Statistica Sinica: Supplement

MOMENT DEVIATION SUBSPACES OF DIMENSION REDUCTION FOR HIGH-DIMENSIONAL DATA WITH CHANGE STRUCTURE

Xuehu Zhu[1], Luoyao Yu[1], Jiaqi Huang[2], Junmin Liu[1] and Lixing Zhu[3,4]

[1] *School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China*

[2] *School of Statistics, Beijing Normal University, Beijing, China*

[3] *Center for Statistics and Data Science, Beijing Normal University, Zhuhai, China*

[4] *Department of Mathematics, Hong Kong Baptist University, Hong Kong, China*

Supplementary Material

The research described herewith was supported by a grant from the National Key R&D Program of China (2022YFA1003803), a grant from the National Social Science Foundation of China (21BTJ048), the Zhongying Young Scholar Program, two grants from the National Scientific Foundation of China (12131006, 62276208) and a grant from the University Grants Council of Hong Kong. As a co-author, Jiaqi Huang made important contributions to the theoretical development during the preparation of this research. The authors thank the editor, the associate editor and two referees for their constructive suggestions that significantly improved an early manuscript.

S1. Central $\kappa-$th moment deviation subspace and part of numerical analysis

S1.1 Central $\kappa-$th moment deviation subspace

In this section, we consider higher moments as we may have an interest in detecting change points in the contemporaneous mean or second- order moment change structures. Assume $X_i = (X_{i1}, \cdots, X_{ip})^\top$, for $i = 1 \cdots, n$, be independent p-dimensional random variable vectors. Define the new high-dimensional variables Z_i based on X_i as:

$$\begin{aligned} Z_i &= (X_{i1}, ..., X_{ip}, X_{i1}^2, X_{i1}X_{i2}, ..., X_{i1}X_{ip}, X_{i2}^2, X_{i2}X_{i3}, ..., X_{i2}X_{ip}, \\ &\quad \cdots X_{i1}^\kappa, X_{i1}^{\kappa-1}X_{i2} \cdots, X_{ip}^\kappa)^\top, \end{aligned} \tag{S1.1}$$

where κ denotes some positive integer. Let p_Z denote the dimension of Z_i. Without loss of generality, assume that the sequence $\{Z_i\}_{i=1}^n$ of all means follows a piecewise constant structure with $K+1$ segments. In other words, there are K change points $1 \leq z_1 < z_2 < ... < z_K \leq n$ such that $E(Z_{z_{k-1}+j}) = \mu_Z^{(k)}$, and $\text{Cov}(Z_{z_{k-1}+j}) = \Sigma_Z^{(k)}$, for $k = 1, \cdots, K+1$ and $1 \leq j \leq z_k - z_{k-1}$ where $z_0 = 0$ and $z_{K+1} = n$.

Definition S1.1. Span$\{\mu_Z^{(k)} - \mu_Z^{(l)},$ for $k, l = 1, \cdots, K+1\}$ is called the central $\kappa-$th moment deviation subspace of the sequence $\{X_i\}_{i=1}^n$ and is written as $S^\kappa_{\{X_i\}_{i=1}^n}$. $q_\kappa = \dim\{S^\kappa_{\{X_i\}_{i=1}^n}\}$ is called the structural dimension of $S^\kappa_{\{X_i\}_{i=1}^n}$.

The following theorem states a similar result as that in Theorem 2.1.

Theorem S1.1. *For any basis matrix $B \in \mathcal{R}^{p_Z \times q_\kappa}$ of $\mathrm{S}^\kappa_{\{X_i\}_{i=1}^n}$, both the sequences of $\{B^\top Z_i\}_{i=1}^n$ and $\{Z_i\}_{i=1}^n$ have the same locations of change points.*

To get a consistent estimator of the basis matrix B about the subspace $S^\kappa_{\{X_i\}_{i=1}^n}$, we also consider the following Mahalanobis matrix of the sequence $\{Z_i\}_{i=1}^n$ as:

$$M_{Z,n} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{i \neq j} (Z_i - Z_j)(Z_i - Z_j)^\top. \tag{S1.2}$$

Compute the expectation of $M_{Z,n}$ to get:

$$\begin{aligned}
E(M_{Z,n}) &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{i \neq j} E\left\{(Z_i - Z_j)(Z_i - Z_j)^\top\right\} \\
&= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{i \neq j} \mathrm{Cov}(Z_i - Z_j) + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{i \neq j} E(Z_i - Z_j) E(Z_i - Z_j)^\top \\
&= \sum_{k=1}^{K+1} \sum_{l \neq k, l \leq K+1} \frac{n_k n_l}{n(n-1)} (\Sigma_Z^{(k)} + \Sigma_Z^{(l)}) + \sum_{k=1}^{K+1} \frac{2n_k(n_k-1)}{n(n-1)} \Sigma_Z^{(k)} \\
&\quad + \sum_{k=1}^{K+1} \sum_{l \neq k, l \leq K+1} \frac{n_k n_l}{n(n-1)} (\mu_Z^{(k)} - \mu_Z^{(l)})(\mu_Z^{(k)} - \mu_Z^{(l)})^\top.
\end{aligned}$$

As $n_k/n \to c_k > 0$, for $k = 1, 2, \cdots, K+1$, and $\sum_{k=1}^{K+1} c_k = 1$, we have

$$\begin{aligned}
E(M_{Z,n}) &\to \sum_{k=1}^{K+1} \sum_{l\neq k, l\leq K+1} c_k c_l (\Sigma_Z^{(k)} + \Sigma_Z^{(l)}) + 2\sum_{k=1}^{K+1} c_k^2 \Sigma_Z^{(k)} \\
&\quad + \sum_{k=1}^{K+1} \sum_{l\neq k, l\leq K+1} c_k c_l (\mu_Z^{(k)} - \mu_Z^{(l)})(\mu_Z^{(k)} - \mu_Z^{(l)})^\top \\
&= 2\sum_{k=1}^{K+1} c_k \Sigma_Z^{(k)} + \sum_{k=1}^{K+1} \sum_{l\neq k, l\leq K+1} c_k c_l (\mu_Z^{(k)} - \mu_Z^{(l)})(\mu_Z^{(k)} - \mu_Z^{(l)})^\top \\
&=: 2\Sigma_{pooled}^Z + \Delta_Z = M_Z.
\end{aligned}$$

Theorem S1.2. *Under the model (S1.1), we have* $\mathrm{Span}(\Delta_Z) = \mathrm{S}^{\kappa}_{\{X_i\}_{i=1}^n}$. *Furthermore, let* $B = (v_1, \cdots, v_{q_\kappa})$ *denote the eigenvectors of* Δ_Z *associated with the nonzero eigenvalues of* Δ_Z*, then* B *is the basis matrix of* $\mathrm{S}^{\kappa}_{\{X_i\}_{i=1}^n}$*, namely,* $\mathrm{Span}(B) = \mathrm{S}^{\kappa}_{\{X_i\}_{i=1}^n}$.

We need to consistently estimate the pooled covariance matrix Σ_{pooled}^Z to estimate the target matrix Δ_Z efficiently. Similarly, we adopt a "divide-and-conquer" strategy. Divide the data into $\tilde{K}$ segments with the subscript as: $\mathcal{S}_m = \{(m-1)\beta_n + 1, \cdots, m\beta_n\}$, for $m = 1, \cdots, \tilde{K} - 1$ and $\mathcal{S}_{\tilde{K}} = \{(\tilde{K}-1)\beta_n + 1, \cdots, n\}$. We calculate the covariance for each segment and then average them to get an estimator of the pooled covariance matrix

Σ^Z_{pooled} as:

$$\Sigma^Z_{pooled,n} = \frac{1}{\tilde{K}} \sum_{m=1}^{\tilde{K}} \hat{\Sigma}_{Zm} \text{ with } \hat{\Sigma}_{Zm} = \frac{1}{\#\{\mathcal{S}_m\} - 1} \sum_{k \in \mathcal{S}_m} (Z_k - \bar{Z}_m)(Z_k - \bar{Z}_m)^T, \tag{S1.3}$$

where $\bar{Z}_m = \frac{1}{\#\{\mathcal{S}_m\}} \sum_{k \in \mathcal{S}_i} Z_k$ and $\#\{\mathcal{S}_m\}$ denotes the cardinality of the set $\mathcal{S}_m$.

Together the formula (S1.2) with (S1.3), we can have an estimator of Δ_Z as:

$$\Delta_{Z,n} = M_{Z,n} - 2\Sigma^Z_{pooled,n}. \tag{S1.4}$$

The basis matrix $B \in \mathcal{R}^{p_Z \times q_\kappa}$ of $S^\kappa_{\{X_i\}^n_{i=1}}$ can be estimated as the eigenvectors $B_n = (\hat{v}_1, \cdots, \hat{v}_{q_\kappa})$ associated with the largest q_κ eigenvalues of $\Delta_{Z,n}$.

Theorem S1.3. *Under the model (S1.1), assume that $X_i - E(X_i)$ are independent random variables, and Assumptions S3.5, S3.6, S3.7 and S3.8 hold. Then,*

$$||\Delta_n - \Delta||_F = O_p\left(\sqrt{\frac{p_Z}{n}}\right) + O_p\left(\frac{\sqrt{p_Z}\beta_n}{n}\right),$$

where $||\cdot||_F$ denotes the Frobenius norm of a matrix. Furthermore, we have

$$||B_n - B||_F = O_p\left(\sqrt{\frac{p_Z}{n}}\right) + O_p\left(\frac{\sqrt{p_Z}\beta_n}{n}\right).$$

Remark S1.1. Theorem S1.1 presents that based on the central 2−moment deviation subspace, it is a direct application to identify the locations of change points in the normal case, namely,

$$X_i \sim N(u_i, \Sigma_i),\ 1 \le i \le n. \tag{S1.5}$$

There are K change points $1 \le z_1 < z_2 < ... < z_K \le n$ such that $E(X_{z_{k-1}+j}) = \mu^{(k)}$, and $\mathrm{Cov}(X_{z_{k-1}+j}) = \Sigma^{(k)}$, for $k = 1, \cdots, K+1$ and $1 \le j \le z_k - z_{k-1}$.

Similarly, we can determine the dimension q_κ of the subspace $S^{\kappa}_{\{X_i\}_{i=1}^n}$ based on TRR in (2.5) of the main body. Then, based on the lower-dimensional sequence $\{B_n^\top Z_i\}_{i=1}^n$, we can estimate the locations $\{\hat{z}_1, \cdots, \hat{z}_{\hat{K}}\}$ and the number $\hat{K}$ by existing methods.

S1.2 Numerical Studies with changes in covariance matrix

Experiment 4: Changes in the covariance matrix. The data are generated from the multivariate normal distributions:

$$G_0 = G_2 = G_4 = N(0_p, \mathrm{I}_{p\times p}), \text{ and } G_1 = G_3 = N(0_p, \Sigma).$$

Consider the following four settings:

- Case 1: $\Sigma = (1-a)\mathrm{I}_{p\times p}$ with $a = 0.3, 0.5$, the change points are located at $100i$ for $i = 1, 2, 3, 4$, respectively;
- Case 2: $\Sigma = (\sigma_{ij})$, where $\sigma_{ij} = I(i = j) + aI(i \neq j)$ with $a = 0.3, 0.5$, the settings of change points are the same as Case 1;
- Case 3: $\Sigma = (\sigma_{ij})$, where $\sigma_{ij} = a^{|i-j|}$ with $a = 0.3, 0.5$, the settings of change points are the same as Case 1;
- Case 4: the setting of $\Sigma = (\sigma_{ij})$ is the same as Case 3, and change points located at 90, 250, 390, and 450.

Here different values of a indicate the magnitudes of changes. We design $p = 5, 10$ associated with $p_Z = 20, 65$, respectively. The results are reported in Table 1. From Table 1, E-Divisive$_{dr}$ performs the best among the competitors. The dimension reduction-based methods perform significantly better than the three methods. The E-Divisive and KCP tend to underestimate the number of change points seriously, but ks-cp3o overestimates it in this experiment. Additionally, the dimension reduction-based methods are relatively robust against the different covariance matrices, while E-Divisive, KCP, and ks-cp3o are very sensitive. Thus, there are significant improvements over the three methods by reducing the dimension.

S2. Genetics data with mean changes

Analyze the array comparative genomic hybridization (aCGH) microarray data set, which was analyzed in Stransky et al. [2006] to detect mean changes in the data structure. The dataset includes 57 individuals with a bladder tumor. We use the processed data in the R package: *ecp* to choose 43 individuals out of the 57 individuals at 2215 different loci's on their genome, namely $p = 43$ and $n = 2215$. This empirical study aims to find unusual chromosomal characteristics.

Because the structural dimension is determined to be $\hat{q} = 1$ via the TR-R criterion, Figure 1 plots the locations of change points using E-Divisive and SBS before and after dimension reduction. We found 34 and 45 change points using E-Divisive$_{dr}$ and SBS$_{dr}$ while E-Divisive, SBS found 97 and 3 changes. It would say that E-Divisive has an overestimation issue, and SBS underestimates the number of change points. It would suggest that the results after dimension reduction are reasonable. Further, when the methods ks-cp3o, Multirank, KCP, Inspect, DCBS and GeomCP are used, we found that they respectively identify 1, 7, 55, 254, 6, and 27 change points. This shows that KCP, and Inspect would overestimate the number of changes. ks-cp3o, Multirank, DCBS would also underestimate the number of change points, GeomCP has a similar result with E-Divisive$_{dr}$ for this data set.

S3. Regularity Conditions and Proofs of the theorems

S3.1 Regularity Conditions

To investigate the asymptotic properties, we list the following assumptions. Let $\epsilon_i = X_i - E(X_i)$, $\tilde{\epsilon}_i = Z_i - E(Z_i)$ and $\tilde{\Sigma}_i = Var(Z_i)$.

Assumption S3.1. $0 < \min_{1\le i\le n} \lambda_{min}(\Sigma_i) \le \max_{1\le i\le n} \lambda_{max}(\Sigma_i) < \infty$.

Assumption S3.2. $0 < \min_{1\le i\le n} \lambda_{min}(Var(\epsilon_i\epsilon_i^\top)) \le \max_{1\le i\le n} \lambda_{max}(Var(\epsilon_i\epsilon_i^\top)) < \infty$.

Assumption S3.3. $0 < \max_{1\le i\le n} \lambda_{max}(E(\epsilon_i\epsilon_i^\top - \Sigma_i)^4) < \infty$.

Assumption S3.4. $0 \le \max_{1\le k\le K+1} \left|\alpha^\top \mu^{(k)}\right| < \infty$ for all $||\alpha|| = 1$.

Assumption S3.5. $0 < \min_{1\le i\le n} \lambda_{min}(\tilde{\Sigma}_i) \le \max_{1\le i\le n} \lambda_{max}(\tilde{\Sigma}_i) < \infty$.

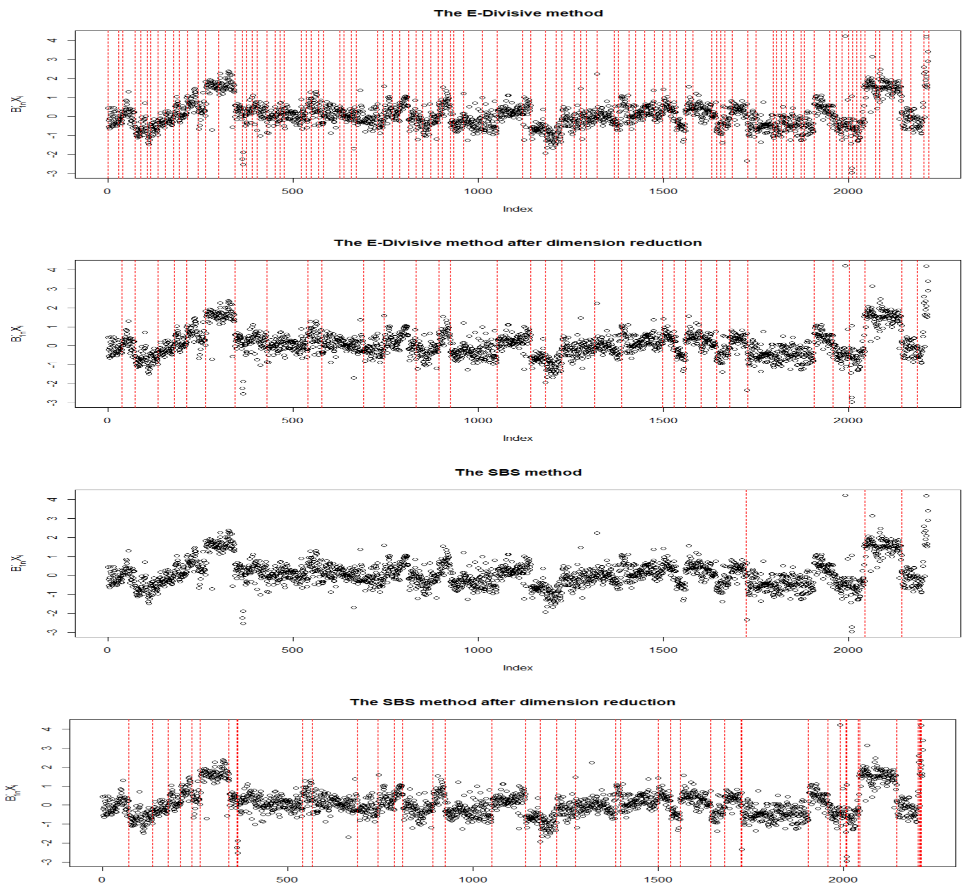

Figure 1: Change point detection for aCGH data, the four figures plot the locations detected by E-Divisive, E-Divisive$_{dr}$, SBS and SBS$_{dr}$, respectively.

Assumption S3.6. $0 < \min_{1\leq i\leq n} \lambda_{min}(Var(\tilde{\epsilon}_i\tilde{\epsilon}_i^\top)) \leq \max_{1\leq i\leq n} \lambda_{max}(Var(\tilde{\epsilon}_i\tilde{\epsilon}_i^\top)) < \infty$.

Assumption S3.7. $0 < \max_{1\leq i\leq n} \lambda_{max}(E(\tilde{\epsilon}_i\tilde{\epsilon}_i^\top - \tilde{\Sigma}_i)^4) < \infty$.

Assumption S3.8. $0 \leq \max_{1\leq k\leq K+1} \left|\alpha^\top \mu_Z^{(i)}\right| < \infty$ for all $||\alpha|| = 1$.

Remark S3.1. These assumptions are satisfied in many situations we are interested in, such as $\Sigma_i = I_p$ and m-dependent cases. See relevant references in the literature such as Dette et al. [2022] and Chen et al. [2010].

S3.2 Appendix. Proofs of the theorems

In this section, we present the proofs of the theoretical results.

Lemma 1. Under the model (2.1), assume that $X_i - E(X_i)$ are independent random variables. Under Assumptions S3.1–S3.4, then

$$\max_{1\le s\le p}\left|\hat{\lambda}_s(\Delta_n) - \lambda_s\right| = O_p\left(\sqrt{\frac{p}{n}}\right) + O_p\left(\frac{\sqrt{p}\beta_n}{n}\right).$$

The proof of this lemma can be similar as that in the proof of Theorem 2.3; we omit it here.

Proof of Theorem 2.1. For any basis matrix $B \in \mathcal{R}^{p\times q}$ of $S_{\{E(X_i)\}_{i=1}^n}$, we have $\text{Span}(B) = S_{\{E(X_i)\}_{i=1}^n}$. Assume there are $\bar{K}$ change points $1 \le \bar{z}_1 < \bar{z}_2 < ... < \bar{z}_{\bar{K}} \le n$ of the sequence $\{Y_i = B^\top X_i\}_{i=1}^n$ such that $E(Y_{\bar{z}_{k-1}+j}) = B^\top \mu_{\bar{z}_{k-1}+j}$, for $k = 1, \cdots, \bar{K}+1$, $1 \le j \le \bar{z}_k - \bar{z}_{k-1}$ and $B^\top \mu_{\bar{z}_k} \ne B^\top \mu_{\bar{z}_k+1}$.

If the locations of change points in the sequence $\{B^\top X_i\}_{i=1}^n$ are not these in the sequence $\{X_i\}_{i=1}^n$, there exist k such that $\mu_{\bar{z}_k} = \mu_{\bar{z}_k+1}$. Then we have $B^\top \mu_{\bar{z}_k} = B^\top \mu_{\bar{z}_k+1}$. However, $\bar{z}_k$ is a change point of the sequence $\{B^\top X_i\}_{i=1}^n$, which implies that $B^\top \mu_{\bar{z}_k} \ne B^\top \mu_{\bar{z}_k+1}$. This is a contradiction. Thus, the locations of change points in the sequence $\{B^\top X_i\}_{i=1}^n$ are those in the sequence $\{X_i\}_{i=1}^n$.

On the other hand, if the locations of change points in the sequence $\{X_i\}_{i=1}^n$ are not those in the sequence $\{B^\top X_i\}_{i=1}^n$, there exists a k such that $B^\top(\mu_{z_k} - \mu_{z_k+1}) = 0$ and $\mu_{z_k} \neq \mu_{z_k+1}$. Therefore, $\mu_{z_k} - \mu_{z_k+1}$ is vertical to the subspace Span(B), namely

$$\mu_{z_k} - \mu_{z_k+1} \perp \mathrm{Span}(B). \tag{S3.6}$$

By the definition of central mean deviation subspace $S_{\{E(X_i)\}_{i=1}^n}$, we have $\mu_{z_k+1} - \mu_{z_k} \in S_{\{E(X_i)\}_{i=1}^n}$. As $\mathrm{Span}(B) = S_{\{E(X_i)\}_{i=1}^n}$, we conclude that

$$\mu_{z_k} - \mu_{z_k+1} \in \mathrm{Span}(B). \tag{S3.7}$$

Altogether the results in (S3.6) and (S3.7), we conclude that $\mu_{z_k} - \mu_{z_k+1} = 0$. This produces the contradiction that z_k is the location of a change point in $\{X_i\}_{i=1}^n$, namely, $\mu_{z_k} \neq \mu_{z_k+1}$. Therefore, the locations of change points in the sequence $\{X_i\}_{i=1}^n$ are these in the sequence $\{B^\top X_i\}_{i=1}^n$. The proof is finished. □

Proof of Theorem 2.2. Recall that

$$\Delta = \sum_{k=1}^{K+1} \sum_{l \neq k, l \leq K+1} c_k c_l (\mu^{(k)} - \mu^{(l)})(\mu^{(k)} - \mu^{(l)})^\top = AA^\top,$$

where $A = (\sqrt{c_1 c_2}\{\mu^{(1)} - \mu^{(2)}\}, \cdots, \sqrt{c_{K+1} c_K}\{\mu^{(K+1)} - \mu^{(K)}\})$, we have $\mathrm{Span}(\Delta) \subseteq \mathrm{Span}(A)$.

As $\mathrm{rank}(A) = \mathrm{rank}(AA^\top)$ and $\mathrm{Span}(AA^\top) = \mathrm{Span}(A)$, we conclude that $\mathrm{Span}(\Delta) = \mathrm{Span}(A)$. By the definition of central mean deviation subspace $S_{\{E(X_i)\}_{i=1}^n}$, $\mathrm{Span}(A) = S_{\{E(X_i)\}_{i=1}^n}$. Therefore, we can get $\mathrm{Span}(\Delta) = S_{\{E(X_i)\}_{i=1}^n}$. □

Proof of Theorem 2.3. By the Taylor expansion, we have (see also, e.g., Sun [1988] that was used in Zhu and Fang [1996])

$$\hat{\lambda}_s(\Delta_n) - \lambda_s(\Delta) = \nu_s^\top (\Delta_n - \Delta)\nu_s + R_{1s}(\Delta^\star)$$

and

$$\hat{\nu}_s(\Delta_n) - \nu_s(\Delta) = \sum_{t=1, t \neq s}^{p} \frac{\nu_t(\Delta)(\nu_s^\top(\Delta)(\Delta_n - \Delta)\nu_s(\Delta))}{\lambda_s(\Delta) - \lambda_t(\Delta)} + R_{2s}(\Delta^\star),$$

where

$$R_{1s}(\Delta^\star) = \nu_s(\Delta^\star)^\top(\Delta_n - \Delta)\nu_s(\Delta^\star) - \nu_s^\top(\Delta)(\Delta_n - \Delta)\nu_s(\Delta),$$
$$R_{2s}(\Delta^\star) = \sum_{t=1,t\neq s}^{p} \frac{\nu_t(\nu_s(\Delta^\star)^\top(\Delta_n - \Delta)\nu_s(\Delta^\star))}{\lambda_s(\Delta) - \lambda_t(\Delta)} - \sum_{t=1,t\neq s}^{p} \frac{\nu_t(\Delta)(\nu_s^\top(\Delta - \Delta)\nu_s)}{\lambda_s(\Delta) - \lambda_t(\Delta)},$$

and $\Delta^\star - \Delta \to 0$ in probability.

Firstly, we know

$$\left|\left|\sum_{t=1,t\neq s}^{p} \frac{\nu_t(\Delta)}{\lambda_s(\Delta) - \lambda_t(\Delta)}\right|\right|_F^2 = O(p).$$

Now we focus on the term $\nu_s^\top(\Delta)(\Delta_n - \Delta)\nu_s(\Delta)$.

Here we give a general conclusion: we consider $\alpha^\top(\Delta_n - \Delta)\alpha$ for any $||\alpha|| = 1$. Recall that

$$\Delta_n - \Delta = M_n - M - 2(\Sigma_{pooled,n} - \Sigma_{pooled}) = M_n - (\Delta - 2\Sigma_{pooled}) - 2(\Sigma_{pooled,n} - \Sigma_{pooled}).$$

Part 1. We calculate the convergence rate of $M_n - \Delta - 2\Sigma_{pooled}$ in the first step. Let $\epsilon_i = X_i - E(X_i)$ and $\delta_{ij} = E(X_i) - E(X_j)$. Then M_n can be decomposed as follows:

$$
\begin{aligned}
M_n =& \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} (X_i - X_j)(X_i - X_j)^\top \\
=& \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} (\epsilon_i - \epsilon_j + \delta_{ij})(\epsilon_i - \epsilon_j + \delta_{ij})^\top \\
=& \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \epsilon_i \epsilon_i^\top + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \epsilon_j \epsilon_j^\top \\
& + \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} (\epsilon_i - \epsilon_j)\delta_{ij}^\top + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \delta_{ij}\delta_{ij}^\top \\
=& \frac{2}{n} \sum_{i=1}^{n} \epsilon_i \epsilon_i^\top + \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} E(X_i)\epsilon_i^\top + \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} E(X_j)\epsilon_j^\top \\
& - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} E(X_i)\epsilon_j^\top - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} E(X_j)\epsilon_i^\top + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \delta_{ij}\delta_{ij}^\top \\
=& \frac{2}{n} \sum_{i=1}^{n} \epsilon_i \epsilon_i^\top + \frac{4}{n} \sum_{i=1}^{n} E(X_i)\epsilon_i^\top - \frac{4}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} E(X_i)\epsilon_j^\top + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \delta_{ij}\delta_{ij}^\top \\
=&: 2M_{1n} + 4M_{2n} - 4M_{3n} + M_{4n}.
\end{aligned}
$$

We now deal with the fourth terms $\alpha^\top M_{in}\alpha$ one by one. For any fixed α with $||\alpha|| = 1$, we have

$$
\alpha^\top M_{1n}\alpha = \frac{1}{n} \sum_{i=1}^{n} \alpha^\top \epsilon_i \epsilon_i^\top \alpha =: \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{ni},
$$

where $Z_{ni} = \frac{1}{\sqrt{n}}\alpha^\top \epsilon_i \epsilon_i^\top \alpha$ is a double array sequence. Then we have

$$E(Z_{ni}) = \frac{1}{\sqrt{n}}\alpha^\top \Sigma_i \alpha, \quad Var\left(\sum_{i=1}^n Z_{ni}\right) = \frac{1}{n}\sum_{i=1}^n Var(\alpha^\top \epsilon_i \epsilon_i^\top \alpha) =: B^2$$

To verify the Lindeberg condition. For any $\eta > 0$, under Assumptions S3.2 and S3.3, we have

$$\begin{aligned}
&\sum_{i=1}^n \frac{1}{B^2}\int_{|Z_{ni}-E(Z_{ni})|>\eta B} (Z_{ni} - E(Z_{ni}))^2 dF_{ni} \\
=&\frac{1}{B^2}\sum_{i=1}^n E\left[(Z_{ni} - E(Z_{ni}))^2 I(|Z_{ni} - E(Z_{ni})| > \eta B)\right] \\
\leq&\frac{n}{B^2}\max_i E\left[(Z_{ni} - E(Z_{ni}))^4\right]^{1/2}\left[P(|Z_{ni} - E(Z_{ni})| > \eta B)\right]^{1/2} \\
=&\frac{1}{B^2}\max_i E\left[(\alpha^\top \epsilon_i \epsilon_i^\top \alpha - \alpha^\top \Sigma_i \alpha)^4\right]^{1/2}\left[P(|Z_{ni} - E(Z_{ni})| > \eta B)\right]^{1/2} \\
\leq&\max_i \frac{E\left[(\alpha^\top \epsilon_i \epsilon_i^\top \alpha - \alpha^\top \Sigma_i \alpha)^4\right]^{1/2}}{B^2}\left[\frac{Var(Z_{ni})}{\eta^2 B^2}\right]^{1/2} \\
\leq&\frac{\max_i \lambda_{max}^{1/2}(E(\epsilon_i \epsilon_i^\top - \Sigma_i)^4)\max_i \lambda_{max}^{1/2}(Var(\epsilon_i \epsilon_i))}{\left[\min_i \lambda_{min}(Var(\epsilon_i \epsilon_i^\top))\right]^{3/2}\eta\sqrt{n}} \\
=&O\left(\frac{1}{\sqrt{n}}\right) \to 0.
\end{aligned}$$

Thus, $\alpha^\top M_{1n}\alpha - \frac{1}{n}\sum_{i=1}^n \alpha^\top \Sigma_i \alpha \overset{\mathcal{D}}{\to} \mathcal{N}\left(0, \frac{1}{n^2}\sum_{i=1}^n Var(\alpha^\top \epsilon_i \epsilon_i^\top \alpha)\right)$ which implies that $\alpha^\top M_{1n}\alpha - 2\alpha^\top \Sigma_{pooled}\alpha = O_p\left(\frac{1}{\sqrt{n}}\right)$ with $\frac{1}{n}\sum_{i=1}^n \alpha^\top \Sigma_i \alpha \overset{a.s.}{\to} \alpha^\top \Sigma_{pooled}\alpha$ and Assumption S3.1.

Next, consider

$$\alpha^\top M_{2n}\alpha = \frac{1}{n}\sum_{i=1}^{n}\alpha^\top E(X_i)\epsilon_i^\top\alpha.$$

Note that if $\alpha^\top E(X_i) = 0$ for all i, then $\alpha^\top M_{2n}\alpha = 0$. Therefore, by the piecewise constant structure of the sequence $E(X_i)$'s, the number of $\alpha^\top E(X_i) \neq 0$ is not smaller than $\min\{n_i\} + 1$, where $\min\{n_i\}$ is the minimum number of data points between any two true changes. Let $\mathcal{N}_n = \{i : \alpha^\top E(X_i) \neq 0\}$. The cardinality $\#\mathcal{N}_n$ of $\mathcal{N}_n$ satisfies that $\min\{n_i\}+1 \leq \#\mathcal{N}_n \leq n$.

$$\alpha^\top M_{2n}\alpha = \frac{1}{n}\sum_{i=1}^{n}\alpha^\top E(X_i)\epsilon_i^\top\alpha = \frac{\#\mathcal{N}_n}{n}\frac{1}{\#\mathcal{N}_n}\sum_{i\in\mathcal{N}_n}\alpha^\top E(X_i)\epsilon_i^\top\alpha.$$

Thus, for every $\xi > 0$, when we take $M = \left[\frac{\max_{1\leq k\leq K}|\alpha^\top\mu^{(k)}|^2\max_{1\leq i\leq n}\lambda_{max}(\Sigma_i)}{\xi}\right]^{1/2} + 1$, then under Assumptions S3.1 and S3.4,

$$\begin{aligned}\Pr\left(\sqrt{\#\mathcal{N}_n}\left|\frac{1}{\#\mathcal{N}_n}\sum_{i\in\mathcal{N}_n}\alpha^\top E(X_i)\epsilon_i^\top\alpha\right| > M\right) \leq& \frac{\sum_{i\in\mathcal{N}_n}\left[\alpha^\top E(X_i)\right]^2\alpha^\top E(\epsilon_i\epsilon_i^\top)\alpha}{\#\mathcal{N}_n M^2}\\ \leq& \frac{\max_{1\leq k\leq K}|\alpha^\top\mu^{(k)}|^2\max_{1\leq i\leq n}\lambda_{max}(\Sigma_i)}{M^2} < \xi.\end{aligned}$$

Thus, $\frac{1}{\#\mathcal{N}_n}\sum_{i\in\mathcal{N}_n}\alpha^\top E(X_i)\epsilon_i^\top\alpha = O_p\left(\frac{1}{\sqrt{\#\mathcal{N}_n}}\right)$. We could also obtain, since

$\#\mathcal{N}_n \le n$,

$$\alpha^\top M_{2n}\alpha = \frac{\#\mathcal{N}_n}{n}\frac{1}{\#\mathcal{N}_n}\sum_{i\in\mathcal{N}_n}\alpha^\top E(X_i)\epsilon_i^\top\alpha = O_p\left(\frac{\sqrt{\#\mathcal{N}_n}}{n}\right) \le O_p\left(\frac{1}{\sqrt{n}}\right).$$

Consider the term $\alpha^\top M_{3n}\alpha$. Under Assumption S3.4, we have

$$\begin{aligned}
\alpha^\top M_{3n}\alpha =& \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\alpha^\top E(X_i)\epsilon_j^\top\alpha \\
=& \frac{1}{n(n-1)}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha^\top E(X_i)\epsilon_j^\top\alpha\right] - \frac{1}{n(n-1)}\sum_{i=1}^{n}\alpha^\top E(X_i)\epsilon_i^\top\alpha \\
=& \frac{n^2}{n(n-1)}\left[\frac{1}{n}\sum_{i=1}^{n}\alpha^\top E(X_i)\right]\left[\frac{1}{n}\sum_{j=1}^{n}\alpha^\top\epsilon_j\right] - \frac{1}{n-1}\alpha^\top M_{2n}\alpha \\
\le& \frac{n}{n-1}\left|\frac{1}{n}\sum_{i=1}^{n}\alpha^\top E(X_i)\right|\left|\frac{1}{n}\sum_{j=1}^{n}\alpha^\top\epsilon_j\right| + O_p\left(\frac{1}{n\sqrt{n}}\right) \\
\le& \frac{n}{n-1}\left|\max_{1\le k\le K}\alpha^\top\mu^{(k)}\right|\left|\frac{1}{n}\sum_{j=1}^{n}\alpha^\top\epsilon_j\right| + O_p\left(\frac{1}{n\sqrt{n}}\right) \\
=& O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{n\sqrt{n}}\right) \\
=& O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Next, we turn to consider the term $\alpha^\top M_{4n}\alpha$ by rewriting it as

$$\begin{aligned}
\alpha^\top M_{4n}\alpha =& \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\alpha^\top\delta_{ij}\delta_{ij}^\top\alpha \\
=& \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\alpha^\top(E(X_i)-E(X_j))(E(X_i)-E(X_j))^\top\alpha \\
=& \frac{1}{n(n-1)}\sum_{k=1}^{K+1}\sum_{l\neq k, l\leq K+1}\sum_{i=z_{k-1}+1}^{z_k}\sum_{j=z_{l-1}+1}^{z_l}\alpha^\top(\mu^{(k)}-\mu^{(l)})(\mu^{(k)}-\mu^{(l)})^\top\alpha \\
=& \frac{1}{n(n-1)}\sum_{k=1}^{K+1}\sum_{l\neq k, l\leq K+1} n_k n_l \alpha^\top(\mu^{(k)}-\mu^{(l)})(\mu^{(k)}-\mu^{(l)})^\top\alpha \\
=& \alpha^\top\Delta\alpha + o(1).
\end{aligned}$$

To sum up, together with all the results of the four terms $\alpha^\top M_{in}\alpha$'s, we conclude that

$$\alpha^\top(M_n - \Delta - 2\Sigma_{pooled})\alpha \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{4}{n^2}\sum_{i=1}^{n} Var(\alpha^\top\epsilon_i\epsilon_i^\top\alpha)\right)$$

and then

$$\alpha^\top(M_n - \Delta - 2\Sigma_{pooled})\alpha = \alpha^\top(M_n - M)\alpha = O_p\left(\frac{1}{\sqrt{n}}\right)$$

for any fixed α.

Part 2. We decompose $\alpha^\top \left(\Sigma_{pooled,n} - \Sigma_{pooled}\right) \alpha$ as

$$\begin{aligned}
&\alpha^\top \left(\Sigma_{pooled,n} - \Sigma_{pooled}\right) \alpha \\
=&\alpha^\top \left[\frac{1}{\tilde{K}} \sum_{m=1}^{\tilde{K}} (\hat{\Sigma}_m - \Sigma_{pooled})\right] \alpha \\
=&\alpha^\top \left(\frac{1}{\tilde{K}} \sum_{m=1}^{\tilde{K}} \frac{1}{2\beta_n(\beta_n - 1)} \sum_{k\in\mathcal{S}_m} \sum_{l\neq k, l\in\mathcal{S}_m} (X_k - X_l)(X_k - X_l)^\top - \Sigma_{pooled}\right) \alpha \\
=&\alpha^\top \left(\frac{1}{\tilde{K}} \sum_{m=1}^{\tilde{K}} \frac{1}{2\beta_n(\beta_n - 1)} \sum_{k\in\mathcal{S}_m} \sum_{l\neq k, l\in\mathcal{S}_m} (\epsilon_k - \epsilon_l + \delta_{kl})(\epsilon_k - \epsilon_l + \delta_{kl})^\top - \Sigma_{pooled}\right) \alpha \\
=&\frac{1}{\tilde{K}} \sum_{m=1}^{\tilde{K}} \frac{1}{\beta_n} \sum_{k\in\mathcal{S}_m} \left(\alpha^\top \epsilon_m \epsilon_m^\top \alpha - \alpha^\top \Sigma_{pooled} \alpha\right) - \frac{1}{\tilde{K}} \sum_{m=1}^{\tilde{K}} \frac{1}{\beta_n(\beta_n - 1)} \sum_{k\in\mathcal{S}_m} \sum_{l\neq k, l\in\mathcal{S}_m} \alpha^\top \epsilon_k \epsilon_l^\top \alpha \\
&+ \frac{1}{\tilde{K}} \sum_{m=1}^{\tilde{K}} \frac{1}{\beta_n(\beta_n - 1)} \sum_{k\in\mathcal{S}_m} \sum_{l\neq k, l\in\mathcal{S}_m} \alpha^\top \delta_{kl} (\epsilon_k - \epsilon_l)^\top \alpha \\
&+ \frac{1}{\tilde{K}} \sum_{m=1}^{\tilde{K}} \frac{1}{2\beta_n(\beta_n - 1)} \sum_{k\in\mathcal{S}_m} \sum_{l\neq k, l\in\mathcal{S}_m} \alpha^\top \delta_{kl} \delta_{kl}^\top \alpha \\
=:& \Sigma_{1n} - \Sigma_{2n} + \Sigma_{3n} + \Sigma_{4n}.
\end{aligned}$$

It is easy to obtain that

$$\begin{aligned}
\Sigma_{1n} =&\frac{1}{\tilde{K}} \sum_{m=1}^{\tilde{K}} \frac{1}{\beta_n} \sum_{k\in\mathcal{S}_m} \left(\alpha^\top \epsilon_k \epsilon_k^\top \alpha - \alpha^\top \Sigma_{pooled} \alpha\right) = \frac{1}{n} \sum_{i=1}^{n} \left(\alpha^\top \epsilon_i \epsilon_i^\top \alpha - \alpha^\top \Sigma_{pooled} \alpha\right) \\
&\xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{n^2} \sum_{i=1}^{n} Var(\alpha^\top \epsilon_i \epsilon_i^\top \alpha)\right),
\end{aligned}$$

which implies that $\Sigma_{1n} = O_p\left(\frac{1}{\sqrt{n}}\right)$.

For Σ_{2n}, we can rewrite it as

$$\begin{aligned}
\Sigma_{2n} =& \frac{1}{\tilde{K}} \sum_{m=1}^{\tilde{K}} \frac{1}{\beta_n(\beta_n - 1)} \sum_{k \in \mathcal{S}_m} \sum_{l \neq k, l \in \mathcal{S}_m} \alpha^\top \epsilon_k \epsilon_l^\top \alpha \\
=& \frac{1}{\sqrt{\tilde{K}\beta_n(\beta_n - 1)}} \sum_{m=1}^{\tilde{K}} \frac{1}{\sqrt{\tilde{K}}} \frac{1}{\sqrt{\beta_n(\beta_n - 1)}} \sum_{k \in \mathcal{S}_m} \sum_{l \neq k, l \in \mathcal{S}_m} \alpha^\top \epsilon_k \epsilon_l^\top \alpha \\
=& \frac{1}{\sqrt{\tilde{K}\beta_n(\beta_n - 1)}} \sum_{m=1}^{\tilde{K}} T_{\tilde{K}m},
\end{aligned}$$

where $T_{\tilde{K}m} = \frac{1}{\sqrt{\tilde{K}\beta_n(\beta_n-1)}} \sum_{k \in \mathcal{S}_m} \sum_{l \neq k, l \in \mathcal{S}_m} \alpha^\top \epsilon_k \epsilon_l^\top \alpha$. We can also derive that

$$\begin{aligned}
E(T_{\tilde{K}m}) =& \frac{1}{\sqrt{\tilde{K}\beta_n(\beta_n - 1)}} \sum_{k \in \mathcal{S}_m} \sum_{l \neq k, l \in \mathcal{S}_m} E(\alpha^\top \epsilon_k) E(\epsilon_l^\top \alpha) = 0, \\
Var(T_{\tilde{K}m}) =& E(T_{\tilde{K}m}^2) = \left[\frac{1}{\sqrt{\tilde{K}\beta_n(\beta_n - 1)}} \sum_{k \in \mathcal{S}_m} \sum_{l \neq k, l \in \mathcal{S}_m} \alpha^\top \epsilon_k \epsilon_l^\top \alpha \right]^2 \\
=& \frac{1}{\tilde{K}\beta_n(\beta_n - 1)} \sum_{k_1 \in \mathcal{S}_m} \sum_{l_1 \neq k_1, l_1 \in \mathcal{S}_m} \sum_{k_2 \in \mathcal{S}_m} \sum_{l_2 \neq k_2, l_2 \in \mathcal{S}_m} E(\alpha^\top \epsilon_{k_1} \alpha^\top \epsilon_{l_1} \alpha^\top \epsilon_{k_2} \alpha^\top \epsilon_{l_2}) \\
=& \frac{2}{\tilde{K}\beta_n(\beta_n - 1)} \sum_{k \in \mathcal{S}_m} \sum_{l \neq k, l \in \mathcal{S}_m} E[(\alpha^\top \epsilon_k)^2] E[(\alpha^\top \epsilon_l)^2] \\
=& \frac{2}{\tilde{K}\beta_n(\beta_n - 1)} \sum_{k \in \mathcal{S}_m} \sum_{l \neq k, l \in \mathcal{S}_m} \left(\alpha^\top \Sigma_k \alpha\right) \left(\alpha^\top \Sigma_l \alpha\right)
\end{aligned}$$

Thus,under Assumptions S3.1 and S3.2, we could calculate

$$\frac{2}{\tilde{K}}\min_{1\leq i\leq n}\lambda^2_{min}(\Sigma_i)\leq Var(T_{\tilde{K}m})\leq\frac{2}{\tilde{K}}\max_{1\leq i\leq n}\lambda^2_{max}(\Sigma_i)$$

and

$$2\min_{1\leq i\leq n}\lambda^2_{min}(\Sigma_i)\leq Var\left(\sum_{i=1}^{\tilde{K}}T_{\tilde{K}i}\right)\leq 2\max_{1\leq i\leq n}\lambda^2_{max}(\Sigma_i).$$

$$\begin{aligned}
E\left[T^4_{\tilde{K}m}\right]=&\frac{1}{[\tilde{K}\beta_n(\beta_n-1)]^2}E\left\{\sum_{k\in\mathcal{S}_m}\sum_{l\neq k,l\in\mathcal{S}_m}\alpha^\top\epsilon_k\epsilon_l^\top\alpha\right\}^4\\
=&\frac{C_1}{[\tilde{K}\beta_n(\beta_n-1)]^2}\sum_{k\in\mathcal{S}_m}\sum_{l\neq k,l\in\mathcal{S}_m}E[(\alpha^\top\epsilon_k)^4]E[(\epsilon_l^\top\alpha)^4]\\
&+\frac{C_2}{[\tilde{K}\beta_n(\beta_n-1)]^2}\sum_{k_1\in\mathcal{S}_m}\sum_{\substack{l_1\neq k_1\\ l_1\in\mathcal{S}_m}}\sum_{\substack{k_2\neq k_1\neq l_1\\ k_2\in\mathcal{S}_m}}\sum_{\substack{l_2\neq l_1\neq k_1\neq l_1\\ l_2\in\mathcal{S}_m}}\\
&\quad E[(\alpha^\top\epsilon_{k_1})^2]E[(\alpha^\top\epsilon_{l_1})^2]E[(\alpha^\top\epsilon_{k_2})^2]E[(\alpha^\top\epsilon_{l_2})^2]\\
=&O\left(\frac{1}{\tilde{K}^2}\right),
\end{aligned}$$

where C_1 and C_2 are positive integers that don't take a lot of effort to calculate. For any $\eta > 0$, with Assumptions S3.1 and S3.2, we have

$$
\begin{aligned}
&\sum_{m=1}^{\tilde{K}} \frac{1}{Var\left(\sum_{m=1}^{\tilde{K}} T_{\tilde{K}m}\right)} \int_{|T_{\tilde{K}m}|>\sqrt{\eta Var\left(\sum_{m=1}^{\tilde{K}} T_{\tilde{K}m}\right)}} T_{\tilde{K}m}^2 dF'_{\tilde{K}m} \\
=&\frac{1}{Var\left(\sum_{i=1}^{\tilde{K}} T_{\tilde{K}m}\right)} \sum_{m=1}^{\tilde{K}} E\left[T_{\tilde{K}m}^2 I\left(|T_{\tilde{K}m}| > \sqrt{\eta Var\left(\sum_{m=1}^{\tilde{K}} T_{\tilde{K}m}\right)}\right)\right] \\
\le&\frac{1}{Var\left(\sum_{m=1}^{\tilde{K}} T_{\tilde{K}m}\right)} \sum_{m=1}^{\tilde{K}} E\left[T_{\tilde{K}m}^4\right]^{1/2} P^{1/2}\left(|T_{\tilde{K}m}| > \sqrt{\eta Var\left(\sum_{m=1}^{\tilde{K}} T_{\tilde{K}m}\right)}\right) \\
\le&\frac{\tilde{K} \max_{1\le m\le \tilde{K}} E\left[T_{\tilde{K}m}^4\right]^{1/2} \max_{1\le m\le \tilde{K}} Var(T_{\tilde{K}m})^{1/2}}{Var\left(\sum_{m=1}^{\tilde{K}} T_{\tilde{K}m}\right)^{3/2} \sqrt{\eta}} \\
=&O\left(\frac{1}{\sqrt{\eta \tilde{K}}}\right) \to 0.
\end{aligned}
$$

Then

$$
\Sigma_{2n} = O_p\left(\frac{1}{\sqrt{\tilde{K}\beta_n(\beta_n - 1)}}\right) = O_p\left(\frac{1}{\sqrt{n\beta_n}}\right).
$$

Recall the definition of $\mathcal{S}_m$ right above equation (2.3) in the main body of the paper. Write all those sets $\left\{\mathcal{S}_m \text{ for } m = 1, \cdots, \tilde{K}\right\}$. Note that all $\mathcal{S}_m$ for $m = 1, \cdots, \tilde{K}$ are disjoint. Then we further split all sets into two disjoint subsets of sets where $\mathcal{S}^c = \{\mathcal{S}_m,$ where $\mathcal{S}_m$ contains at least an index m such that $z_m \in \mathcal{S}_m$ and $z_m + 1 \in \mathcal{S}_m\}$, and the rest sets as $\mathcal{S}$. The number of the set $\mathcal{S}^c$ is less than or equal to the number of change points K.

Based on this definition, we can write Σ_{3n} as

$$\begin{aligned}
\Sigma_{3n} =&\frac{1}{\tilde{K}}\sum_{m=1}^{\tilde{K}}\frac{1}{\beta_n(\beta_n-1)}\sum_{k\in\mathcal{S}_m}\sum_{l\neq k,l\in\mathcal{S}_m}\alpha^\top\delta_{kl}(\epsilon_k-\epsilon_l)^\top\alpha \\
=&\frac{1}{\tilde{K}}\sum_{\mathcal{S}_m\in\mathcal{S}}\frac{1}{\beta_n(\beta_n-1)}\sum_{k\in\mathcal{S}_m}\sum_{l\neq k,l\in\mathcal{S}_m}\alpha^\top\delta_{kl}(\epsilon_k-\epsilon_l)^\top\alpha \\
&+\frac{1}{\tilde{K}}\sum_{\mathcal{S}_m\in\mathcal{S}^c}\frac{1}{\beta_n(\beta_n-1)}\sum_{k\in\mathcal{S}_m}\sum_{l\neq k,l\in\mathcal{S}_m}\alpha^\top\delta_{kl}(\epsilon_k-\epsilon_l)^\top\alpha \\
=&\frac{1}{\tilde{K}}\sum_{\mathcal{S}_m\in\mathcal{S}^c}\frac{1}{\beta_n(\beta_n-1)}\sum_{k\in\mathcal{S}_m}\sum_{l\neq k,l\in\mathcal{S}_m}\alpha^\top\delta_{kl}(\epsilon_k-\epsilon_l)^\top\alpha \\
=&\frac{1}{\tilde{K}}\sum_{\mathcal{S}_m\in\mathcal{S}^c}\frac{1}{\beta_n(\beta_n-1)}\sum_{k\in\mathcal{S}_m}\sum_{l\neq k,l\in\mathcal{S}_m}\alpha^\top(E(X_k)-E(X_l))(\epsilon_k-\epsilon_l)^\top\alpha \\
=&\frac{2}{\tilde{K}}\sum_{\mathcal{S}_m\in\mathcal{S}^c}\frac{1}{\beta_n}\sum_{k\in\mathcal{S}_m}\alpha^\top E(X_k)\epsilon_k^\top\alpha \\
&+\frac{2}{\tilde{K}}\sum_{\mathcal{S}_m\in\mathcal{S}^c}\frac{1}{\beta_n(\beta_n-1)}\left[\sum_{k\in\mathcal{S}_m}\sum_{l\in\mathcal{S}_m}\alpha^\top E(X_k)\epsilon_l^\top\alpha-\sum_{k\in\mathcal{S}_m}\alpha^\top E(X_k)\epsilon_k^\top\alpha\right] \\
=&\frac{2\#\{\mathcal{S}^c\}}{\tilde{K}}O_p\left(\frac{1}{\sqrt{\#\{\mathcal{S}^c\}\beta_n}}\right)+\frac{2\#\{\mathcal{S}^c\}}{\tilde{K}(\beta_n-1)}O_p\left(\frac{1}{\sqrt{\#\{\mathcal{S}^c\}\beta_n}}\right) \\
&+\frac{2}{\tilde{K}}\sum_{\mathcal{S}_m\in\mathcal{S}^c}\left[\frac{1}{\beta_n}\sum_{k\in\mathcal{S}_m}\alpha^\top E(X_k)\right]\left[\frac{\beta_n}{\beta_n-1}\frac{1}{\beta_n}\sum_{l\in\mathcal{S}_m}\alpha^\top\epsilon_l\right] \\
=&\frac{2\#\{\mathcal{S}^c\}}{\tilde{K}}O_p\left(\frac{1}{\sqrt{\#\{\mathcal{S}^c\}\beta_n}}\right)+\frac{2\#\{\mathcal{S}^c\}}{\tilde{K}(\beta_n-1)}O_p\left(\frac{1}{\sqrt{\#\{\mathcal{S}^c\}\beta_n}}\right) \\
&+O(1)\left[\frac{2}{\tilde{K}}\sum_{\mathcal{S}_m\in\mathcal{S}^c}\frac{\beta_n}{\beta_n-1}\frac{1}{\beta_n}\sum_{l\in\mathcal{S}_m}\alpha^\top\epsilon_l\right] \\
=&O_p\left(\frac{\sqrt{\#\{\mathcal{S}^c\}}}{\tilde{K}\sqrt{\beta_n}}\right)=O_p\left(\frac{\sqrt{K}}{\sqrt{\tilde{K}n}}\right)=O_p\left(\frac{\sqrt{K\beta_n}}{n}\right)=O_p\left(\frac{\sqrt{\beta_n}}{n}\right).
\end{aligned}$$

The last term is discussed under Assumption S3.4.

For Σ_{4n}, we have

$$\Sigma_{4n} = \frac{1}{\tilde{K}} \sum_{\mathcal{S}_m \in \mathcal{S}^c} \frac{1}{2\beta_n(\beta_n - 1)} \sum_{k \in \mathcal{S}_m} \sum_{l \neq k, l \in \mathcal{S}_m} \alpha^\top \delta_{kl} \delta_{kl}^\top \alpha = K O_p\left(\frac{1}{\tilde{K}}\right) = K O_p\left(\frac{\beta_n}{n}\right) = O_p\left(\frac{\beta_n}{n}\right).$$

Therefore, together with the results about Σ_{in} for $i = 1, \cdots, 4$, we conclude that

$$\begin{aligned} \alpha^\top(\Sigma_{pooled,n} - \Sigma_{pooled})\alpha =& O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n\beta_n}}\right) + O_p\left(\frac{\sqrt{\beta_n}}{n}\right) + O_p\left(\frac{\beta_n}{n}\right) \\ =& O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{\beta_n}{n}\right). \end{aligned}$$

To sum up, we conclude that

$$\alpha^\top(\Delta_n - \Delta)\alpha = 3\Sigma_{1n} + o_p(1) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{9}{n^2}\sum_{i=1}^n Var(\alpha^\top \epsilon_i \epsilon_i^\top \alpha)\right)$$

when $\beta_n/n \to 0$. Furthermore, we have

$$\begin{aligned} \alpha^\top(\Delta_n - \Delta)\alpha &= \alpha^\top(M_n - M)\alpha - 2\alpha^\top(\Sigma_{pooled,n} - \Sigma_{pooled})\alpha \\ &= O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{\beta_n}{n}\right). \end{aligned}$$

Therefore, we have, recalling the definitions at the beginning of the proof of this theorem,

$$\hat{\lambda}_s(\Delta_n) - \lambda_s(\Delta) = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{\beta_n}{n}\right), s = 1, 2, \cdots, p,$$

and

$$\hat{\nu}_s(\Delta_n) - \nu_s(\Delta) = O_p\left(\sqrt{\frac{p}{n}}\right) + O_p\left(\frac{\sqrt{p}\beta_n}{n}\right).$$

Then we get

$$\begin{aligned} ||\Delta_n - \Delta||_F =& \sqrt{tr(\Delta_n - \Delta)(\Delta_n - \Delta)^\top} = \sqrt{\sum_{k=1}^{p} \lambda_k^2(\Delta_n - \Delta)} \\ =& O_p\left(\sqrt{\frac{p}{n}}\right) + O_p\left(\frac{\sqrt{p}\beta_n}{n}\right), \end{aligned}$$

where $\lambda_k(\Delta_n - \Delta)$ is the eigenvalue of the matrix $\Delta_n - \Delta$. These results imply that

$$\begin{aligned} \max_{1\le s\le p} \left|\hat{\lambda}_s(\Delta_n) - \lambda_s(\Delta)\right| =& ||\Delta_n - \Delta||_2 \\ \le& ||\Delta_n - \Delta||_F \\ =& O_p\left(\sqrt{\frac{p}{n}}\right) + O_p\left(\frac{\sqrt{p}\beta_n}{n}\right), \end{aligned}$$

where $||\cdot||_2$ is L_2 matrix norms. Thus, we

$$||B_n - B||_F = O_p\left(\sqrt{\frac{pq}{n}}\right) + O_p\left(\frac{\sqrt{pq}\beta_n}{n}\right).$$

The proof is finished. □

Proof of Theorem 2.4. We follow the similar arguments of proving Theorem 2.2 in Zhu et al. [2023] to prove this theorem. Write $\tilde{\eta}_n = \max\left\{\sqrt{\frac{p}{n}}, \frac{\sqrt{p}\beta_n}{n}\right\}$. $\hat{\lambda}_s(\Delta_n)$ and $\lambda_s(\Delta)$ as $\hat{\lambda}_s$ and λ_s in short. From Lemma 1, we can get

$$\max_{1\le s\le p}\left|\hat{\lambda}_s - \lambda_s\right| = O_p(\tilde{\eta}_n).$$

The following deduction is in a sense with a probability tending to 1. The above implies there exists a constant C such that the following inequality holds:

$$\max_{1\le s\le p}\left|\hat{\lambda}_s - \lambda_s\right| \le C\tilde{\eta}_n.$$

Then, we have $\lambda_s - C\tilde{\eta}_n \le \hat{\lambda}_s \le \lambda_s + C\tilde{\eta}_n, \forall 1 \le s \le p$. This implies that

$$-C\tilde{\eta}_n \le \min_{q+1\le s\le p} \hat{\lambda}_s \le \max_{q+1\le s\le p} \hat{\lambda}_s \le C\tilde{\eta}_n.$$

Since $\lambda_q > 0$ and $\lambda_{q+1} = 0$, we can obtain

$$\frac{-C\tilde{\eta}_n + c_n}{\lambda_q + C\tilde{\eta}_n + c_n} \leq \frac{\hat{\lambda}_{(q+1)} + c_n}{\hat{\lambda}_q + c_n} \leq \frac{C\tilde{\eta}_n + c_n}{\lambda_q - C\tilde{\eta}_n + c_n}.$$

Due to the conditions $c_n \to 0$ and $c_n/\tilde{\eta}_n \to \infty$, and $c_n/\lambda_q \to 0$, we have

$$\frac{\hat{\lambda}_{(q+1)} + c_n}{\hat{\lambda}_q + c_n} \to 0.$$

Further, since for any $l > q$, $\lambda_l = 0$ and $c_n/\tilde{\eta}_n \to \infty$, we achieve

$$\min_{l>q} \frac{\hat{\lambda}_{(l+1)} + c_n}{\hat{\lambda}_l + c_n} \geq \frac{\min_{l>q} \hat{\lambda}_q + c_n}{\max_{l>q} \hat{\lambda}_q + c_n} \geq \frac{-C\tilde{\eta}_n + c_n}{C\tilde{\eta}_n + c_n} \to 1 > \tau.$$

Therefore, we conclude that $P(\hat{q} = q) \to 1$. □

Proofs of Theorems S1.1, S1.2, and S1.3. The arguments used for proving Theorems 2.1, 2.2, and 2.3 can be used to prove these theorems; we then omit the details here.

Proof of Theorem 4.1. For any basis matrix $B \in \mathcal{R}^{p_Z \times q_\kappa}$ of $S^\kappa_{\{X_i\}_{i=1}^n}$, we have Span$(B) = S^\kappa_{\{X_i\}_{i=1}^n}$. Assume $X_i \in \mathbb{R}^p$ for $i = 1, \cdots, n$ belongs to a union of d categories $\{\mathcal{C}_k\}_{k=1}^d$ and $B^\top Z_i \in \mathbb{R}^{q_\kappa}$ for $i = 1, \cdots, n$ belongs to a union of $\tilde{K}$ categories $\{\tilde{\mathcal{C}}_k\}_{k=1}^{\tilde{K}}$.

First, for any pair X_i and X_j with $i \neq j$ belonging to the same category $\mathcal{C}_k$, we have $\mu_{Z,i} = E(Z_i) = E(Z_j) = \mu_{Z,j}$ and then $E(B^\top Z_i) = E(B^\top Z_j)$. This implies that $B^\top Z_i$ and $B^\top Z_j$ simultaneously belong to the some category $\tilde{\mathcal{C}}_l$. Therefore, we can conclude that any category $\mathcal{C}_k$ can belong to some category $\tilde{\mathcal{C}}_l$.

On the other hand, for any $B^\top Z_i$ and $B^\top Z_j$ with $i \neq j$ in the same category $\tilde{\mathcal{C}}_k$, namely, $E(B^\top Z_i) = E(B^\top Z_j)$. Then we can get that $B^\top \mu_{Z,i} = B^\top \mu_{Z,j}$. Therefore, $\mu_{Z,i} - \mu_{Z,j}$ is vertical to the subspace Span(B):

$$\mu_{Z,i} - \mu_{Z,j} \perp \text{Span}(B). \tag{S3.8}$$

By the definition of the central $\kappa-$moment deviation subspace $\mathcal{S}^{\kappa}_{\{X_i\}_{i=1}^n}$, we have $\mu_{Z,i} - \mu_{Z,j} \in \mathcal{S}^{\kappa}_{\{X_i\}_{i=1}^n}$. As Span$(B) = \mathcal{S}^{\kappa}_{\{X_i\}_{i=1}^n}$, we conclude that

$$\mu_{Z,i} - \mu_{Z,j} \in \text{Span}(B). \tag{S3.9}$$

Together the results in (S3.8) and (S3.9), we can get that $\mu_{Z,i} - \mu_{Z,j} = 0$. This produces that X_i and X_j are simultaneously in some category $\mathcal{C}_l$. Hence, any category $\tilde{\mathcal{C}}_k$ belongs to some category $\mathcal{C}_l$. Therefore, for any basis matrix $B \in \mathcal{R}^{p_Z \times q_\kappa}$ of $\mathcal{S}^{\kappa}_{\{X_i\}_{i=1}^n}$, both the sequences $\{B^\top Z_i\}_{i=1}^n$ and

$\{X_i\}_{i=1}^n$ have the same clustering results.

The argument for proving Theorem 2.2 can be adopted to derive the rest of this theorem. Hence we omit the details here. The proof is finished.

□

Table 1: Changes in the covariance matrix in ***Experiment 4***

Case	p_z	a	method	$\hat{k}$	MSE	RI	p_z	a	method	$\hat{k}$	MSE	RI
1	65	0.3	E-Divisive$_{dr}$	4.133	1.196	0.883	20	0.3	E-Divisive$_{dr}$	2.702	4.073	0.906
			E-Divisive	0.148	15.048	0.239			E-Divisive	0.093	15.391	0.224
			ks-cp3o$_{dr}$	5.460	5.784	0.886			ks-cp3o$_{dr}$	6.031	8.651	0.838
			ks-cp3o	6.298	10.118	0.769			ks-cp3o	6.265	10.550	0.773
			KCP$_{dr}$	3.363	6.695	0.704			KCP$_{dr}$	2.094	11.890	0.488
			KCP	0.006	15.966	0.200			KCP	0.079	15.555	0.212
		0.5	E-Divisive$_{dr}$	4.203	0.257	0.983		0.5	E-Divisive$_{dr}$	4.116	0.144	0.972
			E-Divisive	2.491	4.943	0.718			E-Divisive	0.557	12.807	0.333
			ks-cp3o$_{dr}$	4.061	0.146	0.983			ks-cp3o$_{dr}$	4.487	1.534	0.960
			ks-cp3o	6.248	10.222	0.775			ks-cp3o	6.181	9.751	0.776
			KCP$_{dr}$	4.678	1.944	0.979			KCP$_{dr}$	4.613	2.395	0.949
			KCP	4.013	1.479	0.935			KCP	4.327	11.267	0.785
2	65	0.3	E-Divisive$_{dr}$	4.074	0.092	0.969	20	0.3	E-Divisive$_{dr}$	3.499	1.822	0.841
			E-Divisive	0.311	14.155	0.275			E-Divisive	0.148	15.072	0.238
			ks-cp3o$_{dr}$	4.500	1.673	0.950			ks-cp3o$_{dr}$	5.721	7.487	0.867
			ks-cp3o	6.218	9.774	0.763			ks-cp3o	6.189	9.847	0.766
			KCP$_{dr}$	4.891	7.417	0.891			KCP$_{dr}$	2.427	12.665	0.516
			KCP	0.000	16.000	0.198			KCP	0.060	15.688	0.209
		0.5	E-Divisive$_{dr}$	4.050	0.084	0.980		0.5	E-Divisive$_{dr}$	4.056	0.084	0.969
			E-Divisive	0.600	12.626	0.336			E-Divisive	0.282	14.322	0.268
			ks-cp3o$_{dr}$	4.128	0.319	0.976			ks-cp3o$_{dr}$	4.417	1.394	0.955
			ks-cp3o	6.165	9.655	0.762			ks-cp3o	6.126	9.364	0.769
			KCP$_{dr}$	5.417	7.455	0.964			KCP$_{dr}$	4.974	8.482	0.882
			KCP	0.446	14.512	0.267			KCP	0.546	14.766	0.269
3	65	0.3	E-Divisive$_{dr}$	3.981	1.866	0.845	20	0.3	E-Divisive$_{dr}$	2.263	5.563	0.632
			E-Divisive	0.084	15.466	0.223			E-Divisive	0.099	15.367	0.226
			ks-cp3o$_{dr}$	5.606	7.152	0.863			ks-cp3o$_{dr}$	6.128	9.256	0.809
			ks-cp3o	6.312	9.944	0.772			ks-cp3o	6.317	10.399	0.768
			KCP$_{dr}$	3.146	8.238	0.637			KCP$_{dr}$	2.362	17.662	0.441
			KCP	0.001	15.993	0.199			KCP	0.028	15.818	0.203
		0.5	E-Divisive$_{dr}$	4.152	0.176	0.977		0.5	E-Divisive$_{dr}$	4.023	0.521	0.936
			E-Divisive	0.190	14.886	0.246			E-Divisive	0.178	14.896	0.245
			ks-cp3o$_{dr}$	4.201	0.621	0.967			ks-cp3o$_{dr}$	4.898	3.228	0.930
			ks-cp3o	6.322	10.206	0.774			ks-cp3o	6.240	9.910	0.765
			KCP$_{dr}$	5.299	4.731	0.961			KCP$_{dr}$	3.967	6.703	0.778
			KCP	0.031	15.827	0.205			KCP	0.437	17.323	0.236
4	65	0.3	E-Divisive$_{dr}$	4.137	2.105	0.850	20	0.3	E-Divisive$_{dr}$	2.054	6.200	0.626
			E-Divisive	0.133	15.173	0.268			E-Divisive	0.085	15.422	0.259
			ks-cp3o$_{dr}$	5.797	8.496	0.818			ks-cp3o$_{dr}$	6.327	10.061	0.773
			ks-cp3o	6.291	9.990	0.733			ks-cp3o	6.407	10.753	0.731
			KCP$_{dr}$	4.110	8.951	0.727			KCP$_{dr}$	1.607	12.371	0.438
			KCP	0.000	16.000	0.236			KCP	0.034	15.787	0.243
		0.5	E-Divisive$_{dr}$	4.242	0.433	0.970		0.5	E-Divisive$_{dr}$	3.835	0.639	0.936
			E-Divisive	0.270	14.349	0.300			E-Divisive	0.147	15.140	0.268
			ks-cp3o$_{dr}$	4.281	1.533	0.942			ks-cp3o$_{dr}$	5.053	5.390	0.874
			ks-cp3o	5.977	8.638	0.734			ks-cp3o	6.141	9.828	0.725
			KCP$_{dr}$	6.010	9.077	0.946			KCP$_{dr}$	5.154	9.923	0.845
			KCP	0.000	16.000	0.236			KCP	0.180	15.278	0.260

References

S. Chen, L. Zhang, and P. Zhong. Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819, 2010.

H. Dette, G. M. Pan, and Q. Yang. Estimating a change point in a sequence of very high-dimensional covariance matrices. *Journal of the American Statistical Association*, 117 (537):444–454, 2022.

N. Stransky, C. Vallot, F Reyal, Bernard-Pierrot, et al. Regional copy number–independent deregulation of transcription in cancer. *Nature Genetics*, 38(12):1386–1396, 2006.

S. Sun. Analytic expressions for the derivatives of the eigenvalues and eigenvectors of a matrix. *Advances in Math.(Beijing)*, 17:391–397, 1988.

L. Zhu and K. Fang. Asymptotics for kernel estimate of sliced inverse regression. *The Annals of Statistics*, 24(3):1053–1068, 1996.

X. Zhu, J. Guo, X. Guo, L. Zhu, and J. Zheng. Bandwidth selection for large covariance and precision matrices: a valley-cliff criterion. *Statistica Sinica*, 2023.