

Optimal covariance change point localization in high dimensions

DAREN WANG¹, YI YU² and ALESSANDRO RINALDO³

¹*Department of Statistics, University of Chicago, 5747 S. Ellis Avenue, Jones 120A, Chicago, IL 60637, USA.
E-mail: darenw@galton.uchicago.edu*

²*Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom.
E-mail: yi.yu.2@warwick.ac.uk*

³*Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA.
E-mail: arinaldo@cmu.edu*

We study the problem of change point localization for covariance matrices in high dimensions. We assume that we observe a sequence of independent and centered p -dimensional sub-Gaussian random vectors whose covariance matrices are piecewise constant, and only change at unknown times. We are concerned with the localization task of estimating the positions of the change points. In our analysis, we allow for all the model parameters to change with the sample size n , including the dimension p , the minimal spacing between consecutive change points Δ , the maximal Orlicz- ψ_2 norm B of the sample points and the magnitude κ of the smallest distributional change, defined as the minimal operator norm of the difference between the covariance matrix at a change point and the covariance matrix at the previous time point.

We introduce two procedures, one based on the binary segmentation algorithm and the other on its popular extension known as wild binary segmentation, and demonstrate that, under suitable conditions, both procedures can consistently estimate the change points. In particular, our second algorithm, called Wild Binary Segmentation through Independent Projection (WBSIP), delivers a localization error of order $B^4 \kappa^{-2} \log(n)$, which is shown to be minimax rate optimal, save, possibly, for the $\log(n)$ term. WBSIP requires the model parameters to satisfy the scaling $\Delta \kappa^2 \gtrsim p B^4 \log^{1+\xi}(n)$, for any $\xi > 0$, which we demonstrate to be essentially necessary, in the sense that no algorithm can guarantee consistent localization if $\Delta \kappa^2 \lesssim p B^4$. This result reveals an interesting phase transition effect separating parameter combinations for which the localization task is feasible from the ones for which it is not.

Keywords: binary segmentation; change point detection; high-dimensional covariance testing; independent projection; minimax optimal; wild binary segmentation

1. Introduction

Statistical change point analysis is concerned with identifying abrupt changes in the data, generally observed as a time series or as a realization of a stochastic or spatial process, that are due to actual changes in the underlying distribution and not random fluctuations. Applications of change point analysis are ubiquitous, and include security monitoring, neuroimaging, financial trading, ecological statistics, climate change, medical condition monitoring, sensor networks, disease outbreak risk assessment, flu trend analysis, genetics and various others.

In its most basic form, change point modeling postulates a discrete times series (X_1, \dots, X_n) of variates whose marginal distributions are piecewise constant. Specifically, for some *unknown* increasing subsequence $\{\eta_1, \dots, \eta_K\} \subset \{2, \dots, n\}$ of *change points*,

$$X_t \sim P_k, \quad \text{if } t \in \{\eta_k, \dots, \eta_{k+1} - 1\}, \text{ for } k \in \{0, \dots, K\}, \quad (1)$$

where $\eta_0 = 1$, $\eta_{K+1} = n + 1$, and $\{P_0, \dots, P_K\}$ are probability distributions such that $P_k \neq P_{k-1}$ for all $k = 1, \dots, K$. One of the main inferential goals in change point analysis is the estimation of the positions of the change points, a task we will refer to as *localization*.

The very first, most basic and most studied change point model assumes an independent time series of random variables with piecewise constant means, that is,

$$Y_t = \mu_t + \epsilon_t, \quad t = 1, \dots, n, \tag{2}$$

where $(\mu_1, \dots, \mu_n) \in \mathbb{R}^n$ is such that $\mu_t \neq \mu_{t-1}$ if and only if $t \in \{\eta_1, \dots, \eta_K\}$ and the ϵ_t 's are i.i.d. centered random variables with variance σ^2 . Despite its simplicity, the univariate mean change point model has been a blueprint for studying more complex problems and has led to theoretical and methodological advances. State-of-the-art methods for this model include recently developed change point algorithms, such as PELT of Killick, Fearnhead and Eckley [26], WBS of Fryzlewicz [19] and SMUCE of Frick, Munk and Sieling [18], as well as the renowned binary segmentation (BS) procedure (see, e.g., Venkatraman [37]). Despite the vast literature on this model, a complete theoretical analysis of the associated localization task, encompassing the derivation of minimax rates and the characterization of phase transition effects, has appeared only very recently; see Wang, Yu and Rinaldo [40].

In this paper, we are concerned with the localization task in the change point model (1) whereby the data consist of a sequence of independent centered sub-Gaussian vectors and the distributional changes occurring at the change points take the form of generic changes in the corresponding covariance matrices, measured in the operator norm. We formalize this model next. Below $\|\Sigma\|_{\text{op}}$ denotes the $\ell_2 \rightarrow \ell_2$ -operator norm of a matrix Σ and $\|X\|_{\psi_2}$ the ψ_2 or sub-Gaussian norm of a random vector X (see Section S.1 for a definition).

Assumption 1 (Covariance change point model). Let $X_1, \dots, X_n \in \mathbb{R}^p$ be independent, zero mean random vectors such that $\mathbb{E}(X_t X_t^\top) = \Sigma_t$ and $\|X_t\|_{\psi_2} \leq B$ for all $t = 1, \dots, n$, where $B > 0$. Let $\{\eta_0, \dots, \eta_{K+1}\} \subset \{1, \dots, n + 1\}$ be an increasing subsequence of change points such that $\eta_0 = 1$, $\eta_{K+1} = n + 1$ and $\Sigma_t \neq \Sigma_{t-1}$ if and only if $t \in \{\eta_1, \dots, \eta_K\}$. The minimal spacing between jumps is such that

$$\min_{k=1, \dots, K+1} \{\eta_k - \eta_{k-1}\} = \Delta > 0,$$

and the magnitude of each change is

$$\|\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}\|_{\text{op}} = \kappa_k, \quad k = 1, \dots, K,$$

with $\min_{k=1, \dots, K} \kappa_k = \kappa > 0$, where Δ and κ are positive quantities.

The parameters p , Δ , K , B and κ completely characterize the difficulty of the change point localization problem, which intuitively, for a given n , should be increasing in p , B and K , and decreasing in Δ and κ . In fact, because of the upper bound $K \leq \frac{n}{\Delta}$, we will not be concerned with the parameter K . Throughout, we allow p , Δ , B and κ to be functions of the sample size n , although we do not make this dependence explicit in our notation for ease of readability. This general setting allows us to study the localization problem in growing dimensions, with a growing number of change points and of increasing difficulty, so that, as we gather more data we are able to successfully tackle harder localization tasks. We refer to any relationship holding among all the model parameters (p, Δ, B, κ) and the sample size n as a *scaling*.

For a given scaling, we seek to produce estimators of (η_1, \dots, η_K) of the form

$$(X_1, \dots, X_n) \mapsto (\hat{\eta}_1, \dots, \hat{\eta}_{\hat{K}}) \subset \{2, \dots, n\}$$

such that $\hat{\eta}_1 < \dots < \hat{\eta}_{\hat{K}}$ and, with probability tending to 1 polynomially fast in n as $n \rightarrow \infty$,

$$\hat{K} = K \quad \text{and} \quad \max_{k=1, \dots, K} |\hat{\eta}_k - \eta_k| \leq \epsilon, \tag{3}$$

where $\epsilon = \epsilon(n, p, \Delta, B, \kappa)$. We will refer to the quantity ϵ as the *localization error* of the estimator and we will deem such estimator *consistent* if

$$\lim_{n \rightarrow \infty} \frac{\epsilon}{n} = 0$$

that is, if, with probability tending to one, the number of change points is estimated perfectly and the maximal distance between any true change point and the corresponding estimator is vanishing in sample size. We will refer to the sequence $\{\epsilon/n\}$ as the *localization rate*. Our goals are to derive (i) conditions on the scaling of the model parameters that allow for consistent estimation of the change points and (ii) computationally-feasible estimators that are consistent and in fact optimal, in the sense of achieving minimax localization rates.

We conclude this section by noting that the parameters κ and B are not variation independent, as they satisfy the inequality $\kappa \leq B^2/4$. Indeed,

$$\begin{aligned} \kappa &\leq \max_{k=1}^K \|\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}\|_{\text{op}} \leq 2 \max_{t=1}^n \|\Sigma_t\|_{\text{op}} = 2 \max_{t=1}^n \sup_{v \in \mathcal{S}^{p-1}} \mathbb{E}[(v^\top X_t)^2] \\ &\leq 4 \max_{t=1}^n \|X_t\|_{\psi_2}^2 \leq 4B^2, \end{aligned} \tag{4}$$

where \mathcal{S}^{p-1} is the Euclidean unit sphere in \mathbb{R}^p and the second-to-last inequality follows from (S.1) in Section S.1. In addition to showing that the covariance matrices are all positive definite, the previous chain of inequalities reveals that the larger the smallest magnitude κ of the changes, the higher the variance parameter B should be expected to be – two phenomena affecting the difficulty of the localization task in opposite ways. Interestingly, this feature does not arise in the mean change point localization problem, where the corresponding parameters are decoupled. The combined effect of κ and B may be quantified by κB^{-2} , which should be thought of as a signal-to-noise ratio of sort. Clearly, such quantity, and, as a result, the localization task itself, remains invariant with respect to any multiplicative rescaling of the data by an arbitrary non-zero constant (though of course, not with respect to arbitrary translations, like in change point localization for means).

1.1. Relevant and related literature

The literature on change point detection is extremely rich and covers a large variety of models. Arguably one of the most studied change point problems is the one concerning the localization for a univariate piecewise constant signal corrupted by independent additive noise. The list of contributions in this area is large and includes Yao and Au [44], Wang [43], Lavielle [27], Lavielle and Moulines [28], Davies and Kovac [15], Davis, Lee and Rodriguez-Yam [16], Harchaoui and Lévy-Leduc [21], Qian and Jia [35], Rojas and Wahlberg [36], Frick, Munk and Sieling [18], Lin et al. [31] and Li, Guo and Munk [29].

Among the existing methods, binary segmentation (BS, e.g., Vostrikova [38]) is ‘arguably the most widely used change point search method’ (Killick, Fearnhead and Eckley [26]). The algorithm goes through the whole time course and scans for a change point. If one is detected, then the whole time course is split into two, and the same procedure is deployed separately on the data before and after

the detected change point. The procedure is carried on recursively until no change point is detected, or the remaining time course consists of too few time points to continue. Venkatraman [37] proves the consistency of the BS procedure in the univariate time series mean change point detection, with the number of change points allowed to increase with the number of time points. Fryzlewicz [19] proposes a variant of BS, called wild binary segmentation (WBS), which can be viewed as a flexible moving window technique, or a hybrid of moving window and BS. WBS randomly draws a collection of random time intervals, conducts BS on each of them separately, and returns the time point with the largest CUSUM contrast. This point is deemed a change point and the procedure is then repeated recursively. Generally, WBS is preferable to BS when many change points are present. In the univariate time series mean change point detection problem, Venkatraman [37] shows that in order to achieve the estimating consistency using the BS algorithm, the minimum gap between two consecutive change points should be at least of order $n^{1-\beta}$, where n is the number of time points, and $0 \leq \beta < 1/8$ (therefore the corresponding localization rate is $n^{5/8+\beta} \log(n)$); as claimed in Fryzlewicz [19], this rate can be reduced to $\log(n)$ using the WBS algorithm. Overall, the WBS and the multi-scaled method of Frick, Munk and Sieling [18] are regarded as the state of the art for this problem. Recently, Wang, Yu and Rinaldo [40] have shown that both WBS and the l_0 penalized least square method studied in Boysen et al. [9], which can be efficiently implemented with the PELT algorithm of Killick, Fearnhead and Eckley [26], deliver minimax localization rates for this problem.

The literature mentioned above focuses on univariate time series models. However, in the big data era, data sets are now routinely more complex and of high dimensions. Horváth and Hušková [22] propose a variant of the CUSUM statistic by summing up the square of the CUSUM statistic in each coordinate. Cho and Fryzlewicz [13] transform a univariate non-stationary time series into multi-scale wavelet regime, and conduct BS at each scale in the wavelet context. Jirak [24] allows the dimension p to tend to infinity together with the sample size n , by taking maxima statistics across panels coordinate-wise. Cho and Fryzlewicz [14] propose sparsified binary segmentation method which aggregates the CUSUM statistics across the panel by adding those which exceed a certain threshold. Cho [12] proposes the double CUSUM statistics which, at each time point, picks the coordinate which maximizes the CUSUM statistic, and *de facto* transfers the high-dimensional data to a univariate CUSUM statistics sequence. Aston and Kirch [2] introduce the asymptotic concept of high-dimensional efficiency which quantifies the detection power of different statistics in this setting. Wang and Samworth [42] study the mean change point localization problem in high dimensions under appropriate sparsity assumptions.

As for change point detection in more general scenarios, extensions of the sequential probability ratio test procedure (Wald [39]) can be devised for variance-based change point detection. Based on a generalized likelihood ratio statistic, Baranowski, Chen and Fryzlewicz [5] tackle a range of univariate time series change point scenarios, including the variance change situations, although theoretical results are missing. Picard [34] proposes tests on the existence of change points in terms of spectrum and variance. Inclan and Tiao [23] develop an iterative cumulative sums of squares algorithm to detect the variance changes. Gombay, Horváth and Hušková [20] propose some tests on detection of possible changes in the variance of independent observations and obtain the asymptotic properties under the no-change null hypothesis. Berkes et al. [7], among others, extend the tests and corresponding results to linear processes, as well as ARCH and GARCH processes. Aue et al. [3] consider the problem of variance change point detection in a multivariate time series model, allowing the observations to have m -dependent structures. Note that the consistency results in Aue et al. [3] are in the asymptotic sense that the number of time points diverges and the dimension of the time series remains fixed. Aue et al. [3] also require the existence of good estimators of the covariance and precision matrices, and the conditions thereof are left implicit. Barigozzi, Cho and Fryzlewicz [6] deal with a factor model, which is potentially of high dimension $p/n = O(\log^2(n))$, and use the wavelet transforms to make the data possibly dependent across the timeline. Note that the model in Barigozzi, Cho and Fryzlewicz [6] can

be viewed as a specific covariance change point problem, where the additional structural assumption allows the dimensionality to go beyond the sample size. Dette, Pan and Yang [17] considered a high-dimensional covariance change point detection problem with the dimensionality allowed to exceed the sample size.

As for the problem of hypothesis testing for high dimensional covariance matrices, which corresponds to the problem of change point detection, the literature is also abundant, and includes the work of Anderson [1], Johnstone [25], Birke and Dette [8], Cai and Ma [11], Li and Chen [30], Cai, Liu and Xia [10], to name but only a few.

Our procedures and results have been heavily inspired by two contributions in the recent literature on change point analysis: Fryzlewicz [19] and Wang and Samworth [42]. Below, we highlight the differences between our settings and results and theirs.

Our WBSIP procedure utilizes in a fundamental way the WBS algorithm put forward by Fryzlewicz [19] for change point localization in univariate time series with piecewise constant means. Here we have carried out a new theoretical analysis of the performance of WBS that resolves some issues contained in Fryzlewicz [19] and, in addition, yields an optimal dependence on the model parameters, especially κ . This improvement is non-trivial and may be of independent interest, as it is also directly applicable to the univariate mean change point localization problem itself; see Wang, Yu and Rinaldo [40].

The WBSIP procedure further employs sample splitting and independent projections, much like the methodology `inspect` of Wang and Samworth [42] for change point localization of sparse, high-dimensional means (which too relies on WBS). WBSIP procedure is different in its design, properties and goals from the algorithm `inspect`, which, unlike WBSIP, relies on semidefinite programming and is indirectly targeting the L_2 recovery of the support of a sparse mean vector. In contrast, we focus on the operator norm of the difference of the covariance matrices around the change points. Since the change point models are different, so are the assumptions we impose; in particular, we do not require any eigengap condition. As a result, the theoretical analysis of WBSIP calls for different arguments, and a direct adaptation of the techniques and results of Wang and Samworth [42] to our settings leads to sub-optimal localization rates.

1.2. List of contributions

We propose and analyze two algorithms for covariance change point localization. The first one, called BSOP (Binary Segmentation through Operator Norm), is based on a straightforward adaptation to the covariance setting of the BS algorithm for the univariate mean change point localization problem. The BSOP algorithm is rather simple to implement and relatively fast. Under appropriate assumptions, we show in Theorem 2.1 that BSOP can consistently estimate all the change points, but with a sub-optimal localization rate that exhibits an unfavorable dependence on the dimension p . This finding is consistent with the conclusions of the analysis of the BS algorithm for the univariate mean localization problem contained in Fryzlewicz [19], who showed that the BS algorithm is consistent but possibly sub-optimal.

Our second algorithm, called WBSIP (Wild Binary Segmentation through Independent Projections), is significantly more refined and yields much sharper, in fact almost minimax rate-optimal, localization rates than BSOP under a set of different and milder assumptions. Specifically, we demonstrate a phase transition effect over the space of the model parameters: if $\Delta \lesssim B^4 p \kappa^{-2} \log(n)$ then no consistent estimator of the locations of the change points exists; see Lemma 3.1. On the other hand, Theorem 2.2 shows that, if $\Delta \gtrsim B^4 p \kappa^{-2} \log^{1+\xi}(n)$, for any $\xi > 0$, then WBSIP will yield a localization rate of the order $B^4 \kappa^{-2} \log(n)$, independent of p . In fact, up to a $\log(n)$ term, this rate turns out to be minimax optimal; see Lemma 3.2. Thus, WBSIP delivers optimal performance over nearly all scalings for which consistent localization is possible.

While consistency of change point estimation for high dimensional mean vectors and covariance matrices has been recently studied by several authors (see, e.g., Baranowski, Chen and Fryzlewicz [5], Wang and Samworth [42], Aue et al. [3], Avanesov and Buzun [4]), to the best of our knowledge, neither the phase transition effect nor the minimax rate optimality have been established elsewhere. Overall, our lower bound results and the upper bound on the localization rate afforded by WBSIP procedure provide a complete characterization of the problem of change point localization in the setting considered here.

The above guarantees hold without assuming any structural property of the underlying covariance matrices or of their differences, such as sparsity or low rank form. Should these additional assumptions be made, our procedures and analysis would have to be modified in order to take advantage of such properties and to obtain consistent localization rates that will presumably allow for a larger dimension p .

2. Main results

2.1. The covariance CUSUM statistic and the shadow vector

The bulk of our analysis revolves around studying, as a function of t , the mean and fluctuations of the following matrix-valued statistic, which is a natural generalization to the covariance settings of the renowned univariate CUSUM statistic (e.g., Page [33], Yao and Au [44]) for mean change point detection.

Definition 1 (Covariance CUSUM). For $X_1, \dots, X_n \in \mathbb{R}^p$, a pair of integers (s, e) such that $0 \leq s < e - 1 < n$, and any $t \in \{s + 1, \dots, e - 1\}$, the covariance CUSUM statistic is defined as

$$\tilde{S}_t^{s,e} = \sqrt{\frac{e-t}{(e-s)(t-s)}} \sum_{i=s+1}^t X_i X_i^\top - \sqrt{\frac{t-s}{(e-s)(e-t)}} \sum_{i=t+1}^e X_i X_i^\top.$$

For any such t , we let $\tilde{\Sigma}_t^{s,e} = \mathbb{E}[\tilde{S}_t^{s,e}]$.

One may be tempted to apply the arguments for proving consistency of the BS algorithm for localization in univariate mean change point detection problems, as done in Venkatraman [37], to the covariance CUSUM statistic. Unfortunately, it is non-trivial to prove that the function $t \mapsto \|\tilde{\Sigma}_t^{s,e}\|_{\text{op}}$ achieves its local maxima at the change points. To overcome such difficulty, we study instead, for each pair of integers (s, e) , $0 \leq s < e - 1 < n$ with $e - s > 2p \log(n)$, the univariate time series $\{(v^\top X_t)^2\}_{t=s+1, \dots, e}$ of the squared coefficients of the projection of the data along a one-dimensional linear subspace spanned by a distinguished unit vector v , which we refer to as a *shadow vector*. The shadow vector is simply the leading singular vector of $\tilde{\Sigma}_b^{s,e}$, where $b = \arg \max_{t \in (s+p \log(n), e-p \log(n))} \|\tilde{\Sigma}_t^{s,e}\|_{\text{op}}$. As it turns out, with such a choice of the shadow vector, the local maxima of CUSUM statistic applied to the corresponding one-dimensional time series approximately coincide with the local maxima of the time series of the operator norms of the CUSUM covariance statistics. Thus, for the purpose of identifying the local maxima of the CUSUM covariance statistic, it is enough and in fact much simpler to study the univariate times series of the squared projections onto the corresponding shadow vector. In turn, the shadow vector can be estimated (using sample splitting) by the leading singular vector of $\tilde{S}_b^{s,e}$, where $\hat{b} = \arg \max_{t \in (s+p \log(n), e-p \log(n))} \|\tilde{S}_t^{s,e}\|_{\text{op}}$. Interestingly, in order to yield consistent and in fact optimal localization, such an estimator need not be itself consistent; for this reason, we do not need

Algorithm 1 Binary Segmentation through Operator Norm. BSOP($\{X_t\}_{s+1}^e, \tau$)

INPUT: $\{X_t\}_{t=s+1}^e \subset \mathbb{R}^{p \times (e-s)}, \tau > 0$.

Initial FLAG $\leftarrow 0$
if $e - s > 2p \log(n) + 1$ **and** **FLAG** = 0 **then**
 $a \leftarrow \max_{\lceil s+p \log(n) \rceil \leq t \leq \lfloor e-p \log(n) \rfloor} \|\tilde{S}_t^{s,e}\|_{\text{op}}$
if $a \leq \tau$ **then**
 $\text{FLAG} \leftarrow 1$
else
 $b \leftarrow \arg \max_{\lceil s+p \log(n) \rceil \leq t \leq \lfloor e-p \log(n) \rfloor} \|\tilde{S}_t^{s,e}\|_{\text{op}}$

 add b to the collection of estimated change points

 BSOP($\{X_t\}_{s+1}^{b-1}, \tau$)

 BSOP($\{X_t\}_{b-1}^e, \tau$)

end if
end if
OUTPUT: The collection of estimated change points.

to impose any eigengap condition on the matrices $\tilde{S}_t^{s,e}$. We provide further comments on the uses and interpretation of the shadow vector projections, which can be generally regarded as a dimension reduction scheme, below and in Section S.4.2 in the supplemental article Wang, Yu and Rinaldo [41].

2.2. Consistency of the BSOP algorithm

Our first algorithm, called BSOP, stems from a direct adaptation of BS to the matrix setting based on the distance induced by the operator norm; see Algorithm 1. The BSOP procedure is computationally and conceptually simple: given any time interval (s, e) , BSOP first computes the maximal operator norm of the covariance CUSUM statistics over the time points in $(s + \lceil p \log(n) \rceil, e - \lfloor p \log(n) \rfloor)$; if such maximal value exceeds a predetermined threshold τ , then BSOP will identify the location b of the maximum as a change point. The interval (s, e) is then split into two subintervals at b and the procedure is then iterated separately on each of them until an appropriate stopping condition is met.

The BSOP algorithm differs from the standard BS implementation in one aspect: the maximization of the operator norm of the CUSUM covariance operator is carried out only over the time points in (s, e) that are away by at least $p \log(n)$ from the endpoints of the interval. Such restriction is needed to obtain adequate tail bounds on the operator norm of the covariance CUSUM statistics $\tilde{S}_t^{s,e}$. See Lemma S.1.1 in Section S.1 in the supplemental article Wang, Yu and Rinaldo [41].

To analyze the performance of the BSOP algorithm we will impose the following assumption, which is, for the most part, modeled after Assumption 3.2 in Fryzlewicz [19], whose notation we adopt.

Assumption 2. For an increasing diverging sequence $\{a_n\}_{n=1,2,\dots}$ of positive numbers, a sufficiently large constant $C_\alpha > 0$ and a sufficiently small constant $c_\alpha > 0$, assume that $\Delta \kappa B^{-2} \geq C_\alpha n^\Theta a_n$, $p \leq c_\alpha n^{8\Theta-7} / \log(n)$, where $\Theta \in (7/8, 1]$.

The sequence $\{a_n\}_{n=1,2,\dots}$ may diverge arbitrarily slowly and is only needed to manage the case in which $\Theta = 1$, which arises for instance, when the number of change points is bounded (in n) and κ and B are constants. When the parameters κ and B are fixed, the above assumption requires Δ , the minimal spacing between consecutive change points, to be of at least slightly smaller order than the size of the time series. This is precisely Assumption 3.3 in Fryzlewicz [19] (see also Cho and Fryzlewicz

[14]). The fact that Δ cannot be too small compared to n in order for the BS algorithm to exhibit good performance is well known: see, for example Olshen et al. [32]. In Assumption 2, we require also the dimension p to be upper bounded by $n^{8\Theta-7} \log^{-1}(n)$, which means that p is allowed to diverge as $n \rightarrow \infty$.

Remark 2.1 (Generalizing Assumption 2). In Assumption 2, we impose certain constraints on the scaling of the quantities B, κ, Δ and p in relation to n that are captured by a single parameter Θ , whose admissible values lie in $(7/8, 1]$. The strict lower bound of $7/8$ on Θ is determined by the calculations outlined below in (8) and (9). In fact, Assumption 2 may be generalized by allowing for different types of scaling in n of the signal-to-noise ratio κB^{-2} , the minimal distance Δ between consecutive change points and the dimension p . In detail, we may require that $\kappa B^{-2} \gtrsim n^{\Theta_1}$, $\Delta \gtrsim n^{\Theta_2}$ and $p \log(n) \lesssim n^{\Theta_3}$ for a given triplet of parameters $(\Theta_1, \Theta_2, \Theta_3)$ in an appropriate subset of $[0, 1]^{\otimes 3}$. Such a generalization would then lead to consistency rates in n that depend on all these parameters simultaneously. However, the range of allowable values of $(\Theta_1, \Theta_2, \Theta_3)$ is not a product set due to non-trivial constraints among them. We will refrain from providing details and instead rely on the simpler formulation given in Assumption 2.

Theorem 2.1 (Consistency of BSOP). Under Assumptions 1 and 2, let $\{\hat{\eta}_k\}_{k=1}^{\widehat{K}}$ be the collection of the estimated change points from the BSOP($\{X_t\}_1^n, \tau$) algorithm, where the parameter τ satisfies

$$B^2(\sqrt{p \log(n)} + 2\sqrt{\epsilon_n}) < \tau < C_1 \kappa \Delta n^{-1/2}, \tag{5}$$

for some constant $C_1 \in (0, 1)$. Then, there exists a universal constant $c > 3$ such that

$$\mathbb{P}\left(\widehat{K} = K \text{ and } \max_{k=1, \dots, K} |\eta_k - \hat{\eta}_k| \leq \epsilon_n\right) \geq 1 - 2 \times 9^p n^{3-cp}, \tag{6}$$

where

$$\epsilon_n = C_2 B^2 \kappa^{-1} n^{5/2} \Delta^{-2} \sqrt{p \log(n)}, \tag{7}$$

for some $C_2 > 0$.

The condition (5) on the admissible values of the input parameter τ of the BSOP algorithm is well defined. Indeed, by Assumption 2, for all pairs (s, e) such that $e - s > 2p \log(n)$, we have that

$$B^2 \sqrt{p \log(n)} \leq B^2 c_\alpha^{1/2} n^{4\Theta-7/2} \leq B^2 c_\alpha^{1/2} n^\Theta n^{-1/2} \leq \frac{c_\alpha^{1/2}}{C_\alpha} \kappa \Delta n^{-1/2} \tag{8}$$

and

$$\begin{aligned} 2\sqrt{\epsilon_n} B^2 &= 2C_2^{1/2} B^3 \kappa^{-1/2} n^{5/4} \Delta^{-1} (p \log(n))^{1/4} \leq (2C_2^{1/2} C_\alpha^{-1} c_\alpha^{1/4}) B \kappa^{1/2} n^{5/4+\Theta-7/4} a_n^{-1} \\ &\leq (2C_2^{1/2} C_\alpha^{-1} c_\alpha^{1/4}) \kappa \Delta n^{-1/2} B^{-1} \kappa^{1/2} a_n^{-1} \leq (4C_2^{1/2} C_\alpha^{-1} c_\alpha^{1/4}) \kappa \Delta n^{-1/2}, \end{aligned} \tag{9}$$

where in the chain of inequalities we have used Assumption 2 repeatedly, and the last inequality in (9) relies on the bound (4). It is also worth noting that the difference between the right-hand side and the left-hand side of (5) increases as Θ increases to 1. Finally, we remark that in the proof of Theorem 2.1, we actually let $C_1 = 1/8$, but this is an arbitrary choice and it essentially depends on the constants C_α and c_α from Assumption 2, see, for instance, (8) and (9).

Theorem 2.1 implies, that with high probability, the BSOP algorithm will identify all the change points and estimate their locations consistently, since, due to Assumption 2 and the fact that $\kappa \leq B^2/4$, we have that

$$\begin{aligned} \frac{\epsilon_n}{n} &\lesssim \frac{B^2}{\kappa} \Delta^{-2} n^{3/2} \sqrt{p \log(n)} \lesssim \left(\frac{B^2}{\Delta \kappa}\right)^2 \frac{\kappa}{B^2} n^{3/2} n^{4\Theta-7/2} \\ &\lesssim n^{-2\Theta+3/2+4\Theta-7/2} a_n^{-2} \lesssim a_n^{-2} \rightarrow 0. \end{aligned}$$

As expected, the localization error afforded by BSOP is decreasing in the signal-to-noise ratio parameter κB^{-2} , and in the minimal distance Δ between change points and the dimension p . The above bound yields a family of rates of consistency for BSOP, depending on the scaling of each of the quantities involved in it. For example, in perhaps the simplest and most favorable scenario where B , κ and the dimension p are constants, the bound implies a rate for change point localization of the order

$$\epsilon_n \lesssim n^{-2\Theta+5/2} a_n^{-2} \sqrt{\log(n)},$$

which is decreasing in $\Theta \in (7/8, 1]$. In particular, when the number of change points is also kept constant, we have that $\Theta = 1$, yielding a localization error of order $a_n^{-2} \sqrt{n \log(n)}$.

As we will see in the next subsection, the dependence on the parameters κ/B^2 , p and Δ is sub-optimal. The advantage of BSOP over the rate-optimal algorithm we introduce next, besides its simplicity, is that BSOP only requires one input parameter, the threshold value τ . Furthermore, when the spacing parameter Δ is comparable with the sample size n and when the dimension p of the data grows slowly with respect with n , then BSOP can still deliver good consistency rates.

2.3. Consistency of the WBSIP algorithm

In this section, we introduce a more complex and effective algorithm for covariance change point detection, which we name WBSIP for Wild Binary Segmentation through Independent Projections. The WBSIP algorithm is a generalization of the WBS procedure for mean change point detection and relies on the properties of shadow vectors mentioned in Section 2.1. WBSIP begins by splitting the data into halves and by selecting at random a collection of M pairs of integers (s, e) such that $0 \leq s < e - 1 \leq n$ and $e - s > 2p \log(n) + 1$. This sample splitting can be done by separating data into the odd and even indices subsets. In its second step, WBSIP computes, for each of the M random integer intervals previously generated, a shadow vector using one half of the data and its corresponding univariate time series using the other half. The final step of the procedure is to apply the WBS algorithm over the resulting univariate time series. The details of the algorithm are given in Algorithm 2, which describes the computation of the shadow vectors by principal component, and Algorithm 3, which applies WBS to the resulting univariate time series. Besides the reliance on WBS instead of BS, the other key difference between the WBSIP procedure and the BSOP algorithm considered in the previous section is the use of sample splitting, which allows to remove the dependence on the dimension p in the localization rates for WBSIP. Compare the BSOP rate in (7) to (10) below.

Remark 2.2. Wang and Samworth [42] deploys a combination of the WBS algorithm with sample splitting for the problem of mean change point detection in multivariate settings. Since the authors are concerned with recovering a sparse leading eigenvector in a possibly ultrahigh-dimensional setting, their method is inevitably computationally more expensive, and require appropriate assumptions to yield tight bounds in terms of the sparsity level. This leads to one of the main differences between

Algorithm 2 Principal Component Estimation $PC(\{X_t\}_{t=1}^n, \{(\alpha_m, \beta_m)\}_{m=1}^M)$ **INPUT:** $\{X_t\}_{t=1}^n, \{(\alpha_m, \beta_m)\}_{m=1}^M$ **for** $m = 1, \dots, M$ **do****if** $\beta_m - \alpha_m > 2p \log(n) + 1$ **then**

$$d_m \leftarrow \arg \max_{\lceil \alpha_m + p \log(n) \rceil \leq t \leq \lfloor \beta_m - p \log(n) \rfloor} \|\tilde{S}_t^{\alpha_m, \beta_m}\|_{\text{op}}$$

$$u_m \leftarrow \arg \max_{\|v\|=1} |v^\top \tilde{S}_{d_m}^{\alpha_m, \beta_m} v|$$

else

$$u_m \leftarrow 0$$

end if**end for****OUTPUT:** $\{u_m\}_{m=1}^M$.

our approach and theirs – we do not require the shadow vectors to produce consistent estimators of the eigenspaces related to the population covariance matrices. In particular, our analysis holds without any eigengap assumption.

In order to analyze the performance of the WBSIP procedure, we will impose the following assumption, which is significantly weaker than Assumption 2.

Algorithm 3 Wild Binary Segmentation through Independent Projection. $WBSIP(\{X_t, W_t\}_{t=1}^n, (s, e), \{(\alpha_m, \beta_m)\}_{m=1}^M, \tau)$ **INPUT:** Two independent samples, $\{W_t\}_{t=1}^n$ and $\{X_t\}_{t=1}^n$, and the threshold parameter $\tau > 0$.

$$\{u_m\}_{m=1}^M \leftarrow PC(\{W_t\}_{t=1}^n, \{(\alpha_m, \beta_m)\}_{m=1}^M)$$

for $t \in \{s, \dots, e\}$ **do****for** $m = 1, \dots, M$ **do**

$$Y_t(u_m) \leftarrow (u_m^\top X_t)^2$$

end for**end for****for** $m = 1, \dots, M$ **do**

$$(s_m, e_m) \leftarrow [s, e] \cap [\alpha_m, \beta_m]$$

if $e_m - s_m \geq 2 \log(n) + 1$ **then**

$$b_m \leftarrow \arg \max_{s_m + \log(n) \leq t \leq e_m - \log(n)} |\tilde{Y}_t^{s_m, e_m}(u_m)|$$

▷ Recall Definition 1

$$a_m \leftarrow |\tilde{Y}_{b_m}^{s_m, e_m}(u_m)|$$

else

$$a_m \leftarrow -1$$

end if**end for**

$$m^* \leftarrow \arg \max_{m=1, \dots, M} a_m$$

if $a_{m^*} > \tau$ **then**add b_{m^*} to the set of estimated change points

$$WBSIP(\{X_t, W_t\}_{t=1}^n, (s, b_{m^*}), \{(\alpha_m, \beta_m)\}_{m=1}^M, \tau)$$

$$WBSIP(\{X_t, W_t\}_{t=1}^n, (b_{m^*} + 1, e), \{(\alpha_m, \beta_m)\}_{m=1}^M, \tau)$$

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

Assumption 3. For any $\xi > 0$, there exists a sufficiently large absolute constant $C > 0$ such that

$$\Delta \kappa^2 \geq Cp \log^{1+\xi}(n) B^4.$$

Remark 2.3. We recall that all the parameters Δ, κ, p and B are allowed to depend on n . Since $\kappa \leq B^2$, and assuming without loss of generality that the constant C in the previous assumption is larger than 8, we further have that

$$p \log^{1+\xi}(n) \leq \Delta \kappa^2 B^{-4} C^{-1} \leq \Delta/8,$$

which is used repeatedly below. In fact, in the proof we will set $C = 32\sqrt{2}$, a choice born out of convenience. Finally, the quantity $\log^\xi(n)$ may be replaced by any diverging sequence, such as the sequence $\{a_n\}_{n=1,2,\dots}$ in Assumption 2.

It is similar to the BSOP algorithm that WBSIP also applies a slight modification to the WBS algorithm as originally proposed in Fryzlewicz [19]. When computing the shadow vectors in Algorithm 2, the search for the optimal direction onto which projecting the data is restricted, for any given candidate interval, only to the time points that are at least $p \log(n)$ away from the endpoints of the interval. As remarked in the previous section, this ensures good tail bounds on the operator norms of the matrices involved. We also remark that in Algorithm 3 the WBS procedure could be replaced with the penalized least squares procedure analyzed in Wang, Yu and Rinaldo [40] for mean change point localization in univariate time series, which can be implemented efficiently with dynamic programming using the PELT algorithm of Killick, Fearnhead and Eckley [26]. With an appropriate choice of penalty parameter for PELT, this modification will lead to the same localization rates.

Theorem 2.2 (Consistency of WBSIP). *Let Assumptions 1 and 3 hold and let*

$$\{(\alpha_m, \beta_m)\}_{m=1}^M \subset (0, n)$$

be a collection of intervals whose endpoints are drawn independently and uniformly from $\{1, \dots, n\}$ and such that $\max_{1 \leq m \leq M} (\beta_m - \alpha_m) \leq C \Delta$ for an absolute constant $C > 0$. Set

$$\epsilon_k = C_1 B^4 \log(n) \kappa_k^{-2}, \tag{10}$$

for an absolute constant $C_1 > 0$. Suppose there exist sufficiently small constant $c_2 > 0$ and sufficiently large constant $c_3 > 0$ such that the input parameter τ satisfy

$$c_3 B^2 \sqrt{\log(n)} < \tau < c_2 \kappa \sqrt{\Delta}. \tag{11}$$

Then the collection of the estimated change points $\{\hat{\eta}_k\}_{k=1}^{\hat{K}}$ returned by WBSIP with input parameters of $(0, n)$, $\{(\alpha_m, \beta_m)\}_{m=1}^M$ and τ satisfies

$$\begin{aligned} & \mathbb{P}\{\hat{K} = K; |\eta_k - \hat{\eta}_k| \leq \epsilon_k, \text{ for all } k\} \\ & \geq 1 - 4Mn^{2-c} - 4 \times 9^p n^{3-cp} - \exp\{\log(n/\Delta) - M\Delta/(4Cn)\}, \end{aligned} \tag{12}$$

for some absolute constant $c > 3$.

Remark 2.4 (On the constants in Theorem 2.2). The choice of the constant C is essentially arbitrary but will affect the choice of the constants C_1 , which is linear in C . In addition, the constant C appears

in the last term in (12). A large C results in a smaller probability bound and a larger localization error. This dependence can be tracked in the proof.

The above theorem yields *local rates* $\epsilon_1, \dots, \epsilon_K$, one for each change point, which naturally depend on the magnitudes κ_k 's of the changes. Since $\min_k \kappa_k = \kappa > 0$ by assumption, we can conclude that the WBSIP algorithm, with appropriate inputs, produces an overall vanishing localization rate. Indeed,

$$\frac{\epsilon}{n} \leq \frac{\epsilon}{\Delta} \lesssim \frac{B^4}{\kappa^2 \Delta} \log(n) \lesssim \frac{1}{p \log^\xi(n)} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

In fact, the above display also shows that the WBSIP algorithm achieves consistency in a stronger sense: with probability tending to 1, the estimated change points will be away from the corresponding true change points by an amount that is vanishing in the minimal spacing between change points.

The upper and lower bounds on τ in (11) correspond to the minimal expected magnitude of the CUSUM statistics at the true change points and to a high probability bound on the order of their largest sample fluctuation around their means across the whole time course, respectively. In particular, the lower bound corresponds to the quantity λ_2 defined in the proofs (see the [Appendix](#)).

The fact that the dimension p does not appear explicitly in the localization rates (10) is an interesting, if not perhaps surprising, finding. Of course, the dimension does affect (negatively) the performance of the algorithm through Assumption 3: keeping n and Δ fixed, a larger value of p implies a larger value of $B^4 \kappa^{-2}$ in order for that assumption to hold. In turn, this leads to a larger bound in Theorem 2.2. Furthermore, the dimension p appears in the probability of the event that WBSIP fails to locate all the change points. We remark that, for a different problem of high-dimensional mean change point detection, Wang and Samworth [42] also obtained a localization rate independent of the dimension: see Theorem 3 there. In Section 3 below, we will prove that Assumption 3 is in fact essentially necessary for any algorithm to produce a localization rate of smaller order than n .

Finally, in Theorem 2.2 it is necessary to choose a large enough number of random intervals M to obtain high-probability guarantees. In particular, the probabilistic bound (12) shows that $M \gtrsim n \log(n) \Delta^{-1}$.

3. Minimax lower bounds and the phase transition effect

In Theorem 2.2 above we have shown that, if the distribution of the data $\{X_t\}_{t=1}^n$ follows the model described in Assumption 1, then, under the scaling $\Delta \geq C B^4 \kappa^{-2} p \log^{1+\xi}(n)$ for sufficiently large C as given in Assumption 3, the WBSIP algorithm can, with high probability, estimate all the change points with a localization error of the order $\frac{B^4}{\kappa^2} \log(n)$. Assumption 3 might seem arbitrary at first glance. In fact, we will show next that consistent estimation of the locations of the change points requires $\frac{\Delta \kappa^2}{p B^4}$ to diverge, as required in Assumption 3. Therefore, we may conclude that the WBSIP algorithm guarantees consistent localization under nearly all the parameter scalings for which such a task is feasible, save for a $\log^\xi(n)$ term.

To that effect, we will consider the following class of data generating distribution. For an integer $\Delta \in (1, n)$ and numbers $\kappa > 0$ and $\sigma > 0$, let $\mathcal{P}_{\kappa, \Delta, \sigma, p, n}$ denote the class of all joint distributions of n independent vectors (X_1, \dots, X_n) in \mathbb{R}^p such that

$$X_1, \dots, X_\Delta \stackrel{\text{i.i.d.}}{\sim} N_p(0, \Sigma_1) \quad \text{and} \quad X_{\Delta+1}, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} N_p(0, \Sigma_2),$$

where Σ_1 and Σ_2 are positive definite matrices such that

$$\|\Sigma_1 - \Sigma_2\|_{\text{op}} \geq \kappa \quad \text{and} \quad \max_{j=1,2} \|\Sigma_j\|_{\text{op}} \leq \sigma^2.$$

Notice that each distribution in $\mathcal{P}_{\kappa, \Delta, \sigma, p, n}$ satisfies Assumption 1 with $B = 8\sigma$.

Lemma 3.1. *Consider the class of distributions*

$$\mathcal{P} = \mathcal{P}(n) = \left\{ \mathcal{P}_{\kappa, \Delta, \sigma, p, n} : \Delta \leq \min \left\{ \frac{2\sigma^4 p}{33\kappa^2}, n/3 \right\}, \kappa \leq \sigma^2/4 \right\}.$$

Then,

$$\inf_{\hat{\eta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P (|\hat{\eta} - \eta|) \geq n/6,$$

where the infimum is over all estimators of the change point.

The proof of the previous result is based on the construction used in Cai and Ma [11] to obtain minimax lower bounds for a class of hypothesis testing problems involving covariance matrices. As in other results in the paper, the choice of the constant $2/33$ in the definition of the class \mathcal{P} is made out of convenience.

Lemma 3.1 and Theorem 2.2 together imply that the solution to the covariance change point localization problem undergoes a phase transition in the space of model parameters, which we are able to characterize up to a poly-logarithmic factor in n . Specifically,

- under the scaling $\Delta \gtrsim B^4 p \log^{1+\xi}(n)/\kappa^2$, it is possible to estimate the locations of the change points with a localization rate of smaller order of n ;
- on the other hand, if $\Delta \lesssim B^4 p/\kappa^2$, then the localization rate of any algorithm is, in the worst case, of order n .

In our final result, we prove that the upper bound on the localization error that we have obtained for the WBSIP algorithm, which is of order $B^4 \log(n)\kappa^{-2}$ (see equation (10)), is, up to a logarithmic factor in n , minimax optimal over the set of distributions satisfying Assumption 3. Thus, the WBSIP procedure is essentially minimax rate-optimal across almost all the parameter scalings for which consistent localization is feasible. For the lower bound construction, we will consider the same Gaussian setting introduced above, with one change point.

Lemma 3.2. *Consider the class of distributions*

$$\mathcal{Q} = \mathcal{Q}(n) = \left\{ \mathcal{P}_{\kappa, \Delta, \sigma, p, n} : \Delta \kappa^2 \geq p \log(n)\sigma^4, \kappa \leq \sigma^2/4, 4 \leq \Delta \leq 4/5(n-1) \right\}.$$

Then,

$$\inf_{\hat{\eta}} \sup_{P \in \mathcal{Q}} \mathbb{E}_P (|\hat{\eta} - \eta|) \geq \frac{\sigma^4}{20\kappa^2},$$

where the infimum is over all estimators of the change point.

Table 1. A summary of main results (all in terms of rates)

	Localization error	Conditions	Result
BSOP	$B^2\kappa^{-1}n^{5/2}\Delta^{-2}\sqrt{p\log(n)}$	$\Delta\kappa B^{-2} \gtrsim n^\Theta a_n$ $p \lesssim n^{8\Theta-7}/\log(n)$	Theorem 2.1
WBSIP	$B^4 \log(n)\kappa^{-2}$	$\Delta\kappa^2 \gtrsim \log^{1+\xi}(n)pB^4$	Theorem 2.2
Information theoretic lower bounds	$B^4\kappa^{-2}$	$\Delta\kappa^2 \gtrsim pB^4$	Lemmas 3.1 and 3.2

4. Discussion

In this paper, we have studied the problem of change point localization in a time series of length n of independent p -dimensional random vectors with covariance matrices that are piecewise constant. We allow all the parameters quantifying the difficulty of the problem, namely the dimension p , the minimal spacing Δ , the minimal jump size κ , and the sub-Gaussian variance factor B , to change with the sample size n . We have proposed two procedures based on existing algorithms for change point detection – binary segmentation and wild binary segmentation – both yielding consistent localization. In particular the algorithm WBSIP, which applies wild binary segmentation to carefully chosen univariate projections of the data, produces a localization rate that is, up to a logarithmic factor, minimax optimal. A summary of the main results is collected in Table 1.

We have proved that it is necessary that the dimension p should be of smaller order than n in order for the localization rates of any procedure to grow slower than the length of the time series. One possible future direction is to consider different high dimensional settings whereby p is permitted to grow even faster than n , with additional structural assumptions on the underlying covariance matrices, such as sparsity or being low rank. For instance, we may model the covariance matrices as spiked matrices with sparse leading eigenvectors. Another interesting extension is to apply the entry-wise maximum norm instead of the operator norm to the covariance CUSUM statistics. If the changes are still characterized in the operator norms, then this modification requires a more careful handling and potentially additional assumptions. A second extension of interest would be to replace the operator norm of the difference between consecutive covariance matrices at the change point with the Frobenius norm, in order to capture higher-order spectral changes. This modification would pose non-trivial technical challenges but would allow for more general change point models.

Appendix: Main proofs of Theorems 2.1 and 2.2

In this section, we collect the main proofs of Theorems 2.1 and 2.2. We remark that all the numberings starting with S are the contents in the supplemental article Wang, Yu and Rinaldo [41].

Proof of Theorem 2.1. By induction, it suffices to consider any pair of integers s and e such that $(s, e) \subset (0, n)$ and satisfying

$$\eta_{r-1} \leq s \leq \eta_r \leq \dots \leq \eta_{r+q} \leq e \leq \eta_{r+q+1}, \quad q \geq -1,$$

$$\max\{\min\{\eta_r - s, s - \eta_{r-1}\}, \min\{\eta_{r+q+1} - e, e - \eta_{r+q}\}\} \leq \epsilon_n,$$

where $q = -1$ indicates that there is no change point contained in (s, e) . It follows that, for sufficiently small $c_\alpha > 0$ and sufficiently large $C_\alpha > 0$,

$$\begin{aligned} \frac{\epsilon_n}{\Delta/4} &\leq \frac{C_2 B^2 \kappa^{-1} n^{5/2} \sqrt{p \log(n)} \Delta^{-2}}{\Delta/4} \\ &\leq 4C_2 B^2 \kappa^{-1} n^{5/2} \frac{c_\alpha^{1/2} n^{4\Theta-7/2}}{C_\alpha^3 \kappa^{-3} B^6 n^{3\Theta}} \\ &\leq (4C_2 c_\alpha^{1/2} C_\alpha^{-3}) (\kappa^2 B^{-4}) n^{\Theta-1} \\ &\leq (\kappa^2 B^{-4}) n^{\Theta-1} \\ &\leq 1, \end{aligned}$$

where the second inequality stems from Assumption 2, the third inequality holds by choosing sufficiently small c_α and sufficiently large C_α and the last inequality follows from the fact that $\kappa \leq B^2$. Then, for any change point η_j in (s, e) , it is either the case that

$$|\eta_j - s| \leq \epsilon_n,$$

or that

$$|\eta_j - s| \geq \Delta - \epsilon_n \geq \Delta - \Delta/4 = 3\Delta/4.$$

Similar considerations apply to the other endpoint e . As a consequence, the fact that $\min\{|\eta_j - e|, |\eta_j - s|\} \leq \epsilon_n$ implies that η_j is a detected change point found in the previous induction step, while if $\min\{|\eta_j - s|, |\eta_j - e|\} \geq 3\Delta/4$ we can conclude that $\eta_j \in (s, e)$ is an undetected change point.

In order to complete the induction step, it suffices to show that BSOP($\{X_t\}_{s+1}^e, \tau$) (i) will not find any new change point in the interval (s, e) if there is none, or if all the change points in (s, e) have been already detected and (ii) will identify a location b such that $|\eta_j - b| \leq \epsilon_n$ if there exists at least one undetected change point $\eta_j \in (s, e)$.

Set $\lambda = B^2 \sqrt{p \log(n)}$. Then, the event $\mathcal{A}_1(\{X_i\}_{i=1}^n, \lambda)$ holds with probability at least $1 - 2 \times 9^p n^3 n^{-cp}$, for some universal constant $c > 0$. The proof will be completed in two steps.

Step 1. First, we will show that on the event $\mathcal{A}_1(\{X_i\}_{i=1}^n, \lambda)$, BSOP($\{X_t\}_{s+1}^e, \tau$) can consistently detect or reject the existence of undetected change points within (s, e) .

Suppose there exists $\eta_j \in (s, e)$ such that $\min\{\eta_j - s, \eta_j - e\} \geq 3\Delta/4$. Set $\delta = p \log(n)$. Then $\delta \leq \frac{3}{32} \Delta$, since

$$p \log(n) \leq c_\alpha n^{8\Theta-7} \leq c_\alpha n^\Theta \leq c_\alpha C_\alpha^{-1} \Delta B^{-2} \kappa^1 \leq 3\Delta/32,$$

where the last inequality follows from Assumption 2. With this choice of δ , we apply Lemma S.4.5 in Section S.4 (where we set $c_1 = 3/4$) and obtain that

$$\max_{t=\lceil s+\delta \rceil, \dots, \lfloor e-\delta \rfloor} \|\tilde{\Sigma}_t^{s,e}\|_{\text{op}} \geq (3/8)\kappa \Delta (e-s)^{-1/2}.$$

On the event $\mathcal{A}_1(\{X_i\}_{i=1}^n, \lambda)$,

$$\begin{aligned} \max_{t=\lceil s+\delta \rceil, \dots, \lfloor e-\delta \rfloor} \|\tilde{\mathcal{S}}_t^{s,e}\|_{\text{op}} &\geq \max_{t=\lceil s+\delta \rceil, \dots, \lfloor e-\delta \rfloor} \|\tilde{\Sigma}_t^{s,e}\|_{\text{op}} - \lambda \\ &\geq (3/8)\kappa \Delta (e-s)^{-1/2} - \lambda \geq (1/8)\kappa \Delta (e-s)^{-1/2}, \end{aligned} \tag{13}$$

where the last inequality follows from (8) (in the last step we have set $C_1 = 1/8$). If (5) holds, then, on the event $\mathcal{A}_1(\{X_i\}_{i=1}^n, \lambda)$, BSOP($\{X_t\}_{s+1}^e, \tau$) detects the existence of undetected change points if there are any.

Next, suppose there does not exist any undetected change point within (s, e) . Then, one of the following cases must occur.

- (a) There is no change point within (s, e) ;
- (b) there exists only one change point η_r within (s, e) and $\min\{\eta_r - s, e - \eta_r\} \leq \epsilon_n$;
- (c) there exist two change points η_r, η_{r+1} within (s, e) and that $\max\{\eta_r - s, e - \eta_{r+1}\} \leq \epsilon_n$.

Observe that if case (a) holds, then on the event $\mathcal{A}_1(\{X_i\}_{i=1}^n, \lambda)$, we have that

$$\max_{t=\lceil s+\delta \rceil, \dots, \lfloor e-\delta \rfloor} \|\tilde{S}_t^{s,e}\|_{\text{op}} \leq \max_{t=\lceil s+\delta \rceil, \dots, \lfloor e-\delta \rfloor} \|\tilde{\Sigma}_t^{s,e}\|_{\text{op}} + \lambda = \lambda < \tau,$$

where the last inequality follows from (5). If situation (c) holds, then on the event $\mathcal{A}_1(\{X_i\}_{i=1}^n, \lambda)$, we have

$$\begin{aligned} \max_{t=\lceil s+\delta \rceil, \dots, \lfloor e-\delta \rfloor} \|\tilde{S}_t^{s,e}\|_{\text{op}} &\leq \max_{t=\lceil s+\delta \rceil, \dots, \lfloor e-\delta \rfloor} \|\tilde{\Sigma}_t^{s,e}\|_{\text{op}} + \lambda \\ &\leq \max\{\|\tilde{\Sigma}_{\eta_r}^{s,e}\|_{\text{op}}, \|\tilde{\Sigma}_{\eta_{r+1}}^{s,e}\|_{\text{op}}\} + \lambda \leq 2\sqrt{\epsilon_n}B^2 + \lambda < \tau, \end{aligned}$$

where the first inequality follows from $\mathcal{A}_1(\{X_i\}_{i=1}^n, \lambda)$, the second inequality from Lemma S.4.4, the third inequality from Lemma S.4.8 and the last inequality follows from (5). (Both Lemmas are in Section S.4.2.) Case (b) can be handled in a similar manner. Thus, if (5) holds, then on the event $\mathcal{A}_1(\{X_i\}_{i=1}^n, \lambda)$, BSOP($\{X_t\}_{s+1}^e, \tau$) has no false positives when there are no undetected change points in (s, e) .

Step 2. Assume now that there exists a change point $\eta_j \in (s, e)$ such that $\min\{\eta_j - s, \eta_j - e\} \geq 3\Delta/4$ and let

$$b \in \arg \max_{t=\lceil s+\delta \rceil, \dots, \lfloor e-\delta \rfloor} \|\tilde{S}_t^{s,e}\|_{\text{op}}.$$

To complete the proof it suffices to show that $|b - \eta_j| \leq \epsilon_n$.

Let v be such that

$$v \in \arg \max_{\|u\|=1} |u^\top \tilde{S}_b^{s,e} u|.$$

Consider the univariate time series $\{Y_i(v)\}_{i=1}^n$ and $\{f_i(v)\}_{i=1}^n$ defined in (S.42) and (S.43) in Section S.4.2. By Lemma S.4.6, $b \in \arg \max_{s \leq t \leq e} |\tilde{Y}_t(v)|$. Next, we wish to apply Corollary S.2.1 to the time series $\{Y_i(v)\}_{i=s}^e$ and $\{f_i(v)\}_{i=s}^e$. Towards that end, we first need to ensure that the conditions required for that result to hold are verified. (Notice that in the statement of Corollary S.2.1, the f_i 's are assumed to be uniformly bounded by B_1 , while in this proof the $f_i(v)$'s defined in (S.43) are assumed to be bounded by $2B^2$.) First, the collection of the change points of the time series $\{f_i(v)\}_{i=s+1}^e$ is a subset of $\{\eta_k\}_{k=0}^{K+1} \cap (s, e)$. The condition (S.11) and the inequality $2\sqrt{\delta}B^2 \leq (3c_1/4)\kappa\Delta(e-s)^{-1/2}$ are straightforward consequences of Assumption 2, while (S.19) follows from the fact that

$$|\tilde{f}_t^{s,e}(v) - \tilde{Y}_t^{s,e}(v)| \leq \|\tilde{S}_t^{s,e} - \tilde{\Sigma}_t^{s,e}\|_{\text{op}} \leq \lambda.$$

Similarly, (S.18) stems from the relations

$$\max_{t=\lceil s+\delta \rceil, \dots, \lfloor e-\delta \rfloor} |\tilde{Y}_t^{s,e}(v)| = \max_{t=\lceil s+\delta \rceil, \dots, \lfloor e-\delta \rfloor} \|\tilde{S}_t^{s,e}\|_{\text{op}}$$

$$\geq \max_{t=\lceil s+\delta \rceil, \dots, \lfloor e-\delta \rfloor} \|\tilde{\Sigma}_t^{s,e}\|_{op} - \lambda \geq (1/8)\kappa\Delta(e-s)^{-1/2},$$

where the first inequality holds on the event $\mathcal{A}_1(\{X_i\}_{i=1}^n, \lambda)$ and the second inequality is due to (13) and Assumption 2. Thus, all the assumptions of Corollary S.2.1 are met. An application of that result yields that there exists η_k , a change point of $\{f_i(v)\}_{i=s}^e$ satisfying (S.14), such that

$$|b - \eta_k| \leq C_2\lambda(e-s)^{5/2}\Delta^{-2}\kappa^{-1} \leq \epsilon_n.$$

The proof is complete by observing that (S.18) implies $\min\{\eta_j - s, \eta_j - e\} \geq 3\Delta/4$, as discussed in the argument before *Step 1*. □

Proof of Theorem 2.2. By induction, it suffices to consider any generic $(s, e) \subset (0, n)$ that satisfies

$$\eta_{r-1} \leq s \leq \eta_r \leq \dots \leq \eta_{r+q} \leq e \leq \eta_{r+q+1}, \quad q \geq -1,$$

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

$$\begin{aligned} &\text{either } \eta_r - s \leq \epsilon_r \text{ or } s - \eta_{r-1} \leq \epsilon_{r-1}, \\ &\text{either } \eta_{r+q+1} - e \leq \epsilon_{r+q+1} \text{ or } e - \eta_{r+q} \leq \epsilon_{r+q}. \end{aligned}$$

Note that under Assumption 3, $\epsilon_k \leq \Delta/8$; it, therefore, has to be the case that for any change point $\eta_p \in (0, n)$, either $|\eta_p - s| \leq \epsilon_p$ or $|\eta_p - s| \geq \Delta - \epsilon_p \geq 3\Delta/4$. This means that $\min\{|\eta_p - e|, |\eta_p - s|\} \leq \epsilon_p$ indicates that η_p is a detected change point in the previous induction step, even if $\eta_p \in (s, e)$. We refer to $\eta_p \in [s, e]$ as an undetected change point if $\min\{\eta_p - s, \eta_p - e\} \geq 3\Delta/4$.

In order to complete the induction step, it suffices to show that WBSIP (i) will not detect any new change point in (s, e) if all the change points in that interval have been previously detected, and (ii) will find a point b in (s, e) (in fact, in $(s + \log(n), e - \log(n))$) such that $|\eta_p - b| \leq \epsilon_k$ if there exists at least one undetected change point in (s, e) . Let

$$\{u_m\}_{m=1}^M = PC(\{W_i\}_{i=1}^n, \{(\alpha_m, \beta_m)\}_{m=1}^M).$$

Since the intervals $\{(\alpha_m, \beta_m)\}_{m=1}^M$ are generated independently from $\{X_i\}_{i=1}^n \cup \{W_i\}_{i=1}^n$, the rest of the argument is made on the event \mathcal{M} , which is defined in Equation (S.5) of Section S.1, and which has no effects on the distribution of $\{X_i\}_{i=1}^n \cup \{W_i\}_{i=1}^n$.

Step 1. Let $\lambda_1 = B^2\sqrt{p \log(n)}$. In this step, we are to show that, on the event $\mathcal{A}_1(\{W_i\}_{i=1}^n, \lambda_1)$ and for some $c'_1 > 0$,

$$\sup_{1 \leq m \leq M} |u_m^\top (\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}) u_m| \geq c'_1 \|\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}\|_{op} = c'_1 \kappa_k \quad \text{for every } k = 1, \dots, K+1. \quad (14)$$

On the event \mathcal{M} , for any $\eta_k \in (0, n)$, without loss of generality, there exists $\alpha_k \in [\eta_k - 3\Delta/4, \eta_k - \Delta/2]$ and $\beta_k \in [\eta_k + \Delta/2, \eta_k + 3\Delta/4]$. Thus, $[\alpha_k, \beta_k]$ contains only one change point η_k . Using Lemma S.4.5 in Section S.4 and the inequality $p \log(n) \leq \Delta/8$, we have that

$$\max_{t=\lceil \alpha_k + p \log(n) \rceil, \dots, \lfloor \beta_k - p \log(n) \rfloor} \|\tilde{\Sigma}_t^{\alpha_k, \beta_k}\|_{op} = \|\tilde{\Sigma}_{\eta_k}^{\alpha_k, \beta_k}\|_{op} \geq (1/2)\|\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}\|_{op}\sqrt{\Delta}. \quad (15)$$

Let $b_k \in \arg \max_{t=\lceil \alpha_k + p \log(n) \rceil, \dots, \lfloor \beta_k - p \log(n) \rfloor} \|\tilde{S}_t^{\alpha_k, \beta_k}\|_{op}$, where $\tilde{S}_t^{s,e}$ denote the covariance CUSUM statistics of $\{W_i\}_{i=s+1}^e$ at evaluated t . Since $\|\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}\|_{op} = \kappa_k$, by definition,

$$|u_k^\top \tilde{\Sigma}_{b_k}^{\alpha_k, \beta_k} u_k| \geq |u_k^\top \tilde{S}_{b_k}^{\alpha_k, \beta_k} u_k| - \lambda_1$$

$$\begin{aligned}
 &= \max_{t=\lceil\alpha_k+\delta\rceil, \dots, \lfloor\beta_k-\delta\rfloor} \|\tilde{S}_t^{\alpha_k, \beta_k}\|_{\text{op}} - \lambda_1 \\
 &\geq \max_{t=\lceil\alpha_k+\delta\rceil, \dots, \lfloor\beta_k-\delta\rfloor} \|\tilde{\Sigma}_t^{\alpha_k, \beta_k}\|_{\text{op}} - 2\lambda_1 \\
 &\geq (1/2)\|\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}\|_{\text{op}}\sqrt{\Delta} - 2\lambda_1 \\
 &\geq (1/4)\|\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}\|_{\text{op}}\sqrt{\Delta},
 \end{aligned}$$

where the first and second inequalities hold on the event $\mathcal{A}_1(\{W_i\}_{i=1}^n, \lambda_1)$, the third inequality follows from (15) and the last inequality from Assumption 3. Next, observe that

$$\tilde{\Sigma}_t^{\alpha_k, \beta_k} = \begin{cases} \sqrt{\frac{t - \alpha_k}{(\beta_k - \alpha_k)(\beta_k - t)}}(\beta_k - \eta_k)(\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}), & t \leq \eta_k, \\ \sqrt{\frac{\beta_k - t}{(\beta_k - \alpha_k)(t - \alpha_k)}}(\eta_k - \alpha_k)(\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}), & t \geq \eta_k. \end{cases}$$

Using the above expression, for $b_k \geq \eta_k$, we have that

$$\begin{aligned}
 (1/4)\|\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}\|_{\text{op}}\sqrt{\Delta} &\leq |u_k^\top \tilde{\Sigma}_{b_k}^{\alpha_k, \beta_k} u_k| \\
 &= \sqrt{\frac{\beta_k - b_k}{(\beta_k - \alpha_k)(b_k - \alpha_k)}}(\eta_k - \alpha_k) |u_k^\top (\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}) u_k| \\
 &\leq \sqrt{\frac{(\beta_k - \eta_k)(\eta_k - \alpha_k)}{\beta_k - \alpha_k}} |u_k^\top (\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}) u_k| \\
 &\leq \sqrt{2\Delta} |u_k^\top (\Sigma_{\eta_k} - \Sigma_{\eta_{k-1}}) u_k|.
 \end{aligned}$$

Therefore, (14) holds with $c'_1 = 1/(2\sqrt{2})$. The case of $b_k < \eta_k$ follows from very similar calculations.

Step 2. In this step, we will show that WBSIP will consistently detect or reject the existence of undetected change points within (s, e) , provided that (14) holds and on the two events $\mathcal{B}_1(\{X_i\}_{i=1}^n, \{u_m\}_{m=1}^M, \lambda_2)$, where $\lambda_2 = B^2\sqrt{\log(n)}$, and \mathcal{M} , given in (S.3) and (S.5) in Section S.1, respectively.

Let a_m, b_m and m^* be defined as in WBSIP. Denote

$$Y_i(u_m) = (u_m^\top X_i)^2 \quad \text{and} \quad f_i(u_m) = u_m^\top \Sigma_i u_m.$$

Let $\tilde{Y}_t^{s,e}(u_m)$ and $\tilde{f}_t^{s,e}(u_m)$ be defined as in (S.42) and (S.43) of Section S.4.2, respectively.

Suppose that there exists a change point $\eta_p \in (s, e)$ such that $\min\{\eta_p - s, e - \eta_p\} \geq 3\Delta/4$. Let $p \log(n) \leq 3\Delta/32$. Then, on the event \mathcal{M} , there exists an interval (α_m, β_m) selected by WBSIP such that $\alpha_m \in [\eta_p - 3\Delta/4, \eta_p - \Delta/2]$ and $\beta_m \in [\eta_p + \Delta/2, \eta_p + 3\Delta/4]$.

Denote $[s_m, e_m] = [\alpha_m, \beta_m] \cap [s, e]$ (see details of the WBSIP procedure in Algorithm 3) and we have that $\min\{\eta_p - s_m, e_m - \eta_p\} \geq (1/2)\Delta$. Thus, $[s_m, e_m]$ contains at most one change point of the time series $\{f_i(u_m)\}_{i=1}^n$. A similar calculation as the one shown in the proof of Lemma S.4.5 gives that

$$\max_{\lceil s_m + \log(n) \rceil \leq t \leq \lfloor e_m - \log(n) \rfloor} |\tilde{f}_t^{s_m, e_m}(u_m)| \geq (1/8)\sqrt{\Delta} |u_m^\top (\Sigma_{\eta_p} - \Sigma_{\eta_{p-1}}) u_m|,$$

where $e_m - s_m \leq (3/2)\Delta$ is used in the last inequality. Therefore

$$\begin{aligned} a_m &= \max_{\lceil s_m + \log(n) \rceil \leq t \leq \lfloor e_m - \log(n) \rfloor} |\tilde{Y}_t^{s_m, e_m}(u_m)| \\ &\geq \max_{\lceil s_m + \log(n) \rceil \leq t \leq \lfloor e_m - \log(n) \rfloor} |\tilde{f}_t^{s_m, e_m}(u_m)| - \lambda_2 \\ &\geq (1/8)\sqrt{\Delta} |u_m^\top (\Sigma_{\eta_p} - \Sigma_{\eta_{p-1}}) u_m| - \lambda_2, \end{aligned}$$

where the first inequality holds on the event $\mathcal{B}_1(\{X_i\}_{i=1}^n, \{u_m\}_{m=1}^M, \lambda_2)$. Thus for any undetected change point η_p within (s, e) , it holds that

$$\begin{aligned} a_{m^*} &= \sup_{1 \leq m \leq M} a_m \\ &\geq \sup_{1 \leq m \leq M} (1/8)\sqrt{\Delta} |u_m^\top (\Sigma_p - \Sigma_{p-1}) u_m| - \lambda_2 \geq (c'_1/8)\kappa_p \sqrt{\Delta} - \lambda_2 \\ &\geq (c'_1/16)\kappa_p \sqrt{\Delta}, \end{aligned}$$

where the second inequality follows from (14), and the last inequality from

$$\lambda_2 = B^2 \sqrt{\log(n)} \leq (c'_1/16)\kappa \sqrt{\Delta},$$

by choosing the constant C in Assumption 3 to be at least $4\sqrt{2}$. Therefore,

$$a_{m^*} \geq \kappa_{\max}^{s, e}, \quad (16)$$

where $\kappa_{\max}^{s, e} = \max\{\kappa_p : \min\{\eta_p - s_0, e_0 - \eta_p\} \geq \Delta/16\}$. Then, WBSIP correctly accepts the existence of undetected change points on the events (14),

$$\mathcal{B}_1(\{X_i\}_{i=1}^n, \{u_m\}_{m=1}^M, \lambda_2)$$

and \mathcal{M} .

Suppose there does not exist any undetected change points within (s, e) , then for any $(s_m, e_m) = (\alpha_m, \beta_m) \cap (s, e)$, one of the following situations must hold.

- (a) there is no change point within (s_m, e_m) ;
- (b) there exists only one change point η_r within (s_m, e_m) and $\min\{\eta_r - s_m, e_m - \eta_r\} \leq \epsilon_r$ or
- (c) there exist two change points η_r, η_{r+1} within (s_m, e_m) and $\eta_r - s_m \leq \epsilon_r, e_m - \eta_{r+1} \leq \epsilon_{r+1}$.

Only the calculations of (c) is presented, as the other cases are similar and simple. By Lemma S.4.10,

$$\sup_{s_m \leq t \leq e_m} \|\tilde{\Sigma}_t^{s, e}\|_{\text{op}} \leq \sqrt{\epsilon_r} \kappa_r + \sqrt{\epsilon_{r+1}} \kappa_{r+1} \leq 2\sqrt{C_1 \log(n)} B \leq \lambda_2.$$

Therefore on the event $\mathcal{B}_1(\{X_i\}_{i=1}^n, \{u_m\}_{m=1}^M, \lambda_2)$ given in (S.3), for (s_m, e_m) , satisfying (c), we have

$$\begin{aligned} \max_{\lceil s_m + \log(n) \rceil \leq t \leq \lfloor e_m - \log(n) \rfloor} |\tilde{Y}_t^{s_m, e_m}(u_m)| &\leq \max_{\lceil s_m + \log(n) \rceil \leq t \leq \lfloor e_m - \log(n) \rfloor} |\tilde{f}_t^{s_m, e_m}(u_m)| + \lambda_2 \\ &\leq \sup_{s_m \leq t \leq e_m} \|\tilde{\Sigma}_t^{s_m, e_m}\|_{\text{op}} + \lambda_2 \leq 2\lambda_2. \end{aligned}$$

Therefore if (11) holds, then WBSIP will always correctly reject the existence of undetected change points, on the event $\mathcal{B}_1(\{X_i\}_{i=1}^n, \{u_m\}_{m=1}^M, \lambda_2)$.

Step 3. Assume that there exists a change point $\eta_p \in (s, e)$ such that $\min\{\eta_p - s, \eta_p - e\} \geq 3\Delta/4$. Let a_m, b_m and m^* be defined as in WBSIP($(s, e), \{(\alpha_m, \beta_m)\}_{m=1}^M, \tau$).

To complete the proof it suffices to show that, on the events $\mathcal{B}_1(\{X_i\}_{i=1}^n, \{u_m\}_{m=1}^M, \lambda_2)$ and $\mathcal{B}_2(\{X_i\}_{i=1}^n, \{u_m\}_{m=1}^M, \lambda_2)$ given in (S.3) and (S.4) respectively, there exists a change point $\eta_k \in [s_{m^*}, e_{m^*}]$ such that $\min\{\eta_k - s, \eta_k - e\} \geq 3\Delta/4$ and $|b_{m^*} - \eta_k| \leq \epsilon_n$.

Consider the univariate time series $\{Y_i(u_{m^*})\}_{i=1}^n$ and $\{f_i(u_{m^*})\}_{i=1}^n$ defined in (S.42) and (S.43). Since the collection of the change points of the time series $\{f_i(u_{m^*})\}_{i=s_{m^*}}^{e_{m^*}}$ is a subset of that of $\{\eta_k\}_{k=0}^{K+1} \cap [s, e]$, we may apply Corollary S.2.2 to the time series $\{Y_i(u_{m^*})\}_{i=s_{m^*}}^{e_{m^*}}$ and $\{f_i(u_{m^*})\}_{i=s_{m^*}}^{e_{m^*}}$. To that end, we will need to ensure that the assumptions of Corollary S.2.2 are verified. Let $\delta' = \log(n)$ and $\lambda = \lambda_2$. Observe that (S.33) and (S.34) are straightforward consequences of Assumption 3, (S.31) and (S.32) follow from the definitions of $\mathcal{B}_1(\{X_i\}_{i=1}^n, \{u_m\}_{m=1}^M, \lambda_2)$ and $\mathcal{B}_2(\{X_i\}_{i=1}^n, \{u_m\}_{m=1}^M, \lambda_2)$, and that (S.30) follows from (16).

Thus, all the conditions in Corollary S.2.2 are met, and we therefore conclude that there exists a change point η_k , which is also a change point of $\{f_i(v)\}_{i=s_{m^*}}^{e_{m^*}}$, satisfying

$$\min\{e_{m^*} - \eta_k, \eta_k - s_{m^*}\} > \Delta/4 \tag{17}$$

and

$$|b_{m^*} - \eta_k| \leq \max\{C_3\lambda_2^2\kappa_k^{-2}, \delta'\} \leq \epsilon_n,$$

where the last inequality holds because $\lambda_2^2\kappa_k^{-2} = B^4 \log(n)\kappa_k^{-2} \geq \log(n)$, which is a consequence of the inequality $B^2 \geq \kappa_k$.

The proof is complete with the following two observations: (i) The change points of $\{f_i(u_{m^*})\}_{i=s_{m^*}}^{e_{m^*}}$ belong to $(s, e) \cap \{\eta_k\}_{k=1}^K$; and (ii) Equation (17) and $(s_{m^*}, e_{m^*}) \subset (s, e)$ imply that

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

As discussed in the argument before Step 1, this implies that η_k must be an undetected change point of $\{X_i\}_{i=1}^n$ in the covariance structure. □

Acknowledgements

We are truly grateful to the Editor, the Associate Editor and two referees for making this paper a better one.

Supplementary Material

Supplement to “Optimal covariance change point localization in high dimensions” (DOI: 10.3150/20-BEJ1249SUPP; .pdf). We provide additional technical details.

References

[1] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley Interscience. MR1990662

- [2] Aston, J.A.D. and Kirch, C. (2014). Efficiency of change point tests in high dimensional settings. Preprint. Available at [arXiv:1409.1771](https://arxiv.org/abs/1409.1771).
- [3] Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *Ann. Statist.* **37** 4046–4087. MR2572452 <https://doi.org/10.1214/09-AOS707>
- [4] Avanesov, V. and Buzun, N. (2018). Change-point detection in high-dimensional covariance structure. *Electron. J. Stat.* **12** 3254–3294. MR3861282 <https://doi.org/10.1214/18-EJS1484>
- [5] Baranowski, R., Chen, Y. and Fryzlewicz, P. (2019). Narrowest-over-threshold detection of multiple change points and change-point-like features. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 649–672. MR3961502 <https://doi.org/10.1111/rssb.12322>
- [6] Barigozzi, M., Cho, H. and Fryzlewicz, P. (2018). Simultaneous multiple change-point and factor analysis for high-dimensional time series. *J. Econometrics* **206** 187–225. MR3840788 <https://doi.org/10.1016/j.jeconom.2018.05.003>
- [7] Berkes, I., Gabrys, R., Horváth, L. and Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 927–946. MR2750251 <https://doi.org/10.1111/j.1467-9868.2009.00713.x>
- [8] Birke, M. and Dette, H. (2005). A note on testing the covariance matrix for large dimension. *Statist. Probab. Lett.* **74** 281–289. MR2189467 <https://doi.org/10.1016/j.spl.2005.04.051>
- [9] Boysen, L., Kempe, A., Liebscher, V., Munk, A. and Wittich, O. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *Ann. Statist.* **37** 157–183. MR2488348 <https://doi.org/10.1214/07-AOS558>
- [10] Cai, T., Liu, W. and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618 <https://doi.org/10.1080/01621459.2012.758041>
- [11] Cai, T.T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557 <https://doi.org/10.3150/12-BEJ455>
- [12] Cho, H. (2016). Change-point detection in panel data via double CUSUM statistic. *Electron. J. Stat.* **10** 2000–2038. MR3522667 <https://doi.org/10.1214/16-EJS1155>
- [13] Cho, H. and Fryzlewicz, P. (2012). Multiscale and multilevel technique for consistent segmentation of non-stationary time series. *Statist. Sinica* **22** 207–229. MR2933173 <https://doi.org/10.5705/ss.2009.280>
- [14] Cho, H. and Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 475–507. MR3310536 <https://doi.org/10.1111/rssb.12079>
- [15] Davies, P.L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution. *Ann. Statist.* **29** 1–65. MR1833958 <https://doi.org/10.1214/aos/996986500>
- [16] Davis, R.A., Lee, T.C.M. and Rodriguez-Yam, G.A. (2006). Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* **101** 223–239. MR2268041 <https://doi.org/10.1198/016214505000000745>
- [17] Dette, H., Pan, G. and Yang, Q. (2018). Estimating a change point in a sequence of very high-dimensional covariance matrices. Preprint. Available at [arXiv:1807.10797](https://arxiv.org/abs/1807.10797).
- [18] Frick, K., Munk, A. and Sieling, H. (2014). Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 495–580. MR3210728 <https://doi.org/10.1111/rssb.12047>
- [19] Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42** 2243–2281. MR3269979 <https://doi.org/10.1214/14-AOS1245>
- [20] Gombay, E., Horváth, L. and Hušková, M. (1996). Estimators and tests for change in variances. *Statist. Decisions* **14** 145–159. MR1406521
- [21] Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.* **105** 1480–1493. MR2796565 <https://doi.org/10.1198/jasa.2010.tm09181>
- [22] Horváth, L. and Hušková, M. (2012). Change-point detection in panel data. *J. Time Series Anal.* **33** 631–648. MR2944843 <https://doi.org/10.1111/j.1467-9892.2012.00796.x>
- [23] Inclán, C. and Tiao, G.C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Amer. Statist. Assoc.* **89** 913–923. MR1294735

- [24] Jirak, M. (2015). Uniform change point tests in high dimension. *Ann. Statist.* **43** 2451–2483. MR3405600 <https://doi.org/10.1214/15-AOS1347>
- [25] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 <https://doi.org/10.1214/aos/1009210544>
- [26] Killick, R., Fearnhead, P. and Eckley, I.A. (2012). Optimal detection of changepoints with a linear computational cost. *J. Amer. Statist. Assoc.* **107** 1590–1598. MR3036418 <https://doi.org/10.1080/01621459.2012.737745>
- [27] Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Process. Appl.* **83** 79–102. MR1705601 [https://doi.org/10.1016/S0304-4149\(99\)00023-X](https://doi.org/10.1016/S0304-4149(99)00023-X)
- [28] Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *J. Time Series Anal.* **21** 33–59. MR1766173 <https://doi.org/10.1111/1467-9892.00172>
- [29] Li, H., Guo, Q. and Munk, A. (2019). Multiscale change-point segmentation: Beyond step functions. *Electron. J. Stat.* **13** 3254–3296. MR4010980 <https://doi.org/10.1214/19-ejs1608>
- [30] Li, J. and Chen, S.X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40** 908–940. MR2985938 <https://doi.org/10.1214/12-AOS993>
- [31] Lin, K., Sharpnack, J.L., Rinaldo, A. and Tibshirani, R.J. (2017). A sharp error analysis for the fused lasso, with application to approximate changepoint screening. In *Advances in Neural Information Processing Systems* 30 (I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds.) 6887–6896.
- [32] Olshen, A.B., Venkatraman, E., Lucito, R. and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5** 557–572.
- [33] Page, E.S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. MR0088850 <https://doi.org/10.1093/biomet/41.1-2.100>
- [34] Picard, D. (1985). Testing and estimating change-points in time series. *Adv. in Appl. Probab.* **17** 841–867. MR0809433 <https://doi.org/10.2307/1427090>
- [35] Qian, J. and Jia, J. (2012). On pattern recovery of the fused lasso. Technical Report. Available at [arXiv:1211.5194](https://arxiv.org/abs/1211.5194).
- [36] Rojas, C.R. and Wahlberg, B. (2014). On change point detection using the fused lasso method. Technical Report. Available at [arXiv:1401.5408](https://arxiv.org/abs/1401.5408).
- [37] Venkatraman, E.S. (1992). Consistency results in multiple change-point problems. Ph.D. thesis, Stanford Univ. MR2687536
- [38] Vostrikova, L.J. (1981). Discovery of “discord” in multidimensional random processes. *Dokl. Akad. Nauk SSSR* **259** 270–274. MR0625215
- [39] Wald, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Stat.* **16** 117–186. MR0013275 <https://doi.org/10.1214/aoms/1177731118>
- [40] Wang, D., Yu, Y. and Rinaldo, A. (2020). Univariate mean change point detection: Penalization, CUSUM and optimality. *Electron. J. Stat.* **14** 1917–1961. MR4091859 <https://doi.org/10.1214/20-EJS1710>
- [41] Wang, D., Yu, Y. and Rinaldo, A. (2020). Supplement to “Optimal covariance change point localization in high dimensions.” <https://doi.org/10.3150/20-BEJ1249SUPP>
- [42] Wang, T. and Samworth, R.J. (2018). High dimensional change point estimation via sparse projection. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 57–83. MR3744712 <https://doi.org/10.1111/rssb.12243>
- [43] Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrika* **82** 385–397. MR1354236 <https://doi.org/10.1093/biomet/82.2.385>
- [44] Yao, Y.-C. and Au, S.T. (1989). Least-squares estimation of a step function. *Sankhyā Ser. A* **51** 370–381. MR1175613

Received June 2019 and revised June 2020