < e} \left(\tilde{\epsilon}^b_{s,e}\right)^2 \le 8 \log T \tag{33}$$

and using Lemma 1 and using (21),

$$\begin{aligned}2\left\langle \boldsymbol{\psi}^{b^*}_{c^l_{m^*}, c^r_{k^*}} \langle \boldsymbol{f}, \boldsymbol{\psi}^{b^*}_{c^l_{m^*}, c^r_{k^*}} \rangle - \boldsymbol{\psi}^{r_J}_{c^l_{m^*}, c^r_{k^*}} \langle \boldsymbol{f}, \boldsymbol{\psi}^{r_J}_{c^l_{m^*}, c^r_{k^*}} \rangle, \boldsymbol{\epsilon} \right\rangle \\ \le 2\left\| \boldsymbol{\psi}^{b^*}_{c^l_{m^*}, c^r_{k^*}} \langle \boldsymbol{f}, \boldsymbol{\psi}^{b^*}_{c^l_{m^*}, c^r_{k^*}} \rangle - \boldsymbol{\psi}^{r_J}_{c^l_{m^*}, c^r_{k^*}} \langle \boldsymbol{f}, \boldsymbol{\psi}^{r_J}_{c^l_{m^*}, c^r_{k^*}} \rangle \right\|_2 \sqrt{8 \log T} \\ = 2\sqrt{\Lambda}\sqrt{8 \log T}.\end{aligned} \tag{34}$$

Using (32), (33) and (34), we can conclude that (31) is satisfied if $\Lambda > 8 \log T + \sqrt{2}\sqrt{\Lambda}\sqrt{8 \log T}$ is satisfied, which has solution

$$\Lambda > \left(2\sqrt{2} + 4\right)^2 \log T.$$

From (32) and since

$$\begin{aligned}\frac{|r_J - b^*|(r_J - c^l_{m^*}+1)}{|r_J - b^*| + r_J - c^l_{m^*} + 1} &\ge \frac{|r_J - b^*|(r_J - c^l_{m^*}+1)}{2\max\{|r_J - b^*|, r_J - c^l_{m^*}+1\}} \\ &= \frac{\min\{|r_J - b^*|, r_J - c^l_{m^*}+1\}}{2},\end{aligned}$$

we can conclude that

$$\min\{|r_J - b^*|, r_J - c^l_{m^*}+1\} > \frac{2\left(2\sqrt{2}+4\right)^2 \log T}{\left(\Delta^f_J\right)^2} = C_3 \frac{\log T}{\left(\Delta^f_J\right)^2} \tag{35}$$

implies (30). Now, if

$$\min\{r_J - c^l_{m^*}+1, c^r_{k^*} - r_J\} > C_3 \frac{\log T}{\left(\Delta^f_J\right)^2}, \tag{36}$$

then (35) is restricted to $|r_J - b^*|\left(\Delta^f_J\right)^2 > C_3 \log T$ and this implies (30). So, we conclude that necessarily

$$|r_J - b_J|\left(\Delta^f_J\right)^2 \le C_3 \log T. \tag{37}$$

But (36) must be true since if we assume that $\min\{r_J - c^l_{m^*} + 1, c^r_{k^*} - r_J\} \leq C_3 \log T / \left(\Delta^f_J\right)^2$, then we have that

$$\begin{aligned}
\left|\tilde{X}^{b_J}_{c^l_{m^*}, c^r_{k^*}}\right| &\leq \left|\tilde{f}^{b_J}_{c^l_{m^*}, c^r_{k^*}}\right| + \sqrt{8 \log T} \leq \left|\tilde{f}^{r_J}_{c^l_{m^*}, c^r_{k^*}}\right| + \sqrt{8 \log T} \\
&= \sqrt{\frac{(c^r_{k^*} - r_J)(r_J - c^l_{m^*} + 1)}{c^r_{k^*} - c^l_{m^*} + 1}} \Delta^f_J + \sqrt{8 \log T} \\
&\leq \sqrt{\min\{c^r_{k^*} - r_J, r_J - c^l_{m^*} + 1\}} \Delta^f_J + \sqrt{8 \log T} \\
&\leq \left(\sqrt{C_3} + 2\sqrt{2}\right) \sqrt{\log T} = C_1 \sqrt{\log T} \leq \zeta_T,
\end{aligned}$$

which contradicts $\left|\tilde{X}^{b_J}_{c^l_{m^*}, c^r_{k^*}}\right| > \zeta_T$.

Thus, we have proved that for $\lambda_T \leq \delta_T / 2n$, working under the assumption that both A_T and B_T hold, there will be an interval $[c^l_{m^*}, c^r_{k^*}]$ with $\left|\tilde{X}^{b_J}_{c^l_{m^*}, c^r_{k^*}}\right| > \zeta_T$, where $b_J = \text{argmax}_{c^l_{m^*} \leq t < c^r_{k^*}} \left|\tilde{X}^t_{c^l_{m^*}, c^r_{k^*}}\right|$ is the estimated location for the change-point r_J that satisfies (37).

Step 4: After the detection of the change-point r_J at the estimated location b_J in the interval $[c^l_{m^*}, c^r_{k^*}]$, the process is repeated in the disjoint intervals $[1, c^l_{m^*}]$ and $[c^r_{k^*}, T]$, which contain $r_1, r_2, \ldots, r_{J-1}$ and $r_{J+1}, r_{J+2}, \ldots, r_N$ respectively. This means that we do not check if there are any more change-points in the interval $[c^l_{m^*}, c^r_{k^*}]$, besides the already detected r_J. Since $c^l_{m^*} \geq c^l_{\tilde{m}}$ and $c^r_{k^*} \leq c^r_{\tilde{k}}$, it holds that $[c^l_{m^*}, c^r_{k^*}] \subset [c^l_{\tilde{m}}, c^r_{\tilde{k}}]$ and the argument of (24) can be used to conclude that r_J is isolated in $[c^l_{m^*}, c^r_{k^*}]$.

Step 5: After detecting r_J, the algorithm will first check the interval $[1, c^l_{m^*}]$. So, unless $r_J = r_1$ and $[1, c^l_{m^*}]$ contains no other change-points, the next change-point to get detected will be one of $r_1, r_2, \ldots, r_{J-1}$. The location of the largest difference in the interval $[1, c^l_{m^*}]$, $d_{1, c^l_{m^*}}$, will again determine which change-point will be detected next. If $d_{1, c^l_{m^*}}$ is at a position that the next detected change-point, as explained at the beginning of Step 3, will be one of $r_1, r_2, \ldots, r_{J-2}$, then we can prove exactly as we did for r_J that there will eventually be an interval with end-points far enough from the change-point that allows detection while the change-point is isolated.

Now, we need to discuss the case that the next change-point to get detected is r_{J-1}. This is the closest change-point to the already detected r_J so the end-point of the interval where DAIS is reapplied to depends on the interval where the detection of r_J occurred. As before, for $k_{J-1}, m_{J-1} \in \{0, 1, \ldots, K^{\max}\}$ we will show that r_{J-1} gets detected in $[c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}]$, where $c^l_{m^*_{J-1}} \geq c^l_{m_{J-1}}$ and $c^r_{k^*_{J-1}} \leq c^r_{k_{J-1}} \leq c^l_{m^*}$ and its detection is at location

$$b_{J-1} = \text{argmax}_{c^l_{m^*_{J-1}} \leq t < c^r_{k^*_{J-1}}} \left|\tilde{X}^t_{c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}}\right|,$$

which satisfies $\left|r_{J-1} - b_{J-1}\right| \left(\Delta^f_{J-1}\right)^2 \leq C_3 \log T$. Firstly, r_{J-1} is isolated in the interval $[c^l_{m_{J-1}}, c^r_{k_{J-1}}]$ using the same argument as in (24). Following similar steps as in (26), we have that for $\tilde{b}_{J-1} = \text{argmax}_{c^l_{m_{J-1}} \leq t < c^r_{k_{J-1}}} \left|\tilde{X}^t_{c^l_{m_{J-1}}, c^r_{k_{J-1}}}\right|$,

$$\begin{aligned}
\left|\tilde{X}^{\tilde{b}_{J-1}}_{c^l_{m_{J-1}}, c^r_{k_{J-1}}}\right| &\geq \left|\tilde{f}^{\tilde{r}_{J-1}}_{c^l_{m_{J-1}}, c^r_{k_{J-1}}}\right| - \sqrt{8 \log T} \\
&\geq \sqrt{\frac{\min\{c^r_{k_{J-1}} - r_{J-1}, r_{J-1} - c^l_{m_{J-1}} + 1\}}{2}} \Delta^f_{J-1}. \qquad (38)
\end{aligned}$$

Before we show that $\min\{c^r_{k_{J-1}} - r_{J-1}, r_{J-1} - c^l_{m_{J-1}} + 1\} \geq \delta_T / 2n$, we need show that $c^l_{m^*}$ satisfies $c^l_{m^*} - r_{J-1} \geq \delta_T / 2n$ since $c^l_{m^*} \geq c^r_{k_{J-1}}$. It holds that

$$c^l_{m^*} - r_{J-1} = c^l_{m^*} - r_J + r_J - r_{J-1} \geq c^l_{\tilde{m}} - r_J + \delta_{J-1}$$

We are again concentrating on the worst-case scenario, meaning that both end-points are at least at a distance $\delta_T / 2n$ from the change-point. More precisely, either $c^l_{\tilde{m}} \in I^L_J$ or $c^r_{\tilde{k}} \in I^R_J$, and the other end-point could be further than δ_T / n from the change-point. If $c^l_{\tilde{m}} \in I^L_J$, then, since $n \geq 3/2$, the above inequality becomes

$$c^l_{m^*} - r_{J-1} \geq c^l_{\tilde{m}} - r_J + \delta_{J-1} > -\frac{\delta_T}{n} + 1 + \delta_T > \frac{n-1}{n} \delta_T > \frac{\delta_T}{2n}$$

If instead $c^r_{\tilde{k}} \in I^R_J$, then using (6) and the fact that $d_{1,T} - c^l_{\tilde{m}} < c^r_{\tilde{k}} - d_{1,T}$ due to the way the expansions occur,

$$\begin{aligned} r_J - c^l_{\tilde{m}} &= r_J - d_{1,T} + d_{1,T} - c^l_{\tilde{m}} < \delta^{J-1}_{1,T} + c^r_{\tilde{k}} - d_{1,T} \\ &\leq \delta^{J-1}_{1,T} + c^r_{\tilde{k}} - r_J + r_J - d_{1,T} \leq 2\delta^{J-1}_{1,T} + \frac{\delta_T}{n} \\ &\leq 2\Big(\frac{\delta_{J-1}}{2} - \frac{3\delta_T}{4n}\Big) + \frac{\delta_T}{n} = \delta_{J-1} - \frac{\delta_T}{2n} \end{aligned}$$

and so

$$c^l_{m^*} - r_{J-1} \geq c^l_{\tilde{m}} - r_J + \delta_{J-1} \geq -\delta_{J-1} + \frac{\delta_T}{2n} + \delta_{J-1} = \frac{\delta_T}{2n}$$

This means that the right end-point of the interval could eventually satisfy $c^r_{k_{J-1}} - r_{J-1} \geq \delta_T/2n$ for some k_{J-1} since the largest it can become is $c^r_{k_{J-1}} = c^l_{m^*}$. As in Step 3, the detection could happen before this is satisfied. For the left end-point of the interval, the same holds as in Step 3. In any case, $\min\{c^r_{k_{J-1}} - r_{J-1}, r_{J-1} - c^l_{m_{J-1}} + 1\} \geq \delta_T/2n$ holds and so, from (38) we have that

$$\begin{aligned} \Big|\tilde{X}^{\tilde{b}_{J-1}}_{c^l_{m_{J-1}}, c^r_{k_{J-1}}}\Big| &\geq \sqrt{\frac{\delta_T}{4n}}\Delta^f_{J-1} - \sqrt{8\log T} \geq \sqrt{\frac{\delta_T}{4n}}\underline{f}_T e - \sqrt{8\log T} \\ &\geq \left(\frac{1}{\sqrt{4n}} - \frac{2\sqrt{2}}{\underline{C}}\right)\sqrt{\delta_T}\underline{f}_T = C_2\sqrt{\delta_T}\underline{f}_T > \zeta_T. \end{aligned}$$

Therefore, we have shown that there exists an interval of the form $[c^l_{\tilde{m}_{J-1}}, c^r_{\tilde{k}_{J-1}}]$ with $\max_{c^l_{\tilde{m}_{J-1}} \leq b < c^r_{\tilde{k}_{J-1}}} \Big|\tilde{X}^t_{c^l_{\tilde{m}_{J-1}}, c^r_{\tilde{k}_{J-1}}}\Big| > \zeta_T$.

Now, denote $c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}$ the first points where this occurs and b_{J-1} as defined above with $\Big|\tilde{X}^{b_J-1}_{c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}}\Big| > \zeta_T$. We can show that $\big|r_{J-1} - b_{J-1}\big|\Big(\Delta^f_{J-1}\Big)^2 \leq C_3 \log T$, following exactly the same process as in Step 3.

After detecting r_{J-1} in the interval $[c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}]$, DAIS will restart on intervals $[1, c^l_{m^*_{J-1}}]$ and $[c^r_{k^*_{J-1}}, c^l_{m^*}]$. DAIS will also check $[c^r_{k^*}, T]$, as described in Section 2. Step 5 can be applied to all intervals, as long as there is a change-point. We can conclude that all change-points will get detected, one by one, and their estimated locations will satisfy $\big|r_j - b_j\big|\Big(\Delta^f_j\Big)^2 \leq C_3 \log T, \forall j \in \{1, 2, \ldots, N\}$. There will not be any double detection issues as each interval contains no previously detected change-points.

Step 6: After detecting all the change-points at locations $b_1, b_2, \ldots, b_N$ using the intervals $[c^l_{m^*_j}, c^r_{k^*_j}]$ for $j \in \{1, \ldots, N\}$, the algorithm will check all intervals of the form $[c^r_{k^*_{j-1}}, c^l_{m^*_j}]$ and $[c^r_{k^*_j}, c^l_{m^*_{j+1}}]$, with $c^l_{m^*_0} = 1$ and $c^r_{k^*_{N+1}} = T$. At most $N + 1$ intervals of this form, containing no change-points, will be checked. Denoting by $[s^*, e^*]$ any of those intervals, we can show that DAIS will not detect any change-point as for $b \in \{s^*, s^* + 1, \ldots, e^* - 1\}$,

$$\Big|\tilde{X}^b_{s^*,e^*}\Big| \leq \Big|\tilde{f}^b_{s^*,e^*}\Big| + \sqrt{8\log T} = \sqrt{8\log T} < C_1\sqrt{\log T} \leq \zeta_T.$$

The algorithm will terminate after not detecting any change-points in all intervals.

□

A Further simulations

Some further simulation results are presented in Tables 13 to 15 for piecewise-constant signals and Tables 16 to 18 for continuous piecewise-linear signals. The signals used are:

(S11) *small_dist2*: sequence of length 1000 with 4 change-points at 485, 515, 900 and 930 with values between change-points 0, 1, 0, 1.5, 0. The standard deviation is $\sigma = 1$.

(S12) *small_dist3*: sequence of length 1000 with 6 change-points at 100, 130, 485, 515, 870, 900 with values between change-points 0, 1.5, 0, 1, 0, 1.5, 0. The standard deviation is $\sigma = 1$.

Table 13: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signal (S11). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$									
Method	≤ -3	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS	0	17	2	78	3	0	0	0.0220	0.084	0.004
ID_th	0	3	1	69	10	11	6	0.0228	0.071	0.004
ID_ic	0	26	0	71	3	0	0	0.0230	0.114	0.002
WBS_th	0	0	0	17	18	18	47	0.0349	0.167	0.022
WBS_ic	0	27	0	71	2	0	0	0.0221	0.118	0.022
WBS2	0	6	3	70	8	7	6	0.0221	0.075	0.485
PELT	9	72	0	19	0	0	0	0.0389	0.378	0.001
NOT	0	25	0	72	2	1	0	0.0217	0.111	0.024
MOSUM	0	8	1	85	4	2	0	0.0195	0.055	0.012
MSCP	89	9	1	1	0	0	0	0.0836	0.576	158.370

Table 14: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signal (S12). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$									
Method	≤ -3	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS	0	15	3	80	0	2	0	0.0321	0.067	0.004
ID_th	0	2	2	75	8	7	6	0.0323	0.037	0.003
ID_ic	2	34	1	62	1	0	0	0.0350	0.143	0.002
WBS_th	0	0	0	22	24	19	35	0.0414	0.092	0.022
WBS_ic	2	31	1	65	0	1	0	0.0331	0.132	0.022
WBS2	0	6	5	70	9	6	4	0.0306	0.132	0.479
PELT	26	65	0	9	0	0	0	0.0596	0.437	0.001
NOT	2	30	1	67	0	0	0	0.0333	0.128	0.024
MOSUM	0	13	2	79	3	3	0	0.0287	0.063	0.015
MSCP	98	2	0	0	0	0	0	0.1457	0.632	158.334

(S13) $teeth$: piecewise-constant signal of length 270 with 13 change-points at 11, 31, 51, 71, 91, 111, 131, 151, 171, 191, 211, 231, 251 with values between change-points 0, 1, 0, 1, . . ., 0, 1. The standard deviation is $\sigma = 0.4$.

(S14) $wave4$: continuous piecewise-linear signal of length 200 with 9 change-points at 20, 40, . . ., 180 with changes in slope 1/6, 1/2, -3/4, -1/3, -2/3, 1, 1/4, 3/4, -5/4. The starting intercept is $f_1 = -1$ and slope $f_2 - f_1 = 1/32$. The standard deviation is $\sigma = 0.3$.

(S15) $wave5$: continuous piecewise-linear signal of length 1000 with 19 change-points at 50, 100, . . ., 950 with changes in slope -1/16, -5/16, -5/8, 1, 5/16, 15/32, -5/8, -7/32, -3/4, 13/16, 5/16, 19/32, -1, -5/8, 23/32, 1/2, 15/16, -25/16, -5/4. The starting intercept is $f_1 = -1$ and slope $f_2 - f_1 = 1/32$. The standard deviation is $\sigma = 0.6$.

(S16) $wave6$: continuous piecewise-linear signal of length 350 with 50 change-points at 7, 14, . . ., 343 with changes in slope -2.5, 2.5, . . ., 2.5, -2.5. The starting intercept is $f_1 = -0$ and slope $f_2 - f_1 = 1$. The standard deviation is $\sigma = 1$.

The general conclusions form the tables are that DAIS performs at least as well as the competitors, often having an advantage in computational time and accuracy compared to the best performing competitors. Tables 13 and 14 indicate that in these difficult structures, where pairs of change-points are close and move in opposite directions such that they 'cancel' each-other out, DAIS and MOSUM have the best performance, in terms of accuracy in the number of change-points detected and MSE. For the signal (S13), in Table 15, all algorithms perform similarly, with DAIS being one of the fastest methods. The low computational complexity of DAIS can also be seen in Tables 16, 17, 18.

B Proof of Theorem 2

For the proof of Theorem 2, we require the following two lemmas.

Table 15: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signal (S13). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$									
Method	≤ -3	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS	0	0	1	94	4	1	0	2.72×10^{-2}	0.010	0.002
ID_th	0	0	1	84	12	2	1	2.72×10^{-2}	0.011	0.001
ID_ic	0	0	0	87	12	1	0	2.46×10^{-2}	0.009	0.002
WBS_th	0	0	0	58	16	19	7	2.58×10^{-2}	0.015	0.014
WBS_ic	0	0	0	92	8	0	0	2.35×10^{-2}	0.008	0.014
WBS2	0	0	1	96	3	0	0	2.31×10^{-2}	0.008	0.088
PELT	1	6	3	90	0	0	0	2.63×10^{-2}	0.016	0.001
NOT	0	0	0	95	5	0	0	2.47×10^{-2}	0.009	0.021
MOSUM	0	0	0	95	3	2	0	2.26×10^{-2}	0.008	0.008
MSCP	100	0	0	0	0	0	0	22.49×10^{-2}	0.295	14.850

Table 16: Distribution of $\hat{N} - N$ of 100 simulated data sequences of (S14). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$									
Method	≤ -15	$(-15, -2]$	-1	0	1	$[2, 15)$	≥ 15	MSE	d_H	Time (s)
DAIS	0	0	0	96	3	1	0	0.025	0.116	0.002
ID_th	0	0	0	65	23	12	0	0.020	0.188	0.001
ID_ic	0	0	0	99	1	0	0	0.013	0.108	0.003
CPOP	0	0	0	90	6	4	0	0.083	0.125	0.014
NOT	0	0	1	95	2	2	0	0.014	0.102	0.070
MARS	0	91	7	2	0	0	0	3.990	2.260	0.003
TF	0	0	0	0	0	35	65	0.026	0.432	0.082
TS	0	0	0	99	0	1	0	0.093	0.195	0.079

Table 17: Distribution of $\hat{N} - N$ of 100 simulated data sequences of (S15). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$									
Method	≤ -15	$(-15, -2]$	-1	0	1	$[2, 15)$	≥ 15	MSE	d_H	Time (s)
DAIS	0	0	4	84	12	0	0	0.057	0.147	0.006
ID_th	0	0	0	55	38	7	0	0.059	0.150	0.047
ID_ic	0	0	0	87	13	0	0	0.038	0.108	0.011
CPOP	0	0	0	90	10	0	0	0.089	0.086	0.187
NOT	0	0	60	37	3	0	0	0.043	0.623	0.238
MARS	0	100	0	0	0	0	0	22.000	1.610	0.005
TF	0	0	0	0	0	0	100	0.032	0.467	0.511
TS	0	0	7	85	7	1	0	0.225	0.225	0.468

Table 18: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signal (S16). The average MSE, d_H and computational time are also given.

	$\hat{N} - N$									
Method	≤ -15	$(-15, -2]$	-1	0	1	$[2, 15)$	≥ 15	MSE	d_H	Time (s)
DAIS	0	1	5	94	0	0	0	0.497	0.304	0.007
ID_th	0	0	2	91	7	0	0	0.496	0.290	0.004
ID_ic	37	3	9	51	0	0	0	3.000	18.500	0.022
CPOP	0	0	0	96	3	1	0	1.660	0.316	0.023
NOT	100	0	0	0	0	0	0	6.650	49.000	0.036
MARS	100	0	0	0	0	0	0	6.590	47.800	0.003
TF	0	0	0	0	0	72	28	1.210	0.319	0.102
TS	100	0	0	0	0	0	0	6.650	49.000	0.157

Lemma 2. *Suppose* $\boldsymbol{f} = (f_1, f_2, \ldots, f_T)^T$ *is a piecewise-linear vector and* $r_1, r_2, \ldots, r_N$ *are the locations of the change-points. Suppose* $1 \leq s < e \leq T$*, such that* $r_{j-1} \leq s < r_j < e \leq r_{j+1}$*, for some* $j = 1, 2, \ldots, N$*. Let* $\Delta_j^f = \left|2f_{r_j} - f_{r_j+1} - f_{r_j-1}\right|$ *and* $\eta = \min\{r_j - s, e - r_j\}$*. Then,*

$$C_{s,e}^{r_j}(\boldsymbol{f}) = \max_{s<b<e} C_{s,e}^{b}(\boldsymbol{f}) \begin{cases} \geq \frac{1}{\sqrt{24}}\eta^{3/2}\Delta_j^f, \\ \leq \frac{1}{\sqrt{3}}(\eta+1)^{3/2}\Delta_j^f \end{cases}$$

Proof. See Lemma 5 from Baranowski et al. [2019].

□

Lemma 3. *Suppose* $\boldsymbol{f} = (f_1, f_2, \ldots, f_T)^T$ *is a piecewise-linear vector and* $r_1, r_2, \ldots, r_N$ *are the locations of the change-points. Suppose* $1 \leq s < e \leq T$*, such that* $r_{j-1} \leq s < r_j < e \leq r_{j+1}$*, for some* $j = 1, 2, \ldots, N$*. Let* $\rho = |r - b|, \Delta_j^f = \left|2f_{r_j} - f_{r_j+1} - f_{r_j-1}\right|, \eta_L = r_j - s$ *and* $\eta_R = e - r_j$*. Then,*

$$\|\boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{b}}\langle \boldsymbol{f}, \boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{b}}\rangle - \boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{r}}\langle \boldsymbol{f}, \boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{r}}\rangle\|_2^2 = \left(C_{s,e}^{r_j}(\boldsymbol{f})\right)^2 - \left(C_{s,e}^{b}(\boldsymbol{f})\right)^2.$$

In addition,

1. *for any* $r_j \leq b < e$, $\left(C_{s,e}^{r_j}(\boldsymbol{f})\right)^2 - \left(C_{s,e}^{b}(\boldsymbol{f})\right)^2 = \frac{1}{63}\min\{\rho, \eta_L\}^3\left(\Delta_j^f\right)^2$;
2. *for any* $s \leq b < r_j$, $\left(C_{s,e}^{r_j}(\boldsymbol{f})\right)^2 - \left(C_{s,e}^{b}(\boldsymbol{f})\right)^2 = \frac{1}{63}\min\{\rho, \eta_R\}^3\left(\Delta_j^f\right)^2$.

Proof. See Lemma 7 from Baranowski et al. [2019].

□

The steps of the proof of Theorem 1 are the same as for Theorem 2.

Proof. We will prove the more specific result

$$\mathbb{P}\left(\hat{N} = N, \max_{j=1,2,\ldots,N}\left(|\hat{r}_j - r_j|\left(\Delta_j^f\right)^{2/3}\right) \leq C_3(\log T)^{1/3}\right) \geq 1 - \frac{1}{6\sqrt{\pi}T}, \tag{39}$$

which implies result (12).

Steps 1 & 2: Similar to Theorem 1, denote

$$A_T^* = \left\{ \max_{s,b,e:1\leq s\leq b<e\leq T} \left|C_{s,e}^b(\boldsymbol{X}) - C_{s,e}^b(\boldsymbol{f})\right| \leq \sqrt{8\log T} \right\}$$

$$B_T^* = \left\{ \max_{1\leq s\leq b<e<T} \max_{\substack{r_{j-1}<s\leq r_j \\ r_j<e\leq r_{j+1} \\ s\leq b<e}} \frac{\left|\left\langle \boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{b}}\langle \boldsymbol{f}, \boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{b}}\rangle - \boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{r}}\langle \boldsymbol{f}, \boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{r}}\rangle, \boldsymbol{\epsilon}\right\rangle\right|}{\|\boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{b}}\langle \boldsymbol{f}, \boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{b}}\rangle - \boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{r}}\langle \boldsymbol{f}, \boldsymbol{\phi}_{\boldsymbol{s,e}}^{\boldsymbol{r}}\rangle\|_2} \leq \sqrt{8\log T} \right\}. \tag{40}$$

The same reasoning as in the proof of Theorem 1 leads to $\mathbb{P}(A_T^*) \geq 1 - 1/(12\sqrt{\pi}T)$ and $\mathbb{P}(B_T^*) \geq 1 - 1/(12\sqrt{\pi}T)$. Therefore, it holds that

$$\mathbb{P}(A_T^* \cap B_T^*) \geq 1 - \frac{1}{6\sqrt{\pi}T}$$

Step 3: From now on, we assume that A_T^* and B_T^* both hold. The constants we use are

$$C_1 = \sqrt{\frac{2}{3}}C_3^{\frac{3}{2}} + \sqrt{8}, \quad C_2 = \frac{1}{8\sqrt{3n^3}} - \frac{2\sqrt{2}}{C^*}, \quad C_3 = 63^{\frac{1}{3}}(2\sqrt{2}+4)^{\frac{2}{3}},$$

where C^* satisfies assumption (A3), $\delta_T^{3/2}\underline{f}_T \geq C^*\sqrt{logT}$ and $3/2 \leq n \leq \delta_T/2\lambda_T$. As before, for $j \in \{1, 2, \ldots, N\}$ define I_j^L and I_j^R as in (5). For $d_{s,e} = \text{argmax}_{t\in\{s,s+1,\ldots,e-2\}}\{|X_{t+2} - 2X_{t+1} + X_t|\}$ being the location of the largest difference detected in the interval $[s, e]$, $1 \leq s < e \leq T$, define c_m^l and c_k^r as in (23). Since the length of the intervals in (5) is $\delta_T/2n$ and $\lambda_T \leq \delta_T/2n$, for $m, k \in \{0, 1, \ldots, K^{\max}\}$ we ensure that there exists at least one m and at least one k such that $c_m^l \in I_j^L$ and $c_k^r \in I_j^R$ for all $j \in \{1, 2, \ldots, N\}$.

At the beginning of DAIS, $s = 1, e = T$ and the first change-point that will get detected depends on the value of $d_{1,T}$. As already explained in Section 2.3, the largest difference $d_{s,e}$ will be at most at a distance $\delta_{1,T}^j$ from the nearest change-point, where $\delta_{1,T}^j \leq \frac{\delta_j}{2} - \frac{3\delta_T}{4n}$ for $\delta_j = |r_{j+1} - r_j|$. The first point to get detected will be the point that is closest to the largest difference $d_{1,T}$. The interval where the detection of this change-point occurs, cannot contain more than one change-points, as was explained in Section 2.3.

Without loss of generality, we suppose that the first change-point to get detected is r_J for some $J \in \{1, 2, \ldots, N\}$. We can show that there exists an interval $[c_m^l, c_k^r]$, for $m, k \in \{0, 1, \ldots, K^{\max}\}$, such that r_J is isolated using exactly the same argument as in (24). We will now show that for $\tilde{b}_J = \text{argmax}_{c_m^l \leq t < c_k^r} C_{c_m^l, c_k^r}^t(\boldsymbol{X})$, then $C_{c_m^l, c_k^r}^{\tilde{b}_J}(\boldsymbol{X}) > \zeta_T$. Using (8), we have that

$$C_{c_m^l, c_k^r}^{\tilde{b}_J}(\boldsymbol{X}) \geq C_{c_m^l, c_k^r}^{r_J}(\boldsymbol{X}) \geq C_{c_m^l, c_k^r}^{r_J}(\boldsymbol{f}) - \sqrt{8\log T}. \tag{41}$$

From Lemma 2, we have that

$$C_{c_m^l, c_k^r}^{r_J}(\boldsymbol{f}) \geq \frac{1}{\sqrt{24}}\left(\min\{r_J - c_m^l, c_k^r - r_J\}\right)^{\frac{3}{2}}\Delta_J^f.$$

Showing that

$$\min\{c_k^r - r_J, r_J - c_m^l\} \geq \frac{\delta_T}{2n}. \tag{42}$$

follows the exact same steps as Step 3 in the proof of Theorem 1 and will not be repeated. Now, using assumption (A3) and the results in (41), (42), we have that

$$\begin{aligned} C_{c_m^l, c_k^r}^{\tilde{b}_J}(\boldsymbol{X}) &\geq \frac{1}{\sqrt{24}}\left(\frac{\delta_T}{2n}\right)^{\frac{3}{2}}\Delta_J^f - \sqrt{8\log T} \geq \frac{1}{\sqrt{24}}\left(\frac{\delta_T}{2n}\right)^{\frac{3}{2}}\underline{f}_T - \sqrt{8\log T} \\ &= \left(\frac{1}{8\sqrt{3n^3}} - \frac{2\sqrt{2\log T}}{\delta_T^{3/2}\underline{f}_T}\right)\delta_T^{3/2}\underline{f}_T \geq \left(\frac{1}{8\sqrt{3n^3}} - \frac{2\sqrt{2}}{C^*}\right)\delta_T^{3/2}\underline{f}_T \\ &= C_2\delta_T^{3/2}\underline{f}_T > \zeta_T \end{aligned} \tag{43}$$

and thus, the change-point will get detected.

Therefore, there will be an interval of the form $[c_{\tilde{m}}^l, c_{\tilde{k}}^r]$, such that the interval contains r_J and no other change-point

and $\max_{c^l_{\tilde{m}} \le t < c^r_{\tilde{k}}} C^t_{c^l_{\tilde{m}}, c^r_{\tilde{k}}}(\boldsymbol{X}) > \zeta_T$. For $k^*, m^* \in \{0, 1, \dots K^{\max}\}$, denote by $c^l_{m^*} \ge c^l_{\tilde{m}}$ and $c^r_{k^*} \le c^r_{\tilde{k}}$ the first left- and right-expanding points, respectively, that this happens and let $b_J = \operatorname{argmax}_{c^l_{m^*} \le t < c^r_{k^*}} C^t_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{X})$, with $C^{b_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{X}) > \zeta_T$. Note that b_J cannot be an estimation of $r_j, j \neq J$, as r_J is isolated in the interval where it is detected. Our aim now is to find $\tilde{\gamma}_T > 0$, such that for any $b^* \in \{c^l_{m^*}, c^l_{m^*}+1, \dots, c^r_{k^*}-1\}$ with $|b^* - r_J|\left(\Delta^f_J\right)^2 > \tilde{\gamma}_T$, we have that

$$\left(C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{X})\right)^2 > \left(C^{b^*}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{X})\right)^2. \tag{44}$$

Proving (44) and using the definition of b_J, we can conclude that $|b_J - r_J|\left(\Delta^f_J\right)^2 \le \tilde{\gamma}_T$. Now, using (8), it can be shown, for $\boldsymbol{\phi}^{\boldsymbol{b}}_{\boldsymbol{s},\boldsymbol{e}}$ as defined in (11), that (44) is equivalent to

$$\begin{aligned}\left(C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{f})\right)^2 - \left(C^{b^*}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{f})\right)^2 > \left(C^{b^*}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{\epsilon})\right)^2 - \left(C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{\epsilon})\right)^2 \\ + 2\left\langle \boldsymbol{\phi}^{b^*}_{c^l_{m^*}, c^r_{k^*}} \langle \boldsymbol{f}, \boldsymbol{\phi}^{b^*}_{c^l_{m^*}, c^r_{k^*}} \rangle - \boldsymbol{\phi}^{r_J}_{c^l_{m^*}, c^r_{k^*}} \langle \boldsymbol{f}, \boldsymbol{\phi}^{r_J}_{c^l_{m^*}, c^r_{k^*}} \rangle, \boldsymbol{\epsilon} \right\rangle.\end{aligned} \tag{45}$$

Without loss of generality, assume that $b^* \in [r_J, c^r_{k^*})$ and a similar approach holds when $b^* \in [c^l_{m^*}, r_J)$. Denote

$$\Lambda \colon = \left(C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{f})\right)^2 - \left(C^{b^*}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{f})\right)^2. \tag{46}$$

For the right-hand side of (45) using (40),

$$\left(C^{b^*}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{\epsilon})\right)^2 - \left(C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{\epsilon})\right)^2 \le \max_{s,e,b: s \le b < e} \left(C^b_{s,e}(\boldsymbol{\epsilon})\right)^2 - \left(C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{\epsilon})\right)^2 \le 8 \log T \tag{47}$$

Using Lemma 3, (40) and (46), we have that for the left-hand side,

$$\begin{aligned}&2\left\langle \boldsymbol{\phi}^{b^*}_{c^l_{m^*}, c^r_{k^*}} \langle \boldsymbol{f}, \boldsymbol{\phi}^{b^*}_{c^l_{m^*}, c^r_{k^*}} \rangle - \boldsymbol{\phi}^{r_J}_{c^l_{m^*}, c^r_{k^*}} \langle \boldsymbol{f}, \boldsymbol{\phi}^{r_J}_{c^l_{m^*}, c^r_{k^*}} \rangle, \boldsymbol{\epsilon} \right\rangle \\ &\quad \le 2\left\| \boldsymbol{\phi}^{b^*}_{c^l_{m^*}, c^r_{k^*}} \langle \boldsymbol{f}, \boldsymbol{\phi}^{b^*}_{c^l_{m^*}, c^r_{k^*}} \rangle - \boldsymbol{\phi}^{r_J}_{c^l_{m^*}, c^r_{k^*}} \langle \boldsymbol{f}, \boldsymbol{\phi}^{r_J}_{c^l_{m^*}, c^r_{k^*}} \rangle \right\|_2 \sqrt{8 \log T} \\ &\quad = 2\sqrt{\Lambda}\sqrt{8 \log T}.\end{aligned} \tag{48}$$

Using (46), (47) and (48), we can conclude that (45) is satisfied if $\Lambda > 8 \log T + \sqrt{2}\sqrt{\Lambda}\sqrt{8 \log T}$ is satisfied, which has solution

$$\Lambda > \left(2\sqrt{2} + 4\right)^2 \log T.$$

Using Lemma 3, we can conclude that

$$\begin{aligned}&\Lambda > \left(2\sqrt{2} + 4\right)^2 \log T \\ &\Leftrightarrow \frac{1}{63}\left(\min\{|r_J - b^*|, r_J - c^l_{m^*}\}\right)^3 \left(\Delta^f_J\right)^2 > \left(2\sqrt{2} + 4\right)^2 \log T \\ &\Leftrightarrow \min\{|r_J - b^*|, r_J - c^l_{m^*}\} > \frac{(63 \log T)^{1/3}\left(2\sqrt{2} + 4\right)^{2/3}}{\left(\Delta^f_J\right)^{2/3}} = \frac{C_3 (\log T)^{1/3}}{\left(\Delta^f_J\right)^{2/3}}\end{aligned} \tag{49}$$

Now, if for sufficiently large T

$$\min\{r_J - c^l_{m^*}, c^r_{k^*} - r_J\} > 2^{1/3} C_3 \frac{(\log T)^{1/3}}{\left(\Delta^f_J\right)^{2/3}} - 1, \tag{50}$$

it follows that

$$\min\{r_J - c^l_{m^*}, c^r_{k^*} - r_J\} > C_3 \frac{(\log T)^{1/3}}{\left(\Delta^f_J\right)^{2/3}},$$

and we can deduce that (49) is restricted to

$$|r_J - b^*| > C_3 \frac{(\log T)^{1/3}}{\left(\Delta_J^f\right)^{2/3}}$$

which implies (44). So, we conclude that necessarily

$$|r_J - b_J| \left(\Delta_J^f\right)^{2/3} \leq C_3 \left(\log T\right)^{1/3}. \tag{51}$$

But (50) must be true since if we assume that

$$\min\{r_J - c_{m^*}^l, c_{k^*}^r - r_J\} \leq 2^{1/3} C_3 \frac{(\log T)^{1/3}}{\left(\Delta_J^f\right)^{2/3}} - 1,$$

then, using Lemma 2, we have that

$$\begin{aligned}
C_{c_{m^*}^l, c_{k^*}^r}^{b_J}(\boldsymbol{X}) &\leq C_{c_{m^*}^l, c_{k^*}^r}^{r_J}(\boldsymbol{f}) + \sqrt{8 \log T} \\
&\leq \frac{1}{\sqrt{3}} \left(\min\{r_J - c_{m^*}^l, c_{k^*}^r - r_J\} + 1\right)^{3/2} \Delta_J^f + \sqrt{8 \log T} \\
&\leq \frac{1}{\sqrt{3}} \left(2^{1/3} C_3 \frac{(\log T)^{1/3}}{\left(\Delta_J^f\right)^{2/3}} \right)^{3/2} \Delta_J^f + \sqrt{8 \log T} \\
&= \sqrt{\frac{2}{3}} C_3^{3/2} \sqrt{\log T} + \sqrt{8 \log T} \\
&= \left(\sqrt{\frac{2}{3}} C_3^{3/2} + \sqrt{8} \right) \sqrt{\log T} = C_1 \sqrt{\log T} \leq \zeta_T,
\end{aligned}$$

which contradicts $C_{c_{m^*}^l, c_{k^*}^r}^{b_J}(\boldsymbol{X}) > \zeta_T$.

Thus, we have proved that for $\lambda_T \leq \delta_T / 2n$, working under the assumption that both A_T^* and B_T^* hold, there will be an interval $[c_{m^*}^l, c_{k^*}^r]$ with $C_{c_{m^*}^l, c_{k^*}^r}^{b_J}(\boldsymbol{X}) > \zeta_T$, where $b_J = \text{argmax}_{c_{m^*}^l \leq t < c_{k^*}^r} C_{c_{m^*}^l, c_{k^*}^r}^{t}(\boldsymbol{X})$ is the estimated location for the change-point r_J that satisfies (51).

Step 4: After the detection of the change-point r_J at the estimated location b_J in the interval $[c_{m^*}^l, c_{k^*}^r]$, the process is repeated in the disjoint intervals $[1, c_{m^*}^l]$ and $[c_{k^*}^r, T]$. Proving that there is no other change-point in the interval $[c_{m^*}^l, c_{k^*}^r]$ can be done in exactly the same way as Step 4 in the case of piecewise-constant signals and will not be repeated here.

Step 5: After detecting r_J, the algorithm will first check the interval $[1, c_{m^*}^l]$. So, unless $r_J = r_1$ and $[1, c_{m^*}^l]$ contains no other change-points, the next change-point to get detected will be one of $r_1, r_2, \ldots, r_{J-1}$. The location of the largest difference in the interval $[1, c_{m^*}^l]$, $d_{1, c_{m^*}^l}$, will again determine which change-point will be detected next and as in Step 5 for piecewise-constant signals, we only need to consider the case when the next change-point to get detected is r_{J-1}.

Now, concentrating on the case that the next change-point to get detected is r_{J-1}, we mention that since this is the closest change-point to the already detected r_J, we need to make sure that detection is possible. As before, for $k_{J-1}, m_{J-1} \in \{0, 1, \ldots, K^{\max}\}$ we will show that r_{J-1} gets detected in $[c_{m_{J-1}^*}^l, c_{k_{J-1}^*}^r]$, where $c_{m_{J-1}^*}^l \geq c_{m_{J-1}}^l$ and $c_{k_{J-1}^*}^r \leq c_{k_{J-1}}^r \leq c_{m^*}^l$ and its detection is at location

$$b_{J-1} = \text{argmax}_{c_{m_{J-1}^*}^l \leq t < c_{k_{J-1}^*}^r} C_{c_{m_{J-1}^*}^l, c_{k_{J-1}^*}^r}^{t}(\boldsymbol{X}),$$

which satisfies $\left|r_{J-1}-b_{J-1}\right|\left(\Delta_{J-1}^{f}\right)^{2/3}\leq C_3\left(\log T\right)^{1/3}$. Firstly, r_{J-1} is isolated in the interval $[c^l_{m_{J-1}},c^r_{k_{J-1}}]$ using the same argument as in (24). Using Lemma 3, we have that for $\tilde{b}_{J-1}=\text{argmax}_{c^l_{m_{J-1}}\leq t<c^r_{k_{J-1}}}C^t_{c^l_{m_{J-1}},c^r_{k_{J-1}}}(\boldsymbol{X})$,

$$\begin{aligned}C^{\tilde{b}_{J-1}}_{c^l_{m_{J-1}},c^r_{k_{J-1}}}(\boldsymbol{X})&\geq C^{\tilde{r}_{J-1}}_{c^l_{m_{J-1}},c^r_{k_{J-1}}}(\boldsymbol{X})\geq C^{\tilde{r}_{J-1}}_{c^l_{m_{J-1}},c^r_{k_{J-1}}}(\boldsymbol{f})-\sqrt{8\log T}\\&\geq\frac{1}{\sqrt{24}}\left(\min\{r_{J-1}-c^l_{m_{J-1}},c^r_{k_{J-1}}-r_{J-1}\}\right)^{3/2}\Delta^f_{J-1}-\sqrt{8\log T}.\end{aligned}\tag{52}$$

Before we show that $\min\{c^r_{k_{J-1}}-r_{J-1},r_{J-1}-c^l_{m_{J-1}}\}\geq\delta_T/2n$, we need show that $c^l_{m^*}$ satisfies $c^l_{m^*}-r_{J-1}\geq\delta_T/2n$ The proof is exactly the same as in Step 5 of the proof of Theorem 1. It can be deduced that the right end-point of the interval will satisfy $c^r_{k_{J-1}}-r_{J-1}>\delta_T/2n$ for some k_{J-1}. For the left end-point of the interval, the same holds as in Step 3. In any case, $\min\{c^r_{k_{J-1}}-r_{J-1},r_{J-1}-c^l_{m_{J-1}}\}>\delta_T/2n$ holds and so, from (52), using exactly the same calculations as (43), we have that

$$C^{\tilde{b}_{J-1}}_{c^l_{m_{J-1}},c^r_{k_{J-1}}}(\boldsymbol{X})=C_2(\delta_T)^{3/2}\underline{f}_T>\zeta_T.$$

Therefore, we have shown that there exists an interval of the form $[c^l_{\tilde{m}_{J-1}},c^r_{\tilde{k}_{J-1}}]$ with $\max_{c^l_{\tilde{m}_{J-1}}\leq b<c^r_{\tilde{k}_{J-1}}}C^t_{c^l_{\tilde{m}_{J-1}},c^r_{\tilde{k}_{J-1}}}(\boldsymbol{X})>\zeta_T$.

Now, denote $c^l_{m^*_{J-1}},c^r_{k^*_{J-1}}$ the first points where this occurs and b_{J-1} as defined above with $C^{b_J-1}_{c^l_{m^*_{J-1}},c^r_{k^*_{J-1}}}(\boldsymbol{X})>\zeta_T$. We can show that $\left|r_{J-1}-b_{J-1}\right|\left(\Delta_{J-1}^{f}\right)^{2/3}\leq C_3\left(\log T\right)^{1/3}$, following exactly the same process as in Step 3.

After detecting r_{J-1} in the interval $[c^l_{m^*_{J-1}},c^r_{k^*_{J-1}}]$, DAIS will restart on intervals $[1,c^l_{m^*_{J-1}}]$ and $[c^r_{k^*_{J-1}},c^l_{m^*}]$. Step 5 can be applied to all intervals, as long as there is a change-point. We can conclude that all change-points will get detected, one by one, and their estimated locations will satisfy $\left|r_j-b_j\right|\left(\Delta_{j}^{f}\right)^{2/3}\leq C_3\left(\log T\right)^{1/3},\forall j\in\{1,2,\ldots,N\}$. There will not be any double detection issues as each interval contains no previously detected change-points.

Step 6: After detecting all the change-points at locations $b_1,b_2,\ldots,b_N$ using the intervals $[c^l_{m^*_j},c^r_{k^*_j}]$ for $j\in\{1,\ldots,N\}$, the algorithm will check all intervals of the form $[c^r_{k^*_{j-1}},c^l_{m^*_j}]$ and $[c^r_{k^*_j},c^l_{m^*_{j+1}}]$, with $c^l_{m^*_0}=1$ and $c^r_{k^*_{N+1}}=T$. At most $N+1$ intervals of this form, containing no change-points will be checked. Denoting by $[s^*,e^*]$ any of those intervals, we can show that DAIS will not detect any change-point as for $b\in\{s^*,s^*+1,\ldots,e^*-1\}$,

$$C^b_{s^*,e^*}(\boldsymbol{X})\leq C^b_{s^*,e^*}(\boldsymbol{f})+\sqrt{8\log T}=\sqrt{8\log T}<C_1\sqrt{\log T}\leq\zeta_T.$$

The algorithm will terminate after not detecting any change-points in all intervals.

□

References

Hongyuan Cao and Wei Biao Wu. Changepoint estimation: another look at multiple testing problems. *Biometrika*, 102(4):974–980, 07 2015. ISSN 0006-3444. doi: 10.1093/biomet/asv031. URL https://doi.org/10.1093/biomet/asv031.

Solt Kovács, Housen Li, Peter Bühlmann, and Axel Munk. Seeded binary segmentation: A general methodology for fast and optimal change point detection, 2020. URL https://arxiv.org/abs/2002.06633.

N. Piana Agostinetti and G. Sgattoni. Changepoint detection in seismic double-difference data: application of a trans-dimensional algorithm to data-space exploration. *Solid Earth*, 12(12):2717–2733, 2021. doi: 10.5194/se-12-2717-2021. URL https://se.copernicus.org/articles/12/2717/2021/.

Alexander T. M. Fisch, Idris A. Eckley, and Paul Fearnhead. A linear time method for the detection of point and collective anomalies, 2018. URL https://arxiv.org/abs/1806.01947.

Haeran Cho and Piotr Fryzlewicz. Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica*, 22(1):207–229, 2012. ISSN 10170405, 19968507. URL http://www.jstor.org/stable/24310145.

E Brodsky and Boris S Darkhovsky. *Non-parametric statistical diagnosis: problems and methods*, volume 509. Springer Science & Business Media, 2000.

Haeran Cho and Piotr Fryzlewicz. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):475–507, 2015.

Andreas Anastasiou, Ivor Cribben, and Piotr Fryzlewicz. Cross-covariance isolate detect: A new change-point method for estimating dynamic functional connectivity. *Medical Image Analysis*, 75:102252, 2022.

Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020. ISSN 0165-1684. doi: https://doi.org/10.1016/j.sigpro.2019.107299. URL https://www.sciencedirect.com/science/article/pii/S0165168419303494.

Yi Yu. A review on minimax rates in change point detection and localisation, 2020.

Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978. ISSN 00905364. URL http://www.jstor.org/stable/2958889.

Yi-Ching Yao. Estimating the number of change-points via schwarz' criterion. *Statistics & Probability Letters*, 6 (3):181–189, 1988. ISSN 0167-7152. doi: https://doi.org/10.1016/0167-7152(88)90118-6. URL https://www.sciencedirect.com/science/article/pii/0167715288901186.

Yi-Ching Yao and Siu-Tong Au. Least-squares estimation of a step function. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 370–381, 1989.

Douglas M. Hawkins. Fitting multiple change-point models to data. *Computational Statistics & Data Analysis*, 37(3):323–341, 2001. ISSN 0167-9473. doi: https://doi.org/10.1016/S0167-9473(00)00068-2. URL https://www.sciencedirect.com/science/article/pii/S0167947300000682.

Yoshiyuki Ninomiya. Information criterion for gaussian change-point model. *Statistics & Probability Letters*, 72(3):237–247, 2005. ISSN 0167-7152. doi: https://doi.org/10.1016/j.spl.2004.10.037. URL https://www.sciencedirect.com/science/article/pii/S0167715205000349.

Jie Ding, Yu Xiang, Lu Shen, and Vahid Tarokh. Multiple change point analysis: Fast implementation and strong consistency. *IEEE Transactions on Signal Processing*, 65(17):4495–4510, 2017. doi: 10.1109/TSP.2017.2711558.

B. Jackson, J.D. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and Tun Tao Tsai. An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters*, 12(2): 105–108, 2005. doi: 10.1109/LSP.2001.838216.

R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012. doi: 10.1080/01621459.2012.737745. URL https://doi.org/10.1080/01621459.2012.737745.

Guillem Rigaill. A pruned dynamic programming algorithm to recover the best segmentations with 1 to $k\_\{max\}$ change-points. *Journal de la Société Française de Statistique*, 156(4):180–205, 2015.

Robert Maidstone, Toby Hocking, Guillem Rigaill, and Paul Fearnhead. On optimal multiple changepoint algorithms for large data. *Statistics and computing*, 27(2):519–533, 2017.

Lyudmila Yur'evna Vostrikova. Detecting "disorder" in multidimensional random processes. In *Doklady akademii nauk*, volume 259, pages 270–274. Russian Academy of Sciences, 1981.

Piotr Fryzlewicz. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6): 2243–2281, 2014.

Piotr Fryzlewicz. Detecting possibly frequent change-points: Wild binary segmentation 2 and steepest-drop model selection. *Journal of the Korean Statistical Society*, 49(4):1027–1070, 2020.

Rafal Baranowski, Yining Chen, and Piotr Fryzlewicz. Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(3): 649–672, 2019.

Andreas Anastasiou and Piotr Fryzlewicz. Detecting multiple generalized change-points by isolating single ones. *Metrika*, 85(2):141–174, 2022.

Chia-Shang J Chu, Kurt Hornik, and Chung-Ming Kaun. Mosum tests for parameter constancy. *Biometrika*, 82(3): 603–617, 1995.

Birte Eichinger and Claudia Kirch. A MOSUM procedure for the estimation of multiple random change points. *Bernoulli*, 24(1):526 – 564, 2018. doi: 10.3150/16-BEJ887. URL https://doi.org/10.3150/16-BEJ887.

Tijana Levajković and Michael Messer. Multiscale change point detection via gradual bandwidth adjustment in moving sum processes. *Electronic Journal of Statistics*, 17(1):70 – 101, 2023. doi: 10.1214/22-EJS2101. URL https://doi.org/10.1214/22-EJS2101.

Hock-Peng Chan and Hao Chen. Multi-sequence segmentation via score and higher-criticism tests, 2017. URL https://arxiv.org/abs/1706.07586.

Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66 (1):47–78, 1998. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/2998540.

Seung-Jean Kim, Kwangmoo Koh, Stephen Boyd, and Dimitry Gorinevsky. ℓ_1 trend filtering. *SIAM Review*, 51(2): 339–360, 2009. doi: 10.1137/070690274. URL https://doi.org/10.1137/070690274.

Paul Fearnhead, Robert Maidstone, and Adam Letchford. Detecting changes in slope with an l0 penalty. *Journal of Computational and Graphical Statistics*, 28(2):265–275, 2019. doi: 10.1080/10618600.2018.1512868. URL https://doi.org/10.1080/10618600.2018.1512868.

Jerome H. Friedman. Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1):1 – 67, 1991. doi: 10.1214/aos/1176347963. URL https://doi.org/10.1214/aos/1176347963.

Steven Spiriti, Randall Eubank, Philip W. Smith, and Dennis Young. Knot selection for least-squares and penalized splines. *Journal of Statistical Computation and Simulation*, 83(6):1020–1036, 2013. doi: 10.1080/00949655.2011.647317. URL https://doi.org/10.1080/00949655.2011.647317.

Paul A. Wiggins. An information-based approach to change-point analysis with applications to biophysics and cell biology. *Biophysical Journal*, 109(2):346–354, 2015. ISSN 0006-3495. doi: https://doi.org/10.1016/j.bpj.2015.05.038. URL https://www.sciencedirect.com/science/article/pii/S0006349515005524.

Alberto Sosa-Costa, Izabela K. Piechocka, Lucia Gardini, Francesco S. Pavone, Marco Capitanio, Maria F. Garcia-Parajo, and Carlo Manzo. Plant: A method for detecting changes of slope in noisy trajectories. *Biophysical Journal*, 114(9):2044–2051, 2018. ISSN 0006-3495. doi: https://doi.org/10.1016/j.bpj.2018.04.006. URL https://www.sciencedirect.com/science/article/pii/S0006349518304454.

Housen Li, Axel Munk, and Hannes Sieling. FDR-control in multiscale change-point segmentation. *Electronic Journal of Statistics*, 10(1):918 – 959, 2016. doi: 10.1214/16-EJS1131. URL https://doi.org/10.1214/16-EJS1131.

Piotr Fryzlewicz. Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *The Annals of Statistics*, 46(6B):3390 – 3421, 2018. doi: 10.1214/17-AOS1662. URL https://doi.org/10.1214/17-AOS1662.

Hyeyoung Maeng and Piotr Fryzlewicz. Detecting linear trend changes in data sequences. *Statistical Papers*, pages 1–31, 2023.

Frank R. Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393, 1974. doi: 10.1080/01621459.1974.10482962. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1974.10482962.

Peter J. Rousseeuw and Christophe Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424):1273–1283, 1993. doi: 10.1080/01621459.1993.10476408. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1993.1047640.

Hock Peng Chan and Guenther Walther. Detection with the scan and the average likelihood ratio. *Statistica Sinica*, 23 (1):409–428, 2013. ISSN 10170405, 19968507. URL http://www.jstor.org/stable/24310529.

Edward W Ng and Murray Geller. A table of integrals of the error functions. *Journal of Research of the National Bureau of Standards B*, 73(1):1–20, 1969.

Alexander Meier, Claudia Kirch, and Haeran Cho. mosum: A package for moving sums in change-point analysis. *Journal of Statistical Software*, 97(8):1–42, 2021. doi: 10.18637/jss.v097.i08. URL https://www.jstatsoft.org/index.php/jss/article/view/v097i08.

F. J. Anscombe. The transformation of poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4):246–254, 1948. ISSN 00063444. URL http://www.jstor.org/stable/2332343.