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. The novelty of the proposed algorithm comes from the data-adaptive nature of the methodology. 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. The methodology can be applied to both univariate and multivariate signals. The simulation results presented indicate that DAIS is at least as accurate as state-of-the-art competitors and in many cases significantly less computationally expensive.

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 ([8]), neuroscience ([32]), seismic data ([40]), astronomy ([10]), and finance ([13]).

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. We highlight, however, that the case when X_t is multivariate is also discussed in our paper. We focus on the case of piecewise-constant and continuous piecewise-linear signals f_t, meaning that the changes are in the mean or the slope. However, the proposed algorithm can be extended to other, possibly more complicated, signal structures. Although detecting changes in the mean is a simple case, as noted by [7], 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, [14] and [1] 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 change-point locations identical to those in the original signal. 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. Upon correct choice of a contrast function, we can cover more scenarios such as the case of piecewise-quadratic signals (see [6] for more details). The importance and novelty of our

work lies mainly in the data-adaptive nature of the proposed algorithm itself, which, at each step, starts searching for change-points in the areas that most prominently include one.

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 each 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 [47] and [53].

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 ([44]), which was used by [51] for change-point detection, assuming Gaussian random variables ϵ_t in (1). Under the same assumption, [52] studies an estimator based on least squares. Relaxing this assumption and allowing instead for more general exponential family distributions, [27] introduces a dynamic programming algorithm which uses maximum likelihood estimates of the location of the change-points. [39] introduces an AIC-type criterion for change-point models and [19] 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. [28] employs dynamic programming to guarantee that the exact global optimum, in terms of creating the segments, is found, while [30] proposes an algorithm, called PELT, whose computational cost is linear in the number of observations. [42] introduces the pDPA algorithm, which includes a pruning step towards complexity reduction. [36] combines the ideas from PELT and pDPA leading to two new algorithms, FPOP and SNIP, which have low computational complexity.

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, which is defined in Section 3.1. A method that has received a lot of attention is the Binary Segmentation algorithm, as introduced in [48]. It starts by searching the whole data sequence for one change-point. Subsequently, 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 more than one 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 [23], 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. [25] proposes WBS2 which draws new intervals each time a detection occurs. A different approach is adopted by [6] with the NOT algorithm. By choosing the narrowest interval for which the value of the chosen contrast function exceeds a pre-defined threshold, there is exactly one change-point in each interval with high probability. The ID method in [2] 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, in an interchangeable way. The isolation of each change-point maximizes the detection power. [32] proposes SeedBS which uses a deterministic construction of search intervals which can be pre-computed while [18], [20] and [33] consider moving sum statistics. [24] 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 and [10], through a 'bottom-up' approach, explores reverse segmentation, which 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. A method that does not fall in any of the two categories described above is FDRSeg, proposed by [34], 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.

Methods for detecting more general structural changes can again be split into the same two categories. Focusing on optimization-based methods, [5] considers the estimation of linear models based on the least squares principle. [31] introduces a trend filtering (TF) approach to produce piecewise-linear trend estimates using an ℓ_1 penalty. [21] 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. [22] 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 [46]. [50] 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. [45] 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 algorithm NOT can be applied to both piecewise-linear and quadratic signals, while ID can be applied to piecewise-linear and can be extended to piecewise-quadratic signals. Finally, [35] extends TGUH to piecewise-linear signals and [29] proposes a moving sum methodology.

In this project, we propose a data-adaptive change-point detection method, called Data Adaptive ISolation (labelled DAIS), 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 around a true change-point. The idea behind the data-adaptive nature of the algorithm, and the belief that the algorithm will start close to a change-point, arise from the fact that, for example, 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, therefore the most prominent change-point. Similarly, in the case that f_t is continuous and piecewise-linear, differencing the signal twice will again reveal a signal with spikes near to the true 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. A discussion about the connection between the isolation aspect of the algorithm and the location of the largest difference can be found in Section 2.3, while Section 3.3 complements this discussion with theoretical results in the case that ϵ_t is Gaussian. We want to emphasise that since the largest difference is only used as the starting point of the search for change-points in the data sequence, the true change-points are detected even when the largest difference is far from them. 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 (in absolute value) found, denoted by$d_{s,e}$, 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, as explained before, we have a reason to believe to be around 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 $[s, e]$ 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 data-adaptive nature of DAIS is what differentiates it from its competitors. A key methodological difference with other algorithms is the way the intervals are being checked. WBS and NOT check to the left and right of the interval in which detection has already occurred, using pre-drawn intervals, while WBS2 draws new intervals. Essentially, for both algorithms, the data sequence is split into three parts, one being the interval including the change-point that has just been detected, one to its left and one to its right. After detection, ID re-starts on the left or the right of the interval where detection occurred, splitting the data sequence in two parts. DAIS, similarly to WBS and WBS2, checks intervals both to the left and right of the interval in which detection occurred. However, every time it re-starts, expanding intervals around the location of the largest difference, in absolute value, are created and thus detection of the change-point in an interval in which it is isolated occurs with high probability. The data-adaptivity also gives an advantage in terms of computational complexity, as is explained in Section 4.1.

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, in Section 2.3, we discuss the connection between the location of the largest difference calculated at each step of the algorithm and the guarantee that the change-point can be detected in an interval in which it is isolated. 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 discussion of Section 2.3. Section 4 includes an explanation of how some parameters of DAIS are selected and a note on computational complexity, which is also compared to that of competitors. In Section 5, we provide results on simulated data and a comparison to state-of-the-art competitors. Section 6 discusses extensions to the algorithm in more complicated signal structures, which include temporal correlation, relaxing the Gaussianity assumption on the noise term ϵ_t, and multivariate data sequences. In Section 7, two examples using real world data are presented. The first one is about crime data reported daily and the second one involves weekly data on the Euro to British pound exchange rate. Section 8 concludes the paper with a summary of the most important findings. In Appendix A, we provide the signals used in the simulation study of Section 5 and in Appendix B further simulation results can be found. Finally, the proofs of Theorems 1 and 2 and the outline of the proof of Theorem 3 can be found in Appendix C, D and E, respectively.

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 with length $T = 100$, standard deviation $\sigma = 1$, and a change-point at location $r = 65$ with a jump of magnitude equal to 1.5. The data are presented in the left plot of Figure 1 with the true underlying signal plotted in red. We first calculate the absolute difference of consecutive observations, X_t and X_{t+1} for $t = 1, 2, \ldots, T-1$. 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 ($d_{1,100} = 65$), which 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, where, for the sake of presentation of this example, the expansion parameter is chosen to be $\lambda_T = 10$. The intervals in the order in which they are checked are $[65, 74]$, $[55, 74]$ 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 just at points where the absolute value of the difference between the noise of two consecutive observations is large. In this case, the location of the largest difference is arbitrary and only acts as the starting point for searching the whole interval for change-points to confirm that there is none. The largest differences are now detected at points $d_{1,55} = 46$ and $d_{84,100} = 85$ 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 checked are: $[85, 94]$, $[84, 94]$, 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.

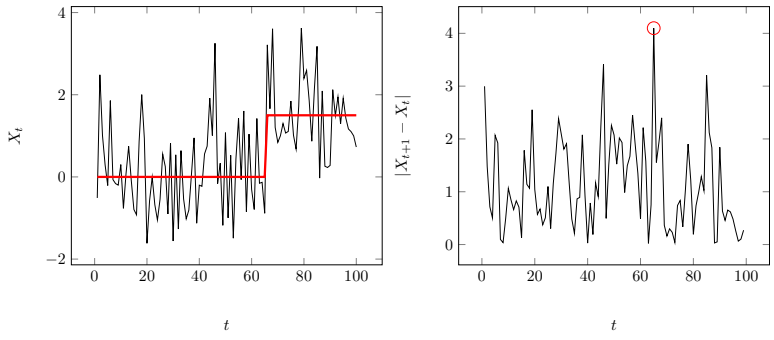

Figure 1: **Left plot:** The data (in black) and the underlying signal (in red) used for the toy example. The change-point is at location 65. **Right plot:** The absolute values of consecutive pairwise differences. The maximum, with location $d_{1,100} = 65$, is marked with a red circle.

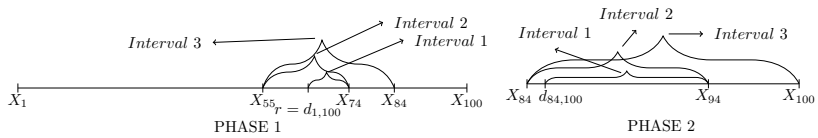

Figure 2: Example with one change-point at $r = 65$ with jump 1.5. **Left plot:** The largest difference is at $d_{1,100} = 65$ and the change-point is detected when the third expanding interval is checked. After detection the next intervals that are checked are $[1, 55]$ and $[84, 100]$. **Right plot:**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

We emphasise that the differences are only used as a way to start the algorithm in an area close to the change-point. In cases with high variance, σ^2, of the noise, the largest difference might not be around the change-point but under the assumptions specified in Section 3, the change-points will still be detected with high probability. The location of the largest difference that guarantees detection of the change-point in an interval in which it is isolated is discussed in Section 2.3, and further results are provided in Section 3.3 for the case that ϵ_t in (1) is Gaussian.

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 and $1 \leq s < e \leq T$. We denote by $N \in \mathbb{N}_0$ and by r_i for $i = 1, 2, \ldots, N$ the number and the locations of the true change-points, respectively, with the latter ones sorted in increasing order, while $\hat{N}$ and $\hat{r}_i$ for $i = 1, 2, \ldots, \hat{N}$ are their estimated values. The true, unknown number, N, of change-points is allowed to grow with the sample size. We denote by $\lambda_T \in \mathbb{N}$ the expansion parameter and $\zeta_T \in \mathbb{R}^+$ the threshold, which is used to decide whether a time-point is a change-point; more details on the choice of their values are given in Section 4.2. In addition, $d_{s,e}$ is the location of the largest difference, in absolute value, in the interval $[s, e]$ and is defined as

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

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 $K = K^{\max} + K^{\min}$, then 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, $L_{d_{s,e},s,e}$ and $R_{d_{s,e},s,e}$, 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 the end-points for the expansions are:

$$\begin{aligned} L_{d_{s,e},s,e} &= \left\{c^l_0, c^l_1, c^l_1, c^l_2, \ldots, c^l_{K^{\min}-1}, c^l_{K^{\min}}, c^l_{K^{\min}+1}, c^l_{K^{\min}+2}, \ldots, c^l_{K^{\max}}\right\}, \\ R_{d_{s,e},s,e} &= \left\{c^r_1, c^r_1, c^r_2, c^r_2, \ldots, c^r_{K^{\min}}, c^r_{K^{\min}}, c^r_{K^{\min}+1}, c^r_{K^{\min}+2}, \ldots, c^r_{K^{\max}}\right\}, \end{aligned} \tag{3}$$

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

$$I_j = \left[s_j, e_j\right] = \left[L_{d_{s,e},s,e}[j], R_{d_{s,e},s,e}[j]\right], \text{ for } j \in \{1, 2, \ldots, K\}, \tag{4}$$

where A[i] denotes the i^{th} element of the ordered set A. It is easy to show that $|L_{d_{s,e},s,e}| = |R_{d_{s,e},s,e}| = K$, 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 $\lceil \frac{e-s+1}{\lambda_T} \rceil + 1$. The intervals I_j defined in (4) are not deterministic. The end-points depend on the location of the largest difference, which depends on the structure of the unknown signal and the noise of the data sequence under consideration.

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). This is the step that makes the algorithm data-adaptive, differentiating it from the competitors, as the algorithm will check different intervals I_j for the same underlying signals (in terms of the type of the change, the length of the sequence and the change-point locations) if they have different observed noise, depending on the value of $d_{1,T}$. Applying the appropriate transformation (as in (2) based on the signal structure) to 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 for the observed data X_t will be at a point near the change-point with the largest magnitude of change. 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 of the point where the maximum absolute difference occurs at every step by an expansion parameter of magnitude λ_T. By doing this, we identify the location of the change-point that is closest to the detected largest difference. 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, for example, the change-point was detected in the subinterval $[s^*, e^*]$, then DAIS will be reapplied to $[1, s^*]$ and $[e^*, T]$. When checking any interval $[s, e]$ the maximum value attained by the contrast function at a point b within the interval, is compared to the chosen threshold ζ_T. If this value exceeds ζ_T, we take b to be a change-point, for $b \in \{s, s+1, \ldots, e-1\}$ in the case of piecewise-constant signals, and $b \in \{s, s+1, \ldots, e-2\}$ in the case of piecewise-linear signals. If no change-point is detected, then the algorithm terminates.

Algorithm 1 DAIS

function DAIS(s, e, λ_T, ζ_T)
if $e - s < 3$ **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)
$b_i = \text{argmax}_{\text{t}\in[\text{s}_\text{i},\text{e}_\text{i})} \text{C}^{\text{t}}_{\text{s}_\text{i},\text{e}_\text{i}}(\boldsymbol{X})$
if $C^{b_i}_{s_i,e_i}(\boldsymbol{X}) > \zeta_T$ **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

As mentioned in the introduction, the algorithm can be applied to more general signals, for example piecewise-polynomial signals of order p. In this case, the location of the largest difference needs to be generalised. In the more general case, it can be defined as the value of t for which the $(p+1)-$times differentiated signal is maximised.

There are two main advantages to the algorithm due to its data-adaptive nature. Firstly, since at each step DAIS starts checking for potential change-points around the largest difference, which as already explained we have reason to believe that it is around the most prominent change-point, then this change-point is detected fast. If the starting point is not near the change-point, detection still occurs with a slower speed. 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 some cases in balanced intervals. Such near balance in the distances of the change-points from the left and right end-points of the intervals under consideration does not necessarily appear in many state-of-the-art competitors, such as Binary Segmentation, ID, NOT, WBS, WBS2 and SeedBS. 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 2.2, for a given data sequence X_t, 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. For different data sequences, DAIS will start searching for change-points in different neighbourhoods of the data, based on where the most prominent change-point (if any) lies. This is exactly the novel data-adaptive nature of the proposed algorithm, compared to the existing algorithms in the literature that exhibit a fixed workflow, not adaptive to the data structure. In this subsection, we now 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 C, the detection of a change-point r_j, for $j \in \{1, \ldots, N\}, N \in \mathbb{N}$ 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 of course 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 $r_0 = 0$ and $r_{N+1} = T$, $\delta_T = \min_{j=1,\ldots,N+1}\{r_j - r_{j-1}\}$ is the smallest distance between two consecutive change-points and λ_T, as explained previously, 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:

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

Setting $n = \frac{\delta_T}{2\lambda_T}$ ensures that the detection happens as quickly as possible. Note that in [2] and [23] the choice $n = 3/2$ was made while in [6] the value of $n = 3$ was used 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 $\delta_T \geq 2n$. For $d_{s,e} \in [1, r_1]$ and $d_{s,e} \in [r_N, T]$, any value of $d_{s,e}$ guarantees isolation, as the left and right expanding intervals will not include a second change-point before detection occurs, besides r_1 and r_N, respectively. Now considering the more complicated case, where the location of the largest difference lies between two change-points, define $\tilde{\delta}_j = r_{j+1} - r_j$ 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 $J \in \{1, \ldots, N-1\}$. Isolation is guaranteed whenever the following holds

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

Note that the lower bound of n ensures that the right hand side of the second inequality in (6) is non-negative. The result 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 (35) and (49) in the Appendix 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 $\tilde{\delta}_j = \mathcal{O}(\delta_T)$ for any $j = 0, 1, \ldots, N$. Also, since $2n\lambda_T \leq \delta_T$, $n\lambda_T = \mathcal{O}(\delta_T)$ and, from (6), 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 $\sqrt{\delta_T}\underline{f}_T \geq \underline{C}\sqrt{\log T}$, where $\underline{f}_T = \min_j |f_{r_j+1} - f_{r_j}|$ is defined as the minimum magnitude of jumps. This assumption implies that when $T \to \infty$ then $\sqrt{\delta_T}\underline{f}_T$ increases. 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. More details on this, in the case that ϵ_t is Gaussian, are provided in Section 3.3. In either case, the probability that $d_{s,e}$ lies within $\delta_{s,e}^J$ of any change-point increases as T becomes large. We can reach the same conclusions using Assumptions (A3) and (A4) for the case of univariate piecewise-linear signals and multivariate signals, respectively.

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

$$d_{s,e} \in \left[1, r_1 + \frac{\tilde{\delta}_1}{2} - \frac{3\delta_T}{4n}\right] \cup \left[r_N - \frac{\tilde{\delta}_{N-1}}{2} + \frac{3\delta_T}{4n}, T\right] \cup \left\{\bigcup_{j=2}^{N-1} B_{n;j}\right\} \tag{7}$$

where

$$B_{n;j} = \left[r_j - \frac{\tilde{\delta}_{j-1}}{2} + \frac{3\delta_T}{4n}, r_j + \frac{\tilde{\delta}_j}{2} - \frac{3\delta_T}{4n}\right]. \tag{8}$$

Note that the length of the interval where $d_{s,e}$ cannot lie under this assumption is $2(\lceil 3\delta_T/4n \rceil - 1)$ around the midpoint of two consecutive change-points. This can be as small as $3\lambda_T - 2$, which 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 $\boldsymbol{x} = (x_1, x_2, \ldots x_T)$, as proposed by [26]. In the case of Gaussian data, for $C = 1.4826$,

$\hat{\sigma}$ is a consistent estimator of the population standard deviation σ as proven by [43]. 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 theoretical consistency results. Without loss of generality, we let $\sigma = 1$ in order to make notation easier in the proof of the theorems presented in this section. The model becomes

$$X_t = f_t + \epsilon_t, \quad t = 1, 2, \ldots, T. \tag{9}$$

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 C and D, respectively. In Section 6.3 we also provide theoretical results for the consistency in the case of multivariate signals and the outline of the proof can be found in Appendix E. In Section 3.3, we return to the discussion of Section 2.3 and present some results in the case when (7) holds, and 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 $t = r_{j-1}+1, \ldots, r_j$ 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

$$C_{s,e}^{b}(\boldsymbol{X}) = \left| \sqrt{\frac{e-b}{\ell(b-s+1)}} \sum_{t=s}^{b} X_t - \sqrt{\frac{b-s+1}{\ell(e-b)}} \sum_{t=b+1}^{e} X_t \right|, \tag{10}$$

where $s \leq b < e$ and $\ell = e - s + 1$. The value of $C_{s,e}^{b}(\boldsymbol{X})$ 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 (7) 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 $\underline{f}_T = \min_j \left| f_{r_j+1} - f_{r_j} \right|$ 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 and is required to ensure that every change-point can be detected in an interval in which it is isolated. It is not a strict assumption as for appropriately chosen values of n and λ_T, (7) includes any point besides the midpoint of two consecutive change-points. The exact probability of the consecutive pairwise difference being larger in absolute value 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. (A2) is not a very restrictive assumption, for example if $\delta_T \geq O(\log T)$, then $\underline{f}_T$ is allowed to decrease as T increases. Further discussion for (A2) is given after Theorem 1.

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 $\{X_t\}_{t=1,2,\ldots,T}$ follow model (9), with f_t being a piecewise-constant 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 assumptions (A1) and (A2) hold. Let N and $r_j, j = 1, 2, \ldots N$ be the number and location of the change-points, while $\hat{N}$ and $\hat{r}_j, j = 1, 2, \ldots \hat{N}$ their estimates, sorted in increasing order. In addition, $\Delta_j^f = \left| f_{r_j+1} - f_{r_j} \right|, j = 1, 2, \ldots N$ is the magnitude of each discontinuity 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 $C_1\sqrt{\log T} \leq \zeta_T < C_2\sqrt{\delta_T}\underline{f}_T$ and for sufficiently large T, we obtain*

$$\mathbb{P}\left(\hat{N} = N, \max_{j=1,2,\ldots,N} \left(|\hat{r}_j - r_j| \left(\Delta_j^f \right)^2 \right) \leq C_3 \log T \right) \geq 1 - \frac{C_4}{T}. \tag{11}$$

The lower bound of the probability in (11) 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 $\mathcal{O}(\log T)$. 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 [11] argues, the lowest possible $\delta_T \underline{f}_T^2$ that allows detection of the change-points, in the case of piecewise-constant signals is $\mathcal{O}(\log T - \log\log T)$. Therefore, the rate obtained by DAIS is optimal up to a rather negligible double logarithmic factor. As mentioned in [53], the rate we obtain has been established in various different papers. Rates for other methods which are the same or comparable to ours can be found in [17].

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. Through a log-likelihood approach, [6] shows that contrast function is $C^b_{s,e}(\boldsymbol{X}) = \left|\langle \boldsymbol{X}, \boldsymbol{\phi^b_{s,e}}\rangle\right|$, where the contrast vector $\boldsymbol{\phi^b_{s,e}} = (\phi^b_{s,e}(1), \ldots, \phi^b_{s,e}(T))$ is given by

$$\phi^b_{s,e}(t) = \begin{cases} \alpha^b_{s,e}\beta^b_{s,e}\big[(e+2b-3s+2)t - (be+bs-2s^2+2s)\big], & t \in \{s,\ldots,b\}, \\ -\frac{\alpha^b_{s,e}}{\beta^b_{s,e}}\big[(3e-2b-s+2)t - (2e^2+2e-be-bs)\big], & t \in \{b+1,\ldots,e\}, \\ 0, & \text{otherwise}, \end{cases} \tag{12}$$

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

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

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

Under the Gaussianity assumption, $C^b_{s,e}(\boldsymbol{X})$, 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 by DAIS 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 (9), 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, $\Delta^f_j = \left|2f_{r_j} - f_{r_j+1} - f_{r_j-1}\right|, j = 1, 2, \ldots N$ is the magnitude of each change of slope in f_t, $\underline{f}_T = \min_j \Delta^f_j$ 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*

$$\mathbb{P}\left(\hat{N} = N, \max_{j=1,2,\ldots,N}\left(|\hat{r}_j - r_j|\left(\Delta^f_j\right)^{2/3}\right) \leq C_3(\log T)^{1/3}\right) \geq 1 - \frac{C_4}{T}. \tag{13}$$

The proof of Theorem 2 can be found in Appendix D. 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 $\sqrt{\delta_T}\underline{f}_T$, which appears in (A2) for the case of piecewise-constant signals. For $\underline{f}_T \sim T^{1/2}$, the change-point detection accuracy is $O\left(T^{-1/3}(\log T)^{1/3}\right)$, which differs from the optimal rate $O(T^{-1/3})$ derived by [41] only by the logarithmic factor. This means that DAIS's change-point detection rate is almost optimal.

3.3 Probability of guaranteed isolation of the change-point at detection

We now return to the discussion of Section 2.3, that concerns the connection between the location of the largest difference and the guarantee that the detection of the change-point closest to it occurs in an interval in which it is isolated. The probability of this happening is given by

$$\mathbb{P}\left(d_{s,e} \in \left[1, r_1 + \frac{\tilde{\delta}_1}{2} - \frac{3\delta_T}{4n}\right] \cup \left[r_N - \frac{\tilde{\delta}_{N-1}}{2} + \frac{3\delta_T}{4n}, T\right] \cup \left\{\bigcup_{j=2}^{N-1} B_{n;j}\right\}\right) \tag{14}$$

where $B_{n;j}$ is defined in (8). 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:

$$\mathbb{P}\left(\bigcup_{j=1}^{N}\bigcap_{t=1}^{T-1}|X_{r_j+1}-X_{r_j}|\geq|X_{t+1}-X_t|\right). \tag{15}$$

It is easy to see that (14) $\geq$ (15). As a first step, we are interested in the probability that the largest difference occurs at the location of one of the N change-points. The focus is put on $t\notin\{r_j-1,r_j,r_j+1\}$ for $j\in\{1,\ldots,N\}$ such that $|X_{r_j+1}-X_{r_j}|$ and $|X_{t+1}-X_t|$ are independent. We will work on

$$\mathbb{P}\left(|X_{r_j+1}-X_{r_j}|\geq|X_{t+1}-X_t|\right), \tag{16}$$

which is the probability that the observations at the change-point r_j, for any $j\in\{1,\ldots,N\}$, 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 $t'\neq r_j, j\in\{1,\ldots,N\}$, (16) can be written as

$$\begin{aligned}\mathbb{P}\left(|X_{r_j+1}-X_{r_j}|\geq|X_{t+1}-X_t|\right)&=\mathbb{P}\left(|f_{r_j+1}-f_{r_j}+\sigma Z_2|\geq|\sigma\tilde{Z}_2|\right)\\&=\mathbb{P}\left(|\alpha_j+Z_2|\geq|\tilde{Z}_2|\right)\end{aligned} \tag{17}$$

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, (17) can be equivalently written as $\mathbb{P}(Y_1\geq Y_2)$ 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, (17) can be written as

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

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

$$\mathbb{P}(Y_1\geq Y_2)=\frac{1}{2}\left[1+\text{erf}\Big(\frac{\alpha_j}{2\sqrt{2}}\Big)^2\right], \tag{18}$$

so we have an explicit expression for (17). It is easy to see now that this probability goes to 1 as α_j gets larger. However, calculating or bounding from below the probability in (15), and subsequently in (14), are very difficult tasks 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 want to note that when $N=1$, the interval in assumption (A1) becomes $[1,T]$, which implies that isolation of the change-point in the interval in which detection occurs is guaranteed for all possible values of the largest difference. Also, in a data sequence with $N\geq 2$ change-points, we expect the order that they are detected to be with decreasing magnitude of change. This means that the change-point with the smallest magnitude of change will be detected last, in an interval in which it is certainly isolated, so the second smallest magnitude of change in the data sequence is the one that needs to be 'large enough' to ensure that all change-points will be isolated in an interval where detection can occur. We performed some simulations to check what values are considered 'large enough'.

In our simulations we consider a data sequence with $\sigma=1$ and two change-points at locations $r_1=20$ and $r_2=T-20$, with equal magnitudes of change, $\alpha=\alpha_1=\alpha_2$. It is easy to see that $\delta_T=20$ and $\tilde{\delta}_1=T-40$. We calculated the Monte Carlo estimates of (14), which is the probability that the location of the largest difference ensured that detection occurred in an interval where the change-point is isolated as T increased, using 10.000 repetitions for each value of T. In other words, choosing $n=\frac{\delta_T}{2\lambda_T}$ for $\lambda_T=3$ we calculated the proportion of times that $d_{1,T}\in\left[1,\frac{T}{2}-4.5\right]\cup\left[\frac{T}{2}+4.5,T\right]$.

In order to be consistent with assumption (A2), we have $\underline{f}_T=c\sqrt{\log T}$, where $c>0$ is a constant; in this case, $c=\frac{1}{2}$ was chosen for the sake of running the simulations and there is no other reason behind this specific choice. 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 can be seen by (18) and using the fact that $\underline{f}_T = \mathcal{O}(\sqrt{\log(T)})$ if δ_T is constant, which is the case in this simulation study. It is worth noting that the proportions are high even for small T and α, which provides evidence that the probability of isolation is extremely high.

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

T	α	proportion
60	0.989	0.951
80	1.047	0.967
100	1.073	0.971
150	1.119	0.981
200	1.151	0.985
250	1.175	0.990
500	1.246	0.994
750	1.286	0.996
1000	1.314	0.997
2000	1.378	0.999

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, as explained before, N is the total number of change-points of the signal, T is the length of the data sequence and λ_T is 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 at every time that detection occurs the already used observations are discarded and the algorithm is restarted on two disjoint subintervals, then the maximum total number of expansions required increases by two (one 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 [2], 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 more on the algorithms ID and DAIS, we need to highlight that the data-adaptive nature of DAIS leads to computational advancements. More specifically, due to the methodology of ID, when the data sequence under consideration has a small number of true change-points, the number of intervals that need to be checked will be close to the maximum possible for ID, which is $2\lceil \frac{T}{\lambda_T} \rceil$. In contrast, DAIS needs to check only K' intervals, which will be much smaller compared to ID. 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. This can be seen in the simulation results of signals (S11) and (S12) in Table 8 in Appendix B, which have 0 and 1 change-points, respectively, when compared with ID_th, which uses the same λ_T as DAIS. When many change-points are present, using again the definition of δ_T and the inequality $\delta_T \geq 2n\lambda_T$, for which the explanation is in Section 2.3, we get $N \leq T/(2n\lambda_T) < T/(2\lambda_T) - 1$. So, K' will always be smaller than the maximum number of intervals that are checked by ID. However, it needs to be noted that the data-adaptivity comes at an additional, but small, computational cost, since, at each step of the algorithm, the location of the largest difference, in absolute value, needs to be re-calculated.

4.2 Parameter selection

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

$$\zeta_T = C\sqrt{\log T}. \tag{19}$$

For the choice of the positive threshold constant C, we ran an extensive simulation study using various signal structures and Gaussian noise. The best behaviour 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. [2] 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.

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.

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 relevant R library. For ID, WBS and SeedBS we report the results for both the information criterion and the thresholding (for the default values $C = 1.56$ for ID and $C = \sqrt{2}$ for WBS and SeedBS) stopping rules. The notation is ID_ic, WBS_ic, SeedBS_ic and ID_th, WBS_th, SeedBS_th, respectively. In order to allow the competitors achieve their best performance, in some cases the inputs of the relevant functions as developed in R were adjusted. The competitors ID, WBS2, PELT and MARS were applied using the default values, while for DP_univar, the values proposed in the examples of the corresponding R function are used. For MOSUM, the function multiscale.localPrune was used, with the default values, which coincide with what is described in the simulation study of [16]. The maximum number of change-points to check for is required for the NOT algorithm and the number was chosen as $\max\{25, N + 1\}$, where 25 is the default value. For MSCP, the value of the minimal window considered is chosen to be $\min\{\lceil \delta_T/2 \rceil, 20\}$, where 20 is the default value. The estimation of the standard deviation for CPOP is set to be done using the MAD method as described in Section 3. TF and SeedBS are employed based on the implementations found in `https://github.com/hadley/l1tf` and `https://github.com/kovacssolt/SeedBinSeg/blob/master/SeedBS.R`, respectively. Change-point detection using local_poly involves performing cross validation for every time series, for the optimal values to be chosen. The degree of the polynomial is set to 1, so that the fitted signals are piecewise-linear. Code on the simulations, as well as on how to implement DAIS, can be found in `https://github.com/Sophia-Loizidou/DAIS`. For the simulations we have used a modification of the DAIS algorithm described in Section 2, in order to make it more accurate. More specifically, DAIS restarts from the detected change-point instead of the start- and end-points of the interval where the detection occurred. This leads to an improvement of the accuracy of the algorithm without affecting the speed. This is also our recommendation on how to use the algorithm in practice.

Table 2: Methods used in the simulations

Type of signal	Method Notation	Reference	R package
Piecewise-constant	ID	[2]	IDetect
	PELT	[30]	changepoint
	WBS	[23]	breakfast
	WBS2	[25]	breakfast
	NOT	[6]	not
	MOSUM	[37]	mosum
	MSCP	[33]	mscp
	SeedBS	[32]	-
	DP_univar	[49]	changepoints
Continuous, piecewise-linear	ID	[2]	IDetect
	CPOP	[21]	cpop
	NOT	[6]	not
	MARS	[22]	earth
	TF	[31]	-
	TS	[35]	trendsegmentR
	local_poly	[54]	changepoints

As a measure of the accuracy of the estimated number of change-points, we present 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 and $\hat{f}_t$ is the ordinary least squares approximation of f_t between two successive change-points. 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 $n_s = \max_{j=1,\ldots,N} \tilde{\delta}_j$ 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 computational time for each method is also reported. The top performing algorithm, in terms of accuracy in the number of change-points detected, is presented in bold, along with all methods that performed within 5%. The signals used can be found in Appendix A and some further simulation results are in Appendix B.

The signals included in this section contain more than one change-point. For signals with no or one change-point, please refer to Tables 8 and 10 in Appendix B. The signal (S1) (Table 3) is a particularly difficult structure for most algorithms because it includes a pair of consecutive change-points that shift the data sequence to opposite directions, which masks 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, which is the case in this signal. The best performing method for (S1) is DAIS, having at least a 7.5% improvement in the number of times the correct number of change-points is detected compared to all the other algorithms. The reason why DAIS performs very well is that it will most likely start the expansion around the change-points, as is discussed in Section 3.3, 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 $[s, e]$. For signals (S2) and (S3), the majority of the algorithms perform well; however, WBS_th and SeedBS_th tend to overestimate and MSCP seems to underestimate the number of the change-points, while DP_univar underestimates in signal (S2) and overestimates in signal (S3). For signals (S4), (S5) and (S6) in Tables 3 and 4, some methods struggle to detect the change-points, which are really close to each other. The signal (S4) includes change-points of different magnitudes and with different distances between them. ID_th and ID_ic perform well in different scenarios, but neither is consistent, which is expected as the thresholding approach is more appropriate for a large number of change-points, while the information criterion is better suited for a smaller number. Even though MOSUM performs well for signals with a small number of change-points, it fails to detect the change-points for (S4), (S5) and (S6). WBS2 performs well but is more computationally expensive compared to DAIS, with DAIS being 10-100 times faster than WBS2 in the simulations. 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 simulation results in Tables 5 and 6 show that in the case of piecewise-linear signals, DAIS performs at least as accurately as the algorithms ID_ic and CPOP, and better than other competitors available in the literature regarding the estimated number of change-points and their locations. The average computational time for DAIS is very low, and most of the times, significantly lower than the best-performing competitors. For signal (S10), DAIS outperforms all competitors except for CPOP in terms of the accuracy for the number of change-points detected. However, CPOP has a higher MSE and is at least four times slower compared to DAIS in all the examples presented.

6 DAIS extensions

6.1 Temporal dependence

We first consider relaxing the assumption of independence for the error terms ϵ_t. Temporal dependence is therefore introduced in the observed data sequence. In such cases, methods that reduce the autocorrelation in the time series can be applied. This can be done using a subsampling technique, which is briefly described below.

For a chosen integer s we subsample the observed time series X_t by choosing every s^{th} observation. This creates s different data sequences, for which the autocorrelation is reduced compared to the original data sequence. If, for example, $s = 5$ and $T = 1000$, the new data sequences are of the form $\{X_1, X_6, \ldots, X_{996}\}$, $\{X_2, X_7, \ldots, X_{997}\}$, $\{X_3, X_8, \ldots, X_{998}\}$, $\{X_4, X_9, \ldots, X_{999}\}$, $\{X_5, X_{10}, \ldots, X_{1000}\}$. DAIS is applied to each one of these smaller data

Table 3: Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signals (S1), (S2), (S3) and (S4). The average MSE, d_H and computational times are also given.

		$\hat{N} - N$									
Method	Signal	≤ 3	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS		0	15	0	**81**	2	2	0	0.014	0.089	0.004
ID_th		0	8	0	72	9	10	1	0.014	0.098	0.004
ID_ic		0	24	0	73	3	0	0	0.016	0.133	0.002
WBS_th		0	0	5	30	18	25	22	0.018	0.196	0.009
WBS_ic		0	30	0	68	2	0	0	0.016	0.166	0.025
WBS2	(S1)	0	10	2	75	7	1	5	0.015	0.091	0.446
PELT		0	77	0	23	0	0	0	0.026	0.397	0.001
NOT		0	26	0	72	2	0	0	0.016	0.146	0.024
MOSUM		0	15	0	74	6	4	1	0.014	0.108	0.006
MSCP		0	75	23	2	0	0	0	0.030	0.394	2.600
SeedBS_th		0	0	4	20	21	22	33	0.019	0.245	0.013
SeedBS_ic		0	32	0	66	2	0	0	0.017	0.176	0.013
DP_univar		0	0	0	6	3	10	81	0.045	0.334	0.138
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	58	28	12	2	0.025	0.018	0.006
WBS_ic		0	0	0	62	26	6	6	0.024	0.017	0.005
WBS2	(S2)	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	0.949	0.182	0.072
SeedBS_th		0	0	1	66	27	6	0	0.025	0.018	0.012
SeedBS_ic		0	0	1	66	27	6	0	0.025	0.018	0.012
DP_univar		100	0	0	0	0	0	0	0.240	0.067	0.001
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	45	18	21	16	1.771	0.036	0.010
WBS_ic		0	0	0	94	6	0	0	1.468	0.009	0.009
WBS2	(S3)	0	0	0	94	5	0	1	1.484	0.012	0.106
PELT		0	0	0	**100**	0	0	0	1.475	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		49	33	16	2	0	0	0	10.770	0.109	0.289
SeedBS_th		0	0	0	45	34	11	10	1.763	0.033	0.011
SeedBS_ic		0	0	0	**96**	4	0	0	1.599	0.010	0.011
DP_univar		0	0	0	0	0	0	100	4.544	0.112	0.004
DAIS		0	0	0	**98**	2	0	0	0.253	0.004	0.0014
ID_th		0	0	0	**96**	4	0	0	0.254	0.006	0.0008
ID_ic		22	11	42	24	1	0	0	1.540	0.111	0.0013
WBS_th		0	0	0	87	8	4	1	0.251	0.007	0.0009
WBS_ic		0	0	0	71	19	9	1	0.270	0.012	0.0018
WBS2	(S4)	0	0	0	**99**	1	0	0	0.235	0.004	0.0018
PELT		4	4	0	92	0	0	0	0.380	0.018	0.0001
NOT		0	0	0	92	7	1	0	0.242	0.007	0.0144
MOSUM		100	0	0	0	0	0	0	6.460	0.893	0.0001
MSCP		100	0	0	0	0	0	0	6.270	0.446	0.0257
SeedBS_th		0	0	3	86	6	5	0	0.267	0.008	0.0108
SeedBS_ic		0	0	0	73	20	5	2	0.265	0.011	0.0108
DP_univar		100	0	0	0	0	0	0	5.790	0.154	0.0001

Table 4: Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signals (S5) and (S6). The average MSE, d_H and computational times are also given.

		$\hat{N} - N$					
Method	Signal	-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	5	**95**	0.714	0.010	0.023
WBS_ic		99	0	1	3.960	0.980	0.023
WBS2	(S5)	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	3.930	0.057	1.400
SeedBS_th		0	0	**100**	0.454	0.009	0.012
SeedBS_ic		100	0	0	4.000	0.990	0.012
DP_univar		0	100	0	3.690	0.026	0.048
DAIS		0	0	**100**	0.234	0.003	0.013
ID_th		0	0	**100**	0.208	0.002	0.009
ID_ic		100	0	0	6.250	0.992	0.023
WBS_th		0	31	69	1.080	0.009	0.020
WBS_ic		100	0	0	6.250	0.992	0.020
WBS2	(S6)	0	0	**100**	0.199	0.001	0.284
PELT		19	38	43	1.670	0.082	0.001
NOT		100	0	0	6.250	0.992	0.041
MOSUM		100	0	0	6.250	0.992	0.003
MSCP		100	0	0	6.220	0.160	1.090
SeedBS_th		0	0	**100**	0.235	0.003	0.012
SeedBS_ic		100	0	0	6.250	0.992	0.013
DP_univar		100	0	0	5.710	0.025	0.030

Table 5: Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signal (S8). The average MSE, d_H and computational times are also given.

		$\hat{N} - N$							
Method	Signal	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS		0	**99**	1	0	0	0.030	0.082	0.024
ID_th		0	71	28	1	0	0.036	0.130	0.024
ID_ic		0	**95**	5	0	0	0.031	0.096	0.008
CPOP		0	**99**	1	0	0	0.013	0.050	5.550
NOT	(S8)	0	**99**	1	0	0	0.016	0.060	0.186
MARS		0	5	39	47	9	0.025	0.194	0.008
TF		0	0	0	0	100	0.018	0.440	0.768
TS		1	**99**	0	0	0	0.096	0.186	1.020
local_poly		0	1	5	12	82	0.054	0.503	125.400

Table 6: Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signals (S9) and (S10). The average MSE, d_H and computational times are also given.

		$\hat{N} - N$									
Method	Signal	≤ -15	$(-15, -2]$	-1	0	1	$[2, 15)$	≥ 15	MSE	d_H	Time (s)
DAIS		0	0	0	**100**	0	0	0	0.266	0.295	0.026
ID_th		0	0	0	67	30	3	0	0.278	0.332	0.014
ID_ic		0	0	0	**96**	4	0	0	0.202	0.251	0.411
CPOP		0	0	0	91	9	0	0	0.381	0.247	0.439
NOT	(S9)	100	0	0	0	0	0	0	4.730	99.00	1.440
MARS		100	0	0	0	0	0	0	4.700	98.50	0.003
TF		0	0	0	0	0	0	100	0.205	0.395	0.645
TS		0	33	37	30	0	0	0	0.854	0.580	1.430
local_poly		100	0	0	0	0	0	0	4.740	0.827	157.80
DAIS		0	0	0	**100**	0	0	0	0.038	0.196	0.016
ID_th		0	0	0	89	11	0	0	0.043	0.271	0.011
ID_ic		2	4	15	79	0	0	0	0.117	2.860	0.365
CPOP		0	0	0	**99**	1	0	0	0.252	0.291	0.062
NOT	(S10)	100	0	0	0	0	0	0	1.060	119.000	0.050
MARS		100	0	0	0	0	0	0	1.060	118.000	0.002
TF		0	0	0	0	0	1	99	0.191	0.307	0.300
TS		100	0	0	0	0	0	0	1.060	119.000	0.480
local_poly		100	0	0	0	0	0	0	1.060	76.700	16.900

sequences and we obtain s sets of detected change-points. We can then apply a majority voting rule, which discards any change-points that are not detected in at least η time series, where $\eta \leq s$ is a positive integer. The change-points are then transformed to represent locations in the original time series. The purpose of taking every s^{th} observation is to not include the observations which have the highest correlation between them. That is, the observations just before or just after each data point. Thus, for larger values of s, we manage to reduce the autocorrelation in the resulting disjoint data sequences, but we obtain a larger number of smaller data sequences, which has a negative impact on the detection accuracy. More details on this can be found in [1]. A different way to reduce autocorrelation is using a pre-averaging technique, such as the one described in Section 6.2, which involves averaging the data sequence before applying the algorithm in order to reduce the correlation between consecutive observations.

6.2 Heavy-tailed noise

In this subsection we relax the assumption that the error term ϵ_t is Gaussian and we instead consider a heavy-tailed distribution for ϵ_t. To improve the performance of DAIS in such cases, the data sequence can be pre-processed by averaging the time series to bring the noise closer to Gaussianity, as described in Section 4.5 of [2]. Using this technique, we take advantage of the Central Limit Theorem, the noise of the observations with large ϵ_t is reduced in absolute value, and we obtain a smaller data sequence with less extreme values for all the data points. The idea is that for a chosen integer s, we define $Q = \lceil T/s \rceil$ and set

$$X_q^* = \frac{1}{s} \sum_{t=(q-1)s+1}^{qs} X_t, \text{ for } t = 1, 2, \ldots, q \quad \text{and} \quad X_Q^* = \frac{1}{T - (Q-1)s} \sum_{t=(Q-1)s+1}^{T} X_t.$$

DAIS can then be applied on X_t^* to obtain change-points $\hat{r}_1^*, \ldots, \hat{r}_{\hat{N}}^*$. The change-points of the original time series X_t can be obtained using the transformation $\hat{r}_i = (\hat{r}_i^* - 1)s + \lfloor \frac{s+1}{2} \rfloor$, for $i = 1, \ldots, \hat{N}$. The choice of the value of s is important. On the one hand, taking larger values brings the data closer to Gaussianity. On the other hand, these values of s return smaller data sequences that have been pre-processed more and so we lose on the accuracy of both the location and the number of change-points detected. In practice we recommend $s = 5$.

The performance of DAIS in the case of heavy-tailed data can be found in Table 7. For the pre-averaging technique we use $s = 5$ and the heavy-tailed noise is chosen to be the Student's t-distribution with degrees of freedom being equal to either 5 or 7. Signals (S1) and (S7) are piecewise-constant and (S8) is piecewise-linear and their description can be found in Appendix A. As expected, the accuracy of DAIS is reduced compared to Section 5. Moreover, the more heavy-tailed the noise, the more the overdetection. This is again expected, as the spikes caused by the heavy-

tailed distribution of ϵ_t can be mistakenly detected as change-points by the algorithm, even after the pre-processing. However, DAIS performs relatively well for signal (S1) even for 5 degrees of freedom.

Table 7: Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signals (S1), (S7) and (S8) with noise following the Student's t-distribution with $d = 5, 7$ degrees of freedom. The average MSE, d_H and computational times including the pre-averaging are also given.

		$\hat{N} - N$									
d	Signal	≤ -3	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (ms)
5	(S1)	0	1	0	87	6	6	0	0.025	0.038	1.0
	(S7)	0	0	0	70	14	12	4	0.016	0.081	1.9
	(S8)	0	0	0	37	22	21	20	2	21.1	10
7	(S1)	0	0	0	92	4	3	1	0.022	0.023	1.0
	(S7)	0	0	0	83	6	8	3	0.015	0.060	0.8
	(S8)	0	0	0	82	8	10	0	2.01	21	10

6.3 Multivariate models

The DAIS algorithm can be extended to multivariate or high dimensional models. The model considered is the following:

$$\boldsymbol{X_t} = \boldsymbol{f_t} + \boldsymbol{\epsilon_t}, t = 1, ..., T, \tag{20}$$

where $\boldsymbol{X_t} \in \mathbb{R}^{d\times 1}$ are the observed data and $\boldsymbol{f_t} \in \mathbb{R}^{d\times 1}$ is the d-dimensional deterministic signal with structural changes at certain points. In the case of piecewise-constant signals, the structure of $\boldsymbol{f_t}$ is given by $\boldsymbol{f_t} = \boldsymbol{\mu_j}$ for $t \in \{r_{j-1}+1, \ldots, r_j\}$ and $\boldsymbol{f_{r_j}} \neq \boldsymbol{f_{r_j+1}}$ where $\boldsymbol{\mu_j} \in \mathbb{R}^{d\times 1}$ for $j = 1, 2, \ldots, N+1$. For continuous piecewise-linear signals, $\boldsymbol{f_t} = \boldsymbol{\mu_{j,1}} + \boldsymbol{\mu_{j,2}} t$ for $t \in \{r_{j-1}+1, \ldots, r_j\}$ and $\boldsymbol{f_{r_j-1}} + \boldsymbol{f_{r_j+1}} \neq 2\boldsymbol{f_{r_j}}$ where $\boldsymbol{\mu_{j,1}}, \boldsymbol{\mu_{j,2}} \in \mathbb{R}^{d\times 1}$. To ensure continuity, we need the additional constraint of $\boldsymbol{\mu_{k,1}} + \boldsymbol{\mu_{k,2}} r_k = \boldsymbol{\mu_{k+1,1}} + \boldsymbol{\mu_{k+1,2}} r_k$ for $k \in \{1, 2, \ldots, N\}$. As with the univariate case, the algorithm can also be applied to more general signal structures.

The calculation of the largest difference, that was used as a starting point for the univariate signal, needs to be adapted for the multivariate case. This can be done using mean-dominant norms $L : \mathbb{R}^d \to \mathbb{R}$, whose definition can be found in [9]. Some examples are

$$L_2(\boldsymbol{X_t}) := \sqrt{\frac{1}{d}\sum_{i=1}^{d} X_{t,i}^2}, \quad L_\infty(\boldsymbol{X_t}) := \sup_{i=1,\ldots,d} |X_{t,i}|. \tag{21}$$

The location of the largest difference of the interval $[s, e]$, $1 \leq s < e \leq T$ can be calculated as

$$d_{s,e}^{\text{multi}} = \begin{cases} \text{argmax}_{t\in\{s,s+1,\ldots,e-1\}} \left\{L(|\boldsymbol{X_{t+1}} - \boldsymbol{X_t}|)\right\}, & \boldsymbol{f_t} \text{ piecewise-constant,} \\ \text{argmax}_{t\in\{s,s+1,\ldots,e-2\}} \left\{L(|\boldsymbol{X_{t+2}} - 2\boldsymbol{X_{t+1}} + \boldsymbol{X_t}|)\right\}, & \boldsymbol{f_t} \text{ piecewise-linear.} \end{cases} \tag{22}$$

We define $\boldsymbol{y_t}$ to be the $\mathbb{R}^{d\times 1}$ vector with entries the value of the chosen contrast function, this being (10) for piecewise-constant signals and (12) for piecewise-linear ones. At each step of the algorithm, the goal is to decide whether the point that maximizes $L(\boldsymbol{y_t})$ is a change-point. The aggregation scheme for the contrast function, as just described, was also proposed by [3]. Algorithm 2 provides a pseudocode for DAIS in the case of multivariate signals, which we call MDAIS. For a discussion on the choice of the mean-dominant norm, we refer the reader to [1, 3, 12, 15].

Theoretical results for this variant of the algorithm about consistency for the number and locations of the estimated change-points can be proven. In the rest of this section, we provide the consistency theorem that concerns the case of a piecewise-constant $\boldsymbol{f_t}$ and the outline of its proof can be found in Appendix E. However, since the focus of this paper is proposing the novel data-adaptive methodology, we do not delve further in this. Following a similar methodology, theoretical results can also be proved for $\boldsymbol{f_t}$ being continuous piecewise-linear. For the statement of the theorem, we require the following assumption, which is equivalent to assumption (A2) for univariate f_t:

(A4) The quantities δ_T and $\underline{f}_T$ are connected by $\sqrt{\delta_T}\underline{f}_T \geq \underline{C}_M\sqrt{\log\left(Td^{1/4}\right)}$, for a large enough constant $\underline{C}_M$.

Similarly to the univariate case, the number of change-points N is allowed to grow with T and d. Note that the threshold constant, $\zeta_{T,d}$ depends on both the length of the data sequence and its dimension.

Theorem 3. *Let $\{\boldsymbol{X_t}\}_{t=1,\dots,T}$ follow model* (20) *with $\boldsymbol{f_t}$ piecewise-constant and let $\boldsymbol{\epsilon_t} \sim \mathcal{N}_d(\mathbf{0}, \Sigma)$, where $\Sigma \in \mathbb{R}^{d\times d}$ is positive definite and also that assumptions (A1) and (A4) 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, $\boldsymbol{\Delta_j} := |\boldsymbol{f_{r_j+1}} - \boldsymbol{f_{r_j}}|, j = 1, 2, \dots N$ is the magnitude of each discontinuity in $\boldsymbol{f_t}$, $\underline{f}_T := \inf_{j=1,\dots,N}\{L(\boldsymbol{\Delta_j})\}$ 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 $C_1\sqrt{\log\left(Td^{1/4}\right)} \le \zeta_{T,d} < C_2\sqrt{\delta_T}\underline{f}_T$ and for sufficiently large T, we obtain*

$$\mathbb{P}\left(\hat{N} = N, \max_{j=1,2,\dots,N}\left\{|\hat{r}_j - r_j|\,\alpha^2\right\} \le C_3\log\left(Td^{1/4}\right)\right) \ge 1 - \frac{C_4}{T}. \tag{23}$$

where $\alpha = \Delta_j^{q_j} = |f_{r_{j+1},q_j} - f_{r_j,q_j}|$ when $L(\cdot) = L_\infty(\cdot)$, and $\alpha = L_2(\boldsymbol{\Delta_j})$ when $L(\cdot) = L_2(\cdot)$, $q_j := \text{argmax}_{k=1,\dots,d}\tilde{X}_{s_j,e_j}^{\hat{r}_j,k}$ for $[s_j, e_j]$ being the interval where $\hat{r}_j$ is obtained and $\tilde{X}_{s,e}^{b,j} = \sqrt{\frac{e-b}{\ell(b-s+1)}}\sum_{t=s}^{b} X_{t,j} - \sqrt{\frac{b-s+1}{\ell(e-b)}}\sum_{t=b+1}^{e} X_{t,j}$, $\ell = e - s + 1$.

The outline of the proof of Theorem 3 is given in Appendix E. Note that this extension of our algorithm is robust to spatial dependence. Theoretical results related to this statement, as well as simulation results can be found in [3], where the extension of ID ([2]), namely the MID algorithm, is introduced.

Algorithm 2 MDAIS

function MDAIS$(s, e, \lambda_T, \zeta_T, L(\cdot))$
if $e - s < 1$ **then**
  STOP
else
  Set $d_{s,e}^{\text{multi}}$ as in (22)
  For $j \in \{1, 2, \dots, K\}$ let $I_j = [s_j, e_j]$ as in (4)
  $i = 1$
  (Main part)
  $b_i = \text{argmax}_{t\in[s_i,e_i)} L\left(C_{s_i,e_i}^t(\boldsymbol{X})\right)$
  if $L\left(C_{s_i,e_i}^{b_i}(\boldsymbol{X})\right) > \zeta_T$ **then**
    add b_i to the list of estimated change-points
    MDAIS$(s, s_i, \lambda_T, \zeta_T)$
    MDAIS$(e_i, e, \lambda_T, \zeta_T)$
  else
    $i = i + 1$
    if $i \le K$ **then**
      Go back to **(Main part)** and repeat
    else
      STOP
    end if
  end if
end if
end function

7 Real data

7.1 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 in Montgomery County, Maryland and can be found in `https://catalog.data.gov/dataset/crime`. We use daily observations starting from 22/01/2020, up to 31/08/2024. 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, [4], $\alpha : \mathbb{N} \to \mathbb{R}$, with $\alpha(x) = 2\sqrt{x + 3/8}$.

The top plot of Figure 3 is a plot of the real data (in black), along with the estimated underlying signal, $\hat{f}_t$, according to the change-points detected by DAIS (plotted in red). 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. 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 behavioural change around the holidays. The number of crimes returns close to the previous levels, on the 28^{th} of May 2021 and a further increase is observed at the end of the summer, 27^{th} of August 2021. This could possibly be due to people going on holidays out of town, leaving their homes more vulnerable to break-ins. 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} of the same month. An increase in the daily number of crimes reported is also detected on the 28^{th} of September 2023. As in the previous years, there is a drop on the 23^{rd} of December 2023 and an increase to a slightly higher average than before on the 23^{rd} of January 2024. Further increase to the number of daily crimes reported is observed on the 26^{th} of February and the 10^{th} of April 2024, with a decrease on the 4^{th} of June 2024. It is clear that the average number of crimes has increased in the last years. 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, 23/12/2022 and 23/12/2023. All four are near the Christmas holidays and a decrease in the daily number of crimes reported is noticed, followed by the trend increasing on the 17/01/2022, 26/12/2022 and 22/01/2024.

The fits obtained by three competitors, ID, MOSUM, and NOT, are shown in the bottom plot of Figure 3, plotted with blue, yellow and green lines, respectively. All methods obtain different signals on the same data, with ID and DAIS providing similar fits and NOT having minor differences with them. ID and NOT detect 11 and 10 out of the 15 change-points that DAIS detects, respectively. They both don't detect the change-points in May 2021 and the last 2 change-points detected by DAIS, in April and June of 2024. ID also does not detect the change-point just before Christmas 2020, while NOT does not detect the 2 change-points around Christmas 2022. MOSUM detects only 5 change-points, with the only change-point around the Christmas period being the one in 2020, which ID missed. By detecting change-points around the same period every year, we could argue that DAIS is detecting the seasonality around the Christmas holidays, which none of the other competitors managed to fully capture.

7.2 Euro to British pound exchange rate

The DAIS algorithm for changes in the slope is applied to the Euro to British pound exchange rate. The data are the weekly close prices adjusted for splits for the period between 07/04/2014 and 12/09/2024. The results can be seen in Figure 4. The first plot is the observed signal, while the other plots are the estimated piecewise-linear signals from DAIS, ID and CPOP. The change-points which are discussed in the next paragraph are indicated by dashed lines in the plot of the observed signal.

All three fits are very similar, with ID detecting the least number of change-points (27) and thus having a smoother signal, while DAIS and CPOP detect 42 and 60, respectively. The first important change-point, detected only by DAIS, on the week of 5^{th} of January of 2015 is the same week as the Charlie Hebdo shooting, the terrorist attack that occurred at the offices of the satirical weekly newspaper in Paris. After this, the exchange rate decreased, indicating that the attack had negative effects to the strength of the Euro. The next important change-point is an increase in the value of the exchange rate starting on the week of 9^{th} of November 2015. This change-point is detected by both DAIS and CPOP, while ID detects one just one week later. The change might have been caused by Prime Minister David Cameron's speech on the 10^{th} of November, during which he repeated his commitment to holding a referendum for Brexit before the end of 2017. A few days later, on the 13^{th}, the terrorist attacks in Paris took place. Both events had an impact on the exchange rate. The next change-point, from which a new increase in the observed value begins, is detected in the week of 13^{th} of June 2016, which is the week just before the referendum was held, on the 23^{rd} of June. It is worth noting that the drop in the week of the 16^{th} of March of 2020, that is detected by DAIS and CPOP, 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, with change-points detected by all methods in two consecutive weeks, on the 6^{th} and 13^{th}, falls around the time when fuel prices soared, according to `https://www.theguardian.com/theguardian/2022/jun/09`.

8 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

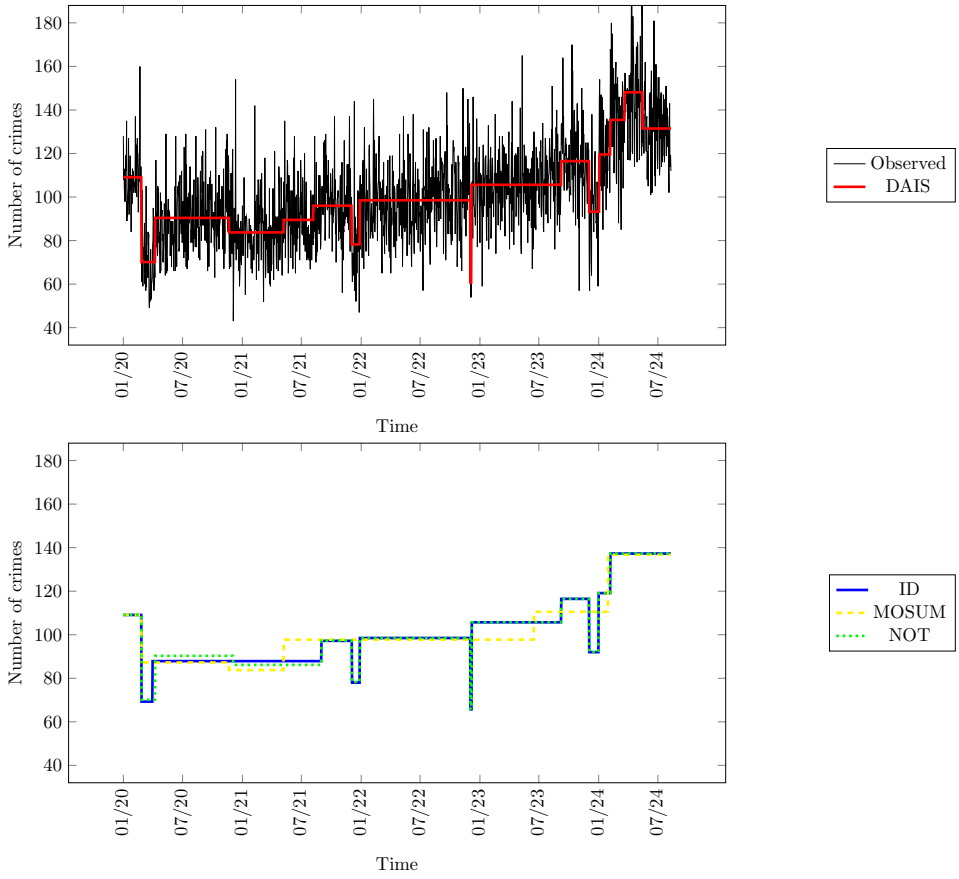

Figure 3: The observed time series data of daily crime reports in Montgomery County and fitted piecewise-constant mean signal obtained by applying DAIS (red), ID (blue), MOSUM (yellow) and NOT (green) to the data. The x-axis labels correspond to the month and year.

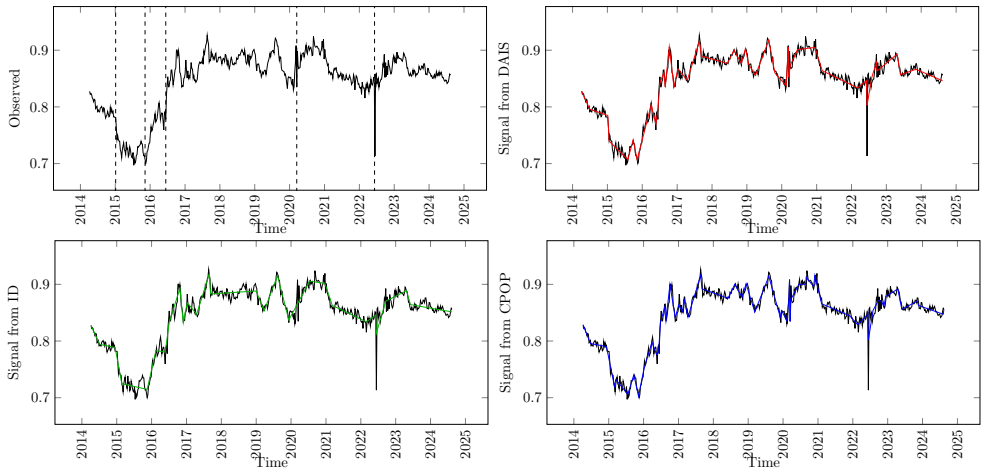

Figure 4: The observed time series data of Euro to British pound sterling exchange rate and fitted piecewise-linear signal obtained by applying DAIS (red), ID (green) and CPOP (blue).

(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. This step is what makes our method data-adaptive and differentiates it from the competitors. Expanding intervals around the location of the largest difference, in absolute value, are used to achieve isolation of the change-points, which enhances the detection power. Theoretical results regarding the consistency of the number of change-points detected and the accuracy of their estimated locations are provided. The simulation results presented 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. 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. In practice, the algorithm performs at least as well as other widely-used algorithms in the literature, exhibiting uniformly good behaviour across different signal structures. Finally, some extensions to more difficult signal structures are discussed, which include heavy tailed noise, temporal dependence and the extension of the method to multivariate signals.

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

(S2) $stairs$: 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 $\sigma = 0.3$.

(S3) 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 $\sigma = 4$.

(S4) $many\_cpts\_mix$: piecewise-constant signal of length 75 with 11 change-points at 5, 12, 17, 25, 31, 38, 44, 50, 56, 61, 67 with values between the change-points 0, 5, 0, 6, 0, 4, 0, 5, 0, 6, 0, 4. The standard deviation is $\sigma = 1$.

(S5) $many\_cpts$: 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 $\sigma = 1$.

(S6) $many\_cpts\_long$: piecewise constant signal of length 600 with 119 change-points at 5, 10, 15, . . ., 595 with values between change-points 0, 5, 0, . . ., 0, 5. The standard deviation is $\sigma = 1$.

(S7) $simple\_signal$: sequence of length 1100 with 1 change-point at 550 with values between change-points 0, 2. The standard deviation is $\sigma = 1$.

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

(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 Further simulations

Some further simulation results are presented in Tables 8 and 9 for piecewise-constant signals and Tables 10 and 11 for continuous piecewise-linear signals. The signals used are:

(S11) $justnoise$: sequence of length 6000 with no change-points. The standard deviation is $\sigma = 1$.

(S12) $long\_signal$: 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 $\sigma = 1$.

(S13) $small\_dist3$: 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 $\sigma = 1$.

(S14) $teeth$: 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 $\sigma = 0.4$.

(S15) $justnoise\_wave$: 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 $\sigma = 1$.

(S16) $wave4$: 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 $\sigma = 0.3$.

(S17) $wave5$: 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. Signal (S11), in Table 8 has 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 signal (S12), in which the big length of the signal 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. Signal (S13) in Table 9 indicates that in this difficult structure, 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 (S14) all algorithms perform similarly, with DAIS being one of the fastest methods. The low computational complexity of DAIS can also be seen in Table 11.

C Proof of Theorem 1

Before proving Theorem 1, we introduce some more notation. For convinience, we denote (10) by $\tilde{X}^b_{s,e} = \sqrt{\frac{e-b}{\ell(b-s+1)}}\sum_{t=s}^{b} X_t - \sqrt{\frac{b-s+1}{\ell(e-b)}}\sum_{t=b+1}^{e} X_t$. Similarly, 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 $\boldsymbol{\psi^b_{s,e}} = \left(\psi^b_{s,e}(1), \psi^b_{s,e}(2), \ldots, \psi^b_{s,e}(T_1)\right)$ is defined through the contrast function

$$\psi^b_{s,e}(t) = \begin{cases} \sqrt{\frac{e-b}{\ell(b-s+1)}}, & t = s, s+1, \ldots, b, \\ -\sqrt{\frac{b-s+1}{\ell(e-b)}}, & t = b+1, b+2, \ldots, e, \\ 0, & \text{otherwise} \end{cases}, \tag{24}$$

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 $\langle \boldsymbol{v}, \boldsymbol{\psi^b_{s,e}} \rangle = \tilde{v}^b_{s,e}$.

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,*

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

In addition,

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

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

Table 8: Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signals (S11) and (S12). The average MSE and computational times are also given.

		$\hat{N} - N$							
Method	Signal	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS		0	**99**	0	1	0	1.87×10^{-4}	-	0.237
ID_th		0	84	8	7	1	5.26×10^{-4}	-	0.450
ID_ic		0	**99**	1	0	0	1.68×10^{-4}	-	0.080
WBS_th		0	52	26	20	2	7.00×10^{-4}	-	0.067
WBS_ic		0	**100**	0	0	0	1.39×10^{-4}	-	0.066
WBS2	(S11)	0	91	5	3	1	2.75×10^{-4}	-	3.51
PELT		0	**100**	0	0	0	1.39×10^{-4}	-	0.003
NOT		0	**100**	0	0	0	1.39×10^{-4}	-	0.066
MOSUM		0	94	2	4	0	2.96×10^{-4}	-	0.028
MSCP		0	**100**	0	0	0	1.39×10^{-4}	-	83.25
SeedBS_th		0	12	23	25	40	12.9×10^{-4}	-	0.027
SeedBS_ic		0	**99**	1	0	0	1.68×10^{-4}	-	0.027
DP_univar		0	0	0	0	100	372×10^{-4}	-	30.70
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	75	19	6	0	5.15×10^{-4}	626.00×10^{-4}	0.107
WBS_ic		0	**100**	0	0	0	4.19×10^{-4}	1.06×10^{-4}	0.107
WBS2	(S12)	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
SeedBS_th		0	10	23	35	42	10.2×10^{-4}	3140×10^{-4}	0.0418
SeedBS_ic		0	**100**	0	0	0	4.19×10^{-4}	1.06×10^{-4}	0.0418
DP_univar		0	0	0	0	100	368×10^{-4}	4840×10^{-4}	190.80

Proof. See Lemma 4 from [6]. □

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

$$\mathbb{P}\left(\hat{N} = N, \max_{j=1,2,\ldots,N}\left(|\hat{r}_j - r_j|\left(\Delta^f_j\right)^2\right) \leq C_3 \log T\right) \geq 1 - \frac{1}{6\sqrt{\pi}T}, \tag{25}$$

which implies result (11).

Table 9: Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signals (S13) and (S14). The average MSE, d_H and computational times are also given.

		$\hat{N} - N$									
Method	Signal	≤ -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	1	31	28	26	14	0.0336	0.088	0.025
WBS_ic		2	33	1	64	0	0	0	0.0337	0.140	0.025
WBS2	(S13)	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		95	5	0	0	0	0	0	0.1390	0.593	2.610
SeedBS_th		0	0	1	18	27	14	40	0.0381	0.097	0.013
SeedBS_ic		1	29	2	63	3	1	1	0.0331	0.122	0.013
DP_univar		0	0	0	2	6	4	88	0.0602	0.126	0.139
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	56	21	19	4	2.58×10^{-2}	0.016	0.008
WBS_ic		0	0	0	**93**	7	0	0	2.35×10^{-2}	0.008	0.008
WBS2	(S14)	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	21.10×10^{-2}	0.261	0.242
SeedBS_th		0	0	1	49	25	15	10	2.68×10^{-2}	0.017	0.011
SeedBS_ic		0	0	0	**93**	5	2	0	2.40×10^{-2}	0.090	0.012
DP_univar		49	18	2	31	0	0	0	8.89×10^{-2}	0.210	0.003

Table 10: Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signal (S15). The average MSE and computational times are also given.

		$\hat{N} - N$					
Method	Signal	0	1	2	≥ 3	MSE	Time (s)
DAIS		**100**	0	0	0	2.12×10^{-3}	0.032
ID_th		83	10	4	3	2.45×10^{-3}	0.066
ID_ic		**100**	0	0	0	2.12×10^{-3}	0.007
CPOP		**100**	0	0	0	2.45×10^{-3}	3.220
NOT	(S15)	**100**	0	0	0	2.31×10^{-3}	0.062
MARS		0	100	0	0	5.25×10^{-3}	0.006
TF		7	44	37	12	5.29×10^{-3}	0.429
TS		**100**	0	0	0	2.12×10^{-3}	0.618
local_poly		10	5	17	68	6.47×10^{-3}	33.000

Table 11: Distribution of $\hat{N} - N$ over 100 simulated data sequences of the signals (S16) and (S17). The average MSE, d_H and computational times are also given.

		$\hat{N} - N$									
Method	Signal	≤ -15	$(-15, -2]$	-1	0	1	$[2, 15)$	≥ 15	MSE	d_H	Time (s)
DAIS		0	0	0	**96**	4	0	0	0.022	0.104	0.003
ID_th		0	0	0	76	19	5	0	0.021	0.149	0.002
ID_ic		0	0	0	**99**	1	0	0	0.013	0.097	0.004
CPOP		0	0	0	87	12	1	0	0.081	0.122	0.017
NOT	(S16)	0	0	1	**96**	1	2	0	0.015	0.097	0.082
MARS		0	84	13	3	0	0	0	3.720	2.160	0.003
TF		0	0	0	0	0	21	79	0.024	0.433	0.097
TS		0	0	0	**96**	4	0	0	0.094	0.198	0.110
local_poly		0	100	0	0	0	0	0	0.331	0.956	0.095
DAIS		0	1	2	**96**	1	0	0	0.461	0.300	0.012
ID_th		0	0	0	**92**	8	0	0	0.482	0.294	0.006
ID_ic		39	1	15	44	1	0	0	3.100	19.50	0.025
CPOP		0	0	0	89	11	0	0	1.650	0.323	0.025
NOT	(S17)	100	0	0	0	0	0	0	6.650	49.00	0.039
MARS		100	0	0	0	0	0	0	6.590	48.00	0.003
TF		0	0	0	0	0	82	18	1.200	0.323	0.110
TS		100	0	0	0	0	0	0	6.650	49.00	0.197
local_poly		100	0	0	0	0	0	0	6.610	5.220	0.639

Step 1: Allow us to denote by

$$A_T = \left\{ \max_{s,b,e:1\leq s\leq b<e\leq T} \left| \tilde{X}^b_{s,e} - \tilde{f}^b_{s,e} \right| \leq \sqrt{8\log T} \right\}. \tag{26}$$

We will show that $\mathbb{P}(A_T) \geq 1 - 1/(12\sqrt{\pi}T)$. From (9) and (10), 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 (24), we denote by

$$B_T = \left\{ \max_{1\leq s\leq b<e<T} \max_{\substack{r_{j-1}<s\leq r_j \\ r_j<e\leq r_{j+1} \\ s\leq b<e}} \frac{\left| \left\langle \boldsymbol{\psi^b_{s,e}}\langle \boldsymbol{f}, \boldsymbol{\psi^b_{s,e}}\rangle - \boldsymbol{\psi^r_{s,e}}\langle \boldsymbol{f}, \boldsymbol{\psi^r_{s,e}}\rangle, \boldsymbol{\epsilon} \right\rangle \right|}{\|\boldsymbol{\psi^b_{s,e}}\langle \boldsymbol{f}, \boldsymbol{\psi^b_{s,e}}\rangle - \boldsymbol{\psi^r_{s,e}}\langle \boldsymbol{f}, \boldsymbol{\psi^r_{s,e}}\rangle\|_2} \leq \sqrt{8\log T} \right\}. \tag{27}$$

Because

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

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

$$C_1 = \sqrt{C_3} + \sqrt{8}, C_2 = \frac{1}{\sqrt{4n}} - \frac{2\sqrt{2}}{\underline{C}}, C_3 = 2(2\sqrt{2}+4)^2, \tag{28}$$

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_j^L and I_j^R 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

$$\begin{aligned} c_m^l &= \max\{d_{s,e} - m\lambda_T, s\}, \quad m = 0, 1, \ldots, K^{\max}, \\ c_k^r &= \min\{d_{s,e} + k\lambda_T - 1, e\}, \quad k = 1, 2, \ldots, K^{\max}. \end{aligned} \tag{29}$$

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 $j \in \{1, 2, \ldots, N\}$ 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, for $j \in \{1, \ldots, N-1\}$, the largest difference $d_{s,e}$ will be at most at a distance $\delta_{1,T}^j$ from the nearest change-point r_j or r_{j+1}, where $\delta_{1,T}^j \leq \frac{\tilde{\delta}_j}{2} - \frac{3\delta_T}{4n}$ for $\tilde{\delta}_j = r_{j+1} - r_j$. 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 (30).

We will show that there exists an interval $[c_m^l, c_k^r]$, for $m, k \in \{0, 1, \ldots, K^{\max}\}$, such that the first change-point to get detected, r_J, is isolated, assuming that $d_{1,T} > r_J$. The same approach works for $d_{1,T} \leq r_J$. Without loss of generality, we suppose that $J \in \{1, 2, \ldots, N-1\}$. The result in the case that $J = N$ is easier to prove and so it is skipped. Note that since the closest change-point to the largest difference is r_J, it must hold that $d_{1,T}^J = d_{1,T} - r_J$. 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 $m = k$, $c_m^l \in I_J^L$ and $c_k^r - r_J > \delta_T/2n$, but it doesn't necessarily hold that $c_k^r \in I_J^R$. So, $r_J - c_m^l \leq \delta_T/n$ and $c_k^r - d_{1,T} + 1 = d_{1,T} - c_m^l$. It follows that

$$\begin{aligned} c_k^r - c_m^l &= c_k^r - d_{1,T} + d_{1,T} - c_m^l = 2(d_{1,T} - c_m^l) - 1 < 2(d_{1,T} - r_J + r_J - c_m^l) \\ &< 2\Big(\delta_{1,T}^J + \frac{\delta_T}{n}\Big) \leq \tilde{\delta}_J + \frac{\delta_T}{2n}. \end{aligned} \tag{30}$$

Now, since $r_J - c_m^l \leq \delta_T/n$, there can be at most one change-point in the interval $[c_m^l, r_J]$. Considering the interval $[r_J, c_k^r]$, using (30) and $r_J - c_m^l > \delta_T/2n$, it holds that $c_k^r - r_J = c_k^r - c_m^l + c_m^l - r_J < \tilde{\delta}_J$, so $c_k^r < r_{J+1}$ and r_J is isolated in $[r_J, c_k^r]$. The two results combined prove that the change-point is isolated in the interval $[c_m^l, c_k^r]$.

Now, we consider $J \in \{1, \ldots, N\}$. We will show that for $\tilde{b} = \text{argmax}_{c_m^l \leq t < c_k^r} |\tilde{X}_{c_m^l, c_k^r}^t|$, it holds that $\left|\tilde{X}_{c_m^l, c_k^r}^{\tilde{b}}\right| > \zeta_T$. Using (26), we have that

$$\left|\tilde{X}_{c_m^l, c_k^r}^{\tilde{b}}\right| \geq \left|\tilde{X}_{c_m^l, c_k^r}^{r_J}\right| \geq \left|\tilde{f}_{c_m^l, c_k^r}^{r_J}\right| - \sqrt{8\log T}. \tag{31}$$

But,

$$\left|\tilde{f}^{r_J}_{c^l_m,c^r_k}\right| = \left|\sqrt{\frac{c^r_k - r_J}{(c^r_k - c^l_m + 1)(r_J - c^l_m + 1)}}(r_J - c^l_m + 1)f_{r_J} - \sqrt{\frac{r_J - c^l_m + 1}{(c^r_k - c^l_m + 1)(c^r_k - r_J)}}(c^r_k - r_J)f_{r_J+1}\right| \tag{32}$$

$$= \left|\sqrt{\frac{(c^r_k - r_J)(r_J - c^l_m + 1)}{(c^r_k - c^l_m + 1)}}f_{r_J} - \sqrt{\frac{(r_J - c^l_m + 1)(c^r_k - r_J)}{(c^r_k - c^l_m + 1)}}f_{r_J+1}\right|$$

$$= \sqrt{\frac{(c^r_k - r_J)(r_J - c^l_m + 1)}{c^r_k - c^l_m + 1}}\Delta^f_J \geq \sqrt{\frac{(c^r_k - r_J)(r_J - c^l_m + 1)}{2\max\{c^r_k - r_J, r_J - c^l_m + 1\}}}\Delta^f_J$$

$$= \sqrt{\frac{\min\{c^r_k - r_J, r_J - c^l_m + 1\}}{2}}\Delta^f_J. \tag{33}$$

If we show that

$$\min\{c^r_k - r_J, r_J - c^l_m + 1\} \geq \frac{\delta_T}{2n} \tag{34}$$

then using assumption (A2) and the results in (31) (32), (34), we have that

$$\left|\tilde{X}^{\tilde{b}}_{c^l_m,c^r_k}\right| \geq \sqrt{\frac{\delta_T}{4n}}\Delta^f_J - \sqrt{8\log T} \geq \sqrt{\frac{\delta_T}{4n}}\underline{f_T} - \sqrt{8\log T}$$

$$= \left(\frac{1}{\sqrt{4n}} - \frac{2\sqrt{2\log T}}{\sqrt{\delta_T}\underline{f}_T}\right)\sqrt{\delta_T}\underline{f}_T \geq \left(\frac{1}{\sqrt{4n}} - \frac{2\sqrt{2}}{\underline{C}}\right)\sqrt{\delta_T}\underline{f}_T$$

$$= C_2\sqrt{\delta_T}\underline{f}_T > \zeta_T \tag{35}$$

and thus, the change-point will get detected.

But (34) 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 $r_J - c^l_m + 1 = r_J - d_{1,T} + d_{1,T} - c^l_m + 1 > c^r_k - d_{1,T} \geq c^r_k - r_J > \delta_T/2n$.
- 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 $k^* \in \{1, 2, \ldots K^{\max}\}$ 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 $b_J = \text{argmax}_{c^l_{m^*} \leq t < c^r_{k^*}}\left|\tilde{X}^t_{c^l_{m^*},c^r_{k^*}}\right|$, 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

$$\left(\tilde{X}^{r_J}_{c^l_{m^*},c^r_{k^*}}\right)^2 > \left(\tilde{X}^{b^*}_{c^l_{m^*},c^r_{k^*}}\right)^2. \tag{36}$$

Proving (36) and using the definition of b_J, we can conclude that $|b_J - r_J|\left(\Delta^f_J\right)^2 \leq \gamma_T$. Now, using (9), it can be shown, for $\boldsymbol{\psi}^{\boldsymbol{b}}_{\boldsymbol{s},\boldsymbol{e}}$ as defined in (24), that (36) is equivalent to

$$\left(\tilde{f}^{r_J}_{c^l_{m^*},c^r_{k^*}}\right)^2 - \left(\tilde{f}^{b^*}_{c^l_{m^*},c^r_{k^*}}\right)^2 > \left(\tilde{\epsilon}^{b^*}_{c^l_{m^*},c^r_{k^*}}\right)^2 - \left(\tilde{\epsilon}^{r_J}_{c^l_{m^*},c^r_{k^*}}\right)^2$$

$$+ 2\left\langle \boldsymbol{\psi}^{\boldsymbol{b^*}}_{\boldsymbol{c^l_{m^*}},\boldsymbol{c^r_{k^*}}}\langle \boldsymbol{f}, \boldsymbol{\psi}^{\boldsymbol{b^*}}_{\boldsymbol{c^l_{m^*}},\boldsymbol{c^r_{k^*}}}\rangle - \boldsymbol{\psi}^{\boldsymbol{r_J}}_{\boldsymbol{c^l_{m^*}},\boldsymbol{c^r_{k^*}}}\langle \boldsymbol{f}, \boldsymbol{\psi}^{\boldsymbol{r_J}}_{\boldsymbol{c^l_{m^*}},\boldsymbol{c^r_{k^*}}}\rangle, \boldsymbol{\epsilon}\right\rangle. \tag{37}$$

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 (37),

$$\left(\tilde{f}^{r_J}_{c^l_{m^*},c^r_{k^*}}\right)^2 - \left(\tilde{f}^{b^*}_{c^l_{m^*},c^r_{k^*}}\right)^2 = \frac{|r_J - b^*|(r_J - c^l_{m^*}+1)}{|r_J - b^*| + (r_J - c^l_{m^*}+1)}\left(\Delta^f_J\right)^2 :\ = \Lambda. \tag{38}$$

Also, for the right-hand side

$$\left(\tilde{\epsilon}^{b^*}_{c^l_{m^*},c^r_{k^*}}\right)^2 - \left(\tilde{\epsilon}^{r_J}_{c^l_{m^*},c^r_{k^*}}\right)^2 \leq \max_{s,e,b:s\leq b<e}\left(\tilde{\epsilon}^b_{s,e}\right)^2 \leq 8\log T \tag{39}$$

and using Lemma 1 and using (27),

$$\begin{aligned}
&2\Big\langle \boldsymbol{\psi}^{\boldsymbol{b^*}}_{\boldsymbol{c^l_{m^*}},\boldsymbol{c^r_{k^*}}}\langle \boldsymbol{f}, \boldsymbol{\psi}^{\boldsymbol{b^*}}_{\boldsymbol{c^l_{m^*}},\boldsymbol{c^r_{k^*}}}\rangle - \boldsymbol{\psi}^{\boldsymbol{r_J}}_{\boldsymbol{c^l_{m^*}},\boldsymbol{c^r_{k^*}}}\langle \boldsymbol{f}, \boldsymbol{\psi}^{\boldsymbol{r_J}}_{\boldsymbol{c^l_{m^*}},\boldsymbol{c^r_{k^*}}}\rangle, \boldsymbol{\epsilon}\Big\rangle \\
&\qquad \leq 2\big\|\boldsymbol{\psi}^{\boldsymbol{b^*}}_{\boldsymbol{c^l_{m^*}},\boldsymbol{c^r_{k^*}}}\langle \boldsymbol{f}, \boldsymbol{\psi}^{\boldsymbol{b^*}}_{\boldsymbol{c^l_{m^*}},\boldsymbol{c^r_{k^*}}}\rangle - \boldsymbol{\psi}^{\boldsymbol{r_J}}_{\boldsymbol{c^l_{m^*}},\boldsymbol{c^r_{k^*}}}\langle \boldsymbol{f}, \boldsymbol{\psi}^{\boldsymbol{r_J}}_{\boldsymbol{c^l_{m^*}},\boldsymbol{c^r_{k^*}}}\rangle\big\|_2\sqrt{8\log T} \\
&\qquad = 2\sqrt{\Lambda}\sqrt{8\log T}.
\end{aligned} \tag{40}$$

Using (38), (39) and (40), we can conclude that (37) 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 (38) and since

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

we can conclude that

$$\min\{|r_J - b^*|, r_J - c^l_{m^*}+1\} > \frac{2\left(2\sqrt{2}+4\right)^2\log T}{\left(\Delta^f_J\right)^2} = C_3\frac{\log T}{\left(\Delta^f_J\right)^2} \tag{41}$$

implies (36). Now, if

$$\min\{r_J - c^l_{m^*}+1, c^r_{k^*} - r_J\} > C_3\frac{\log T}{\left(\Delta^f_J\right)^2}, \tag{42}$$

then (41) is restricted to $|r_J - b^*|\left(\Delta^f_J\right)^2 > C_3\log T$ and this implies (36). So, we conclude that necessarily

$$|r_J - b_J|\left(\Delta^f_J\right)^2 \leq C_3\log T. \tag{43}$$

But (42) 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 (43).

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 (30) 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 $k_{J-1} \in \{1, \ldots, K^{\max}\}$ and $m_{J-1} \in \{k_{J-1} - 1, k_{J-1}\}$ 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 (30). Following similar steps as in (32), we have that for $\tilde{b}_{J-1} = \text{argmax}_{c^l_{m_{J-1}} \leq t < c^r_{k_{J-1}}} \left| \tilde{X}^t_{c^l_{m_{J-1}}, c^r_{k_{J-1}}} \right|$,

$$\begin{aligned} \left| \tilde{X}^{\tilde{b}_{J-1}}_{c^l_{m_{J-1}}, c^r_{k_{J-1}}} \right| &\geq \left| \tilde{f}^{\tilde{r}_{J-1}}_{c^l_{m_{J-1}}, c^r_{k_{J-1}}} \right| - \sqrt{8 \log T} \\ &\geq \sqrt{\frac{\min\{c^r_{k_{J-1}} - r_{J-1}, r_{J-1} - c^l_{m_{J-1}} + 1\}}{2}} \Delta^f_{J-1}. \end{aligned} \tag{44}$$

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

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

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

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

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,

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

and so

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

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 (44) we have that

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

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 $\max_{c^l_{\tilde{m}_{J-1}} \leq b < c^r_{\tilde{k}_{J-1}}} \left|\tilde{X}^t_{c^l_{\tilde{m}_{J-1}}, c^r_{\tilde{k}_{J-1}}}\right| > \zeta_T$.

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 $\left|\tilde{X}^{b_J-1}_{c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}}\right| > \zeta_T$. We can show that $\left|r_{J-1} - b_{J-1}\right|\left(\Delta^f_{J-1}\right)^2 \leq C_3\log T$, 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 $\left|r_j - b_j\right|\left(\Delta^f_j\right)^2 \leq C_3\log T$, $\forall j \in \{1, 2, \ldots, N\}$. 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 $b \in \{s^*, s^*+1, \ldots, e^*-1\}$,

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

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

□

D Proof of Theorem 2

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

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 $\Delta^f_j = \left|2f_{r_j} - f_{r_j+1} - f_{r_j-1}\right|$ and $\eta = \min\{r_j - s, e - r_j\}$. Then,*

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

Proof. See Lemma 5 from [6].

□

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 $\rho = |r - b|$, $\Delta^f_j = \left|2f_{r_j} - f_{r_j+1} - f_{r_j-1}\right|$, $\eta_L = r_j - s$ and $\eta_R = e - r_j$. Then,*

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

In addition,

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

2. *for any* $s \le b < r_j$, $\left(C^{r_j}_{s,e}(\boldsymbol{f})\right)^2 - \left(C^{b}_{s,e}(\boldsymbol{f})\right)^2 = \frac{1}{63}\min\{\rho, \eta_R\}^3\left(\Delta^f_j\right)^2$.

Proof. See Lemma 7 from [6]. □

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

Proof. We will prove the more specific result

$$\mathbb{P}\left(\hat{N} = N, \max_{j=1,2,\ldots,N}\left(|\hat{r}_j - r_j|\left(\Delta^f_j\right)^{2/3}\right) \le C_3(\log T)^{1/3}\right) \ge 1 - \frac{1}{6\sqrt{\pi}T}, \tag{45}$$

which implies result (13).

Steps 1 & 2: Similar to Theorem 1, denote

$$A^*_T = \left\{\max_{s,b,e:1\le s\le b<e\le T}\left|C^b_{s,e}(\boldsymbol{X}) - C^b_{s,e}(\boldsymbol{f})\right| \le \sqrt{8\log T}\right\}$$

$$B^*_T = \left\{\max_{1\le s\le b<e<T}\max_{\substack{r_{j-1}<s\le r_j\\ r_j<e\le r_{j+1}\\ s\le b<e}}\frac{\left|\left\langle \boldsymbol{\phi}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\langle \boldsymbol{f}, \boldsymbol{\phi}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\rangle - \boldsymbol{\phi}^{\boldsymbol{r}}_{\boldsymbol{s,e}}\langle \boldsymbol{f}, \boldsymbol{\phi}^{\boldsymbol{r}}_{\boldsymbol{s,e}}\rangle, \boldsymbol{\epsilon}\right\rangle\right|}{\|\boldsymbol{\phi}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\langle \boldsymbol{f}, \boldsymbol{\phi}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\rangle - \boldsymbol{\phi}^{\boldsymbol{r}}_{\boldsymbol{s,e}}\langle \boldsymbol{f}, \boldsymbol{\phi}^{\boldsymbol{r}}_{\boldsymbol{s,e}}\rangle\|_2} \le \sqrt{8\log T}\right\}. \tag{46}$$

The same reasoning as in the proof of Theorem 1 leads to $\mathbb{P}\left(A^*_T\right) \ge 1 - 1/(12\sqrt{\pi}T)$ and $\mathbb{P}\left(B^*_T\right) \ge 1 - 1/(12\sqrt{\pi}T)$. Therefore, it holds that

$$\mathbb{P}\left(A^*_T \cap B^*_T\right) \ge 1 - \frac{1}{6\sqrt{\pi}T}$$

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), $\delta_T^{3/2}\underline{f}_T \ge C^*\sqrt{\log T}$ and $3/2 \le n \le \delta_T/2\lambda_T$. As before, for $j \in \{1, 2, \ldots, N\}$ define I^L_j and I^R_j 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 $[s, e]$, $1 \le s < e \le T$, define c^l_m and c^r_k as in (29). Since the length of the intervals in (5) is $\delta_T/2n$ and $\lambda_T \le \delta_T/2n$, 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 $j \in \{1, 2, \ldots, N\}$.

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, for $j \in \{1, \ldots, N-1\}$, the largest difference $d_{s,e}$ will be at most at a distance $\delta^j_{1,T}$ from the nearest change-point r_j or r_{j+1}, where $\delta^j_{1,T} \le \frac{\tilde{\delta}_j}{2} - \frac{3\delta_T}{4n}$ for $\tilde{\delta}_j = r_{j+1} - r_j$. 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 $[c^l_m, c^r_k]$, for $m, k \in \{0, 1, \ldots, K^{\max}\}$, such that r_J is isolated using exactly the same argument as in (30). We will now show that for $\tilde{b}_J = \text{argmax}_{c^l_m \le t < c^r_k} C^t_{c^l_m, c^r_k}(\boldsymbol{X})$, then $C^{\tilde{b}_J}_{c^l_m, c^r_k}(\boldsymbol{X}) > \zeta_T$. Using (46), we have that

$$C^{\tilde{b}_J}_{c^l_m, c^r_k}(\boldsymbol{X}) \ge C^{r_J}_{c^l_m, c^r_k}(\boldsymbol{X}) \ge C^{r_J}_{c^l_m, c^r_k}(\boldsymbol{f}) - \sqrt{8\log T}. \tag{47}$$

From Lemma 2, we have that

$$C^{r_J}_{c^l_m, c^r_k}(\boldsymbol{f}) \ge \frac{1}{\sqrt{24}}\left(\min\{r_J - c^l_m, c^r_k - r_J\}\right)^{\frac{3}{2}}\Delta^f_J.$$

Showing that

$$\min\{c_k^r - r_J, r_J - c_m^l\} \geq \frac{\delta_T}{2n}. \tag{48}$$

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 (47), (48), we have that

$$\begin{aligned} C^{\tilde{b}_J}_{c^l_m, c^r_k}(\boldsymbol{X}) &\geq \frac{1}{\sqrt{24}}\left(\frac{\delta_T}{2n}\right)^{\frac{3}{2}} \Delta_J^f - \sqrt{8\log T} \geq \frac{1}{\sqrt{24}}\left(\frac{\delta_T}{2n}\right)^{\frac{3}{2}} \underline{f}_T - \sqrt{8\log T} \\ &= \left(\frac{1}{8\sqrt{3n^3}} - \frac{2\sqrt{2\log T}}{\delta_T^{3/2}\underline{f}_T}\right)\delta_T^{3/2}\underline{f}_T \geq \left(\frac{1}{8\sqrt{3n^3}} - \frac{2\sqrt{2}}{C^*}\right)\delta_T^{3/2}\underline{f}_T \\ &= C_2\delta_T^{3/2}\underline{f}_T > \zeta_T \end{aligned} \tag{49}$$

and thus, the change-point will get detected.

Therefore, 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 $\max_{c^l_{\tilde{m}} \leq t < c^r_{\tilde{k}}} C^t_{c^l_{\tilde{m}}, c^r_{\tilde{k}}}(\boldsymbol{X}) > \zeta_T$. For $k^*, m^* \in \{0, 1, \ldots K^{\max}\}$, 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 $b_J = \operatorname{argmax}_{c^l_{m^*} \leq t < c^r_{k^*}} C^t_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{X})$, 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 $b^* \in \{c^l_{m^*}, c^l_{m^*}+1, \ldots, c^r_{k^*}-1\}$ with $|b^* - r_J|\left(\Delta_J^f\right)^2 > \tilde{\gamma}_T$, we have that

$$\left(C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{X})\right)^2 > \left(C^{b^*}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{X})\right)^2. \tag{50}$$

Proving (50) and using the definition of b_J, we can conclude that $|b_J - r_J|\left(\Delta_J^f\right)^2 \leq \tilde{\gamma}_T$. Now, using (9), it can be shown, for $\boldsymbol{\phi}^{\boldsymbol{b}}_{\boldsymbol{s},\boldsymbol{e}}$ as defined in (12), that (50) is equivalent to

$$\begin{aligned} \left(C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{f})\right)^2 - \left(C^{b^*}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{f})\right)^2 &> \left(C^{b^*}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{\epsilon})\right)^2 - \left(C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{\epsilon})\right)^2 \\ &+ 2\left\langle \boldsymbol{\phi}^{\boldsymbol{b^*}}_{\boldsymbol{c^l_{m^*}, c^r_{k^*}}} \langle \boldsymbol{f}, \boldsymbol{\phi}^{\boldsymbol{b^*}}_{\boldsymbol{c^l_{m^*}, c^r_{k^*}}} \rangle - \boldsymbol{\phi}^{\boldsymbol{r_J}}_{\boldsymbol{c^l_{m^*}, c^r_{k^*}}} \langle \boldsymbol{f}, \boldsymbol{\phi}^{\boldsymbol{r_J}}_{\boldsymbol{c^l_{m^*}, c^r_{k^*}}} \rangle, \boldsymbol{\epsilon} \right\rangle. \end{aligned} \tag{51}$$

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

$$\Lambda \colon = \left(C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{f})\right)^2 - \left(C^{b^*}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{f})\right)^2. \tag{52}$$

For the right-hand side of (51) using (46),

$$\left(C^{b^*}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{\epsilon})\right)^2 - \left(C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{\epsilon})\right)^2 \leq \max_{s,e,b: s \leq b < e} \left(C^b_{s,e}(\boldsymbol{\epsilon})\right)^2 - \left(C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{\epsilon})\right)^2 \leq 8\log T \tag{53}$$

Using Lemma 3, (46) and (52), we have that for the left-hand side,

$$\begin{aligned} &2\left\langle \boldsymbol{\phi}^{\boldsymbol{b^*}}_{\boldsymbol{c^l_{m^*}, c^r_{k^*}}} \langle \boldsymbol{f}, \boldsymbol{\phi}^{\boldsymbol{b^*}}_{\boldsymbol{c^l_{m^*}, c^r_{k^*}}} \rangle - \boldsymbol{\phi}^{\boldsymbol{r_J}}_{\boldsymbol{c^l_{m^*}, c^r_{k^*}}} \langle \boldsymbol{f}, \boldsymbol{\phi}^{\boldsymbol{r_J}}_{\boldsymbol{c^l_{m^*}, c^r_{k^*}}} \rangle, \boldsymbol{\epsilon} \right\rangle \\ &\quad \leq 2\left\| \boldsymbol{\phi}^{\boldsymbol{b^*}}_{\boldsymbol{c^l_{m^*}, c^r_{k^*}}} \langle \boldsymbol{f}, \boldsymbol{\phi}^{\boldsymbol{b^*}}_{\boldsymbol{c^l_{m^*}, c^r_{k^*}}} \rangle - \boldsymbol{\phi}^{\boldsymbol{r_J}}_{\boldsymbol{c^l_{m^*}, c^r_{k^*}}} \langle \boldsymbol{f}, \boldsymbol{\phi}^{\boldsymbol{r_J}}_{\boldsymbol{c^l_{m^*}, c^r_{k^*}}} \rangle \right\|_2 \sqrt{8\log T} \\ &\quad = 2\sqrt{\Lambda}\sqrt{8\log T}. \end{aligned} \tag{54}$$

Using (52), (53) and (54), we can conclude that (51) 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

$$\begin{aligned} &\Lambda > \left(2\sqrt{2} + 4\right)^2 \log T \\ &\Leftrightarrow \frac{1}{63}\left(\min\{|r_J - b^*|, r_J - c^l_{m^*}\}\right)^3 \left(\Delta_J^f\right)^2 > \left(2\sqrt{2} + 4\right)^2 \log T \\ &\Leftrightarrow \min\{|r_J - b^*|, r_J - c^l_{m^*}\} > \frac{(63\log T)^{1/3}\left(2\sqrt{2}+4\right)^{2/3}}{\left(\Delta_J^f\right)^{2/3}} = \frac{C_3\left(\log T\right)^{1/3}}{\left(\Delta_J^f\right)^{2/3}} \end{aligned} \tag{55}$$

Now, if for sufficiently large T

$$\min\{r_J - c^l_{m^*}, c^r_{k^*} - r_J\} > 2^{1/3} C_3 \frac{(\log T)^{1/3}}{\left(\Delta^f_J\right)^{2/3}} - 1, \tag{56}$$

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 (55) is restricted to

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

which implies (50). So, we conclude that necessarily

$$|r_J - b_J| \left(\Delta^f_J\right)^{2/3} \leq C_3 \left(\log T\right)^{1/3}. \tag{57}$$

But (56) must be true since if we assume that

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

then, using Lemma 2, we have that

$$\begin{aligned}
C^{b_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{X}) &\leq C^{r_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{f}) + \sqrt{8 \log T} \\
&\leq \frac{1}{\sqrt{3}} \left(\min\{r_J - c^l_{m^*}, c^r_{k^*} - r_J\} + 1\right)^{3/2} \Delta^f_J + \sqrt{8 \log T} \\
&\leq \frac{1}{\sqrt{3}} \left(2^{1/3} C_3 \frac{(\log T)^{1/3}}{\left(\Delta^f_J\right)^{2/3}} \right)^{3/2} \Delta^f_J + \sqrt{8 \log T} \\
&= \sqrt{\frac{2}{3}} C_3^{3/2} \sqrt{\log T} + \sqrt{8 \log T} \\
&= \left(\sqrt{\frac{2}{3}} C_3^{3/2} + \sqrt{8} \right) \sqrt{\log T} = C_1 \sqrt{\log T} \leq \zeta_T,
\end{aligned}$$

which contradicts $C^{b_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{X}) > \zeta_T$.

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 $[c^l_{m^*}, c^r_{k^*}]$ with $C^{b_J}_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{X}) > \zeta_T$, where $b_J = \text{argmax}_{c^l_{m^*} \leq t < c^r_{k^*}} C^t_{c^l_{m^*}, c^r_{k^*}}(\boldsymbol{X})$ is the estimated location for the change-point r_J that satisfies (57).

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 $[c^r_{k^*}, T]$. Proving that there is no other change-point in the interval $[c^l_{m^*}, c^r_{k^*}]$ 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 $[1, c^l_{m^*}]$. 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 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 $k_{J-1} \in \{1, \ldots, K^{\max}\}$ and $m_{J-1} \in \{k_{J-1}-1, k_{J-1}\}$ 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

$$b_{J-1} = \text{argmax}_{c^l_{m^*_{J-1}} \leq t < c^r_{k^*_{J-1}}} C^t_{c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}}(\boldsymbol{X}),$$

which satisfies $\left|r_{J-1} - b_{J-1}\right| \left(\Delta^f_{J-1}\right)^{2/3} \leq C_3 (\log T)^{1/3}$. 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 (30). Using Lemma 3, we have that for $\tilde{b}_{J-1} = \text{argmax}_{c^l_{m_{J-1}} \leq t < c^r_{k_{J-1}}} C^t_{c^l_{m_{J-1}}, c^r_{k_{J-1}}}(\boldsymbol{X})$,

$$\begin{aligned} C^{\tilde{b}_{J-1}}_{c^l_{m_{J-1}}, c^r_{k_{J-1}}}(\boldsymbol{X}) &\geq C^{\tilde{r}_{J-1}}_{c^l_{m_{J-1}}, c^r_{k_{J-1}}}(\boldsymbol{X}) \geq C^{\tilde{r}_{J-1}}_{c^l_{m_{J-1}}, c^r_{k_{J-1}}}(\boldsymbol{f}) - \sqrt{8 \log T} \\ &\geq \frac{1}{\sqrt{24}} \left(\min\{r_{J-1} - c^l_{m_{J-1}}, c^r_{k_{J-1}} - r_{J-1}\}\right)^{3/2} \Delta^f_{J-1} - \sqrt{8 \log T}. \end{aligned} \tag{58}$$

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 (58), using exactly the same calculations as (49), 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 $\left|r_{J-1} - b_{J-1}\right| \left(\Delta^f_{J-1}\right)^{2/3} \leq C_3 (\log T)^{1/3}$, 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 $\left|r_j - b_j\right| \left(\Delta^f_j\right)^{2/3} \leq C_3 (\log T)^{1/3}, \forall j \in \{1, 2, \ldots, N\}$. 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.

□

E Outline of the proof of Theorem 3

Before providing the outline of the proof, we need some notation. For $1 \leq s \leq b < e \leq T$, we denote by $\tilde{\boldsymbol{X}}^{\boldsymbol{b}}_{\boldsymbol{s,e}} := (\tilde{X}^{b,1}_{s,e}, \tilde{X}^{b,2}_{s,e}, \ldots, \tilde{X}^{b,d}_{s,e})$, where $\tilde{X}^{b,j}_{s,e} = \sqrt{\frac{e-b}{\ell(b-s+1)}} \sum_{t=s}^{b} X_{t,j} - \sqrt{\frac{b-s+1}{\ell(e-b)}} \sum_{t=b+1}^{e} X_{t,j}$, which is equivalent to (10), $\ell = e - s + 1$ and $X_{t,j}$ is the j^{th} component of $\boldsymbol{X_t}$, for $j \in \{1, \ldots, d\}$. Furthermore, $\tilde{\boldsymbol{f}}^{\boldsymbol{b}}_{\boldsymbol{s,e}} := (\tilde{f}^{b,1}_{s,e}, \tilde{f}^{b,2}_{s,e}, \ldots, \tilde{f}^{b,d}_{s,e})$, where $\tilde{f}^{b,j}_{s,e} = \sqrt{\frac{e-b}{n(b-s+1)}} \sum_{t=s}^{b} f_{t,j} - \sqrt{\frac{b-s+1}{n(e-b)}} \sum_{t=b+1}^{e} f_{t,j}$ is the value of the CUSUM statistic at the location b, when we work in the interval from s up to e, for the j^{th} underlying component signal. Also, for $L(\cdot)$ any mean-dominant norm as those in (21), and for $|\boldsymbol{x}|$ being the vector of the absolute value of

the elements of an $\boldsymbol{x} \in \mathbb{R}^d$, $C^b_{s,e} = L\left(\left|\tilde{\boldsymbol{X}}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\right|\right)$ and $D^b_{s,e} = L\left(\left|\tilde{\boldsymbol{f}}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\right|\right)$. Proving Theorem 3 involves the following steps:

Step 1: We show that the values of the mean-dominant norm of the contrast function of the observed and unobserved signals, $L\left(\tilde{\boldsymbol{X}}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\right)$ and $L\left(\tilde{\boldsymbol{f}}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\right)$, are uniformly close, for all $1 \leq s \leq b < e \leq T$. Using the mean-dominance property as given in p.190 of [9], it holds that $0 \leq L_1(\boldsymbol{x}) \leq L(\boldsymbol{x}) \leq L_\infty(\boldsymbol{x}), \forall \boldsymbol{x} \in (\mathbb{R}^d)^+$ and so, for $A^*_T = \left\{\max_{1\leq s\leq b<e\leq T} L\left(\left|\tilde{\boldsymbol{X}}^{\boldsymbol{b}}_{\boldsymbol{s,e}} - \tilde{\boldsymbol{f}}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\right|\right) \leq \sqrt{8\log\left(Td^{\frac{1}{4}}\right)}\right\}$ and $A_T = \left\{\max_{1\leq s\leq b<e\leq T}\left|C^b_{s,e} - D^b_{s,e}\right| \leq \sqrt{8\log\left(Td^{\frac{1}{4}}\right)}\right\}$, it holds that $A^*_T \subseteq A_T$. From the definition of our model in (20), we have that $\boldsymbol{\epsilon_t} \sim \mathcal{N}_d(\boldsymbol{0}, \Sigma), \forall t \in \{1, \ldots, T\}$ and, using the Bonferroni inequality, we can show, following similar steps as those in [3], that $\mathbb{P}\left((A^*_T)\right) \leq 1/(12\sqrt{\pi}T)$, which implies $\mathbb{P}\left(A_T\right) \geq \mathbb{P}\left(A^*_T\right) \geq 1/(12\sqrt{\pi}T)$.

Step 2: We control the distance between $L\left(\tilde{\boldsymbol{X}}^{\boldsymbol{b_1}}_{\boldsymbol{s,e}}\right) - L\left(\tilde{\boldsymbol{X}}^{\boldsymbol{b_2}}_{\boldsymbol{s,e}}\right)$ and $L\left(\tilde{\boldsymbol{f}}^{\boldsymbol{b_1}}_{\boldsymbol{s,e}}\right) - L\left(\tilde{\boldsymbol{f}}^{\boldsymbol{b_2}}_{\boldsymbol{s,e}}\right)$, for all possible combinations of s, e, b_1, b_2, where $1 \leq s < e \leq T$. For $\boldsymbol{f_k} = (f_{1,k}, \ldots, f_{T,k})$ and for $[s, e]$ being any interval that contains only one true change-point, namely r_j, we denote

$$A^b_{s,e}(k, r_j) := \frac{\left\langle \boldsymbol{\psi}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\left\langle \boldsymbol{f_k}, \boldsymbol{\psi}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\right\rangle - \boldsymbol{\psi}^{\boldsymbol{r_j}}_{\boldsymbol{s,e}}\left\langle \boldsymbol{f_k}, \boldsymbol{\psi}^{\boldsymbol{r_j}}_{\boldsymbol{s,e}}\right\rangle, \boldsymbol{\epsilon}\right\rangle}{\left\|\boldsymbol{\psi}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\left\langle \boldsymbol{f_k}, \boldsymbol{\psi}^{\boldsymbol{b}}_{\boldsymbol{s,e}}\right\rangle - \boldsymbol{\psi}^{\boldsymbol{r_j}}_{\boldsymbol{s,e}}\left\langle \boldsymbol{f_k}, \boldsymbol{\psi}^{\boldsymbol{r_j}}_{\boldsymbol{s,e}}\right\rangle\right\|_2}, \tag{59}$$

where $\psi^b_{s,e}$ is defined in (24). In (59) above, $k \in \{1, \ldots, d\}$, while $\|\cdot\|_2$ is the Euclidean norm. Due to $\epsilon_{t,k}, t = 1, \ldots, T$ following the standard normal distribution for all $k \in \{1, \ldots, d\}$ (with the univariate error terms being independent over time, but possibly spatially dependent), then straightforward calculations lead to $A^b_{s,e}(k, r_j) \sim \mathcal{N}(0, 1), \forall k \in \{1, \ldots, d\}$. For

$$B_T = \left\{\max_{\substack{j=1,\ldots,N\\ k=1,\ldots,d}} \max_{\substack{r_{j-1}<s\leq r_j\\ r_j<e\leq r_{j+1}\\ s\leq b<e}} \left|A^b_{s,e}(k, r_j)\right| \leq \sqrt{8\log\left(Td^{\frac{1}{4}}\right)}\right\}, \tag{60}$$

we can show that $\mathbb{P}\left(B_T\right) \geq 1 - 1/(12\sqrt{\pi}T)$ using a similar procedure as in Step 1. From Steps 1 and 2, we conclude that $\mathbb{P}\left(A^*_T \cap B_T\right) \geq 1 - \frac{1}{6\sqrt{\pi}T}$.

Step 3: Working under the assumption that both A^*_T (and thus also A_T) and B_T hold, we show that detection in an interval where the change-point is isolated will occur and so the proof can be restricted to an interval with a single change-point and we also prove that the estimated and true change-points are close to each other. The values used in this step for the positive constants C_1, C_2, and C_3 of Theorem 3, are $C_1 = \sqrt{C_3} + \sqrt{8}, C_2 = \frac{1}{\sqrt{4n}} - \frac{2\sqrt{2}}{C_M}, C_3 = 2(2\sqrt{2}+4)^2$. The order with which the change-points are detected depends on the location of the largest difference at each step of the algorithm and so, without loss of generality, we assume that the first change-point to get detected is r_J for $J \in \{1, \ldots, N\}$. For c^l_m, c^r_k as defined in (29), $c^l_m \in I^L_J$ or $c^r_k \in I^R_J$ hold (or both), with I^L_J, I^R_J as defined in (5), and showing that the change-point will be isolated in $[c^l_m, c^r_k]$ is the same as for Theorem1 Step 3. It can be shown that for $\tilde{b} = \text{argmax}_{c^l_m \leq t < c^r_k} C^t_{c^l_m, c^r_k}$, $C^{\tilde{b}}_{c^l_m, c^r_k} > \zeta_{T,d}$, so r_J can be detected in the interval.

Now, for $k^* \in \{1, \ldots, K^{\max}\}$ and $m^* \in \{k^* - 1, k^*\}$, we denote by $c^l_{m^*} \geq c^l_m$ and $c^r_{k^*} \leq c^r_k$ the first left- and right-expanding points, respectively, where detection of r_J occurs. Showing that the estimated and true change-points are close to each other requires working on the two mean-dominant norms of (21) separately. Denoting by $q_J := \text{argmax}_{j=1,\ldots,d}\left|\tilde{X}^{b_J,j}_{c^l_{m^*}, c^r_{k^*}}\right|$ and $\alpha = \begin{cases}\Delta^{q_J}_J, \text{ when } L(\cdot) = L_\infty(\cdot)\\ L_2(\boldsymbol{\Delta_J}), \text{ when } L(\cdot) = L_2(\cdot)\end{cases}$, we show that $|b_J - r_J|\,\alpha^2 \leq C_3 \log\left(Td^{\frac{1}{4}}\right)$. This can be done using a contradiction argument, based on the fact that for any $b^* \in \left\{c^l_{m^*}, \ldots, c^r_{k^*} - 1\right\}$ with $|b^* - r_J|\,\alpha^2 > \log\left(Td^{\frac{1}{4}}\right)$, it holds that $\left(\tilde{X}^{r_J,q_J}_{c^l_{m^*}, c^r_{k^*}}\right)^2 > \left(\tilde{X}^{b^*,q_J}_{c^l_{m^*}, c^r_{k^*}}\right)^2$, which, using the definition of b_J, cannot be true. The detailed proof of this is skipped, but we highlight that it basically relies on a

combination of the strategy followed in [3] that the consistency of the MID algorithm was proved, and the approach followed in our Theorem 1, Step 3.

Step 4: We show that when the algorithm re-starts, no change-points are left undetected and that the new intervals allow for the detection of the remaining change-points. The first part is immediate, as from Step 3 we know that the change-point was isolated in the interval in which detection occurred. Since the change-point was detected in the interval $[c^l_{m^*}, c^r_{k^*}]$, MDAIS restarts on $[1, c^l_{m^*}]$ and $[c^r_{k^*}, T]$. The proof that detection of r_j occurs is the same as in Step 3 for all $j \in \{1, \ldots, J-2, J+2, \ldots, N\}$. We focus on $j = J-1$ and similar steps hold for $j = J+1$. For $k_{J-1} \in \{1, \ldots, K^{\max}\}$ and $m_{J-1} \in \{k_{J-1}-1, k_{J-1}\}$, 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 $c^r_{k^*_{J-1}} \leq c^r_{k_{J-1}} \leq c^l_{m^*}$. The upper bound of the last inequality is the reason why r_{J-1} should be considered separately, as the right end-point of the interval where detection will occur is upper bounded by the left end-point of the interval in which r_J was detected. The detection of r_{J-1} is at $b_{J-1} = \operatorname{argmax}_{c^l_{m^*_{J-1}} \leq t < c^r_{k^*_{J-1}}} C^t_{c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}}$, which, for $q_{J-1} := \operatorname{argmax}_{j=1,2,\ldots,d} \left| \tilde{X}^{b_{J-1},j}_{c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}} \right|$, satisfies

$$C_3 \log\left(Td^{1/4}\right) \geq \begin{cases} |b_{J-1} - r_{J-1}| \left(\Delta^{q_{J-1}}_{J-1}\right)^2, \text{ when } L(\cdot) = L_\infty(\cdot) \\ |b_{J-1} - r_{J-1}| \, L_2^2(\boldsymbol{\Delta_{J-1}}), \text{ when } L(\cdot) = L_2(\cdot) \end{cases}.$$

Following similar steps as in Step 3, we have that $C^{b_{J-1}}_{c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}} > \zeta_{T,d}$. Now, with

$$\alpha_{J-2} := \begin{cases} \Delta^{q_{J-1}}_{J-1}, \text{ when } L(\cdot) = L_\infty(\cdot) \\ L_2(\boldsymbol{\Delta_{J-1}}), \text{ when } L(\cdot) = L_2(\cdot) \end{cases},$$

it is straightforward to show, following the same process as Step 3 of the proof, that $|b_{J-2} - r_{J-2}| \, \alpha^2_{J-2} \leq C_3 \log\left(Td^{1/4}\right)$.

Step 5: Finally, when no change-points are left, the algorithm terminates. 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, for $b \in \{s^*, s^*+1, \ldots, e^*-1\}$,

$$C^b_{s^*,e^*} \leq D^b_{s^*,e^*} + \sqrt{8 \log\left(Td^{\frac{1}{4}}\right)} = \sqrt{8 \log\left(Td^{\frac{1}{4}}\right)} < C_1 \sqrt{\log\left(Td^{\frac{1}{4}}\right)} \leq \zeta_{T,d}$$

and the algorithm terminates.

References

[1] A. Anastasiou, I. Cribben, and P. Fryzlewicz. Cross-covariance isolate detect: A new change-point method for estimating dynamic functional connectivity. *Medical Image Analysis*, 75:102252, 2022.

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

[3] A. Anastasiou and A. Papanastasiou. Generalized multiple change-point detection in the structure of multivariate, possibly high-dimensional, data sequences. *Statistics and Computing*, 33(5):94, 2023.

[4] F. J. Anscombe. The transformation of poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4):246–254, 1948.

[5] J. Bai and P. Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78, 1998.

[6] R. Baranowski, Y. Chen, and P. 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.

[7] B. Brodsky and B. Darkhovsky. *Non-Parametric Statistical Diagnosis Problems and Methods*. Springer Netherlands, Dordrecht, 2000.

[8] H. Cao and W. Biao Wu. Changepoint estimation: another look at multiple testing problems. *Biometrika*, 102(4):974–980, 2015.

[9] E. Carlstein. Nonparametric Change-Point Estimation. *The Annals of Statistics*, 16(1):188 – 197, 1988.

[10] H.-P. Chan and H. Chen. Multi-sequence segmentation via score and higher-criticism tests, 2022.

[11] H. P. Chan and G. Walther. Detection with the scan and the average likelihood ratio. *Statistica Sinica*, 23(1):409–428, 2013. `http://www.jstor.org/stable/24310529`.

[12] H. Cho. Change-point detection in panel data via double CUSUM statistic. *Electronic Journal of Statistics*, 10(2):2000–2038, 2016.

[13] H. Cho and P. Fryzlewicz. Multiscale and multilevel technique for consistent segmentation of nonstationary time series. *Statistica Sinica*, 22(1):207–229, 2012. `http://www.jstor.org/stable/24310145`.

[14] H. Cho and P. 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.

[15] H. Cho and P. 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.

[16] H. Cho and C. Kirch. Two-stage data segmentation permitting multiscale change points, heavy tails and dependence. *Annals of the Institute of Statistical Mathematics*, 74(4):653–684, 2022.

[17] H. Cho and C. Kirch. Data segmentation algorithms: Univariate mean change and beyond. *Econometrics and Statistics*, 30:76–95, 2024.

[18] C.-S. J. Chu, K. Hornik, and C.-M. Kaun. Mosum tests for parameter constancy. *Biometrika*, 82(3):603–617, 1995.

[19] J. Ding, Y. Xiang, L. Shen, and V. Tarokh. Multiple change point analysis: Fast implementation and strong consistency. *IEEE Transactions on Signal Processing*, 65(17):4495–4510, 2017.

[20] B. Eichinger and C. Kirch. A MOSUM procedure for the estimation of multiple random change points. *Bernoulli*, 24(1):526 – 564, 2018.

[21] P. Fearnhead, R. Maidstone, and A. Letchford. Detecting changes in slope with an L_0 penalty. *Journal of Computational and Graphical Statistics*, 28(2):265–275, 2019.

[22] J. H. Friedman. Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1):1 – 67, 1991.

[23] P. Fryzlewicz. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243–2281, 2014. `https://www.jstor.org/stable/43556493`.

[24] P. Fryzlewicz. Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *The Annals of Statistics*, 46(6B):3390 – 3421, 2018.

[25] P. 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.

[26] F. R. Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393, 1974.

[27] D. M. Hawkins. Fitting multiple change-point models to data. *Computational Statistics & Data Analysis*, 37(3):323–341, 2001.

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

[29] H.-S. O. Joonpyo Kim and H. Cho. Moving sum procedure for change point detection under piecewise linearity. *Technometrics*, 66(3):358–367, 2024.

[30] 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.

[31] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky. ℓ_1 trend filtering. *SIAM Review*, 51(2):339–360, 2009.

[32] S. Kovács, P. Bühlmann, H. Li, and A. Munk. Seeded binary segmentation: a general methodology for fast and optimal changepoint detection. *Biometrika*, 110(1):249–256, 2023.

[33] T. Levajković and M. Messer. Multiscale change point detection via gradual bandwidth adjustment in moving sum processes. *Electronic Journal of Statistics*, 17(1):70 – 101, 2023.

[34] H. Li, A. Munk, and H. Sieling. FDR-control in multiscale change-point segmentation. *Electronic Journal of Statistics*, 10(1):918 – 959, 2016.

[35] H. Maeng and P. Fryzlewicz. Detecting linear trend changes in data sequences. *Statistical Papers*, pages 1–31, 2023.

[36] R. Maidstone, T. Hocking, G. Rigaill, and P. Fearnhead. On optimal multiple changepoint algorithms for large data. *Statistics and computing*, 27(2):519–533, 2017.

[37] A. Meier, C. Kirch, and H. Cho. mosum: A package for moving sums in change-point analysis. *Journal of Statistical Software*, 97(8):1–42, 2021.

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

[39] Y. Ninomiya. Information criterion for gaussian change-point model. *Statistics & Probability Letters*, 72(3):237–247, 2005.

[40] 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.

[41] M. Raimondo. Minimax estimation of sharp change points. *The Annals of Statistics*, 26(4):1379 – 1397, 1998.

[42] G. 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. `http://www.numdam.org/item/JSFS_2015__156_4_180_0/`.

[43] P. J. Rousseeuw and C. Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424):1273–1283, 1993.

[44] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978. `http://www.jstor.org/stable/2958889`.

[45] A. Sosa-Costa, I. K. Piechocka, L. Gardini, F. S. Pavone, M. Capitanio, M. F. Garcia-Parajo, and C. Manzo. Plant: A method for detecting changes of slope in noisy trajectories. *Biophysical Journal*, 114(9):2044–2051, 2018.

[46] S. Spiriti, R. Eubank, P. W. Smith, and D. Young. Knot selection for least-squares and penalized splines. *Journal of Statistical Computation and Simulation*, 83(6):1020–1036, 2013.

[47] C. Truong, L. Oudre, and N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

[48] L. Y. Vostrikova. Detecting "disorder" in multidimensional random processes. In *Doklady akademii nauk*, volume 259, pages 270–274. Russian Academy of Sciences, 1981.

[49] D. Wang, Y. Yu, and A. Rinaldo. Univariate mean change point detection: Penalization, CUSUM and optimality. *Electronic Journal of Statistics*, 14(1):1917 – 1961, 2020.

[50] P. Wiggins. An information-based approach to change-point analysis with applications to biophysics and cell biology. *Biophysical Journal*, 109(2):346–354, 2015.

[51] Y.-C. Yao. Estimating the number of change-points via schwarz' criterion. *Statistics & Probability Letters*, 6(3):181–189, 1988.

[52] Y.-C. Yao and S.-T. Au. Least-squares estimation of a step function. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 370–381, 1989. `https://www.jstor.org/stable/25050759`.

[53] Y. Yu. A review on minimax rates in change point detection and localisation, 2020.

[54] Y. Yu, S. Chatterjee, and H. Xu. Localising change points in piecewise polynomials of general degrees. *Electronic Journal of Statistics*, 16(1):1855–1890, 2022.