Data-adaptive algorithm for multiple change-point detection

Sophia Loizidou

A project submitted for the fulfillment of
the requirements of DSC573

Under the supervision of
Dr. Andreas Anastasiou

Department of Mathematics and Statistics
University of Cyprus

Fall Semester 2022

1 Introduction

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

Change-point detection has two main directions; sequential (or online) and a posteriori (or offline). We focus on the latter, where the goal is to estimate the number and locations of changes in data, for which the total number of observations is known in advance. 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 are concerned with piecewise-constant signals f_t, meaning that the changes are in the mean, but the algorithm can also be extended to other signals. Although detecting changes in the mean is a simple case, as noted by Brodsky and Darkhovsky [2000] more complex change-point problems that allow changes in properties other than the mean, can be reduced to the segmentation problem we are studying. This is done by applying a suitable transformation to the data that reveals the changes as those in the mean of the transformed data. One example is finding the points where the slope of the underlying signal, f_t, changes when f_t is a piecewise-linear signal. The appropriate transformation in this case is to take pairwise differences in the data so that detecting changes in the mean of the sequence $Y_t = X_{t+1} - X_t$ for $t = 2, 3, \ldots, T$ is equivalent to detecting changes in the slope of X_t. A more complicated example is described by Cho and Fryzlewicz [2015] who detect multiple change-points in the second-order (i.e. autocovariance and cross-covariance) structure of possibly high dimensional time series. This is done using Haar wavelets as building blocks. The signal is transformed into a piecewise-constant one with identical change-point locations.

The various methods available for solving the problem of change-point detection can be split into two main categories, according to whether the change-points are detected all at once or one at a time. The former includes global optimization, in the sense that it uses optimization based models that estimate the signal using least squares or log-likelihood, while trying to avoid overfitting using a penalization function. The latter involves scanning the sequence in different ways in an attempt to detect the change-points, one at a time, using local testing methods. Starting with the methods of detecting all change-points at once, the most common example of a penalty function is the Schwarz Information Criterion (Schwarz [1978]), which was firstly used by Yao [1988] in the context of change-point detection for i.i.d. Gaussian ϵ_t, with ϵ_t as in (1). Under the same assumption, both non-penalized and penalized least squares approaches are considered in Yao and Au [1989], while Lavielle [2005] proposes an adaptive choice of the penalty function for automatically estimating the number of change-points. Lee [1997] and Hawkins [2001] relax the Gaussianity assumption of the noise to exponential family distributions. More general forms of Schwarz-like penalties are studied, for example, in Wu [2008] and Ciuperca [2011]. Wu [2008] proposes an information-based criterion for carrying out change-point analysis and variable selection simultaneously in linear models with a possible change-point, while Ciuperca [2011] introduces a general criterion to determine the number K of the change-points in a parametric nonlinear multi-response model. Moving on to dynamic programming approaches, Auger and Lawrence [1989] introduce the Segment Neighbourhood (SN) approach which requires an upper bound on the number of change-points. Rigaill [2015] proposes a new representation of the segmentations, called pruned dynamic programming algorithm (pDPA), where the cost of a segmentation is represented as a function of a possibly multidimensional parameter. The dynamic algorithm Optimal Partitioning (OP) (Jackson et al. [2005]) is guaranteed to find the exact global optimum, while Killick et al. [2012] improve OP's computational cost by adding a pruning step within the dynamic program, with the method called Pruned Exact Linear Time (PELT). Haynes et al. [2017] introduce NP.PELT, a non-parametric adaptation of PELT, for which no distributional assumptions are needed. Maid-

stone et al. [2017] combine the ideas from PELT and pDPA leading to two new algorithms, Functional Pruning Optimal Partitioning (FPOP) and Segment Neighbourhood with Inequality Pruning (SNIP). FPOP, which uses functional pruning to solve the penalised minimisation problem, has low computational complexity and always prunes more than PELT. SNIP uses inequality based pruning to solve the constrained minimisation problem and thus has increased speed. Li et al. [2016] propose a multiscale segmentation method, FDRSeg, which controls the false discovery rate (FDR) in the sense that the number of false jumps is bounded linearly by the number of true jumps. Ciuperca [2014] uses the LASSO technique which allows the parametric estimation, including the detection of change-points, and automatic variable selection simultaneously of a linear regression model. Finally, Ninomiya [2005] introduces an AIC-type information criterion for change-point models for which the bias cannot be estimated by the number of parameters and Arlot and Celisse [2011] attempt time series segmentation using Cross Validation as a penalization function, without assuming constant variance.

In the category where change-points are detected one at a time, a method that has attracted much attention is Binary Segmentation (BS), introduced by Vostrikova [1981]. It is a general approach that can be viewed as a greedy approach with a low computational complexity and performs an iterative binary splitting of the data on intervals determined by the previously obtained splits using a CUSUM-like procedure. Venkatraman [1992] established the consistency of BS while allowing the number of change-points to diverge with the length of the data sequence, T. The fact that each stage of BS involves searching for a single change-point means that it may be unsuitable for signals that contain multiple change-points in certain configurations, for example signals with frequent change-points; thus it has suboptimal accuracy. Circular Binary Segmentation (CBS) and Wild Binary Segmentation (WBS) are adaptations of BS, by Olshen et al. [2004] and Fryzlewicz [2014], respectively, in an attempt to solve this problem. CBS considers each segment to be spliced at the two ends to form a circle and searches for at most two change-points at each step, while WBS attempts to eliminate the weaknesses of BS by randomly drawing a number of smaller intervals from $t = 1$ up to $t = T$ and comparing the maximum CUSUM over all intervals against a threshold. Further improvement is given by Fryzlewicz [2020] who introduces WBS2, which iteratively draws random intervals and produces an ordered set according to their importance measured by the associated CUSUM statistics. While WBS draws many intervals at the start of the algorithm, WBS2 draws a smaller number of intervals every time a detection occurs. Wang et al. [2020] consider WBS with the lengths of the chosen intervals bounded from above. This ensures that each background interval contains at most a single change-point when the signal-to-noise ratio diverges with T at the rate of at least $\sqrt{\log T}$, leading to an improvement of the localisation rate. A different adaptation is to select the shortest interval over which the associated CUSUM statistic exceeds a given threshold at each recursive iteration, and this is what Baranowski et al. [2019] use in the algorithm Narrowest-Over-Threshold (NOT). Moving sum procedures (MOSUM), where the data are scanned for change-points using local intervals of the form $(t - G, t + G)$ for every time point t and a chosen bandwidth G are considered in Chu et al. [1995] and Eichinger and Kirch [2018]. Considering bottom-up and sequential approaches, Fryzlewicz [2018] achieves a multiscale decomposition of the data with respect to an adaptively chosen unbalanced Haar wavelet basis. Similar ideas are introduced by Chan and Chen [2017] and Jun Shin et al. [2020] that explore reverse segmentation and a backward procedure (BWD), respectively. The reverse segmentation involves creating a 'solution path' by deleting the change-point with the smallest CUSUM value in the segment determined by its closest left and right neighbours, in order to obtain a hierarchy of nested models. BWD starts by considering that every point is a change-point, for $t = 1, 2, \ldots, T - 1$ and works by repeatedly merging two neighbouring groups into one until a desired stopping rule is satisfied. Finally, Anastasiou and Fryzlewicz [2022] propose a method called Isolate-Detect (ID) that uses expanding intervals in a sequential way, starting from both the beginning and the end of the data sequence. This method achieves isolation of the change-points and is thus consistent and accurate at detecting their number and locations.

In this project, we propose a data-adaptive change-point detection method, called DAIS (Data Adaptive ISolation), that tries to imitate the isolation aspect of ID, but taking into account the potential true locations of the change-points. The idea behind the data-adaptive behaviour of the algorithm arises from the fact that, for the unobserved true signal f_t in (1), taking pairwise differences between

consecutive time steps will reveal a constant signal at 0, with 'spikes' where the change-points occur. This means that the largest 'spike', in absolute value, is the location of the change-point with the largest jump in the sequence. We try to use this fact in the observed signal, X_t. In the algorithm, we identify this 'spike', which we will be referring to as 'largest difference' for the rest of the project, and test around it for possible change-points. We check in expanding intervals in such a way that we achieve isolation of the change-points. In such cases, the detection power of the contrast function is maximised. More details on the choice of the contrast function can be found in Section 3.

The DAIS algorithm is split into two parts. The first one is to use left- and right-expanding intervals around the location of the largest difference found, $d_{s,e}$ in the interval $[s, e]$. This is done in a sequential way, expanding once either only to the left or only to the right at each step. For a sequence of length T, using an expansion parameter λ_T (more details in Section 4.1), at most $\lceil T/\lambda_t \rceil + 1$ steps are necessary for this first step. During this first part, we are checking expanding intervals around what we have a reason to believe to be the change-point with the largest jump in the sequence. However, since we are observing the noisy sequence X_t, the largest difference may not be in an area close to the change-point, maybe at a position that two-sided expansions around $d_{s,e}$ do not allow for the detection of any change-points. This can happen if $d_{s,e}$ is between two change-points and neither is isolated in an interval where detection is possible, for example when $d_{s,e}$ is the midpoint between the change-points. That is why, if no change is detected in the first part of the algorithm, we move on to the second part, which involves left-expanding intervals with fixed end-point $d_{s,e}$ and right-expanding intervals with fixed start-point $d_{s,e}$. Using these expansions, detection of each change-point in an interval where it is isolated is guaranteed with high probability. 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 interval. A more detailed explanation of the algorithm can be found in Section 2.

The project is organized as follows. Section 2 provides details on the methodology of DAIS. We present a simple example and a detailed explanation of the algorithm follows. In Section 3, we provide theoretical results regarding the consistency of the number of change-points detected and the accuracy of their estimated locations. Section 4 includes an explanation of how some parameters of DAIS are selected and some modifications to achieve better performance. In Section 5, we provide results on simulated data and a comparison to state-of-the-art competitors. Section 6 is an application of DAIS on real data and Section 7 concludes the report with a summary of the important findings.

2 Methodology

We begin by presenting a toy example of how DAIS works in practice. We have a sequence of length 100, so $T = 100$, and a change-point at location $r = 65$. For the sake of presentation of this example, we take the expansion parameter to be $\lambda_T = 10$. We first calculate the absolute difference of consecutive observations, X_t, for $t = 1, 2, \ldots, T$. The largest difference in the unobserved signal f_t is at $t = 65$ and so we expect that the largest difference in the observed X_t will be in an area around the true change-point, say at location $d_{1,100} = 63$. This means that $\text{argmax}_{t=1,2,\ldots,99}|X_{t+1} - X_t| = 63$. DAIS initially creates right- and left-expanding intervals around this point and so, as a first step, checks the interval $[63, 72]$, as shown in Figure 1. The change-point is identified and then the algorithm restarts using the intervals $[1, 63]$ and $[72, 100]$. As these two intervals contain no change-points, the largest differences are at points where the absolute value of the difference between the noise of two consecutive observations is large. Suppose that the largest differences are now detected at points X_2 and X_{86} in the two intervals, respectively. For the interval $[72, 100]$ the left- and right-expanding subintervals checked are: $[86, 95]$, $[76, 95]$, $[76, 100]$, $[72, 100]$, which is shown on the left plot of Figure 2. Since no change-point is detected, the algorithm now uses only one sided expansions and so checks intervals $[76, 86]$, $[72, 86]$, $[86, 95]$, $[86, 100]$, as shown on the right plot of Figure 2. In a similar way, the expansions for the interval $[1, 63]$ are performed. No change-point is detected and so the algorithm terminates.

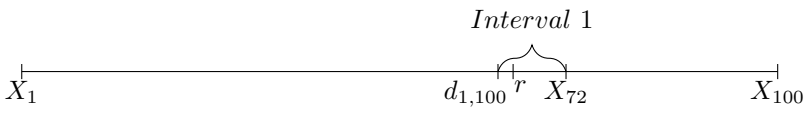

Figure 1: Example with one change-point at $r = 65$ and largest difference at $d_{1,100} = 63$. The change-point is detected when the first interval is checked.

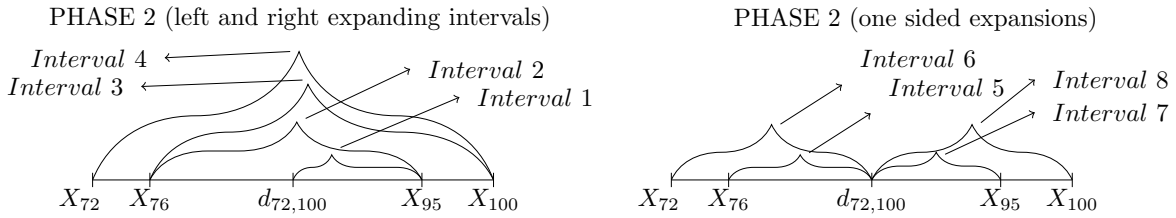

Figure 2: Example with one change-point at $r = 65$ after detection occurs in the interval $[63, 72]$. The next intervals that are checked are $[1, 63]$ and $[72, 100]$. The left plot shows the left- and right-expanding intervals around the location of the largest difference, $d_{72,100} = 86$, when checking $[72, 100]$. The right plot shows the only left- and only right-expanding intervals around $d_{72,100} = 86$. No change-point is detected.

For the rest of the project, we use s, e to denote the start and end point of the interval respectively. $T \in \mathbb{N}$ denotes the total length of the sequence we are given. We have $N \in \mathbb{N}_0$ and r_i for $i = 1, 2, \ldots, N$ are the number and location of the true change-points, while $\hat{N}$, $\hat{r}_i$ for $i = 1, 2, \ldots, \hat{N}$ are the estimated change-point number and locations. We denote by $\lambda_T \in \mathbb{N}$ the expansion parameter and $\zeta_T \in \mathbb{R}^+$ the threshold. In addition, $d_{s,e} = \text{argmax}_{t \in \{s, s+1, \ldots, e-1\}}\{|X_{t+1} - X_t|\}$ is the location of the largest difference in the interval $[s, e]$ and $C^b_{s,e}(\boldsymbol{X})$ is the contrast function applied on point X_b using the values $\{X_s, X_s + 1, \ldots, X_e\}$ for $b \in [s, e)$. 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 $K = K^l + K^r$, the start- and end-points of the intervals that will be checked by the algorithm are given by

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

These are collected in the ordered sets defined in (2), which are used to create the intervals that are checked by DAIS. $L^b_{d_{s,e},s,e}$ and $R^b_{d_{s,e},s,e}$ consist of the left and right end-points of the intervals respectively, for the two sided expansions. Since the expansions occur 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^b_j in (3). The sets $L^o_{d_{s,e},s,e}$ and $R^o_{d_{s,e},s,e}$ are the end-points of the left only and right only expanding intervals, respectively, which are defined in (4) and (5). The ordered sets of end-points for all the expansions are:

$$\begin{aligned} L^b_{d_{s,e},s,e} &= \{c^l_0, c^l_1, c^l_1, c^l_2, c^l_2, \ldots, c^l_{K^{\min}-1}, c^l_{K^{\min}-1}, c^l_{K^{\min}}, c^l_{K^{\min}+1}, \ldots, c^l_{K^{\max}}\}, \\ R^b_{d_{s,e},s,e} &= \{c^r_1, c^r_1, c^r_2, c^r_2, \ldots, c^r_{K^{\min}}, c^r_{K^{\min}}, c^r_{K^{\min}+1}, c^r_{K^{\min}+2}, \ldots, c^r_{K^{\max}}\}, \\ L^o_{d_{s,e},s,e} &= \{c^l_1, c^l_2, \ldots, c^l_{K^l}\}, \quad R^o_{d_{s,e},s,e} = \{c^r_1, c^r_2, \ldots, c^r_{K^r}\}. \end{aligned} \tag{2}$$

It is easy to show that $|L^b_{d_{s,e},s,e}| = |R^b_{d_{s,e},s,e}| = K$, where $|A|$ denotes the cardinality of the set A. The

ordered sets of the intervals checked are:

$$I_j^b = \left[L^b_{d_{s,e},s,e}[j], R^b_{d_{s,e},s,e}[j]\right], \text{ for } j \in \{1, 2, \ldots, K\} \tag{3}$$

$$I_j^l = \left[L^o_{d_{s,e},s,e}[j], d_{s,e}\right] \text{ for } j \in \{1, 2, \ldots, K^l\} \tag{4}$$

$$I_j^r = \left[d_{s,e}, R^o_{d_{s,e},s,e}[j]\right] \text{ for } j \in \{1, 2, \ldots, K^r\} \tag{5}$$

where A[i] denotes the i^{th} element of the ordered set A and finally, all intervals to be checked are collected in the ordered set I_j, as follows:

$$I_j = [s_j, e_j] = \begin{cases} I_j^b, & j = 1, \ldots, K, \\ I_j^l, & j = K+1, K+2, \ldots, K+K^l, \\ I_j^r, & j = K+K^l+1, K+K^l+2, \ldots, 2K \end{cases} . \tag{6}$$

Algorithm 1 DAIS

function DAIS(s, e, λ_T, ζ_T)
if $e - s < 1$ **then**
  STOP
else
  $d_{s,e} = \text{argmax}_{t \in \{s, s+1, \ldots, e-1\}}\{|X_{t+1} - X_t|\}$
  For $j \in \{1, 2, \ldots, 2K\}$ let $I_j = [s_j, e_j]$ as in (6)
  $i = 1$
  (Main part)
  $b_i = \text{argmax}_{t \in [s_i, e_i)} C^t_{s_i, e_i}(\mathbf{X})$
  if $C^{b_i}_{s_i, e_i}(\mathbf{X}) > \zeta_T$ **then**
    add b_i to the list of estimated change-points
    DAIS($s, s_i, \lambda_T, \zeta_T$)
    DAIS($e_i, e, \lambda_T, \zeta_T$)
  else
    $i = i + 1$
    if $i \leq 2K$ **then**
      Go back to **(Main part)** and repeat
    else
      STOP
    end if
  end if
end if
end function

The first step of DAIS, as described in Algorithm 1, is to identify the location of the largest first difference $d_{1,T}$ within the observed data, X_t, $t = 1, 2, \ldots, T$. This is done by taking the value of t that maximises the pairwise absolute differences between consecutive observations. Applying this on the unobserved true signal f_t in (1), will reveal a constant signal at 0, with 'spikes' where the change-points occur, so we expect that, with high probability, the largest difference of X_t will be at a point near the change-point with the jump of the largest magnitude. Since DAIS starts checking for potential change-points in that area, the change-point will be detected fast. This step is what makes the algorithm data adaptive. After calculating $d_{1,T}$, expanding intervals are used to check for possible change-points. There are two different phases to the expansions and both are needed to ensure that every change-point can be detected in an interval where it is isolated. The first one is to expand the intervals once from the right and once from the left at every step by an expansion parameter of magnitude λ_T. This is done in order to identify potential change-points near the location of the detected largest difference. This expansion is performed for at most $K = \lceil \frac{T}{\lambda_T} \rceil + 1$ steps to both directions. If no change-point is detected, one-sided expansions are used, first to the left and then

to the right of $d_{1,T}$. These are required to make sure that each change-point can be detected in an interval where it is isolated, and thus achieve the desirable rate of convergence in Theorem 1. For these expansions, the intervals have a fixed right- or left-endpoint respectively and so they have the form $[s_j, d_{1,T}]$ and $[d_{1,T}, e_j]$. These intervals use the same expansion parameter. Upon the detection of a change-point, the algorithm restarts after discarding the previously used data. If the change was detected in the subinterval $[s^*, e^*]$ then DAIS will be re-applied on $[s, s^*]$ and $[e^*, e]$. When checking any interval $[s, e]$, a contrast function is applied to the data, checking for possible change-points at location b, where $b \in \{s, s+1, \ldots, e-1\}$. The contrast function is denoted as $C^b_{s,e}(\boldsymbol{X})$ and the maximum value attained in the interval is compared with a threshold ζ_T. If $C^b_{s,e}(\boldsymbol{X})$ exceeds the threshold, we take b to be a change-point. The contrast function used in the case of piecewise-constant signal f_t in (1), is explained in Section 3 and the choice of the parameters ζ_T and λ_T is explained in Section 4.1.

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 Hampel [1974]. In the case of Gaussian data, for $C = 1.4826$ it is a consistent estimator of the population standard deviation σ as proven by Rousseeuw and Croux [1993]. In this section, we assume that ϵ_t in (1) are i.i.d. random variables following the $\mathcal{N}(0,1)$ distribution. This assumption is used in order to make notation easier in the proof of the theorem presented in this section and is not a requirement for DAIS. The model becomes

$$X_t = f_t + \epsilon_t, \quad t = 1, 2, \ldots, T. \tag{7}$$

We are interested in the case of abrupt changes in the mean of X_t, meaning that the unobserved signal f_t satisfies $f_t = \mu_j$ for $t = r_{j-1}+1, \ldots, r_j$ and $f_{r_j} \neq f_{r_j+1}$ for change-points r_j, $j \in \{1, 2, \ldots, N\}$ and $r_0 = 1$, $r_{N+1} = T$. The statistic used is the absolute value of the CUSUM statistic, which is defined as

$$C^b_{s,e}(\mathbf{X}) = \left|\tilde{X}^b_{s,e}\right| = \left| \sqrt{\frac{e-b}{n(b-s+1)}} \sum_{t=s}^{b} X_t - \sqrt{\frac{b-s+1}{n(e-b)}} \sum_{t=b+1}^{e} X_t \right|, \tag{8}$$

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

We are now ready to give the theoretical result for the consistency of the number and location of the estimated change-points. First, we make the following assumption.

(A1) The minimum distance, δ_T, between two successive change-points and the minimum magnitude of jumps $\underline{f}_T$ are connected by $\sqrt{\delta_T}\underline{f}_T \geq \underline{C}\sqrt{logT}$ for a large enough constant $\underline{C}$.

Below, we provide the relevant theorem for the consistency of DAIS in accurately estimating the true number and the locations of the change-points.

Theorem 1. *Let $\{X_t\}_{t=1,2,\ldots,T}$ follow model (7), 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 (A1) holds. 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, $\Delta^f_j = \left|f_{r_j+1} - f_{r_j}\right|, j = 1, 2, \ldots N$ is the magnitude of each discontinuity in f_t, $\underline{f}_T = \min_j \Delta^f_j$ 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 $C_1\sqrt{\log T} \leq \zeta_T < C_2\sqrt{\delta_T}\underline{f}_T$ and for sufficiently large T, we obtain*

$$\mathbb{P}\left(\hat{N} = N, \max_{j=1,2,\ldots,N}\left(|\hat{r}_j - r_j|\left(\Delta^f_j\right)^2\right) \leq C_3 \log T\right) \geq 1 - \frac{C_4}{T}. \tag{9}$$

From Theorem 1, we can conclude that in order to be able to match the estimated change-points with the true ones, δ_T needs to be larger than $\max_{j=1,2,\ldots,N} |\hat{r}_j - r_j|$, which means that δ_T must be at least of order $\mathcal{O}(\log T)$. Also, assumption (A1) ensures that the rate attained for $\delta_T \underline{f}_T^2$ is $\mathcal{O}(\log T)$ and, as Chan and Walther [2013] argues, the smallest possible $\delta_T \underline{f}_T^2$ that allows detection of the change-points is $\mathcal{O}(\log T - \log\log T)$ so the rate obtained by DAIS is nearly optimal.

Before proving Theorem 1, we introduce some more notation. Similar to (8), we denote the CUSUM of the unobserved signal as $\tilde{f}_{s,e}^b = \sqrt{\frac{e-b}{n(b-s+1)}} \sum_{t=s}^{b} f_t - \sqrt{\frac{b-s+1}{n(e-b)}} \sum_{t=b+1}^{e} f_t$. From now on, the contrast vector $\boldsymbol{\psi}_{\boldsymbol{s,e}}^{\boldsymbol{b}} = \left(\psi_{s,e}^b(1), \psi_{s,e}^b(2), \ldots, \psi_{s,e}^b(T_1)\right)$ is defined through the contrast function

$$\psi_{s,e}^b(t) = \begin{cases} \sqrt{\frac{e-b}{n(b-s+1)}}, & t = s, s+1, \ldots, b, \\ -\sqrt{\frac{b-s+1}{n(e-b)}}, & t = b+1, b+2, \ldots, e, \\ 0, & \text{otherwise} \end{cases}, \tag{10}$$

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

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,*

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

In addition,

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

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

The proof of Theorem 1 consists of 6 steps. Step 1 is to show that the observed $|\tilde{X}_{s,e}^b|$ is uniformly close to the unobserved $|\tilde{f}_{s,e}^b|$, 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, which we are interested in. In Step 2, we control the distance between $|\tilde{X}_{s,e}^{b_1}| - |\tilde{X}_{s,e}^{b_2}|$ and $|\tilde{f}_{s,e}^{b_1}| - |\tilde{f}_{s,e}^{b_2}|$ 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. 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

$$\mathbb{P}\left(\hat{N} = N, \max_{j=1,2,\ldots,N} \left(|\hat{r}_j - r_j| \left(\Delta_j^f\right)^2\right) \leq C_3 \log T\right) \geq 1 - \frac{1}{6\sqrt{\pi}T}, \tag{11}$$

which implies result (9).

Step 1: Allow us to denote by

$$A_T = \left\{ \max_{s,b,e:1\le s\le b<e\le T} \left| \tilde{X}^b_{s,e} - \tilde{f}^b_{s,e} \right| \le \sqrt{8\log T} \right\}. \tag{12}$$

We will show that $\mathbb{P}(A_T) \ge 1 - 1/(12\sqrt{\pi}T)$. From (7) and (8), 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

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

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 (10), we denote by

$$B_T = \left\{ \max_{1\le s\le b<e<T} \frac{\left| \left\langle \boldsymbol{\psi^b_{s,e}}\langle \boldsymbol{f}, \boldsymbol{\psi^b_{s,e}}\rangle - \boldsymbol{\psi^r_{s,e}}\langle \boldsymbol{f}, \boldsymbol{\psi^r_{s,e}}\rangle, \boldsymbol{\epsilon} \right\rangle \right|}{\|\boldsymbol{\psi^b_{s,e}}\langle \boldsymbol{f}, \boldsymbol{\psi^b_{s,e}}\rangle - \boldsymbol{\psi^r_{s,e}}\langle \boldsymbol{f}, \boldsymbol{\psi^r_{s,e}}\rangle\|_2} \le \sqrt{8\log T} \right\}. \tag{13}$$

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 $\mathbb{P}\left((B_T)^c\right) \le \frac{1}{12\sqrt{\pi}T}$. Therefore, Steps 1 and 2 lead to

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

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

$$C_1 = \sqrt{C_3} + \sqrt{8}, C_2 = \frac{1}{\sqrt{12}} - \frac{2\sqrt{2}}{\underline{C}}, C_3 = 2(2\sqrt{2}+4)^2, \tag{14}$$

where $\underline{C}$ satisfies $\sqrt{\delta_T}\underline{f}_T \ge \underline{C}\sqrt{\log T}$. We take $\lambda_T \le \delta_T/6$, but the proof is similar for more general $\lambda_T \le \delta_T/m$, for $m > 1$. For $j \in \{1,2,\ldots,N\}$, define

$$I^L_j = \left(r_j - \frac{\delta_T}{3}, r_j - \frac{\delta_T}{6} \right], \quad I^R_j = \left[r_j + \frac{\delta_T}{6}, r_j + \frac{\delta_T}{3} \right). \tag{15}$$

For an interval $[s,e]$, $1 \le s < e \le T$, for $d_{s,e} = \text{argmax}_{t\in\{s,s+1,\ldots,e-1\}}\{|X_{t+1} - X_t|\}$ being the location of the largest difference detected in the interval and $K^l = \lceil \frac{d_{s,e}-s+1}{\lambda_T} \rceil$, $K^r = \lceil \frac{e-d_{s,e}+1}{\lambda_T} \rceil$, $K^{\max} = \max\{K^l, K^r\}$, define

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

Since the length of the intervals in (15) is $\delta_T/6$ and $\lambda_T \le \delta_T/6$, 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 $d_{1,T}$. If $d_{1,T} = r_j$ for any $j \in \{1,2,\ldots,N\}$, then r_j will be the first change-point to get

detected. If $d_{1,T} < r_1$ or $d_{1,T} > r_N$, the first change-point to get detected will be r_1 or r_N, respectively. Finally, if $r_j < d_{1,T} < r_{j+1}$, for $j \in \{1, 2, \ldots, N-1\}$ then the change-point that is closest to $d_{1,T}$ will be detected first during the initial phase of the algorithm, where we expand from both directions interchangeably, if detection is possible. Otherwise, r_j will be detected first within a left-expanding interval. In either case, isolation is guaranteed because if detection is not possible during the first phase of DAIS, in any left- and right-expanding intervals in which r_j is isolated, it will certainly be detected in the second phase in a one-sided expanding interval.

Without loss of generality, we suppose that the first change-point to get detected is r_J for $J \in \{1, 2, \ldots, N\}$. There exists an interval $[c_m^l, c_k^r]$, for $m, k \in \{0, 1, \ldots, K^{\max}\}$, such that r_J is the only change-point included in the interval. We will now show that for $\tilde{b} = \text{argmax}_{c_m^l \le t < c_k^r} |\tilde{X}^t_{c_m^l, c_k^r}|$, then $\left|\tilde{X}^{\tilde{b}}_{c_m^l, c_k^r}\right| > \zeta_T$. Using (7), we have that

$$\left|\tilde{X}^{\tilde{b}}_{c_m^l, c_k^r}\right| \ge \left|\tilde{X}^{r_J}_{c_m^l, c_k^r}\right| \ge \left|\tilde{f}^{r_J}_{c_m^l, c_k^r}\right| - \sqrt{8 \log T}. \tag{16}$$

But,

$$\begin{aligned}
\left|\tilde{f}^{r_J}_{c_m^l, c_k^r}\right| &= \left| \sqrt{\frac{c_k^r - r_J}{(c_k^r - c_m^l + 1)(r_J - c_m^l + 1)}} (r_J - c_m^l + 1) f_{r_J} - \sqrt{\frac{r_J - c_m^l + 1}{(c_k^r - c_m^l + 1)(c_k^r - r_J)}} (c_k^r - r_J) f_{r_J + 1} \right| \\
&= \left| \sqrt{\frac{(c_k^r - r_J)(r_J - c_m^l + 1)}{(c_k^r - c_m^l + 1)}} f_{r_J} - \sqrt{\frac{(r_J - c_m^l + 1)(c_k^r - r_J)}{(c_k^r - c_m^l + 1)}} f_{r_J + 1} \right| \\
&= \sqrt{\frac{(c_k^r - r_J)(r_J - c_m^l + 1)}{c_k^r - c_m^l + 1}} \Delta_J^f \ge \sqrt{\frac{(c_k^r - r_J)(r_J - c_m^l + 1)}{2 \max\{c_k^r - r_J, r_J - c_m^l + 1\}}} \Delta_J^f \\
&= \sqrt{\frac{\min\{c_k^r - r_J, r_J - c_m^l + 1\}}{2}} \Delta_J^f.
\end{aligned} \tag{17}$$

If we show that

$$\min\{c_k^r - r_J, r_J - c_m^l + 1\} \ge \frac{\delta_T}{6} \tag{18}$$

then using assumption (A1) and the results in (16) (17), (18), we have that

$$\begin{aligned}
\left|\tilde{X}^{\tilde{b}}_{c_m^l, c_k^r}\right| &\ge \sqrt{\frac{\delta_T}{12}} \Delta_J^f - \sqrt{8 \log T} \ge \sqrt{\frac{\delta_T}{12}} \underline{f}_T - \sqrt{8 \log T} \\
&= \left(\frac{1}{\sqrt{12}} - \frac{2\sqrt{2 \log T}}{\sqrt{\delta_T} \underline{f}_T} \right) \sqrt{\delta_T} \underline{f}_T \ge \left(\frac{1}{\sqrt{12}} - \frac{2\sqrt{2}}{\underline{C}} \right) \sqrt{\delta_T} \underline{f}_T \\
&= C_2 \sqrt{\delta_T} \underline{f}_T > \zeta_T
\end{aligned}$$

and thus, the change-point will get detected.

Now, we consider the different cases for the location of the change-point and the possible intervals, to show that (18) holds. The two possible cases are $r_J \in \left[d_{1,T} - \delta_T/6, d_{1,T} + \delta_T/6\right]$ and $r_J \in [1, d_{1,T} - \delta_T/6) \cup (d_{1,T} + \delta_T/6, T)$. For both cases, we explain why isolation is guaranteed and prove that (18) holds, which implies that detection is achieved.

Case 1: If $r_J \in \left[d_{1,T} - \frac{\delta_T}{6}, d_{1,T} + \frac{\delta_T}{6}\right]$, then the change-point will necessarily get detected in a left- and right-expanding interval. First, we show that the change-point will be isolated. We will have an interval $[c_m^l, c_k^r]$ with $c_m^l \ge d_{1,T} - \delta_T/3$, $c_k^r \le d_{1,T} + \delta_T/3$ and since the length of the interval is at most $2\delta_T/3$, the change-point is indeed isolated. To show that (18) holds, we consider two cases

- if $r_J \in \left[d_{1,T} - \delta_T/6, d_{1,T}\right]$ then the interval is of the form $[c_k^l, c_k^r]$ with $c_k^l \in I_J^L$ for some $k \in \{1, 2, \ldots, K^{\max}\}$ and so $r_J - c_k^l + 1 > \delta_T/6$ and $c_k^r - r_J > c_k^r - d_{1,T} - 1 = d_{1,T} - c_k^l \ge r_J - c_k^l \ge \delta_T/6$.

- if $r_J \in \big(d_{1,T}, d_{1,T}+\delta_T/6\big]$ then again the interval is of the form $[c_k^l, c_k^r]$ but this time $c_k^r \in I_J^R$ for some $k \in \{1,2,\ldots,K^{\max}\}$ and so $c_k^r - r_J \geq \delta_T/6$ and $r_J - c_k^l + 1 > d_{1,T} - c_k^l + 1 = c_k^r - d_{1,T} > c_k^r - r_J \geq \delta_T/6$.

Case 2: If $r_J \in [1, d_{1,T} - \frac{\delta_T}{6}) \cup (d_{1,T} + \frac{\delta_T}{6}, T)$, then detection may be possible in a left- and right-expanding interval where r_J is isolated, but this is not guaranteed. If the change-point has not been detected yet, it will get detected in a one-sided expanding interval. So, we have the following cases:

- If $r_J \in [1, d_{1,T} - \frac{\delta_T}{6}) \cup (d_{1,T} + \frac{\delta_T}{6}, T)$ and detection occurs in a left- and right-expanding interval, then we again have two different cases depending on $r_J \leq d_{1,T}$ or $r_J > d_{1,T}$ and the steps are similar to Case 1 and so will not be repeated.

- If $r_J \in [1, d_{1,T} - \frac{\delta_T}{6})$ and $[c_m^l, c_k^r]$ is a left-expanding interval, then the interval is of the form $[c_m^l, c_0^r] = [c_m^l, d_{1,T}]$ for some $m \in \{1,2,\ldots,K^l\}$, with $c_m^l \in I_J^L$ and so $r_J - c_m^l + 1 > \delta_T/6$, $c_k^r - r_J > \delta_T/6$.

- If $r_J \in (d_{1,T} + \frac{\delta_T}{6}, T)$ and $[c_m^l, c_k^r]$ is a right-expanding interval, then the interval is of the form $[c_0^l, c_k^r] = [d_{1,T}, c_k^l]$ for some $k \in \{1,2,\ldots,K^r\}$, with $c_k^r \in I_J^R$ and so $r_J - c_m^l + 1 > \delta_T/6$, $c_k^r - r_J > \delta_T/6$.

Therefore, there will be an interval of the form $[c_{\tilde{m}}^l, c_{\tilde{k}}^r]$, such that the interval contains r_J and no other change-point and $\max_{c_{\tilde{m}}^l \leq t < c_{\tilde{k}}^r} \left|\tilde{X}^b_{c_{\tilde{m}}^l, c_{\tilde{k}}^r}\right| > \zeta_T$. For $k^*, m^* \in \{0,1,\ldots K^{\max}\}$, denote by $c_{m^*}^l \geq c_{\tilde{m}}^l$ and $c_{k^*}^r \leq c_{\tilde{k}}^r$ the first left- and right-expanding points, respectively, that this happens and let $b_J = \operatorname{argmax}_{c_{m^*}^l \leq t < c_{k^*}^r} \left|\tilde{X}^t_{c_{m^*}^l, c_{k^*}^r}\right|$, with $\left|\tilde{X}^{b_J}_{c_{m^*}^l, c_{k^*}^r}\right| > \zeta_T$. Our aim now is to find $\gamma_T > 0$, such that for any $b^* \in \{c_{m^*}^l, c_{m^*}^l + 1, \ldots, c_{k^*}^r - 1\}$ with $|b^* - r_J|\Big(\Delta_J^f\Big)^2 > \gamma_T$, we have that

$$\Big(\tilde{X}^{r_J}_{c_{m^*}^l, c_{k^*}^r}\Big)^2 > \Big(\tilde{X}^{b^*}_{c_{m^*}^l, c_{k^*}^r}\Big)^2. \tag{19}$$

Proving (19) and using the definition of b_J, we can conclude that $|b_J - r_J|\Big(\Delta_J^f\Big)^2 \leq \gamma_T$. Now, using (7), it can be shown, for $\boldsymbol{\psi}^{\boldsymbol{b}}_{\boldsymbol{s,e}}$ as defined in (10), that (19) is equivalent to

$$\begin{aligned}\Big(\tilde{f}^{r_J}_{c_{m^*}^l, c_{k^*}^r}\Big)^2 - \Big(\tilde{f}^{b^*}_{c_{m^*}^l, c_{k^*}^r}\Big)^2 > \Big(\tilde{\epsilon}^{b^*}_{c_{m^*}^l, c_{k^*}^r}\Big)^2 - \Big(\tilde{\epsilon}^{r_J}_{c_{m^*}^l, c_{k^*}^r}\Big)^2 \\ + 2\Big\langle \boldsymbol{\psi}^{b^*}_{c_{m^*}^l, c_{k^*}^r} \langle \boldsymbol{f}, \boldsymbol{\psi}^{b^*}_{c_{m^*}^l, c_{k^*}^r}\rangle - \boldsymbol{\psi}^{r_1}_{c_{m^*}^l, c_{k^*}^r} \langle \boldsymbol{f}, \boldsymbol{\psi}^{r_1}_{c_{m^*}^l, c_{k^*}^r}\rangle, \boldsymbol{\epsilon}\Big\rangle.\end{aligned} \tag{20}$$

W.l.o.g. assume that $b^* \in [r_J, c_{k^*}^r)$ and a similar approach holds when $b^* \in [c_{m^*}^l, r_J)$. Using Lemma 1, we have that for the left-hand side of (20),

$$\Big(\tilde{f}^{r_J}_{c_{m^*}^l, c_{k^*}^r}\Big)^2 - \Big(\tilde{f}^{b^*}_{c_{m^*}^l, c_{k^*}^r}\Big)^2 = \frac{|r_J - b^*|(r_J - c_{m^*}^l + 1)}{|r_J - b^*| + (r_J - c_{m^*}^l + 1)}\Big(\Delta_J^f\Big)^2 := \Lambda. \tag{21}$$

Also, for the right-hand side

$$\Big(\tilde{\epsilon}^{b^*}_{c_{m^*}^l, c_{k^*}^r}\Big)^2 - \Big(\tilde{\epsilon}^{r_J}_{c_{m^*}^l, c_{k^*}^r}\Big)^2 \leq \max_{s,e,b: s \leq b < e} \Big(\tilde{\epsilon}^b_{s,e}\Big)^2 \leq 8 \log T \tag{22}$$

and using Lemma 1 and using (13),

$$\begin{aligned}&2\Big\langle \boldsymbol{\psi}^{b^*}_{c_{m^*}^l, c_{k^*}^r} \langle \boldsymbol{f}, \boldsymbol{\psi}^{b^*}_{c_{m^*}^l, c_{k^*}^r}\rangle - \boldsymbol{\psi}^{r_1}_{c_{m^*}^l, c_{k^*}^r} \langle \boldsymbol{f}, \boldsymbol{\psi}^{r_1}_{c_{m^*}^l, c_{k^*}^r}\rangle, \boldsymbol{\epsilon}\Big\rangle \\ &\quad \leq 2\big\|\boldsymbol{\psi}^{b^*}_{c_{m^*}^l, c_{k^*}^r} \langle \boldsymbol{f}, \boldsymbol{\psi}^{b^*}_{c_{m^*}^l, c_{k^*}^r}\rangle - \boldsymbol{\psi}^{r_1}_{c_{m^*}^l, c_{k^*}^r} \langle \boldsymbol{f}, \boldsymbol{\psi}^{r_1}_{c_{m^*}^l, c_{k^*}^r}\rangle\big\|_2 \sqrt{8 \log T} \\ &\quad = 2\sqrt{\Lambda}\sqrt{8 \log T}.\end{aligned} \tag{23}$$

Using (21), (22) and (23), we can conclude that (20) 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 (21) and since

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

we can conclude that

$$\min\{|r_J - b^*|, r_J - c^l_{m^*}+1\} > \frac{2\left(2\sqrt{2}+4\right)^2 \log T}{\left(\Delta^f_J\right)^2} = C_3 \frac{\log T}{\left(\Delta^f_J\right)^2} \tag{24}$$

implies (19). Now, if

$$\min\{r_J - c^l_{m^*}+1, c^r_{k^*} - r_J\} > C_3 \frac{\log T}{\left(\Delta^f_J\right)^2}, \tag{25}$$

then (24) is restricted to $|r_J - b^*|\left(\Delta^f_J\right)^2 > C_3 \log T$ and this implies (19). So, we conclude that necessarily

$$|r_J - b_J|\left(\Delta^f_J\right)^2 \leq C_3 \log T. \tag{26}$$

But (25) 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 $\lambda_T \leq \delta_T/6$, 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 (26).

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. Now, we will prove that there is no other change-point in the interval and to do that we need to consider two different cases, depending on the location of the true change-point compared to the estimated one. We only show the proof for the case when $c^l_{m^*} \leq b_J < r_J < c^r_{k^*}$ and note that the proof when $c^l_{m^*} \leq r_J \leq b_J < c^r_{k^*}$ is very similar.

We need to show that there is no change-point in either of the intervals $[c^l_{m^*}, b_J)$ and $[b_J + 1, c^r_{k^*})$ besides r_J. Starting with the first one, for $c^l_{\tilde{m}} \in I^L_J$ as was defined in Step 3,

$$r_J - c^l_{m^*} \leq r_J - c^l_{\tilde{m}} < \frac{\delta_T}{3}.$$

Since $r_J - r_{J-1} \geq \delta_T$ and $b_J < r_J$, there is no change-point in the interval $[c^l_{m^*}, b_J)$. Now, using the assumption of Theorem 1, we have that

$$C_1\sqrt{\log T} \leq \zeta_T < C_2\sqrt{\delta_T}\underline{f}_T \Rightarrow \underline{C}C_2 > C_1 \Leftrightarrow \underline{C}\left(\frac{1}{\sqrt{12}} - \frac{2\sqrt{2}}{\underline{C}}\right) > \sqrt{C_3} + 2\sqrt{2}$$
$$\Leftrightarrow \underline{C} > 3\sqrt{2}\left(\sqrt{C_3} + 4\sqrt{2}\right)$$
$$\Rightarrow \underline{C} > 3\sqrt{2C_3} \Rightarrow \underline{C} > \sqrt{\frac{3}{2}C_3}$$

and by assumption (A1),

$$\delta_T \geq \frac{\underline{C}^2 \log T}{\underline{f}_T^2} > \frac{3}{2}\frac{C_3 \log T}{\underline{f}_T^2} \geq \frac{3C_3 \log T}{2\left(\Delta_J^f\right)^2}.$$

So, we have that necessarily

$$\delta_T > \frac{3}{2}C_3\frac{\log T}{\left(\Delta_J^f\right)^2}, \tag{27}$$

Combining (27) with (26), we have that for $c^r_{\tilde{k}} \in I_J^R$ as in Step 3,

$$c^r_{k^*} - b_J \leq c^r_{\tilde{k}} - b_J = c^r_{\tilde{k}} - r_J + r_J - b_J \leq \frac{\delta_T}{3} + \frac{C_3 \log T}{\left(\Delta_J^f\right)^2} < \delta_T$$

and since $r_J \in [b_J+1, c^r_{k^*})$ and $r_{J+1} - r_J \geq \delta_T$, there is no other change-point in the interval $[b_J+1, c^r_{k^*})$ besides r_J.

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}. As before, for $k_{J-1}, m_{J-1} \in \{0, 1, \ldots, K^{\max}\}$ we will show that r_{J-1} gets detected in $[c^l_{m^*_{J-1}}, c^r_{k^*_{J-1}}]$, where $c^l_{m^*_{J-1}} \geq c^l_{m_{J-1}}$ and $c^r_{k^*_{J-1}} \leq c^r_{k_{J-1}} \leq c^l_{m^*}$ and its detection is at location

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

which satisfies $\left|r_{J-1} - b_{J-1}\right|\left(\Delta^f_{J-1}\right)^2 \leq C_3 \log T$. Following similar steps as in (17), we have that for $\tilde{b}_{J-1} = \text{argmax}_{c^l_{m_{J-1}} \leq t < c^r_{k_{J-1}}} \left|\tilde{X}^t_{c^l_{m_{J-1}}, c^r_{k_{J-1}}}\right|$,

$$\left|\tilde{X}^{\tilde{b}_{J-1}}_{c^l_{m_{J-1}}, c^r_{k_{J-1}}}\right| \geq \left|\tilde{f}^{\tilde{r}_{J-1}}_{c^l_{m_{J-1}}, c^r_{k_{J-1}}}\right| - \sqrt{8\log T} \geq \sqrt{\frac{\min\{c^r_{k_{J-1}} - r_{J-1}, r_{J-1} - c^l_{m_{J-1}} + 1\}}{2}}\Delta^f_{J-1}. \tag{28}$$

Now, to show that $\min\{c^r_{k_{J-1}} - r_{J-1}, r_{J-1} - c^l_{m_{J-1}} + 1\} > \delta_T/6$, we need to take 2 cases as in Step 3. For all cases, $r_{J-1} - c^l_{m_{J-1}} + 1 > \delta_T/6$ for exactly the same reasons as $r_J - c^l_m + 1 > \delta_T/6$. Now, for $c^r_{k_{J-1}}$, we can consider two different scenarios

- If $c^r_{k_{J-1}} < c^l_{m^*}$, then the lower bound can be calculated for each case exactly as in Step 3.

- If $c^r_{k_{J-1}} = c^l_{m^*}$, then $c^r_{k_{J-1}} \geq c^l_{\tilde{m}}$, and so $c^r_{k_{J-1}} - r_{J-1} \geq c^l_{\tilde{m}} + \delta_T - r_J = \delta_T - (r_J - c^l_{\tilde{m}}) > \delta_T - \delta_T/3 = 2\delta_T/3 > \delta_T/6$

In any case, $\min\{c^r_{k_{J-1}} - r_{J-1}, r_{J-1} - c^l_{m_{J-1}} + 1\} > \delta_T/6$ holds and so, from (28) we have that

$$\left|\tilde{X}^{\tilde{b}_{J-1}}_{c^l_{m_{J-1}}, c^r_{k_{J-1}}}\right| \geq \sqrt{\frac{\delta_T}{12}}\Delta^f_{J-1} - \sqrt{8\log T} \geq \sqrt{\frac{\delta_T}{12}}\underline{f}_T - \sqrt{8\log T}$$
$$\geq \left(\frac{1}{\sqrt{12}} - \frac{2\sqrt{2}}{\underline{C}}\right)\sqrt{\delta_T}\underline{f}_T = C_2\sqrt{\delta_T}\underline{f}_T > \zeta_T.$$

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 $[c^r_{k^*_{J-1}}, c^l_{m^*}]$. 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 $c^r_{k^*_{N+1}} = T$. In total, $N + 1$ 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. □

4 Computational complexity and practicalities

4.1 Parameter selection

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

$$\zeta_T = C\sqrt{\log T}. \tag{29}$$

For the choice of the threshold constant C, we ran a simulation study using various signal structures and Gaussian noise. The best behaviour occurred for $C = 1.7$ and this is what was used for the simulations in Section 5. Anastasiou and Fryzlewicz [2022] prove in Corollary 1 that as $T \to \infty$, the threshold can be taken to be at most $\sqrt{3\log T}$ meaning that $C \leq \sqrt{3} \approx 1.73$ and so our choice of the constant does not violate this result.

Choice of the expanding parameter λ_T: As can be seen by the proof of Theorem 1, in a given signal, detection will occur for any $\lambda_T \leq \delta_T/6$. 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 lead to large computational complexity.

4.2 Computational complexity and variants

DAIS needs to check at most $2K$ intervals, where $K = \lceil \frac{d_{1,T}}{\lambda_T} \rceil + \lceil \frac{T-d_{1,T}+1}{\lambda_T} \rceil$, as defined in Section 2, and so $K \in \{\lceil \frac{T}{\lambda_T} \rceil, \lceil \frac{T}{\lambda_T} \rceil + 1\}$. Half of these intervals will be left- and right-expanding intervals, while the other half will be only left- and only right-expanding intervals, with the exact number for each determined by the location of the largest difference, $d_{1,T}$. This can be easily deduced from (6). Since we take $\lambda_T < \delta_T$, $K > \lceil T/\delta_T \rceil$ and so the lower bound of K is $\mathcal{O}(T/\delta_T)$. As explained by Anastasiou and Fryzlewicz [2022], the lower bound for the maximum number of intervals that need to be checked in ID is $\mathcal{O}(T/\delta_T)$ and in WBS and NOT it is $\mathcal{O}(T^2/\delta_T^2)$ up to a logarithmic factor, which means that DAIS is faster than the latter two and has the same order as ID. However, in ID the intervals checked will be generally larger than the intervals DAIS checks. When no change-points are detected, ID will check a large number of big chunks of the data, since for example it checks twice the whole length of the data, while DAIS will check it only once, in the left- and right-expanding intervals. The one-sided expanding intervals will be as big as the whole sequence only in the unlucky case that the largest difference is at an endpoint of the interval. This gives a computational advantage to DAIS over ID and thus all aforementioned competitors.

In practice, we use two modifications to the algorithm as described in Section 2, in order to make it faster and more accurate. The first one is to only check up to 100 expansions to either side of the location of the largest difference instead of all the possible expansions. When $\lambda_T = 3$, as described in Section 4.1, the biggest interval checked in this phase of the algorithm has length equal to 600, which in practice is large enough to detect any change-points around $d_{s,e}$. The second alteration is to restart DAIS from the detected change-point $\hat{r}$ instead of the start- and end-points of the interval where the detection occurred. This leads to the improvement of the accuracy of the algorithm without affecting the speed.

5 Simulations

This section compares the performance of DAIS with state-of-the-art competitors that are widely used nowadays. The competitors can be found in Table 1, along with the R library from which each function was obtained. The competitors were applied using the default values of their functions as developed in R. For WBS we report the results for both the information criterion and the thresholding (for $C = \sqrt{2}$) stopping rules. The notation is WBSIC and WBSTH, respectively.

Method Notation	Reference	R package
ID	Anastasiou and Fryzlewicz [2022]	IDetect
PELT	Killick et al. [2012]	changepoint
WBS	Fryzlewicz [2014]	wbs
WBS2	Fryzlewicz [2020]	breakfast
NOT	Baranowski et al. [2019]	not

Table 1: Methods used in the simulations

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

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

where T is the total length of the sequence and $\hat{f}_t$ is the ordinary least square approximation of f_t between two successive change-points. We also report the scaled Hausdorff distance (d_H), 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 is the length of the largest segment between successive change-points. d_H is given for all signals except for the constant signal *justnoise*, given in Table 3, for which d_H is uninformative. Finally, the average time taken for each method to find the change-points is given.

The signals used are the following:

- *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 $\sigma = 1$.
- *small_dist*2: sequence of length 1000 with 4 change-points at 485, 515, 900 and 930 with values between change-points 0, 1, 0, 1.5, 0. The standard deviation is $\sigma = 1$.
- *small_dist*3: 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 $\sigma = 1$.
- *justnoise*: sequence of length 6000 with no change-points. The standard deviation is $\sigma = 1$.
- *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 $\sigma = 1$.
- *stairs*: sequence 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 $\sigma = 0.3$.

Table 2 shows a significant advantage of DAIS over all competitors. These signals are particularly difficult for most algorithms because we have pairs of consecutive change-points that shift the data sequence to opposite directions, leading to masking the existence of the change-points. Detection is more difficult when the change-points are close to each other and their isolation is unlikely (e.g. for WBS) or occurs only in a limited number of intervals (e.g. for ID). Three examples of the signals used can be seen in Figure 3. Since DAIS will most likely start the expansion around the change-points, it has an advantage in detecting them, as the power of the contrast function $C^b_{s,e}(\boldsymbol{X})$ is maximised when b is at the midpoint of $[s, e]$. For *small_dist*, DAIS has an 11% improvement over the best performing competitor in the number of times that the correct number of change-points is detected, which is done with extremely low computational time. For *small_dist*2 and *small_dist*3, the improvement over the best performing competitor is even bigger, with 14% and 16% respectively.

For the signal *justnoise*, all algorithms, except for WBSTH, have very good results with negligible differences in the accuracy in the number of change-points detected. However, DAIS is the second fastest method. The signal *long_signal* shows a significant reduction in computation time compared to ID, which Anastasiou and Fryzlewicz [2022] showed that outperforms the other competitors in most signal structures. Finally, the signal *stairs* provides evidence that DAIS works well in signals with a lot of change-points, even when there is limited spacing between them.

6 Real data

In this section, we apply DAIS to real data. The chosen dataset includes daily crime reports. The data was obtained by *dataMontgomery*, Montgomery County's open data website, which provides the public with direct access to crime statistic databases. The data used are derived from reported crimes classified according to the National Incident-Based Reporting System (NIBRS) of the Criminal Justice Information Services (CJIS) Division Uniform Crime Reporting (UCR) Program and documented by approved police incident reports. The data can be found here. We use daily observations starting from 22/01/2020, up to 17/10/2022. Since the data involve counts of crimes, they are positive integer-valued and thus we perform a transformation to bring them closer to Gaussian data with constant variance. This is done using the Anscombe transform, $\alpha : \mathbb{N} \to \mathbb{R}$, with $\alpha(x) = 2\sqrt{x + 3/8}$ (Anscombe [1948]). Figure 4 is a plot of the true data, before the transformation, along with the estimated underlying

Method	Model	$\hat{N} - N$							MSE	d_H	Time (s)
		≤ -3	-2	-1	0	1	2	≥ 3			
DAIS		0	14	1	80	3	2	0	0.0150	0.091	0.00593
ID		0	31	0	67	2	0	0	0.0166	0.169	0.00679
WBSTH		0	0	0	18	17	19	46	0.0277	0.259	0.0229
WBSIC	*small_dist*	0	29	0	70	1	0	0	0.0164	0.159	0.0229
WBS2		0	4	3	72	10	6	5	0.0150	0.0895	0.501
PELT		0	80	0	20	0	0	0	0.0266	0.413	0.00068
NOT		0	30	0	68	2	0	0	0.0161	0.164	0.0247
DAIS		0	13	1	82	4	0	0	0.0210	0.0657	0.00599
ID		0	29	0	68	3	0	0	0.0273	0.126	0.00677
WBSTH		0	0	0	17	18	18	47	0.0349	0.167	0.0232
WBSIC	*small_dist*2	0	27	0	71	2	0	0	0.0221	0.118	0.0232
WBS2		0	6	3	70	8	7	6	0.0221	0.0752	0.474
PELT		9	72	0	19	0	0	0	0.0389	0.378	0.0071
NOT		0	25	0	72	2	1	0	0.0217	0.111	0.0255
DAIS		0	10	5	81	1	3	0	0.0310	0.0510	0.00525
ID		2	35	1	61	1	0	0	0.0352	0.147	0.00615
WBSTH		0	0	0	22	24	19	35	0.0414	0.0922	0.0232
WBSIC	*small_dist*3	2	31	1	65	0	1	0	0.0331	0.132	0.0232
WBS2		0	6	5	70	9	6	4	0.0306	0.132	0.470
PELT		26	65	0	9	0	0	0	0.0596	0.437	0.00075
NOT		2	30	1	67	0	0	0	0.0333	0.128	0.0262

Table 2: Distribution of $\hat{N} - N$ of 100 simulated data sequences of the signals *small_dist*, *small_dist*2 and *small_dist*3. The average MSE, d_H and computational time are also given.

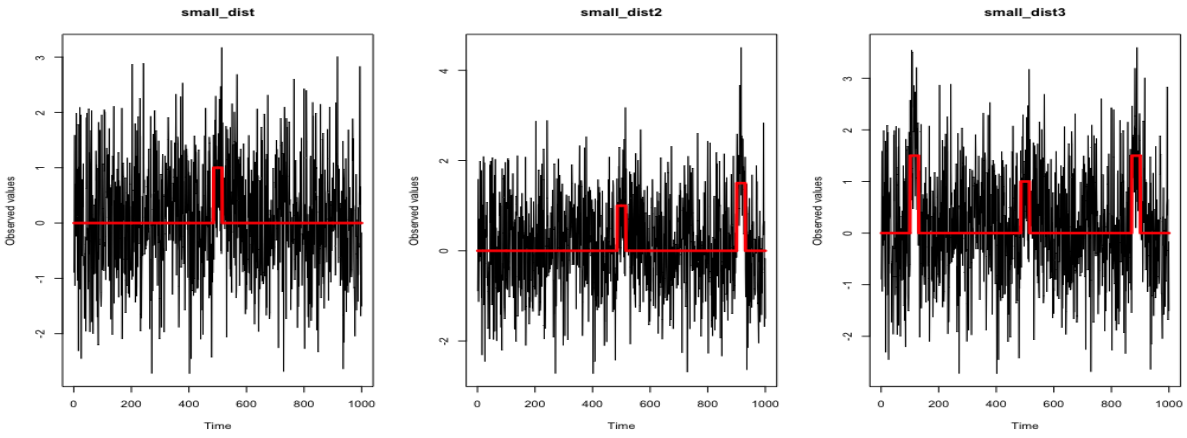

Figure 3: Examples of signals used in the simulations. The true signal, f_t, is plotted in red

Method	Model	$\hat{N} - N$ 0	1	2	≥ 3	MSE	Time (s)
DAIS		98	1	1	0	1.98×10^{-4}	0.261
ID		99	1	0	0	1.68×10^{-4}	0.353
WBSTH		22	16	31	31	31.7×10^{-4}	0.0859
WBSIC	*justnoise*	99	1	0	0	1.68×10^{-4}	0.0859
WBS2		91	5	3	1	2.75×10^{-4}	3.25
PELT		100	0	0	0	1.39×10^{-4}	0.00215
NOT		100	0	0	0	1.39×10^{-4}	0.0667

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

Method	Model	$\hat{N} - N$ -1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS		0	99	0	1	0	4.33×10^{-4}	43.1×10^{-4}	0.298
ID		0	100	0	0	0	4.25×10^{-4}	1.09×10^{-4}	0.657
WBSTH		0	43	14	24	19	16×10^{-4}	186×10^{-4}	0.148
WBSIC	*long_signal*	0	100	0	0	0	4.19×10^{-4}	1.06×10^{-4}	0.148
WBS2		0	92	3	4	1	5.46×10^{-4}	17.6×10^{-4}	6.27
PELT		0	100	0	0	0	4.19×10^{-4}	1.06×10^{-4}	0.00464
NOT		0	100	0	0	0	4.19×10^{-4}	1.06×10^{-4}	0.112

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

Method	Model	$\hat{N} - N$ ≤ -3	-2	-1	0	1	2	≥ 3	MSE	d_H	Time (s)
DAIS		0	0	2	94	4	0	0	0.0229	0.0135	0.00220
ID		0	0	0	87	12	1	0	0.0204	0.125	0.00346
WBSTH		0	0	0	66	27	6	1	0.0246	0.0163	0.0125
WBSIC	*stairs*	0	0	0	62	25	7	6	0.0246	0.167	0.0125
WBS2		0	0	2	92	5	1	0	0.0250	0.0153	0.0357
PELT		0	0	7	93	0	0	0	0.0216	0.0143	0.00044
NOT		0	0	0	93	7	0	0	0.0211	0.123	0.0504

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

signal, f_t plotted in red. A total of 8 change-points have been detected.

We now attempt to provide a possible explanation about the estimated change-points that express the most important movements in the data. The first one, on the 17^{th} of March 2020 corresponds to the first days of positive cases of COVID-19 and when the first rules were imposed to limit the spread of the virus among the community in the state of Maryland, where Montgomery County is located. As expected, since people were forced to stay at home, the number of crimes reported dropped significantly. The second change-point is on the 28^{th} of April 2020, when the official authorities started lifting the restrictions and the number of crimes increased. Some small decrease around the 24^{th} of December 2020 can be explained by new measures taken for the pandemic around the holidays. There is a drop in the reported crimes during the months June and July 2021, possibly due to the fact that people left the county to go on holidays, followed by an increase in September 2021, when people returned home. As has already been observed for 2020, there is again a reduced number of reported crimes around the Christmas holidays of 2021, starting from the 23^{rd} of December 2021 until the 30^{th} of January 2022.

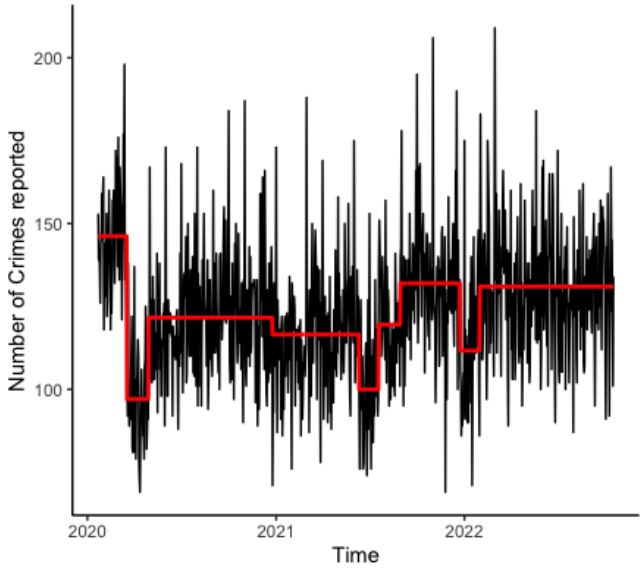

Figure 4: The time series and fitted piecewise-constant mean signal obtained by applying DAIS to the data of daily crime reports in Montgomery County.

7 Conclusions

In this project, DAIS is introduced, which is a new method for detecting structural changes in a given data sequence. In the case of changes in the mean, covered in this research project, the method first identifies the location of the largest difference in the sequence in an effort to start searching from an area around the most apparent change-point in the case that there is one. Expanding intervals are used to achieve isolation of the change-points, which provides an advantage over competitors in the detection power of the algorithm. The theoretical results regarding the consistency of the number of change-points detected and the accuracy of their estimated locations are given in Section 3. In Section 5, the advantage of DAIS compared to the state-of-the-art competitors is highlighted in Table 2, as the data-adaptive nature of the algorithm allows for the detection of the true change-points in these difficult structures, where 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 mid-point of these intervals, and so the power of the contrast function is maximized. DAIS outperforms the best performing competitor, with the difference in the amount of times the estimated and the actual

number of change-points agree being as large as 16% in these examples.

As with any algorithm, our method has drawbacks. Firstly, in practice the algorithm works really well using only the left- and right-expanding intervals. However, this does not guarantee isolation in theory and thus we need to add the only left-expanding and only right-expanding intervals feature in order to achieve the desired lower bound in Theorem 1. Also, even though DAIS is faster than the best-performing competitor, the expansions can cause the algorithm to be computationally expensive in cases of large signals with no change-points. We try to minimize this problem in Section 4.2 by introducing an upper bound to the number of the left- and right-expanding intervals used in Phase 1 of the algorithm.

References

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

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

Sylvain Arlot and Alain Celisse. Segmentation of the mean of heteroscedastic data via cross-validation. *Statistics and Computing*, 21(4):613–632, 2011.

Ivan E Auger and Charles E Lawrence. Algorithms for the optimal identification of segment neighborhoods. *Bulletin of mathematical biology*, 51(1):39–54, 1989.

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

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

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

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

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

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

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

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

Gabriela Ciuperca. A general criterion to determine the number of change-points. *Statistics & probability letters*, 81(8):1267–1275, 2011.

Gabriela Ciuperca. Model selection by lasso methods in a change-point model. *Statistical Papers*, 55 (2):349–374, 2014.

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

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

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

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

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

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

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

Kaylea Haynes, Paul Fearnhead, and Idris A Eckley. A computationally efficient nonparametric approach for changepoint detection. *Statistics and computing*, 27(5):1293–1305, 2017.

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

Seung Jun Shin, Yichao Wu, and Ning Hao. A backward procedure for change-point detection with applications to copy number variation detection. *Canadian Journal of Statistics*, 48(3):366–385, 2020.

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

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

Marc Lavielle. Using penalized contrasts for the change-point problem. *Signal Processing*, 85(8):1501–1510, 2005. ISSN 0165-1684. doi: https://doi.org/10.1016/j.sigpro.2005.01.012. URL https://www.sciencedirect.com/science/article/pii/S0165168405000381.

Chung-Bow Lee. Estimating the number of change points in exponential families distributions. *Scandinavian Journal of Statistics*, 24(2):201–210, 1997.

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

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

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

Adam B Olshen, E Seshan Venkatraman, Robert Lucito, and Michael Wigler. Circular binary segmentation for the analysis of array-based dna copy number data. *Biostatistics*, 5(4):557–572, 2004.

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

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

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

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

Ennapadam Seshan Venkatraman. *Consistency results in multiple change-point problems*. Stanford University, 1992.

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

Daren Wang, Yi Yu, and Alessandro Rinaldo. Univariate mean change point detection: Penalization, CUSUM and optimality. *Electronic Journal of Statistics*, 14(1):1917 – 1961, 2020. doi: 10.1214/20-EJS1710. URL https://doi.org/10.1214/20-EJS1710.

Yuehua Wu. Simultaneous change point analysis and variable selection in a regression problem. *Journal of Multivariate Analysis*, 99(9):2154–2171, 2008.

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

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