Alternating Dynamic Programming for Multiple Epidemic Change-Point Estimation

Zifeng Zhao
Department of Applied and Computational Mathematics and Statistics
University of Notre Dame

Chun Yip Yau
Department of Statistics
Chinese University of Hong Kong

July 17, 2019

Abstract
In this paper, we study the problem of multiple change-point detection for a univariate sequence under the epidemic setting, where the behavior of the sequence alternates between a common normal state and different epidemic states. This is a non-trivial generalization of the classical (single) epidemic change-point testing problem. To explicitly incorporate the alternating structure of the problem, we propose a novel model selection based approach for simultaneous inference on both number and locations of change-points and alternating states. Using the same spirit as profile likelihood, we develop a two-stage alternating pruned dynamic programming algorithm, which conducts efficient and exact optimization of the model selection criteria and has $O(n^2)$ as the worst case computational cost. As demonstrated by extensive numerical experiments, compared to classical multiple change-point detection procedures, the proposed method improves accuracy for both change-point estimation and model parameter estimation. We further show promising applications of the proposed algorithm to multiple testing with locally clustered signals, and demonstrate its superior performance over existing methods in large scale multiple testing and in DNA copy number variation detection through real data.

Keywords: Change-point, Pruned dynamic programming, Multiple testing, DNA copy number variation, Model selection, Epidemic detection
1 Introduction

Change-point detection is an active research area in statistics and has been studied extensively due to its broad applications in many fields such as finance, genetics and meteorology among others. There is vast literature in change-point detection, for example, see Yao (1987), Davis et al. (2006), Fearnhead (2006), Killick et al. (2012), Chan et al. (2014), Zou et al. (2014), Matteson and James (2014), Fearnhead and Rigail (2017) and references therein.

A less studied yet important type of change-point detection problem is the epidemic change-point detection, which is first proposed and studied in Levin and Kline (1985). Let $y_{1:n} = (y_1, \ldots, y_n)$ be a sequence of independently distributed univariate observations. Roughly speaking, under the (classical) epidemic change-point setting, there exist two change-points $0 < \tau_1 < \tau_2 < n$ such that $y_{1:\tau_1}$ and $y_{(\tau_2+1):n}$ follow the same distribution and $y_{(\tau_1+1):\tau_2}$ follows a different distribution. The two segments on the sides $y_{1:}\tau_1$ and $y_{(\tau_2+1):n}$ are referred to as the normal state and the central segment $y_{(\tau_1+1):\tau_2}$ is referred to as the epidemic state. We call this setting the single epidemic change-point setting.

In the literature, the epidemic change-point detection is typically formulated as a hypothesis testing problem, where different test statistics have been proposed to test the null hypothesis of no change-point against the above defined epidemic alternative with two change-points, see Yao (1993), Guan (2004), Arias-Castro et al. (2005) and Ning et al. (2012) for examples. Moreover, the existing literature focus on the single epidemic change-point setting where the data is assumed to start at the normal state and only one single epidemic state is allowed. The more realistic setting of detecting multiple epidemic change-points, however, has not been explored.

In this paper, we propose a model selection based framework on multiple epidemic change-points estimation. Specifically, we assume that under the epidemic alternative, there exist $m$ unknown change-points $0 < \tau_1 < \tau_2 < \ldots < \tau_m < n$ such that the distribution of $y_t$ alternates between a (common) normal state and (different) epidemic states. Note that we do not require the assumption that the data starts at the normal state. For a concrete example, let the number of change-points $m$ be even and the data $y_{1:n}$ start at the normal state. Denote $\tau_0 = 0$ and $\tau_{m+1} = n$. Under the multiple epidemic change-point setting, the $m/2 + 1$ odd-numbered segments $y_{(\tau_{2k+1}):\tau_{2k+1}}, k = 0, \ldots, m/2$ are at the (common) normal state and the $m/2$ even-numbered segments $y_{(\tau_{2k-1+1}):\tau_{2k}}, k = 1, \ldots, m/2$ are at (different) epidemic states. In other words, the data $y_{1:n}$
alternates between the normal state and epidemic states.

The multiple epidemic change-point setting incorporates the aforementioned single epidemic setting as a special case and is more realistic in that it allows the observations $y_{1:n}$ to move back and forth between the common normal state and epidemic states. One motivating example for multiple epidemic change-point setting is the DNA copy number variation (see e.g. Olshen et al. 2004; Niu and Zhang 2012), where the observations $y_{1:n}$ are the log-ratios of the copy number of genes between the test and reference sequence. For most genes, there is no variation (common normal state) and the mean log-ratio is a common constant (e.g. 0). When there is variation (epidemic state), depending on the duplication or deletion of certain genes, the mean log-ratio can be either larger or smaller than that of the normal state. Another important example is large scale multiple testing with locally clustered signal as considered in Cao and Wu (2015), where a sequence of $p$-values $p_{1:n}$ are observed with $p_i$ being the $p$-value for the $i$th test, and we need to perform $n$ hypothesis tests based on $p_{1:n}$. The signal is locally clustered in the sense that the sequence of $p$-values can be partitioned into alternating blocks of signal (epidemic state, where $p$-values do not follow $U(0, 1)$) and noise (common normal state, where $p$-values follow $U(0, 1)$). The two examples are later discussed in detail in Section 5.

Compared to the conventional multiple change-point detection problem, the unique aspect of the multiple epidemic change-point setting is that there is an underlying alternating structure on the behavior of the observation $y_{1:n}$. Same as all the other change-point detection problems, our primary interest is to recover the unknown number and locations of change-points. In addition, a further interest is to recover the underlying alternating states of the observation $y_{1:n}$. Specifically, the goal is to assign a normal or epidemic label to each estimated segment.

The unique alternating structure of states and the shared common behavior among all normal state segments impose both challenges and opportunities for change-point detection. Specifically, existing efficient multiple change-point detection algorithms such as PELT in Killick et al. (2012) and FPOP in Maidstone et al. (2017) cannot directly recover the underlying alternating states and thus require additional post-analysis on the estimation results. Moreover, intuitively, if an algorithm can explicitly incorporate and exploit the alternating structure and the knowledge that segments at the normal state share the same behavior, improved estimation accuracy should be expected due to the additional information on the structure of the estimation problem.
Motivated by the above observations, in this paper, we propose a novel alternating dynamic programming algorithm, named aPELT, to efficiently solve the multiple epidemic change-point problem. The proposed approach is based on the seminal work of PELT in Killick et al. (2012), but involves an explicit treatment of the alternating structure and common normal state behavior. Similar to PELT, it can be applied to find change-points under a range of statistical criteria such as penalized likelihood, quasi-likelihood and cumulative sum of squares, and enjoys the same computational efficiency of PELT, thus it can be applied to segment large data sets.

The advantages of aPELT are two-fold. First, by incorporating the shape-constraint explicitly, aPELT achieves simultaneous inference on both change-points and alternating states of the sequence, thus does not require any post processing of the estimation result. Moreover, as demonstrated by extensive numerical experiments and real data applications, the explicit treatment further helps to improve accuracy for both change-point estimation and parameter estimation. The proposed aPELT has useful applications in multiple testing problems with locally clustered signals and to DNA copy number variation detection (see Section 5 for more details), where superior performance over existing methods is observed.

A related yet different stream of literature is the constrained dynamic optimization according to Hocking et al. (2015) and Hocking et al. (2018), which is motivated by mean changes in ChIP-seq data. The authors propose efficient algorithms to solve a model selection problem under the constraint that a decrease in mean must be followed by an increase, and vice versa. Note that similar to Hocking et al. (2015) and Hocking et al. (2018), we also face a constrained optimization problem. However, under the multiple epidemic change-point setting, we do not impose directional relation on the normal state and epidemic state behavior. Moreover, our use of common normal state parameter poses further difficulty on the optimization, since the constraint is not only on two neighboring segments but on all normal state segments.

The rest of the paper is organized as follows. In Section 2, we formulate the multiple epidemic change-point detection problem and review the model selection approach for general change-point problems. For a tailored and efficient solution, an alternating dynamic programming algorithm (aPELT) is proposed in Section 3. The efficiency and accuracy of the proposed method are demonstrated via extensive numerical experiments in Section 4. Applications of aPELT to DNA copy number variation and multiple testing with locally clustered signals are presented in Sec-
tion 5, where results show the superior performance of aPELT over existing methods. The paper concludes with a discussion. Additional simulations and technical materials can be found in the supplementary material.

2 Background and Existing Solutions

2.1 Basic setting

Roughly speaking, change-point detection can be considered as the identification of points within a dataset where the statistical properties change. In this paper, we assume that \( y_1:n = (y_1, \ldots, y_n) \) is a sequence of independently distributed univariate observations. There are \( m \) change-points \( 0 < \tau_1 < \ldots < \tau_m < n \) that split the data into \( m + 1 \) segments. Define \( \tau_0 = 0 \) and \( \tau_{m+1} = n \), we have that the \( i \)th segment contains data \( y_{(\tau_{i-1}+1):\tau_i} \). We remark that extensions to settings with multivariate observations or dependence within segments is straightforward.

We assume that the distribution of \( y_t \) belongs to a parametric family \( f(y|\theta, \gamma) \), where \( \theta \in \mathbb{R}^{p_1} (p_1 \geq 1) \) denotes the parameter of interest and \( \gamma \in \mathbb{R}^{p_2} (p_2 \geq 0) \) denotes the nuisance parameter. For example, in mean change detection for independent Gaussian observations, \( \theta \) is the mean and \( \gamma \) is the variance of \( y_t \), and in change-point detection for independent Poisson counts, \( \theta \) is the intensity of \( y_t \) and there is no \( \gamma \). Denote the parameters for the \( i \)th segment \( y_{(\tau_{i-1}+1):\tau_i} \) as \( (\theta^i, \gamma^i) \), we have \( \theta^i \neq \theta^{i+1} \) for \( i = 1, \ldots, m \). Note that we do not put any restriction on the nuisance parameters \( \gamma^i \) except assuming that they are unknown.

Under the multiple epidemic change-point setting, we further assume that the parameter of interest \( \{\theta^i\}_{i=1}^{m+1} \) alternates between a common normal state \( \theta^o \) and different epidemic states. More formally, for any \( i = 1, \ldots, m \), if the \( i \)th segment \( y_{(\tau_{i-1}+1):\tau_i} \) follow \( f(y|\theta^o) \) (or \( f(y|\theta) \) for some \( \theta \neq \theta^o \)), then the \((i + 1)\)th segment \( y_{(\tau_{i+1}+1):\tau_{i+1}} \) will follow \( f(y|\theta) \) for some \( \theta \neq \theta^o \) (or \( f(y|\theta^o) \)). The only requirement for an epidemic state is that \( \theta \neq \theta^o \) without any directional constraint. For the multiple epidemic change-point setting, our inference interests are two-fold: 1. to recover the unknown number and locations of change-points, 2. to recover the alternating states of the observation \( y_{1:n} \).

As mentioned in Section 1, existing literature focus on single epidemic change-point detection with \( m = 2 \) and assume that the data starts at the normal state. Under such setting, typically a test
statistic in the form of \( \max_{1 \leq i < j < n} Z(i, j) \) is constructed for change-point detection and estimation via hypothesis testing. With \( m \) and initial state of the data being unknown, a direct generalization of this testing procedure to the multiple setting is not obvious. Moreover, the computational cost for obtaining such test statistic is \( O(n^2) \), which makes it less suitable for change-point detection in large data set. Thus, we instead tackle the multiple epidemic change-point detection via a model selection approach.

### 2.2 Optimal Partitioning and PELT

The multiple epidemic change-point detection is a special type of multiple change-point detection problem. In this section, we review two existing model selection based detection algorithms, which serves as the basis for our proposed alternating change-point detection procedure. For the moment, assume that we are doing classical multiple change-point detection, thus the only requirement is \( \theta^i \neq \theta^{i+1} \) for \( i = 1, \ldots, m \).

Given the observation \( y_{1:n} \), denote \( T(n) = \{ \tau : 0 = \tau_0 < \tau_1, \ldots, \tau_m < \tau_{m+1} = n \} \) as the candidate set of all possible vectors of change-points. The model selection approach estimates the true change-points \( \tau^o \) by minimizing a penalized loss function

\[
F(n) = \min_{\tau \in T(n)} \left\{ \sum_{i=1}^{m+1} \left[ \min_{\theta^i, \gamma^i} C(y_{(\tau_{i-1}+1):\tau_i}) + P \right] \right\} = \min_{\tau \in T(n)} \left\{ \sum_{i=1}^{m+1} \left[ C(y_{(\tau_{i-1}+1):\tau_i}) + P \right] \right\} ,
\]

where \( C(y_{(\tau_{i-1}+1):\tau_i}) = \sum_{t=\tau_{i-1}+1}^{\tau_i} g(y_t|\theta^i, \gamma^i) \) denotes the measure of model fit such as twice the negative log-likelihood, \( C(y_{(\tau_{i-1}+1):\tau_i}) = \min_{\theta^i, \gamma^i} C(y_{(\tau_{i-1}+1):\tau_i}) \), and \( P \) denotes the penalty for model complexity such as BIC or MDL.

The optimization of (1) is in general difficult. Using dynamic programming, Jackson et al. (2005) propose the Optimal Partitioning (OP) algorithm which obtains the exact solution of (1) with \( O(n^2) \) computational complexity. The essential idea is the recursive relationship where for any \( s \leq n \),

\[
F(s) = \min_{\tau \in T(s)} \left\{ \sum_{i=1}^{m+1} \left[ C(y_{(\tau_{i-1}+1):\tau_i}) + P \right] \right\} = \min_{t < s} \left\{ \min_{\tau \in T(t)} \sum_{i=1}^{m} \left[ C(y_{(\tau_{i-1}+1):\tau_i}) + P \right] + C(y_{(t+1):s}) + P \right\} = \min_{t < s} \left\{ F(t) + C(y_{(t+1):s}) + P \right\} .
\]
This provides a recursion which gives the minimal cost $F(s)$ of $y_{1:s}$ in terms of the minimal cost $F(t)$ of $y_{1:t}$ for $t < s$, and thus can be solved in turn for $s = 1, 2, \ldots, n$. Note that the essential condition for the recursive relationship (2) to hold is that the optimization of $C(y_{(t+1):s}) = \min_{\theta, \gamma} \sum_{i=t+1}^{s} g(y_{i}|\theta, \gamma)$ is independent across different segments, which is true under the classical multiple change-point setting.

Assuming the existence of a constant $K$ such that for all $t < s < n$, $C(y_{(t+1):s}) + C(y_{(s+1):n}) + K \leq C(y_{(t+1):n})$, [Killick et al. (2012)] propose the PELT algorithm, which further reduces the computational complexity and can solve $F(n)$ in linear time under mild conditions. The central observation is that for the calculation of $F(s)$, we do not need to consider all $\{t| t < s\}$ but only a pruned subset $\{t| t < s\} \setminus \{t| \text{there exists } t < t' < s \text{ such that } F(t) + C(y_{(t+1):t'}) + K \geq F(t')\}$, and thus achieve a lower computational cost.

3 Exact Multiple Epidemic Change-point Detection via Alternating Dynamic Programming

Compared to the classical setting, the epidemic change-point setting imposes an implicit shape-constraint on the model parameter $\{\theta_{i}\}_{i=1}^{m+1}$ where $\{\theta_{i}\}_{i=1}^{m+1}$ alternates between the normal state and epidemic states, and all normal state $\theta$s are the same.

One primary interest of inference is to recover the label of each segment, in other words, we would like the estimated parameter $\{\hat{\theta}_{i}\}_{i=1}^{m+1}$ to possess the alternating structure. However, neither OP nor PELT can directly recover the alternating structure since the shape-constraint is not explicitly considered in the penalized loss function $F(n)$ in [1], where $\tau$ only determines the number and locations of the change-points but does not restrict the state of each segment.

To impose the alternating structure of $\{\theta_{i}\}_{i=1}^{m+1}$, we propose to modify the penalized loss function in [1] by explicitly assigning states to segments. Note that due to the alternating structure, for any given $\tau$, once the state of the last segment is determined, all the other states are fixed. Given
alternating structure of programming based algorithm. However, due to the presence of the common parameter \(\sum\) of segments at the normal state, the recursive relationship in (2) no longer holds since the optimization of knowledge of the initial state of \(S\) segments and all segments in last segment to be at the epidemic state. Moreover, \(F\) state segment, there is no penalty for the parameter estimation of \(\theta\)P segment at epidemic state. Note that the modified penalized loss function which explicitly incorporates the alternating shape-constraint of \(m\) and \(\tau\), we define four index sets

\[
S_m^{oo} = \{i|i\text{th segment normal, } (m+1)\text{th segment normal}\}, \\
S_m^{ol} = \{i|i\text{th segment epidemic, } (m+1)\text{th segment normal}\}, \\
S_m^{lo} = \{i|i\text{th segment normal, } (m+1)\text{th segment epidemic}\}, \\
S_m^{11} = \{i|i\text{th segment epidemic, } (m+1)\text{th segment epidemic}\}.
\]

Depending on the state of the last segment, the four index sets assign states and group the segments of \(\tau\) by normal or epidemic states. For example, for \(m = 5\), \(S_5^{oo} = \{2, 4, 6\}, S_5^{ol} = \{1, 3, 5\}, S_5^{lo} = \{1, 3, 5\}\) and \(S_5^{11} = \{2, 4, 6\}\).

Based on the four index sets, we further define two penalized loss functions

\[
F^*_o(n) = \min_{\tau \in T(n)} \left\{ \min_{\theta^o} \sum_{i \in S_m^{oo}} \left[ \min_{\gamma^i} C(y(\tau_{i+1}); \tau_i | \theta^o, \gamma^i) + P^o \right] + \sum_{i \in S_m^{ol}} \left[ \min_{\theta^o, \gamma^i} C(y(\tau_{i+1}); \tau_i | \theta^i, \gamma^i) + P^l \right] \right\}, \quad (3)
\]

\[
F^*_1(n) = \min_{\tau \in T(n)} \left\{ \min_{\theta^o} \sum_{i \in S_m^{lo}} \left[ \min_{\gamma^i} C(y(\tau_{i+1}); \tau_i | \theta^o, \gamma^i) + P^o \right] + \sum_{i \in S_m^{11}} \left[ \min_{\theta^o, \gamma^i} C(y(\tau_{i+1}); \tau_i | \theta^i, \gamma^i) + P^l \right] \right\}, \quad (4)
\]

where \(P^o\) denotes the penalty for the segment at normal state and \(P^l\) denotes the penalty for the segment at epidemic state. Note that \(P^o\) and \(P^l\) may take different values, since for the normal state segment, there is no penalty for the parameter estimation of \(\theta^o\).

By design, \(F^*_o(n)\) forces the last segment of \(\tau\) to be at the normal state and \(F^*_1(n)\) forces the last segment to be at the epidemic state. Moreover, \(F^*_o(n)\) and \(F^*_1(n)\) explicitly incorporate the alternating structure of \(\{\theta^i\}_{i=1}^{m+1}\) since \(S_m\) enforces alternating normal and epidemic states among segments and all segments in \(S_m^{oo}\) or \(S_m^{11}\) share a common \(\theta^o\).

Thus, for the simultaneous inference of change-points and alternating states, we can then solve the modified penalized loss function

\[
F^*(n) = \min(F^*_o(n), F^*_1(n)), \quad (5)
\]

which explicitly incorporates the alternating shape-constraint of \(\{\theta^i\}_{i=1}^{m+1}\) and does not require the knowledge of the initial state of \(y_{1:n}\).

To solve (5) efficiently, a recursive relationship similar to (2) is required for an efficient dynamic programming based algorithm. However, due to the presence of the common parameter \(\theta^o\) across all segments at the normal state, the recursive relationship in (2) no longer holds since the optimization of \(\sum_{i \in S_m^{oo}} \left[ \min_{\gamma^i} C(y(\tau_{i+1}); \tau_i | \theta^o, \gamma^i) + P^o \right] \) in \(F^*_o(n)\) and \(\sum_{i \in S_m^{lo}} \left[ \min_{\gamma^i} C(y(\tau_{i+1}); \tau_i | \theta^i, \gamma^i) + P^l \right] \) in
In this section, for a given $\theta$, we describe the two-stage optimization procedure in detail. The key observation is that the optimization of the common $\theta$ causes the breakdown of the recursive relationship (2). To bypass this obstacle, we propose a two-stage optimization procedure for $F^*(n)$, which separates the optimization of $\theta$ and other model parameters ($\gamma, \tau$). This procedure shares the same spirit as profile likelihood. For any fixed $\theta$, we define

$$F^*_o(n; \theta) = \min_{\tau \in T(n)} \left\{ \sum_{i \in S_{m0}^o} [\min C(y_{(\tau_{i-1}+1):\tau_i} | \theta^o, \gamma^i) + P^o] + \sum_{i \in S_{m1}^o} [\min C(y_{(\tau_{i-1}+1):\tau_i} | \theta^i, \gamma^i) + P^i] \right\},$$

$$F^*_1(n; \theta) = \min_{\tau \in T(n)} \left\{ \sum_{i \in S_{m0}^1} [\min C(y_{(\tau_{i-1}+1):\tau_i} | \theta^o, \gamma^i) + P^o] + \sum_{i \in S_{m1}^1} [\min C(y_{(\tau_{i-1}+1):\tau_i} | \theta^i, \gamma^i) + P^i] \right\}.$$

Rearrange the order of optimization between $\theta$ and $\tau$ in (3) and (4), we have

$$F^*_o(n) = \min_{\theta^o} \min_{\tau \in T(n)} \left\{ \sum_{i \in S_{m0}^o} [\min C(y_{(\tau_{i-1}+1):\tau_i} | \theta^o, \gamma^i) + P^o] + \sum_{i \in S_{m1}^o} [\min C(y_{(\tau_{i-1}+1):\tau_i} | \theta^i, \gamma^i) + P^i] \right\} = \min_{\theta^o} F^*_o(n; \theta^o),$$

$$F^*_1(n) = \min_{\theta^o} \min_{\tau \in T(n)} \left\{ \sum_{i \in S_{m0}^1} [\min C(y_{(\tau_{i-1}+1):\tau_i} | \theta^o, \gamma^i) + P^o] + \sum_{i \in S_{m1}^1} [\min C(y_{(\tau_{i-1}+1):\tau_i} | \theta^i, \gamma^i) + P^i] \right\} = \min_{\theta^o} F^*_1(n; \theta^o).$$

Denote $F^*(n; \theta^o) = \min(F^*_o(n; \theta^o), F^*_1(n; \theta^o))$, we have

$$F^*(n) = \min(F^*_o(n), F^*_1(n)) = \min_{\theta^o} F^*(n; \theta^o).$$

Thus if we can solve $F^*(n; \theta^o)$ efficiently for each given $\theta^o$ and $F^*(n; \theta^o)$ is a smooth function of $\theta^o$, then we can efficiently solve $F^*(n)$ in a profile-likelihood fashion. In the following two subsections, we describe the two-stage optimization procedure in detail.

### 3.2 Alternating PELT (aPELT) under known normal state parameter $\theta^o$

In this section, for a given $\theta^o$, we propose an efficient alternating dynamic programming algorithm for solving $F^*(n; \theta^o) = \min(F^*_o(n; \theta^o), F^*_1(n; \theta^o))$, based on an alternating recursion between
to obtain $F^*_0(n; \theta^o)$ and $F^*_1(n; \theta^o)$. Later, we further extend the algorithm to the case where $\theta^o$ is unknown. Denote $C^o(y_{(t+1):s}) = \min_{\gamma} C(y_{(t+1):s}|\theta^o, \gamma)$ and $C^1(y_{(t+1):s}) = \min_{\theta, \gamma} C(y_{(t+1):s}|\theta, \gamma)$. Under the epidemic change-point setting, a normal state is always followed by an epidemic state and vice versa. Thus, there is an implicit alternating recursion between $F^*_0(s; \theta^o)$ and $F^*_1(s; \theta^o)$ where

$$F^*_0(s; \theta^o) = \min_{t<s} \{ F^*_1(t; \theta^o) + C^o(y_{(t+1):s}) + P^o \}, \quad F^*_1(s; \theta^o) = \min_{t<s} \{ F^*_0(s; \theta^o) + C^1(y_{(t+1):s}) + P^1 \}. \quad (6)$$

Equations (6) provides a recursive relationship between the minimal cost $F^*_0(s; \theta^o)$ for $y_{1:s}$ and the minimal cost $F^*_1(t; \theta^o)$ for $y_{1:t}$ with $t < s$, and similarly between $F^*_1(s; \theta^o)$ and $F^*_0(t; \theta^o)$. Thus, to obtain $F^*(n; \theta^o) = \min( F^*_0(n; \theta^o), F^*_1(n; \theta^o) )$, we can solve $F^*_0(s; \theta^o)$ and $F^*_1(s; \theta^o)$ simultaneously by recursion in turn for $s = 1, 2, \ldots, n$. The computational cost of the algorithm is $O(n^2)$.

To further reduce the computational cost for large data set, we propose an alternating PELT (aPELT) by extending the idea of PELT in Killick et al. (2012). The central idea is that when calculating $F^*_0(s; \theta^o)$ and $F^*_1(s; \theta^o)$ via recursion (6), we do not need to consider all $t < s$. Instead, we only need to consider a subset of $t < s$ by adding a pruning step. The theoretical guarantee for the pruning step is stated in Theorem 1.

**Theorem 1.** Given $\theta^o$, assume that there exists a constant $K_o$ such that for all $t < s < n$,

$$C^o(y_{(t+1):s}) + C^o(y_{(s+1):n}) + K_o \leq C^o(y_{(t+1):n}).$$

Then if

$$F^*_1(t; \theta^o) + C^o(y_{(t+1):s}) + K_o > F^*_1(s; \theta^o)$$

holds, at a future time $n > s$, $t$ can never be the optimal last change-point for $F^*_0(n; \theta^o)$ prior to $n$.

**Similarly assume that there exists a constant $K_1$ such that for all $t < s < n$,**

$$C^1(y_{(t+1):s}) + C^1(y_{(s+1):n}) + K_1 \leq C^1(y_{(t+1):n}).$$

Then if

$$F^*_0(t; \theta^o) + C^1(y_{(t+1):s}) + K_1 > F^*_0(s; \theta^o),$$

holds, at a future time $n > s$, $t$ can never be the optimal last change-point for $F^*_1(n; \theta^o)$ prior to $n$.

Based on Theorem 1, the pseudo-code of aPELT with known normal state parameter $\theta^o$ is given in Algorithm 1 and we name it aPELT($\theta^o$). An interesting phenomenon in Theorem 1 is that the pruning for $F^*_0(n; \theta^o)$ requires the values of $F^*_1(s; \theta^o)$ and vice versa, which again requires
simultaneous calculation of $F^*_0(n; \theta^o)$ and $F^*_1(n; \theta^o)$.

Remark 1: If $C(y_{(t+1):s}; \theta, \gamma)$ is the log-likelihood function of $y_{(t+1):s}$, it can be easily shown that the constant $K_o$ and $K_1$ exist and can be set to 0 for any $\theta^o$.

Algorithm 1 aPELT($\theta^o$): aPELT algorithm with known normal state parameter $\theta^o$

1: Initialize: Set $F^*_o(0; \theta^o) = F^*_1(0; \theta^o) = 0$, $c^o(0) = c^o(0) = NULL$, $R^o_0 = R^1_1 = \{0\}$.
2: Iterate: for $s = 1, \ldots, n$
3: (i). Calculate $F^*_o(s; \theta^o) = \min_{t \in R^o_0} \{ F^*_1(t; \theta^o) + C^o(y_{(t+1):s}) + P^o \}$,
4: $F^*_1(s; \theta^o) = \min_{t \in R^1_1} \{ F^*_o(t; \theta^o) + C^1(y_{(t+1):s}) + P^1 \}$.
5: (ii). Let $t^*_o = \arg \min_{t \in R^o_0} \{ F^*_1(t; \theta^o) + C^o(y_{(t+1):s}) + P^o \}$,
6: $t^*_1 = \arg \min_{t \in R^1_1} \{ F^*_o(t; \theta^o) + C^1(y_{(t+1):s}) + P^1 \}$.
7: (iii). Set $c^o(s) = \{ c^o(t^*_o), t^*_o \}$ and $c^1(s) = \{ c^o(t^*_1), t^*_1 \}$.
8: (iv). Set $R^o_{s+1} = \{ t \in R^o_0 \cup \{ s \} : F^*_1(t; \theta^o) + C^o(y_{(t+1):s}) + K_o \leq F^*_1(s; \theta^o) \}$,
9: $R^1_{s+1} = \{ t \in R^1_1 \cup \{ s \} : F^*_o(t; \theta^o) + C^1(y_{(t+1):s}) + K_1 \leq F^*_o(s; \theta^o) \}$.
10: Output: If $F^*_o(n; \theta^o) \leq F^*_1(n; \theta^o)$, return $c^o(n)$ and alternating states with last state normal.
11: Otherwise, return $c^1(n)$ and alternating states with last state epidemic.

3.3 Alternating PELT under unknown normal state parameter $\theta^o$

For many applications, the normal state parameter $\theta^o$ is naturally known and thus aPELT($\theta^o$) proposed in Section 3.2 is sufficient. For example, in DNA copy number variation, the mean log-ratio between the test and reference sequence is typically 0 when there is no variation; in multiple testing with locally clustered signals, the normal state is uniform distribution $U(0, 1)$. See Section 5 for more details of the above two examples. Nevertheless, for the sake of generality, it is of interest to cover the case of unknown $\theta^o$. In this section, we discuss two extensions of aPELT, namely aPELT_profile and aPELT_plugin, to handle such situation.

3.3.1 Profile aPELT

The proposed aPELT($\theta^o$) in Section 3.2 can find the exact minimum of $F^*(n; \theta^o)$ for a given normal state parameter $\theta^o$, thus if $F^*(n; \theta^o)$ is a smooth function of $\theta^o$, we can solve $F^*(n) = \min_{\theta^o} F^*(n; \theta^o)$ by a standard optimization algorithm such as gradient descent, and as a byproduct, $\theta^o$ can be estimated by $\hat{\theta}^o = \arg \min_{\theta^o} F^*(n; \theta^o)$. This two-stage procedure shares the same spirit as profile likelihood, thus we name it aPELT_profile.

To justify aPELT_profile, we investigate the behavior of $F^*(n; \theta^o)$ as a function of $\theta^o$ and show
that in general a gradient-based algorithm can be used to solve $\min_{\theta^o} F^*(n; \theta^o)$. We have

$$F^*(n; \theta^o) = \min(F^*_o(n; \theta^o), F^*_n(n; \theta^o))$$

$$= \min \left\{ \min_{\tau \in \mathcal{T}(n)} \left[ \sum_{i \in S_{\theta^o,n}^m} \left[ \min_{\gamma^i} C(y(\tau_{i-1}+1):\tau_i|\theta^o, \gamma^i) + P^o \right] + \sum_{i \in S_{\theta^o,n}^1} \left[ \min_{\gamma^i} C(y(\tau_{i-1}+1):\tau_i|\theta^o, \gamma^i) + P^1 \right] \right] \right\},$$

For a given $\tau$, $\sum_{i \in S_{\theta^o,n}^m} \left[ \min_{\theta^i, \gamma^i} C(y(\tau_{i-1}+1):\tau_i|\theta^i, \gamma^i) + P^1 \right]$ and $\sum_{i \in S_{\theta^o,n}^1} \left[ \min_{\gamma^i} C(y(\tau_{i-1}+1):\tau_i|\theta^i, \gamma^i) + P^1 \right]$ are constants, thus we denote $C_1(\tau) = \sum_{i \in S_{\theta^o,n}^m} \left[ \min_{\theta^i, \gamma^i} C(y(\tau_{i-1}+1):\tau_i|\theta^i, \gamma^i) + P^1 \right]$ and $C_2(\tau) = \sum_{i \in S_{\theta^o,n}^1} \left[ \min_{\gamma^i} C(y(\tau_{i-1}+1):\tau_i|\theta^i, \gamma^i) + P^1 \right]$. Therefore, we have

$$F^*(n; \theta^o) = \min \left\{ \min_{\tau \in \mathcal{T}(n)} \left[ g_1(\theta^o; \tau) + C_1(\tau) \right], \min_{\tau \in \mathcal{T}(n)} \left[ g_2(\theta^o; \tau) + C_2(\tau) \right] \right\}.$$ 

In other words, $F^*(n; \theta^o)$ is the minimum of $2|\mathcal{T}(n)|$ functions of $\theta^o$, where $|\cdot|$ denotes the cardinality of a set. Intuitively, if for each $\tau$, $g_1(\theta^o; \tau)$ and $g_2(\theta^o; \tau)$ are smooth functions of $\theta^o$, $F^*(n; \theta^o)$ should also be a (piecewise) smooth function of $\theta^o$.

In the following, denote $\Theta \in \mathbb{R}^{p_1}$ as the parameter space of the normal state parameter $\theta^o$ and denote $\hat{\Theta}$ as the interior of $\Theta$. Before stating Theorem 2, we first state two assumptions on the behavior of the $2|\mathcal{T}(n)|$ functions in $\{g_1(\theta^o; \tau), g_2(\theta^o; \tau)|\tau \in \mathcal{T}(n)\}$.

**Assumption 1 (Smoothness).** Any function in $\{g_1(\theta^o; \tau), g_2(\theta^o; \tau)|\tau \in \mathcal{T}(n)\}$ is a differentiable function of $\theta^o$ and has a unique global minimizer in $\hat{\Theta}$. WLOG, further assume that the global minimizers and the minimum values of different functions are different.

**Assumption 2 (Finite Partition).** There exists a finite partition $\{\Theta_1, \ldots, \Theta_{N(n)}\}$ of $\Theta$ where each $\Theta_i$ is a connected set in $\mathbb{R}^p$ and there is no intersection among functions in $\{g_1(\theta^o; \tau), g_2(\theta^o; \tau)|\tau \in \mathcal{T}(n)\}$ in $\hat{\Theta}_i$, for $i = 1, \ldots, N(n)$.

Both Assumptions 1 and 2 are mild and are expected to hold for common loss functions $C(y(\tau_{i+1}):\gamma^i)$ such as log-likelihood functions. For example, for $\Theta \subset \mathbb{R}$, a sufficient condition for Assumption 2 to hold is that all functions in $\{g_1(\theta^o; \tau), g_2(\theta^o; \tau)|\tau \in \mathcal{T}(n)\}$ intersect finite times.
with each other. Assumption 2 is used to evoke intermediate value theorem in the proof and show that $F^*(n; \theta^o)$ is “piecewise” differentiable on $\Theta$. In Section 3.5, we verify Assumptions 1 and 2 for some classical change-point settings.

**Theorem 2.** Under Assumption 1, $F^*(n; \theta^o)$ is a continuous function of $\theta^o$ and has a unique global minimizer $\theta^o^*$ in $\hat{\Theta}$, and there exists an open neighborhood $N(\theta^o^*)$ of $\theta^o^*$ such that $F^*(n; \theta^o)$ is differentiable on $N(\theta^o^*)$. If in addition Assumption 2 holds, then we further have that $F^*(n; \theta^o)$ is differentiable in every $\hat{\Theta}$ and $\theta^o^* \in \hat{\Theta}$ for some $i = 1, \ldots, N(n)$.

By Theorem 2 under Assumption 1 with a proper starting point, combined with a standard optimization algorithm such as gradient descent, aPELT_profile can successfully locate the global minimizer of $F^*(n; \theta^o)$, and thus simultaneously estimate the unknown normal state parameter $\theta^o$ and the unknown change-points $\tau^o$. With the additional Assumption 2, we know that $F^*(n; \theta^o)$ is “piecewise” differentiable on $\Theta$ and is differentiable in every $\hat{\Theta}$, which further justifies the use of a gradient-based optimization method such as gradient descent.

Based on Theorem 2 and Algorithm 1, the pseudo-code of aPELT_profile with unknown normal state parameter is given in Algorithm 2.

**Algorithm 2 aPELT_profile: aPELT algorithm with unknown normal state parameter**

1: Given an initial point $\theta^o_0$
2: Run gradient descent on $F^*(n; \theta^o)$ starting from $\theta^o_0$, where the function value of $F^*(n; \theta^o)$ is evaluated via aPELT($\theta^o$).
3: **Output:** Return $\hat{\theta}^o = \arg \min_{\theta^o} F^*(n; \theta^o)$ given by gradient descent and output of aPELT($\hat{\theta}^o$).

**Remark 2:** In practice, as for the starting point of $\theta^o$, we can either select a set of different starting points across $\Theta$ and run aPELT_profile in parallel or we can start aPELT_profile from an estimated $\hat{\theta}^o$ for $\theta^o$ as is discussed in Section 3.3.2.

### 3.3.2 Plug-in aPELT

When the normal state parameter $\theta^o$ is unknown, another natural way to proceed is to first obtain an estimated $\hat{\theta}^o$ and then run aPELT($\hat{\theta}^o$) as if it is the true parameter. We call this method aPELT_plugin. As expected, the performance of aPELT_plugin is closely related to the accuracy of $\hat{\theta}^o$. With an accurate estimator for $\theta^o$, aPELT_plugin should have decent performance.
The estimator $\hat{\theta}^o$ can be chosen based on specific cases. For example, if $y_{1:n}$ is a sequence of univariate Gaussian random variables with epidemic mean change and a normal state mean $\mu^o$, then one possible estimator $\tilde{\mu}^o$ is the median of estimated mean from a sequence of short screening-windows. This should work well if the epidemic state lasts shorter than the normal state.

### 3.4 Computational cost of aPELT

For a fixed $\theta^o$, aPELT$(\theta^o)$ essentially runs two PELT algorithms simultaneously in an alternating fashion. Thus, the computational properties for aPELT$(\theta^o)$ are largely the same as PELT.

For any given $\theta^o$, the upper bound for solving $F^*(n; \theta^o)$ via aPELT$(\theta^o)$ is $O(n^2)$. Similar to PELT, aPELT$(\theta^o)$ has a linear computational cost $O(n)$ if the number of change-points increases linearly with $n$. As is demonstrated by numerical experiments in Section 4, the computational cost of aPELT$(\theta^o)$ and aPELT_plugin (which is aPELT$(\hat{\theta}^o)$) are almost the same as PELT. The computational cost for aPELT_profile depends further on the smoothness of $F^*(n; \theta^o)$ and the second-stage optimization algorithm. Based on the extensive numerical experiments, we empirically find that with gradient descent, the computational time for aPELT_profile is around 10 to 20 times that of PELT or aPELT_plugin, regardless of $n$. In other words, the computational cost of aPELT_profile is at the same scale with aPELT and PELT.

### 3.5 Example of aPELT for epidemic mean change and variance change

In this section, we give a detailed example of aPELT for epidemic change-point detection in mean or variance of a sequence of independent Gaussian random variables. Consider a sequence of random variables $y_{1:n}$ such that

$$y_t = \mu_t + \epsilon_t,$$

where $\{\epsilon_t\}$ is a sequence of independent Gaussian noise following $N(0, \sigma_t^2)$. We consider two cases:

1) There are epidemic changes in mean $\mu_{1:n}$ with unknown variance, and
2) There are epidemic changes in variance $\sigma_{1:n}^2$ with unknown mean. For epidemic mean change, $\mu$ is the parameter of interest and $\sigma^2$ is the nuisance parameter, while for epidemic variance change, $\sigma^2$ is the parameter of interest and $\mu$ is the nuisance parameter.

For both cases, we set $C(y_{(t+1):s} \mid \mu, \sigma)$ to be twice the negative log-likelihood of Gaussian random
variables, where
\[ C(y_{(t+1);s}|\mu, \sigma) = (s - t) \log(2\pi\sigma^2) + \sum_{i=t+1}^{s} \frac{(y_i - \mu)^2}{\sigma^2}. \]  

We set \( P_0 = 2\log n \) and \( P_1 = 3\log n \) since for an extra epidemic state, we need to record the location of the change-point and two parameters \((\mu, \sigma)\), while for an extra normal state, we do not need to record the common normal state parameter. Proposition 1 characterizes the behavior of \( F^*(n; \mu^o) \) and \( F^*(n; \sigma^o) \) as functions of \( \mu^o \) and \( \sigma^o \).

**Proposition 1.** For epidemic change-point detection in mean \( \mu \), Assumptions 1 and 2 hold for \( F^*(n; \mu^o) \). Thus, \( F^*(n; \mu^o) \) is a continuous and piecewise differentiable function of \( \mu^o \) and \( F^*(n; \mu^o) \) has a global minimizer \( \mu^{o*} \). More specifically, there exists a finite number \( N = N(n) \) and a sequence \( -\infty = \mu_0 < \mu_1 < \cdots < \mu_{N+1} = \infty \) such that \( F^*(n; \mu^o) \) is differentiable on each \((\mu_i, \mu_{i+1})\) and \( \mu^{o*} \) is an interior point of one of the intervals. Moreover, the same result holds for \( F^*(n; \sigma^o) \) of epidemic change-point detection in variance.

In Figure 1 and Figure 2, we give sample plots for \( F^*(n; \mu^o) \) and \( F^*(n; \sigma^o) \), which confirms the continuous and piecewise differentiable claims in Proposition 1.

**Remark 3:** Following the same argument, it is easy to prove that similar results as the one in Proposition 1 hold for other distributions from the exponential family, such as Poisson, Binomial and Exponential distributions.

4 Simulation Study

In this section, we investigate the performance of PELT, aPELT(\( \theta^o \)), aPELT_plugin and aPELT_profile under various change-point settings for the univariate sequence \( y_{1:n} \) in (7) in Section 3.5. We present the numerical result for epidemic mean change. The result for epidemic variance change is similar and can be found in Section §3 of the supplementary material.

4.1 Basic simulation setting

For PELT, we set \( C(y_{(s+1);s}|\mu, \sigma) \) as the log-likelihood function (8) with unknown \((\mu, \sigma)\). For aPELT(\(\theta^o\)), we assume that the normal state parameter \( \mu^o \) or \( \sigma^o \) is known. For aPELT_plugin, we estimate the normal state parameter \( \mu^o \) or \( \sigma^o \) by the median of sample mean or standard deviation.
from a sequence of screening-windows of size 10; for example, \( \hat{\mu}^o = \text{median}\{\text{mean}(y_{t+1:t+10}), t = 0, \ldots, n - 10\} \). Under the typical scenario where the sequence \( y_{1:n} \) is mostly at the normal state, one would expect \( \hat{\mu}^o \) to be an accurate estimator for \( \mu^o \). For aPELT_profile, we set the starting point of optimization for \( \mu^o \) or \( \sigma^o \) at the estimated \( \hat{\mu}^o \) or \( \hat{\sigma}^o \) as in aPELT_plugin. For the aPELT methods, we employ BIC-penalty with \( P^o = 2 \log(n) \) for the normal state and \( P^1 = 3 \log(n) \) for the epidemic state, and for PELT, the penalty is always \( P = 3 \log(n) \).

Denote \( m \) as the number of true change-points in the sequence \( y_{1:n} \), WLOG, we set \( m \) to be an even number. For the location of the true change-points \( \tau^o \) and the alternating states, we first assign the last segment \( ([n \frac{n+1}{m+2} + 1] : n \) of \( y_{1:n} \) to be the normal state. We then divide the rest of the sequence into \( m/2 \) equal-length segments and for each of the \( m/2 \) segments, we randomly divide them into a normal state and an epidemic state, where the ratio between the length of the epidemic state and the normal state is randomly sampled from \( U(0.2, 0.5) \), i.e. the normal state is expected to last longer than the epidemic state. Under this setting, the alternating state of \( y_{1:n} \) starts and ends with the normal state.

To evaluate the performance of change-point estimation, similar to Killick et al. (2012), we report the true positive rate (TPR) and false positive rate (FPR) of each algorithm. For each simulated sequence \( y_{1:n} \), we define

\[
\text{TPR} = \frac{\text{number of correctly detected true change-point (CP)}}{m},
\]

\[
\text{FPR} = \frac{\text{total number of detected CP} - \text{number of correctly detected true CP}}{\text{total number of detected CP}},
\]

where for any true change-point \( \tau^o_i \in \tau^o \), we consider it to be correctly detected if there is an estimated change-point \( \hat{\tau}^o_i \) within a distance of 20 points from \( \tau^o_i \).

To evaluate the performance of parameter estimation, we report the MSE of the estimated parameter as

\[
\text{MSE}(\theta) = \sqrt{\frac{\sum_{t=1}^{n} (\hat{\theta}_t - \theta_t)^2}{n}},
\]

where \( \theta_{1:n} \) denotes the true parameter of interest and \( \theta = \mu \) for epidemic mean change.

### 4.2 Epidemic mean change

We present the result for epidemic mean change. We set \( n = 1000, 2000, 10000 \) and \( m = \sqrt{n/10} = 10, 14, 32 \), i.e. the number of change-points grows sublinearly with \( n \). Numerical experiments for
the case where the number of change-points \( m \) grows linearly with \( n \) are also conducted. The result is similar and can be found in Section §2 and §4 of the supplementary material.

For each sequence \( y_{1:n} \), we simulate its change-point locations and alternating states as described in Section 4.1. As for \( \mu^i \) of the \( i \)th segment, we set

\[
\mu^i = \begin{cases} 
\mu^o = 0 & \text{if } \text{ith segment is at the normal state}, \\
(1 - 2 \cdot \text{Ber}_i(0.5)) \cdot U_i(a, a + 0.2) & \text{if } \text{ith segment is at the epidemic state},
\end{cases}
\]

where \( \text{Ber}_i(0.5) \) are i.i.d. Bernoulli distribution with mean 0.5 and \( U_i(a, a + 0.2) \) are i.i.d. uniform distribution with mean \( a + 0.1 \).

We fix variance \( \sigma = 1 \) and change the signal level by varying \( a \) through \( a = 0.7 \) (low), 0.9 (medium), 1.1 (high). For each combination of \( (n, a) \), we repeat the simulation 100 times using PELT and the three versions of aPELT. In addition, for each simulation, we calculate the “true” minimum of \( F^*(n, \mu^o) \) by conducting grid search on \( \mu = \{-3, -2.95, \cdots, 2.95, 3\} \). Note that the locations of true change-points \( \tau^o \) and alternating mean structures \( \{\mu^i\}_{i=1}^{m+1} \) are simulated separately for each experiment instead of being fixed for all.

We report the average TPR, FPR, number of estimated CP, MSE of estimated parameters \( \hat{\mu}_{1:n} \) and computational time for different algorithms in Table 1.

Table 1: Performance of aPELT(\( \theta^o \)), aPELT_plugin, aPELT_profile and PELT under epidemic mean change for \( n = 1000, 2000, 10000 \) with \( m = \sqrt{n/10} \).

| \( n = 1000 \), Low signal | \( \text{TPR} \) | \( \text{FPR} \) | \( \text{Number of CP} \) | \( \text{MSE} \) | \( \text{Time} \) |
|-----------------------------|----------------|----------------|----------------|--------------------|----------------|
| aPELT(\( \theta^o \))       | 0.514          | 0.054          | 5.390          | 0.291              | 16.1           |
| aPELT_plugin                | 0.461          | 0.063          | 4.900          | 0.307              | 17.1           |
| aPELT_profile               | 0.431          | 0.046          | 4.530          | 0.302              | 279.4          |
| PELT                        | 0.284          | 0.081          | 3.040          | 0.324              | 13.7           |

| \( n = 2000 \), Low signal | \( \text{TPR} \) | \( \text{FPR} \) | \( \text{Number of CP} \) | \( \text{MSE} \) | \( \text{Time} \) |
|-----------------------------|----------------|----------------|----------------|--------------------|----------------|
| aPELT(\( \theta^o \))       | 0.616          | 0.054          | 9.080          | 0.242              | 26.7           |
| aPELT_plugin                | 0.597          | 0.047          | 8.780          | 0.252              | 27.6           |
| aPELT_profile               | 0.559          | 0.070          | 8.300          | 0.251              | 484.1          |
| PELT                        | 0.384          | 0.078          | 5.800          | 0.283              | 21.8           |

| \( n = 10000 \), Low signal | \( \text{TPR} \) | \( \text{FPR} \) | \( \text{Number of CP} \) | \( \text{MSE} \) | \( \text{Time} \) |
|-----------------------------|----------------|----------------|----------------|--------------------|----------------|
| aPELT(\( \theta^o \))       | 0.946          | 0.046          | 31.700         | 0.110              | 145.4          |
| aPELT_plugin                | 0.940          | 0.047          | 31.560         | 0.115              | 147.4          |
| Algorithm | TPR  | FPR  | Number of CP | MSE  | Time |
|-----------|------|------|--------------|------|------|
| aPELT(θ₀) | 0.847 | 0.015 | 8.590        | 0.238| 9.9  |
| aPELT_plugin | 0.814 | 0.022 | 8.340        | 0.260| 10.6 |
| aPELT_profile | 0.809 | 0.021 | 8.280        | 0.253| 169.7|
| PELT      | 0.620 | 0.026 | 6.350        | 0.309| 9.1  |

| Algorithm | TPR  | FPR  | Number of CP | MSE  | Time |
|-----------|------|------|--------------|------|------|
| aPELT(θ₀) | 0.920 | 0.022 | 13.160       | 0.182| 14.5 |
| aPELT_plugin | 0.904 | 0.024 | 12.970       | 0.192| 14.8 |
| aPELT_profile | 0.915 | 0.019 | 13.050       | 0.184| 245.5|
| PELT      | 0.771 | 0.019 | 11.020       | 0.240| 13.0 |

| Algorithm | TPR  | FPR  | Number of CP | MSE  | Time |
|-----------|------|------|--------------|------|------|
| aPELT(θ₀) | 0.986 | 0.008 | 13.920       | 0.159| 10.4 |
| aPELT_plugin | 0.989 | 0.008 | 13.950       | 0.164| 10.5 |
| aPELT_profile | 0.984 | 0.008 | 13.880       | 0.161| 172.9|
| PELT      | 0.961 | 0.007 | 13.550       | 0.181| 9.3  |

| Algorithm | TPR  | FPR  | Number of CP | MSE  | Time |
|-----------|------|------|--------------|------|------|
| aPELT(θ₀) | 0.996 | 0.004 | 32.000       | 0.103| 103.7|
| aPELT_plugin | 0.996 | 0.004 | 32.000       | 0.107| 104.0|
| aPELT_profile | 0.996 | 0.004 | 32.000       | 0.103| 1829.0|
| PELT      | 0.996 | 0.005 | 32.010       | 0.111| 92.8 |

From Table 1 with n grows, all algorithms perform better. (For comparison, in the supple-
mentary material, we give the simulation results with $n = m/100$, where we observe that the performance of all algorithms get worse with $n$.) aPELT($\theta^o$) performs the best in terms of TPR, FPR, number of estimated CP and MSE, since it explicitly utilizes the alternating structure of the epidemic mean change and the true normal state parameter $\mu^o = 0$. As expected, aPELT($\theta^o$) and aPELT_plugin have almost the same computational time as PELT.

In general, all variations of aPELT give better performance than PELT in terms of TPR, FPR, number of estimated CP and MSE, especially under the low signal scenario. Under most cases, aPELT_profile gives the second best performance after aPELT($\theta^o$). The TPR and MSE of aPELT_profile are noticeably better than PELT in all scenarios. On the other hand, aPELT_profile takes significantly more computational time than aPELT($\theta^o$) and aPELT_plugin since there is a second-stage optimization algorithm involved. Empirically, the computational time for aPELT_profile is around 10 to 20 times that of aPELT($\theta^o$) or aPELT_plugin, regardless of $n$. In other words, the computational cost of aPELT_profile is at the same scale with PELT.

As for the optimization of function $F^*(n, \mu^o)$ by aPELT_profile, Table 2 reports the proportion of experiments where the functional value attained by aPELT_profile is within 0.01 of the minimal value attained by grid search. Table 2 shows that, as $n$ grows or as the signal level $a$ increases, aPELT_profile with starting point at $\hat{\mu}^o$ can almost always find the global optimizer.

| $n$  | Low signal | Medium signal | High signal |
|------|------------|---------------|-------------|
| 1000 | 0.76       | 0.86          | 0.97        |
| 2000 | 0.87       | 0.99          | 1.00        |
| 10000| 1.00       | 1.00          | 1.00        |

Table 2: The proportion of experiments where the functional value attained by aPELT_profile is within 0.01 of the value attained by grid search under epidemic mean change for $n = 1000, 2000, 10000$ with $m = \sqrt{n/10}$.

To build more intuition, Figure 1 shows the examples of both cases where aPELT_profile successfully finds the global minimizer and where aPELT_profile stops at a local minimizer. Observe that $F^*(n, \mu^o)$ is indeed piecewise smooth and whether or not aPELT_profile stops at the global minimizer largely depends on the starting point $\hat{\mu}^o$. 

Figure 1: (a) An example that $aPELT$ stops at the global minimizer of $F^*(n, \mu^o)$ with $n = 10000$ and high signal level. (b) An example that $aPELT$ stops at a local minimizer of $F^*(n, \mu^o)$ with $n = 2000$ and low signal level.

5 Applications of aPELT

In this section, we provide applications of aPELT to large-scale multiple testing problems and to change-point detection of DNA copy number variation, and demonstrate the superior performance given by aPELT over existing methods.

5.1 Multiple testing with locally clustered signals

In this subsection, we apply aPELT to solve the multiple testing problem with locally clustered signals, which is an important special case of multiple testing and is studied by [Zhang et al. (2011)] and [Cao and Wu (2015)].

Under the basic setting, a sequence of independent $p$-values $p_1, \ldots, p_n$ are observed and we need to choose between the null hypothesis $p_1, \ldots, p_n \sim U(0, 1)$ and an alternative hypothesis with locally clustered signals. Specifically, the alternative hypothesis is formulated as follows: there exist change-points $0 < \tau_1 < \ldots < \tau_m < n$ such that

$$p_1, \ldots, p_{\tau_1} \sim U(0, 1), p_{\tau_1+1}, \ldots, p_{\tau_2} \not\sim U(0, 1), p_{\tau_2+1}, \ldots, p_{\tau_3} \sim U(0, 1), p_{\tau_3+1}, \ldots, p_{\tau_4} \not\sim U(0, 1), \ldots$$
or vice versa. Note that the alternative hypothesis shares a similar alternating structure as the multiple epidemic change-point problem, where the behavior of p-values alternates between a known common normal state (i.e. $U(0,1)$) and epidemic states, and thus can be potentially solved by aPELT($\theta^o$).

To operationalize aPELT($\theta^o$) for multiple testing with locally clustered signals, we model the p-values $p_t$ via the beta distribution Beta($\alpha, \beta$). Specifically, the p-values on the epidemic state are modeled by a beta distribution with parameters $\theta = (\alpha, \beta) \neq (1, 1)$ and the normal state p-values follow Beta$(1, 1) = U(0,1)$. In other words, aPELT($\theta^o$) is employed with $\theta^o = (1, 1)$ and the loss function $C(p_{(\tau_{i-1}+1):\tau_i}|\theta^i)$ is set to be twice the negative log-likelihood of $p_{(\tau_{i-1}+1):\tau_i}$ based on the beta distribution. After applying aPELT($\theta^o$) to $p_{1:n}$, we then reject the hypotheses for all the cases that are classified as epidemic states by aPELT. Note that no post-processing is needed since aPELT simultaneously estimates both the change-points and the alternating states of the sequence of p-values. We remark this is a general approach and can be directly applied to solve multiple testing with locally clustered signals regardless of the underlying true distribution of $p_t$.

We borrow the simulation setting from Cao and Wu (2015) where $p_{1:n}$ is a sequence of p-values generated by two-sided tests for mean. Specifically, we have $p_t = 2(1 - \Phi(|y_t|))$, where $y_{1:n}$ is a sequence of independent Gaussian random variables with variance $\sigma = 1$ and mean $\mu_t$ exhibited in Table 3, and $\Phi(\cdot)$ is the cdf of standard normal distribution. In other words, each $p_t$ is the p-value of the two-sided test for $E(y_t) = 0$.

| Segment (% among $n$) | 2.5 | 2.5 | 30 | 2.5 | 30 | 2.5 | 30 |
|-----------------------|-----|-----|----|-----|----|-----|----|
| Signal strength (mean level $\mu_t$) | 1 | -1.5 | 0 | 1.5 | 0 | -1.5 and 1 alternating | 0 |
| State (normal/epidemic) | E | N | E | N | E | N |

Table 3: Mean structure with locally clustered signals that alternates between epidemic (E) and normal (N) states.

The behavior of p-values $p_{1:n}$ alternates between the normal state $U(0,1)$ and different epidemic states at $\tau^o = n(0.05, 0.35, 0.375, 0.675, 0.7)$ depending on whether $\mu_t = 0$. Note that $\mu_t$ is not a constant on some epidemic states, for example, the first epidemic state $[1,0.05n]$ and the third epidemic state $[0.675n + 1,0.7n]$. Thus, the distribution of $p_t$ is not identical on those epidemic states. Furthermore, $p_t = 2(1 - \Phi(|y_t|))$ on epidemic states does not exactly follow the beta distribution. Hence aPELT($\theta^o$) encounters model misspecification under the current simulation...
setting. However, as is shown by the numerical experiments, with the flexible beta distribution, aPELT still gives robust and efficient performance under such misspecification.

We set \( n = 1000, 2000, 10000 \) and compare the performance of aPELT(\( \theta^o \)) and the proposed procedure in Cao and Wu (2015) (hereafter CW). For each \( n \), we repeat the simulation 100 times. We emphasize that for both procedures (aPELT and CW), the observed sequence is \( p\)-values \( p_1:n \) instead of \( y_1:n \). CW detects change-points and conducts multiple testing by thresholding a sequence of local scan statistics calculated via a screening window of size \( k \). As a local screening method, CW has a computational cost of \( O(n) \). The window size \( k \) is a tuning parameter and we try both \( k = \sqrt{n} \) and \( k = (\log n)^2 \) as suggested by Cao and Wu (2015).

As a multiple testing problem, the ultimate interest is the realized error rate of the performed tests instead of the accuracy of the estimated change-points. Thus, we evaluate the performance of the two algorithms by the false discovery rate (FDR), the false non-discovery rate (FNR) and the missed discovery rate (MDR) as suggested by Cao and Wu (2015). Specifically, the FDR follows the standard definition, the FNR is defined as the ratio of falsely accepted hypotheses and total accepted hypotheses and the MDR is defined as the ratio of falsely accepted hypotheses and total alternative hypotheses. The FNR and MDR can be used to describe the power of a multiple testing procedure, similar to the type II error rate in a single hypothesis testing setup.

The results are summarized in Table 4. In general, aPELT provides significantly smaller FNR and MDR than CW while having similar FDR, indicating that aPELT is more powerful under the current simulation setting and is robust under model misspecification. This is not surprising since aPELT is a likelihood-based global change-point detection procedure and explicitly incorporates the alternating behavior and the known normal state of \( p_1:n \). This example demonstrates the important application of aPELT to multiple testing and suggests that aPELT can serve as a promising and efficient tool for solving multiple testing problems with locally clustered signals.

### 5.2 DNA copy number variation

In this subsection, we apply aPELT, profile and PELT to detect mean change in DNA copy number variation. The data set is publicly available from R package “DNAcopy”, which contains two CGH arrays (Coriell.05296 and Coriell.13330) of Corriel cell lines from Snijders et al. (2001). We give the detailed result for Coriell.05296 here. The result for Coriell.13330 is qualitatively similar to...
Table 4: Average FDR, FNR and MDR of aPELT and two CWs for \( n = 1000, 2000, 10000 \).

|       | \( n \) | FDR  | FNR  | MDR  |
|-------|--------|------|------|------|
| aPELT | 1000   | 7.31 | 3.85 | 36.22|
|       | 2000   | 5.23 | 1.50 | 13.85|
|       | 10000  | 0.91 | 0.22 | 2.02 |
| CW \( k = \sqrt{n} \) | 1000   | 2.39 | 8.05 | 79.07|
|       | 2000   | 4.29 | 4.49 | 41.08|
|       | 10000  | 2.19 | 2.24 | 18.37|
| CW \( k = (\log n)^2 \) | 1000   | 2.36 | 8.09 | 79.45|
|       | 2000   | 6.12 | 4.74 | 41.98|
|       | 10000  | 5.97 | 3.12 | 22.53|

Coriell.05296 and thus is omitted.

The estimated change-point locations by PELT and aPELT_profile are given in Figure 2. Observe that the estimated change-points by the two algorithms are very similar to each other, while aPELT_profile has one extra change-points. As expected, the estimated mean vector \( \hat{\mu}_{1:n} \) by both algorithms exhibit the alternating structure, which further justifies the use of aPELT_profile. In terms of model selection, the BIC achieved by the final estimated model given by PELT is \(-4292.76\) and the BIC achieved by aPELT_profile is \(-4318.67\), thus aPELT_profile gives a better change-point model by explicitly incorporating the epidemic structure.

For the optimization result, the normal state parameter estimated by aPELT_profile via \( \tilde{\mu} = \arg \min_{\mu} F^*(n; \mu) \) along with the function \( F^*(n; \mu) \) are plotted in Figure 3. It confirms that \( F^*(n; \mu) \) is piecewise smooth and that aPELT_profile successfully achieves the minimal value of \( F^*(n; \mu) \), with the estimate \( \tilde{\mu} = -0.07 \), which is very close to 0. In fact, aPELT(\( \theta^o \)) with \( \mu^o = 0 \) gives exactly the same estimated change-points with aPELT_profile, which further indicates the robustness of the estimation result.

6 Discussion

In this paper, we generalize the classical (single) epidemic change-point detection problem to a more realistic multiple epidemic change-point setting. The inference interests are to recover the unknown number and locations of change-points and also to recover the alternating states of the observed sequence.

To solve the new problem, we develop a modified penalized loss function for multiple change-
point estimation which incorporates explicit treatment of the alternating states and common normal state parameter. The modified loss function cannot be solved directly by existing dynamic optimization algorithms such as PELT, which motivates us to propose a novel alternating algorithm, named aPELT, for the efficient and exact optimization of the loss function. The proposed aPELT employs a two-stage optimization procedure and shares the same spirit as profile likelihood. Empirically, the computational cost of aPELT is at the same scale of PELT, which is nearly linear with an upper bound $O(n^2)$. The efficiency and superior performance of aPELT over existing methods is demonstrated via extensive numerical experiments and relevant applications to multiple testing and DNA copy number variation.

The aPELT can be extended to more complicated epidemic change-point setting. For example,
Figure 3: The $F^*(n; \mu^o)$ function and its global minimizer $\hat{\mu}^o$ estimated by aPELT PROFILE for Coriell.05296.

instead of assuming a constant change for the epidemic state, a nonparametric curve can be used for better approximation. Another possible extension is to generalize the proposed epidemic detection procedure to multivariate or high dimensional sequences.
References

Arias-Castro, E., Donoho, D. L., and Huo, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Transactions on Information Theory*, 51(7):2402–2425.

Cao, H. and Wu, W. B. (2015). Change-point estimation: another look at multiple testing problems. *Biometrika*, 102(4):974–980.

Chan, N. H., Yau, C. Y., and Zhang, R. M. (2014). Group lasso for structural break time series. *Journal of the American Statistical Association*, 109(506):590–599.

Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101:223–239.

Fearnhead, P. (2006). Exact and efficient bayesian inference for multiple changepoint problems. *Statistics and Computing*, 16:203–213.

Fearnhead, P. and Rigaill, G. (2017). Changepoint detection in the presence of outliers. *Journal of the American Statistical Association*.

Guan, Z. (2004). A semiparametric changepoint model. *Biometrika*, 91(4):849–862.

Hocking, T. D., Rigaill, G., and Bourque, G. (2015). Peakseg: constrained optimal segmentation and supervised penalty learning for peak detection in count data. *Proceedings of the 32nd International Conference on Machine Learning*.

Hocking, T. D., Rigaill, G., Fearnhead, P., and Bourque, G. (2018). Generalized functional pruning optimal partitioning (gfpop) for constrained changepoint detection in genomic data. *arXiv:1810.00117v1*.

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

Killick, R., P., F., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of American Statistical Association*, 107:1590–1598.
Levin, B. and Kline, J. (1985). The cusum test of homogeneity with an application in spontaneous abortion epidemiology. *Statistics in Medicine*, 4:469–488.

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

Matteson, D. S. and James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, 109(505):334–345.

Ning, W., Pailden, J., and Gupta, A. (2012). Empirical likelihood ratio test for the epidemic change model. *Journal of Data Science*, 10:107–127.

Niu, Y. S. and Zhang, H. (2012). The screening and ranking algorithm to detect dna copy number variations. *The Annals of Applied Statistics*, 6(3):1306–1326.

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

Snijders, A., Nowak, N., Segraves, R., and Blackwood, S. (2001). Assembly of microarrays for genome-wide measurement of dna copy number. *Nature Genetics*, (29):263–264.

Yao, Q. (1993). Tests for change-points with epidemic alternatives. *Biometrika*, 80(1):179–191.

Yao, Y. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *Annals of Statistics*, 15:1321–1328.

Zhang, C., Fan, J., and Yu, T. (2011). Multiple testing via fdrl for large-scale imaging data. *Annals of Statistics*, (39):613–642.

Zou, C., Yin, G., Feng, L., and Wang, Z. (2014). Nonparametric maximum likelihood approach to multiple change-point problems. *The Annals of Statistics*, 42(3):970–1002.