

# Consistency of a range of penalised cost approaches for detecting multiple changepoints

Chao Zheng<sup>†,\*</sup>, Idris Eckley<sup>†</sup> and Paul Fearnhead<sup>†</sup>

<sup>†</sup> *Department of Mathematics and Statistics  
Lancaster University*

<sup>\*</sup> *School of Mathematical Sciences  
University of Southampton*

**Abstract:** A common approach to detect multiple changepoints is to minimise a measure of data fit plus a penalty that is linear in the number of changepoints. This paper shows that the general finite sample behaviour of such a method can be related to its behaviour when analysing data with either none or one changepoint. This property results in simpler conditions for verifying whether the method will consistently estimate the number and locations of the changepoints. We apply and demonstrate the usefulness of these simple conditions for a range of changepoint problems. Our new results include a weaker requirement on the choice of penalty to have consistency in a change-in-slope model; and the first results for the accuracy of recently-proposed methods for detecting spikes.

**MSC2020 subject classifications:** Primary 62F30; secondary 60K35.

**Keywords and phrases:** Change-point detection, consistency, local region condition, changes in slope, spike plus exponential decay .

## 1. Introduction

Detecting changepoints is a long-standing problem in statistics, dating at least as far back as [Page \(1955\)](#). In recent years, there has been an explosion of research into methods for detecting multiple changes in data, stimulated by the growing need across many and diverse application areas. For example, changepoint detection methods have been applied in finance ([Preuß et al., 2015](#)), bioinformatics ([Cribben and Yu, 2017](#)), network traffic ([Lung-Yut-Fong et al., 2012](#)) and climatology ([Itoh and Kurths, 2010](#)).

There are a number of generic ways to detect multiple changepoints, many of which are based on recursively or repeatedly applying a method that detects and locates single changepoint. These include binary segmentation ([Vostrikova, 1981](#)) and its variants such as wild binary segmentation ([Fryzlewicz, 2014](#)) and circular binary segmentation ([Olshen et al., 2004](#)); simultaneous multi-scale changepoint estimation ([Frick et al., 2014](#); [Li et al., 2016](#)); scan statistics ([Eichinger and Kirch, 2018](#)); and the narrowest-over-threshold approach ([Baranowski et al., 2016](#)). This paper considers a different class of popular changepoint algorithms, that aims to jointly detect multiple changepoints through minimising a penalised cost. This cost involves a measure of fit to the data together

with a penalty that increases proportionally with the number of changepoints. These penalised cost ideas are closely related to penalised likelihood methods and model choice approaches (Birgé and Massart, 2001) for linear models that employ an  $L_0$  penalty.

In practice, different applications often require the ability to detect different types of change, or have different types of data structure. The simplest case is detecting multiple changes-in-mean in Gaussian data, where the goal is equivalent to find the optimal piecewise-constant mean function. Fast algorithms for penalised cost approaches with an expected run time that can be linear in the amount of data (Killick et al., 2012; Maidstone et al., 2018) have been developed. These algorithms have been widely used in applications from RNA sequencing (Cleynen et al., 2014) to analysis of historical warfare (Fagan et al., 2019). For detecting other types of change, for example, changes in slope (Fearnhead et al., 2018), spikes plus exponential decay (Jewell and Witten, 2018), and changes in mean of heavy-tailed data (Fearnhead and Rigai, 2019), efficient algorithms have also been proposed, and have been implemented to solve real-world problems such as determining the exact moment in time at which a neuron spikes in calcium image data (Jewell et al., 2019).

In terms of theoretical analysis, existing theory is well-developed for penalised cost approaches for detecting change-in-mean (Yao, 1988; Lavielle and Moulines, 2000; Tickle et al., 2018). For example, a specific penalty value is known to give consistent estimates of the number of changepoints and empirical results suggest that this penalty choice is tight: with lower values frequently leading to over-estimating the number of changes. However, separate results need to be carried out if we want to look at different types of change. Whilst detecting different types of change seems to be similar statistical problems to detecting changes in mean, it is fundamentally more challenging in certain cases due to the complexity of underlying signals. Asymptotic results about consistent estimates of the number and locations of changepoints in those cases are much more limited and weaker, if not unavailable, in existing literatures, e.g., often opposing an unspecific value of penalty that depends on a loosely defined large enough constant (Fearnhead et al., 2018) and/or requiring a finite upper bound on the number of changes as in Yao (1988). To our knowledge, the most general results are in Boysen et al. (2009a) which provides a simple argument that demonstrates consistency for a wide class of changepoint models if we measure fit via a residual sum of squares, and have a penalty that increases faster than the logarithm of the number of data points. In this paper we develop a general statistical theory for penalised cost approaches of detecting different types of changes, and, in particular, provide theoretical guidance on how the penalty for each additional changepoint, should be chosen.

As the first contribution of this paper, we establish a general framework of penalised cost approaches, which applies to all multiple changepoint models that admit an additive cost as the measure of fit, and show that the performance of the approaches is related to their behaviour when analysing data with either no change or one changepoint. We call these two specific cases as local regions. Analysing the behaviour of penalised cost methods in local regions is much sim-

pler than for the case of multiple changes. Informally our results show, subject to additional conditions, that a choice of penalty that leads to consistent estimates of the number of changes when there is one or no changepoints will directly lead to a consistent estimator when there are multiple changes. Therefore, for different types of changepoint problem, we only need to focus on their properties in local regions. We propose a set of general conditions of detectability of changes across all local regions, that will be sufficient to provide consistent estimator of the number and locations of multiple changes for the entire data.

As the second contribution, we apply the proposed framework to different multiple changepoint problems as applications, showing that the local region conditions are satisfied, and hence obtain the theoretical results on the choice of penalty and the corresponding consistent estimates of number of changepoints and their locations. These lead to new results for detecting changes-in-slope, with a weaker condition on the penalty that is needed to obtain consistent estimates comparing to existing results in [Fearnhead et al. \(2018\)](#); and lead to the first theoretical results for the problem of detecting spikes in an exponentially decaying signal. Using our framework, similar consistency results can be easily derived for many other multiple changepoint detection problems.

## 2. General Framework

### 2.1. Problem setup

Consider a general changepoint model with  $T$  observations  $x_{1:T} = \{x_1, \dots, x_T\}$  ordered in sequence, for example by time or position along a chromosome. Assume that there are a set of  $m$  changepoints, at ordered locations,  $\tau_{1:m} = \{\tau_1, \dots, \tau_m\}$ , with  $0 < \tau_1 < \tau_2 < \dots < \tau_m < T$ . This will partition the data into  $m + 1$  distinct segments with the  $j$ -th segment including the observations  $x_{\tau_{j-1}+1:\tau_j} = \{x_{\tau_{j-1}+1}, \dots, x_{\tau_j}\}$ , where we write  $\tau_0 = 0$  and  $\tau_{m+1} = T$ . In other words, there is a common structure for the data within each single segment, but the nature of this structure can change between segments. We characterise this structure using segment specific parameters  $\boldsymbol{\theta}_{1:m+1} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{m+1}\}$ , which depending on the application, could, for example, be the mean changes of the data between segments, or the variance changes, or both, among many other possibilities.

We wish to estimate both the number and locations of the changepoints. To this end we focus on the  $L_0$  penalised cost methods. These introduce a segment cost, which measures the fit to data within the segment. Often appropriate costs are specified by modelling the data, and setting the cost to be the negative of the log-likelihood under such a model. For a segment with data  $x_{s:e}$  for some  $e > s$ , we have a segment cost  $\mathcal{C}(x, s, e; \boldsymbol{\theta})$  that will depend on the segment specific parameters  $\boldsymbol{\theta}$ . We will assume that the cost is additive over data points:

$$\mathcal{C}(x, s, e; \boldsymbol{\theta}) = \sum_{t=s}^e c_t(x_{1:T}, \boldsymbol{\theta}). \quad (2.1)$$

where,  $c_t(x_{1:T}, \boldsymbol{\theta})$ , the cost associated with data point  $x_t$  can also depend on other data-points, which allows dependency between  $x_{1:T}$ . This additive assumption is a quite generally property for segment cost, which holds for most problems in changepoint literature.

We then define the cost for fitting a set of  $m$  changepoints,  $\tau_{1:m}$ , with associated parameters  $\boldsymbol{\theta}_{1:m+1}$  as

$$\sum_{j=1}^{m+1} \mathcal{C}(x, \tau_{j-1} + 1, \tau_j; \boldsymbol{\theta}_j).$$

We minimise over the parameters  $\boldsymbol{\theta}_{1:m+1}$  simultaneously to define the cost associated with the segmentation:

$$\mathcal{L}(x_{1:T}; \tau_{1:m}) = \min_{\boldsymbol{\theta}_{1:m+1}} \sum_{j=1}^{m+1} \mathcal{C}(x, \tau_{j-1} + 1, \tau_j; \boldsymbol{\theta}_j),$$

where, potentially, the minimisation can be subject to constraints on the segment parameters. For simplicity, here we assume the constraints fix the relationship between the parameters of neighbouring segments. For example, we may enforce strict monotonicity in a change-in-mean model, i.e,  $\theta_{j+1} \in (\theta_j, \infty)$ , where  $\theta_j$  is mean value in  $j$ -th segment; or in a change-in-slope model  $\theta_j$  will specify a linear function for the mean signal within  $j$ -th segment, and the constraints would enforce continuity at the changepoints for two consecutive linear functions. Our results in this paper will remain valid for more general constraints as long as (2.3) and (2.4) in Lemma 2.1, given below, still hold.

We can extend our definition of cost so that it applies to a subset of data  $x_{s:e}$  and a series of changepoint locations  $s \leq \tau_{u:v} < e$ . Using the notation that  $\tau_{u-1} = s - 1$  and  $\tau_{v+1} = e$ , we have

$$\mathcal{L}(x_{s:e}; \tau_{u:v}) = \min_{\boldsymbol{\theta}_{u:v+1}} \sum_{j=u}^{v+1} \mathcal{C}(x, \tau_{j-1} + 1, \tau_j; \boldsymbol{\theta}_j).$$

We take the convention that if the set of changepoints contains changes outside the region of data, then these changes are ignored. For example, if  $\tau < s$  or  $\tau \geq e$  then  $\mathcal{L}(x_{s:e}; \tau_{u:v}, \tau) = \mathcal{L}(x_{s:e}; \tau_{u:v})$ . Let  $\mathcal{L}(x_{s:e}; \emptyset)$  denote the segmentation cost where there is no changepoint between  $s$  and  $e$ .

If we know the number of changepoints,  $m$ , it would be natural to estimate their locations by the set  $\hat{\tau}_{1:m}$ , which minimises  $\mathcal{L}(x_{1:T}; \tau_{1:m})$ . However, in practice we need to also estimate the number of changepoints. Therefore we consider methods that estimate  $m$  and  $\tau_{1:m}$  simultaneously as the value that minimises the  $L_0$  penalised cost:

$$\arg \min_{(m, \tau_{1:m})} \{ \mathcal{L}(x_{1:T}; \tau_{1:m}) + \beta m \}, \quad (2.2)$$

where  $\beta > 0$  is a user-defined tuning parameter that penalises the addition of each changepoint. We call (2.2) minimisation of  $L_0$  penalised cost as  $m\beta$  can

be viewed as an  $L_0$  penalty on the difference in segment parameters associated with neighbouring time-points.

We will use the superscript  $*$  to denote the true changepoint locations and parameter values. That is, the true model will have  $m^*$  changepoints, at locations  $\tau_{1:m^*}^*$ , and with segment parameters  $\theta_{1:m^*+1}^*$ . If we impose constraints when calculating the cost of a segmentation, we require the true segment parameters to satisfy those constraints. We will also define  $\mathcal{L}^*(x_{s:e})$  to be the cost of fitting data  $x_{s:e}$  with the true set of changepoints and the true parameters, that is

$$\mathcal{L}^*(x_{s:e}) = \mathcal{C}(x, s, \tau_u^*; \theta_u^*) + \mathcal{C}(x, \tau_v^*, e; \theta_{v+1}^*) + \sum_{j=u+1}^v \mathcal{C}(x, \tau_{j-1}^*, \tau_j^*; \theta_j^*),$$

where  $u$  and  $v$  are defined so that  $\tau_u^*$  and  $\tau_v^*$  are, respectively, the first and last true change between  $s$  and  $e$ ; and if  $u = v$  we can set the summation part to be 0. If  $x_{s:e}$  does not contain a change then  $\mathcal{L}^*(x_{s:e}) = \mathcal{C}(x, s, e; \theta_u^*)$  where  $u$  is the index of the true segment that  $x_{s:e}$  lie in.

The following properties of the cost, which follow from the additive assumption (2.1), will be important for results in subsequent sections.

**Lemma 2.1.** *Assume (2.1) holds, then for any  $s \leq r < e$  and  $s \leq \tau_{u:v} < e$ , we have*

$$\mathcal{L}(x_{s:e}; \tau_{u:v}) \geq \mathcal{L}(x_{s:r}; \tau_{u:v}) + \mathcal{L}(x_{r+1:e}; \tau_{u:v}), \text{ and} \quad (2.3)$$

$$\mathcal{L}^*(x_{s:e}) = \mathcal{L}^*(x_{s:r}) + \mathcal{L}^*(x_{r+1:e}). \quad (2.4)$$

## 2.2. Local region conditions

Our aim is to build general conditions under which estimating the number and locations of the changepoints via a  $L_0$  penalised cost approach will be consistent, and to quantify the accuracy within which the locations are estimated. We will achieve this by relating properties of penalised cost approach when analysing data with multiple changes in  $x_{1:T}$  to its properties when analysing data with either zero or one changepoint in a local region, i.e.  $x_{t+1:t+n}$  or  $x_{t+1:t+2n}$ , respectively, where  $n$  or  $2n$  is the number of data points the local region.

To this end we introduce the following conditions that govern the value of the penalised cost procedure when fitting data simulated with either zero or one true changepoint with either the correct or too many number of changepoints. These conditions need to apply for our assumed data generating mechanism and our choice of penalised cost. We assume the data generating mechanism is parameterised by the set of changepoints and the segment parameters.

**Condition 2.1.** There exists increasing positive constants  $\gamma_n^{(1)}$  and  $\gamma_n^{(2)}$ , such that  $\gamma_n^{(1)}$  and  $\gamma_n^{(2)} \rightarrow \infty$  as  $n \rightarrow \infty$ ; and there exists positive numbers  $a(\gamma, n)$  and  $b(\gamma, n)$  increasing in  $\gamma$ , and positive functions,  $p_j(\gamma, n)$  such that  $p_j(\gamma, n) \rightarrow 0$  as  $\gamma \rightarrow \infty$ , for  $j \in \{1, 2, 3, 4\}$ , such that:

- (i) Let  $\mathbf{S}_{1,n}(t) = x_{t+1:t+n}$  be a segment between  $t+1$  and  $t+n$  that includes no changepoint. If  $\gamma \geq \gamma_n^{(1)}$ ,

$$\max_t \mathbb{P} \left( \min_{k \geq 1, \tau_{1:k}} \{ \mathcal{L}(\mathbf{S}_{1,n}(t); \tau_{1:k}) + k\gamma \} - \mathcal{L}^*(\mathbf{S}_{1,n}(t)) \leq a(\gamma, n) \right) \leq p_1(\gamma, n),$$

$$\max_t \mathbb{P} \left( \mathcal{L}^*(\mathbf{S}_{1,n}(t)) - \mathcal{L}(\mathbf{S}_{1,n}(t); \emptyset) \geq b(\gamma, n) \right) \leq p_2(\gamma, n),$$

where probability is with respect to the data generating mechanism for  $\mathbf{S}_{1,n}(t)$ .

- (ii) Let  $\mathbf{S}_{2,n}(t) = x_{t+1:t+2n}$  be a segment between  $t+1$  and  $t+2n$  that has a single changepoint which is at  $t+n$ . If  $\gamma \geq \gamma_n^{(2)}$ , we have

$$\max_t \mathbb{P} \left( \min_{k \geq 2, \tau_{1:k}} \{ \mathcal{L}(\mathbf{S}_{2,n}(t); \tau_{1:k}) + (k-1)\gamma \} - \mathcal{L}^*(\mathbf{S}_{2,n}(t)) \leq a(\gamma, 2n) \right) \leq p_3(\gamma, n),$$

$$\max_t \mathbb{P} \left( \mathcal{L}^*(\mathbf{S}_{2,n}(t)) - \min_{\tau_1} \mathcal{L}(\mathbf{S}_{2,n}(t); \tau_1) \geq b(\gamma, 2n) \right) \leq p_4(\gamma, n),$$

where probability is with respect to the data generating mechanism for  $\mathbf{S}_{2,n}(t)$ .

The above condition bounds the reduction in the penalised cost, if we have a penalty of  $\gamma$  for adding a changepoint, that can be obtained by fitting too many changes. Note that this penalty of  $\gamma$ , is specific for local regions  $\mathbf{S}_{1,n}(t)$  and  $\mathbf{S}_{2,n}(t)$ , thus is dependent on  $n$  not  $T$ . Whilst it is different from the global penalty  $\beta$ , it plays an essential role in determining the value of  $\beta$ .

The above probabilities  $p_j(\gamma, n)$  are for the worst case over possible parameters of the data generating mechanism and time-points such that the specified region has no change or one change in the middle. In most situations, for example  $x_{1:T}$  are independent and identically distributed, the probabilities will be the same for all choices. Note that by considering the maximum over the set  $\{\mathbf{S}_{2,n}(t)\}$ , we must have  $m^* \geq 1$ , that is, there exists at least one true changepoint.

Moreover, we need another condition on the cost function if we do not fit a change near a true changepoint, which is given below.

**Condition 2.2.** Let  $\mathbf{S}_{\Delta,n}(t) = x_{t+1:t+2n}$  be a segment between  $t+1$  and  $t+2n$  that has a single changepoint which is at  $t+n$ , and, with  $\Delta = \text{dist}(\boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(2)})$ . Here  $\boldsymbol{\theta}_{(1)}$  and  $\boldsymbol{\theta}_{(2)}$  are the parameters associated with the segments immediately before and after  $t+n$ , and  $\text{dist}(\cdot, \cdot)$  is a suitable measure of distance in parameter space that may be differing across applications. Then we have

$$\max_t \mathbb{P} \left( \mathcal{L}(\mathbf{S}_{\Delta,n}(t), \emptyset) - \mathcal{L}^*(\mathbf{S}_{\Delta,n}(t)) \leq z \right) \leq p_5(\mathcal{S}(\Delta, n), z),$$

where  $\mathcal{S}(\Delta, n)$  is a function of signal strength, and we require  $p_5(y, z) \rightarrow 0$  as  $y \rightarrow \infty$  and  $y/z \rightarrow \infty$ , and the probability is with-respect to the data generating mechanism.

Condition 2.2 indirectly defines  $\mathcal{S}(\Delta, n)$  as the signal strength of a change from segment parameter  $\boldsymbol{\theta}_{(1)}$  to  $\boldsymbol{\theta}_{(2)}$  with data of length  $n$  on either side of the change point, where  $\Delta$  is some appropriate measure of distance between the two segment parameters. The idea is that the reduction in cost of not fitting the change will be of the order of this signal strength. Again we bound the worst-case probability, but in many cases the probability will be the same for all segments in the set of  $\{\mathbf{S}_{\Delta, n}(t)\}$ .

### 2.3. Global changepoint consistency

In this section, based on the introduced local region conditions, we establish consistency of estimates of the number and locations for the changepoints under the penalised cost approach when applied to data simulated with a general number  $m^*$  of changes. The result is not limited to the type of changes and the underlying data generating mechanism. Thus it builds a general framework to obtain consistency theories for a broad class of changepoint problems.

As earlier, we denote the location of the changes by  $\tau_{1:m^*}^*$ , and we will denote the true segment lengths by

$$\delta_j = \tau_j^* - \tau_{j-1}^*, \quad j = 1, \dots, m^* + 1,$$

with, as before,  $\tau_0^* = 0$  and  $\tau_{m^*+1}^* = T$ . Let the size of each change be denoted by

$$\Delta_j = \text{dist}(\boldsymbol{\theta}_{j+1}^*, \boldsymbol{\theta}_j^*), \quad j = 1, \dots, m^*,$$

where the distance  $\|\cdot\|$  is defined in Condition 2.2. In addition, we define  $\delta_T = \min_j \delta_j$  and  $\Delta_T = \min_j \Delta_j$ .

Given a set of integers  $n_{1:m^*}$  satisfying  $0 < n_j \leq \min\{\delta_j, \delta_{j+1}\}$ , we can partition the data into  $2m^* + 1$  regions  $\{\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_{2m^*+1}\}$ , such that:

$$\begin{cases} \mathbf{S}_{2j+1} = x_{(\tau_j^*+n_j+1):(\tau_{j+1}^*-n_j+1)}, \\ \mathbf{S}_{2j} = x_{(\tau_j^*-n_j+1):(\tau_j^*+n_j)}, \end{cases} \quad (2.5)$$

where we define  $n_0 = n_{m^*+1} = 0$ . See the top plot of Figure 1 for an example of this partitioning of the data. In this way, each region  $\mathbf{S}_j$  with an odd index  $j$  does not contain a true changepoint, and each region with an even index has exactly one true changepoint in the middle, satisfying our definition of local regions. Therefore we can verify if Conditions 2.1 and 2.2 are satisfied on  $\{\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_{2m^*+1}\}$ , depending on specific problem.

We are now ready to give a unified result for the changepoint estimation under our penalised cost criteria in the following theorem.

**Theorem 2.1.** *Assume Conditions 2.1 and 2.2 hold. For data  $x_{1:T}$ , let  $\hat{m}$  and  $\hat{\tau}_{1:\hat{m}}$  be the number and locations of changepoints we obtain by minimising the penalised cost (2.2) with penalty  $\beta$ . For any  $n_{1:m^*}$  where  $0 < n_j < \min\{\delta_j, \delta_{j+1}\}$ , define an event  $\mathcal{E}(\beta, n_{1:m^*})$  as*

$$\mathcal{E}(\beta, n_{1:m^*}) = \{\hat{m} = m^* \text{ and } |\hat{\tau}_j - \tau_j^*| \leq n_j, j = 1, \dots, m^*\}, \quad (2.6)$$

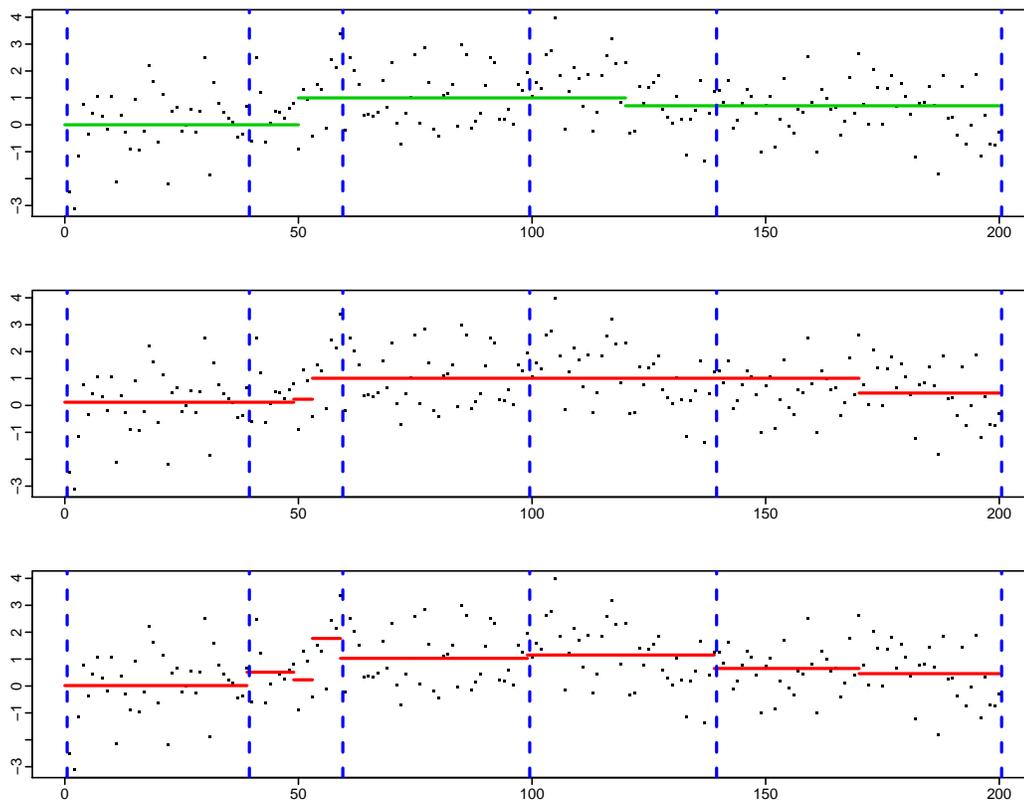


FIG 1. Top: Example partitioning of the data, for the univariate change-in-mean problem. The green-line is the true mean, which has two changes. Thus our partition has 5 regions,  $\mathbf{S}_{1:5}$ , with the even regions containing a change and the odd-regions containing no change. The size of the even regions can be chosen based on the size of the change, with larger regions around smaller changes. Middle: a segmentation of the data, together with fitted mean (red-line), that violates the event (2.6). Such a segmentation will have errors within at least one region, in this case the fitted segmentation fits too many changes in  $\mathbf{S}_2$ , misses the change in  $\mathbf{S}_4$  and erroneously fits a change in  $\mathbf{S}_5$ . To show the penalised cost of such a segmentation will not be optimal, under our event  $E_1$  (see the proof in Appendix) we bound the difference in the cost between such a segmentation and the true segmentation, by the difference of the cost if we fit the segmentation separately within each region (bottom figure) and the cost of the true segmentation with the true parameters (top figure). This difference is simply the sum of the differences of the fits in each region. The key idea is that for regions where a putative segmentation makes an error this difference will be sufficiently large that putative segmentation can not be optimal.

and a minimum signal strength as  $\mathcal{S} = \min_{j=1}^{m^*} \mathcal{S}(\Delta_j, n_j)$ . Then if

$$\beta \geq \max\{\gamma_T^{(1)}, \gamma_{\max_i n_j}^{(2)}\} \text{ and } a(\beta, T) > 2m^*b(\beta, T),$$

we have

$$\begin{aligned} \mathbb{P}\left(\mathcal{E}(\beta, n_{1:m^*})\right) &\geq 1 - (m^* + 1)p_1(\beta, T) - (m^* + 1)p_2(\beta, T) - m^*p_3\left(\beta, \max_j n_j\right) \\ &\quad - m^*p_4\left(\beta, \max_j n_j\right) - m^*p_5(\mathcal{S}, \beta + a(\beta, T)). \end{aligned} \quad (2.7)$$

The proof of Theorem 2.1 is deferred to the appendix. Here we give a pictorial outline of the proof, using change-in-mean problem as an example in Figure 1.

Theorem 2.1 provides the probability bound on the event that the estimated number of changes is correct and the estimated location of each change  $\tau_j^*$  is within the accuracy of  $n_j$ . The global penalty  $\beta$  is chosen by collecting the maximum of  $\gamma_T^{(1)}$  and  $\gamma_{n_j}^{(2)}$  for all  $1 \leq j \leq m^* + 1$ . If we specify any asymptotic regime such that

$$\frac{\mathcal{S}}{\beta + a(\beta, T)} \rightarrow \infty, \quad \text{as } T \rightarrow \infty.$$

Then if  $m^*$  is finite or diverging with  $T$  in a slow rate, we have

$$m^* \min \left\{ p_1(\beta, T), p_2(\beta, T), p_3(\beta, \max_j n_j), p_4(\beta, \max_j n_j), p_5(\mathcal{S}, \beta + a(\beta, T)) \right\} \rightarrow 0.$$

Hence such an event will hold with probability going to 1.

In the event such that  $m^* = 0$ , we do not need to split the data into local regions but treat  $x_{1:T}$  as in scenario (i) in Condition 2.1; therefore, if  $\beta \geq \gamma_T^{(1)}$ , we have (2.7) still holds. Also, it is simple to adapt the proof to show that we can replace  $\beta > \gamma_T$  with  $\beta > \gamma_{n_{\max}}$ , where  $n_{\max} = \max_j \{\delta_j\}$  is the maximum true segment length. This suggests the possibility of using smaller penalties in situations where the maximum segment length is known and is much shorter than  $T$ . For a given value of the minimum signal strength,  $\mathcal{S}$ , we can optimise the choice of  $n_{1:m^*}$  that bound the accuracy of our estimates of the locations of each changepoint and make it to be tighter. Specifically we can choose  $n_j$  to be the smallest value such that  $\mathcal{S}(\Delta_j, n_j) \geq \mathcal{S}$  for each  $j = 1, \dots, m^*$ .

### 3. Applications

In this section, to show the usefulness and broad applicability of the general framework developed in Section 2, we apply it to the estimation of change-in-mean, change-in-slope as well as changepoint in spike and exponential decay problems. The obtained consistency results for change-in-mean is not new, but it is more flexible than that in Yao (1988), allowing the number of changes,  $m^*$ , to diverge with  $T$ . For change-in-slope, the results is stronger than previous consistency results, as it carefully specifies the value of the penalty,  $\beta$ , that ensures

consistency. This provides very important information on how to select  $\beta$  and make the methods more accessible to practitioners. For spike and exponential decay, we provide the first consistency result of this problem.

Note that in this section we always assume  $m^* \geq 1$ , meaning that there exists at least one changepoint. Otherwise, we only need to verify scenario (i) of Condition 2.1 and the global changepoint consistency holds trivially in all the three problems. Proofs for results in this section are deferred to the appendix of this paper.

### 3.1. Change-in-mean problem

First, we revisit the canonical problem of detecting change-in-mean. Suppose we observe data  $x_{1:T}$  with underlying decomposition,  $x_t = \mu_t + \varepsilon_t$ , where  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$  are independent and identical distributed (i.i.d.) Gaussian, and  $\mu_t$  are piecewise constant means, i.e.:

$$\mu_t = \theta_j^*, \quad \text{if } \tau_{j-1}^* + 1 \leq t \leq \tau_j^*, \quad \text{for all } j = 1, \dots, m^* + 1.$$

To estimate the set of changepoints, we use the square error loss as the cost function to measure fit to the data. That is, fitting a set of points  $x_{s:e}$  with the same segment parameter,  $\theta$ , has cost function in the following form:

$$\mathcal{C}(x, s, e; \theta) = \sum_{t=s}^e \frac{(x_t - \theta)^2}{\sigma^2}. \quad (3.1)$$

Note that in this application, no constraint is imposed on the parameters that minimise the cost function. Therefore, in fact we wish to minimise over  $m$  and  $\tau_{1:m}$ , for the following penalised cost:

$$\sum_{j=1}^{m+1} \sum_{t=\tau_{j-1}+1}^{\tau_j} \frac{(x_t - \bar{x}_{\tau_{j-1}+1:\tau_j})^2}{\sigma^2} + m\beta, \quad (3.2)$$

where  $\bar{x}_{\tau_{j-1}+1:\tau_j} = \sum_{t=\tau_{j-1}+1}^{\tau_j} x_t / (\tau_j - \tau_{j-1})$ .

The above minimisation assumes knowledge of the noise variance  $\sigma^2$ , which can be regarded as a nuisance parameter. In practice if the variance is unknown, it is a common procedure in the literature to plug in, for example, the Median Absolute Deviation (MAD) estimator (Hampel, 1974) applied to the differenced data (Baranowski et al., 2016), without any effect on the correctness of the theoretical analysis below. More specifically, for this change-in-mean example we can set

$$\hat{\sigma} = \frac{\text{median}\{|x_2 - x_1|, \dots, |x_T - x_{T-1}|\}}{\sqrt{2}\Phi^{-1}(3/4)},$$

where  $\Phi^{-1}(\cdot)$  is the quantile function of the standard normal distribution.

It is also possible to include the effect of the consistent estimation of variance parameter in the theoretical analysis. To this end, a simple approach is to use

some sample-splitting procedure that will be helpful to avoid dependency in the analysis and simplify the results. For example, using the odd data to estimate the variance, and use even data to estimate the changepoints. A drawback of this method is we may increase the variance of our changepoint estimation.

As in Section 2.3, we define the size of the change at  $j$ -th true changepoint  $\tau_j^*$  as

$$\Delta_j = \text{dist}(\theta_{j+1}^*, \theta_j^*) = |\theta_{j+1}^* - \theta_j^*|, \quad j = 1, \dots, m^*,$$

which is the absolute mean difference in two consecutive segments. The following propositions show Conditions 2.1 and 2.2 are satisfied for this change-in-mean application.

**Proposition 3.1.** *Consider the following choices of  $\gamma_n^{(1)}$  and  $\gamma_n^{(2)}$*

$$\begin{aligned} \gamma_n^{(1)} &= \max \left\{ (2 + \epsilon) \log n, 2 \log n + \kappa_{1,1} \sqrt{\log n}, 2 \log n + \kappa_{1,2} m^* \right\}, \\ \gamma_n^{(2)} &= \max \{ \kappa_{1,3} m^* \log(2n), 2 \log(2n) + \kappa_{1,4} m^* \}, \end{aligned}$$

for some large enough constants  $\kappa_{1,1}, \kappa_{1,2}, \kappa_{1,3}, \kappa_{1,4}$ , and  $\epsilon$  is an positive constant that can be arbitrarily small. Moreover, let

$$a(\gamma, n) = \frac{\gamma - 2 \log n}{4} \quad \text{and} \quad b(\gamma, n) = \frac{\gamma - 2 \log n}{4(2m^* + 1)}. \quad (3.3)$$

We have the Conditions 2.1 are satisfied with

$$\begin{aligned} p_1(\gamma, n) &= 2 \exp \left( -\frac{\gamma - 2 \log n}{4} \right), \quad p_2(\gamma, n) = \exp \left( -\frac{\gamma - 2 \log n}{16(2m^* + 1)} \right), \\ p_3(\gamma, n) &= \exp \left( -\frac{\gamma - 8 \log(2n)}{4} \right), \quad p_4(\gamma, n) = \exp \left( -\frac{\gamma - (8m^* + 6) \log(2n)}{16(2m^* + 1)} \right). \end{aligned}$$

**Proposition 3.2.** *Let  $\mathcal{S}(\Delta, n) = n\Delta^2/2$ , we have Condition 2.2 is satisfied if  $\mathcal{S}(\Delta, n)/4 \geq z \geq 5$ , with  $p_5(\mathcal{S}(\Delta, n), z) = 2 \exp(-z/20)$ .*

In this application the data mechanism is uniform at all time-point, therefore the probability is the same for all elements in  $\{\mathbf{S}_{1,n}(t)\}$ ,  $\{\mathbf{S}_{2,n}(t)\}$  and  $\{\mathbf{S}_{\Delta,n}(t)\}$ . Here we remark that the above propositions illustrate as an example that Conditions 2.1 and 2.2 are satisfied for the change-in-mean problem, where the constants in  $\gamma_n^{(i)}$ ,  $a(\gamma, n)$  and  $p_j(\gamma, n)$ , where  $i = 1, 2$  and  $j = 1, \dots, 5$ , are not optimised. Note that by Proposition 3.1 we always have  $a(\gamma, n) > 2m^*b(\gamma, n)$ .

**Theorem 3.1.** *If  $\beta = (2 + \epsilon) \log T$ , where  $\epsilon > 0$  is an arbitrarily small constant; and if*

$$\delta_T \Delta_T^2 \geq (16 + 10\epsilon) \log T \quad \text{and} \quad \Delta_T^2 \geq \frac{(32 + 20\epsilon) \log T}{T^{1/(4m^* + 3)}}, \quad (3.4)$$

then for large enough  $T$ , with probability at least  $1 - (7m^* + 3)T^{-\epsilon/(32m^* + 16)}$ ,

$$\hat{m} = m^*, \quad \max_{j=1, \dots, m^*} |\hat{\tau}_j - \tau_j| \Delta_j^2 \leq (16 + 10\epsilon) \log T. \quad (3.5)$$

This result can be viewed as a finite-sample version of the existing consistency results for the change-in-mean problem (Yao, 1988; Tickle et al., 2018). In a recent work, Wang et al. (2020) provides a similar finite-sample result for change-in-mean problem and shows that the localisation rate in (3.5) is minimax optimal. In comparison, the extra condition on  $\Delta_T$  and the inclusion of  $m^*$  in the convergence rate in our Theorem 3.1 is due to the fact that we specify the value of penalty to be  $2 + \epsilon$ , other than a lack-of-track large constant, times  $\log T$ . The following corollary gives consistency for a specific asymptotic regime.

**Corollary 3.1.** *If  $\beta = (2 + \epsilon) \log T$ , where  $\epsilon > 0$  is an arbitrarily small constant, assume that  $m^* = o(\log T)$ ,  $\delta_T \Delta_T^2 \geq c_1 \log T$  and  $\Delta_T^2 \geq c_2 T^{-1/(4m^*+3)} \log T$ , we have*

$$\mathbb{P}\left(\hat{m} = m^*, \max_{j=1, \dots, m^*} |\hat{\tau}_j - \tau_j| \Delta_j^2 \leq c_1 \log T\right) \rightarrow 1$$

as  $T \rightarrow \infty$ , where  $c_1$  and  $c_2$  are absolute constants that only depend on  $\epsilon$ .

For the standard in-fill asymptotic regime with a fixed number of true changes  $m^*$  and constant size of minimum jump in the signals, i.e.,  $\Delta_T = O(1)$ , from above corollary we can obtain a changepoint localisation rate of  $O_p(\log T)$ , which is the same order as in the classic results (Yao, 1988). Note that this order on accuracy of the changepoint locations could be further improved to basically  $O_p(1)$  using the argument in Yao and Au (1989). Also, similar improved accuracy of consistent estimates is given in Boysen et al. (2007) and Boysen et al. (2009a) for piecewise constant least squares regression and general class of functions based on the solution to an  $L_0$  least squares (Potts functional), under the assumption of the piecewise-constant mean function, fixed number of change points  $m$ , constant size of the minimum jump in the signal  $\Delta_T$  and minimum segment length  $\delta_T$  is of order  $T$ .

In this paper, we have a slight generalisation of these existing consistency results as we allow all the parameters  $m$ ,  $\delta_T$  and  $\Delta_T$  to change with the sample size  $T$ . Our localisation rate also matches the minimax optimal rate. Note that some extra restrictions on  $m^*$  (only slowly diverging with  $T$ ) and  $\Delta_T$  (lower bounded) is posed in our results due to the use of the argument of constructing global consistency from local region conditions, where in Wang et al. (2020) and Verzelen et al. (2020) they are actually not required.

Finally, we remark that similar results to Theorem 3.1 and Corollary 3.1 could be attained if we relax the Gaussianity assumption on noise to having an exponential or lighter tail. This can be done by adapting the chi-square bound in Lemma B.1 to sub-gamma concentration inequalities, see similar argument in Baranowski et al. (2016). By taking a robust cost function and penalisation term  $\beta$ , the general framework should also work for the mean plus heavy-tailed noises model as considered in Fearnhead and Rigall (2019).

### 3.2. Change-in-slope problem

For the change-in-slope application, we have the following decomposition of observations:

$$x_t = f_t + \varepsilon_t, \quad 1 \leq t \leq T \quad (3.6)$$

where  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$  are i.i.d Gaussian noises, and  $f_t$  denote the piecewise linear mean signals, that is, for  $j = 1, \dots, m^* + 1$ :

$$f_t = \theta_{j-1}^* + \frac{\theta_j^* - \theta_{j-1}^*}{\tau_j^* - \tau_{j-1}^*} (t - \tau_{j-1}^*), \quad \tau_{j-1}^* + 1 \leq t \leq \tau_j^*, \quad (3.7)$$

In the above parameterisation,  $\theta_{0:m^*+1}^*$  are values of the linear function at the changes  $\tau_{0:m^*+1}$ . As a consequence, this means we directly introduce the continuity constraint that enforces the value at the end of one segment to be equal to the value at the start of next segment.

We take the negative log-likelihood as the cost function, thus to fit a set of points  $x_{s:e}$  such that  $\tau_{j-1} + 1 \leq s < e \leq \tau_j$ , with the same bivariate structure parameter  $\boldsymbol{\theta} = (\theta^{(1)}, \theta^{(2)})$ , the cost function is as follows:

$$\mathcal{C}(x, s, e; \boldsymbol{\theta}) = \sum_{t=s}^e \frac{1}{\sigma^2} \left[ x_t - \theta^{(2)} - \frac{\theta^{(1)} - \theta^{(2)}}{\tau_j - \tau_{j-1}} (t - \tau_{j-1}) \right]^2. \quad (3.8)$$

Due to requirement of continuity at the changepoint, we need a constrained minimisation on the corresponding overall cost function for fitting a set of changes  $\tau_{1:m}$ , i.e.,  $\theta_j^{(2)} = \theta_{j-1}^{(1)}$  for all  $j = 1, \dots, m$  and  $\boldsymbol{\theta}_j = (\theta_j^{(1)}, \theta_j^{(2)})$ . Therefore the overall cost function takes the following form:

$$\begin{aligned} \mathcal{L}(x_{1:T}, \tau_{1:m}) &= \min_{\boldsymbol{\theta}_{1:m+1}, \theta_j^{(2)} = \theta_{j+1}^{(1)}} \sum_{j=1}^{m+1} \mathcal{C}(x_{\tau_{j-1}+1:\tau_j}; \boldsymbol{\theta}_j) \\ &= \min_{\boldsymbol{\theta}_{0:m+1}} \sum_{j=1}^{m+1} \sum_{t=\tau_{j-1}+1}^{\tau_j} \frac{1}{\sigma^2} \left[ x_t - \theta_{j-1} - \frac{\theta_j - \theta_{j-1}}{\tau_j - \tau_{j-1}} (t - \tau_{j-1}) \right]^2. \end{aligned} \quad (3.9)$$

In this application, we define the size of change at the  $j$ -th changepoint,  $\tau_j^*$ , as

$$\Delta_j = \text{dist}(\boldsymbol{\theta}_{j+1}^*, \boldsymbol{\theta}_j^*) = \left| \frac{\theta_{j+1}^* - \theta_j^*}{\tau_{j+1}^* - \tau_j^*} - \frac{\theta_j^* - \theta_{j-1}^*}{\tau_j^* - \tau_{j-1}^*} \right|, \quad (3.10)$$

which is the absolute difference of slopes in two consecutive segments. Again, the nuisance noise variance  $\sigma^2$ , if unknown, can be robustly estimated ([Fearnhead et al., 2018](#)), for example we can set:

$$\hat{\sigma} = \frac{\text{Median}\{|x_1 - 2x_2 + x_3|, \dots, |x_{T-2} - 2x_{T-1} + x_T|\}}{\sqrt{6}\Phi^{-1}(3/4)}.$$

In order to study the property of the changepoint detection for this application, we first need to verify Conditions 2.1 and 2.2 are satisfied, which is shown in the following propositions.

**Proposition 3.3.** *Consider following choice of  $\gamma_n^{(1)}$  and  $\gamma_n^{(2)}$ :*

$$\begin{aligned}\gamma_n^{(1)} &= \max \left\{ (2 + \epsilon) \log n, 2 \log n + \kappa_{2,1} \sqrt{\log n}, 2 \log n + \kappa_{2,2} m^* \right\}, \\ \gamma_n^{(2)} &= \max \left\{ (3 + \epsilon) \log(2n), 2 \log(2n) + \kappa_{2,3} \log(\log(2n)), 2 \log(2n) + \kappa_{2,4} m^* \right\},\end{aligned}$$

for some large enough constants  $\kappa_{2,1}, \kappa_{2,2}, \kappa_{2,3}, \kappa_{2,4}$ , and  $\epsilon$  is an positive constant that can be arbitrarily small. Moreover, let

$$a(\gamma, n) = \frac{\gamma - 2 \log n}{6} \quad \text{and} \quad b(\gamma, n) = \frac{\gamma - 2 \log n}{6(2m^* + 1)}. \quad (3.11)$$

We consequently find that Conditions 2.1 is satisfied by:

$$\begin{aligned}p_1(\gamma, n) &= 2 \exp \left( -\frac{\gamma - 2 \log n}{6} \right), \quad p_2(\gamma, n) = \exp \left( -\frac{\gamma - 2 \log n}{24(2m^* + 1)} \right), \\ p_3(\gamma, n) &= \frac{9}{4} \exp \left( -\frac{\gamma - 3 \log(2n)}{3} \right), \quad p_4(\gamma, n) = \exp \left( -\frac{\gamma - 2 \log(2n)}{24(2m^* + 1)} \right).\end{aligned}$$

**Proposition 3.4.** *Let  $\mathcal{S}(\Delta, n) = n^3 \Delta^2 / 25$ , then Condition 2.2 is satisfied if  $\mathcal{S}(\Delta, n) / 4 \geq z \geq 8$  and  $n \geq 2$ , with  $p_5(\mathcal{S}(\Delta, n), z) = 2 \exp(-z/20)$ .*

Propositions 3.3 and 3.4 have strong similarity with Propositions 3.1 and 3.2 for the change-in-mean problem, where again the constants are not optimised for simplicity purposes. Based on Theorem 2.1 we can obtain the following theorem.

**Theorem 3.2.** *If  $\beta = (2 + \epsilon) \log T$ , where  $\epsilon > 0$  is an arbitrarily small constant; and if*

$$\delta_T^3 \Delta_T^2 \geq (200 + 350\epsilon/3) \log T \quad \text{and} \quad \Delta_T^2 \geq \frac{(1600 + 2000\epsilon/3) \log T}{T^2}, \quad (3.12)$$

then for large enough  $T$ , with probability at least  $1 - (33m^*/4 + 3)T^{-\epsilon/(48m^* + 24)}$ ,

$$\hat{m} = m^*, \quad \max_{j=1, \dots, m^*} |\hat{\tau}_j - \tau_j|^3 \Delta_j^2 \leq (200 + 350\epsilon/3) \log T. \quad (3.13)$$

In terms of asymptotics, we have the following corollary.

**Corollary 3.2.** *If  $\beta = (2 + \epsilon) \log T$ , where  $\epsilon > 0$  is an arbitrarily small constant, assume that  $m^* = o(\log T)$ ,  $\delta_T^3 \Delta_T^2 \geq c_3 \log T$  and  $\Delta_T^2 \geq c_4 \log T / T^2$ , we have*

$$\mathbb{P} \left( \hat{m} = m^*, \quad \max_{j=1, \dots, m^*} |\hat{\tau}_j - \tau_j|^3 \Delta_j^2 \leq c_3 \log T \right) \rightarrow 1$$

as  $T \rightarrow \infty$ , where  $c_3$  and  $c_4$  are a absolute constants that only depend on  $\epsilon$ .

For the standard in-fill asymptotic regime with a fixed number of true changes  $m^*$ , we would have  $\Delta_T = O(T^{-1})$ . In such case we get a bound on the error of location estimates that is just a logarithmic factor worse than the minimax rate of  $T^{2/3}$  (Raimondo, 1998). The results also holds when  $m^*$  is diverging at a rate slower than  $\log T$ .

We emphasize that this result is stronger than previous consistency results of the change-in-slope model derived in Baranowski et al. (2016) and Fearnhead et al. (2018), as it specifies the value of the penalty  $\beta$  that ensures consistency. This is a non-trivial technical improvement as shown in the Appendix E. In this case to get tighter results for the choice of penalty, we need to take account of the positive dependency in the reduction of cost of similar segmentations. We need these tighter bounds because for the in-fill asymptotic regime the accuracy of estimating a change-in-slope is polynomial in  $T$  rather than logarithmic in  $T$ . This accuracy impacts, and increases, the number of possible segmentations we can fit to data in our local region that have one changepoint in the middle location. We believed similar arguments can be applied to refine the value of penalty in other changepoint detection problems.

### 3.3. Changepoint in spike and exponential decay problem

In this application, the observations  $x_{1:T}$  have an underlying decomposition  $x_t = c_t + \varepsilon_t$ , where  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$  are i.i.d. Gaussian innovations, and the mean function,  $c_t$ , follows a piecewise spike and exponential decay model. That is, for  $j = 1, \dots, m^* + 1$ ,

$$c_t = \theta_j^* \alpha^{t - \tau_{j-1}^* - 1}, \quad \tau_{j-1}^* + 1 \leq t \leq \tau_j^*,$$

where  $0 < \alpha < 1$  is the decay rate. When  $\alpha = 1$  this reduces to the change-in-mean problem in Section 3.1.

We take the square error loss as the cost function, so to fit a set of points  $x_{s:e}$  such that  $\tau_{j-1} + 1 \leq s < e \leq \tau_j$  with the same parameter  $\theta_j$ , we have the following cost:

$$\mathcal{C}(x, s, e; \theta_j) = \sum_{t=s}^e \frac{1}{\sigma^2} (x_t - \theta_j \alpha^{t - \tau_{j-1} - 1})^2. \quad (3.14)$$

Therefore, the corresponding cost function is

$$\mathcal{L}(x_{1:T}, \tau_{1:m}) = \min_{\theta_{j+1} \neq \theta_j \alpha^{\tau_j - \tau_{j-1}}} \sum_{j=1}^{m+1} \sum_{t=\tau_{j-1}+1}^{\tau_j} \frac{1}{\sigma^2} (x_t - \theta_j \alpha^{t - \tau_{j-1} - 1})^2, \quad (3.15)$$

where  $\tau_0 = 0$  and  $\tau_{m+1} = T$ . In this application, we minimises over  $m$ ,  $\tau_{1:m}$  and  $\theta_{1:m+1}$  in the penalised cost  $\mathcal{L}(x_{1:T}; \tau_{1:m}) + m\beta$  to estimate the number of changepoints and their positions.

We define the size of  $j$ -th changepoint at  $\tau_j^*$  as

$$\Delta_j = \text{dist}(\theta_{j+1}^*, \theta_j^*) = |\theta_{j+1}^* - \theta_j^* \alpha^{\tau_j^* - \tau_{j-1}^* - 1}|, \quad (3.16)$$

which is the size of the jump in the signal from the end of  $j$ -th segment to the beginning of the  $(j + 1)$ -th segment. Note that solving the problem (3.15) requires knowledge of the decay rate  $\alpha$  and the noise variance  $\sigma^2$ . For methods to estimate these see Jewell et al. (2019) and Jewell and Witten (2018).

In the following propositions, we show that local Conditions 2.1 and 2.2 are satisfied for this application.

**Proposition 3.5.** *Let*

$$\begin{aligned}\gamma_n^{(1)} &= \max \left\{ (2 + \epsilon) \log n, 2 \log n + \kappa_{3,1} \sqrt{\log n}, 2 \log n + \kappa_{3,2} m^* \right\}, \\ \gamma_n^{(2)} &= \max \left\{ \kappa_{3,3} \log(2n), 2 \log(2n) + \kappa_{3,4} m^* \right\},\end{aligned}$$

for some large enough constants  $\kappa_{3,1}, \kappa_{3,3}, \kappa_{3,3}, \kappa_{3,4}$ , and  $\epsilon$  is an positive constant that can be arbitrarily small. Moreover, let  $a(\gamma, n) = \frac{\gamma - 2 \log n}{4}$  and  $b(\gamma, n) = \frac{\gamma - 2 \log n}{4(2m^* + 1)}$ . We have that Condition 2.1 is satisfied with

$$\begin{aligned}p_1(\gamma, n) &= 2 \exp \left( -\frac{\gamma - 2 \log n}{4} \right), & p_2(\gamma, n) &= \exp \left( -\frac{\gamma - 2 \log n}{16(2m^* + 1)} \right), \\ p_3(\gamma, n) &= \exp \left( -\frac{\gamma - 8 \log(2n)}{4} \right), & p_4(\gamma, n) &= \exp \left( -\frac{\gamma - (8m^* + 6) \log(2n)}{16(2m^* + 1)} \right).\end{aligned}$$

**Proposition 3.6.** *Let  $\mathcal{S}(\Delta, n) = \frac{\Delta^2}{(1 - \alpha^{2n})(1 - \alpha^2)}$ , then Condition 2.2 is satisfied if  $\mathcal{S}(\Delta, n)/4 \geq z \geq 5$ , with*

$$p_5 \left( \mathcal{S}(\Delta, n), z \right) = 2 \exp \left( -\frac{z}{20} \right).$$

Propositions 3.5 and 3.6 is quite similar to Propositions 3.1 and 3.2 for the change-in-mean problem. The only difference lies in the form of signal strength. Based on Theorem 2.1 we can obtain the following theorem.

**Theorem 3.3.** *If  $\beta = (2 + \epsilon) \log T$ , where  $\epsilon > 0$  is an arbitrarily small constant; and if we have*

$$\frac{\Delta_T^2}{(1 - \alpha^{2\delta_T})(1 - \alpha^2)} \geq (8 + 5\epsilon) \log T \quad (3.17)$$

and

$$\log_\alpha \left( 1 - \frac{\Delta_T^2}{(1 - \alpha^2)(8 + 5\epsilon) \log T} \right) \leq T^{2/(8m^* + 6 + \epsilon)}, \quad (3.18)$$

then for large enough  $T$ , with probability at least  $1 - (7m^* + 3)T^{-\epsilon/(32m^* + 16)}$

$$\hat{m} = m, \quad \min_{j=1, \dots, m^*} \frac{\Delta_j^2}{(1 - \alpha^2) \left( 1 - \alpha^{2|\hat{\tau}_j - \tau_j^*|} \right)} \geq (8 + 5\epsilon) \log T. \quad (3.19)$$

In terms of asymptotic regime, note that the signal strength in this application  $\mathcal{S}(\Delta, n) = O(\Delta^2)$  will become finite if  $\alpha$  is fixed. A more natural asymptotic regime is to assume that we are able to obtain data at a higher frequency, which would correspond to  $\alpha \rightarrow 1$  as  $T \rightarrow \infty$ .

**Corollary 3.3.** *If  $\beta = (2+\epsilon) \log T$ , where  $\epsilon > 0$  is an arbitrarily small constant.*

*Assume that  $c_5 \log T \leq \frac{\Delta_T^2}{(1-\alpha^{2\delta_T})(1-\alpha^2)}$ , and there exists universal positive constant  $D < 1$  such that  $\log_\alpha D \leq T^{2/(8m^*+6+\epsilon)}$ . Let  $m^* = o(\log T)$  and  $1-\alpha \geq c_6(\Delta_T^2/\log T)$ , we have*

$$\mathbb{P}\left(\hat{m} = m^*, \min_{j=1, \dots, m^*} \frac{\Delta_j^2}{(1-\alpha^2)(1-\alpha^{2|\hat{\tau}_j - \tau_j^*|})} \geq c_5 \log T\right) \rightarrow 1$$

as  $T \rightarrow \infty$ , where  $c_5$  and  $c_6$  are absolute constants only depends on  $\epsilon$ .

For the standard in-fill asymptotic regime with a fixed number of true changes  $m^*$ , we would have  $\Delta_T = O(1)$ . In such case if  $\alpha = \exp(-c/T)$  for some positive constant  $c$ , we get a bound on the error of changepoint location estimate,  $\max_j |\hat{\tau}_j - \tau_j^*|$ , of order not greater than  $\log T$  with probability going to 1.

#### 4. Discussion

In this paper our key motivation is to show that for a class of changepoint methods, statistical properties for detecting multiple changepoints using penalised cost approaches can be derived from the behaviour of the method when analysing data with either no changepoint or a single changepoint, i.e., on the so-called local region. Therefore, we propose a general framework to prove consistency results of a broad class of penalised cost approaches for detecting multiple changepoints, only assuming the properties on the local regions.

These properties on the local regions are often easier to verify; for example the results for the three applications we considered in Section 3 almost all follow from bounds on chi-squared random variables and the use of a Bonferroni correction. The one exception is for the change-in-slope model where to get sharper results on the choice of penalty, we need to consider the dependency between the cost of fitting similar segmentations. The related techniques are presented in Appendix E, which we believe to be of independent interest.

Our focus on the theoretical aspects of different types of change, such as of change-in-slope and spike plus exponential decay problem, without carrying out numerical studies is due to the fact that fast computational algorithms based on dynamic programming and (functional) pruning have been well-developed, and their numerical performance and applications to real-world data have been extensively analysed in the literature. Using the technique of local region conditions, this paper closes the theoretical gap in these applications. More importantly, the technique provides a powerful tool to derive consistency results for many other multiple changepoint applications.

In addition to the applications considered in Section 3, the general framework can be readily applied to more complicated cases, for example, the signal within the segments is non-linear (Yu et al., 2022), structure changes in the moments or quantiles (Fisch et al., 2022; Baranowski et al., 2016), and other structure changes such as autocovariance in time series. Moreover, as mentioned earlier the gaussianity condition on the noise can be relaxed as well (Fearhead and Rigaiil, 2019).

The theoretical results given in this paper could be further improved in two regards. First the constants we obtained have not been optimised. More importantly, as mentioned earlier the results on accuracy of the changepoint locations could be improved using arguments similar to Yao and Au (1989) and . The idea is to leverage the result that shows we accurately estimate the location of  $\tau_{j-1}$  and  $\tau_{j+1}$  and then show that the error in estimating  $\tau_j$  converges in distribution to the accuracy of estimating the location of a single changepoint from data in the region between  $\tau_{j-1}$  and  $\tau_{j+1}$ .

## Appendix A: Proofs of Section 2

*Proof of Lemma 2.1.* It is trivial to verify (2.4) as once we fix the segment parameters the cost,  $\mathcal{L}^*(\cdot)$ , is additive over data-points.

Property (2.3) follows because we are minimising the cost on the left-hand side over a more constrained space for the segment parameters. If  $r$  is not a changepoint location, then

$$\begin{aligned}
& \mathcal{L}(x_{s:r}; \tau_{u:v}) + \mathcal{L}(x_{r+1:e}; \tau_{u:v}) \\
&= \min_{\boldsymbol{\theta}_{u:k-1}, \boldsymbol{\theta}_k^{(1)}} \left( \sum_{j=u}^{k-1} \mathcal{C}(x_{\tau_{j-1}+1:\tau_j}; \boldsymbol{\theta}_j) + \mathcal{C}(x_{\tau_{k-1}+1:r}; \boldsymbol{\theta}_k^{(1)}) \right) \\
& \quad + \min_{\boldsymbol{\theta}_k^{(2)}, \boldsymbol{\theta}_{k+1:v}} \left( \mathcal{C}(x_{r+1:\tau_k}; \boldsymbol{\theta}_k^{(2)}) + \sum_{j=k+1}^v \mathcal{C}(x_{\tau_{j-1}+1:\tau_j}; \boldsymbol{\theta}_j) \right) \\
&\leq \min_{\boldsymbol{\theta}_{u:v}} \left( \sum_{j=u}^v \mathcal{C}(x_{\tau_{j-1}+1:\tau_j}; \boldsymbol{\theta}_j) \right) \\
&= \mathcal{L}(x_{s:e}; \tau_{u:v}),
\end{aligned}$$

where all minimisations include any constraints on segment parameters in neighbouring segments. The inequality comes from the fact that for  $\mathcal{L}(x_{s:r}; \tau_{u:v}) + \mathcal{L}(x_{r+1:e}; \tau_{u:v})$  we have no constraint between  $\boldsymbol{\theta}_k^{(1)}$  and  $\boldsymbol{\theta}_k^{(2)}$ , but for  $\mathcal{L}(x_{s:e}; \tau_{u:v})$  we have  $\boldsymbol{\theta}_k^{(1)} = \boldsymbol{\theta}_k^{(2)} := \boldsymbol{\theta}_k$ . A similar argument applies if  $r$  is a changepoint, as  $\mathcal{L}(x_{s:r}; \tau_{u:v}) + \mathcal{L}(x_{r+1:e}; \tau_{u:v})$  will not apply any constraint between the segment parameters for the segments immediately before and after  $r$ .  $\square$

*Proof of Theorem 2.1.* Consider a split of the data as (2.5) for the specified  $n_{1:m^*}$ . Let  $l_j$  be the number of data in region  $\mathbf{S}_j$ , therefore  $l_{2k+1} = \delta_k - n_k - n_{k+1}$  and

$l_{2k} = 2n_k$ . We define an event,  $E_1$ , based on this split, such that the following holds jointly for regions  $\mathbf{S}_{1:2m^*+1}$ :  
if  $j$  is odd,

$$\min_{k \geq 1, \tau_{1:k}} \{\mathcal{L}(\mathbf{S}_j; \tau_{1:k}) + k\beta\} - \mathcal{L}^*(\mathbf{S}_j) \geq a(\beta, l_j), \text{ and}$$

$$\mathcal{L}^*(\mathbf{S}_j) - \mathcal{L}(\mathbf{S}_j; \emptyset) \leq b(\beta, l_j);$$

if  $j$  is even,

$$\min_{k \geq 2, \tau_{1:k}} \{\mathcal{L}(\mathbf{S}_j; \tau_{1:k}) + (k-1)\beta\} - \mathcal{L}^*(\mathbf{S}_j) \geq a(\beta, l_j),$$

$$\mathcal{L}^*(\mathbf{S}_j) - \min_{x_\tau \in \mathbf{S}_j} \mathcal{L}(\mathbf{S}_j; \tau) \leq b(\beta, l_j), \text{ and}$$

$$\mathcal{L}(\mathbf{S}_j, \emptyset) - \mathcal{L}^*(\mathbf{S}_j) \geq a(\beta, l_j) + \beta;$$

In what follows, we will condition on this event holding. Since  $\beta \geq \max_k \{\gamma_{l_{2k+1}}^{(1)}, \gamma_{l_{2k}}^{(2)}\}$ , due to the Conditions 2.1 and 2.2, by a simple union bound we have the probability of this is lower bounded by

$$1 - (m^* + 1)p_1(\beta, T) - (m^* + 1)p_2(\beta, T) - m^*p_3(\beta, \max_j n_j) \\ - m^*p_4(\beta, \max_j n_j) - m^*p_5(\mathcal{S}, \beta + a(\beta, T)).$$

Now for any segmentation  $\tau_{1:m}$  we can compare the penalised cost of that segmentation with the penalised cost of the true segmentation. Trivially, we have

$$\{\mathcal{L}(x_{1:T}; \tau_{1:m}^*) + \beta m^*\} - \{\mathcal{L}(x_{1:T}; \tau_{1:m}) + \beta m\} \\ \leq \mathcal{L}^*(x_{1:T}) - \mathcal{L}(x_{1:T}; \tau_{1:m}) + (m^* - m)\beta,$$

as  $\mathcal{L}(x_{1:T}; \tau_{1:m}^*)$  minimises over the segment parameters whereas  $\mathcal{L}^*(x_{1:T})$  fixes them to their true values. Using Lemma 2.1 we have

$$\mathcal{L}^*(x_{1:T}) - \mathcal{L}(x_{1:T}; \tau_{1:m}) \leq \sum_{j=1}^{2m^*+1} \{\mathcal{L}^*(\mathbf{S}_j) - \mathcal{L}(\mathbf{S}_j; \tau_{1:m})\}.$$

We can partition the set of regions into  $\mathcal{A}$ ,  $\mathcal{B}$  and  $\mathcal{D}$ , which are defined as the regions where the putative segmentation,  $\tau_{1:m}$ , fits too many changes, too few changes, or the correct number of changes. Let  $k_j$  and  $k_j^*$ , respectively, denote the number of changepoints in region  $j$  in the putative and the true segmentation. Thus

$$\{\mathcal{L}(x_{1:T}; \tau_{1:m}^*) + \beta m^*\} - \{\mathcal{L}(x_{1:T}; \tau_{1:m}) + \beta m\} \\ \leq \sum_{j \in \mathcal{A}} \{\mathcal{L}^*(\mathbf{S}_j) - \mathcal{L}(\mathbf{S}_j; \tau_{1:m}) + (k_j^* - k_j)\beta\} \\ + \sum_{j \in \mathcal{B}} \{\mathcal{L}^*(\mathbf{S}_j) - \mathcal{L}(\mathbf{S}_j; \tau_{1:m}) + (k_j^* - k_j)\beta\} \\ + \sum_{j \in \mathcal{D}} \{\mathcal{L}^*(\mathbf{S}_j) - \mathcal{L}(\mathbf{S}_j; \tau_{1:m}) + (k_j^* - k_j)\beta\}$$

Conditional on our event  $E_1$  holding, terms in the first two sums can be bounded above by  $-a(\beta, T)$ , while terms in the final sum can be bounded above by  $b(\beta, T)$ . If we let  $m_D$  denote the number of terms in the final sum we have

$$\begin{aligned} & \{\mathcal{L}(x_{1:T}; \tau_{1:m^*}^*) + \beta m^*\} - \{\mathcal{L}(x_{1:T}; \tau_{1:m}) + \beta m\} \\ & \leq m_D b(\beta, T) - (2m^* + 1 - m_D) a(\beta, T). \end{aligned}$$

For any segmentation  $\tau_{1:m}$  with which  $\mathcal{E}$  does not hold when  $\hat{m} = m$  and  $\hat{\tau}_{1:\hat{m}} = \tau_{1:m}$  we have  $m_D \leq 2m^*$ . Thus as  $a(\beta, T) > 2m^* b(\beta, T)$  the right-hand side of this equality will be strictly less than 0. Hence, conditional on event  $E_1$  holding, no such segmentation can minimise our penalised cost.  $\square$

## Appendix B: Proofs of Section 3.1

Lemma B.1 below is a direct adaptation of Lemma 1 in [Laurent and Massart \(2000\)](#) and Lemma 8.1 in [Birgé \(2001\)](#). We will use it repeatedly.

**Lemma B.1.** *Let  $\chi_k^2$  be a central chi-square statistic with  $k$  degrees of freedom and  $\chi_k^2(\nu)$  a chi-square statistic with  $k$  degrees of freedom and non-centrality parameter  $\nu$ . For any  $x > k$  and  $y < k + \nu$ , we have*

$$\mathbb{P}(\chi_k^2 \geq x) \leq \exp\left(-\frac{x - \sqrt{k(2x - k)}}{2}\right), \quad (\text{B.1})$$

$$\mathbb{P}(\chi_k^2(\nu) \leq y) \leq \exp\left(-\frac{(k + \nu - y)^2}{4k + 8\nu}\right). \quad (\text{B.2})$$

### B.1. Proof of Propositions 3.1 and 3.2

**Lemma B.2.** *For any  $t$  and any model  $\mathcal{N}_n(t)$ , if  $\mathbf{S} = x_{t+1:t+n}$  is a region that contains no true changepoint, then for any  $\gamma \geq \gamma_n^{(1)}$ , where*

$$\gamma_n^{(1)} = \max\left\{(2 + \epsilon) \log n, 2 \log n + 8\sqrt{16 + 2 \log n} + 32, 2 \log n + 32(2m^* + 1)\right\} \quad (\text{B.3})$$

with  $\epsilon > 0$  is a constant, we have

$$\mathbb{P}\left(\min_{1 \leq k, \tau_{1:k}} \{\mathcal{L}(\mathbf{S}; \tau_{1:k}) + k\gamma\} - \mathcal{L}^*(\mathbf{S}) \leq \frac{\gamma - 2 \log n}{4}\right) < 2 \exp\left(-\frac{\gamma - 2 \log n}{4}\right),$$

and

$$\mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \emptyset) \geq \frac{\gamma - 2 \log n}{4(2m^* + 1)}\right) \leq \exp\left(-\frac{\gamma - 2 \log n}{16(2m^* + 1)}\right).$$

*Proof.* For any  $t$  and model  $\mathcal{N}_n(t)$  such that  $\mathbf{S} = x_{t+1:t+n}$  is a region that contains no true changepoint, since  $\varepsilon_{t+1:t+n}$  are i.i.d, it is easy to derive that  $\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k}) \sim \chi_{k+1}^2$  and  $\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \emptyset) \sim \chi_1^2$ .

Letting  $a(\gamma, n) = (\gamma - 2 \log n)/4$ , we have

$$\begin{aligned} & \mathbb{P} \left( \min_{k, \tau_{1:k}} \{ \mathcal{L}(\mathbf{S}; \tau_{1:k}) + k\gamma - \mathcal{L}^*(\mathbf{S}) \} \leq a(\gamma, n) \right) \\ & \leq \sum_{k=1}^n \mathbb{P} \left( \max_{\tau_{1:k}} [ \mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k}) ] \geq k\gamma - a(\gamma, n) \right) \\ & \leq \sum_{k=1}^{\infty} \binom{n}{k} \mathbb{P} (\chi_{k+1}^2 \geq k\gamma - a(\gamma, n)) \end{aligned}$$

Combined with (B.1) in Lemma B.1,

$$\begin{aligned} & \binom{n}{k} \mathbb{P} (\chi_{k+1}^2 \geq k\gamma - a(\gamma, n)) \\ & < \frac{n^k}{k!} \exp \left( - \frac{k\gamma - a(\gamma, n) - \sqrt{(k+1)(2k\gamma - 2a(\gamma, n) - k - 1)}}{2} \right) \\ & < \frac{1}{k!} \exp \left( - \frac{(k-1/4)(\gamma - 2 \log n) - \sqrt{2k(k+1)\gamma}}{2} \right) \end{aligned}$$

As long as  $\gamma \geq 2 \log n + 8\sqrt{16 + 2 \log n} + 32$ , we have

$$\sqrt{2k(k+1)\gamma} \leq \frac{k(\gamma - 2 \log n)}{4}. \quad (\text{B.4})$$

Hence

$$\begin{aligned} & \mathbb{P} \left( \min_{k, \tau_{1:k}} \{ \mathcal{L}(\mathbf{S}; \tau_{1:k}) + k\gamma - \mathcal{L}^*(\mathbf{S}) \} \leq a(\gamma, n) \right) \\ & < \sum_{k=1}^{\infty} \frac{1}{k!} \exp \left( - \frac{(3k/4 - 1/4)(\gamma - 2 \log n)}{2} \right) \\ & < \sum_{k=1}^{\infty} \frac{1}{2^{k-1}} \exp \left( - \frac{\gamma - 2 \log n}{4} \right) \\ & < 2 \exp \left( - \frac{\gamma - 2 \log n}{4} \right). \end{aligned}$$

Next we prove the second inequality. For any specified  $m^*$ , define  $b(\gamma, n) = (\gamma - 2 \log n)/(8m^* + 4)$ . Since  $\gamma > 2 \log n + 32(2m^* + 1)$ , we have  $b(\gamma, n) \geq 8$ ,

which leads to  $\sqrt{2b(\gamma, n)} \leq b(\gamma, n)/2$ . Applying (B.1) in Lemma B.1, we obtain

$$\begin{aligned} \mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \emptyset) \geq b(\gamma, n)\right) &= \mathbb{P}\left(\chi_1^2 \geq b(\gamma, n)\right) \\ &< \exp\left(-\frac{b(\gamma, n) - \sqrt{2b(\gamma, n)}}{2}\right) \\ &\leq \exp\left(-\frac{b(\gamma, n)}{4}\right) \\ &= \exp\left(-\frac{\gamma - 2\log n}{16(2m^* + 1)}\right). \end{aligned}$$

□

**Lemma B.3.** For any  $t$  and any model, if  $\mathbf{S} = x_{t+1:t+2n}$  is a region that contains a single changepoint at  $\tau^* = t + n$ , for any  $\gamma \geq \gamma_n^{(2)}$ , where

$$\gamma_n^{(2)} = \max\{(8m^* + 6 + \epsilon)\log(2n), 2\log(2n) + 64(2m^* + 1)\}, \quad (\text{B.5})$$

with  $\epsilon > 0$  is a constant, we have

$$\mathbb{P}\left(\min_{k \geq 2, \tau_{1:k}} \{\mathcal{L}(\mathbf{S}; \tau_{1:k}) + (k-1)\gamma\} - \mathcal{L}^*(\mathbf{S}) \leq \frac{\gamma - 2\log(2n)}{4}\right) < \exp\left(-\frac{\gamma - 8\log(2n)}{4}\right),$$

and

$$\mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_1} \mathcal{L}(\mathbf{S}; \tau_1) \geq \frac{\gamma - 2\log(2n)}{4(2m^* + 1)}\right) \leq \exp\left(-\frac{\gamma - (8m^* + 6)\log(2n)}{16(2m^* + 1)}\right).$$

*Proof.* Note that for any  $\tau_{1:k}$  on  $\mathbf{S}$ ,

$$\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k}) \leq \mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k}, \tau^*) \sim \chi_{k+2}^2, \text{ and}$$

$$\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_1) \sim \chi_2^2.$$

$$\text{Let } a(\gamma, 2n) = \frac{\gamma - 2\log(2n)}{4} \leq \gamma/4 \text{ and } b(\gamma, n) = (\gamma - 2\log n)/(8m^* + 4).$$

Since  $\gamma \geq 64(2m^* + 1) + 2\log(2n)$ , we obtain  $b(\gamma, 2n) \geq 16$  and  $\sqrt{2(k-1)(k+2)}\gamma \leq (k-1)\gamma/4$ .

Similar to the proof of Lemma B.2, by Bonferroni correction, we have the probability  $\mathbb{P}(\min_{k \geq 2, \tau_{1:k}} \{\mathcal{L}(\mathbf{S}; \tau_{1:k}) + (k-1)\gamma\} - \mathcal{L}^*(\mathbf{S}) \leq a(\gamma, 2n))$  is upper bounded by  $\sum_{k=2}^{\infty} \binom{2n}{k} \mathbb{P}(\chi_{k+2}^2 \geq (k-1)\gamma - a(\gamma, 2n))$

Applying again (B.1) in Lemma B.1,

$$\begin{aligned}
& \binom{2n}{k} \mathbb{P}(\chi_{k+2}^2 \geq (k-1)\gamma - a(\gamma, 2n)) \\
& < \frac{(2n)^k}{k!} \exp\left(-\frac{(k-1)\gamma - a(\gamma, 2n) - \sqrt{2(k+2)(k-1)\gamma}}{2}\right) \\
& \leq \frac{(2n)^k}{k!} \exp\left(-\frac{(k-1)\gamma}{4}\right) \\
& \leq \frac{1}{k!} \exp\left(-\frac{(k-1)\gamma - 4k \log(2n)}{4}\right) \\
& \leq \frac{1}{2^{k-1}} \exp\left(-\frac{(k-1)(\gamma - 8 \log(2n))}{4}\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
& \mathbb{P}\left(\min_{k \geq 2, \tau_{1:k}} \{\mathcal{L}(\mathbf{S}; \tau_{1:k}) + (k-1)\gamma\} - \mathcal{L}^*(\mathbf{S}) \leq a(\gamma, 2n)\right) \\
& < \sum_{k=2}^{\infty} \frac{1}{2^{k-1}} \exp\left(-\frac{(k-1)(\gamma - 8 \log(2n))}{4}\right) = \exp\left(-\frac{\gamma - 8 \log(2n)}{4}\right).
\end{aligned}$$

Moreover, since  $b(\gamma, n) \geq 16$ , which leads to  $2\sqrt{b(\gamma, n)} \leq b(\gamma, n)/2$ , therefore

$$\begin{aligned}
\mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_1} \mathcal{L}(\mathbf{S}; \tau_1) \geq b(\gamma, 2n)\right) & \leq (2n-1) \mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_1) \geq b(\gamma, 2n)\right) \\
& = (2n-1) \mathbb{P}(\chi_2^2 \geq b(\gamma, 2n)) \\
& < 2n \exp\left(-\frac{b(\gamma, 2n) - 2\sqrt{b(\gamma, 2n)}}{2}\right) \\
& \leq \exp\left(-\frac{b(\gamma, 2n) - 4 \log(2n)}{4}\right) \\
& = \exp\left(-\frac{\gamma - (8m^* + 6) \log(2n)}{16(2m^* + 1)}\right),
\end{aligned}$$

which completes the proof.  $\square$

**Lemma B.4.** Consider any  $t$  and model such that  $\mathbf{S} = x_{t+1:t+2n}$  is a region then contains a single changepoint at  $\tau^* = t+n$  and  $\Delta$  is the absolute difference between the true means before and after the change. For any  $5 \leq z \leq n\Delta^2/8$  we have

$$\mathbb{P}(\mathcal{L}(\mathbf{S}; \emptyset) - \mathcal{L}^*(\mathbf{S}) \leq z) \leq 2 \exp\left(-\frac{z}{20}\right).$$

*Proof.* It is straightforward to show that  $\mathcal{L}(\mathbf{S}; \emptyset) - \mathcal{L}(\mathbf{S}, \tau^*) \sim \chi_1^2(\nu)$  with non-centrality parameter  $\nu = n\Delta^2/2$ , and  $\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}, \tau^*) \sim \chi_2^2$ . Therefore as long as  $5 \leq z \leq n\Delta^2/8$ ,

$$\mathbb{P}(\mathcal{L}(\mathbf{S}; \emptyset) - \mathcal{L}^*(\mathbf{S}) \leq z) \quad (\text{B.6})$$

$$\begin{aligned} &\leq \mathbb{P}(\mathcal{L}(\mathbf{S}; \emptyset) - \mathcal{L}^*(\mathbf{S}) \leq z + \mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}, \tau^*)) \\ &\leq \mathbb{P}(\mathcal{L}(\mathbf{S}; \emptyset) - \mathcal{L}(\mathbf{S}, \tau^*) \leq 2z) + \mathbb{P}(\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}, \tau^*) \geq z) \\ &\leq \mathbb{P}(\chi_1^2(\nu) \leq 2z) + \mathbb{P}(\chi_2^2 \geq z) \\ &\leq \exp\left(-\frac{(1 + \nu - 2z)^2}{4 + 8\nu}\right) + \exp\left(-\frac{z - 2\sqrt{z}}{2}\right) \\ &< 2 \exp\left(-\frac{z}{20}\right), \end{aligned} \quad (\text{B.7})$$

where the second last inequality follows from Lemma B.1.  $\square$

As Lemmas B.2, B.3 and B.4 hold, respectively, for any  $\mathbf{S} \in \{\mathbf{S}_{1,n}(t)\}$ ,  $\{\mathbf{S}_{2,n}(t)\}$  and  $\{\mathbf{S}_{\Delta,n}(t)\}$ , it is straightforward to obtain Propositions 3.1 and 3.2.

### B.2. Proof of Theorem 3.1

We take  $\beta = (2 + \epsilon) \log T$ , with a suitable choice of  $n_{1:m^*}$  such that

$$n_j = \min \left\{ \frac{8(\beta + a(\beta, T))}{\Delta_j^2}, \delta_j, \delta_{j+1} \right\},$$

where  $a(\beta, T) = (\beta - 2 \log T)/4$  as indicates in Proposition 3.2. Let  $b(\beta, T) = (\beta - 2 \log T)/4(2m^* + 1)$ .

First, we show that the choice of  $\beta$  satisfy the requirements in Theorem 2.1. As we require  $\delta_T \Delta_T^2 \geq (16 + 10\epsilon) \log T$ , it follows that

$$n_j = \frac{8(\beta + a(\beta, T))}{\Delta_j^2} = \frac{(16 + 10\epsilon) \log T}{\Delta_j^2}. \quad (\text{B.8})$$

If  $T \geq \max \{ \exp(64/\epsilon + 128/\epsilon^2), \exp((128m^* + 64)/\epsilon) \}$ , we have  $\beta \geq \gamma_T^{(1)}$ , where

$$\gamma_T^{(1)} = \max \left\{ (2 + \epsilon) \log T, 2 \log n + 8\sqrt{16 + 2 \log T} + 32, 2 \log T + 32(2m^* + 1) \right\},$$

as defined in Proposition 3.2.

Moreover, note that if  $T$  is large enough and

$$\Delta_T^2 \geq \frac{2(16 + 10\epsilon) \log T}{T^{1/(4m^*+3)}},$$

we have  $\epsilon \log T \geq 64(2m^* + 1)$  and  $2 \log T \geq (8m^* + 6) \log(2n_j)$  for all  $j = 1, 2, \dots, m^*$ . Therefore  $\beta \geq \max_j \gamma_{n_j}^{(2)}$ , where each of  $\gamma_{n_j}^{(2)}$  has the following form

$$\gamma_{n_j}^{(2)} = \max \{ (8m^* + 6 + \epsilon) \log(2n_j), 2 \log(2n_j) + 64(2m^* + 1) \},$$

as defined in Proposition 3.2.

Altogether, as long as  $T$  is large enough, we have  $\beta \geq \max \left\{ \gamma_T^{(1)}, \max_j \gamma_{n_j}^{(2)} \right\}$  and  $a(\beta, T) > 2m^*b(\beta, T)$ .

Next, we give the probability bound for local region conditions. By Proposition 3.1, we can derive the formula of  $p_1(\gamma, n), p_2(\gamma, n), p_3(\gamma, n)$  and  $p_4(\gamma, n)$ . Since  $\beta = (2 + \epsilon) \log T$ , it is straightforward to verify that

$$\min \left\{ p_1(\beta, T)/2, p_2(\beta, T), p_3 \left( \beta, \max_j n_j \right), p_4 \left( \beta, \max_j n_j \right) \right\} \leq T^{-\epsilon/16(2m^*+1)}.$$

In addition, note that  $\bar{\mathcal{S}} = \min_j \Delta_j^2 n_j / 2$ , using equation (B.8), we have  $\bar{\mathcal{S}} \geq 20$  as  $T$  is large enough. Combined with Proposition 3.2 we obtain

$$p_5(\bar{\mathcal{S}}, \beta + a(\beta, T)) = 2 \exp \left( -\frac{\beta + a(\beta, T)}{20} \right) = 2T^{-\frac{2+5/4\epsilon}{20}} \leq 2T^{-\frac{\epsilon}{32m^*+16}}.$$

Hence, following Theorem 2.1, we have

$$\begin{aligned} & \mathbb{P}(\hat{m} = m, |\hat{\tau}_j - \tau_j^*| \leq n_j \text{ for all } j = 1, \dots, m^*) \\ & \geq 1 - (m^* + 1)p_1(\beta, T) - (m^* + 1)p_2(\beta, T) - m^*p_3 \left( \beta, \max_j n_j \right) \\ & \quad - m^*p_4 \left( \beta, \max_j n_j \right) - m^*p_5(\bar{\mathcal{S}}, \beta + a(\beta, T)) \\ & \geq 1 - (7m^* + 3)T^{-\epsilon/(32m^*+16)}. \end{aligned}$$

## Appendix C: Proofs of Section 3.2

### C.1. Proof for Proposition 3.3 and 3.4

**Lemma C.1.** For any  $t$  and any model, if  $\mathbf{S} = x_{t+1:t+n}$  is a region that contains no true changepoint, then for any  $\gamma \geq \gamma_n^{(1)}$ , where

$$\gamma_n^{(1)} = \max \left\{ (2 + \epsilon) \log n, 2 \log n + 4\sqrt{9 + 3 \log n} + 12, 2 \log n + 96(2m^* + 1) \right\}, \quad (\text{C.1})$$

where  $\epsilon > 0$ , we have

$$\mathbb{P} \left( \min_{1 \leq k, \tau_{1:k}} \{ \mathcal{L}(\mathbf{S}; \tau_{1:k}) + k\gamma \} - \mathcal{L}^*(\mathbf{S}) \leq \frac{\gamma - 2 \log n}{6} \right) < 2 \exp \left( -\frac{\gamma - 2 \log n}{6} \right),$$

and

$$\mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \emptyset) \geq \frac{\gamma - 2 \log n}{6(2m^* + 1)}\right) \leq \exp\left(-\frac{\gamma - 2 \log n}{24(2m^* + 1)}\right).$$

*Proof.* Note that by Lemma E.2, for any  $k$  and  $\tau_{1:k}$ , we have  $\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k}) \sim \chi_{k+2}^2$  and  $\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \emptyset) \sim \chi_2^2$ .

Therefore, similar to the proof of Lemma B.2, let  $a(\gamma, n) = \frac{\gamma - 2 \log n}{6}$ ,

$$\begin{aligned} & \mathbb{P}\left(\min_{k, \tau_{1:k}} \{\mathcal{L}(\mathbf{S}; \tau_{1:k}) + k\gamma - \mathcal{L}^*(\mathbf{S})\} \leq a(\gamma, n)\right) \\ & \leq \sum_{k=1}^n \mathbb{P}\left(\max_{\tau_{1:k}} [\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k})] \geq k\gamma - a(\gamma, n)\right) \\ & \leq \sum_{k=1}^{\infty} \binom{n}{k} \mathbb{P}(\chi_{k+2}^2 \geq k\gamma - a(\gamma, n)) \end{aligned}$$

Applying (B.1) in Lemma B.1, we have

$$\begin{aligned} \binom{n}{k} \mathbb{P}(\chi_{k+2}^2 \geq k\gamma - a(\gamma, n)) & \leq \frac{n^k}{k!} \exp\left(-\frac{k\gamma - (\gamma - 2 \log n)/6 - \sqrt{2k(k+2)\gamma}}{2}\right) \\ & \leq \frac{1}{k!} \exp\left(-\frac{(k-1/6)(\gamma - 2 \log n) - \sqrt{2k(k+2)\gamma}}{2}\right) \end{aligned}$$

Note that as long as  $\gamma \geq 2 \log n + 4\sqrt{9+3 \log n} + 12$ , we have

$$\sqrt{2k(k+2)\gamma} \leq \frac{k(\gamma - 2 \log n)}{2}, \quad (\text{C.2})$$

which leads to

$$\begin{aligned} \mathbb{P}\left(\min_{k, \tau_{1:k}} \{\mathcal{L}(\mathbf{S}; \tau_{1:k}) + k\gamma - \mathcal{L}^*(\mathbf{S})\} \leq a(\gamma, n)\right) & \leq \sum_{k=1}^{\infty} \frac{1}{k!} \exp\left(\frac{(k/2 - 1/6)(\gamma - 2 \log n)}{2}\right) \\ & \leq \sum_{k=1}^{\infty} \frac{1}{2^{k-1}} \exp\left(-\frac{\gamma - 2 \log n}{6}\right) \\ & = 2 \exp\left(-\frac{\gamma - 2 \log n}{6}\right). \end{aligned}$$

Let  $b(\gamma, n) = (\gamma - 2 \log n)/(12m^* + 6)$ . If  $\gamma \geq 2 \log n + 96(2m^* + 1)$ , then

$b(\gamma, n) \geq 16$ , and, as a result,  $\sqrt{4b(\gamma, n)} \leq b(\gamma, n)/2$ . Hence

$$\begin{aligned} \mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \emptyset) \geq b(\gamma, n)\right) &= \mathbb{P}\left(\chi_2^2 \geq b(\gamma, n)\right) \\ &\leq \exp\left(-\frac{b(\gamma, n) - \sqrt{4b(\gamma, n)}}{2}\right) \\ &\leq \exp\left(-\frac{b(\gamma, n)}{4}\right) \\ &= \exp\left(-\frac{\gamma - 2 \log n}{24(2m^* + 1)}\right). \end{aligned}$$

□

**Lemma C.2.** For any  $t$ , any  $n \geq 4$  and any model, if  $\mathbf{S} = x_{t+1:t+2n}$  is a region that contains a single changepoint at  $\tau^* = t + n$ , then for any  $\gamma \geq \gamma_n^{(2)}$  where

$$\gamma_n^{(2)} = \max\{(3 + \epsilon) \log(2n), 2 \log(2n) + 32 \log(C \log(2n)), 2 \log(2n) + 972(2m^* + 1), 3240\}$$

where  $\epsilon > 0$  and  $C$  is a positive constant not depend on  $n$ , we have

$$\mathbb{P}\left(\min_{k \geq 2, \tau_{1:k}} \{\mathcal{L}(\mathbf{S}; \tau_{1:k}) + (k-1)\gamma\} - \mathcal{L}^*(\mathbf{S}) \leq \frac{\gamma - 2 \log(2n)}{6}\right) < \frac{9}{4} \exp\left(-\frac{\gamma - 3 \log(2n)}{3}\right),$$

and

$$\mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_1} \mathcal{L}(\mathbf{S}; \tau_1) \leq \frac{\gamma - 2 \log(2n)}{6(2m^* + 1)}\right) \leq \exp\left(-\frac{\gamma - 2 \log(2n)}{24(2m^* + 1)}\right).$$

*Proof.* By Lemma E.1, note that for any  $\tau_{1:k}$  on  $\mathbf{S}$  and  $k \geq 1$ ,

$$\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k}) \leq \mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k}, \tau^*) \sim \chi_{k+3}^2, \text{ and}$$

$$\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau^*) \sim \chi_3^2.$$

Define  $a(\gamma, 2n) = (\gamma - 2 \log(2n))/6$ . Since  $\gamma > \max\{(3 + \epsilon) \log(2n), 3240\}$ , we have

$$a(\gamma, 2n) \leq \frac{\gamma}{6} \quad \text{and} \quad \sqrt{2(k-1)(k+3)\gamma} \leq \frac{(k-1)\gamma}{18}. \quad (\text{C.3})$$

Therefore, similar to the proof of Lemma B.3, for  $k \geq 4$  we have

$$\begin{aligned} &\mathbb{P}\left(\min_{k \geq 4, \tau_{1:k}} \{\mathcal{L}(\mathbf{S}; \tau_{1:k}) + (k-1)\gamma\} - \mathcal{L}^*(\mathbf{S}) \leq a(\gamma, 2n)\right) \\ &\leq \sum_{k=4}^{\infty} \mathbb{P}\left(\max_{\tau_{1:k}} (\mathcal{L}(\mathbf{S})^* - \mathcal{L}(\mathbf{S}; \tau_{1:k}, \tau^*)) \geq (k-1)\gamma - a(\gamma, 2n)\right) \\ &< \sum_{k=4}^{\infty} \binom{2n}{k} \exp\left(-\frac{(k-1)\gamma - a(\gamma, 2n) - \sqrt{2(k-1)(k+3)\gamma}}{2}\right) \end{aligned}$$

where the last inequality is due to Bonferroni correction and (B.1) in Lemma B.1. Together with (C.3), we have

$$\begin{aligned}
& \mathbb{P} \left( \min_{k \geq 4, \tau_{1:k}} \{ \mathcal{L}(\mathbf{S}; \tau_{1:k}) + (k-1)\gamma \} - \mathcal{L}^*(\mathbf{S}) \leq a(\gamma, 2n) \right) \\
& \leq \sum_{k=4}^{\infty} \binom{2n}{k} \exp \left( - \frac{(k-1)\gamma - \gamma/6 - (k-1)\gamma/18}{2} \right) \\
& \leq \sum_{k=4}^{\infty} \frac{1}{k!} \exp \left( - \frac{[18(k-1)/17 - 1/6]\gamma - 2k \log(2n)}{2} \right) \\
& < \sum_{k=4}^{\infty} \frac{1}{k!} \exp \left( - \frac{4(\gamma - 3 \log(2n))}{3} \right) \\
& < \frac{1}{4} \exp \left( - \frac{4(\gamma - 3 \log(2n))}{3} \right) \leq \frac{1}{4} \exp \left( - \frac{\gamma - 3 \log(2n)}{3} \right)
\end{aligned}$$

Now we only need to handle the case when  $k = 2$  and  $3$ , note that

$$\max_{\tau_1, \tau_2} \{ \mathcal{L}^*(\mathbf{S}) - \mathcal{L}_n(\mathbf{S}; \tau_1, \tau_2) \} \leq \max_{\tau_1, \tau_2} \{ \mathcal{L}^*(\mathbf{S}) - \mathcal{L}_n(\mathbf{S}; \tau_{1:2}, \tau^*) \},$$

and

$$\max_{\tau_{1:3}} \{ \mathcal{L}^*(\mathbf{S}) - \mathcal{L}_n(\mathbf{S}; \tau_{1:3}) \} \leq \max_{\tau_{1:3}} \{ \mathcal{L}^*(\mathbf{S}) - \mathcal{L}_n(\mathbf{S}; \tau_{1:3}, \tau^*) \}.$$

Using the results from Lemmas E.7 and E.8, if  $\gamma > \max \left\{ 240, 24 \log(C \log(2n)) \right\}$ , where  $C = \max \{ C_1''', C_2''', C_3'''' \}$  is a positive constant and  $C_1''', C_2''', C_3''''$  are constants introduced in Section E.3, we have both the events

$$\max_{\tau_1, \tau_2} \{ \mathcal{L}^*(\mathbf{S}) - \mathcal{L}_n(\mathbf{S}; \tau_{1:2}, \tau^*) \} \geq \gamma - a(\gamma, 2n)$$

and

$$\max_{\tau_1, \tau_2, \tau_3} \{ \mathcal{L}^*(\mathbf{S}) - \mathcal{L}_n(\mathbf{S}; \tau_{1:3}, \tau^*) \} \geq 2\gamma - a(\gamma, 2n)$$

hold with probability less than  $\exp \left( - \frac{\gamma - 3 \log(2n)}{3} \right)$ . Therefore, by the union bound, we have

$$\mathbb{P} \left( \min_{k \geq 2, \tau_{1:k}} \{ \mathcal{L}(\mathbf{S}; \tau_{1:k}) + (k-1)\gamma \} - \mathcal{L}^*(\mathbf{S}) \leq \frac{\gamma - 2 \log(2n)}{6} \right) < \frac{9}{4} \exp \left( - \frac{\gamma - 3 \log(2n)}{3} \right).$$

Let  $b(\gamma, n) = (\gamma - 2 \log n)/(12m^* + 6)$ , by Lemma E.6, if

$$\gamma > \max \left\{ 2 \log(2n) + 32 \log(C \log(2n)), 2 \log(2n) + 972(2m^* + 1) \right\},$$

we have

$$\begin{aligned} \mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_1} \mathcal{L}(\mathbf{S}; \tau_1) \geq b(\gamma, 2n)\right) &\leq \mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_1} \mathcal{L}(\mathbf{S}; \tau_1, \tau^*) \geq b(\gamma, 2n)\right) \\ &\leq \exp\left(-\frac{\gamma - 2 \log(2n)}{24(2m^* + 1)}\right), \end{aligned}$$

which completes the proof.  $\square$

**Lemma C.3.** *For any  $t$  and any model, if  $\mathbf{S} = x_{t+1:t+2n}$  is a region that contains a single changepoint at  $\tau^* = t + n$  and  $\Delta$  is the absolute difference between the true slopes before and after the change, then for any  $8 \leq z \leq n^3 \Delta^2/100$  and  $n \geq 2$ , we have*

$$\mathbb{P}\{\mathcal{L}(\mathbf{S}; \emptyset) - \mathcal{L}^*(\mathbf{S}) \leq z\} \leq \exp(-z/18).$$

*Proof.* Lemma E.1 suggests that  $\mathcal{L}(\mathbf{S}, \emptyset) - \mathcal{L}(\mathbf{S}, \tau^*)$  and  $\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}, \tau^*)$  follows  $\chi_1^2(\nu)$  and  $\chi_3^2$ , respectively, where

$$\nu = \Delta_n^2 \frac{n(n+1)(n-1)}{24} \frac{4n^2 + 2}{4n^2 - 1} \geq \frac{\Delta_n^2 n^3}{25}.$$

Following the same argument as in the proof of Lemma B.4, as long as  $8 \leq z \leq n^3 \Delta^2/100$ ,

$$\begin{aligned} &\mathbb{P}\left(\mathcal{L}(\mathbf{S}; \emptyset) - \mathcal{L}^*(\mathbf{S}) \leq z + \mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}, \tau^*)\right) \\ &\leq \mathbb{P}\left(\mathcal{L}(\mathbf{S}; \emptyset) - \mathcal{L}(\mathbf{S}; \tau^*) \leq 2z\right) + \mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}, \tau^*) \geq z\right) \\ &\leq \mathbb{P}\left(\chi_1^2(\nu) \leq 2z\right) + \mathbb{P}\left(\chi_3^2 \geq z\right) \\ &\leq \exp\left(-\frac{(1 + \nu - 2z)^2}{4 + 8\nu}\right) + \exp\left(-\frac{z - \sqrt{6z}}{2}\right) \tag{C.4} \\ &\leq 2 \exp\left(-\frac{z}{20}\right), \end{aligned}$$

where the second inequality follows from Lemma B.1.  $\square$

Note that Lemmas C.1, C.2 and C.3 hold for any  $\mathbf{S} \in \{\mathbf{S}_{1,n}(t)\}$ ,  $\{\mathbf{S}_{2,n}(t)\}$  and  $\{\mathbf{S}_{\Delta,n}(t)\}$ , respectively. Therefore, it is straightforward to obtain Propositions 3.3 and 3.4.

### C.2. Proof of Theorem 3.2

Similar to the proof of Theorem 3.1, we take  $\beta = (2 + \epsilon) \log T$  and

$$n_j = \min \left\{ \left( \frac{100(\beta + a(\beta, T))}{\Delta_j^2} \right)^{1/3}, \delta_j, \delta_{j+1} \right\},$$

where  $a(\beta, T) = (\beta - 2 \log T)/6$  as indicates in Proposition 3.3 and let  $b(\beta, T) = (\beta - 2 \log T)/6(2m^* + 1)$ .

Therefore, since  $\delta_T^3 \Delta_T^2 \geq (200 + 350\epsilon/3) \log T = 100(\beta + a(\beta, T))$ , we have

$$n_j = \left( \frac{(200 + 350\epsilon/3) \log T}{\Delta_j^2} \right)^{1/3}.$$

Combined with the assumption that  $\Delta_T^2 \geq \frac{8(200 + 250\epsilon/3) \log T}{T^2}$ , thus  $\max_j n_j \leq T^{2/3}/2$ , which leads to  $\log(2 \max_j n_j) \leq 2/3 \log T$ .

Therefore, it is straightforward to verify that as long as  $T$  is large enough, we have  $\beta = (2 + \epsilon) \log T \geq \max\{\gamma_T^{(1)}, \max_j \gamma_{n_j}^{(2)}\}$ , where

$$\gamma_T^{(1)} = \max \left\{ (2 + \epsilon) \log T, 2 \log t + 4\sqrt{9 + 3 \log t} + 12, 2 \log T + 96(2m^* + 1) \right\},$$

$$\gamma_{n_j}^{(2)} = \max \left\{ (3 + \epsilon) \log(2n_j), 2 \log(2n_j) + 32 \log(C \log(2n_j)), 2 \log(2n_j) + 972(2m^* + 1), 3240 \right\},$$

as defined in Proposition 3.3. Therefore we can verify that

$$\min \left\{ p_1(\beta, T)/2, p_2(\beta, T), 4p_3 \left( \beta, \max_j n_j \right) /9, p_4 \left( \beta, \max_j n_j \right) \right\} \leq T^{-\epsilon/24(2m^*+1)},$$

where  $p_1(\gamma, n), p_2(\gamma, n), p_3(\gamma, n), p_4(\gamma, n)$  can be obtained from Lemmas C.1, C.2 and C.3.

Applying Proposition 3.4, note that if  $\bar{\mathcal{S}} \geq 4(\beta + a(\beta, T)) \geq 32$ , which is true as  $T$  is large enough, we have

$$p_5(\bar{\mathcal{S}}, \beta + a(\beta, T)) \leq 2T^{-(2+7\epsilon/6)/20} \leq 2T^{-\epsilon/(48m^*+24)}.$$

Therefore,

$$\begin{aligned} & \mathbb{P} \left( \hat{m} = m, \max_j |\hat{\tau}_j - \tau_j^*|^3 \Delta_j^2 \leq (200 + 350\epsilon/3) \log T \right) \\ & \geq 1 - (m^* + 1)p_1(\beta, T) - (m^* + 1)p_2(\beta, T) - m^* p_3 \left( \beta, \max_j n_j \right) \\ & \quad - m^* p_4 \left( \beta, \max_j n_j \right) - m^* p_5(\bar{\mathcal{S}}, \beta + a(\beta, T)) \\ & \geq 1 - (33m^*/4 + 3)T^{-\epsilon/(48m^*+24)}. \end{aligned}$$

□

## Appendix D: Proofs of in Section 3.3

### D.1. Proofs for Propositions 3.5 and 3.6

**Lemma D.1.** *For any  $t$  and any model, if  $\mathbf{S} = x_{t+1:t+n}$  is a region contains no true changepoint, for any*

$$\gamma \geq \max \left\{ (2 + \epsilon) \log n, 2 \log n + 8\sqrt{16 + 2 \log n} + 32, 2 \log n + 32(2m^* + 1) \right\},$$

where  $\epsilon$  is an arbitrarily small positive constant, we have

$$\mathbb{P} \left( \min_{1 \leq k, \tau_{1:k}} \{ \mathcal{L}(\mathbf{S}; \tau_{1:k}) + k\gamma \} - \mathcal{L}^*(\mathbf{S}) \leq \frac{\gamma - 2 \log n}{4} \right) < 2 \exp \left( -\frac{\gamma - 2 \log n}{4} \right),$$

and

$$\mathbb{P} \left( \mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \emptyset) \geq \frac{\gamma - 2 \log n}{4(2m^* + 1)} \right) \leq \exp \left( -\frac{\gamma - 2 \log n}{16(2m^* + 1)} \right).$$

*Proof.* We only need to consider the case that  $t + 1 \leq \tau_{1:k} \leq t + n$ . Without loss of generality, let  $t = 0$ , i.e., consider  $\mathbf{S} = x_{1:n}$ , where  $x_s = \theta^* \alpha^{s-1} + \varepsilon_s$ .

Note that  $\mathcal{L}^*(\mathbf{S}) = \sum_{s=1}^n \varepsilon_s^2 / \sigma^2$ .

For any  $1 \leq \tau_{1:k} < n$ , let  $\tau_0 = 0$  and  $\tau_{k+1} = n$ , it is easy to derive

$$\mathcal{L}(\mathbf{S}; \tau_{1:k}) = \sum_{j=1}^{k+1} \sum_{s=\tau_{j-1}+1}^{\tau_j} \frac{1}{\sigma^2} \left( x_s - \hat{\theta}_j \alpha^{s-\tau_{j-1}-1} \right)^2,$$

where

$$\hat{\theta}_j = \frac{\sum_{s=\tau_{j-1}+1}^{\tau_j} x_s \alpha^{s-\tau_{j-1}-1}}{\sum_{s=\tau_{j-1}+1}^{\tau_j} \alpha^{2(s-\tau_{j-1}-1)}} = \theta^* \alpha^{\tau_{j-1}} + \frac{\sum_{s=\tau_{j-1}+1}^{\tau_j} \varepsilon_s \alpha^{s-\tau_{j-1}-1}}{\sum_{s=\tau_{j-1}+1}^{\tau_j} \alpha^{2(s-\tau_{j-1}-1)}}.$$

Therefore for each  $j = 1, \dots, m + 1$ ,

$$\begin{aligned} & \sum_{s=\tau_{j-1}+1}^{\tau_j} \frac{1}{\sigma^2} \varepsilon_s^2 - \sum_{s=\tau_{j-1}+1}^{\tau_j} \frac{1}{\sigma^2} \left( x_s - \hat{\theta}_{j+1} \alpha^{s-\tau_{j-1}-1} \right)^2 \\ &= \sum_{s=\tau_{j-1}+1}^{\tau_j} \frac{1}{\sigma^2} \varepsilon_s^2 - \sum_{s=\tau_{j-1}+1}^{\tau_j} \frac{1}{\sigma^2} \left( \varepsilon_s - \alpha^{s-\tau_{j-1}-1} \frac{\sum_{s=\tau_{j-1}+1}^{\tau_j} \varepsilon_s \alpha^{s-\tau_{j-1}-1}}{\sum_{s=\tau_{j-1}+1}^{\tau_j} \alpha^{2(s-\tau_{j-1}-1)}} \right)^2 \\ &= \frac{\left( \sum_{s=\tau_{j-1}+1}^{\tau_j} \varepsilon_s \alpha^{s-\tau_{j-1}-1} \right)^2}{\sigma^2 \sum_{s=\tau_{j-1}+1}^{\tau_j} \alpha^{2(s-\tau_{j-1}-1)}}. \end{aligned}$$

Note that  $\frac{\sum_{s=\tau_{j-1}+1}^{\tau_j} \varepsilon_s \alpha^{s-\tau_{j-1}-1}}{\sigma \left( \sum_{s=\tau_{j-1}+1}^{\tau_j} \alpha^{2(s-\tau_{j-1}-1)} \right)^{1/2}} \sim N(0, 1)$ . Since  $\{\varepsilon_s\}_{s=1}^T$  are i.i.d, we

have  $\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k}) \sim \chi_{k+1}^2$ .

Similarly, we have

$$\mathcal{L}(\mathbf{S}, \emptyset) = \sum_{s=1}^n \frac{1}{\sigma^2} \left( \varepsilon_s - \alpha^{s-1} \frac{\sum_{s=1}^n \varepsilon_s \alpha^{(s-1)}}{\sum_{s=1}^n \alpha^{2(s-1)}} \right)^2 = \sum_{s=1}^n \frac{\varepsilon_s^2}{\sigma^2} - \frac{\left( \sum_{s=1}^n \varepsilon_s \alpha^{(s-1)} \right)^2}{\sigma^2 \sum_{s=1}^n \alpha^{2(s-1)}},$$

which leads to  $\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \emptyset) \sim \chi_1^2$ .

Using the same argument in the proof of Lemma B.2 completes the proof.  $\square$

**Lemma D.2.** *For any  $t$  and any model, if  $\mathbf{S} = x_{t+1:t+2n}$  is a region contains a single changepoint at  $\tau^* = t + n$ , for any*

$$\gamma \geq \max\{(8m^* + 6 + \epsilon) \log(2n), 2 \log(2n) + 64(2m^* + 1)\},$$

where  $\epsilon$  is an arbitrarily small positive constant, for  $n \geq 4$ , we have

$$\mathbb{P} \left( \min_{k \geq 2, \tau_{1:k}} \{\mathcal{L}(\mathbf{S}; \tau_{1:k}) + (k-1)\gamma\} - \mathcal{L}^*(\mathbf{S}) \leq \frac{\gamma - 2 \log(2n)}{4} \right) < \exp \left( -\frac{\gamma - 8 \log n}{4} \right),$$

and

$$\mathbb{P} \left( \mathcal{L}^*(\mathbf{S}) - \min_{\tau_1} \mathcal{L}(\mathbf{S}; \tau_1) \leq \frac{\gamma - 2 \log(2n)}{4(2m^* + 1)} \right) \leq \exp \left( -\frac{\gamma - (8m^* + 6) \log(2n)}{16(2m^* + 1)} \right).$$

*Proof.* For any  $\tau_{1:k}$ , note that

$$\begin{aligned} \mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k}) &\leq \mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k}, \tau^*) \\ &= \mathcal{L}^*(x_{(t+1):(t+n)}) + \mathcal{L}^*(x_{(t+n+1):(t+2n)}) \\ &\quad - \mathcal{L}(x_{(t+1):(t+n)}; \tau_{1:k}) - \mathcal{L}(x_{(t+n+1):(t+2n)}; \tau_{1:k}) \end{aligned} \quad (\text{D.1})$$

From the proof of Lemma D.1, we have (D.1) follows a chi-square distribution with degrees of freedom  $k + 2$ . Similarly, we have

$$\mathcal{L}^*(\mathbf{S}) - \min_{\tau_1} \mathcal{L}(\mathbf{S}; \tau_1) \leq \mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau^*) \sim \chi_2^2. \quad (\text{D.2})$$

Using the same argument as in the proof of Lemma B.3, we obtain the results.  $\square$

**Lemma D.3.** *If  $\mathbf{S} = x_{t+1:t+2n}$  is a region contains a single changepoint at  $\tau^* = t + n$  and  $\Delta$  be the absolute difference between the true means before and after the change. For any  $5 \leq z \leq \frac{\Delta^2}{(1 - \alpha^{2n})(1 - \alpha^2)}$ , we have*

$$\mathbb{P} \{ \mathcal{L}(\mathbf{S}; \emptyset) - \mathcal{L}^*(\mathbf{S}) \leq z \} \leq \exp(-z/20).$$

*Proof.* Without loss of generality we let  $t = 0$ , i.e.,  $\mathbf{S} = x_{1:2n}$  with  $\tau^* = n$  is a changepoint. Therefore, we write that

$$\begin{cases} x_s = \theta_1^* \alpha^{s-1} + \varepsilon_s, & 1 \leq s \leq n \\ x_s = \theta_2^* \alpha^{s-n-1} + \varepsilon_s, & n+1 \leq s \leq 2n. \end{cases}$$

First, we have  $\mathcal{L}^*(\mathbf{S}) = \sum_{s=1}^{2n} \varepsilon_s^2 / \sigma^2$  and

$$\mathcal{L}(\mathbf{S}; \tau^*) = \sum_{s=1}^n \frac{1}{\sigma^2} \left( x_s - \hat{\theta}_1 \alpha^{s-1} \right)^2 + \sum_{s=n+1}^{2n} \frac{1}{\sigma^2} \left( x_s - \hat{\theta}_2 \alpha^{s-n-1} \right)^2, \quad (\text{D.3})$$

where

$$\hat{\theta}_1 = \theta_1^* + \frac{\sum_{s=1}^n \varepsilon_s \alpha^{s-1}}{\sum_{s=1}^n \alpha^{2(s-1)}}, \quad \text{and} \quad \hat{\theta}_2 = \theta_2^* + \frac{\sum_{s=n+1}^{2n} \varepsilon_s \alpha^{s-n-1}}{\sum_{s=n+1}^{2n} \alpha^{2(s-n-1)}}.$$

Using the similar argument as in the proof of Lemma D.1, we can rewrite (D.3) as

$$\begin{aligned} \mathcal{L}(\mathbf{S}; \tau^*) &= \sum_{s=1}^n \frac{1}{\sigma^2} \left( \varepsilon_s - \alpha^{s-1} \frac{\sum_{s=1}^n \varepsilon_s \alpha^{(s-1)}}{\sum_{s=1}^n \alpha^{2(s-1)}} \right)^2 \\ &\quad + \sum_{s=n+1}^{2n} \frac{1}{\sigma^2} \left( \varepsilon_s - \alpha^{s-n-1} \frac{\sum_{s=n+1}^{2n} \varepsilon_s \alpha^{(s-n-1)}}{\sum_{s=n+1}^{2n} \alpha^{2(s-n-1)}} \right)^2 \\ &= \sum_{s=1}^{2n} \frac{1}{\sigma^2} \varepsilon_s^2 - \frac{\left( \sum_{s=1}^n \varepsilon_s \alpha^{s-1} \right)^2}{\sigma^2 \sum_{s=1}^n \alpha^{2(s-1)}} - \frac{\left( \sum_{s=n+1}^{2n} \varepsilon_s \alpha^{s-n-1} \right)^2}{\sigma^2 \sum_{s=n+1}^{2n} \alpha^{2(s-n-1)}}. \end{aligned}$$

Therefore,  $\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau^*) \sim \chi_2^2$ . Moreover, note  $\mathcal{L}(\mathbf{S}, \emptyset) = \sum_{s=1}^{2n} (x_s - \hat{\theta} \alpha^{s-1})^2 / \sigma^2$ , where we omit the changepoint and as there is a single parameter  $\theta$  to estimate, let  $\eta = \alpha^n$ , we have

$$\hat{\theta} = \frac{\theta_1}{1 + \eta^2} + \frac{\theta_2 \eta}{1 + \eta^2} + \frac{\sum_{s=1}^{2n} \varepsilon_s \alpha^{s-1}}{\sum_{s=1}^{2n} \alpha^{2(s-1)}}.$$

Thus, by simply algebra calculation, we obtain  $\mathcal{L}(\mathbf{S}, \emptyset) - \mathcal{L}(\mathbf{S}, \tau^*) \sim \chi_1^2(\nu)$ , where the non-centrality parameter

$$\nu = \frac{(\theta_2 - \eta \theta_1)^2 (1 - \eta^2)}{(1 + \eta^2)(1 - \alpha^2)} = \frac{\Delta^2 (1 + \alpha^{2n})}{(1 - \alpha^{2n})(1 - \alpha^2)} \geq \frac{\Delta^2}{(1 - \alpha^{2n})(1 - \alpha^2)} \quad (\text{D.4})$$

Follow the same argument in the proof of Lemma B.4 we obtain that, as long as  $5 \leq z \leq \nu/4$ ,

$$\begin{aligned} &\mathbb{P}(\mathcal{L}(\mathbf{S}; \emptyset) - \mathcal{L}^*(\mathbf{S}) \leq z + \mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}, \tau^*)) \\ &\leq \mathbb{P}(\mathcal{L}(\mathbf{S}; \emptyset) - \mathcal{L}(\mathbf{S}; \tau^*) \leq 2z) + \mathbb{P}(\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}, \tau^*) \geq z) \\ &\leq \mathbb{P}(\chi_1^2(\nu) \leq 2z) + \mathbb{P}(\chi_2^2 \geq z) \\ &\leq \exp\left(-\frac{(1 + \nu - 2z)^2}{4 + 8\nu}\right) + \exp\left(-\frac{z - 2\sqrt{z}}{2}\right) \\ &\leq 2 \exp\left(-\frac{z}{20}\right), \end{aligned}$$

□

Note that Lemmas D.1, D.2 and D.3 hold for any  $\mathbf{S} \in \{\mathbf{S}_{1,n}(t)\}, \{\mathbf{S}_{2,n}(t)\}$  and  $\{\mathbf{S}_{\Delta,n}(t)\}$ , respectively. Therefore, it is straightforward to obtain Propositions 3.5 and 3.6.

### D.2. Proof of Theorem 3.3

Similar to the proof of Theorem 3.1, we take  $\beta = (2 + \epsilon) \log T$  and

$$n_j = \min \left\{ \frac{1}{2} \log_\alpha \left( 1 - \frac{\Delta_j^2}{4(1 - \alpha^2)(\beta + a(\beta, T))} \right), \delta_j \cdot \delta_{j+1} \right\},$$

where  $a(\beta, T) = (\beta - 2 \log T)/4$  as indicates in Proposition 3.3 and let  $b(\beta, T) = (\beta - 2 \log T)/4(2m^* + 1)$ .

Therefore, since we assume  $\frac{\Delta_T^2}{(1 - \alpha^{2\delta^T})(1 - \alpha^2)} \geq (8 + 5\epsilon) \log T$ , we will have each  $n_j$  achieves the minimum value at

$$\frac{1}{2} \log_\alpha \left( 1 - \frac{\Delta_j^2}{4(1 - \alpha^2)(\beta + a(\beta, T))} \right) = \frac{1}{2} \log_\alpha \left( 1 - \frac{\Delta_j^2}{(1 - \alpha^2)(8 + 5\epsilon) \log T} \right). \quad (\text{D.5})$$

Combined with the assumption that

$$\log_\alpha \left( 1 - \frac{\Delta_T^2}{4(1 - \alpha^2)(\beta + a(\beta, T))} \right) \leq T^{2/(8m^* + 6 + \epsilon)},$$

which leads to  $2 \max_j n_j \leq T^{1/4}$ . Therefore, it is straightforward to verify that as long as  $T$  is large enough, we have  $\beta = (2 + \epsilon) \log T \geq \max\{\gamma_T^{(1)}, \max_j \gamma_{n_j}^{(2)}\}$ , where

$$\gamma_T^{(1)} = \max \left\{ (2 + \epsilon) \log T, 2 \log n + 8\sqrt{16 + 2 \log T} + 32, 2 \log T + 32(2m^* + 1) \right\},$$

$$\gamma_{n_j}^{(2)} = \max \left\{ (8m^* + 6 + \epsilon) \log(2n_j), 2 \log(2n_j) + 64(2m^* + 1) \right\},$$

as defined in Proposition 3.5.

Next, by Proposition 3.5, we can work out  $p_1(\gamma, n), p_2(\gamma, n), p_3(\gamma, n)$  and  $p_4(\gamma, n)$ . Since  $\beta = (2 + \epsilon) \log T$ , it is straightforward that

$$\min \left\{ p_1(\beta, T)/2, p_2(\beta, T), p_3 \left( \beta, \max_j n_j \right), p_4 \left( \beta, \max_j n_j \right) \right\} \leq T^{-\epsilon/16(2m^* + 1)}.$$

Moreover, note that  $\bar{\mathcal{S}} = \min_j \frac{\Delta_j^2}{(1 - \alpha^{2n_j})(1 - \alpha^2)} \geq 4(\beta + a(\beta, T)) \geq 20$  as  $T$  is large enough, we have

$$p_5(\bar{\mathcal{S}}, \beta + a(\beta, T)) \leq 2T^{-(2+5\epsilon/4)/20} \leq 2T^{-\epsilon/(32m^* + 16)}.$$

Therefore,

$$\begin{aligned}
& \mathbb{P}\left(\hat{m} = m, \min_j \frac{\Delta_j^2}{(1-\alpha^2)(1-\alpha^{2|\hat{\tau}_j-\tau_j^*|})} \geq (8+5\epsilon)\log T\right) \\
& \geq 1 - (m^*+1)p_1(\beta, T) - (m^*+1)p_2(\beta, T) - m^*p_3\left(\beta, \max_j n_j\right) \\
& \quad - m^*p_4\left(\beta, \max_j n_j\right) \\
& \quad - m^*p_5(\mathcal{S}, \beta + a(\beta, T)) \\
& \geq 1 - (7m^*+3)T^{-\epsilon/(32m^*+16)}.
\end{aligned}$$

□

### D.3. Proof of Corollary 3.3

Let  $c_5 \geq 8 + 5\epsilon$ , and  $0 < c_6^{-1} \leq (1-D)c_5$ , since  $\alpha^{T^{2/(8m^*+6+\epsilon)}} \leq D < 1$ , we have (3.17) and (3.18) hold. Applying Theorem 3.3 with  $c_5 \geq 8 + 5\epsilon$ , since  $m^* = o(\log T)$ , as  $T \rightarrow \infty$ , we obtain

$$\begin{aligned}
& \mathbb{P}\left(\hat{m} = m, \min_j \frac{\Delta_j^2}{(1-\alpha^2)(1-\alpha^{2|\hat{\tau}_j-\tau_j^*|})} \geq c_5 \log T\right) \\
& \geq 1 - (7m^*+3)T^{-\epsilon/(32m^*+16)} \rightarrow 1.
\end{aligned}$$

□

## Appendix E: Orthogonal basis techniques for change-in-slope

This appendix provides additional technical lemmas needed for the change-in-slope problem. We believe the orthogonal basis representation and maxima inequality of correlated Gaussian variables to be of independent interest and therefore present them in this separate appendix.

Without loss of generality, we re-index the  $2n$  points in a local segment  $\mathbf{S} = x_{t+1:t+2n}$  as  $\mathbf{x} = (x_1, \dots, x_{2n})^T$  with a single true changepoint at  $\tau^* = n$ . Let  $\mathbf{f} = (f_1, \dots, f_{2n})^T$  denotes the vector of the linear signals with a change of slope at  $\tau^*$ , e.g

$$f_i = \begin{cases} \theta_0 + \frac{\theta_1 - \theta_0}{n}i, & i = 1, \dots, n; \\ \theta_1 + \frac{\theta_2 - \theta_1}{n}(i - n), & i = n + 1, \dots, 2n, \end{cases}$$

where  $\theta_0$ ,  $\theta_1$  and  $\theta_2$  are unknown parameters, and  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_{2n})^T$  denote the vector of Gaussian stochastic noises. Therefore  $\boldsymbol{x} = \boldsymbol{f} + \boldsymbol{\varepsilon}$ . The following basis representation in the  $2n$ -dimensional vector space will be used to approximate  $\boldsymbol{x}$ .

### E.1. Orthogonal basis.

By algebra calculation, we can sequentially calculate the following basis representation for the  $2n$ -vector  $\boldsymbol{x}$ .

BASIS REPRESENTATION:

1. Constant basis representation:  $\psi_{(C)} = (\psi_{(C)}(1), \dots, \psi_{(C)}(2n))^T$  with  $\psi_{(C)}(i) = (2n)^{-1/2}$ .
2. Linear basis:  $\psi_{(L)} = (\psi_{(L)}(1), \dots, \psi_{(L)}(2n))^T$ , with

$$\psi_{(L)}(i) = \sqrt{\frac{12}{2n(2n-1)(2n+1)}} \left( i - \frac{2n+1}{2} \right).$$

Note that  $\psi_{(L)}$  is orthonormal to  $\psi_{(C)}$ .

3. Basis corresponding to  $\tau^* = n$ :  $\psi_{(\tau^*)} = (\psi_{(\tau^*)}(1), \dots, \psi_{(\tau^*)}(2n))^T$ , with

$$\psi_{(\tau^*)}(i) = \begin{cases} -\sqrt{\frac{3(n+1)}{n(4n^2-1)(2n^2+1)(n-1)}} \left[ (4n-1)i - n(2n+1) \right], & i = 1, \dots, n; \\ \sqrt{\frac{3(n-1)}{n(4n^2-1)(2n^2+1)(n+1)}} \left[ (4n+1)i - 3n(2n+1) \right], & i = n+1, \dots, 2n. \end{cases}$$

Note that  $\psi_{(\tau^*)}$  is orthonormal to both  $\psi_{(C)}$  and  $\psi_{(L)}$ .

4. Basis  $\psi_{(\tau_1)}$  corresponding to adding an additional change  $\tau_1$  on  $\mathbf{S}$ , where  $2 \leq \tau_1 \leq 2n$  and  $\tau_1 \neq \tau^*$ . For example, if  $2 \leq \tau_1 \leq n-1$ , then

$$\psi_{(\tau_1)}(i) = \begin{cases} -A_n \sqrt{\frac{n-\tau_1}{n(n-1)(2n^2+1)\tau_1(\tau_1-1)}} \left( a_n i - b_n \right), & i = 1, \dots, \tau_1; \\ A_n \sqrt{\frac{\tau_1(\tau_1-1)}{n(2n^2+1)(n-\tau_1)(n-1)}} \left( c_n i - d_n \right), & i = \tau_1+1, \dots, n; \\ -A_n \sqrt{\frac{\tau_1(\tau_1-1)(n-\tau_1)(n-1)}{n(2n^2+1)}} \left[ 3i - (5n+1) \right], & i = n+1, \dots, 2n. \end{cases}$$

where

$$\begin{aligned}
A_n &= \sqrt{3}(8n^3\tau_1 - 4n^3 - 13n^2\tau_1^2 + 9n^2\tau_1 + 4n^2 + 5n\tau_1^3 \\
&\quad - 6n\tau_1^2 + 5n\tau_1 - 2n + \tau_1^3 - 5\tau_1^2 + 2\tau_1 + 2)^{-1/2}, \\
a_n &= (4n^3 + 4n^2\tau_1 - 4n^2 - 5n\tau_1^2 + 5n\tau_1 + 2n - \tau_1^2 + 3\tau_1 - 2), \\
b_n &= \tau_1(4n^3 - 3n^2\tau_1 + 3n^2 - 2n\tau_1 + 4n - \tau_1 + 1), \\
c_n &= 9n^2 - 5n\tau_1 + n - \tau_1 + 2, \\
d_n &= 7n^3 - 3n^2\tau_1 + 2n^2 - 2n\tau_1 + 3n - \tau_1.
\end{aligned}$$

Similarly we can write  $\psi_{(\tau_1)}$  for  $n+1 \leq \tau_1 \leq 2n$ . Note that  $\psi_{(\tau_1)}$  is orthonormal to  $\psi_{(C)}$ ,  $\psi_{(L)}$  and  $\psi_{(\tau^*)}$ .

5. Basis  $\psi_{(\tau_j)}$ ,  $j = 2, 3, \dots$ , corresponding to adding a  $j$ -th change  $\tau_j$  on  $\mathbf{S}$  after  $\tau_1, \dots, \tau_{j-1}$ , where  $2 \leq \tau_j \leq 2n$  and  $\tau_j \neq \tau_{1:(j-1)}$  or  $n$ . Moreover  $\psi_{(\tau_j)}$  is orthonormal to  $\psi_{(C)}$ ,  $\psi_{(L)}$ ,  $\psi_{(\tau^*)}$  and  $\psi_{(\tau_{1:(j-1)})}$ .

The formulas for  $\psi_{(C)}$ ,  $\psi_{(L)}$  and  $\psi_{(\tau^*)}$  were also given in [Baranowski et al. \(2016\)](#). We derive  $\psi_{(\tau_1)}$  as it will be used in the proof of [Lemmas E.4](#) and [E.5](#). The formulas for  $\psi_{(\tau_j)}$  can be calculated by applying the Gram-Schmidt procedure to make the vector  $\nu_{(\tau_j)}$  (linear with a kink at  $\tau_j$ ) orthogonal to  $\psi_{(C)}$ ,  $\psi_{(L)}$ ,  $\psi_{(\tau^*)}$  and  $\psi_{(\tau_1)}, \dots, \psi_{(\tau_{j-1})}$ , where

$$\nu_{(\tau_j)}(i) = \begin{cases} 0, & i = 1, 2, \dots, \tau_j \\ i - \tau_j & i = \tau_j + 1, \dots, 2n, \end{cases}$$

We define  $\mathbf{S}_{\tau_1} = \{1, \dots, 2n\} \setminus \{1, n\}$ ;  $\mathbf{S}_{\tau_2} = \{1, \dots, 2n\} \setminus \{1, \tau_1, n\}$  for any given  $\tau_1$ ; and

$$\mathbf{S}_{\tau_{k+1}} = \{1, \dots, 2n\} \setminus \{1, \tau_1, \tau_2, \dots, \tau_k, n\}, \quad \text{given } \tau_1, \dots, \tau_k;$$

which are the sets of possible locations for  $\tau_1$ ,  $\tau_2$  and  $\tau_{k+1}$  on  $\mathbf{S}$ , respectively. To distinguish each of  $\psi_{(\tau_j)}$ , we write  $\psi_{(i,j)}$  as the basis formulas for  $\psi_{(\tau_j)}$  at locations  $i$ , where  $j \in \{1, 2, \dots, k\}$  and  $i \in \mathbf{S}_{\tau_j}$ .

For each orthogonal basis  $\psi_{(\cdot)}$ , define the coefficients that correspond to  $\mathbf{x}$  projected onto it as  $x_{(\cdot)} = \langle \mathbf{x}, \psi_{(\cdot)} \rangle = \langle \mathbf{f}, \psi_{(\cdot)} \rangle + \langle \boldsymbol{\varepsilon}, \psi_{(\cdot)} \rangle = f_{(\cdot)} + \varepsilon_{(\cdot)}$ . We have the following straightforward properties for the signal components  $f_{(\cdot)}$  and the noise components  $\varepsilon_{(\cdot)}$ .

- (1)  $f_{(\tau_j)} = 0$  for  $j = 1, 2, \dots, k$ .
- (2) Each of  $\varepsilon_{(\cdot)}$  are i.i.d with distribution  $\mathcal{N}(0, \sigma^2)$ . Without loss of generality, we assume  $\sigma = 1$  for the rest of the section.
- (3) For any two possible locations  $i$  and  $j$  in  $\mathbf{S}_{\tau_k}$ , we have  $\mathbb{E}\{\varepsilon_{(i,k)}, \varepsilon_{(j,k)}\} = \text{corr}\{\varepsilon_{(i,k)}, \varepsilon_{(j,k)}\} = \langle \psi_{(i,k)}, \psi_{(j,k)} \rangle$ .

The cost function of fitting changes within  $\mathbf{S}$  therefore can be expressed using

above basis representation, for example:

$$\begin{aligned}\mathcal{L}^*(\mathbf{S}) &= \|\mathbf{x} - f_{(C)}\psi_{(C)} - f_{(L)}\psi_{(L)} - f_{(\tau^*)}\psi_{(\tau^*)}\|^2, \\ \mathcal{L}(\mathbf{S}, \emptyset) &= \|\mathbf{x} - x_{(C)}\psi_{(C)} - x_{(L)}\psi_{(L)}\|^2, \\ \mathcal{L}(\mathbf{S}, \tau^*) &= \|\mathbf{x} - x_{(C)}\psi_{(C)} - x_{(L)}\psi_{(L)} - x_{(\tau^*)}\psi_{(\tau^*)}\|^2, \\ \mathcal{L}(\mathbf{S}; \tau_{1:k}, \tau^*) &= \left\| \mathbf{x} - x_{(C)}\psi_{(C)} - x_{(L)}\psi_{(L)} - x_{(\tau^*)}\psi_{(\tau^*)} - \sum_{j=1}^k x_{(\tau_j)}\psi_{(\tau_j)} \right\|^2.\end{aligned}$$

**Lemma E.1.** *The following differences in cost follow chi-square distributions:*

$$\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau^*) \sim \chi_3^2$$

,

$$\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k}, \tau^*) \sim \chi_{k+3}^2$$

and

$$\mathcal{L}(\mathbf{S}, \emptyset) - \mathcal{L}(\mathbf{S}, \tau^*) \sim \chi_1^2(\nu),$$

where

$$\nu = \frac{(\theta_2 - 2\theta_1 + \theta_0)^2 n(n+1)(n-1)(2n^2+1)}{n^2 12(2n-1)(2n+1)}.$$

*Proof.* Applying the above properties and basis representation of loss functions, it is straightforward that

$$\begin{aligned}\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau^*) &= \|\mathbf{x} - f_{(C)}\psi_{(C)} - f_{(L)}\psi_{(L)} - f_{(\tau^*)}\psi_{(\tau^*)}\|^2 \\ &\quad - \left\| \mathbf{x} - x_{(C)}\psi_{(C)} - x_{(L)}\psi_{(L)} - x_{(\tau^*)}\psi_{(\tau^*)} \right\|^2 \\ &= \varepsilon_{(C)}^2 + \varepsilon_{(L)}^2 + \varepsilon_{(\tau^*)}^2 \sim \chi_3^2,\end{aligned}$$

and

$$\begin{aligned}\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_{1:k}, \tau^*) &= \|\mathbf{x} - f_{(C)}\psi_{(C)} - f_{(L)}\psi_{(L)} - f_{(\tau^*)}\psi_{(\tau^*)}\|^2 \\ &\quad - \left\| \mathbf{x} - x_{(C)}\psi_{(C)} - x_{(L)}\psi_{(L)} - x_{(\tau^*)}\psi_{(\tau^*)} - \sum_{j=1}^k x_{(\tau_j)}\psi_{(\tau_j)} \right\|^2 \\ &= \varepsilon_{(C)}^2 + \varepsilon_{(L)}^2 + \varepsilon_{(\tau^*)}^2 + \sum_{j=1}^k \varepsilon_{(\tau_j)}^2 \sim \chi_{k+3}^2.\end{aligned}$$

In addition,

$$\begin{aligned}\mathcal{L}(\mathbf{S}, \emptyset) - \mathcal{L}(\mathbf{S}, \tau^*) &= \|\mathbf{x} - f_{(C)}\psi_{(C)} - f_{(L)}\psi_{(L)} - f_{(\tau^*)}\psi_{(\tau^*)}\|^2 \\ &\quad - \|\mathbf{x} - x_{(C)}\psi_{(C)} - x_{(L)}\psi_{(L)}\|^2 \\ &= x_{(\tau^*)}^2 = \{f_{(\tau^*)} + \varepsilon_{(\tau^*)}\}^2 \sim \chi_1^2(f_{(\tau^*)}^2),\end{aligned}$$

where  $f_{\tau^*}^2 = \langle \mathbf{f}, \psi_{(\tau^*)} \rangle = (2n^2 + 1) \frac{(\theta_2 - 2\theta_1 + \theta_0)^2}{n^2} \frac{n(n+1)(n-1)}{12(2n-1)(2n+1)}$ .  $\square$

Moreover, one should noticed that if we consider  $\mathbf{S}'$  as a local region consisting no true changepoint, by similar arguments, we have the following lemma.

**Lemma E.2.**  $\mathcal{L}^*(\mathbf{S}') - \mathcal{L}(\mathbf{S}'; \tau_{1:k}) \sim \chi_{k+2}^2$  and  $\mathcal{L}^*(\mathbf{S}') - \mathcal{L}(\mathbf{S}'; \emptyset) \sim \chi_2^2$ .

### E.2. Maxima of correlated Gaussian variables.

In this step, we prove a lemma that provides the upper bound of probability tail for the maxima of a series of Gaussian random variables, i.e.,  $\max_{\tau_1 \in \mathbf{S}_{\tau_1}} \{\varepsilon_{(\tau_1)}\}$ ,  $\max_{\tau_2 \in \mathbf{S}_{\tau_2}} \{\varepsilon_{(\tau_2)}\}$  for given  $\tau_1$ , and  $\max_{\tau_3 \in \mathbf{S}_{\tau_3}} \{\varepsilon_{(\tau_3)}\}$  for given  $\tau_1$  and  $\tau_2$ .

We first introduce the following Lemma E.3, which is a direct adaptation from a result in Davies (1977).

**Lemma E.3.** Let  $\mathcal{G}(t)$  be a Gaussian process indexed by  $t \in [a, b]$ , with expectation 0 and covariance function  $\mathbb{E}[\mathcal{G}(t_1)\mathcal{G}(t_2)] = \rho(t_1, t_2)$ . Let

$$\rho_{11}(t_1) = \left. \frac{\partial^2 \rho(t_1, t_2)}{\partial t_2^2} \right|_{t_2=t_1}.$$

Then for any  $z > 0$ :

$$\mathbb{P}\left(\sup_t \mathcal{G}(t) > z\right) \leq \Phi(-z) + \frac{1}{2\pi} \exp\left(-\frac{z^2}{2}\right) \int_a^b |\rho_{11}(t)|^{1/2} dt, \quad (\text{E.1})$$

where  $\Phi(\cdot)$  denotes the cumulative distribution function of  $\mathcal{N}(0, 1)$ .

Based on Lemma E.3, we can prove the following useful lemmas.

**Lemma E.4.** There exists positive constants  $C_1$ , such that for any  $z > 0$ ,

$$\mathbb{P}\left(\max_{\tau_1 \in \mathbf{S}_{\tau_1}} \varepsilon_{(\tau_1)} > z\right) < C_1 \exp\left(-\frac{z^2}{2}\right) \log(2n), \quad (\text{E.2})$$

*Proof.* Note that the collection of random variables  $\{\varepsilon_{(i,1)}\}$  for  $i \in \mathbf{S}_{\tau_1}$ , which are  $\varepsilon$  projecting onto all the possible locations of  $\tau_1$ , are jointly Gaussian with covariance  $\mathbb{E}\{\varepsilon_{(i,1)}, \varepsilon_{(j,1)}\} = \langle \psi_{(i,1)}, \psi_{(j,1)} \rangle$ , as each of them is a linear combination of i.i.d Gaussian variables  $\varepsilon_1, \dots, \varepsilon_{2n}$ .

Let  $\mathbf{R}_{\tau_1} = [2, n-1] \cup [n+1, 2n]$ . Define a function  $\rho(x, y)$  on  $\mathbf{R}_{\tau_1} \times \mathbf{R}_{\tau_1}$  with continuous second derivatives with respect to both components, such that for any pair  $(i, j) \in \mathbf{S}_{\tau_1} \times \mathbf{S}_{\tau_1}$ , we have  $\rho(i, j) = \text{corr}(\varepsilon_{(i,1)}, \varepsilon_{(j,1)})$ . For example, we could let  $\rho(x, y)$  be the function that replace the discrete pair of variables  $(i, j)$  in the formula of  $\langle \psi_{(i,1)}, \psi_{(j,1)} \rangle$  by continuous pair of variables  $(x, y)$ . In this

way, by algebra calculation, we have

$$\rho(x, y) = \begin{cases} \sqrt{\frac{x(x-1)(n-y)}{y(y-1)(n-x)}} \frac{D_{n,1}(x, y)}{\sqrt{D_{n,1}(x)D_{n,1}(y)}}, & 2 \leq x \leq y \leq n-1; \\ \sqrt{\frac{(n-y)(x-\tau_1)}{(n-x)(y-\tau_1)}} \frac{E_{n,1}(x, y)}{\sqrt{E_{n,1}(x)E_{n,1}(y)}}, & x \leq n-1 \text{ and } y \geq n+1; \\ \sqrt{\frac{(2n-y)(2n-y+1)(x-n)}{(2n-x)(2n-x+1)(x-n)}} \frac{F_{n,1}(x, y)}{\sqrt{F_{n,1}(x)F_{n,1}(y)}}, & n+1 \leq x \leq y \leq 2n, \end{cases} \quad (\text{E.3})$$

where we write

$$D_{n,1}(x, y) = 2x - 2n - 3xy - 2xn + 7yn + xy^2 + 4xn^2 - 4xn^3 + 5yn^2 - y^2n + 12yn^3 - 2y^2 + 4n^2 - 4n^3 - 9y^2n^2 - 4xyn^2 + 5xy^2n - 5xyn + 2;$$

$$D_{n,1}(x) = D_{n,1}(x, x), \quad D_{n,1}(y) = D_{n,1}(y, y);$$

$$E_{n,1}(x, y) = 2x + 2y - 4n - xy - 7xn + 9yn + 4xn^2 + 4xn^3 + 4yn^2 + 36yn^3 + 7n^2 - 8n^3 - 12n^4 - 20xyn^2 + 2,$$

$$E_{n,1}(x) = E_{n,1}(x, x), \quad E_{n,1}(y) = E_{n,1}(y, y);$$

$$F_{n,1}(x, y) = 2y - 2n - 3xy + 7xn + 8yn + x^2y - xn^2 - x^2n - 8xn^3 - 2yn^2 - 8yn^3 - 2x^2 - 6n^2 + 2n^3 + 4n^4 + x^2n^2 + 16xyn^2 - 5x^2yn + xyn + 2;$$

$$F_{n,1}(x) = F_{n,1}(x, x), \quad F_{n,1}(y) = F_{n,1}(y, y).$$

If  $x > y$ , note that  $\rho(x, y) = \rho(y, x)$ , which completes the definition of  $\rho(x, y)$ . Therefore, we can construct a Gaussian process  $\mathcal{G}(t)$  indexed by  $\mathbf{R}_{\tau_1}$ , with mean 0 and covariance function  $\mathbb{E}[\mathcal{G}(t_1)\mathcal{G}(t_2)] = \rho(t_1, t_2)$ , such that  $\mathcal{G}(i) = \varepsilon_{(i)}$  for  $i \in \mathbf{S}_{\tau_1}$ .

Notice that for any  $2 \leq x \leq n-1$ , we obtain

$$\begin{aligned} \rho_{11}(x) &= \frac{1}{4} \left( \frac{1}{n-x} + \frac{1}{x} + \frac{1}{x-1} \right)^2 - \frac{1}{2} \left\{ \frac{1}{(n-x)^2} - \frac{1}{x^2} - \frac{1}{(x-1)^2} \right\} \\ &\quad + \frac{1}{2D_{n,1}(x)} \left( \frac{1}{n-x} + \frac{1}{x} + \frac{1}{x-1} \right) \left\{ \frac{\partial D_{n,1}(x)}{\partial x} - 2 \frac{\partial D_{n,1}(x, y)}{\partial y} \Big|_{y=x} \right\} \\ &\quad + \frac{1}{2D_{n,1}(x)} \left\{ 2 \frac{\partial^2 D_{n,1}(x, y)}{\partial y^2} \Big|_{y=x} - \frac{\partial^2 D_{n,1}(x)}{\partial x^2} \right\} \\ &\quad - \frac{1}{2D_{n,1}^2(x)} \left( \frac{1}{n-x} + \frac{1}{x} + \frac{1}{x-1} \right) \left\{ 2 \frac{\partial D_{n,1}(x)}{\partial x} \frac{\partial D_{n,1}(x, y)}{\partial y} \Big|_{y=x} - \frac{3}{2} \left( \frac{\partial D_{n,1}(x)}{\partial x} \right)^2 \right\}. \end{aligned}$$

Let  $g(x) = \frac{2}{x-1} + \frac{2}{n-x}$ , applying the Maple program **Psdgcd** (Han et al. , 2016) which can prove polynomial inequalities using symbolic computation, we

prove that:

$$\left| \frac{\partial D_{n,1}(x, y)/\partial y|_{y=x}}{D_{n,1}(x)} \right| \leq g(x), \quad \left| \frac{\partial D_{n,1}(x)/\partial x}{D_{n,1}(x)} \right| \leq g(x), \quad (\text{E.4})$$

and

$$\left| \frac{\partial^2 D_{n,1}(x, y)/\partial y^2|_{y=x}}{D_{n,1}(x)} \right| \leq g(x), \quad \left| \frac{\partial^2 D_{n,1}(x)/\partial x^2}{D_{n,1}(x)} \right| \leq g(x). \quad (\text{E.5})$$

Hence, for  $2 \leq x \leq n-1$ , the following inequality holds for some  $c > 0$ ,

$$|\rho_{11}(x)| < c \max \left\{ \frac{1}{(x-1)^2}, \frac{1}{(n-x)^2} \right\}.$$

In a similar way, we can also obtain that for  $\tau_1 + 1 \leq x \leq n-1$ , the following inequality holds for some  $c > 0$ ,

$$|\rho_{11}(x)| < c \max \left\{ \frac{1}{(x-n)^2}, \frac{1}{(2n-x)^2} \right\}.$$

Altogether, we have that there exists an absolute constant  $C_1$  that does not depend on  $n, \tau_1$ , such that for any  $x \in \mathbf{R}_{\tau_1}$  the following inequality holds:

$$|\rho_{11}(x)| \leq 4\pi^2 C_1^2 \max \left\{ \frac{1}{(x-1)^2}, \frac{1}{(x-n)^2}, \frac{1}{(2n-x)^2} \right\}.$$

By Lemma E.3, we have

$$\begin{aligned} \mathbb{P} \left( \sup_{t \in \mathbf{R}_{\tau_1}} \mathcal{G}(t) > z \right) &= \frac{1}{2\pi} \exp \left( -\frac{z^2}{2} \right) \int_{\mathbf{R}_{\tau_1}} |\rho_{11}(t)|^{1/2} dt \\ &= C_1 \exp \left( -\frac{z^2}{2} \right) \log(2n-4). \end{aligned}$$

As a result, we obtain

$$\mathbb{P} \left( \max_{\tau_2 \in \mathbf{S}_{\tau_1}} \varepsilon_{(\tau_2)} > z \right) \leq \mathbb{P} \left( \sup_{t \in \mathbf{R}_{\tau_1}} \mathcal{G}(t) > z \right) = C_1 \exp \left( -\frac{z^2}{2} \right) \log(2n-4),$$

which proves (E.2). □

**Lemma E.5.** *There exists absolute constants  $C_2, C_3$ , such that for any given  $\tau_1$ ,*

$$\mathbb{P} \left( \max_{\tau_2 \in \mathbf{S}_{\tau_2}} \varepsilon_{(\tau_2)} > z \right) < C_2 \exp \left( -\frac{z^2}{2} \right) \log(2n), \quad (\text{E.6})$$

and for any given  $\tau_1, \tau_2$ ,

$$\mathbb{P} \left( \max_{\tau_3 \in \mathbf{S}_{\tau_3}} \varepsilon_{(\tau_3)} > z \right) < C_3 \exp \left( -\frac{z^2}{2} \right) \log(2n). \quad (\text{E.7})$$

*Proof.* For a given  $\tau_1$ , note that  $\{\varepsilon_{(i,2)}\}_{i \in \mathbf{s}_{\tau_2}}$  are jointly Gaussian with covariance  $\mathbb{E}\{\varepsilon_{(i,2)}, \varepsilon_{(j,2)}\} = \langle \psi_{(i,2)}, \psi_{(j,2)} \rangle$ . Due to the symmetry of the local region  $\mathbf{x}$  (having a change in the middle), we only need to deal with  $2 \leq \tau_1 \leq n-1$ .

Let  $\mathbf{R}_{\tau_2} = [2, \tau_1 - 1] \cup [\tau_1 + 1, n - 1] \cup [n + 1, 2n]$ . Define  $\rho(x, y) : \mathbf{R}_{\tau_2} \times \mathbf{R}_{\tau_2} \rightarrow [-1, 1]$  as the function that replace the discrete pair  $(i, j)$  in the formula of  $\langle \psi_{(i,2)}, \psi_{(j,2)} \rangle$  by continuous pair  $(x, y)$ . In this way, by algebra calculation, we have

$$\rho(x, y) = \begin{cases} \sqrt{\frac{x(x-1)(\tau_1-y)}{y(y-1)(\tau_1-x)}} \frac{D_{n,2}(x, y)}{\sqrt{D_{n,2}(x)D_{n,2}(y)}}, & 2 \leq x \leq y \leq \tau_1 - 1; \\ \sqrt{\frac{(n-y)(x-\tau_1)}{(n-x)(y-\tau_1)}} \frac{E_{n,2}(x, y)}{\sqrt{E_{n,2}(x)E_{n,2}(y)}}, & \tau_1 + 1 \leq x \leq y \leq n - 1; \\ \sqrt{\frac{(2n-y)(2n-y+1)(x-n)}{(2n-x)(2n-x+1)(x-n)}} \frac{F_{n,2}(x, y)}{\sqrt{F_{n,2}(x)F_{n,2}(y)}}, & n + 1 \leq x \leq y \leq 2n. \end{cases} \quad (\text{E.8})$$

where we write

$$\begin{aligned} D_{n,2}(x, y) = & 2n - 11n^2\tau_1^2 + 11n^2\tau_1^3 - 22n^3\tau_1^2 - 5n^2\tau_1^4 + 13n^3\tau_1^3 - 8n^4\tau_1^2 + 2n^2y^2 - \\ & 4n^3y^2 + 4n^4y^2 - 3\tau_1^2y^2 + 6\tau_1^3y^2 - 3\tau_1^4y^2 + 2nx - 7n\tau_1^2 + 7n^2\tau_1 + 6n\tau_1^3 + 5n^3\tau_1 - \\ & n\tau_1^4 + 12n^4\tau_1 - 2n^2x + 4n^3x - 4n^4x - 2ny^2 + 3\tau_1^2y - 6\tau_1^3y + 3\tau_1^4y - 2n^2 + \\ & 4n^3 - 4n^4 - 7n\tau_1^2x + 7n^2\tau_1x + 6n\tau_1^3x + 5n^3\tau_1x - n\tau_1^4x + 12n^4\tau_1x + 15n\tau_1^2y - \\ & 4n^2\tau_1y - 28n\tau_1^3y + 8n^3\tau_1y + 9n\tau_1^4y - 8n^4\tau_1y - 2nxy^2 + 3\tau_1xy^2 + 6\tau_1^2xy - 3\tau_1^3xy - \\ & 11n^2\tau_1^2x + 11n^2\tau_1^3x - 22n^3\tau_1^2x - 5n^2\tau_1^4x + 13n^3\tau_1^3x - 8n^4\tau_1^2x - 4n\tau_1^2y^2 - 7n^2\tau_1y^2 + \\ & 29n^2\tau_1^2y + 12n\tau_1^3y^2 - 32n^2\tau_1^3y - 5n^3\tau_1y^2 + 31n^3\tau_1^2y - 6n\tau_1^4y^2 + 15n^2\tau_1^4y - 39n^3\tau_1^3y - \\ & 12n^4\tau_1y^2 + 24n^4\tau_1^2y + 2n^2xy^2 - 4n^3xy^2 + 4n^4xy^2 - 6\tau_1^2xy^2 + 3\tau_1^3xy^2 + 4n\tau_1y - \\ & 3\tau_1xy - 8n^2\tau_1^2y^2 + 9n^2\tau_1^3y^2 + 9n^3\tau_1^2y^2 - 11n\tau_1xy - 14n^2\tau_1^2xy^2 + 9n\tau_1xy^2 + \\ & 18n\tau_1^2xy - 19n^2\tau_1xy - 7n\tau_1^3xy - 13n^3\tau_1xy - 8n^4\tau_1xy - 13n\tau_1^2xy^2 + 14n^2\tau_1xy^2 + \\ & 20n^2\tau_1^2xy + 6n\tau_1^3xy^2 - 5n^2\tau_1^3xy + 4n^3\tau_1xy^2 + 13n^3\tau_1^2xy, \end{aligned}$$

$$D_{n,2}(x) = D_{n,2}(x, x), D_{n,2}(y) = D_{n,2}(y, y);$$

$$\begin{aligned} E_{n,2}(x, y) = & 10x^2y^2n^2\tau_1 - 5x^2y^2n^2 - 6x^2y^2n\tau_1^2 + 8x^2y^2n\tau_1 - x^2y^2n - 3x^2y^2\tau_1^2 + \\ & 3x^2y^2\tau_1 - 8x^2yn^3\tau_1 + 4x^2yn^3 - 6x^2yn^2\tau_1^2 - 4x^2yn^2\tau_1 + 5x^2yn^2 + 6x^2yn\tau_1^3 - \\ & 3x^2yn\tau_1^2 - 9x^2yn\tau_1 + 3x^2yn + 3x^2y\tau_1^3 - 3x^2y\tau_1 - 8x^2n^4\tau_1 + 4x^2n^4 + 21x^2n^3\tau_1^2 - \\ & 13x^2n^3\tau_1 - 4x^2n^3 - 9x^2n^2\tau_1^3 + 15x^2n^2\tau_1^2 - 10x^2n^2\tau_1 + 2x^2n^2 - 6x^2n\tau_1^3 + 15x^2n\tau_1^2 - \\ & 5x^2n\tau_1 - 2x^2n - 3x^2\tau_1^3 + 3x^2\tau_1^2 - 18xy^2n^3\tau_1 + 9xy^2n^3 + 10xy^2n^2\tau_1^2 - 17xy^2n^2\tau_1 + \\ & 6xy^2n^2 + 8xy^2n\tau_1^2 - 13xy^2n\tau_1 + 3xy^2n + 3xy^2\tau_1^2 - 3xy^2\tau_1 + 24xyn^4\tau_1 - 12xyn^4 - \\ & 8xyn^3\tau_1^2 + 22xyn^3\tau_1 - 9xyn^3 - 4xyn^2\tau_1^2 + 23xyn^2\tau_1 - 12xyn^2 - 6xyn\tau_1^3 - 3xyn\tau_1^2 + \\ & 12xyn\tau_1 - 3xyn - 3xy\tau_1^3 + 3xy\tau_1 - 8xn^4\tau_1^2 + 4xn^4\tau_1 - 13xn^3\tau_1^2 + 17xn^3\tau_1 + \\ & 9xn^2\tau_1^3 - 19xn^2\tau_1^2 + 8xn^2\tau_1 + 6xn\tau_1^3 - 11xn\tau_1^2 + 7xn\tau_1 + 3x\tau_1^3 - 3x\tau_1^2 + 9y^2n^3\tau_1^2 - \\ & 9y^2n^3 - 5y^2n^2\tau_1^3 + y^2n^2\tau_1^2 + 5y^2n^2\tau_1 - y^2n^2 - y^2n\tau_1^3 + 2y^2n\tau_1^2 + y^2n\tau_1 - 2y^2n - \\ & 12yn^4\tau_1^2 + 12yn^4 + 4yn^3\tau_1^3 - 5yn^3\tau_1^2 - 4yn^3\tau_1 + 5yn^3 + 5yn^2\tau_1^3 - 7yn^2\tau_1^2 - 5yn^2\tau_1 + \\ & 7yn^2 + 3yn\tau_1^3 - 3yn\tau_1 + 4n^4\tau_1^3 + 4n^4\tau_1^2 - 4n^4\tau_1 - 4n^4 - 4n^3\tau_1^3 - 4n^3\tau_1^2 + 4n^3\tau_1 + \end{aligned}$$

$$4n^3 + 2n^2\tau_1^3 + 2n^2\tau_1^2 - 2n^2\tau_1 - 2n^2 - 2n\tau_1^3 - 2n\tau_1^2 + 2n\tau_1 + 2n;$$

$$E_{n,2}(x) = E_{n,2}(x, x), \quad E_{n,2}(y) = E_{n,2}(y, y);$$

$$\begin{aligned} F_{n,2}(x, y) = & 2y - 4n + 2\tau_1 + 17n^2\tau_1^2 - 17n^2\tau_1^3 + 53n^3\tau_1^2 - 32n^3\tau_1^3 + 32n^4\tau_1^2 + 16n^4\tau_1^3 - \\ & 52n^5\tau_1^2 + 28n^5\tau_1^3 - 44n^6\tau_1^2 - 3xy + 7xn + 6yn + 2y\tau_1 + 3n\tau_1 + x^2y - 8xn^2 + x^2n + \\ & 7xn^3 - 8xn^4 - 14xn^5 + 16xn^6 - 6yn^2 + 6yn^3 - 12yn^4 - 12yn^5 + 16yn^6 - 2x^2\tau_1 - 5y\tau_1^2 + \\ & y\tau_1^3 - n\tau_1^2 - 2n^2\tau_1 + 4n\tau_1^3 - 23n^3\tau_1 - 44n^4\tau_1 + 4n^5\tau_1 + 44n^6\tau_1 + 16n^7\tau_1 - 2x^2 - 6n^4 + \\ & 12n^5 + 4n^6 - 8n^7 - 5\tau_1^2 + \tau_1^3 - 2x^2n^2 + x^2n^3 + 4x^2n^4 - 2x^2n^5 + 5x^2\tau_1^2 - x^2\tau_1^3 + 9xyn^2 - \\ & 6x^2yn - 8xyn^3 + 30xyn^4 - 32xyn^5 + 12xy\tau_1^2 + x^2y\tau_1 - 6xy\tau_1^3 - 22xn\tau_1^2 + 21xn^2\tau_1 - \\ & 6x^2n\tau_1 + 8xn\tau_1^3 + 39xn^3\tau_1 - xn^4\tau_1 - 58xn^5\tau_1 - 32xn^6\tau_1 - 35yn\tau_1^2 + 36yn^2\tau_1 + \\ & 18yn\tau_1^3 + 49yn^3\tau_1 - 6yn^4\tau_1 - 62yn^5\tau_1 - 32yn^6\tau_1 + 7x^2yn^2 - 12x^2yn^3 + 10x^2yn^4 - \\ & 7x^2y\tau_1^2 + 5x^2y\tau_1^3 - 41xn^2\tau_1^2 + 13x^2n\tau_1^2 - 15x^2n^2\tau_1 + 35xn^2\tau_1^3 - 36xn^3\tau_1^2 - 10x^2n\tau_1^3 - \\ & 10x^2n^3\tau_1 - 11xn^3\tau_1^3 + 53xn^4\tau_1^2 + 5x^2n^4\tau_1 - 38xn^4\tau_1^3 + 70xn^5\tau_1^2 + 4x^2n^5\tau_1 - \\ & 59yn^2\tau_1^2 + 37yn^2\tau_1^3 - 35yn^3\tau_1^2 - 18yn^3\tau_1^3 + 64yn^4\tau_1^2 - 38yn^4\tau_1^3 + 70yn^5\tau_1^2 + 4xyn - \\ & 3xy\tau_1 + 7xn\tau_1 + 13yn\tau_1 + 18x^2n^2\tau_1^2 - 2x^2n^2\tau_1^3 - x^2n^3\tau_1^2 + 7x^2n^3\tau_1^3 - 11x^2n^4\tau_1^2 - \\ & 11xyn\tau_1 + 13x^2yn^2\tau_1^2 - 17x^2yn^2\tau_1^3 + 37x^2yn^3\tau_1^2 + 29xyn\tau_1^2 - 13xyn^2\tau_1 + 2x^2yn\tau_1 - \\ & 25xyn\tau_1^3 + 37xyn^3\tau_1 + 94xyn^4\tau_1 + 64xyn^5\tau_1 - 10xyn^2\tau_1^2 + 5x^2yn\tau_1^2 - 8x^2yn^2\tau_1 + \\ & 15xyn^2\tau_1^3 - 77xyn^3\tau_1^2 - 23x^2yn^3\tau_1 + 58xyn^3\tau_1^3 - 122xyn^4\tau_1^2 - 20x^2yn^4\tau_1 + 2, \end{aligned}$$

$$F_{n,2}(x) = F_{n,2}(x, x), \quad F_{n,2}(y) = F_{n,2}(y, y).$$

The formula of  $p(x, y)$  for other cases, such as  $x < \tau_1 < y$  and  $x < n < y$  can be derived similarly. Moreover, if  $x > y$ , note that  $\rho(x, y) = \rho(y, x)$ . Therefore, we can define a Gaussian process  $\mathcal{G}(t)$  indexed by  $\mathbf{R}_{\tau_2}$ , with mean 0 and covariance function  $\mathbb{E}[\mathcal{G}(t_1)\mathcal{G}(t_2)] = \rho(t_1, t_2)$ , such that  $\mathcal{G}(i) = \varepsilon_{(i,2)}$  for  $i \in \mathbf{S}_{\tau_2}$ .

Notice that for any  $2 \leq x \leq \tau_1 - 1$ , we obtain

$$\begin{aligned} \rho_{11}(x) = & \frac{1}{4} \left( \frac{1}{\tau_1 - x} + \frac{1}{x} + \frac{1}{x-1} \right)^2 - \frac{1}{2} \left\{ \frac{1}{(\tau_1 - x)^2} - \frac{1}{x^2} - \frac{1}{(x-1)^2} \right\} \\ & + \frac{1}{2D_{n,2}(x)} \left( \frac{1}{\tau_1 - x} + \frac{1}{x} + \frac{1}{x-1} \right) \left\{ \frac{\partial D_{n,2}(x)}{\partial x} - 2 \frac{\partial D_{n,2}(x, y)}{\partial y} \Big|_{y=x} \right\} \\ & + \frac{1}{2D_{n,2}(x)} \left\{ 2 \frac{\partial^2 D_{n,2}(x, y)}{\partial y^2} \Big|_{y=x} - \frac{\partial^2 D_{n,2}(x)}{\partial x^2} \right\} \\ & - \frac{1}{2D_n^2(x)} \left( \frac{1}{\tau_1 - x} + \frac{1}{x} + \frac{1}{x-1} \right) \left\{ 2 \frac{\partial D_{n,2}(x)}{\partial x} \frac{\partial D_{n,2}(x, y)}{\partial y} \Big|_{y=x} - \frac{3}{2} \left( \frac{\partial D_{n,2}(x)}{\partial x} \right)^2 \right\}. \end{aligned}$$

Let  $g(x) = \frac{2}{x-1} + \frac{2}{\tau_1 - x}$ , with the help of **Psdgcd**, we can prove that:

$$\left| \frac{\partial D_{n,2}(x, y)/\partial y|_{y=x}}{D_{n,2}(x)} \right| \leq g(x), \quad \left| \frac{\partial D_{n,2}(x)/\partial x}{D_{n,2}(x)} \right| \leq g(x), \quad (\text{E.9})$$

and

$$\left| \frac{\partial^2 D_{n,2}(x, y)/\partial y^2|_{y=x}}{D_{n,2}(x)} \right| \leq g(x), \quad \left| \frac{\partial^2 D_{n,2}(x)/\partial x^2}{D_{n,2}(x)} \right| \leq g(x). \quad (\text{E.10})$$

Hence, for  $2 \leq x \leq \tau_1 - 1$ , the following inequality holds,

$$|\rho_{11}(x)| < c \max \left\{ \frac{1}{(x-1)^2}, \frac{1}{(\tau_1-x)^2} \right\}$$

where  $c > 0$  is a constant. In a similar way, we can also obtain that for  $\tau_1 + 1 \leq x \leq n - 1$

$$|\rho_{11}(x)| < c \max \left\{ \frac{1}{(n-x)^2}, \frac{1}{(x-\tau_1)^2} \right\}$$

and for  $n + 1 \leq x < y \leq 2n$

$$|\rho_{11}(x)| < c \max \left\{ \frac{1}{(x-n)^2}, \frac{1}{(2n-x)^2} \right\},$$

where  $c$  is some universal positive constants.

Altogether, we have that there exists an absolute constant  $C_2$  that does not depend on  $n, \tau_1$ , such that for any  $x \in \mathbf{R}_{\tau_2}$  the following inequality holds:

$$|\rho_{11}(x)| \leq 4\pi^2 C_2^2 \max \left\{ \frac{1}{(x-1)^2}, \frac{1}{(\tau_1-x)^2}, \frac{1}{(x-n)^2}, \frac{1}{(2n-x)^2} \right\}.$$

By Lemma E.3, we have

$$\begin{aligned} \mathbb{P} \left( \sup_{t \in \mathbf{R}_{\tau_2}} \mathcal{G}(t) > z \right) &= \frac{1}{2\pi} \exp \left( -\frac{z^2}{2} \right) \int_{\mathbf{R}_{\tau_2}} |\rho_{11}(t)|^{1/2} dt \\ &= C_2 \exp \left( -\frac{z^2}{2} \right) \log(2n-6). \end{aligned}$$

As a result, we obtain

$$\mathbb{P} \left( \max_{\tau_2 \in \mathbf{S}_{\tau_2}} \varepsilon_{(\tau_2)} > z \right) \leq \mathbb{P} \left( \sup_{t \in \mathbf{R}_{\tau_2}} \mathcal{G}(t) > z \right) = C_2 \exp \left( -\frac{z^2}{2} \right) \log(2n-6),$$

which proves (E.6).

The proof of (E.7) can be obtained similarly, thus is omitted here.  $\square$

### E.3. Tight probabilistic bounds for fitting too many changes.

In this step, we provide the following two lemmas, which is much tighter than simply applying Bonferroni correction for the probability of maximum over the events that we fit too many changes.

**Lemma E.6.** *As long as  $\gamma > \max \left\{ 2 \log(2n) + 32 \log(C_1''' \log(2n)), 2 \log(2n) + 972(2m^* + 1) \right\}$ , where  $C_1'''$  is a positive constant only related to  $C_1$  in (E.2), we have*

$$\mathbb{P} \left( \max_{\tau_1} \{ \mathcal{L}^*(\mathbf{S}) - \mathcal{L}_n(\mathbf{S}; \tau_1, \tau^*) \} \geq \frac{\gamma - 2 \log(2n)}{6} \right) \leq \exp \left( -\frac{\gamma - 2 \log(2n)}{24(2m^* + 1)} \right)$$

*Proof.* By the orthonormality of the basis and the properties (i),

$$\begin{aligned}\mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_1) &\leq \mathcal{L}^*(\mathbf{S}) - \mathcal{L}(\mathbf{S}; \tau_1, \tau^*) \\ &= \varepsilon_{(C)}^2 + \varepsilon_{(L)}^2 + \varepsilon_{(\tau^*)}^2 + \varepsilon_{(\tau_1)}^2\end{aligned}$$

This implies

$$\begin{aligned}\mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_1 \in \mathbf{S}_{\tau_1}} \mathcal{L}(\mathbf{S}; \tau_1) \geq z\right) &\leq \mathbb{P}\left(\varepsilon_{(C)}^2 + \varepsilon_{(L)}^2 + \varepsilon_{(\tau^*)}^2 + \max_{\tau_1 \in \mathbf{S}_{\tau_1}} \left(\varepsilon_{(\tau_1)}^2\right) \geq z\right) \\ &\leq \mathbb{P}\left(\varepsilon_{(C)}^2 + \varepsilon_{(L)}^2 + \varepsilon_{(\tau^*)}^2 + \varepsilon_{(\tau_1)}^2 + \sup_{t \in \mathbf{R}_{\tau_1}} \mathcal{G}^2(t) \geq z\right),\end{aligned}$$

where  $\mathcal{G}(t)$  is the continuous Gaussian process constructed based on  $\varepsilon_{(\tau_1)}, \tau_1 \in \mathbf{S}_{\tau_1}$  as in the proof of Lemma E.4.

Let  $Z_1 = \varepsilon_{(C)}^2 + \varepsilon_{(L)}^2 + \varepsilon_{(\tau^*)}^2$  and  $Z_2 = \sup_{t \in \mathbf{R}_{\tau_1}} \mathcal{G}^2(t)$ . Note that  $\varepsilon_{(C)}, \varepsilon_{(L)}$  and  $\varepsilon_{(\tau^*)}$  are i.i.d random variables with  $\mathcal{N}(0, 1)$  distribution, and are all independent to  $\varepsilon_{(\tau_1)}$  for any  $\tau_1 \in \mathbf{S}_{\tau_1}$ . Therefore  $Z_1 \sim \chi_3^2$  and is independent to  $Z_2$ .

Using the arguments from the proof of Lemma 1 in [Laurent and Massart \(2000\)](#), we can upper bound the logarithm of Laplace transform of  $Z_1$ :

$$\log\{\mathbb{E} \exp(u(Z_1 - 3))\} \leq \frac{3u^2}{1 - 2u}, \quad \text{for } 0 < u < 1/2.$$

Note that for all  $z$

$$\mathbb{P}\left(\sup_{t \in \mathbf{R}_{\tau_1}} \mathcal{G}(t) \geq z\right) \leq \mathbb{P}\left(\sup_{t \in \mathbf{R}_{\tau_1}} |\mathcal{G}(t)| \geq z\right) \leq 2\mathbb{P}\left(\sup_{t \in \mathbf{R}_{\tau_1}} \mathcal{G}(t) \geq z\right).$$

Therefore we have the probability density function  $f(z)$  for  $\sup_{t \in \mathbf{R}_{\tau_1}} |\mathcal{G}(t)|$  is upper bounded by  $C'_1 \log(2n - 6)x \exp(-z^2/2)$  when  $z$  is large enough, where  $C'_1$  is a positive constant only depending on  $C_1$ . This further leads to an upper bound on the Laplace transform of  $Z_2$ :

$$\log\{\mathbb{E} \exp(uZ_2)\} \leq \log(C''_1 \log(2n - 6)) - \frac{\log(1 - 2u)}{2}.$$

For  $0 < u < 1/2$ , we have

$$\begin{aligned}\log \mathbb{E} \exp[u\{Z_2 - 1 - \log(C''_1 \log(2n - 6))\}] &\leq -u - \frac{\log(1 - 2u)}{2} \\ &\leq \frac{u^2}{1 - 2u}\end{aligned}$$

As a result, let  $Z = Z_1 + Z_2 - 4 - \log(C''_1 \log(2n - 6))$ , then

$$\log\{\mathbb{E} \exp(u(Z))\} \leq \frac{5u^2}{1 - 2u}$$

By Lemma 8.2 in Birgé (2001), if

$$\log(\mathbb{E}e^{uZ}) \leq \frac{v^2 u^2}{1 - bu}, \quad \text{for } 0 < t < b^{-1}$$

then for any positive  $x$ ,

$$\mathbb{P}(Z \geq bx + 2v\sqrt{x}) \leq \exp(-x).$$

Hence, we have for any given  $\tau_1$ , as long as  $z \geq 162$ ,

$$\begin{aligned} \mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_1 \in \mathbf{S}_{\tau_1}} \mathcal{L}(\mathbf{S}; \tau_1, \tau_2) \geq z\right) &\leq \mathbb{P}(Z \geq z - 4 - \log(C_1'' \log(2n - 4))) \\ &\leq \mathbb{P}(Z \geq z - (C_1'' + 4) - \log \log(2n)) \\ &\leq C_1''' \log(2n) \exp\left(-\frac{4z}{9}\right), \end{aligned}$$

where  $C_1'''$  is a positive constant only depend on  $C_1$ .

Taking  $z = \frac{\gamma - 2 \log(2n)}{6(2m^* + 1)}$ , then as long as

$$\gamma > \max \left\{ 2 \log(2n) + 32 \log(C_1''' \log(2n)), 2 \log(2n) + 972(2m^* + 1) \right\}$$

we have

$$C_1''' \log(2n) \exp\left(-\frac{4z}{9}\right) \leq \exp\left(-\frac{\gamma - 2 \log(2n)}{24(2m^* + 1)}\right),$$

which completes the proof.  $\square$

**Lemma E.7.** *As long as  $\gamma > \max\{240, 24 \log(2C_2''' \log(2n))\}$ , where  $C_2'''$  is a positive constant only related to  $C_2$  in (E.6), we have*

$$\mathbb{P}\left(\max_{\tau_1, \tau_2} \{\mathcal{L}^*(\mathbf{S}) - \mathcal{L}_n(\mathbf{S}; \tau_1, \tau_2)\} \geq \gamma - \frac{\gamma - 2 \log(2n)}{6}\right) \leq \exp\left(-\frac{\gamma - 3 \log(2n)}{3}\right)$$

*Proof.* Using a similar argument to that of the proof of Lemma E.6, we have that for any given  $\tau_1$  and  $\tau_2$ , as long as  $z \geq 200$ ,

$$\mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_2 \in \mathbf{S}_{\tau_1}} \mathcal{L}(\mathbf{S}; \tau_1, \tau_2) \geq z\right) \leq C_2''' \log(2n) \exp\left(-\frac{9z}{20}\right),$$

where  $C_2'''$  is a positive constant only depends on  $C_2$ .

Consider all the  $2n - 2$  possible locations for the first change  $\tau_1$ , by Bonferroni correction,

$$\begin{aligned} \mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_1, \tau_2} \mathcal{L}(\mathbf{S}; \tau_1, \tau_2) \geq z\right) &\leq (2n - 2) \mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_2 \in \mathbf{S}_{\tau_1}} \mathcal{L}(\mathbf{S}; \tau_1, \tau_2) \geq z\right) \\ &\leq C_2''' 2n \log(2n) \exp\left(-\frac{9z}{20}\right). \end{aligned}$$

Taking  $z = \gamma - \frac{\gamma - 2 \log(2n)}{6} = \frac{5\gamma + 2 \log(2n)}{6}$ , then as long as

$$\gamma > 24 \log(2C_2''' \log(2n)),$$

we have

$$C_2''' 2n \log(2n) \exp\left(-\frac{9z}{20}\right) \leq \exp\left(-\frac{\gamma - 3 \log(2n)}{3}\right),$$

which completes the proof.  $\square$

**Lemma E.8.** *As long as  $\gamma > \max\{132, 24 \log(2C_3''' \log(2n))\}$ , where  $C_3'''$  is a positive constant only related to  $C_3$  in (E.7) we have*

$$\mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_{1:3}} \mathcal{L}(\mathbf{S}; \tau_{1:3}, \tau^*) \geq 2\gamma - \frac{\gamma - 2 \log(2n)}{6}\right) \leq \exp\left(-\frac{\gamma - 3 \log(2n)}{3}\right).$$

*Proof.* Using the similar argument as in the proof of Lemma E.7, we have for any given  $\tau_1$  and  $\tau_2$ , as long as  $z \geq 242$ ,

$$\mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_3} \mathcal{L}(\mathbf{S}; \tau_1, \tau_2) \geq z\right) \leq C_3''' \log(2n) \exp\left(-\frac{10z}{22}\right),$$

where  $C_3'''$  is a positive constant only depends on  $C_3$ .

Consider all the  $(2n-2) \times (2n-3)$  possible locations for the first two changes  $\tau_1$  and  $\tau_2$ , by Bonferroni correction,

$$\begin{aligned} \mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_1, \tau_2, \tau_3} \mathcal{L}(\mathbf{S}; \tau_1, \tau_2, \tau_3) \geq z\right) &\leq (2n-2) \mathbb{P}\left(\mathcal{L}^*(\mathbf{S}) - \min_{\tau_3} \mathcal{L}(\mathbf{S}; \tau_1, \tau_2, \tau_3) \geq z\right) \\ &\leq C_3''' (2n)^2 \log(2n) \exp\left(-\frac{10z}{22}\right). \end{aligned}$$

Taking  $z = 2\gamma - \frac{\gamma - 2 \log(2n)}{6} = \frac{11\gamma + 2 \log(2n)}{6}$ , as long as

$$\gamma > 24 \log(2C_3''' \log(2n)),$$

we have

$$C_3''' (2n)^2 \log(2n) \exp\left(-\frac{10z}{22}\right) \leq \exp\left(-2\frac{(\gamma - 3 \log(2n))}{3}\right),$$

which completes the proof.  $\square$

## References

- BARANOWSKI, R., CHEN, Y. AND FRYZLEWICZ, P. (2019). Narrowest-Over-Threshold detection of multiple change-points and change-point-like features. *J. Royal. Statist. Soc. Ser. B.* **81**, 649–672.

- BIRGÉ, L. (2001). An alternative point of view on Lepski's method. In *State of the Art in Probability and Statistics. IMS Lecture Notes*. **37**, 113–133.
- BIRGÉ, L. AND MASSART, P. (2001). Gaussian model selection. *Ann. J. Eur. Math. Soc.* **3**, 203–268.
- BOYSEN, L., KEMPE, A., LIEBSCHER, V., MUNK, A., AND WITTICH, O. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *Ann. Statist.* **37**, 157–183.
- BOYSEN, L., BRUNS, S. AND MUNK, A. (2009) Jump estimation in inverse regression. *Electron. J. Statist.* **3**, 1322–1359.
- BOYSEN, L., LIEBSCHER, V., MUNK, A., AND WITTICH, O. (2007) Scale space consistency of piecewise constant least squares estimators another look at the regressogram. *IMS Lecture Notes - Monograph Series 2007*, 65–84.
- CLEYNEN, A., DUDOIT, S., AND ROBIN, S. (2014). Comparing segmentation methods for genome annotation based on RNA-seq data. *J. Agric. Biol. Environ. Stat.* **19**, 101–118
- CRIBBEN, I. AND YU, Y. (2017). Estimating whole brain dynamics using spectral clustering. *J. Royal. Statist. Soc. Ser. C.* **66**, 607–627
- DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika.* **64**, 247–254.
- EICHINGER, B. AND C. KIRCH (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli.* **24**, 526–564.
- FAGAN, B. T., KNIGHT, M.I., MACKAY, N. J., AND WOOD, A. J. (2019) Change-point analysis of historical battle deaths. *arXiv preprint*. arXiv:1909.02626.
- FEARNHEAD, P., MAIDSTONE, R., AND LETCHFORD, A. (2019). Detecting changes in slope With an  $L_0$  penalty. *J. Comput. Graph. Statist.* **28**, 265–275
- FEARNHEAD, P. AND G. RIGAILL (2019). Change-point detection in the presence of outliers. *J. Amer. Statist. Assoc.* **114**, 169–183.
- FISCH, A.T.M., ECKLEY, I.A., FEARNHEAD P. (2022) A linear time method for the detection of collective and point anomalies *Stat. Anal. Data Min.* **15**, 494–508.
- FRICK, K., MUNK, A., SIELING, H. (2014). Multiscale Change-Point Inference. *J. Royal. Statist. Soc. Ser. B.* **76**, 495–580.
- FRYZLEWICZ, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42**, 2243–2281.
- HAMPEL, F. R. (1987). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69**, 383–393.
- HAN, J. JIN, Z. AND XIA, B. (2016). Proving Inequalities and Solving Global Optimization Problems via Simplified CAD Projection. *J. Symbolic Comput.* **72**, 206–230.
- ITOH, N. AND KURTHS, J. (2010). Change-point detection of climate time series by nonparametric method. In *Proceedings of the world congress on engineering and computer science.* **1**, 445–448.
- JEWELL, S. HOCKING, T. D., FEARNHEAD, P., AND WITTEN, D. (2019). Fast nonconvex deconvolution of calcium imaging data. *Biostatistics.* to appear.
- JEWELL, S. AND WITTEN, D. (2018). Exact spike train inference via  $\ell_0$  opti-

- mization *Ann. Appl. Statist.* **12**, 2457–2482.
- KILLICK, R., FEARNHEAD, P., AND ECKLEY, I. A. (2012) Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107**, 15901598.
- LAURENT, B., AND MASSART, P. (2000). Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28**, 1302–1338.
- LAVIELLE, M. AND MOULINES, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *J. Time Ser. Anal.* **21**, 33–59.
- LI, H., MUNK, A., SIELING, H. (2016). FDR-control in multiscale change-point segmentation. *Electron. J. Statist.* **10**, 918–959.
- LUNG-YUT-FONG, A., LÉVY-LEDUC, C. AND CAPPÉ, O. (2012). Distributed detection/localization of change-points in high-dimensional network traffic data. *Statist. Comput.* **22**, 485–496.
- MAIDSTONE, R., HOCKING, T, RIGAILL, G., AND FEARNHEAD, P. (2018). On optimal multiple changepoint algorithms for large data. *Statist. Comput.* **27**, 519–533.
- OLSHEN, A. B., VENKATRAMAN, E. , LUCITO, R., AND WIGLER, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics.* **5**, 557–572.
- PAGE, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika.* **42**, 523–527.
- PREUSS, P., PUCHSTEIN, R. AND DETTE, H. (2015). Detection of multiple structural breaks in multivariate time series. *J. Amer. Statist. Assoc.* **110**, 654–668.
- RAIMONDO, M. (1998). Minimax estimation of sharp change points. *Ann. Statist.* **26**, 1379–1397.
- TICKLE, S. O., ECKLEY, I. A., FEARNHEAD, P., AND HAYNES, K. (2019). Parallelisation of a common changepoint detection method. *J. Comput. Graph. Statist.*, to appear.
- VERZELEN, N., FROMONT, M., LERASLE, M., AND REYNAUD-BOURET, P. (2020). Optimal Change-Point Detection and Localization. *ArXiv preprint*. arXiv:2010.11470.
- VOSTRIKOVA, L. (1981). Detecting disorder in multidimensional random processes. *Soviet Math. Dokl.* **24**, 55–59.
- WANG, D., YU, Y. AND RINALDO, A. (2020). Univariate mean change point detection: penalization, CUSUM and optimality. *Electron. J. Statist.*, **14**, 1917–1961.
- YAO, Y.-C. (1988). Estimating the number of change-points via Schwarz’ criterion. *Statist. Probab. Lett.* **6**, 181–189.
- YAO, Y.-C. AND AU, S. (1989). Least-Squares Estimation of a Step Function. *Sankhya A.* **51**, 370–381.
- YU, Y., CHATTERJEE, S., AND XU, H. (2022) Localising change points in piecewise polynomials of general degrees. *Electron. J. Statist.*, **16**, 18551890.