# Optimal Nonparametric Multivariate Change Point Detection and Localization

Oscar Hernan Madrid Padilla, Yi Yu, Daren Wang, and Alessandro Rinaldo

***Abstract*— We study the multivariate nonparametric change point detection problem, where the data are a sequence of independent $p$-dimensional random vectors whose distributions are piecewise-constant with Lipschitz densities changing at unknown times, called change points. We quantify the size of the distributional change at any change point with the supremum norm of the difference between the corresponding densities. We are concerned with the localization task of estimating the positions of the change points. In our analysis, we allow for the model parameters to vary with the total number of time points, including the minimal spacing between consecutive change points and the magnitude of the smallest distributional change. We provide information-theoretic lower bounds on both the localization rate and the minimal signal-to-noise ratio required to guarantee consistent localization. We formulate a novel algorithm based on kernel density estimation that nearly achieves the minimax lower bound, save possibly for logarithm factors. We have provided extensive numerical evidence to support our theoretical findings.**

***Index Terms*— Multivariate, nonparametric, kernel density estimation, CUSUM, binary segmentation.**

## I. Introduction

We STUDY the nonparametric multivariate change point detection problem, where we are given a sequence of independent random vectors $\{X(t)\}_{t=1}^T \subset \mathbb{R}^p$ with unknown distributions $\{P_t\}_{t=1}^T$ such that, for an unknown sequence of *change points* $\{\eta_k\}_{k=1}^K \subset \{2, \ldots, T\}$ with $1 = \eta_0 < \eta_1 < \ldots < \eta_K \leq T < \eta_{K+1} = T+1$, we have

$$P_t \neq P_{t-1} \quad \text{if and only if} \quad t \in \{\eta_1, \ldots, \eta_K\}. \qquad (1)$$

Our goal is to accurately estimate the number of change points $K$ and their locations.

Change point localization problems of this form arise in a variety of application areas, including finance [1] and [2], economics [18], neuroscience [9] and [35], climatology [15], biology [16] and [34], medical sciences [25] and [28], to name but a few. As a concrete example, we consider stock price data, ranging from Jan-1-2016 to Aug-11-2019 and consisting of 20 companies with highest average stock prices from the S&P500 market. Applying the change point detection method to be proposed in this paper, we are able to detect a number of change points, all of which correspond to some key dates in the US-China trade war. For instance, the date Dec-21-2017 is a detected change point and it is close to Jan 2018, when Mr. Trump imposed threats and tariffs to China. For more details about this data analysis, see Section V-B.

Due to the high demand from real-life applications, change point detection is a well-established topic in statistics with a rich literature. Some early efforts include seminal works by Wald [53], Yao [57], Yao and Au [59], Yao and Davis [58]. More recently, the change point detection literature has been brought back to the spotlight due to significant methodological and theoretical advances, including Aue *et al.* [4], Killick *et al.* [31], Fryzlewicz [20], Frick *et al.* [19], Cho [13], Wang and Samworth [56], Wang *et al.* [55], Verzelen *et al.* [52], among many others, in different aspects of parametric change point detection problems. See Wang *et al.* [54] for a more comprehensive review.

Most of the exiting results in the change point localization literature rely on parametric assumptions on the underlying distributions and on the nature of their changes. Despite the popularity and applicability of parametric change point detection methods, it is also important to develop more general and flexible change point localization procedures that are applicable over larger, possibly nonparametric, classes of distributions. Several efforts in this direction have been recently made for univariate data. Pein *et al.* [43] proposed a version of the SMUCE algorithm (Frick *et al.* [19]) that is sensitive to mean changes, but robust to changes in variance; Zou *et al.* [62] introduced a nonparametric estimator that can detect general distributions shifts; as an extension of Zou *et al.* [62], Haynes *et al.* [27] simplified the loss function in Zou *et al.* [62] and adopted the pruned exact linear time algorithm (Killick *et al.* [31]) to improve the computational efficiency; Padilla *et al.* [39] considered a nonparametric procedure for sequential change point detection; Fearnhead and Rigaill [17] focused on univariate mean change point detection constructing an estimator that is robust to outliers; Jula Vanegas *et al.* [30] proposed an estimator for detecting changes in pre-specified quantiles of the generative model; and Padilla *et al.* [39] developed a nonparametric version of binary segmentation (e.g. Scott and Knott [47]) based on the Kolmogorov–Smirnov statistic.

Manuscript received May 11, 2020; revised September 16, 2021; accepted November 8, 2021. Date of publication December 2, 2021; date of current version February 17, 2022. This work was supported in part by NSF under Grant DMS 2015489 and in part by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/V013432/1. *(Corresponding author: Oscar Hernan Madrid Padilla.)*

Oscar Hernan Madrid Padilla is with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA 90095 USA (e-mail: oscar.madrid@stat.ucla.edu).

Yi Yu is with the Department of Statistics, University of Warwick, Coventry CV4 7AL, U.K. (e-mail: yi.yu.2@warwick.ac.uk).

Daren Wang is with the Department of ACMS, University of Notre Dame, Notre Dame, IN 46556 USA (e-mail: dwang24@nd.edu).

Alessandro Rinaldo is with the Department of Statistics and Data Science, Carnegie Mellon University, Baker Hall Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: arinaldo@cmu.edu).

Communicated by E. Gassiat, Associate Editor for Probability and Statistics.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2021.3130330.

Digital Object Identifier 10.1109/TIT.2021.3130330

In multivariate nonparametric settings, the literature on change point analysis is comparatively limited. Arlot *et al.* [3] considered a penalized kernel least squares estimator, originally proposed by Harchaoui and Cappé [26], for multivariate change point problems and derive an oracle inequality. Garreau and Arlot [21] obtained an upper bound on the localization rate afforded by this method, which is further improved computationally in Celisse *et al.* [8]. We remark that this method is a very convenient and general nonparametric change point detection method. The theoretical guarantees are shown based on a transformation of the original data into a univariate mean change point detection, with the jump size corresponding to the maximum mean discrepancy (e.g. Gretton *et al.* [24]), therefore the nonparametric and multivariate complexity of the problems are not directly shown. In addition, as pointed out in Garreau and Arlot [21], the success relies on a proper choice of the embedding. Matteson and James [36] also proposed a methodology for multivariate nonparametric change point localization and show that it can consistently estimate the change points. Zhang *et al.* [61] provided a computationally-efficient algorithm, based on a pruning routine based on dynamic programming. Chen [10] proposed a multivariate change point testing method based on a graph-based testing technique (e.g. Chen and Zhang [11] and Chen *et al.* [12]), focusing on the limiting distribution of a test statistic in an asymptotic sense.

In this paper we investigate the multivariate change point localization problem in fully nonparametric settings where the underlying distributions are only assumed to have piecewise and uniformly (in $T$, the total number of time points) Lipschitz continuous densities and the magnitudes of the distributional changes are measured by the supremum norm of the differences between the corresponding densities. We formally introduce our model next.

*Assumption 1 (Model Setting):* Let $\{X(t)\}_{t=1}^T \subset \mathbb{R}^p$ be a sequence of independent vectors satisfying (1). Assume that, for each $t = 1, \ldots, T$, the distribution $P_t$ has a bounded Lebesgue density function $f_t : \mathbb{R}^p \to \mathbb{R}$ such that

$$\max_{t=1,\ldots,T} \left| f_t(s_1) - f_t(s_2) \right| \leq C_{\mathrm{Lip}} \|s_1 - s_2\|, \qquad (2)$$

for all $s_1, s_2 \in \mathcal{X}$, where $\mathcal{X} \subset \mathbb{R}^p$ is the union of the supports of all the density functions $f_t$, $\|\cdot\|$ represents the $\ell_2$-norm, and $C_{\mathrm{Lip}} > 0$ is an absolute constant. We let

$$\Delta = \min_{k=1,\ldots,K+1} \{\eta_k - \eta_{k-1}\} \leq T$$

denote the minimal spacing between any two consecutive change points. For each $k = 1, \ldots, K$, we set

$$\kappa_k = \sup_{z \in \mathbb{R}^p} \left| f_{\eta_k}(z) - f_{\eta_k - 1}(z) \right| = \|f_{\eta_k} - f_{\eta_k - 1}\|_\infty$$

as the size of the change at the $k$th change point. Finally, we let

$$\kappa = \min_{k=1,\ldots,K} \kappa_k > 0, \qquad (3)$$

be the minimal such change.

The uniform Lipschitz condition (2) is a rather mild requirement on the smoothness of the underlying densities. The use of the supremum distance is a natural choice in nonparametric density estimation settings (e.g. Tsybakov [49]). If we assume the domain $\mathcal{X}$ to be compact, then the supremum distance is stronger than the $L_1$ distance (total variation distance). Thus, If the domain $\mathcal{X}$ is compact, then without any further assumptions on the function $f(\cdot)$, it holds that

$$\|f\|_1 \leq \mathrm{Vol}(\mathcal{X}) \|f\|_\infty,$$

while a reverse inequality with constants independent of $f(\cdot)$ is not available. Due to this observation, we see that if one assumes $\|f_1\| \geq \kappa$, then it always holds that $\|f\|_\infty \geq \kappa / \mathrm{Vol}(\mathcal{X})$. One can just simply apply our theory and the results follow.

Comparing the supremum distance with the Kolmogorov–Smirnov distance, we notice that the latter is not yet widely used in the theoretical and applied literature. In the multivariate goodness-of-fit literature, Polonik [44] pointed out that the multivariate Kolmogorov–Smirnov distance "*is not enough to a goodness-of-fit test to be consistent against all (or at least a large class) of alternatives*". As for the goodness-of-fit test, Polonik [44] argued that the right choice of a test statistic should be based on the level sets, which is in turn based on the supremum distance.

The model parameters $\Delta$ and $\kappa$ are allowed to change with the total number of time points $T$. This modeling choice allows us to consider change point models for which it becomes increasingly difficult to identify and estimate the change point locations accurately as we acquire more data. For simplicity, we will not explicitly express the dependence of $\Delta$ and $\kappa$ on $T$ in our notation. The dimension $p$ is instead treated as a fixed constant, as is customary in nonparametric literature. We will refer to any relationship among $\Delta$ and $\kappa$ that holds as $T$ tends to infinity as a *parameter scaling* of the model in Assumption 1.

The change point localization task can be formally stated as follows. We seek to construct change point estimators $1 < \hat{\eta}_1 < \ldots < \hat{\eta}_{\widehat{K}} \leq T$ of the true change points $\{\eta_k\}_{k=1}^K$ such that, with probability tending to 1 as $T \to \infty$,

$$\widehat{K} = K \quad \text{and} \quad \max_{k=1,\ldots,K} |\hat{\eta}_k - \eta_k| \leq \epsilon,$$

where $\epsilon = \epsilon(T, \Delta, \kappa)$. We say that the change point estimators $\{\hat{\eta}_k\}_{k=1}^{\widehat{K}}$ are consistent if the above holds with

$$\lim_{T \to \infty} \epsilon / \Delta = 0. \qquad (4)$$

We refer to $\epsilon$ as the localization error and to the sequence $\{\epsilon/\Delta\}$ as the localization rate.

### A. Summary of the Results

The contributions of this paper are as follows.

- We show that the difficulty of the localization task can be completely characterized in terms of the signal-to-noise ratio $\kappa^{p+2}\Delta$. Specifically, the space of the model parameters $(T, \Delta, \kappa)$ can be separated into an infeasible region, characterized by the scaling

$$\kappa^{p+2}\Delta \lesssim 1 \qquad (5)$$

and where no algorithm is guaranteed to produce consistent estimators of the change points (see Lemma 2), and a feasible region, in which

$$\kappa^{p+2}\Delta \gtrsim \log^{1+\xi}(T), \quad \text{for any } \xi > 0. \tag{6}$$

Under the feasible scaling, we develop the MNP (multivariate nonparametric) change point estimator, given in Algorithm 1, that is provably consistent. The gap between (5) and (6) is a poly-logarithmic factor in $T$, which implies that our procedure is consistent under nearly all scalings for which this task is feasible.

- We show that the localization error achieved by the MNP procedure is of order $\log(T)\kappa^{-(p+2)}$ across the entire feasibility region given in (6); see Theorem 1. We verify that this rate is nearly minimax optimal by deriving an information-theoretic lower bound on the localization error, showing that if $\kappa^{p+2}\Delta \gtrsim \zeta_T$, for any sequence $\{\zeta_T\}$ satisfying $\lim_{T\to\infty} \zeta_T = \infty$, then the localization error is larger than $\kappa^{-(p+2)}$, up to constants; see Lemma 3. Interestingly, the dependence on the dimension $p$ is exponential, and matches the optimal dependence in the multivariate density estimation problems assuming Lipschitz-continuous densities. We elaborate on this point further in Section III-B. The numerical experiments in Section V confirm the good performance of our algorithm.
- The MNP estimator is a procedure for nonparametric change point localization in multivariate settings that runs in polynomial time and can be considered a multivariate nonparametric extension of the binary segmentation methodology (Scott and Knott [47]) and its, variant wild binary segmentation (Fryzlewicz [20]). The MNP estimator deploys a version of the CUSUM statistic (Page [41]) based on kernel density estimators. We remark that some of our auxiliary results on consistency of kernel density estimators are obtained through non-trivial adaptation of existing techniques that allow for non-i.i.d. data and may be of independent interest.

The rest of the paper is organized as follows. In Section II we introduce the MNP procedure and in Section III we study its consistency and optimality. Section IV presents a discussion of choice of tuning parameters in practice. Simulation experiments demonstrating the effectiveness of the MNP algorithm and its competitive performance relative to existing procedures are reported in Section V. The proofs and technical details are left in the Appendices.

## II. Methodology

Our procedures for change point detection and localization is a nonparametric extension of the traditional CUSUM statistic and it relies on kernel density estimators.

*Definition 1 (Multivariate Nonparametric CUSUM):* Let $\{X(i)\}_{i=1}^T$ be a sample in $\mathbb{R}^p$. For any integer triplet $(s,t,e)$ satisfying $0 \le s < t < e \le T$ and any $x \in \mathbb{R}^p$, the multivariate nonparametric CUSUM statistic is defined as the function

$$x \in \mathbb{R}^p \mapsto$$

$$\widetilde{Y}_t^{s,e}(x) = \sqrt{\frac{(t-s)(e-t)}{e-s}} \left\{ \hat{f}_{s+1,t,h}(x) - \hat{f}_{t+1,e,h}(x) \right\},$$

where

$$\hat{f}_{s,e,h}(x) = \frac{h^{-p}}{e-s} \sum_{i=s+1}^{e} \mathcal{k}\left(\frac{x - X(i)}{h}\right) \tag{7}$$

and $\mathcal{k}(\cdot)$ is a kernel function (see e.g. Parzen, 1962). In addition, define

$$\widetilde{Y}_t^{s,e} = \max_{i=1,\ldots,T} \left| \widetilde{Y}_t^{s,e}(X(i)) \right|. \tag{8}$$

*Remark 1:* The statistic $\widetilde{Y}_t^{s,e}$ can be seen as an estimator of

$$\sup_{z\in\mathbb{R}^p} \left| \widetilde{Y}_t^{s,e}(z) \right|.$$

**Algorithm 1** Multivariate Nonparametric Change Point Detection. MNP $((s,e), \{(\alpha_r,\beta_r)\}_{r=1}^R, \tau, h)$

**INPUT:** Sample $\{X(t)\}_{t=s}^e \subset \mathbb{R}^p$, collection of intervals $\{(\alpha_r,\beta_r)\}_{r=1}^R$, tuning parameter $\tau > 0$ and bandwidth $h > 0$.

```
for r = 1, ..., R do
    (s_r, e_r) ← [s, e] ∩ [α_r, β_r]
    if e_r − s_r > 2h^{−p} + 1 then
        b_r ← argmax_{s_r + h^{−p} ≤ t ≤ e_r − h^{−p}} Ỹ_t^{s_r,e_r}
        a_r ← Ỹ_{b_r}^{s_r,e_r}
    else
        a_r ← −1
    end if
end for
r* ← argmax_{r=1,...,R} a_r
if a_{r*} > τ then
    add b_{r*} to the set of estimated change points
    MNP((s, b_{r*}), {(α_r, β_r)}_{r=1}^R, τ)
    MNP((b_{r*} + 1, e), {(α_r, β_r)}_{r=1}^R, τ)
end if
```

**OUTPUT:** The set of estimated change points.

Algorithm 1 below presents a multivariate nonparametric version of the univariate nonparametric change point detection method proposed in Padilla *et al.* [40], wild binary segmentation (Fryzlewicz [20]) and binary segmentation (BS) (e.g. Scott and Knott [47]). The resulting procedure consists of repeated application of the BS algorithm over random time intervals and using the multivariate nonparametric CUSUM statistic in Definition 1. The inputs of Algorithm 1 are a sequence $\{X(t)\}_{t=1,\ldots,T}$ of random vectors in $\mathbb{R}^p$, a tuning parameter $\tau > 0$ and a bandwidth $h > 0$. Detailed theoretical requirements on the values of $\tau$ and $h$ are discussed below in Section III, and Section IV offers guidance on how to select them in practice. In particular, the lengths of the sub-intervals are of order at least $h^{-p}$, where $h > 0$ is the value of the bandwidth used to define the multivariate nonparametric CUSUM statistic. This is to ensure that each sub-interval will contain enough points to yield a reliable density estimator.

Furthermore, in Algorithm 1 we scan through all time points between $s_r + h^{-p}$ and $e_r - h^{-p}$ in the interval $(s_r, e_r)$. This is done for technical reasons, to avoid working with intervals that have insufficient data, which would be the case when $e_r - s_r$ is small.

Finally, the computational cost of the algorithm is of order $O(T^2R \cdot \text{kernel})$, where $R$ is the number of random intervals and "kernel" stands for the computational cost of calculating the value of the kernel function evaluated at one data point. The dependence on the dimension $p$ is only through the evaluation of the kernel function. Additionally, Algorithm 1 has the worst-case memory consumption of order $O(T^2R)$. We highlight that our method is indeed computationally intensive which can be a problem in practice. One possible way to overcome this is to use parallel computing, as the calculations of $\{\widetilde{Y}_t^{s,e}\}$ can be done in parallel. Alternatively, as pointed out by one reviewer, one can compute $\mathcal{k}(X(i) - X(j))/h)$ for $i, j = 1, \ldots, T$, and then computing (7) only requires to calculate cumulative sums. The overall cost becomes $O(T^2\text{kernel} + RT^2)$. However, this requires a memory allocation of $O(T^2p + T^2R)$.

# III. Theory

In this section we prove that the change point estimator MNP returned by Algorithm 1 is consistent based on the model described in Assumption 1, under the parameter scaling

$$\kappa^{p+2}\Delta \gtrsim \log^{1+\xi}(T),$$

for any $\xi > 0$; see Theorem 1. In addition, we show in Lemma 2 that no consistent estimator exists if the above scaling condition is not satisfied, up to a poly-logarithmic factor. Finally, in Lemma 3, we demonstrate that the localization rate returned by the MNP procedure is nearly minimax rate-optimal.

## A. Optimal Change Point Localization

We begin by stating some assumptions on the kernel $\mathcal{k}(\cdot)$ used to compute the kernel density estimators involved in the definition of the multivariate nonparametric CUSUM statistic.

*Assumption 2 (The Kernel Function):* Let $\mathcal{k} : \mathbb{R}^p \to \mathbb{R}$ be a kernel function with $\|\mathcal{k}\|_\infty, \|\mathcal{k}\|_2 < \infty$ such that,

(i) the class of functions

$$\mathcal{F}_{\mathcal{k},[l,\infty)} = \left\{ \mathcal{k}\left(\frac{x - \cdot}{h}\right) : x \in \mathcal{X}, h \geq l \right\}$$

from $\mathbb{R}^p$ to $\mathbb{R}$ is separable in $L_\infty(\mathbb{R}^p)$, and is a uniformly bounded VC-class with dimension $\nu$, i.e. there exist positive numbers $A$ and $\nu$ such that, for every positive measure $Q$ on $\mathbb{R}^p$ and for every $u \in (0, \|\mathcal{k}\|_\infty)$, it holds that

$$\mathcal{N}(\mathcal{F}_{\mathcal{k},[l,\infty)}, L_2(Q), u) \leq \left(\frac{A\|\mathcal{k}\|_\infty}{u}\right)^\nu;$$

(ii) for a fixed $m > 0$,

$$\int_0^\infty t^{p-1} \sup_{\|x\| \geq t} |\mathcal{k}(x)|^m \, dt \; < \infty.$$

(iii) there exists a constant $C_{\mathcal{k}} > 0$ such that

$$\int_{\mathbb{R}^p} \mathcal{k}(z)\|z\| \, dz \; \leq C_{\mathcal{k}}.$$

Assumption 2 (i) and (ii) correspond to Assumptions 4 and 3 in Kim *et al.* [32] and are fairly standard conditions used in the nonparametric density estimation literature, see Giné and Guillou [22], Giné and Guillou [23], Sriperumbudur and Steinwart [48]. They hold for most commonly used kernels, such as uniform, Epanechnikov and Gaussian kernels. Assumption 2 (iii) is a mild integrability assumption on the kernel.

Next, we require the following signal-to-noise condition on the parameters of the model in order to guarantee that the MNP estimator is consistent.

*Assumption 3:* Assume that for a given $\xi > 0$, there exists an absolute constant $C_{\text{SNR}} > 0$ such that

$$\kappa^{p+2}\Delta > C_{\text{SNR}} \log^{1+\xi}(T). \tag{9}$$

Assumption 3 can be relaxed by only requiring that $\kappa^{p+2}\Delta > C_{\text{SNR}} \log(T) e_T$, for any arbitrary sequence $\{e_T\}$ diverging to infinity, as $T$ goes unbounded. As we will see later, the above scaling is not only sufficient for consistent localization but almost necessary, aside for a poly-logarithmic factor in $T$; see Lemma 2. This implies that the MNP estimator is consistent for nearly all parameter scalings for which the localization task is possible.

*Theorem 1:* Assume that the sequence $\{X_t\}_{t=1}^T$ satisfies the model described in Assumption 1 and the signal-to-noise ratio condition Assumption 3. Let $\mathcal{k}(\cdot)$ be a kernel function satisfying Assumption 2. Then, there exist positive universal constants $C_R$, $c_{\tau,1}$, $c_{\tau,2}$ and $c_h$, such that if Algorithm 1 is applied to the sequence $\{X_t\}_{t=1}^T$ using any collection $\{(\alpha_r, \beta_r)\}_{r=1}^R \subset \{1, \ldots, T\}$ of random time intervals with endpoints drawn independently and uniformly from $\{1, \ldots, T\}$ with

$$\max_{r=1,\ldots,R} (\beta_r - \alpha_r) \leq C_R\Delta \quad \text{almost surely,} \tag{10}$$

tuning parameter $\tau$ satisfying

$$c_{\tau,1} \max\left\{h^{-p/2}\log^{1/2}(T),\, h\Delta^{1/2}\right\} \leq \tau \leq c_{\tau,2}\kappa\Delta^{1/2} \tag{11}$$

and bandwidth $h$ given by

$$h = c_h\kappa, \tag{12}$$

then the resulting change point estimator $\{\hat{\eta}_k\}_{k=1}^{\widehat{K}}$ satisfies

$$\begin{aligned}\mathbb{P}\Bigg\{&\widehat{K} = K \quad \text{and} \quad \epsilon_k = |\hat{\eta}_k - \eta_k| \leq C_\epsilon \kappa_k^{-2}\kappa^{-p}\log(T), \\ &\forall k = 1, \ldots, K\Bigg\} \\ &\geq 1 - 3T^{-c} - \exp\left\{\log\left(\frac{T}{\Delta}\right) - \frac{R\Delta}{4C_RT}\right\},\end{aligned}$$

for universal positive constants $C_\epsilon$ and $c$.

The constants in Theorem 1 are well-defined provided that the constant $C_{\text{SNR}}$ in the signal-to-noise ratio Assumption 3 is sufficiently large. Their dependence can be tracked in the proof of the Theorem 1, given in Section B. In particular, it must hold that $c_{\tau,1}\max\{1, c_h^{-p/2}\} < c_{\tau,2}$. We would like to point out that the choice $h = c_h\kappa$ is merely to prompt theoretical optimality on localization error rate. We discuss a wider range of choices on $h$ and their theoretical implications in Section III-B. Practical guidance on the choice of $h$ is collected in Section IV.

It is worth emphasizing that we provide individual localization errors $\epsilon_k$, one for each true change point, in order to avoid false positives in the iterative search of change points in Algorithm 1. Using (1) and setting

$$\epsilon = \max_{k=1,\ldots,K} \epsilon_k,$$

our result further yields the general localization consistency guarantee defined in (4) since, as $T \to \infty$,

$$\frac{\epsilon}{\Delta} \leq C_\epsilon \frac{\log(T)}{\Delta \kappa^{p+2}} \leq \frac{C_\epsilon}{C_{\mathrm{SNR}}} \frac{\log(T)}{\log^{1+\xi}(T)} \to 0,$$

where the second inequality follows from the definition of $\kappa$ in (3), and convergence follows from Assumption 3.

The tuning parameter $\tau$ plays the role of a threshold for detecting change points in Algorithm 1. In particular, for the time points with maximal CUSUM statistics, if their CUSUM statistic values exceed $\tau$, then they are included in the change point estimators. This means that, with large probability, the upper bound in (11) ought be smaller than the smallest population CUSUM statistics at the true change points, and the lower bound in (11) should be larger than the largest sample CUSUM statistics when there are no change points. In detail, the upper bound is determined in Lemma 10, and the lower bound comes from Lemmas 7 and 8. Lemma 7 is dedicated to the variance of the kernel density estimators at the observations, whereas Lemma 8 focuses on the deviance between the sample and population maxima. Lastly, the set of values for $\tau$ is not empty, by the inequalities

$$c_{\tau,1} h^{-p/2} \log^{1/2}(T) \leq c_{\tau,1} c_h^{-p/2} \kappa^{-p/2} \log^{1/2}(T) < c_{\tau,2} \kappa \Delta^{1/2}$$

and

$$c_{\tau,1} \max\{1,\, c_h^{-p/2}\} < c_{\tau,2}.$$

The lower bound on the probability in (1) tends to 1, as $T$ grows unbounded, provided that the number $R$ of random intervals $(\alpha_r, \beta_r)$ is such that

$$R \gtrsim \frac{T}{\Delta} \log\left(\frac{T}{\Delta}\right).$$

With this, we remark that the computational complexity is therefore of order $O(\Delta^2 \cdot T/\Delta \cdot \log(T/\Delta) \cdot \mathrm{kernel})$, by noticing that the interval lengths are upper bounded by $C_R\Delta$. Since the procedures in random intervals are parallelable, one may run Algorithm 1 in parallel and the computational complexity is of order $O(\Delta^2 \cdot \mathrm{kernel})$.

The assumption (10) is imposed to guarantee that each of the random intervals used in the MNP procedure contains a bounded number of change points. Thus, if $K = O(1)$, this assumption can be discarded. More generally, it is possible to drop this assumption even when $\Delta = o(T)$, in which case the MNP estimator would still yield consistent localization, albeit with a localization error inflated by a polynomial factor in $T/\Delta$, under a stronger signal-to-noise ratio condition. Assumptions of this nature are commonly used in the analysis of the WBS procedure. For a discussion on the necessity of assumption (10) in order to derive optimal rates, see Padilla *et al.* [40].

*Remark 2 (When $\kappa = 0$):* Theorem 1 builds upon the assumption that $\kappa > 0$, which implies that there exists at least one change point. In fact, an immediate consequence of Step 1 in the proof of Theorem 1 is the consistency for the simpler task of merely deciding if there are change points or not. To be specific, if there are no true change points, then with the bandwidth and tuning parameter satisfying

$$h > (\log(T)/T)^{1/p}$$

and

$$\tau \geq c_{\tau,1} \max\left\{h^{-p/2} \log^{1/2}(T),\, h\, T^{1/2}\right\},$$

it holds that

$$\mathbb{P}\{\widehat{K} = 0\} \to 1,$$

as $T$ goes unbounded.

### *B. Change Point Localization Versus Density Estimation*

We now discuss how the change point localization problem relates to the classical task of optimal density estimation. For simplicity, assume equally-spaced change points, so that the data consist of $K$ independent samples of size $\Delta$ from each of the underlying distributions.

If we knew the locations of the change points – or, equivalently, the number of change points – then we could compute $K$ kernel density estimators, one for each sample. Recalling that we assume the underlying densities to be Lipschitz and using well-known results about minimax density estimation, choosing the bandwidth to be of order

$$h_1 \asymp \left(\frac{\log(\Delta)}{\Delta}\right)^{1/(p+2)}$$

would yield $K$ kernel density estimators that are minimax rate-optimal in the $L_\infty$-norm for each of the underlying densities. In contrast, the choice of the bandwidth for the change point detection task is

$$h_{\mathrm{opt}} \asymp \kappa,$$

as given in (12). In light of the minimax results established in the next section, such a choice of $h_{\mathrm{opt}}$ further guarantees that the localization rate afforded by the MNP algorithm is almost minimax rate-optimal.

In virtue of Assumption 3 and the boundedness assumption on the densities, it holds that

$$h_1 \lesssim h_{\mathrm{opt}}.$$

The choice of bandwidth for optimal change point localization in the present problem is no smaller than the choice for optimal density estimation. In particular, the two bandwidth coincides, i.e. $h_1 \asymp h_{\mathrm{opt}}$, when the signal-to-noise ratio is smallest, i.e. when Assumption 3 is an equality. As we will see below in Lemma 2, change point localization is not possible when the signal-to-noise ratio Assumption 3 fails, up to a slack factor that is poly-logarithmic in $T$. As a result, $h_1$ and $h_{\text{opt}} \log^{\xi}(\Delta)$ are of the same order (up to a poly-logarithmic term in $T$) only under (nearly) the worst possible condition for localization. On the other hand, if $\kappa$ is vanishing in $T$ at a rate slower than $(\log(\Delta)/\Delta)^{1/(p+2)}$ (while still fulfilling Assumption 3), then change point localization can be solved optimally using kernel density estimators that are suboptimal for density estimation, since they are based on bandwidths that are larger than the ones needed for optimality. Thus we conclude that the optimal sample complexity for the localization problem is strictly better than the optimal sample complexity needed for estimating all the underlying densities, unless the difficulty of the change localization problem is maximal, in which case they coincide. At the opposite end of the spectrum, if $\kappa$ is bounded away from 0, then the optimal change point localization can still be achieved using *biased* kernel density estimators with bandwidths bounded away from zero.

More generally, and quite interestingly, our analysis reveals that there is a rather simple and intuitive way of describing how the difficulty of density estimation problem relates to the difficulty of consistent change point localization, at least in our problem. Indeed, it follows from the proof of Theorem 1 (see also (11) in the statement of Theorem 1) that, in order for MNP to return a consistent – and, as we will see shortly, nearly minimax optimal – estimator of the change point, the following should hold:

$$\kappa\sqrt{\Delta} \gtrsim \gamma_{\mathcal{A}} + \gamma_{\mathcal{B}} \asymp h^{-p/2}\log^{1/2}(T) + h\sqrt{\Delta}. \tag{13}$$

Assuming for simplicity $\log(\Delta) \asymp \log(T)$, the right hand side of the previous expression divided by $\sqrt{\Delta}$ precisely corresponds to the sum of the magnitudes of the bias and of the random fluctuation for the kernel density estimator over each sub-interval, both measured in the $L_\infty$-norm. From this we immediately see that the MNP procedure will estimate the change points optimally provided that $\kappa$, the smallest magnitude of the distributional change at the change point, is larger than the $L_\infty$ error in estimating the underlying densities via kernel density estimation, *assuming full knowledge of the change point locations.* Though simple, we believe that this characterization is non-trivial and illustrates nicely the differences between the task of density estimation of that of change point localization.

We conclude this section by providing some rationale as to why the optimal choice of $h$ for the purpose of change point localization happens to be $\kappa$, which in light of the inequality (13), is the largest value $h$ is allowed to take in order for MNP to be consistent. We offer two different perspectives.

- (Localization error). It can be seen in Lemma 15 or in inequality (69) in the proof of Theorem 1 that the localization error is such that

$$\epsilon_k \lesssim \frac{\gamma_{\mathcal{A}}^2}{\kappa_k^2} = \frac{\log(T)}{\kappa_k^2 h^p}, \quad k \in \{1, \ldots, K\}.$$

  Therefore, the larger the bandwidth $h$ is, the smaller the localization error.

- (Signal-to-noise ratio). Since we require $\gamma_{\mathcal{A}} \lesssim \kappa\sqrt{\Delta}$, it needs to hold that

$$\kappa^2 h^p \Delta \gtrsim \log(T);$$

  since in (30) in the proof of Lemma 8 we require

$$\kappa_k \sqrt{C_\epsilon \log(T) V_p^2 \kappa_k^{-2} \kappa^{-p}} \le \gamma_{\mathcal{B}},$$

  where $V_p = \pi^{p/2}(\Gamma(p/2+1))^{-1}$ is the volume of a unit ball in $\mathbb{R}^p$, it needs to hold that

$$\kappa^p h^2 \Delta \gtrsim \log(T).$$

  Therefore, the larger the bandwidth $h$ is, the smaller $\kappa$ and $\Delta$ can be.

### C. Minimax Lower Bounds

For the model given in Assumption 1, we will describe low signal-to-noise ratio parameter scalings for which consistent localization is not feasible. These scalings are complementary to the ones in Assumption 3, which, by Theorem 1, are sufficient for consistent localization.

*Lemma 2:* Let $\{X(t)\}_{t=1}^T$ be a sequence of random vectors satisfying Assumption 1 with one and only one change point and let $P_{\kappa,\Delta}^T$ denote the corresponding joint distribution. Then, there exist universal positive constants $C_1$, $C_2$ and $c < \log(2)$ such that, for all $T$ large enough,

$$\inf_{\hat{\eta}} \sup_{P \in \mathcal{Q}} \mathbb{E}_P\left(|\hat{\eta} - \eta(P)|\right) \ge \Delta/4,$$

where

$$\mathcal{Q} = \mathcal{Q}(C_1, C_2, c) = \\ \left\{P_{\kappa,\Delta}^T : \Delta < T/2,\, \kappa < C_1,\, \kappa^{p+2}\Delta \le c,\, C_{\text{Lip}} \le C_2\right\},$$

the quantity $\eta(P)$ denotes the true change point location of $P \in \mathcal{Q}$ and the infimum is over all possible estimators of the change point location.

The above result offers an information theoretic lower bound on the minimal signal-to-noise ratio required for localization consistency. It implies that Assumption 3 used by the MNP procedure, is, save for a poly-logarithmic term in $T$, the weakest possible scaling condition on the model parameters any algorithm can afford. Thus, Lemma 2 and Theorem 1 together reveal a phase transition over the parameter scalings, separating the impossibility regime in which no algorithm is consistent from the one in which MNP accurately estimates the change point locations. We conjecture that the logarithmic gap is due to a loose lower bound and the upper bound is tight. It remains an open problem to close this gap.

Our next result shows that the localization rate achieved by Algorithm 1, under the tuning parameters specified in Theorem 1, is indeed almost minimax optimal, aside possibly for a poly-logarithmic factor, over all scalings for which consistent localization is possible.

*Lemma 3:* Let $\{X(t)\}_{t=1}^T$ be a sequence of random vectors satisfying Assumption 1 with one and only one change point and let $P_{\kappa,\Delta}^T$ denote the corresponding joint distribution.

Then, there exist universal positive constants $C_1$ and $C_2$ such that, for any sequence $\{\zeta_T\}$ satisfying $\lim_{T\to\infty}\zeta_T = \infty$,

$$\inf_{\hat\eta}\sup_{P\in\mathcal{Q}} \mathbb{E}_P\left(|\hat\eta - \eta(P)|\right) \geq \max\left\{1, \frac{1}{4}\left\lceil \frac{1}{V_p^2\kappa^{p+2}}\right\rceil e^{-2}\right\},$$

where $V_p = \pi^{p/2}(\Gamma(p/2+1))^{-1}$ is the volume of a unit ball in $\mathbb{R}^p$,

$$\mathcal{Q} = \mathcal{Q}(C_1, C_2, \{\zeta_T\}) = \left\{P^T_{\kappa,\Delta} : \Delta < T/2,\ \kappa < C_1,\ \kappa^{p+2}V_p^2\Delta \geq \zeta_T,\ C_{\mathrm{Lip}} \leq C_2\right\},$$

the quantity $\eta(P)$ denotes the true change point location of $P \in \mathcal{Q}$ and the infimum is over all possible estimators of the change point locations.

The previous result demonstrates that the performance of the MNP procedure is essentially non-improvable, except possibly for a poly-logarithmic term in $T$. In particular, adapting choosing the bandwidth in a way that depend on the lengths of the working intervals is not going to bring significant improvements over a fixed choice.

## IV. Choice of Tuning Parameters in Practice

In this section, we discuss the choice of the tuning parameters involved in Algorithm 1. The first tuning parameter is $R$, the number of random intervals. Based on Theorem 1, we should choose $R \gtrsim \frac{T}{\Delta}\log\left(\frac{T}{\Delta}\right)$. If $\Delta \asymp T$ then this becomes $R \gtrsim 1$. However, both $\Delta$ and the constants in these inequalities are unknown. In all of our experiments we set $R = 50$ and notice that Algorithm 1 is not sensitive to $R$ for the examples considered here. However, guided by our theory, for problems where $T$ increases one might wish to choose $R$ as a linear function of $T$, assuming that $\Delta = O(1)$. Furthermore, it is worth to address that increasing $R$ only helps the performance and hence a conservative choice is to choose large $R$, if computational resources permit. Once $R$ is fixed, we independently draw $\alpha_r$ uniformly from the set $\{1,\dots,T\}$ for $r = 1,\dots,R$. Then we generate $\beta_r$ uniformly from $\{\alpha_r,\dots,T\}$ for $r = 1,\dots,R$.

With regards to the other implementation details of the MNP method described in Algorithm 1, we use the Gaussian kernel and set $h = 5\times(30\log(T)/T)^{1/(p+2)}$. The intuition behind this choice comes from Theorem 1. As stated there, one needs to choose $h \asymp \kappa$. However, by Assumption 3 we require $\kappa \gtrsim (\log(T)\ /\Delta)^{1/(p+2)} = (30\log(T)\ /T)^{1/(p+2)}$ if $\Delta \approx T/30$. Here, the constants 30 and 5 are ad-hoc that we find to work well in practice.

With fixed $h$, we then run Algorithm 1 with $\tau = 0$ to produce a binary tree where every node corresponds to a potential change point. This is useful since then for any $\tau > 0$ we can run Algorithm 1 by simply pruning the tree. We then choose $m = 30$ and sequence of values of $\tau$ corresponding to the $m$ largest CUSUM statistics in the tree. This produces a sequence of nested sets

$$S_0 = \emptyset \subset S_1 \subset \ldots \subset S_m,$$

corresponding to the different values of $\tau$. We then borrow some inspiration from the selection procedure in Padilla *et al.* [40]. Specifically, we start from $S_i$, with $i = m$, and for every $\hat\eta \in S_i\backslash S_{i-1}$ we decide whether $\hat\eta$ is a change point or not. If at least one element $\hat\eta \in S_i\backslash S_{i-1}$ is declared as a change point, then we stop and set $\widehat{\mathcal{C}} = S_i$ as the set of estimated change points. Otherwise, we set $i = m-1$ and repeat the same procedure. We continue iteratively until the procedure stops, or $i = 0$ in which case $\widehat{\mathcal{C}} = \emptyset$. The only remaining ingredient is how to decide if $\hat\eta \in S_i\backslash S_{i-1}$ is a change point or not. To that end, we let $\hat\eta_{(1)}, \hat\eta_{(2)} \in S_{i-1}$, such that

$$\hat\eta \in [\hat\eta_{(1)}, \hat\eta_{(2)}] \ \text{ and } \ (\hat\eta_{(1)}, \hat\eta_{(2)}) \cap S_{i-1} = \emptyset.$$

If $\hat\eta < \hat\eta'$ ($\hat\eta > \hat\eta'$) for all $\hat\eta' \in S_{i-1}$, then we set $\hat\eta_{(1)} = 1$ ($\hat\eta_{(2)} = T$). Then, we independently draw $v_l$ uniformly from $\{v \in \mathbb{R}^p : \|v\| = 1\}$ for $l = 1,\dots,N$, and we calculate the Kolmogorov–Smirnov (KS) statistic (for instance, see Padilla *et al.* [40])

$$a_l = \mathrm{KS}(\{v_l^\top X(t)\}_{\hat\eta_{(1)}+1}^{\hat\eta}, \{v_l^\top X(t)\}_{\hat\eta+1}^{\hat\eta_{(2)}}),$$

and the corresponding $p$-value which we compute as $P_l = \exp(-2a_l^2)$. Next we sort $\{P_l\}$ as $P_{(1)} \leq P_{(2)} \leq \ldots \leq P_{(N)}$. Inspired by Benjamini and Hochberg [5], we declare $\hat\eta$ as change point if there exists $k \in \{1,\dots,N\}$ such that

$$P_{(k)} \leq \frac{k}{N}\alpha, \tag{14}$$

with $\alpha = 0.0005$. The choice $\alpha = 0.0005$ is due to the fact that we do multiple tests for different values of $\tau$ and their corresponding estimated change points. The number of tests is in principle random. We choose the value 0.0005 since $(1-0.0005)^{20} \approx 0.99$, and we use as an ad-hoc rule that there are at most 20 change points. Also, in our experiments, we set $N = 200$. We acknowledge that when $p$ is large, better ways of choosing $N$ might be necessary, as there are more directions to choose from for larger $p$.

Furthermore, it is worth pointing out that our approach above, based on false discovery rate using ideas from Benjamini and Hochberg [5], is heuristic and does not have statistical guarantees. In fact, the test statistics that we calculate are not independent, hence the rule in (14) is not theoretically justified, even though the resulting approach works well in practice as it can be seen in the next section.

Finally, if necessary, one may further improve the numerical performance by cross-validation on a grid of $h$.

## V. Experiments

In this section we describe several computational experiments illustrating the effectiveness of the MNP procedure for estimating change point locations across a variety of scenarios. We organize our experiments into two subsections, one consisting of examples with simulated data and the other based on a real data example. Code implementing our method can be found in the R (R Core Team [45]) package available at https://github.com/hernanmp/RMNCP.

### A. Simulations

We start our experiments section by assessing the performance of Algorithm 1 in a wide range of situations.

We compare our MNP procedure against the energy based method (EMNCP) from Matteson and James [36], the sparsified binary segmentation (SBS) method from Cho and Fryzlewicz [14], the double CUSUM binary segmentation estimator (DCBS) from Cho [13], and the kernel change point detection procedure (KCPA)(Celisse *et al.* [8]; Arlot *et al.* [3]).

As a measure of performance we use the absolute error $|\widehat{K} - K|$, averaged over 100 Monte Carlo simulations, where $\widehat{K}$ is the estimated number of change points returned by the estimators. In addition, we use the one-sided Hausdorff distance

$$d(\widehat{\mathcal{C}}|\mathcal{C}) = \max_{\eta \in \mathcal{C}} \min_{x \in \widehat{\mathcal{C}}} |x - \eta|,$$

where $\mathcal{C} = \{\eta_1, \ldots, \eta_K\}$ is the set of true change points and $\widehat{\mathcal{C}}$ is the set of estimated change points. We report the medians of both $d(\widehat{\mathcal{C}}|\mathcal{C})$ and $d(\mathcal{C}|\widehat{\mathcal{C}})$ over 100 Monte Carlo simulations. We use the convention that when $\widehat{\mathcal{C}} = \emptyset$, we define $d(\widehat{\mathcal{C}}|\mathcal{C}) = \infty$ and $d(\mathcal{C}|\widehat{\mathcal{C}}) = -\infty$.

With regards to the implementation of the EMNCP method, we use the R (R Core Team [45]) package `ecp` (James and Matteson [29]). The calculation of the change points is done via the function `e.divisive()` with its default choice of tuning parameters. Furthermore, the methods SBS and DCBS methods we use the R (R Core Team [45]) package `hdbinseg` via the functions `sbs.alg()` and `dcbs.alg()`. Furthermore, for KCPA we use the R (R Core Team [45]) package `KernSeg` and the function `KernSeg_MultiD()` with the choices $Kmax = 20$ (maximum number of points), $min.size = 2$ (minimum size between change points), and with the choice of the Gaussian Kernel.

As for the MNP method described in Algorithm 1 we use the tuning parameters as described in Section IV.

To evaluate the quality of the competing estimators, we construct several change point models. In each case, we make choice of $K$ and consider $T \in \{150, 300\}$ and $p \in \{10, 20\}$.

*a) Scenario 1:* We generate data as

$$X(t) = \mu(t) + \epsilon(t), \quad t \in \{1, \ldots, T\},$$

where $\epsilon(t) \sim N(0, I_p)$ and $I_p$ is the $p \times p$ identity matrix. Moreover, the mean vectors satisfy

$$\mu(t) = \begin{cases} v^{(0)} & t \in A_1 \cup A_3, \\ v^{(1)} & \text{otherwise}, \end{cases}$$

where $A_1 = [1, \lfloor T/3 \rfloor]$, $A_3 = [1 + \lfloor 2T/3 \rfloor, T]$, $v^{(0)} = 0 \in \mathbb{R}^p$, and $v^{(1)}_j = 1$ for $j \in \{1, \ldots, p/2\}$ and $v^{(1)}_j = 0$ otherwise.

*b) Scenario 2.:* We define $A_j = [1 + (j - 1)\lfloor T/7 \rfloor, j\lfloor T/7 \rfloor]$ for $j = 1, \ldots, 6$ and $A_7 = [1 + 6\lfloor T/7 \rfloor, T]$. This gives seven roughly evenly spaced segments that we use to generate data as

$$X(t) = \mu(t) + \epsilon(t), \quad t \in \{1, \ldots, T\},$$

where

$$\mu(t) = \begin{cases} 0 & t \in A_j, \ \text{for } j \text{ odd}, \\ v^{(1)} & \text{otherwise}, \end{cases}$$

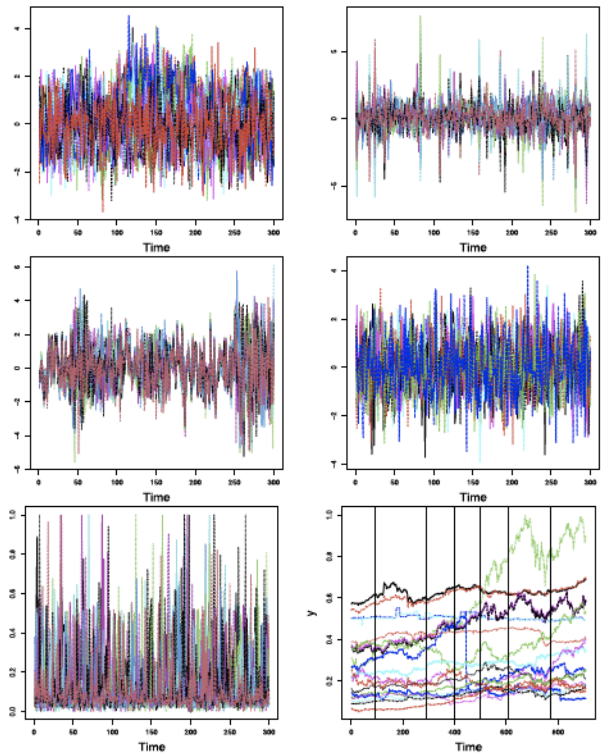


Fig. 1. From left to right and from top to bottom, the first five plots illustrate raw data generated from Scenarios 1 to 5, respectively, with one realization each. In each case, $T = 300$ and $p = 20$, with the x-axis representing the time horizon, and the y-axis the values of each measurement. Different curves in each plot are associated with different coordinates of the vector $X(t)$. The right panel in the third row illustrates the raw data and estimated change points by MNP for the example in Section V-B.

with $v^{(1)} = (0.2, \ldots, 0.2)^\top \in \mathbb{R}^p$. Furthermore, the errors satisfy $\sqrt{3}\,\epsilon(1), \ldots, \sqrt{3}\,\epsilon(T) \overset{\text{i.i.d.}}{\sim} \text{Mt}(I_p, 3)$, where $\text{Mt}(I_p, 3)$ is the multivariate $t$-distribution with the scale matrix $I_p$ and the degrees of freedom three.

*c) Scenario 3:* We generate observations from the model

$$X(t) \overset{\text{i.i.d.}}{\sim} N(0, \Sigma(t)), \quad t \in \{1, \ldots, T\},$$

where

$$\Sigma(t) = \begin{cases} 0.1 I_p + 0.9 \mathbb{1}\mathbb{1}^\top & t \in A_1 \cup A_4 \cup A_6, \\ 2 I_p + 0.8 \mathbb{1}\mathbb{1}^\top & t \in A_2 \cup A_7, \\ I_p & t \in A_3 \cup A_5, \end{cases}$$

with $A_1 = [1, \lfloor T/7 \rfloor]$, $A_2 = [\lfloor T/7 \rfloor + 1, \lfloor T/7 \rfloor + \lfloor T/14 \rfloor]$, $A_3 = [\lfloor T/7 \rfloor + \lfloor T/14 \rfloor + 1, \lfloor T/7 \rfloor + 3\lfloor T/14 \rfloor]$, $A_4 = [\lfloor T/7 \rfloor + 3\lfloor T/14 \rfloor + 1, \lfloor T/7 \rfloor + 4\lfloor T/14 \rfloor]$, $A_5 = [\lfloor T/7 \rfloor + 4\lfloor T/14 \rfloor + 1, \lfloor T/7 \rfloor + 6\lfloor T/14 \rfloor]$, $A_6 = [\lfloor T/7 \rfloor + 6\lfloor T/14 \rfloor + 1, 6\lfloor T/7 \rfloor]$, and $A_7 = [6\lfloor T/7 \rfloor + 1, T]$. Thus, the segments between the change points are of size roughly $T/7, T/14, T/7, T/14, T/7, 2T/7$ and $T/7$.

*d) Scenario 4:* Let $A_1[1, \lfloor T/3 \rfloor]$, $A_2 = [\lfloor T/3 \rfloor + 1, 2\lfloor T/3 \rfloor]$, and $A_3 = [2\lfloor T/3 \rfloor + 1, T]$. Then the observations are constructed as $X(t) \overset{\text{i.i.d.}}{\sim} N(0, 1.25 I_p)$ for $t \in A_1 \cup A_3$, and for $t \in A_2$ we have

$$X(t)|\{u_t = 1\} \overset{\text{i.i.d.}}{\sim} N(0.5 \cdot \mathbb{1}, I_p)$$

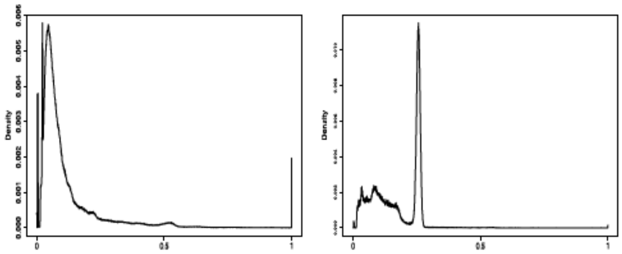

Fig. 2. Densities taken from Padilla *et al.* [38] and used in Scenario 5.

TABLE I
SCENARIO 1

| Method | Metric | T = 300 p = 20 | T = 300 p = 10 | T = 150 p = 20 | T = 150 p = 10 |
|---|---|---|---|---|---|
| MNP | $\lvert K-\widehat{K}\rvert$ | **0.0** | **0.0** | **0.0** | **0.0** |
| EMNCP | $\lvert K-\widehat{K}\rvert$ | 0.1 | **0.0** | **0.0** | **0.0** |
| KCPA | $\lvert K-\widehat{K}\rvert$ | 1.4 | 0.9 | 1.3 | 1.0 |
| SBS | $\lvert K-\widehat{K}\rvert$ | 2.0 | 2.0 | 2.0 | 2.0 |
| DCBS | $\lvert K-\widehat{K}\rvert$ | **0.0** | **0.0** | **0.0** | 0.1 |
| MNP | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | 1.0 | 2.0 | 1.0 | 2.0 |
| EMNCP | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | **0.0** | **1.0** | **0.0** | **0.0** |
| KCPA | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | 175.0 | 119.0 | 60.0 | 62.0 |
| SBS | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | $\infty$ | $\infty$ | $\infty$ | $\infty$ |
| DCBS | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | 1.0 | **1.0** | 1.0 | 3.0 |
| MNP | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | 1.0 | 2.0 | 1.0 | 2.0 |
| EMNCP | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | **0.0** | **1.0** | **0.0** | **0.0** |
| KCPA | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | 1.0 | 19 | 1.0 | 13.0 |
| SBS | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | $-\infty$ | $-\infty$ | $-\infty$ | $-\infty$ |
| DCBS | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | 1.0 | **1.0** | 1.0 | 3.0 |

and

$$X(t)|\{u_t = 2\} \overset{\text{i.i.d.}}{\sim} N(-0.5 \cdot 1, I_p),$$

where the i.i.d. random variables $\{u_t\}$ satisfy $\mathbb{P}(u_t = 1) = \mathbb{P}(u_t = 2) = 1/2$.

*e) Scenario 5:* The vector $X(t)$ satisfies $X_j(t) \sim g_1$ for $t \in A_1 \cap A_3$ and for all $j \in \{1, \ldots, p\}$. In contrast, if $t \in A_2$ we have that

$$X_j(t) \sim \begin{cases} g_2, & j \in \{1, 2\}, \\ g_1, & \text{otherwise.} \end{cases}$$

Here $g_1$ and $g_2$ are the densities shown in the left and right panels in Figure 2, respectively. Moreover, the sets $A_1$, $A_2$ and $A_3$ are the same as in Scenario 4.

Figure 1 illustrates examples of data generated from each of the scenarios that we consider. This is complemented by the results in Tables I–V. Specifically, we observe that for Scenario 1, a setting with mean changes, the best methods seem to be MNP, DCBS and EMNCP.

Interestingly, from Table II, we see that KCPA and MNP outperform the other methods. This setting presents a bigger challenge than Scenario 1, as it involves a heavy-tailed distribution of the errors and smaller changes in mean.

Scenario 3 posses a situation where the mean remains constant and the covariance structure changes. From Table III, we observe that MNP ttains the best performance, followed by EMNCP.

TABLE II
SCENARIO 2

| Method | Metric | T = 300 p = 20 | T = 300 p = 10 | T = 150 p = 20 | T = 150 p = 10 |
|---|---|---|---|---|---|
| MNP | $\lvert K-\widehat{K}\rvert$ | **3.4** | **2.1** | 5.7 | **4.8** |
| EMNCP | $\lvert K-\widehat{K}\rvert$ | 3.5 | 4.7 | 5.9 | 5.9 |
| KCPA | $\lvert K-\widehat{K}\rvert$ | 4.9 | 4.6 | **5.0** | 5.1 |
| SBS | $\lvert K-\widehat{K}\rvert$ | 6.0 | 6.0 | 6.0 | 6.0 |
| DCBS | $\lvert K-\widehat{K}\rvert$ | 6.0 | 6.0 | 6.0 | 5.9 |
| MNP | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | **160.0** | **45** | $\infty$ | 85 |
| EMNCP | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | 167.0 | $\infty$ | $\infty$ | $\infty$ |
| KCPA | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | 212.0 | 225 | **107** | 111.0 |
| SBS | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | $\infty$ | $\infty$ | $\infty$ | $\infty$ |
| DCBS | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | $\infty$ | $\infty$ | $\infty$ | $\infty$ |
| MNP | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | 12.0 | 18.0 | $-\infty$ | **3.0** |
| EMNCP | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | **3.0** | $-\infty$ | $-\infty$ | $-\infty$ |
| KCPA | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | 4.0 | **17** | **3.0** | 4.0 |
| SBS | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | $-\infty$ | $-\infty$ | $-\infty$ | $-\infty$ |
| DCBS | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | $-\infty$ | $-\infty$ | $-\infty$ | $-\infty$ |

TABLE III
SCENARIO 3

| Method | Metric | T = 300 p = 20 | T = 300 p = 10 | T = 150 p = 20 | T = 150 p = 10 |
|---|---|---|---|---|---|
| MNP | $\lvert K-\widehat{K}\rvert$ | **4.5** | **4.6** | **5.4** | **5.2** |
| EMNCP | $\lvert K-\widehat{K}\rvert$ | **4.5** | 4.9 | 5.8 | 5.7 |
| KCPA | $\lvert K-\widehat{K}\rvert$ | 6.0 | 6.0 | 6.0 | 5.9 |
| SBS | $\lvert K-\widehat{K}\rvert$ | 6.0 | 6.0 | 6.0 | 6.0 |
| DCBS | $\lvert K-\widehat{K}\rvert$ | 6.0 | 6.0 | 6.0 | 6.0 |
| MNP | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | **208.0** | **208.0** | **103** | **104** |
| EMNCP | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | 210.0 | 210.0 | $\infty$ | $\infty$ |
| KCPA | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | $\infty$ | $\infty$ | $\infty$ | $\infty$ |
| SBS | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | $\infty$ | $\infty$ | $\infty$ | $\infty$ |
| DCBS | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | $\infty$ | $\infty$ | $\infty$ | $\infty$ |
| MNP | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | **1.0** | 2.0 | **2.0** | **2.0** |
| EMNCP | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | **1.0** | **1.0** | $-\infty$ | $-\infty$ |
| KCPA | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | $-\infty$ | $-\infty$ | $-\infty$ | $-\infty$ |
| SBS | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | $-\infty$ | $-\infty$ | $-\infty$ | $-\infty$ |
| DCBS | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | $-\infty$ | $-\infty$ | $-\infty$ | $-\infty$ |

In Table IV, we also see the advantage of the MNP method which is the best at estimating the number of change points. This is in the context of Scenario 4 where the mean and covariance remain unchanged and the jumps happen in the shape of the distribution.

Finally, Scenario 5 is an example of a model that does not belong to a usual parametric family. In such setting, Table V shows that MNP, KCPA and EMNCP seem to provide better estimation of the number of change points and their locations as compared to the other two methods.

Overall, we can see that SBS and DCBS, two methods designed for mean change point detection, are not robust in the cases where the changes are not in mean or the noise is not sub-Gaussian. MNP, EMNCP and KCPA are, arguably, the best performing methods. KCPA and EMNCP are competitive and sometimes outperform MNP.

TABLE IV
SCENARIO 4

| Method | Metric | T = 300 p = 20 | T = 300 p = 10 | T = 150 p = 20 | T = 150 p = 10 |
|---|---|---|---|---|---|
| MNP | $\|K - \widehat{K}\|$ | **0.7** | **0.9** | **1.1** | 1.4 |
| EMNCP | $\|K - \widehat{K}\|$ | 1.8 | 1.8 | 2.0 | 1.8 |
| KCPA | $\|K - \widehat{K}\|$ | 1.2 | 1.1 | 1.4 | **1.2** |
| SBS | $\|K - \widehat{K}\|$ | 2.0 | 2.0 | 2.0 | 2.0 |
| DCBS | $\|K - \widehat{K}\|$ | 2.0 | 2.0 | 1.9 | 2.0 |
| MNP | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | **38.0** | **68.0** | **43.0** | **65.0** |
| EMNCP | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | $\infty$ | $\infty$ | $\infty$ | $\infty$ |
| KCPA | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | 99.0 | 112.0 | 72.0 | 81.0 |
| SBS | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | $\infty$ | $\infty$ | $\infty$ | $\infty$ |
| DCBS | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | $\infty$ | $\infty$ | $\infty$ | $\infty$ |
| MNP | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | 36.0 | **33.0** | 7.0 | **6.0** |
| EMNCP | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | $-\infty$ | $-\infty$ | $-\infty$ | $\infty$ |
| KCPA | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | **3.0** | 34 | **1.0** | 22.0 |
| SBS | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | $-\infty$ | $-\infty$ | $-\infty$ | $-\infty$ |
| DCBS | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | $-\infty$ | $-\infty$ | $-\infty$ | $-\infty$ |

TABLE V
SCENARIO 5

| Method | Metric | T = 300 p = 20 | T = 300 p = 10 | T = 150 p = 20 | T = 150 p = 10 |
|---|---|---|---|---|---|
| MNP | $\|K - \widehat{K}\|$ | **0.8** | **0.7** | 1.3 | **1.0** |
| EMNCP | $\|K - \widehat{K}\|$ | 1.4 | 0.8 | 1.6 | 1.2 |
| KCPA | $\|K - \widehat{K}\|$ | 1.1 | 1.2 | **1.1** | 1.1 |
| SBS | $\|K - \widehat{K}\|$ | 2.0 | 2.0 | 2.0 | 2.0 |
| DCBS | $\|K - \widehat{K}\|$ | 2.0 | 2.0 | 2.0 | 1.9 |
| MNP | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | **26** | 46.0 | **50.0** | **49.0** |
| EMNCP | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | $\infty$ | **22.0** | $\infty$ | 66.0 |
| KCPA | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | 170.0 | 156.0 | 69 | 55.0 |
| SBS | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | $\infty$ | $\infty$ | $\infty$ | $\infty$ |
| DCBS | $d(\widehat{\mathcal{C}}\|\mathcal{C})$ | $\infty$ | $\infty$ | $\infty$ | $\infty$ |
| MNP | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | **26** | 34.0 | **6.0** | 9.0 |
| EMNCP | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | $-\infty$ | **1.0** | $-\infty$ | 1.0 |
| KCPA | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | 41.0 | 8.0 | 17 | 3.0 |
| SBS | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | $-\infty$ | $-\infty$ | $-\infty$ | $-\infty$ |
| DCBS | $d(\mathcal{C}\|\widehat{\mathcal{C}})$ | $-\infty$ | $-\infty$ | $-\infty$ | $-\infty$ |

### B. Real Data Example

The experiments section concludes with an example using financial data. Specifically, our data consist of the daily close stock price, from Jan-1-2016 to Aug-11-2019, of the 20 companies with highest average stock price from the S&P500 market. The data can be downloaded from Microsoft Corp. (MSFT) [37]. Our final dataset is then a matrix $X \in \mathbb{R}^{T \times p}$, with $T = 907$ and $p = 20$.

We then run MNP, KCPA, and procedure and the estimator from Matteson and James [36]. The implementation and details are the same as those in Section V-A. Our goal is to detect potential change points in the period aforementioned and determine if they might have a financial meaning.

We find that our estimator localizes change points at the dates May-17-2016, Mar-2-2017, Aug-7-2017, Dec-21-2017, Jun-1-2018 and Jan-24-2019. The first change point seems to correspond with the moment when President Donald Trump, while still a presidential candidate, outlined his plan for the USA vs. China trade war (see e.g. Burns *et al.*, 2019). The second change point, Feb-21-2017, might be associated with Trump signing two executive orders increasing tariffs on the trade with China; the date Aug-7-2017 could correspond to the bipartite agreement on Jul-19 2017 to reduce USA deficit with China; the date Dec-21-2017 could be explained by the threats and tariffs imposed by Mr. Trump to China in January of 2018. The other two dates are also relatively close to important dates in the USA vs. China trade war time-line. The raw data, scaled to the interval $[0, 1]$, and the estimated change points can be seen in the right panel in the third row in Figure 1.

As for EMNCP, we find a total of 22 change points with spacings between 30 and 58 units of time. This might suggest that some of the change point are spurious as the minimum spacing parameter of the function `e.divisive()` is by default set to 30.

Finally, when comparing with KCPA, we notice that the scores that the function `KernSeg_MultiD()` outputs do not have an inflection point. We suspect that this could be avoided with a better choice of the tuning parameters. Nevertheless, we look at the model provided by KCPA consisting of six change points, the same number of change points that MNP produced. The six change points estimated by KCPA corresponds to the dates Jun-29-2016, Apr-27-2017, Oct-5-2017, Jan-10-2018, Jun-1-2018, and Mar-14-2019. These are roughly similar to the dates estimated by MNP with an exact match on Jun-1-2018.

## VI. DISCUSSIONS

In this paper, we tackle a multivariate nonparametric change point detection problem, which aims to provide with change point estimators robust against model mis-specification. The computational-efficient method we propose has matched minimax lower bounds, off by logarithm factors, in terms of both the signal-to-noise ratio condition and the localization rate. The lower bounds are also presented in this paper, which is self-contained. The theoretical findings are backed up by extensive numerical experiments, including a real data example.

The distributions in this paper are assumed to have Lebesgue densities. For distributions of discrete data, one may consider using other kernels designed to estimate discrete distributions (e.g. Rajagopalan and Lall [46] and Kokonendji and Kiesse [33]). The algorithms developed in this paper can be straightforwardly extended to the discrete distributions by using these kernels, but the theoretical results would rely on different techniques from those in Section A.

Another possible extension of this this paper is to characterize change points by other measures, instead of the supreme norm of the density function differences. Different measures would require different methods, the algorithmic efficiency and theoretical optimality are remained interesting and open.

Finally, we would like to reiterate that the purpose of this paper is to estimate the locations of the change points accurately. If one only wishes to estimate the number of change points accurately, without requiring upper bounding the localization errors, then this can be formed as a testing problem, for which we conjecture that a consistent result may hold under a weaker signal-to-noise ratio condition.

## APPENDIX A
## LARGE PROBABILITY EVENTS

In this section, we deal with all the large probability events occurred in the proof of Theorem 1. Lemma 4 is almost identical to Theorem 2.1 in Bousquet [6] and therefore we omit the proof. Lemma 5 is an adaptation of Theorem 2.3 in Bousquet [6] and Proposition 8 in Kim *et al.* [32], but we allow for non-i.i.d. cases. Lemma 6 is a non-i.i.d. version of Proposition 2.1 in Giné and Guillou [23]. Lemma 7 is to control the deviance between the sample and population quantities and provides an lower bound on a large probability event. Lemma 8 is to provide a lower bound on the probability of the event that the data can reach the maxima closely enough. Lemma 9 is identical to Lemma 13 in Wang *et al.* [54], controlling the random intervals selected in Algorithm 1.

*Lemma 4:* Let $\mathcal{D}$ be the $\sigma$-field generated by $\{X(i)\}_{i=1}^T$, $\mathcal{D}_T^t$ be the $\sigma$-field generated by $\{X(i)\}_{i=1}^T \setminus \{X(t)\}$ and $\mathbb{E}_T^t(\cdot)$ be the conditional expectation given $\mathcal{D}_T^t$, for all $t \in \{1, \ldots, T\}$. Let $(Z, Z_1', \ldots, Z_T')$ be a sequence of $\mathcal{D}$-measurable random variables, and $\{Z_k\}_{k=1}^T$ be a sequence of random variables such that $Z_k$ measurable with respect to $\mathcal{D}_T^k$, for all $k$. Assume that there exists $u > 0$ such that for all $k = 1, \ldots, T$, the following inequalities hold

$$Z_k' \leq Z - Z_k \text{ a.s.}, \quad \mathbb{E}_T^k(Z_k') \geq 0 \quad \text{and} \quad Z_k' \leq u \quad \text{a.s.}. \quad (15)$$

Let $\sigma$ be a real value satisfying $\sigma^2 \geq \sum_{k=1}^T \mathbb{E}_T^k\{(Z_k')^2\}$ almost surely and let $\nu = (1+u)\mathbb{E}(Z) + \sigma^2$. If

$$\sum_{k=1}^T (Z - Z_k) \leq Z \quad \text{a.s.}, \quad (16)$$

then for all $x > 0$,

$$\mathbb{P}\{Z \geq \mathbb{E}(Z) + \sqrt{2\nu x} + x/3\} \leq e^{-x}.$$

*Lemma 5:* Assume that $\{X(i)\}_{i=1}^T$ satisfy Assumption 1. Let $\mathcal{F}$ be a class of functions from $\mathbb{R}^p$ to $\mathbb{R}$ that is separable in $L_\infty(\mathbb{R}^p)$. Suppose all functions $g \in \mathcal{F}$ are measurable with respect to $P_{\eta_k}$, $k \in \{1, \ldots, K+1\}$, and there exist $B$, $\sigma > 0$ such that for all $g \in \mathcal{F}$

$$\mathbb{E}_{P_{\eta_k}}\{g^2\} - (\mathbb{E}_{P_{\eta_k}}\{g\})^2 \leq \sigma^2 \quad \text{and} \quad \|g\|_\infty \leq B.$$

Let $Z = \sup_{g \in \mathcal{F}} \left|\sum_{i=1}^T w_i[g(X(i)) - \mathbb{E}_{P_i}\{g(X(i))\}]\right|$, with $\sum_{i=1}^T w_i^2 = 1$ and $\max_{i=1,\ldots,T} |w_i| = w$. Then for any $\varepsilon > 0$ , we have

$$\mathbb{P}\left\{Z \geq \mathbb{E}(Z) + \sqrt{2\{(1+wB)\mathbb{E}(Z) + \sigma^2\}x} + x/3\right\} \leq e^{-x}.$$

*Proof:* For all $k \in \{1, \ldots, T\}$, define

$$Z_k = \sup_{g \in \mathcal{F}} \left|\sum_{i \neq k} w_i[g(X(i)) - \mathbb{E}_{P_i}\{g(X(i))\}]\right|$$

and

$$Z_k' = \left|\sum_{i=1}^T w_i[g_k(X(i)) - \mathbb{E}_{P_i}\{g_k(X(i))\}]\right| - Z_k,$$

where $g_k$ denotes the function for which the supremum is obtained in $Z_k$. We then have

$$\begin{aligned} Z_k' &\leq Z - Z_k \\ &\leq |\textstyle\sum_{i=1}^T w_i[g_0(X(i)) - \mathbb{E}_{P_i}\{g_0(X(i))\}]| - \\ &\quad |\textstyle\sum_{i \neq k} w_i[g_0(X(i)) - \mathbb{E}_{P_i}\{g_0(X(i))\}]| \\ &\leq |w_k[g_0(X(k)) - \mathbb{E}_{P_k}\{g_0(X(k))\}]| \leq wB \quad \text{a.s.}, \end{aligned}$$

where $g_0$ is the function for which the supremum is obtained in $Z$. Moreover, we have

$$\begin{aligned} \mathbb{E}_T^k(Z_k') &\geq \left|\sum_{i=1}^T \mathbb{E}_T^k\{w_i(g_k(X(i)) - \mathbb{E}_{P_i}\{g_k(X(i))\})\}\right| - Z_k \\ &= 0, \end{aligned}$$

which concludes the proof of (15) with $u = B$. In addition,

$$\begin{aligned} (T-1)Z &= \left|\sum_{k=1}^T \sum_{i \neq k} w_i[g_0(X(i)) - \mathbb{E}_{P_i}\{g_k(X(i))\}]\right| \\ &\leq \sum_{k=1}^T \left|\sum_{i \neq k} w_i[g_0(X(i)) - \mathbb{E}_{P_i}\{g_k(X(i))\}]\right| \leq \sum_{k=1}^T Z_k, \end{aligned}$$

which leads to (16). Finally, since

$$\begin{aligned} \textstyle\sum_{k=1}^T \mathbb{E}_T^k\{(Z_k')^2\} &\leq \textstyle\sum_{k=1}^T \mathrm{Var}_T^k\{w_k g_k(X(k))\} \\ &\leq \max_k \sup_g \mathrm{Var}\{g(X(k))\} \\ &\leq \sigma^2, \end{aligned}$$

it follows due to Lemma 4 that

$$\mathbb{P}\left\{Z \geq \mathbb{E}(Z) + \sqrt{2\{(1+wB)\mathbb{E}(Z) + \sigma^2\}x} + x/3\right\} \leq e^{-x},$$

for all $x > 0$. □

*Lemma 6:* Let $\mathcal{F}$ be a uniformly bounded VC class of functions, and measurable with respect to all $P_{\eta_k}$, $k = 1, \ldots, K+1$. Suppose

$$\sup_{g \in \mathcal{F}} \mathrm{Var}_{P_{\eta_k}}(g) \leq \sigma^2, \quad \sup_{g \in \mathcal{F}} \|g\|_\infty \leq B, \quad \text{and} \quad 0 < \sigma \leq B.$$

Then there exist positive constants $A$ and $\nu$ depending on $\mathcal{F}$ but not on $\{P_{\eta_k}\}_{k=1}^{K+1}$ or $T$, such that for all $T \in \mathbb{N}$,

$$\begin{aligned} &\sup_{g \in \mathcal{F}} \mathbb{E}\left\|\sum_{i=1}^T w_i\{g(X_i) - \mathbb{E}(g(X_i))\}\right\| \\ &\leq C\left\{\nu wB \log(2AwB/\sigma) + \sqrt{\nu}\sigma\sqrt{\log(2AwB/\sigma)}\right\}, \end{aligned}$$

where $C$ is a universal constant, $\sum_{i=1}^T w_i^2 = 1$ and $\max_{i=1,\ldots,T} |w_i| = w$.

The proof of Lemma 6 is almost identical to that of Propostion 2.1 in Giné and Guillou [23], except noticing that $\sum_{i=1}^T w_i^2 = 1$.

For any $x \in \mathbb{R}^p$, $0 \le s < t < e \le T$ and $h > 0$, define

$$\widetilde{f}^{s,e}_{t,h}(x) = \sqrt{\frac{e-t}{(e-s)(t-s)}} \sum_{j=s+1}^{t} f_{j,h}(x) - \sqrt{\frac{t-s}{(e-s)(e-t)}} \sum_{j=t+1}^{e} f_{j,h}(x), \qquad (17)$$

where

$$f_{j,h}(x) = h^{-p}\mathbb{E}\left\{ \mathcal{k}\left(\frac{x - X(j)}{h}\right)\right\}$$

and the expectation is taken with respect to the distribution $P_j$.

*Lemma 7:* Define the events

$$\mathcal{A}_1(\gamma, h) = \left\{ \max_{0 \le s < t - h^{-p} < t + h^{-p} < e \le T} \sup_{z \in \mathbb{R}^p} \left| \widetilde{Y}^{s,e}_t(z) - \widetilde{f}^{s,e}_{t,h}(z) \right| \le \gamma \right\}$$

and

$$\mathcal{A}_2(\gamma, h) = \left\{ \max_{0 \le s < t - h^{-p} < t + h^{-p} < e \le T} \sup_{z \in \mathbb{R}^p} \frac{1}{\sqrt{e-s}} \left| \sum_{j=s+1}^{e} (\hat{f}_{j,h}(z) - f_{j,h}(z)) \right| \le \gamma \right\}.$$

Under Assumptions 1 and 2, we have that

$$\mathbb{P}\left\{ \mathcal{A}_1\left( Ch^{-p/2}\sqrt{\log(T)}, h\right)\right\} \ge 1 - T^{-c}$$

and

$$\mathbb{P}\left\{ \mathcal{A}_2\left( Ch^{-p/2}\sqrt{\log(T)}, h\right)\right\} \ge 1 - T^{-c},$$

where $C, c > 0$ are absolute constants depending on $\|\mathcal{k}\|_\infty$, $A$ and $\nu$.

We remark that the proof here is an adaptation of Theorem 12 in Kim *et al.* [32].

*Proof:* For any fixed $x \in \mathbb{R}^p$, it holds that

$$\widetilde{Y}^{s,e}_t(x) - \widetilde{f}^{s,e}_{t,h}(x) = \sum_{j=s+1}^{e} w_j \left[ h^{-p}\mathcal{k}\left(\frac{x - X(j)}{h}\right) - \mathbb{E}\left\{ h^{-p}\mathcal{k}\left(\frac{x - X(j)}{h}\right)\right\}\right], \qquad (18)$$

where

$$w_j = \begin{cases} \sqrt{\frac{e-t}{(e-s)(t-s)}}, & j = s+1, \ldots, t, \\ -\sqrt{\frac{t-s}{(e-s)(e-t)}}, & j = t+1, \ldots, e, \end{cases}$$

satisfying that

$$\sum_{j=s+1}^{e} w_j^2 = 1 \quad \text{and} \quad \max_{j=s+1,\ldots,e} |w_j| \le h^{p/2}.$$

**Step 1.** Let $\mathcal{K}_{x,h} : \mathbb{R}^p \to \mathbb{R}$ be $\mathcal{K}_{x,h}(\cdot) = \mathcal{k}(h^{-1}x - h^{-1}\cdot)$ and

$$\widetilde{\mathcal{F}}_{\mathcal{k},h} = \{ h^{-p}\mathcal{K}_{x,h} : x \in \mathcal{X}\}$$

be a class of normalized kernel functions centred on $\mathcal{X}$ and bandwidth $h$. It follows from (18) that, for each $s, t, e$,

$$\sup_{x \in \mathcal{X}} \left| \widetilde{Y}^{s,e}_t(x) - \widetilde{f}^{s,e}_{t,h}(x)\right| = \sup_{g \in \widetilde{\mathcal{F}}_{\mathcal{k},h}} \left| \sum_{j=s+1}^{e} w_j \left[ g(X(j)) - \mathbb{E}\{g(X(j))\}\right]\right| = W_{s,t,e}.$$

It is immediate to check that for any $g \in \widetilde{\mathcal{F}}_{\mathcal{k},h}$,

$$\|g\|_\infty \le h^{-p}\|\mathcal{k}\|_\infty.$$

Due to the arguments used in Theorem 12 in Kim *et al.* [32] and Assumption 2 (i), for every probability measure $Q$ on $\mathbb{R}^p$ and for every $\zeta \in (0, h^{-p}\|\mathcal{k}\|_\infty)$, the covering number $\mathcal{N}(\widetilde{F}_{\mathcal{k},h}, L_2(Q), \zeta)$ is upper bounded as

$$\sup_Q \mathcal{N}(\widetilde{F}_{\mathcal{k},h}, L_2(Q), \zeta) \le \left( \frac{2Ap\|\mathcal{k}\|_\infty}{h^p \zeta}\right)^{\nu+2}.$$

Under Assumption 2, due to Lemma 11 in Kim *et al.* [32], it holds that for any $j = 1, \ldots, T$,

$$\mathbb{E}\left\{ \left( h^{-p}\mathcal{K}_{x,h}(X(j))\right)^2\right\} \le C_1 h^{-p},$$

where $C_1$ is an absolute constant.

It follows from Lemma 5 that for any $x > 0$,

$$\mathbb{P}\Big\{ W_{s,t,e} < \mathbb{E}(W_{s,t,e}) + \sqrt{2\{(1 + h^{-p/2}\|\mathcal{k}\|_\infty)\mathbb{E}(W_{s,t,e}) + C_1 h^{-p}\}x} + x/3 \Big\} \ge 1 - e^{-x}. \qquad (19)$$

**Step 2.** We then need to bound $\mathbb{E}(W_{s,t,e})$, where the expectation is taken on the product of $P_1 \otimes \ldots \otimes P_T$. Let $\check{\mathcal{F}} = \{ g - a : g \in \widetilde{\mathcal{F}}_{\mathcal{k},h}, a \in [-h^{-p}\|\mathcal{k}\|_\infty, h^{-p}\|\mathcal{k}\|_\infty]\}$. Then for any $a \in [-h^{-p}\|\mathcal{k}\|_\infty, h^{-p}\|\mathcal{k}\|_\infty]$, it follows from the proof of Theorem 30 in Kim *et al.* [32] that

$$\sup_P \mathcal{N}(\check{\mathcal{F}}, L_2(P), a) \le (2Ah^{-p}\|\mathcal{k}\|_\infty / a)^{\nu+1}.$$

Applying Lemma 6, we have

$$\mathbb{E}(W_{s,t,e}) \le C\left\{ (\nu+1)\frac{\|\mathcal{k}\|_\infty}{h^{p/2}} \log\left( \frac{8Ah^{-p/2}\|\mathcal{k}\|_\infty}{C_1^{1/2}h^{-p/2}}\right) + h^{-p/2}\sqrt{(\nu+1)\log\left( \frac{8Ah^{-p/2}\|\mathcal{k}\|_\infty}{C_1^{1/2}h^{-p/2}}\right)}\right\}. \qquad (20)$$

**Step 3.** We now plug (20) into (19) and take $x = \log(T^m)$, with $m > 4$, resulting in

$$\mathbb{P}\left\{ W_{s,t,e} < C_2 h^{-p/2}\log^{1/2}(T)\right\} \ge 1 - C_3 T^{-m},$$

where $C_2, C_3 > 0$ are absolute constants depending on $\|\mathcal{k}\|_\infty$, $A$ and $\nu$. The final claims follow with a union bound argument over $s, t, e$.

□

*Lemma 8:* Under Assumptions 1, 2 and 3, for $s < t < e$, define

$$z^*_{s,e,t} \in \arg\max_{z \in \mathbb{R}^p} \left| \widetilde{f}^{s,e}_t(z)\right|.$$

With $h = c_h\kappa$, define the event

$$\mathcal{B}(\gamma) = \left\{ \max_{\substack{0\le s<t<e\le T\\ e-s\le C_R\Delta}} \left| \max_{j=1,\ldots,T} \left|\widetilde{f}^{s,e}_{t,h}(X(j))\right| - \left|\widetilde{f}^{s,e}_{t,h}(z^*_{s,e,t})\right| \right| \le \gamma,\ (s,e) \text{ satisfies Condition } \mathcal{SE} \right\},$$

where Condition $\mathcal{SE}$ is defined as follows: the interval $(s,e)$ is such that either

(a) there is no true change point in $(s,e)$; or

(b) there exists at least one true change point in $\eta_k \in (s,e)$ satisfying

$$\min\left\{\min_{\eta_k>s}\{\eta_k - s\},\ \min_{\eta_k<e}\{e-\eta_k\}\right\} > c_1\Delta,$$

for some $c_1 > 0$;

(c) there exists one and only one change point $\eta_k \in (s,e)$ satisfying

$$\min\{\eta_k - s,\ e-\eta_k\} \le C_\epsilon \log(T) V_p^2 \kappa^{-p}\kappa_k^{-2};$$

or

(d) there exist exactly two change points $\eta_k, \eta_{k+1} \in (s,e)$ with $\eta_k < \eta_{k+1}$ satisfying

$$\eta_k - s \le C_\epsilon \log(T) V_p^2\kappa^{-p}\kappa_k^{-2}, \quad \text{and}$$
$$e - \eta_{k+1} \le C_\epsilon \log(T) V_p^2\kappa^{-p}\kappa_k^{-2}.$$

Then for

$$\gamma = C_\gamma h\sqrt{\Delta}, \tag{21}$$

with

$$C_\gamma > 2C_{\mathrm{Lip}}\sqrt{C_R}, \tag{22}$$

it holds that

$$\mathbb{P}\left\{\mathcal{B}(\gamma)\right\} \ge 1 - T^3\exp\left\{-\frac{\Delta}{8}\left(\frac{c\gamma}{4\sqrt{C_R\Delta}C_{\mathrm{Lip}}}\right)^{p+1} V_p\right\},$$

for some constant $c > 0$.

*Proof:* Fix $0 \le s < t < e \le T$ with $e-s \le C_R\Delta$.

For case (a), it holds that $\widetilde{f}^{s,e}_{t,h}(x) = 0$, for all $x \in \mathbb{R}^p$, and the claim holds consequently.

For case (b), if $\left|\widetilde{f}^{s,e}_{t,h}(z^*_{s,e,t})\right| < \gamma$, then by the definition of $z^*_{s,e,t}$, we have that

$$\begin{aligned}
&\left| \max_{j=1,\ldots,T}\left|\widetilde{f}^{s,e}_{t,h}(X(j))\right| - \left|\widetilde{f}^{s,e}_{t,h}(z^*_{s,e,t})\right| \right| \\
&= \left|\widetilde{f}^{s,e}_{t,h}(z^*_{s,e,t})\right| - \max_{j=1,\ldots,T}\left|\widetilde{f}^{s,e}_{t,h}(X(j))\right| \\
&\le \left|\widetilde{f}^{s,e}_{t,h}(z^*_{s,e,t})\right| < \gamma,
\end{aligned}$$

which implies that

$$\mathbb{P}\left\{\left| \max_{j=1,\ldots,T}\left|\widetilde{f}^{s,e}_{t,h}(X(j))\right| - \left|\widetilde{f}^{s,e}_{t,h}(z^*_{s,e,t})\right| \right| > \gamma\right\} = 0. \tag{23}$$

If $\left|\widetilde{f}^{s,e}_{t,h}(z^*_{s,e,t})\right| > \gamma$, then

$$\gamma < \left|\widetilde{f}^{s,e}_{t,h}(z^*_{s,e,t})\right| \le 2\min\{\sqrt{t-s},\ \sqrt{e-t}\}\max_{j=1,\ldots,T}|f_{j,h}(z^*_{s,e,t})|, \tag{24}$$

there exists $j_0 \in \{1,\ldots,K+1\}$ such that

$$\begin{aligned}
f_{\eta_{j_0}}(z^*_{s,e,t}) &\ge f_{\eta_{j_0},h}(z^*_{s,e,t}) - C_{\mathrm{Lip}}h \\
&\ge \frac{\gamma}{2\min\{\sqrt{t-s},\ \sqrt{e-t}\}} - C_{\mathrm{Lip}}h \\
&\ge \frac{c\gamma}{2\min\{\sqrt{t-s},\ \sqrt{e-t}\}},
\end{aligned} \tag{25}$$

where $0 < c < 1$ is an absolute constant, the first inequality follows from (34), the second inequality follows from (24), and the last inequality follows from Assumption 3 and the choice of $\gamma$.

As for the function $\widetilde{f}^{s,e}_{t,h}(\cdot)$, for any $x_1, x_2 \in \mathbb{R}^p$, it holds that

$$\begin{aligned}
&\left|\widetilde{f}^{s,e}_{t,h}(x_1) - \widetilde{f}^{s,e}_{t,h}(x_2)\right| \\
&= \Bigg|\sqrt{\frac{e-t}{(e-s)(t-s)}}\sum_{j=s+1}^{t}\int_{\mathbb{R}^p} k(y)\{f_j(x_1-hy) - f_j(x_2-hy)\}\,dy \ - \\
&\quad \sqrt{\frac{t-s}{(e-s)(e-t)}}\sum_{j=t+1}^{e}\int_{\mathbb{R}^p} k(y)\{f_j(x_1-hy) - f_j(x_2-hy)\}\,dy\Bigg| \\
&\le 2\min\{\sqrt{e-t},\ \sqrt{t-s}\}C_{\mathrm{Lip}}\|x_1-x_2\|,
\end{aligned} \tag{26}$$

where the last inequality follows from Assumption 1. As a result, the function $\widetilde{f}^{s,e}_{t,h}(\cdot)$ is Lipschitz with constant $2\min\{\sqrt{e-t},\ \sqrt{t-s}\}C_{\mathrm{Lip}}$. Furthermore, defining

$$d_{j_0} = \left|\left\{j \in \{\eta_{j_0-1}+1,\ldots,\eta_{j_0}\} :\ \|X(j) - z^*_{s,e,t}\| \le \frac{\gamma}{2\min\{\sqrt{t-s},\sqrt{e-t}\}C_{\mathrm{Lip}}}\right\}\right|,$$

and noticing that

$$d_{j_0} \sim \text{Binomial}\left(\eta_{j_0+1} - \eta_{j_0},\ \int_{B(z^*_{s,e,t},a)} f_{\eta_{j_0}}(z)\,dz\right),$$

where

$$a := \frac{\gamma}{2\min\{\sqrt{t-s},\sqrt{e-t}\}C_{\mathrm{Lip}}},$$

we arrive at

$$\begin{aligned}
&\mathbb{P}\left\{\left| \max_{j=1,\ldots,T}\left|\widetilde{f}^{s,e}_{t,h}(X(j))\right| - \left|\widetilde{f}^{s,e}_{t,h}(z^*_{s,e,t})\right| \right| > \gamma\right\} \\
&= \mathbb{P}\left\{\min_{j=1,\ldots,T}\left|\widetilde{f}^{s,e}_{t,h}(X(j)) - \widetilde{f}^{s,e}_{t}(z^*_{s,e,t})\right| > \gamma\right\} \\
&\le \mathbb{P}\left\{\min_{j=1,\ldots,T}\|X(j) - z^*_{s,e,t}\| > \frac{\gamma}{2\min\{\sqrt{e-t},\ \sqrt{t-s}\}C_{\mathrm{Lip}}}\right\} \\
&\le \mathbb{P}\{d_{j_0} = 0\},
\end{aligned} \tag{27}$$

where the identity follows from the definition of $z^*_{s,e,t}$, the first inequality follows from (26) and the second inequality follows from the definition of $d_{j_0}$.

In addition, we have that

$$\begin{aligned}
&\int_{B\left(z^*_{s,e,t},\, \frac{\gamma}{2\min\{\sqrt{t-s},\, \sqrt{e-t}\}C_{\mathrm{Lip}}}\right)} f_{\eta_{j_0}}(z)\, dz \\
&\geq \int_{B\left(z^*_{s,e,t},\, \frac{c\gamma}{4\min\{\sqrt{t-s},\, \sqrt{e-t}\}C_{\mathrm{Lip}}}\right)} f_{\eta_{j_0}}(z)\, dz \\
&\geq \int_{B\left(z^*_{s,e,t},\, \frac{c\gamma}{4\min\{\sqrt{t-s},\, \sqrt{e-t}\}C_{\mathrm{Lip}}}\right)} \Big\{ f_{\eta_{j_0}}(z^*_{s,e,t}) - \\
&\quad \frac{c\gamma}{4\min\{\sqrt{t-s},\, \sqrt{e-t}\}C_{\mathrm{Lip}}} \Big\}\, dz \\
&\geq \left( \frac{c\gamma}{4\min\{\sqrt{t-s},\, \sqrt{e-t}\}C_{\mathrm{Lip}}} \right)^{p+1} V_p, \qquad (28)
\end{aligned}$$

where the last inequality is due to (25). Therefore,

$$\begin{aligned}
&\mathbb{P}\{d_{j_0} = 0\} \\
&\leq \mathbb{P}\Big\{ d_{j_0} \leq \frac{\Delta}{2} \left( \frac{c\gamma}{4\min\{\sqrt{t-s},\, \sqrt{e-t}\}C_{\mathrm{Lip}}} \right)^{p+1} V_p \Big\} \\
&\leq \mathbb{P}\left\{ d_{j_0} \leq \frac{(\eta_{j_0} - \eta_{j_0-1})}{2} \int_{B(z^*_{s,e,t},\, a)} f_{\eta_{j_0}}(z)\, dz \right\} \\
&\leq \exp\left\{ - \frac{(\eta_{j_0} - \eta_{j_0-1})}{8} \int_{B(z^*_{s,e,t},\, a)} f_{\eta_{j_0}}(z)\, dz \right\} \\
&\leq \exp\left\{ - \frac{\Delta}{8} \left( \frac{c\gamma}{4\sqrt{C_R \Delta} C_{\mathrm{Lip}}} \right)^{p+1} V_p \right\}, \qquad (29)
\end{aligned}$$

where the second and the fourth inequality follow from (28), and the third by the Chernoff bound (e.g. Mitzenmacher and Upfal [38]). Combining (23), (27) and (29) results in

$$\begin{aligned}
&\mathbb{P}\left\{ \left| \max_{j=1,\ldots,T} \left| \widetilde{f}^{s,e}_{t,h}(X(j)) \right| - \left| \widetilde{f}^{s,e}_{t,h}(z^*_{s,e,t}) \right| \right| > \gamma \right\} \\
&\leq \exp\left\{ -\frac{\Delta}{8} \left( \frac{c\gamma}{4\sqrt{C_R \Delta} C_{\mathrm{Lip}}} \right)^{p+1} V_p \right\}.
\end{aligned}$$

The conclusion follows from a union bound.

Cases (c) and (d) are similar, and we only deal with case (c) here. Note that

$$\left| \widetilde{f}^{s,e}_t(z^*_{s,e,t}) \right| \leq \kappa_k \sqrt{C_\epsilon \log(T) V_p^2 \kappa_k^{-2} \kappa^{-p}} \leq \gamma, \qquad (30)$$

where the first inequality follows from Lemma 13 (i) and the second follows from Assumption 3. The final claim holds due to the fact that $\widetilde{f}^{s,e}_{t,h}$ is a smoothed version of $\widetilde{f}^{s,e}_t$. □

We independently select at random from $\{1, \ldots, T\}$ two sequences $\{\alpha_m\}_{m=1}^{M_1}, \{\beta_m\}_{m=1}^{M_1}$, then we keep the pairs which satisfy $\beta_m - \alpha_m \leq C_R \Delta$, with $C_R \geq 3/2$. For notational simplicity, we label them as $\{\alpha_r\}_{r=1}^{R}, \{\beta_r\}_{r=1}^{R}$. Let

$$\mathcal{M} = \bigcap_{k=1}^{K} \big\{ \alpha_r \in \mathcal{S}_k, \beta_r \in \mathcal{E}_k, \text{ for some } r \in \{1, \ldots, R\} \big\}, \qquad (31)$$

where $\mathcal{S}_k = [\eta_k - 3\Delta/4, \eta_k - \Delta/2]$ and $\mathcal{E}_k = [\eta_k + \Delta/2, \eta_k + 3\Delta/4]$, $k = 1, \ldots, K$. In the lemma below, we give a lower bound on the probability of $\mathcal{M}$.

*Lemma 9:* For the event $\mathcal{M}$ defined in (31), we have

$$\mathbb{P}(\mathcal{M}) \geq 1 - \exp\left\{ \log\left( \frac{T}{\Delta} \right) - \frac{R\Delta}{4 C_R T} \right\}.$$

See Lemma S.24 in Wang *et al.* [55] for the proof of Lemma 9.

## Appendix B
## Change Point Detection Lemmas and the Proof of Theorem 1

Lemma 10 below provides a lower bound on the maximum of the population CUSUM statistic when there exists a true change point. Lemma 11 shows that the maxima of the population CUSUM statistic are the true change points. Lemma 13 is a collection of results on the population quantities. Lemma 14 provides an initial upper bound for the localization error. Lemma 15 is the key lemma to provide the final localization rate. The proof of Theorem 1 is collected at the end of this section.

In the rest of this section, we will adopt the notation

$$\begin{aligned}
\widetilde{f}^{s,e}_t(x) = &\sqrt{\frac{e-t}{(e-s)(t-s)}} \sum_{j=s+1}^{t} f_j(x) - \\
&\sqrt{\frac{t-s}{(e-s)(e-t)}} \sum_{j=t+1}^{e} f_j(x),
\end{aligned}$$

for all $0 \leq s < t < e \leq T$ and $x \in \mathbb{R}^p$.

*Lemma 10:* Under Assumptions 1-3, let $(s, e)$ be an interval such that $e - s \leq C_R \Delta$ and there exists a true change point $\eta_k \in (s, e)$ with

$$\min\{\eta_k - s,\, e - \eta_k\} > c_1 \Delta,$$

where $c_1 > 0$ is a large enough constant, depending on all the other absolute constants. Then for any $h$ such that

$$(\log(T)/\Delta)^{1/p} \leq h \leq \frac{c_1}{C_R C_{\mathrm{Lip}} C_{\mathcal{k}}} \kappa, \qquad (32)$$

it holds that

$$\max_{s + h^{-p} < t < e - h^{-p}} \sup_{z \in \mathbb{R}^p} \left| \widetilde{f}^{s,e}_{t,h}(z) \right| \geq \frac{c_1 \kappa \Delta}{4\sqrt{e-s}}.$$

*Proof:* Let $z_1 \in \arg\max_{z \in \mathbb{R}^p} \left| f_{\eta_k}(z) - f_{\eta_{k+1}}(z) \right|$. Due to Assumption 1, we have that

$$\left| f_{\eta_k}(z_1) - f_{\eta_{k+1}}(z_1) \right| \geq \kappa_k \geq \kappa.$$

Then by the argument in Lemma 2.4 of Venkatraman [51], we have that

$$\max_{t \in \{\eta_k + c_1\Delta/2,\, \eta_k - c_1\Delta/2\}} \left| \widetilde{f}^{s,e}_t(z_1) \right| \geq \frac{c_1 \kappa \Delta}{2\sqrt{e-s}}. \qquad (33)$$

Next, for any $x \in \mathbb{R}^p$, $h > 0$ and $j \in \{1, \ldots, T\}$, we have

$$\begin{aligned}
&\left| f_j(x) - f_{j,h}(x) \right| \\
&= \left| \int_{\mathbb{R}^p} \frac{1}{h^p} \mathcal{k}(y/h) \{ f_j(x-y) - f_j(x) \}\, dy \right| \\
&\leq \frac{C_{\mathrm{Lip}}}{h^p} \int_{\mathbb{R}^p} \mathcal{k}(y/h) \|y\|\, dy \\
&\leq h C_{\mathrm{Lip}} \int_{\mathbb{R}^p} \mathcal{k}(z) \|z\|\, dz \ \leq C_{\mathrm{Lip}} C_{\mathcal{k}} h, \qquad (34)
\end{aligned}$$

where the last inequality follows from Assumption 2 (iii). Hence, for $t \in \{\eta_k + c_1\Delta/2, \eta_k - c_1\Delta/2\}$

$$\left|\widetilde{f}^{s,e}_{t,h}(z_1) - \widetilde{f}^{s,e}_{t}(z_1)\right| \leq C_{\mathrm{Lip}} C_k h \sqrt{\frac{(e-t)(t-s)}{e-s}} \leq \sqrt{(e-s)} C_{\mathrm{Lip}} C_k h \leq \frac{c_1 \kappa \Delta}{4\sqrt{e-s}}, \tag{35}$$

which follows from (32). Finally, the claim follows combining (33) and (35). □

*Lemma 11:* Under Assumption 1, for any interval $(s, e) \subset (0, T)$ satisfying

$$\eta_{k-1} \leq s \leq \eta_k \leq \ldots \leq \eta_{k+q} \leq e \leq \eta_{k+q+1}, \quad q \geq 0.$$

Let

$$b \in \underset{t=s+1,\ldots,e}{\arg\max} \sup_{x \in \mathbb{R}^p} \left|\widetilde{f}^{s,e}_{t,h}(x)\right|.$$

If

$$h \leq \frac{\kappa}{4 C_{\mathrm{Lip}} C_k},$$

then $b \in \{\eta_1, \ldots, \eta_K\}$.

For any fixed $z \in \mathbb{R}^p$, if $\widetilde{f}^{s,e}_{t,h}(z) > 0$ for some $t \in (s,e)$, then $\widetilde{f}^{s,e}_{t,h}(z)$ is either strictly monotonic or decreases and then increases within each of the interval $(s, \eta_k), (\eta_k, \eta_{k+1}), \ldots, (\eta_{k+q}, e)$.

*Proof:* We prove by contradiction. Assume that $b \notin \{\eta_1, \ldots, \eta_K\}$. Let $z_1 \in \arg\max_{x\in\mathbb{R}^p} \left|\widetilde{f}^{s,e}_{b,h}(x)\right|$. Due to the definition of $b$, we have

$$b \in \underset{t=s+1,\ldots,e}{\arg\max} \left|\widetilde{f}^{s,e}_{t,h}(z_1)\right|.$$

It is easy to see that the collection of the change points of $\{f_{t,h}(z_1)\}_{t=s+1}^{e}$ is a subset of the change points of $\{f_{t,h}\}_{t=s+1}^{e}$. In addition, due to (34), it holds that

$$\min_{k=1,\ldots,K+1} \|f_{\eta_k,h} - f_{\eta_{k-1},h}\|_\infty \geq \kappa - 2C_{\mathrm{Lip}} C_k h \geq \kappa/2,$$

which implies that the collection of the change points of $\{f_{t,h}\}_{t=s+1}^{e}$ is the collection of the change points of $\{f_t\}_{t=s+1}^{e}$.

It follows from Lemma 2.2 in Venkatraman [51] that

$$\widetilde{f}^{s,e}_{b,h}(z_1) < \max_{j\in\{k,\ldots,k+q\}} \widetilde{f}^{s,e}_{\eta_j,h}(z_1) \leq \max_{t=s+1,\ldots,e} \sup_{x\in\mathbb{R}^p} \left|\widetilde{f}^{s,e}_{t,h}(x)\right|,$$

which is a contradiction. □

Recall that in Algorithm 1, when searching for change points in the interval $(s, e)$, we actually restrict to values $t \in (s + h^{-p}, e - h^{-p})$. We now show that for intervals satisfying condition $\mathcal{SE}$ from Lemma 8, taking the maximum of the CUSUM statistic over $(s + h^{-p}, e - h^{-p})$ is equivalent to searching on $(s, e)$, when there are change points in $(s + h^{-p}, e - h^{-p})$.

*Lemma 12:* Suppose that Assumptions 1 and 3 hold, and the events $\mathcal{A}_1(\gamma_{\mathcal{A}})$ and $\mathcal{B}(\gamma_{\gamma_{\mathcal{B}}})$ happens where

$$\gamma_{\mathcal{A}} = C h^{-p/2}\sqrt{\log(T)}, \quad \text{and} \quad \gamma_{\mathcal{B}} = C_\gamma h \sqrt{\Delta}$$

with $C$ as in Lemma 7, and $C_\gamma$ as in (21). Let $(s, e) \subset (0, T)$ satisfy $e - s \leq C_R \Delta$. Assume that Condition $\mathcal{SE}$ from Lemma 8 holds, and that

$$\eta_{k-1} \leq s \leq \eta_k \leq \ldots \leq \eta_{k+q} \leq e \leq \eta_{k+q+1}, \quad q \geq 0.$$

Then

$$\underset{t=s+h^{-p},\ldots,e-h^{-p}}{\arg\max} \sup_{x\in\mathbb{R}^p} \left|\widetilde{f}^{s,e}_{t,h}(x)\right| = \underset{t=s+1,\ldots,e}{\arg\max} \sup_{x\in\mathbb{R}^p} \left|\widetilde{f}^{s,e}_{t,h}(x)\right|, \tag{36}$$

and

$$\arg\max_{t=s+h^{-p},\ldots,e-h^{-p}} \max_{j=1,\ldots,T} \left|\widetilde{Y}^{s,e}_t(X(j))\right| = \arg\max_{t=s+1,\ldots,e} \max_{j=1,\ldots,T} \left|\widetilde{Y}^{s,e}_t(X(j))\right|. \tag{37}$$

*Proof:* Firs notice that, due to Lemma 10, there exists $\eta_k \in (s, e)$ such that

$$\sup_{z\in\mathbb{R}^p} \left|\widetilde{f}^{s,e}_{\eta_k,h}(z)\right| \geq \frac{c_1\kappa\Delta}{4\sqrt{e-s}}.$$

Furthermore, if

$$t \in (s,e) \backslash (s + \max\{h^{-p}, C_\epsilon \log(T) V_p^2 \kappa_k^{-2} \kappa^{-p}\}, e - \max\{h^{-p}, C_\epsilon \log(T) V_p^2 \kappa_k^{-2} \kappa^{-p}\}), \tag{38}$$

then

$$\begin{aligned} &\sup_{z\in\mathbb{R}^p} \left|\widetilde{f}^{s,e}_{t,h}(z)\right| \\ &\leq 2\sqrt{\min\{e-t,\, t-s\}} \max_{t=,1\ldots,T} \sup_{z\in\mathbb{R}^p} |f_{t,h}(z)| \\ &\leq 2\max\left\{h^{-p/2}, \sqrt{C_\epsilon \log(T) V_p^2 \kappa_k^{-2} \kappa^{-p}}\right\} \cdot \\ &\quad \max_{t=,1\ldots,T} \sup_{z\in\mathbb{R}^p} |f_{t,h}(z)| < \frac{c_1\kappa\Delta}{32\sqrt{e-s}}, \end{aligned}$$

where the last inequality follows from Assumption 3. Therefore, (36) follows.

As for (37), we notice that

$$\begin{aligned} \max_{j=1,\ldots,T} \left|\widetilde{Y}^{s,e}_{\eta_k}(X(j))\right| &\geq \sup_{z\in\mathbb{R}^p} \left|\widetilde{f}^{s,e}_{\eta_k,h}(z)\right| - \gamma_{\mathcal{A}} - \gamma_{\mathcal{B}} \\ &\geq \frac{c_1\kappa\Delta}{4\sqrt{e-s}} - \gamma_{\mathcal{A}} - \gamma_{\mathcal{B}} \\ &\geq \frac{c_1\kappa\Delta}{8\sqrt{e-s}}. \end{aligned}$$

Moreover, for $t$ satisfying (38), we have

$$\begin{aligned} &\max_{j=1,\ldots,T} \left|\widetilde{Y}^{s,e}_t(X(j))\right| \\ &\leq \sup z \in \mathbb{R}^p \left|\widetilde{f}^{s,e}_{t,h}(z)\right| + \gamma_{\mathcal{A}} + \gamma_{\mathcal{B}} \\ &\leq 2\sqrt{\min\{e-t,\, t-s\}} \max_{t=1,\ldots,T} \sup_{z\in\mathbb{R}^p} |f_{t,h}(z)| + \gamma_{\mathcal{A}} + \gamma_{\mathcal{B}} \\ &\leq 2\max\left\{h^{-p/2}, \sqrt{C_\epsilon \log(T) V_p^2 \kappa_k^{-2} \kappa^{-p}}\right\} \cdot \\ &\quad \max_{t=1.,\ldots,T} \sup_{z\in\mathbb{R}^p} |f_{t,h}(z)| + \gamma_{\mathcal{A}} + \gamma_{\mathcal{B}} \\ &< \frac{c_1\kappa\Delta}{16\sqrt{e-s}}, \end{aligned}$$

and the claim follows once again using Assumption 3. □

*Lemma 13:* Under Assumptions 1 and 2, the following statements hold.

(i) If $\eta_k$ is the only change point in $(s,e)$, then for any $h$,

$$\sup_{x\in\mathbb{R}^p}\left|\widetilde{f}^{s,e}_{\eta_k,h}(s)\right| \leq \kappa_k \min\left\{\sqrt{s-\eta_k},\, \sqrt{e-\eta_k}\right\}. \quad (39)$$

(ii) Suppose $e-s \leq C_R\Delta$, where $C_R > 0$ is an absolute constant, and that

$$\eta_{k-1} \leq s \leq \eta_k \leq \ldots \leq \eta_{k+q} \leq e \leq \eta_{k+q+1}, \quad q \geq 0. \quad (40)$$

Denote

$$\kappa^{s,e}_{\max} = \max\left\{\sup_{x\in\mathbb{R}^p}\left|f_{\eta_p}(x) - f_{\eta_{p-1}}(x)\right| : k \leq p \leq k+q\right\}.$$

Then for any $k-1 \leq p \leq k+q$, it holds that

$$\sup_{x\in\mathbb{R}^p}\left|\frac{1}{e-s}\sum_{i=s+1}^{e} f_{i,h}(x) - f_{\eta_p,h}(x)\right| \leq C_R\kappa^{s,e}_{\max}. \quad (41)$$

(iii) Assume (40) and $q \geq 1$. If

$$\eta_k - s \leq c_1\Delta, \quad (42)$$

for $c_1 > 0$, then for any $h$,

$$\begin{aligned}\sup_{z\in\mathbb{R}^p} |\widetilde{f}^{s,e}_{\eta_k,h}(z)| \leq \sqrt{c_1}\sup_{z\in\mathbb{R}^p}|\widetilde{f}^{s,e}_{\eta_{k+1},h}(z)| \\ + 2\kappa_k\sqrt{\eta_k - s} \\ + 4\sqrt{\eta_k - s}C_{\mathrm{Lip}}C_{\mathcal{k}}h, \quad (43)\end{aligned}$$

where $C_{\mathcal{k}} > 0$ is an absolute constant only depending on the kernel function.

(iv) Assume (40) and $q = 1$, then

$$\begin{aligned}&\max_{t=s+1,\ldots,e}\sup_{z\in\mathbb{R}^p}\left|\widetilde{f}^{s,e}_{t,h}(z)\right| \\ &\leq 2\sqrt{e-\eta_k}\kappa_{k+1} + 2\sqrt{\eta_k - s}\kappa_k + \\ &\quad 4\sqrt{\eta_k - s}C_{\mathrm{Lip}}C_{\mathcal{k}}h + 4\sqrt{e-\eta_k}C_{\mathrm{Lip}}C_{\mathcal{k}}h.\end{aligned}$$

*Proof:* Note that for (i),

$$\begin{aligned}&\sup_{x\in\mathbb{R}^p}\left|\widetilde{f}^{s,e}_{\eta_k,h}(x)\right| \\ &= \sqrt{\frac{(e-\eta_k)(\eta_k-s)}{e-s}}\sup_{x\in\mathbb{R}^p}\left|\int_{\mathbb{R}^p}\mathcal{k}(y)\{f_{\eta_k}(x-hy) - \right. \\ &\quad \left. f_{\eta_{k+1}}(x-hy)\}\,dy\right| \\ &\leq \kappa_k\min\left\{\sqrt{s-\eta_k},\, \sqrt{e-\eta_k}\right\}.\end{aligned}$$

The claim (ii) follows from the same arguments used in showing (i) and Lemmas 17 and 19 in Wang *et al.* [54]. For the claim (iii), we define

$$\widetilde{g}^{s,e}_{t,h} = \begin{cases} f_{\eta_{k+1},h}, & t = s+1,\ldots,\eta_k, \\ f_{t,h}, & t = \eta_k + 1,\ldots,e. \end{cases}$$

Thus,

$$\begin{aligned}&\left|\widetilde{f}^{s,e}_{\eta_k,h}\right| \leq \left|\widetilde{g}^{s,e}_{\eta_k,h}\right| + \sqrt{\frac{(e-\eta_k)(\eta_k-s)}{e-s}}(f_{\eta_{k+1},h} - f_{\eta_k,h}) \\ &\leq \sqrt{\frac{(\eta_k-s)(e-\eta_{k+1})}{(\eta_{k+1}-s)(e-\eta_k)}}\left|\widetilde{g}^{s,e}_{\eta_{k+1},h}\right| + \\ &\quad \sqrt{\frac{(e-\eta_k)(\eta_k-s)}{e-s}}(f_{\eta_{k+1},h} - f_{\eta_k,h}) \\ &\leq \sqrt{c_1}\left|\widetilde{g}^{s,e}_{\eta_{k+1},h}\right| + \sqrt{\frac{(e-\eta_k)(\eta_k-s)}{e-s}}(f_{\eta_{k+1},h} - f_{\eta_k,h}) \\ &\leq \sqrt{c_1}\left|\widetilde{f}^{s,e}_{\eta_{k+1},h}\right| + 2\sqrt{\frac{(e-\eta_k)(\eta_k-s)}{e-s}}(f_{\eta_{k+1},h} - f_{\eta_k,h}) \\ &\leq \sqrt{c_1}\left|\widetilde{f}^{s,e}_{\eta_{k+1},h}\right| + 2\sqrt{\eta_k - s}\kappa_k + 4\sqrt{\eta_k - s}C_{\mathrm{Lip}}C_{\mathcal{k}}h,\end{aligned}$$

where the first, second and fourth inequalities follow from the definition of $\widetilde{g}^{s,e}_{t,h}$, the second follows from (42) and the last follows from (34).

As for (iv), we define

$$\widetilde{q}^{s,e}_{t,h} = \begin{cases} f_{\eta_k,h}, & t = s+1,\ldots,\eta_k, \\ f_t, & t = \eta_k + 1,\ldots,e. \end{cases}$$

For any $t \geq \eta_k$, it holds that

$$\widetilde{f}^{s,e}_{t,h} - \widetilde{q}^{s,e}_{t,h} = \sqrt{\frac{e-t}{(e-s)(t-s)}}(\eta_k - s)(f_{\eta_k,h} - f_{\eta_{k-1},h}).$$

Therefore, for $t \geq \eta_k$,

$$\begin{aligned}&\max_{t=s+1,\ldots,e}|\widetilde{f}^{s,e}_{t,h}| \\ &\leq \max\{|\widetilde{f}^{s,e}_{\eta_k,h}|, |\widetilde{f}^{s,e}_{\eta_{k+1},h}|\} \leq \max_{t=s+1,\ldots,e}|\widetilde{q}^{s,e}_{t,h}| \\ &\quad + 2\sqrt{\eta_k - s}\kappa_k + 4\sqrt{\eta_k - s}C_{\mathrm{Lip}}C_{\mathcal{k}}h \\ &\leq 2\sqrt{e-\eta_k}\kappa_{k+1} + 2\sqrt{\eta_k - s}\kappa_k \\ &\quad + 4\sqrt{\eta_k - s}C_{\mathrm{Lip}}C_{\mathcal{k}}h + 4\sqrt{e-\eta_k}C_{\mathrm{Lip}}C_{\mathcal{k}}h.\end{aligned}$$

□

*Lemma 14:* Let $z_0 \in \mathbb{R}^p$, $(s,e) \subset (0,T)$. Suppose that there exits a true change point $\eta_k \in (s,e)$ such that

$$\min\{\eta_k - s,\, e - \eta_k\} \geq c_1\Delta, \quad (44)$$

and

$$\left|\widetilde{f}^{s,e}_{\eta_k,h}(z_0)\right| \geq (c_1/4)\frac{\kappa\Delta}{\sqrt{e-s}}, \quad (45)$$

where $c_1 > 0$ is a sufficiently small constant. In addition, assume that

$$\max_{t=s+1,\ldots,e}\left|\widetilde{f}^{s,e}_{t,h}(z_0)\right| - \left|\widetilde{f}^{s,e}_{\eta_k,h}(z_0)\right| \leq c_2\Delta^4(e-s)^{-7/2}\kappa, \quad (46)$$

where $c_2 > 0$ is a sufficiently small constant.

Then for any $d \in (s,e)$ satisfying

$$|d - \eta_k| \leq c_1\Delta/32, \quad (47)$$

it holds that

$$\left|\widetilde{f}^{s,e}_{\eta_k,h}(z_0)\right| - \left|\widetilde{f}^{s,e}_{d,h}(z_0)\right| > c|d-\eta_k|\Delta\left|\widetilde{f}^{s,e}_{\eta_k,h}(z_0)\right|(e-s)^{-2},$$

where $c > 0$ is a sufficiently small constant, depending on all the other absolute constants.

*Proof:* Without loss of generality, we assume that $d \geq \eta_k$ and $\widetilde{f}^{s,e}_{\eta_k,h}(z_0) \geq 0$. Following the arguments in Lemma 2.6 in Venkatraman [51], it suffices to consider two cases: (i) $\eta_{k+1} > e$ and (ii) $\eta_{k+1} \leq e$.

**Case (i).** Note that

$$\widetilde{f}^{s,e}_{\eta_k,h}(z_0) = \sqrt{\frac{(e-\eta_k)(\eta_k-s)}{e-s}}\left\{f_{\eta_k,h}(z_0) - f_{\eta_{k+1},h}(z_0)\right\}$$

and

$$\widetilde{f}^{s,e}_{d,h}(z_0)=(\eta_k-s)\sqrt{\frac{e-d}{(e-s)(d-s)}}\left\{f_{\eta_k,h}(z_0)-f_{\eta_{k+1},h}(z_0)\right\}.$$

Therefore, it follows from (44) that

$$\widetilde{f}^{s,e}_{\eta_k,h}(z_0) - \widetilde{f}^{s,e}_{d,h}(z_0) = \left(1-\sqrt{\frac{(e-d)(\eta_k-s)}{(d-s)(e-\eta_k)}}\right)\widetilde{f}^{s,e}_{\eta_k,h}(z_0) \geq c\Delta|d-\eta_k|(e-s)^{-2}\widetilde{f}^{s,e}_{\eta_k,h}(z_0).$$

The inequality follows from the following arguments. Let $u=\eta_k-s$, $v=e-\eta_k$ and $w=d-\eta_k$. Then

$$\begin{aligned}
&1-\sqrt{\frac{(e-d)(\eta_k-s)}{(d-s)(e-\eta_k)}} - c\Delta|d-\eta_k|(e-s)^2\\
=&1-\sqrt{\frac{(v-w)u}{(u+w)v}} - c\frac{\Delta w}{(u+v)^2}\\
=&\frac{w(u+v)}{\sqrt{(u+w)v}(\sqrt{(v-w)u}+\sqrt{(u+w)v})} - c\frac{\Delta w}{(u+v)^2}.
\end{aligned}$$

The numerator of the above equals

$$\begin{aligned}
&w(u+v)^3 - c\Delta w(u+w)v - c\Delta w\sqrt{uv(u+w)(v-w)}\\
\geq&2c_1\Delta w\left\{(u+v)^2-\frac{c(u+w)v}{2c_1}-\frac{c\sqrt{uv(u+w)(v-w)}}{2c_1}\right\}\\
\geq&2c_1\Delta w\left\{(1-c/(2c_1))(u+v)^2 - 2^{-1/2}c/c_1 uv\right\} > 0,
\end{aligned}$$

as long as

$$c < \frac{\sqrt{2}c_1}{4+1/(\sqrt{2}c_1)}.$$

**Case (ii).** Let $g = c_1\Delta/16$. We can write

$$\widetilde{f}^{s,e}_{\eta_k,h}(z_0) = a\sqrt{\frac{e-s}{(\eta_k-s)(e-\eta_k)}},$$
$$\widetilde{f}^{s,e}_{\eta_k+g,h}(z_0) = (a+g\theta)\sqrt{\frac{e-s}{(e-\eta_k-g)(\eta_k+g-s)}},$$

where

$$a = \sum_{j=s+1}^{\eta_k}\left\{f_{j,h}(z_0) - \frac{1}{e-s}\sum_{j=s+1}^{e}f_{j,h}(z_0)\right\},$$
$$\theta = \frac{a\sqrt{(\eta_k+g-s)(e-\eta_k-g)}}{g}\left\{\frac{1}{\sqrt{(\eta_k-s)(e-\eta_k)}} - \frac{1}{(\eta_k+g-s)(e-\eta_k-g)} + \frac{b}{a\sqrt{e-s}}\right\},$$

and $b = \widetilde{f}^{s,e}_{\eta_k+g,h}(z_0) - \widetilde{f}^{s,e}_{\eta_k,h}(z_0)$.

To ease notation, let $d-\eta_k = l \leq g/2$, $N_1 = \eta_k - s$ and $N_2 = e-\eta_k-g$. We have

$$E_l = \widetilde{f}^{s,e}_{\eta_k,h}(z_0) - \widetilde{f}^{s,e}_{d,h}(z_0) = E_{1l}(1+E_{2l}) + E_{3l}, \tag{48}$$

where

$$E_{1l} = \frac{al(g-l)\sqrt{e-s}}{\sqrt{N_1(N_2+g)}\sqrt{(N_1+l)(g+N_2-l)}} \cdot \frac{1}{\left(\sqrt{(N_1+l)(g+N_2-l)}+\sqrt{N_1(g+N_2)}\right)}$$
$$E_{2l} = \frac{(N_2-N_1)(N_2-N_1-l)}{\left(\sqrt{(N_1+l)(g+N_2-l)}+\sqrt{(N_1+g)N_2}\right)} \cdot \frac{1}{\left(\sqrt{N_1(g+N_2)}+\sqrt{(N_1+g)N_2}\right)},$$

and

$$E_{3l} = -\frac{bl}{g}\sqrt{\frac{(N_1+g)N_2}{(N_1+l)(g+N_2-l)}}.$$

Next, we notice that $g-l \geq c_1\Delta/32$. It holds that

$$E_{1l} \geq c_{1l}|d-\eta_k|\Delta\widetilde{f}^{s,e}_{\eta_k,h}(z_0)(e-s)^{-2}, \tag{49}$$

where $c_{1l} > 0$ is a sufficiently small constant depending on $c_1$. As for $E_{2l}$, due to (47), we have

$$E_{2l} \geq -1/2. \tag{50}$$

As for $E_{3l}$, we have

$$\begin{aligned}
E_{3l} &\geq -c_{3l,1}b|d-\eta_k|(e-s)\Delta^{-2}\\
&\geq -c_{3l,2}b|d-\eta_k|\Delta^{-3}(e-s)^{3/2}\widetilde{f}^{s,e}_{\eta_k,h}(z_0)\kappa^{-1}\\
&\geq -c_{1l}/2|d-\eta_k|\Delta\widetilde{f}^{s,e}_{\eta_k,h}(z_0)(e-s)^{-2},
\end{aligned}$$

where the second inequality follows from (45) and the third inequality follows from (46), $c_{3l,1}, c_{3l,2} > 0$ are sufficiently small constants, depending on all the other absolute constants.

Combining (48), (49), (50) and (51), we have

$$\widetilde{f}^{s,e}_{\eta_k,h}(z_0) - \widetilde{f}^{s,e}_{d,h}(z_0) \geq c|d-\eta_k|\Delta\widetilde{f}^{s,e}_{\eta_k,h}(z_0)(e-s)^{-2}, \tag{51}$$

where $c > 0$ is a sufficiently small constant.

In view of (B) and (51), the proof is complete. □

*Lemma 15:* Under Assumptions 1, 2 and 3, let $(s_0, e_0)$ be an interval with $e_0 - s_0 \leq C_R\Delta$ and containing at least one change point $\eta_l$ such that

$$\eta_{l-1} \leq s_0 \leq \eta_l \leq \ldots \leq \eta_{l+q} \leq e_0 \leq \eta_{l+q+1}, \quad q \geq 0.$$

Suppose that there exists $k'$ such that

$$\min\{\eta_{k'} - s_0,\ e_0 - \eta_{k'}\} \geq \Delta/16.$$

Let

$$\kappa^{\max}_{s_0,e_0} = \max\{\kappa_p : \min\{\eta_p - s_0,\ e_0 - \eta_p\} \geq \Delta/16\}.$$

Consider any generic $(s,e) \subset (s_0,e_0)$, satisfying

$$\min_{l:\,\eta_l\in(s,e)}\min\{\eta_l - s_0, e_0 - \eta_l\} \geq \Delta/16.$$

Let

$$b \in \underset{t=s+h^{-p},\ldots,e-h^{-p}}{\arg\max}\max_{j=1,\ldots,T}\left|\widetilde{Y}^{s,e}_t(X(j))\right|.$$

Assume

$$h \le \frac{\kappa}{16 C_R C_{\mathrm{Lip}} C_{k}}, \tag{52}$$

where $C_{k} > 0$ is an absolute constant depending only on the kernel function. For some $c_1 > 0$ and $\gamma > 0$, suppose that

$$\max_{j=1,\ldots,T} \left| \widetilde{Y}_b^{s_0,e_0}(X(j)) \right| \ge c_1 \kappa_{s,e}^{\max} \sqrt{\Delta}. \tag{53}$$

Then on the event $\mathcal{A}_1(\gamma_{\mathcal{A}}) \cap \mathcal{A}_2(\gamma_{\mathcal{A}}) \cap \mathcal{B}(\gamma_{\mathcal{B}})$, defined in Lemmas 7 and 8, where

$$\max\{\gamma_{\mathcal{A}},\, \gamma_{\mathcal{B}}\} \le c_2 \kappa \sqrt{\Delta}, \tag{54}$$

with a sufficiently small constant $0 < c_2 < c_1/4$, there exists a change point $\eta_k \in (s,e)$ such that

$$\min\{e - \eta_k,\, \eta_k - s\} \ge \Delta/4 \quad \text{and} \quad |\eta_k - b| \le C \kappa_k^{-2} \gamma_{\mathcal{A}}^2,$$

where $C > 0$ is a sufficiently large constant depending on all the other absolute constants.

*Proof:* Let $z_1 \in \arg\max_{z \in \mathbb{R}^p} \left| \widetilde{f}_{b,h}^{s,e}(z) \right|$. Without loss of generality, assume that $\widetilde{f}_{b,h}^{s,e}(z_1) > 0$ and that $\widetilde{f}_{b,h}^{s,e}(z_1)$ as a function of $t$ is locally decreasing at $b$. Observe that there has to be a change point $\eta_k \in (s,b)$, or otherwise $\widetilde{f}_{b,h}^{s,e}(z_1) > 0$ implies that $\widetilde{f}_{t,h}^{s,e}(z_1)$ is decreasing, as a consequence of Lemma 11.

Thus, there exists a change point $\eta_k \in (s,b)$ satisfying that

$$\begin{aligned}
\sup_{z \in \mathbb{R}^p} \left| \widetilde{f}_{\eta_k,h}^{s,e}(z) \right| &\ge \left| \widetilde{f}_{\eta_k,h}^{s,e}(z_1) \right| > \left| \widetilde{f}_{b,h}^{s,e}(z_1) \right| \\
&\ge \max_{j=1,\ldots,T} \left| \widetilde{f}_{b,h}^{s,e}(X(j)) \right| - \gamma_{\mathcal{B}} \\
&\ge \max_{j=1,\ldots,T} \left| \widetilde{Y}_b^{s,e}(X(j)) \right| - \gamma_{\mathcal{A}} - \gamma_{\mathcal{B}} \\
&\ge c \kappa_k \sqrt{\Delta},
\end{aligned}$$

where the second inequality follows from Lemma 11, the third and fourth inequalities hold on the events $\mathcal{A}_1(\gamma_{\mathcal{A}}, h) \cap \mathcal{A}_2(\gamma_{\mathcal{A}}, h) \cap \mathcal{B}(\gamma_{\mathcal{B}})$, and $c > 0$ is an absolute constant.

Observe that $e - s \le e_0 - s_0 \le C_R \Delta$ and that $(s,e)$ has to contain at least one change point or otherwise $\sup_{z \in \mathbb{R}} |\widetilde{f}_{\eta_k,h}^{s,e}(z)| = 0$ which contradicts (55).

**Step 1.** In this step, we are to show that

$$\min\{\eta_k - s,\, e - \eta_k\} \ge \min\{1, c_1^2\} \Delta/16. \tag{55}$$

Suppose that $\eta_k$ is the only change point in $(s,e)$. Then (55) must hold or otherwise it follows from (39) that

$$\sup_{z \in \mathbb{R}^p} \left| \widetilde{f}_{\eta_k,h}^{s,e}(z) \right| \le \kappa_k \frac{c_1 \sqrt{\Delta}}{4},$$

which contradicts (55).

Suppose $(s,e)$ contains at least two change points. Then arguing by contradiction, if $\eta_k - s < \min\{1,\, c_1^2\}\Delta/16$, it must be the cast that $\eta_k$ is the left most change point in $(s,e)$. Therefore

$$\begin{aligned}
&\sup_{z \in \mathbb{R}^p} \left| \widetilde{f}_{\eta_k,h}^{s,e}(z) \right| \\
&\le c_1/4 \sup_{z \in \mathbb{R}^p} |\widetilde{f}_{\eta_{k+1},h}^{s,e}(z)| + \\
&\quad 2\kappa_k \sqrt{\eta_k - s} + 4\sqrt{\eta_k - s}\, C_{\mathrm{Lip}} C_{k} h \\
&< c_1/4 \max_{s+h^{-p} < t < e-h^{-p}} \sup_{z \in \mathbb{R}^p} |\widetilde{f}_{t,h}^{s,e}(z)| \\
&\quad + \frac{\sqrt{\Delta}}{2} c_1 \kappa_k \\
&\le c_1/4 \max_{s+h^{-p} < t < e-h^{-p}} \max_{j=1,\ldots,T} |\widetilde{f}_{t,h}^{s,e}(X(j))| + \\
&\quad c_1/4 \gamma_{\mathcal{B}} + \frac{\sqrt{\Delta}}{2} c_1 \kappa_k \\
&\le c_1/4 \max_{s+h^{-p} < t < e-h^{-p}} \max_{j=1,\ldots,T} \left| \widetilde{Y}_t^{s,e}(X(j)) \right| + \\
&\quad c_1/4 \gamma_{\mathcal{A}} + c_1/4 \gamma_{\mathcal{B}} + \frac{\sqrt{\Delta}}{2} c_1 \kappa_k \\
&\le \max_{j=1,\ldots,T} \left| \widetilde{Y}_b^{s,e}(X(j)) \right| - \gamma_{\mathcal{A}} - \gamma_{\mathcal{B}},
\end{aligned}$$

where the first inequality follows from (43), the second follows from (52), the third from the definition of the event $\mathcal{B}$, the fourth from the definition of the event $\mathcal{A}$ and the last from (53). The last display contradicts (55), thus (55) must hold.

**Step 2.** Let

$$z_0 \in \arg\max_{z \in \mathbb{R}^p} \left| \widetilde{f}_{\eta_k,h}^{s,e}(z) \right|.$$

It follows from Lemma 14 that there exits $d \in (\eta_k, \eta_k + c_1\Delta/32)$ such that

$$\widetilde{f}_{\eta_k,h}^{s,e}(z_0) - \widetilde{f}_{d,h}^{s,e}(z_0) \ge 2\gamma_{\mathcal{A}} + 2\gamma_{\mathcal{B}}. \tag{56}$$

We claim that $b \in (\eta_k, d) \subset (\eta_k, \eta_k + c_1\Delta/16)$. By contradiction, suppose that $b \ge d$. Then

$$\begin{aligned}
\widetilde{f}_{b,h}^{s,e}(z_0) &\le \widetilde{f}_{d,h}^{s,e}(z_0) \\
&\le \max_{s<t<e} \sup_{z \in \mathbb{R}^p} \left| \widetilde{f}_{t,h}^{s,e}(z) \right| - 2\gamma_{\mathcal{A}} - 2\gamma_{\mathcal{B}} \\
&\le \max_{j=1,\ldots,T} \left| \widetilde{Y}_b^{s,e}(X(j)) \right| - \gamma_{\mathcal{A}} - \gamma_{\mathcal{B}},
\end{aligned}$$

where the first inequality follows from Lemma 11, the second follows from (56) and the third follows from the definition of the event $\mathcal{A}_1(\gamma_{\mathcal{A}}, h) \cap \mathcal{A}_2(\gamma_{\mathcal{A}}, h) \cap \mathcal{B}(\gamma_{\mathcal{B}})$. Note that (B) is a contradiction to the bound in (55), therefore we have $b \in (\eta_k, \eta_k + c_1\Delta/32)$.

**Step 3.** Let

$$\begin{aligned}
&j^* \in \arg\max_{j=1,\ldots,T} \left| \widetilde{Y}_b^{s,e}(X(j)) \right|, \\
&\quad f^{s,e} = (f_{s+1,h}(X(j^*)), \ldots, f_{e,h}(X(j^*)))^\top \in \mathbb{R}^{(e-s)}
\end{aligned}$$

and

$$\begin{aligned}
&Y^{s,e} \\
&= \left( \frac{1}{h^p} k\left( \frac{X(j^*) - X(s)}{h} \right), \ldots, \frac{1}{h^p} k\left( \frac{X(j^*) - X(e)}{h} \right) \right) \in \mathbb{R}^{(e-s)}.
\end{aligned}$$

By the definition of $b$, it holds that

$$\begin{aligned}
&\left\| Y^{s,e} - \mathcal{P}_b^{s,e}(Y^{s,e}) \right\|^2 \\
&\le \left\| Y^{s,e} - \mathcal{P}_{\eta_k}^{s,e}(Y^{s,e}) \right\|^2 \le \left\| Y^{s,e} - \mathcal{P}_{\eta_k}^{s,e}(f^{s,e}) \right\|^2,
\end{aligned}$$

where the operator $\mathcal{P}_{\cdot}^{s,e}(\cdot)$ is defined in Lemma 20 in Wang *et al.* [54]. For the sake of contradiction, throughout the rest of this argument suppose that, for some sufficiently large constant $C_3 > 0$ to be specified,

$$\eta_k + C_3 \gamma_{\mathcal{A}}^2 \kappa_k^{-2} < b. \tag{57}$$

We will show that this leads to the bound

$$\left\| Y^{s,e} - \mathcal{P}_b^{s,e}(Y^{s,e}) \right\|^2 > \left\| Y^{s,e} - \mathcal{P}_{\eta_k}^{s,e}(f^{s,e}) \right\|^2, \tag{58}$$

which is a contradiction. If we can show that

$$2\langle Y^{s,e} - f^{s,e}, \mathcal{P}_b^{s,e}(Y^{s,e}) - \mathcal{P}_{\eta_k}^{s,e}(f^{s,e})\rangle < \left\|f^{s,e} - \mathcal{P}_b^{s,e}(f^{s,e})\right\|^2 - \left\|f^{s,e} - \mathcal{P}_{\eta_k}^{s,e}(f^{s,e})\right\|^2, \quad (59)$$

then (58) holds.

To derive (59) from (57), we first note that $\min\{e-\eta_k, \eta_k - s\} \geq \min\{1, c_1^2\}\Delta/16$ and that $|b-\eta_k| \leq c_1\Delta/32$ implies that

$$\min\{e-b, b-s\} \geq \min\{1, c_1^2\}\Delta/16 - c_1\Delta/32 \geq \min\{1, c_1^2\}\Delta/32.$$

As for the right-hand side of (59), we have

$$\begin{aligned}&\left\|f^{s,e} - \mathcal{P}_b^{s,e}(f^{s,e})\right\|^2 - \left\|f^{s,e} - \mathcal{P}_{\eta_k}^{s,e}(f^{s,e})\right\|^2 \\ &= \left(\widetilde{f}_{\eta_k,h}^{s,e}(X(j^*))\right)^2 - \left(\widetilde{f}_{b,h}^{s,e}(X(j^*))\right)^2 \\ &\geq \left(\widetilde{f}_{\eta_k,h}^{s,e}(X(j^*)) - \widetilde{f}_{b,h}^{s,e}(X(j^*))\right)\left|\widetilde{f}_{\eta_k,h}^{s,e}(X(j^*))\right|.\end{aligned}$$

On the event $\mathcal{A}_1(\gamma_\mathcal{A}, h) \cap \mathcal{A}_2(\gamma_\mathcal{A}, h) \cap \mathcal{B}(\gamma_\mathcal{B})$, we are to use Lemma 14. Note that (45) holds due to the fact that here we have

$$\begin{aligned}\left|\widetilde{f}_{\eta_k,h}^{s,e}(X(j^*))\right| &\geq \left|\widetilde{f}_{b,h}^{s,e}(X(j^*))\right| \\ &\geq \left|\widetilde{Y}_b^{s,e}(X(j^*))\right| - \gamma_\mathcal{A} \\ &\geq c_1\kappa_k\sqrt{\Delta} - \gamma_\mathcal{A} \geq (c_1)/2\kappa_k\sqrt{\Delta}, \quad (60)\end{aligned}$$

where the first inequality follows from the fact that $\eta_k$ is a true change point, the second inequality holds due to the event $\mathcal{A}_1(\gamma_\mathcal{A}, h)$, the third inequality follows from (53), and the final inequality follows from (54). Towards this end, it follows from Lemma 14 that

$$\begin{aligned}&\left|\widetilde{f}_{\eta_k,h}^{s,e}(X(j^*))\right| - \left|\widetilde{f}_{b,h}^{s,e}(X(j^*))\right| \\ &> c|b-\eta_k|\Delta\left|\widetilde{f}_{\eta_k,h}^{s,e}(X(j^*)))\right|(e-s)^{-2}. \quad (61)\end{aligned}$$

Combining (60), (60) and (61), we have

$$\begin{aligned}&\left\|f^{s,e} - \mathcal{P}_b^{s,e}(f^{s,e})\right\|^2 - \left\|f^{s,e} - \mathcal{P}_{\eta_k}^{s,e}(f^{s,e})\right\|^2 \\ &\geq \frac{cc_1^2}{4}\Delta^2\kappa_k\mathcal{A}_1(\gamma_\mathcal{A}, h)^2(e-s)^{-2}|b-\eta_k|. \quad (62)\end{aligned}$$

The left-hand side of (59) can be decomposed as follows.

$$\begin{aligned}&2\langle Y^{s,e} - f^{s,e}, \mathcal{P}_b^{s,e}(Y^{s,e}) - \mathcal{P}_{\eta_k}^{s,e}(f^{s,e})\rangle \\ &= 2\langle Y^{s,e} - f^{s,e}, \mathcal{P}_b^{s,e}(Y^{s,e}) - \mathcal{P}_b^{s,e}(f^{s,e})\rangle + \\ &\quad 2\langle Y^{s,e} - f^{s,e}, \mathcal{P}_b^{s,e}(f^{s,e}) - \mathcal{P}_{\eta_k}^{s,e}(f^{s,e})\rangle \\ &= (I) + \\ &\quad 2\left(\sum_{i=1}^{\eta_k - s} + \sum_{i=\eta_k-s+1}^{b-s} + \sum_{i=b-s+1}^{e-s}\right)\left(Y^{s,e} - f^{s,e}\right)_i \cdot \\ &\quad \left(\mathcal{P}_b^{s,e}(f^{s,e}) - \mathcal{P}_{\eta_k}^{s,e}(f^{s,e})\right)_i \\ &= (I) \ + (II.1) + (II.2) + (II.3).\end{aligned}$$

As for the term (I), we have

$$(I) \ \leq 2\gamma_\mathcal{A}^2. \quad (63)$$

As for the term (II.1), we have

$$\begin{aligned}(II.1) = 2\sqrt{\eta_k - s}&\left\{\frac{1}{\sqrt{\eta_k - s}}\sum_{i=1}^{\eta_k - s}\left(Y^{s,e} - f^{s,e}\right)_i\right\} \\ &\left\{\frac{1}{b-s}\sum_{i=1}^{b-s}(f^{s,e})_i - \frac{1}{\eta_k - s}\sum_{i=1}^{\eta_k - s}(f^{s,e})_i\right\}.\end{aligned}$$

In addition, it holds that

$$\begin{aligned}&\left|\frac{1}{b-s}\sum_{i=1}^{b-s}(f^{s,e})_i - \frac{1}{\eta_k - s}\sum_{i=1}^{\eta_k - s}(f^{s,e})_i\right| \\ &= \frac{b-\eta_k}{b-s}\left|-\frac{1}{\eta_k - s}\sum_{i=1}^{\eta_k - s} f_{i,h}(X(j^*)) + f_{\eta_{k+1},h}(X(j^*))\right| \\ &\leq \frac{b-\eta_k}{b-s}(C_R + 1)\kappa_{s_0,e_0}^{\max},\end{aligned}$$

where the inequality follows from (41). Combining with Lemma 7, it leads to that

$$\begin{aligned}(II.1) &\leq 2\sqrt{\eta_k - s}\frac{b-\eta_k}{b-s}(C_R + 1)\kappa_{s_0,e_0}^{\max}\gamma_\mathcal{A} \\ &\leq 2\frac{4}{\min\{1,\, c_1^2\}}\Delta^{-1/2}\gamma_\mathcal{A}|b-\eta_k|(C_R + 1)\kappa_{s_0,e_0}^{\max}. \quad (64)\end{aligned}$$

As for the term (II.2), it holds that

$$(II.2) \leq 2\sqrt{|b-\eta_k|}\gamma_\mathcal{A}(2C_R + 3)\kappa_{s_0,e_0}^{\max}. \quad (65)$$

As for the term (II.3), it holds that

$$(II.3) \leq 2\frac{4}{\min\{1,\, c_1^2\}}\Delta^{-1/2}\gamma_\mathcal{A}|b-\eta_k|(C_R + 1)\kappa_{s_0,e_0}^{\max}. \quad (66)$$

Therefore, combining (62), (63), (63), (64), (65) and (65), we have that (59) holds if

$$\begin{aligned}&\Delta^2\kappa_k^2(e-s)^{-2}|b-\eta_k| \\ &\gtrsim \max\left\{\gamma_\mathcal{A}^2,\, \Delta^{-1/2}\gamma_\mathcal{A}|b-\eta_k|\kappa_k,\, \sqrt{|b-\eta_k|}\gamma_\mathcal{A}\kappa_k\right\}.\end{aligned}$$

The second inequality holds due to Assumption 3, the third inequality holds due to (57) and the first inequality is a consequence of the third inequality and Assumption 3. □

*Proof of Theorem 1:* Let $\epsilon_k = C_\epsilon \log^{1+\xi}(T)\kappa_k^{-2}\kappa^{-p} \leq \epsilon = C_\epsilon \log^{1+\xi}(T)\kappa^{-(p+2)}$. Since $\epsilon$ is the upper bound of the localization error, by induction, it suffices to consider any interval $(s, e) \subset (0, T)$ that satisfies

$$\eta_{k-1} \leq s \leq \eta_k \leq \ldots \leq \eta_{k+q} \leq e \leq \eta_{k+q+1}, \quad q \geq -1,$$

and

$$\max\{\min\{\eta_k - s,\, s - \eta_{k-1}\},\, \min\{\eta_{k+q+1} - e,\, e - \eta_{k+q}\}\} \leq \epsilon,$$

where $q = -1$ indicates that there is no change point contained in $(s, e)$.

By Assumption 3, it holds that $\epsilon \leq \Delta/4$. It has to be the case that for any change point $\eta_k \in (0, T)$, either $|\eta_k - s| \leq \epsilon$ or $|\eta_k - s| \geq \Delta - \epsilon \geq 3\Delta/4$. This means that $\min\{|\eta_k - s|, |\eta_k - e|\} \leq \epsilon$ indicates that $\eta_k$ is a detected change point in the previous induction step, even if $\eta_k \in (s, e)$. We refer to $\eta_k \in (s,e)$ an undetected change point if $\min\{|\eta_k - s|, |\eta_k - e|\} \geq 3\Delta/4$.

In order to complete the induction step, it suffices to show that we (i) will not detect any new change point in $(s,e)$ if all the change points in that interval have been previous detected, and (ii) will find a point $b \in (s,e)$, such that $|\eta_k - b| \leq \epsilon$ if there exists at least one undetected change point in $(s,e)$.

Define

$$\mathcal{S} = \bigcap_{k=1}^{K} \Big\{ \alpha_s \in [\eta_k - 3\Delta/4, \eta_k - \Delta/2], \quad \beta_s \in [\eta_k + \Delta/2, \eta_k + 3\Delta/4], \text{ for some } s = 1, \ldots, S \Big\}.$$

The rest of the proof assumes the event $\mathcal{A}_1(\gamma_\mathcal{A}) \cap \mathcal{A}_2(\gamma_\mathcal{A}) \cap \mathcal{B}(\gamma_\mathcal{B}) \cap \mathcal{M}$, with

$$\gamma_\mathcal{A} = C_{\gamma_\mathcal{A}} h^{-p/2} \sqrt{\log(T)} \quad \text{and} \quad \gamma_\mathcal{B} = C_{\gamma_\mathcal{B}} h\sqrt{\Delta},$$

and $C_{\gamma_\mathcal{A}}, C_{\gamma_\mathcal{A}} > 0$ are absolute constants. The probability of the event $\mathcal{A}_1(\gamma_\mathcal{A}) \cap \mathcal{A}_2(\gamma_\mathcal{A}) \cap \mathcal{B}(\gamma_\mathcal{B}) \cap \mathcal{M}$ is lower bounded in Lemmas 7, 8 and 9.

**Step 1.** In this step, we will show that we will consistently detect or reject the existence of undetected change points within $(s,e)$. Let $a_r$, $b_r$ and $r^*$ be defined as in Algorithm 1. Suppose there exists a change point $\eta_k \in (s,e)$ such that $\min\{\eta_k - s, e - \eta_k\} \geq 3\Delta/4$. In the event $\mathcal{S}$, there exists an interval $(\alpha_r, \beta_r)$ selected such that $\alpha_r \in [\eta_k - 3\Delta/4, \eta_k - \Delta/2]$ and $\beta_r \in [\eta_k + \Delta/2, \eta_k + 3\Delta/4]$. Following Algorithm 1, $[s_r, e_r] = [\alpha_r, \beta_r] \cap [s,e]$. We have that $\min\{\eta_k - s_r, e_r - \eta_k\} \geq (1/4)\Delta$ and $[s_r, e_r]$ contains at most one true change point.

It follows from Lemma 10, Lemma 12, and Assumption 3, with $c_1$ there chosen to be $1/4$, that

$$\max_{s_r + h^{-p} < t < e_r - h^{-p}} \sup_{z \in \mathbb{R}^p} \left| \widetilde{f}_{t,h}^{s,e}(z) \right| \geq \frac{\kappa\Delta}{16\sqrt{e-s}}.$$

Therefore

$$\begin{aligned} a_r &= \max\nolimits_{s_r + h^{-p} < t < e_r - h^{-p}} \max\nolimits_{j=1,\ldots,T} \left| \widetilde{Y}_t^{s_r,e_r}(X(j)) \right| \\ &\geq \max\nolimits_{s_r + h^{-p} < t < e_r - h^{-p}} \max\nolimits_{j=1,\ldots,T} \left| \widetilde{f}_{t,h}^{s_r,e_r}(X(j)) \right| - \gamma_\mathcal{A} \\ &\geq \max\nolimits_{s_r + h^{-p} < t < e_r - h^{-p}} \sup_{z \in \mathbb{R}^p} \left| \widetilde{f}_{t,h}^{s_r,e_r}(z) \right| - \gamma_\mathcal{A} - \gamma_\mathcal{B} \geq \frac{\kappa\Delta}{16\sqrt{e-s}} - \gamma_\mathcal{A} - \gamma_\mathcal{B}, \end{aligned}$$

where $\gamma_\mathcal{A}$ and $\gamma_\mathcal{B}$ are the same as in (54). Thus for any undetected change point $\eta_k \in (s,e)$, it holds that

$$a_{m^*} = \sup_{1 \leq m \leq S} a_m \geq \frac{\kappa\Delta}{16\sqrt{e-s}} - \gamma_\mathcal{A} - \gamma_\mathcal{B} \geq c_{\tau,2} \kappa \Delta^{1/2}, \tag{67}$$

where $c_{\tau,2} > 0$ is achievable with a sufficiently large $C_{\mathrm{SNR}}$ in Assumption 3. This means we accept the existence of undetected change points.

Suppose that there is no any undetected change point within $(s,e)$, then for any $(s_r, e_r) = (\alpha_r, \beta_r) \cap (s,e)$, one of the following situations must hold.

(a) There is no change point within $(s_r, e_r)$;
(b) there exists only one change point $\eta_k \in (s_r, e_r)$ and $\min\{\eta_k - s_r, e_r - \eta_k\} \leq \epsilon_k$; or
(c) there exist two change points $\eta_k, \eta_{k+1} \in (s_r, e_r)$ and $\eta_k - s_r \leq \epsilon_k$, $e_r - \eta_{k+1} \leq \epsilon_{k+1}$.

Observe that if (a) holds, then we have

$$\begin{aligned} &\max\nolimits_{s_r + h^{-p} < t < e_r - h^{-p}} \max\nolimits_{j=1,\ldots,T} \left| \widetilde{Y}_t^{s_r,e_r}(X(j)) \right| \\ &\leq \max\nolimits_{s_r + h^{-p} < t < e_r - h^{-p}} \sup\nolimits_{z \in \mathbb{R}^p} \left| \widetilde{f}_{t,h}^{s_r,e_r}(z) \right| + \gamma_\mathcal{A} + \gamma_\mathcal{B} \\ &= \gamma_\mathcal{A} + \gamma_\mathcal{B}. \end{aligned}$$

Cases (b) and (c) can be dealt with using similar arguments. We will only work on (c) here. It follows from Lemma 13 (iv) that

$$\begin{aligned} &\max\nolimits_{s_r + h^{-p} < t < e_r - h^{-p}} \max\nolimits_{j=1,\ldots,T} \left| \widetilde{Y}_t^{s_r,e_r}(X(j)) \right| \\ &\leq \max\nolimits_{s_r < t < e_r} \sup\nolimits_{z \in \mathbb{R}^p} \left| \widetilde{f}_{t,h}^{s_r,e_r}(z) \right| + \gamma_\mathcal{A} + \gamma_\mathcal{B} \\ &\leq 2\sqrt{e - \eta_k}\kappa_{k+1} + 2\sqrt{\eta_k - s}\kappa_k + 8\sqrt{\eta_k - s} C_{\mathrm{Lip}} C_k h + \\ &\quad \gamma_\mathcal{A} + \gamma_\mathcal{B} \leq 2(\gamma_\mathcal{A} + \gamma_\mathcal{B}). \end{aligned}$$

Under (11), we will always correctly reject the existence of undetected change points.

**Step 2.** Assume that there exists a change point $\eta_k \in (s,e)$ such that $\min\{\eta_k - s, \eta_k - e\} \geq 3\Delta/4$. Let $s_r$, $e_r$ and $r^*$ be defined as in Algorithm 1. To complete the proof it suffices to show that, there exists a change point $\eta_k \in (s_{r*}, e_{r*})$ such that $\min\{\eta_k - s_{r^*}, \eta_k - e_{r*}\} \geq \Delta/4$ and $|b_{r*} - \eta_k| \leq \epsilon$.

To this end, we are to ensure that the assumptions of Lemma 15 are verified. Note that (53) follows from (67), and (54) follows from Assumption 3.

Thus, all the conditions in Lemma 15 are met, and we therefore conclude that there exists a change point $\eta_k$, satisfying

$$\min\{e_{r*} - \eta_k, \eta_k - s_{r*}\} > \Delta/4 \tag{68}$$

and

$$|b_{r*} - \eta_k| \leq C\kappa_k^{-2}\gamma_\mathcal{A}^2 \leq \epsilon, \tag{69}$$

where the last inequality holds from the choice of $\gamma_\mathcal{A}$ and Assumption 3.

The proof is complete by noticing the fact that (68) and $(s_{r*}, e_{r*}) \subset (s,e)$ imply that

$$\min\{e - \eta_k, \eta_k - s\} > \Delta/4 > \epsilon.$$

As discussed in the argument before **Step 1**, this implies that $\eta_k$ must be an undetected change point. □

## Appendix C
## Proofs of Lemmas 2 and 3

*Proof of Lemma 3:* Consider distributions $F$ and $G$ in $\mathbb{R}^p$ with densities $f$ and $g$, respectively, constructed as follows. The density $f$ is a test function, thus it has compact support and it is infinitely differentiable. Note also that we can take $f$ constant in $B(0, V_p^{-1/p} 2^{-1/p})$, with $f(0) = 1/2$, and with

$$\max\{\|f\|_\infty, \max_x \|\nabla f(x)\|\} \leq \frac{1}{2}. \tag{70}$$

Then, by construction, $f$ is 1-Lipschitz. Let $c_1$ be a constant such that

$$0 < c_1 < V_p^{-1/p} 2^{-1-1/p}, \tag{71}$$

for all $p$, which is possible since $V_p^{-1/p} 2^{p-1-1/p} \to \infty$ as $p \to \infty$. Then define $g$ as

$$g(x) = \begin{cases} \frac{1}{2} + \kappa - c_1^{-1}\|x - p_1\| & \text{if } \|x - p_1\| < \kappa c_1 \\ \frac{1}{2} - \kappa + c_1^{-1}\|x - p_2\| & \text{if } \|x - p_2\| < \kappa c_1 \\ f(x) & \text{otherwise.} \end{cases}$$

where $p_1 = (V_p^{-1/p} 2^{-1/p-1}, 0, \dots, 0) \in \mathbb{R}^p$ and $p_2 = (-V_p^{-1/p} 2^{-1/p-1}, 0, \dots, 0) \in \mathbb{R}^p$. Notice that $g$ is well defined since (71) implies $\kappa c_1 \le C_1 c_1 < V_p^{-1/p} 2^{-1-1/p}$.

Furthermore, by the triangle inequality and (70), $g$ is $C$-Lipschitz for a universal constant $C$. Moreover,

$$\sup_{z \in \mathbb{R}^p} |f(z) - g(z)| = \kappa.$$

Let $P_1$ denote the joint distribution of the independent random variables $\{X(t)\}_{t=1}^T$, where

$$X(1), \dots, X(\Delta) \overset{i.i.d.}{\sim} F \quad \text{and} \quad X(\Delta+1), \dots, X(T) \overset{i.i.d.}{\sim} G;$$

and, similarly, let $P_0$ be the joint distribution of the independent random variables $\{Z(t)\}_{t=1}^T$ such that

$$Z(1), \dots, Z(\Delta+\xi) \overset{i.i.d.}{\sim} F, \quad \text{and}$$
$$Z(\Delta+\xi+1), \dots, Z(T) \overset{i.i.d.}{\sim} G,$$

where $\xi$ is a positive integer no larger than $n - 1 - \Delta$.

Observe that $\eta(P_0) = \Delta$ and $\eta(P_1) = \Delta + \xi$. By Le Cam's Lemma (e.g. Yu [60]) and Lemma 2.6 in Tsybakov [50], it holds that

$$\begin{aligned} \inf_{\hat\eta} \sup_{P \in \mathcal{Q}} \mathbb{E}_P\left(|\hat\eta - \eta|\right) &\ge \xi\{1 - d_{\mathrm{TV}}(P_0, P_1)\} \\ &\ge \frac{\xi}{2} \exp\left(-\mathrm{KL}(P_0, P_1)\right). \end{aligned} \tag{72}$$

Since

$$\mathrm{KL}(P_0, P_1) = \sum_{i \in \{\Delta+1, \dots, \Delta+\xi\}} \mathrm{KL}(P_{0i}, P_{1i}) = \xi \mathrm{KL}(F, G).$$

However,

$$\begin{aligned} &\mathrm{KL}(F, G) \\ &= \frac{1}{2} \int_{B(p_1, \kappa c_1)} \log\left(\frac{1/2}{\frac{1}{2} + \kappa - c_1^{-1}\|x - p_1\|}\right) dx + \\ &\quad \frac{1}{2} \int_{B(p_2, \kappa c_1)} \log\left(\frac{1/2}{\frac{1}{2} - \kappa + c_1^{-1}\|x - p_2\|}\right) dx \\ &= -\frac{1}{2} \int_{B(0, \kappa c_1)} \log\left(1 + 2\kappa - 2c_1^{-1}\|x\|\right) dx - \\ &\quad \frac{1}{2} \int_{B(0, \kappa c_1)} \log\left(1 - 2\kappa + 2c_1^{-1}\|x\|\right) dx \\ &= -\frac{1}{2} \int_{B(0, \kappa c_1)} \log\left(1 - (2\kappa - 2c_1^{-1}\|x\|)^2\right) dx \\ &\le 4\kappa^2 V_p (\kappa c_1)^p \le 4\kappa^{p+2} V_p, \end{aligned}$$

by the inequality $-\log(1 - x) \le 2x$ for $x \in [0, 1/2]$. Therefore,

$$\inf_{\hat\eta} \sup_{P \in \mathcal{Q}} \mathbb{E}_P\left(|\hat\eta - \eta|\right) \ge \frac{\xi}{2} \exp\left(-4\xi\kappa^{p+2} V_p\right) \tag{73}$$

Next, set $\xi = \min\{\lceil \frac{1}{4V_p\kappa^{p+2}} \rceil, T - 1 - \Delta\}$. By the assumption on $\zeta_T$, for all $T$ large enough we must have that $\xi = \lceil \frac{1}{4V_p^2 \kappa^{2(p+1)}} \rceil$. □

*Proof of Lemma 2:* **Step 1.** Let $f_1, f_2 : \mathbb{R}^p \to \mathbb{R}^+$ be two densities such that

$$f_1(x) = \begin{cases} \lambda - \kappa + \|x - x_1\|_2, & x \in B(x_1, \kappa), \\ \lambda, & x \in B(x_2, \kappa), \\ g(x), & \text{otherwise,} \end{cases}$$

$$f_2(x) = \begin{cases} \lambda - \kappa + \|x - x_2\|_2, & x \in B(x_2, \kappa), \\ \lambda, & x \in B(x_1, \kappa), \\ g(x), & \text{otherwise.} \end{cases}$$

where $g$ is a function such that $f_1$ and $f_2$ are density functions, $\lambda$ is a constant, and $\kappa$ is a model parameter that can change with $T$. Note that for small enough $\kappa$ and $\lambda$,

$$\int_{B(x_1, \kappa)} f_1(x)\, dx \le 1.$$

Set $\|x_1 - x_2\| \ge 2\kappa$ to be any two fixed points. The excess probability mass can be place at $(B(x_1, \kappa) \cup B(x_2, \kappa))^c$. Since $f_1 = f_2$ in this region, it does not affect $KL(f_1, f_2)$ no matter how the functions are defined in this region.

Observe that, by integrating in polar coordinate and using symmetry

$$\begin{aligned} \mathrm{KL}(f_1, f_2) &= 2pV_p \int_0^\kappa \left\{\lambda \log\left(\frac{\lambda}{\lambda - \kappa + r}\right) r^{p-1} \right. \\ &\quad \left. + (\lambda - \kappa + r) \log\left(\frac{\lambda - \kappa + r}{\lambda}\right) r^{p-1}\right\} dr \\ &= 2pV_p \int_0^\kappa (\kappa - r) \log\left(\frac{\lambda}{\lambda - \kappa + r}\right) r^{p-1}\, dr \\ &\le 2pV_p \int_0^\kappa (\kappa - r) \frac{\kappa - r}{\lambda - \kappa + r} r^{p-1}\, dr \\ &\le 2pV_p \int_0^\kappa (\kappa - r) \frac{\kappa - r}{\lambda - \kappa + r} r^{p-1}\, dr \\ &\le 2pV_p \kappa^2 \lambda^{-1} \int_0^\kappa r^{p-1}\, dr \le C_p \kappa^{p+2} \end{aligned}$$

**Step 2.** Define $\mathcal{P}_T^1$ to be the joint density of $(X(1), \dots, X(T))$ such that $X(1), \dots, X(\Delta) \overset{\text{i.i.d.}}{\sim} f_1$ and $X(\Delta + 1), \dots, X(T) \overset{\text{i.i.d.}}{\sim} f_2$. Define $\mathcal{P}_T^2$ to be the joint density of $(X(1), \dots, X(T))$ such that $X(1), \dots, X(T - \Delta - 1) \overset{\text{i.i.d.}}{\sim} f_2$ and $X(T - \Delta), \dots, X(T) \overset{\text{i.i.d.}}{\sim} f_1$. We have that

$$\begin{aligned} \inf_{\hat\eta} \sup_{\mathcal{P}_n} \mathbb{E}\{|\widehat\eta - \eta(P)|\} &\ge (T - 2\Delta) d_{\mathrm{TV}}(\mathcal{P}_T^1, \mathcal{P}_T^2) \\ &\ge (T/4) \exp\{-\mathrm{KL}(\mathcal{P}_T^1, \mathcal{P}_T^2)\}. \end{aligned}$$

Note that

$$\mathrm{KL}(\mathcal{P}_T^1, \mathcal{P}_T^2) \le 2\Delta \mathrm{KL}\ (f_1, f_2) = C_p' \kappa^{p+2} \Delta.$$

TABLE VI
VALUES OF $\kappa$ FOR THE DIFFERENT SCENARIOS AND INSTANCES CONSIDERED IN THE PAPER

| Scenario | $\kappa$ | $p$ |
|---|---|---|
| 1 | $5.79 \times 10^{-10}$ | 20 |
| 1 | $4.78 \times 10^{-5}$ | 10 |
| 2 | 0.17 | 20 |
| 2 | 0.33 | 10 |
| 3 | $7.53 \times 10^{-10}$ | 20 |
| 3 | $6.37 \times 10^{-5}$ | 10 |
| 4 | $3.90 \times 10^{-10}$ | 20 |
| 4 | $1.23 \times 10^{-5}$ | 10 |
| 5 | $1.09 \times 10^{-42}$ | 20 |
| 5 | $2.58 \times 10^{-20}$ | 10 |

Since $\Delta\kappa^{p+2} \leq c < \log(2)$, we have

$$\exp(-\mathrm{KL}(\mathcal{P}_T^1, \mathcal{P}_T^2)) \geq \exp(-c) \geq 1/2$$

see e.g. Tsybakov [50]. In addition, noticing that $\Delta < T/2$, we reach the final claim. □

## APPENDIX D
## PARAMETER $\kappa$ FOR DIFFERENT SCENARIOS

We now display in Table VI the value of $\kappa$ for each scenario and instance of the experiments section. In the different scenarios, we evaluate the probability density functions before and after each change point in a set consisting of 2000 points, with 1000 samples drawn from each distribution. This allows us to compute $\kappa$.

## REFERENCES


[1] R. Aggarwal, C. Inclan, and R. Leal, "Volatility in emerging stock markets," *J. Financial Quant. Anal.*, vol. 34, no. 1, pp. 33–55, 1999. [Online]. Available: http://www.jstor.org/stable/2676245

[2] E. Andreou and E. Ghysels, "Detecting multiple breaks in financial market volatility dynamics," *J. Appl. Econometrics*, vol. 17, no. 5, pp. 579–600, 2002. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/jae.684/abstract

[3] S. Arlot, A. Celisse, and Z. Harchaoui, "A kernel multiple change-point algorithm via model selection," *J. Mach. Learn. Res.*, vol. 20, no. 162, pp. 1–56, 2019.

[4] A. Aue, S. Hörmann, L. Horváth, and M. Reimherr, "Break detection in the covariance structure of multivariate time series models," *Ann. Statist.*, vol. 37, no. 6B, pp. 4046–4087, Dec. 2009.

[5] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Stat. Soc. B, Methodol.*, vol. 57, no. 1, pp. 289–300, Jan. 1995.

[6] O. Bousquet, "A Bennett concentration inequality and its application to suprema of empirical processes," *Comptes Rendus Math.*, vol. 334, no. 6, pp. 495–500, 2002.

[7] D. Burns, J. Ekblom, and A. Shalal. (2019). *Timeline: Key Dates in the U.S.-China Trade War*. [Online]. Available: https://www.reuters.com/article/us-usa-trade-china-timeline/timeline-key-dates-in-the-us-china-trade-war-idUSKCN1UZ24U

[8] A. Celisse, G. Marot, M. Pierre-Jean, and G. J. Rigaill, "New efficient algorithms for multiple change-point detection with reproducing kernels," *Comput. Statist. Data Anal.*, vol. 128, pp. 200–220, Dec. 2018.

[9] F. Chamroukhi, S. Mohammed, D. Trabelsi, L. Oukhellou, and Y. Amirat, "Joint segmentation of multivariate time series with hidden process regression for human activity recognition," *Neurocomputing*, vol. 120, pp. 633–644, Nov. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231213004086

[10] H. Chen, "Change-point detection for multivariate and non-Euclidean data with local dependency," 2019, *arXiv:1903.01598*.

[11] H. Chen and N. Zhang, "Graph-based change-point detection," *Ann. Statist.*, vol. 43, no. 1, pp. 139–176, Feb. 2015.

[12] H. Chen, X. Chen, and Y. Su, "A weighted edge-count two-sample test for multivariate and object data," *J. Amer. Stat. Assoc.*, vol. 113, no. 523, pp. 1146–1155, Jul. 2018.

[13] H. Cho, "Change-point detection in panel data via double CUSUM statistic," *Electron. J. Statist.*, vol. 10, no. 2, pp. 2000–2038, Jan. 2016.

[14] H. Cho and P. Fryzlewicz, "Multiple-change-point detection for high dimensional time series via sparsified binary segmentation," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 77, no. 2, pp. 475–507, Mar. 2015.

[15] J. B. Elsner, X. Niu, and T. H. Jagger, "Detecting shifts in hurricane rates using a Markov chain Monte Carlo approach," *J. Climate*, vol. 17, no. 13, pp. 2652–2666, 2004.

[16] C. Erdman and J. W. Emerson, "A fast Bayesian change point analysis for the segmentation of microarray data," *Bioinformatics*, vol. 24, no. 19, pp. 2143–2148, Oct. 2008. [Online]. Available: http://bioinformatics.oxfordjournals.org/content/24/19/2143.short

[17] P. Fearnhead and G. Rigaill, "Changepoint detection in the presence of outliers," *J. Amer. Stat. Assoc.*, vol. 114, pp. 1–15, Jan. 2018.

[18] V. Fernandez, "The impact of major global events on volatility shifts: Evidence from the Asian crisis and 9/11," *Econ. Syst.*, vol. 30, no. 1, pp. 79–97, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0939362506000094

[19] K. Frick, A. Munk, and H. Sieling, "Multiscale change point inference," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 76, no. 3, pp. 495–580, Jun. 2014.

[20] P. Fryzlewicz, "Wild binary segmentation for multiple change-point detection," *Ann. Statist.*, vol. 42, no. 6, pp. 2243–2281, 2014.

[21] D. Garreau and S. Arlot, "Consistent change-point detection with kernels," *Electron. J. Statist.*, vol. 12, no. 2, pp. 4440–4486, Jan. 2018.

[22] E. Giné and A. Guillou, "Laws of the iterated logarithm for censored data," *Ann. Probab.*, vol. 27, no. 4, pp. 2042–2067, 1999.

[23] E. Giné and A. Guillou, "On consistency of kernel density estimators for randomly censored data: Rates holding uniformly over adaptive intervals," *Annales de l'IHP Probabilités et Statistiques*, vol. 37, no. 4, pp. 503–522, 2001.

[24] A. Gretton *et al.*, "Optimal kernel choice for large-scale two-sample tests," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1205–1213.

[25] S. W. Han, R. C. Mesquita, T. M. Busch, and M. E. Putt, "A method for choosing the smoothing parameter in a semi-parametric model for detecting change-points in blood flow," *J. Appl. Statist.*, vol. 41, no. 1, pp. 26–45, Jan. 2014.

[26] Z. Harchaoui and O. Cappé, "Retrospective mutiple change-point estimation with kernels," in *Proc. IEEE/SP 14th Workshop Stat. Signal Process.*, Aug. 2007, pp. 768–772.

[27] K. Haynes, P. Fearnhead, and I. A. Eckley, "A computationally efficient nonparametric approach for changepoint detection," *Statist. Comput.*, vol. 27, no. 5, pp. 1293–1305, Sep. 2017.

[28] R. Henderson and J. N. S. Matthews, "An investigation of changepoints in the annual number of cases of haemolytic uraemic syndrome," *Appl. Statist.*, vol. 42, no. 3, pp. 461–471, 1993. [Online]. Available: http://www.jstor.org/stable/2986325

[29] N. A. James and D. S. Matteson, "ecp: An R package for nonparametric multiple change point analysis of multivariate data," *J. Stat. Softw.*, vol. 62, no. 7, pp. 1–25, 2014. [Online]. Available: http://www.jstatsoft.org/v62/i07/

[30] L. J. Vanegas, M. Behr, and A. Munk, "Multiscale quantile segmentation," *J. Amer. Stat. Assoc.*, to be published.

[31] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *J. Amer. Statist. Assoc.*, vol. 107, no. 500, pp. 1590–1598, 2012.

[32] J. Kim, J. Shin, A. Rinaldo, and L. Wasserman, "Uniform convergence rate of the kernel density estimator adaptive to intrinsic volume dimension," 2018, *arXiv:1810.05935*.

[33] C. C. Kokonendji and T. S. Kiesse, "Discrete associated kernels method and extensions," *Stat. Methodol.*, vol. 8, no. 6, pp. 497–516, 2011.

[34] D. Kwon, M. Vannucci, J. J. Song, J. Jeong, and R. M. Pfeiffer, "A novel wavelet-based thresholding method for the pre-processing of mass spectrometry data that accounts for heterogeneous noise," *Proteomics*, vol. 8, no. 15, pp. 3019–3029, Aug. 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2855839/

[35] M. A. Lindquist, C. Waugh, and T. D. Wager, "Modeling state-related fMRI activity using change-point theory," *NeuroImage*, vol. 35, no. 3, pp. 1125–1141, Apr. 2007.

[36] D. S. Matteson and N. A. James, "A nonparametric approach for multiple change point analysis of multivariate data," *J. Amer. Stat. Assoc.*, vol. 109, no. 505, pp. 334–345, 2014.

[37] Microsoft Corp. (MSFT). (2019). *Yahoo! Finance*. [Online]. Available: https://finance.yahoo.com

[38] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[39] O. H. M. Padilla, A. Athey, A. Reinhart, and J. G. Scott, "Sequential nonparametric tests for a change in distribution: An application to detecting radiological anomalies," *J. Amer. Stat. Assoc.*, vol. 114, pp. 1–15, Apr. 2018.

[40] O. H. M. Padilla, Y. Yu, D. Wang, and A. Rinaldo, "Optimal nonparametric change point detection and localization," 2019, *arXiv:1905.10019*.

[41] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, nos. 1–2, pp. 100–115, 1954.

[42] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.

[43] F. Pein, H. Sieling, and A. Munk, "Heterogeneous change point inference," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 79, no. 4, pp. 1207–1227, Sep. 2017.

[44] W. Polonik, "Concentration and goodness-of-fit in higher dimensions: (Asymptotically) distribution-free methods," *Ann. Statist.*, vol. 27, no. 4, pp. 1210–1229, 1999.

[45] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. Accessed: 2019. [Online]. Available: https://www.R-project.org/

[46] B. Rajagopalan and U. Lall, "A kernel estimator for discrete distributions," *J. Nonparam. Statist.*, vol. 4, no. 4, pp. 409–426, Jan. 1995.

[47] J. A. Scott and M. Knott, "A cluster analysis method for grouping means in the analysis of variance," *Biometrics*, vol. 30, pp. 507–512, Sep. 1974.

[48] "Consistency and rates for clustering with dbscan," in *Proc. Bharath Sriperumbudur, Ingo Steinwart 15th Int. Conf. Artif. Intell. Statist.*, vol. 22. PMLR, 2012, pp. 1090–1098.

[49] A. B. Tsybakov, *Introduction à l'Estimation Non-Paramétrique*. Berlin, Germany: Springer, 2004.

[50] B. A. Tsybakov, *Introduction to Nonparametric Estimation*. New York, NY, USA: Springer, 2009.

[51] E. S. Venkatraman, "Consistency results in multiple change-point problems," Ph.D. dissertation, Dept. Statist., Stanford Univ., Stanford, CA, USA, 1992.

[52] N. Verzelen, M. Fromont, M. Lerasle, and P. Reynaud-Bouret, "Optimal change-point detection and localization," 2020, *arXiv:2010.11470*.

[53] A. Wald, "Sequential tests of statistical hypotheses," *Ann. Math. Statist.*, vol. 16, no. 2, pp. 117–186, 1945.

[54] D. Wang, Y. Yu, and A. Rinaldo, "Univariate mean change point detection: Penalization, CUSUM and optimality," 2018, *arXiv:1810.09498*.

[55] D. Wang, Y. Yu, and A. Rinaldo, "Optimal change point detection and localization in sparse dynamic networks," *Ann. Statist.*, vol. 49, no. 1, pp. 203–232, 2021.

[56] T. Wang and R. J. Samworth, "High dimensional change point estimation via sparse projection," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 80, no. 1, pp. 57–83, Jan. 2018.

[57] Y.-C. Yao, "Estimating the number of change-points via Schwarz' criterion," *Statist. Probab. Lett.*, vol. 6, no. 3, pp. 181–189, Feb. 1988.

[58] Y. C. Yao and R. A. Davis, "The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates," *Sankhyā, Indian J. Statist. A*, vol. 48, no. 3, pp. 339–353, Oct. 1986.

[59] Y.-C. Yao and S.-T. Au, "Least-squares estimation of a stop function," *Sankhyā, Indian J. Statist., Series A*, vol. 51, no. 3, pp. 370–381, Oct. 1989.

[60] B. Yu, "Festschrift for lucien le cam," in *Assouad, Fano, and Le Cam*, vol. 423. New York, NY, USA: Springer, 1997, p. 435.

[61] W. Zhang, N. A. James, and D. S. Matteson, "Pruning and nonparametric multiple change point detection," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 288–295.

[62] C. Zou, G. Yin, L. Feng, and Z. Wang, "Nonparametric maximum likelihood approach to multiple change-point problems," *Ann. Statist.*, vol. 42, no. 3, pp. 970–1002, 2014.


**Oscar Hernan Madrid Padilla** received the B.S. degree in mathematics from the Universidad de Guanajuato, Mexico, in April 2013, and the Ph.D. degree in statistics from The University of Texas at Austin in May 2017. From July 2017 to June 2019, he was a Neyman Visiting Assistant Professor with the Department of Statistics, University of California at Berkeley, Berkeley, CA, USA. He is currently an Assistant Professor with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA. His research interests include high-dimensional statistics, nonparametric statistics, change point detection, causal inference, quantile regression, Bayesian and empirical Bayes methodology, and graphical models.

**Yi Yu** received the Ph.D. degree in mathematical statistics from Fudan University. She is currently a Reader in statistics at the Department of Statistics, University of Warwick. Before joining the University of Warwick, she was a Lecturer in statistical sciences at the School of Mathematics, University of Bristol. Her main research interest includes various aspects of high-dimensional statistics.

**Daren Wang** received the Ph.D. degree in statistics from Carnegie Mellon University. He is currently an Assistant Professor with the Department of ACMS, University of Notre Dame. Before joining the University of Notre Dame, he was a Post-Doctoral Researcher with the Department of Statistics, University of Chicago. His main research interests include time series analysis and functional data analysis.

**Alessandro Rinaldo** received the Ph.D. degree in statistics from Carnegie Mellon in 2005 under the supervision of Stephen E. Fienberg. He is currently a Professor with the Department of Statistics and Data Science, Carnegie Mellon University. His research interests include the theoretical properties of statistical and machine learning models for high-dimensional data under various structural assumptions, such as sparsity or intrinsic low dimensionality. In his research work, he has investigated a broad range of issues related to the feasibility of statistical inference in a variety of problems, including high-dimensional regression, time series, privacy, categorical data analysis and graphical modeling, statistical network analysis, algebraic statistics, density-based clustering, topological and geometric data analysis, and change-point detection. Though predominantly of theoretical nature, his research work is motivated by highly practical problems in data science and has direct methodological implications, as it aims at deriving theoretical guarantees in support of existing methods and algorithm.