Statistica Sinica Preprint No: SS-2022-0384

Title	Moment Deviation Subspaces of Dimension Reduction for High-Dimensional Data With Change Structure
Manuscript ID	SS-2022-0384
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0384
Complete List of Authors	Xuehu Zhu, Luoyao Yu, Jiaqi Huang, Junmin Liu and Lixing Zhu
Corresponding Authors	Lixing Zhu
E-mails	lzhu@hkbu.edu.hk

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

Statistica Sinica

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 at Zhuhai, China*

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

Abstract: This paper introduces the notion of moment deviation subspaces of dimension reduction for high-dimensional data with change structure. We propose a novel estimation method to identify subspaces by combining the Mahalanobis matrix and the pooled covariance matrix. The theoretical properties are investigated to show that the change point detection and clustering can be equivalently implemented in the dimension reduction subspaces, whether the data structure is dense or sparse, whenever the dimension divided by the sample size goes to zero. We propose an iterative algorithm based on dimension reduction subspaces that can be applied for data clustering of high-dimensional data. The numerical studies on synthetic and real data sets suggest that the dimension reduction versions of existing methods of change point detection and clustering methods

*Corresponding author (L. Zhu). Email address: lzhu@hkbu.edu.hk.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

significantly improve the performances of existing approaches in finite sample scenarios.

Key words and phrases: Clustering; Dimension reduction; Moment changes; Moment deviation subspace.

1. Introduction

This research is motivated by detecting structural changes and clustering of high-dimensional data. For change point detection, there are several proposals available in the literature. For instance, Jirak [2015] suggested a coordinate-wise CUSUM-statistic; Cho and Fryzlewicz [2015] proposed the sparsified binary segmentation (SBS) method; Cho [2016] used a double CUSUM statistic for panel data; Wang and Samworth [2018] developed a projection-based method; Enikeeva and Harchaoui [2019] developed a scan-statistic-based algorithm; Grundy et al. [2020] proposed a method via a geometrically inspired mapping; and Dette et al. [2022] proposed a two-stage approach for the covariance matrix structure; and Wang et al. [2022] applied a self-normalized U-statistic to replace the CUSUM statistics.

Without sparsity structure, the dimensionality problem challenges most existing methods. Dimension reduction with no loss of the information provided by the original data is then an important technique to alleviate this

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

challenge. In a different but relevant research field with supervised learning, sufficient dimension reduction introduced first by Li [1991] can achieve this goal by projecting original predictors onto a lower-dimensional subspace called the central subspace. In the last three decades, several promising methods have been developed, such as inverse regression methods (e.g., [Li, 1991, Cook and Weisberg, 1991, Zhu et al., 2010]), forward regression methods (e.g., [Xia et al., 2002]). This paper introduces the notion of central moment deviation subspaces of dimension reduction and verifies the equivalence between the changes in the dimension reduction subspace and the original data space. We develop a novel method to construct a subspace estimation by combining the Mahalanobis matrix and the pooled covariance matrix. As the detection is performed on the lower-dimensional subspace, we could significantly enhance the performances of existing methods. When the primary interest is on the mean structure, our method needs not to assume the homoscedasticity of observations. When we are interested in detecting the number of change points and their locations under the contemporaneous mean and second-order moment structures, we can extend the method to handle higher central moment deviation subspace. For space-saving, we put the results in Supplementary Materials.

Unlike change point analysis, when the clustering analysis is considered,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

there is no sufficient information on the details of the subscript over the data. Hence we can not directly estimate the pooled covariance matrix. To overcome this difficulty, we then develop an iterative subspace clustering algorithm to improve some classical clustering methods, such as the K-means algorithm.

For the estimated dimension reduction subspaces, we show the consistency whenever the dimension is fixed or divergent at a certain rate as the sample size goes to infinity. The asymptotic results apply to both dense and sparse data structures. But the current method has a limitation in that the method can not be used to handle ultra-high dimension cases. If we wish to study the properties in those cases, the estimation procedure for the dimension reduction subspaces needs to modify, say, using a method for dimension reduction with simultaneous variable selection, see, e.g., [Wang et al., 2018, Lin et al., 2019, Qian et al., 2019]. Some technical issues remain to be unsolved; thus, the research is beyond the scope of this paper and deserves further study.

The remainder of the paper is organized as follows. Subsection 2.1 introduces the notion of central mean deviation subspace and proposes a novel method to identify it. Subsection 2.2 suggests a criterion to determine the subspace dimension. Section 3 contains the dimension reduction method

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

for clustering and suggests an iterative algorithm. Section 4 includes simulation studies and illustrative analyses of Genetics data and Financial data. Section 5 discusses the merits and limitations of the new method and some other research topics. For space-saving, we, in Supplementary Materials, discuss an extension of central mean deviation subspace to central $\kappa-$moment deviation subspace to handle more general issues such as covariance matrices with change structure. Supplementary Materials also include part of the simulations with changes in the covariance matrix, the regularity conditions, and technical proofs for the theorems.

2. Central mean deviation subspace

Before giving the detail of the notion and the constructions of this subspace and its estimation, we point out that the methods and results described in this section can be extended to develop the general central κ-th moment deviation subspace when we want to consider the contemporaneous mean or second-order moment change structures. The results can be used for clustering analysis, as described in Section 3. To save space, the details can be found in Supplementary Materials.

2.1 The subspace identification

Let $X_i = (X_{i1}, \cdots, X_{ip})^\top$, for $i = 1 \cdots, n$, be independent p-dimensional random vectors as

$$X_i = \mu_i + \epsilon_i, 1 \leq i \leq n, \tag{2.1}$$

where $\mu_i = E(X_i)$ and $\Sigma_i = \text{Cov}(X_i)$. The primary interest in this section is on the means μ_i's. Assume that the sequence $\{\mu_i\}_{i=1}^n$ follows a piecewise constant structure with $K+1$ segments. That is, there are K change points $1 \leq z_1 < z_2 < ... < z_K \leq n$ such that $\mu_{z_{k-1}+j} = \mu^{(k)}$, $\Sigma_{z_{k-1}+j} = \Sigma^{(k)}$ and $\mu^{(k)} \neq \mu^{(k+1)}$, for $k = 1, \cdots, K$ and $1 \leq j \leq z_k - z_{k-1}$, with $z_0 = 0$ and $z_{K+1} = n$. Let $\text{Span}\{\mu^{(k)} - \mu^{(l)}, \text{ for } k, l = 1, 2, \cdots, K+1\}$ denote the column space spanned by $\{\mu^{(k)} - \mu^{(l)}, \text{ for } k, l = 1, 2, \cdots, K+1\}$.

Definition 2.1. $\text{Span}\{\mu^{(k)} - \mu^{(l)}, \text{ for } k, l = 1, 2, \cdots, K+1\}$ is called the central mean deviation subspace of the sequence $\{X_i\}_{i=1}^n$ and is written as $S_{\{E(X_i)\}_{i=1}^n}$. For this subspace, $q = \dim\{S_{\{E(X_i)\}_{i=1}^n}\}$ is called the structural dimension of $S_{\{E(X_i)\}_{i=1}^n}$.

The following theorem states the equivalence between the change structures of the original data sequence and the low-dimensional data sequence.

Theorem 2.1. *For any basis matrix* $B \in \mathcal{R}^{p\times q}$ *of* $\mathrm{S}_{\{E(X_i)\}_{i=1}^n}$ *with* $q \leq \min\{p, K\}$, *both the sequences* $\{B^\top X_i\}_{i=1}^n$ *and* $\{X_i\}_{i=1}^n$ *have the same locations of changes.*

Hence, Theorem 2.1 persuasively offers a way to detect change points by using the sequence projected $\{B^\top X_i\}_{i=1}^n$. Motivated by Xiang et al. [2008], we estimate the projection matrix B using the following Mahalanobis matrix as the target matrix:

$$M_n = \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{i\neq j}(X_i - X_j)(X_i - X_j)^\top. \tag{2.2}$$

Compute the expectation of M_n to see that

$$\begin{aligned}
E(M_n) &= \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{i\neq j} E\left\{(X_i - X_j)(X_i - X_j)^\top\right\} \\
&= \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{i\neq j} \mathrm{Cov}(X_i - X_j) \\
&\quad + \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{i\neq j} E(X_i - X_j)E(X_i - X_j)^\top \\
&= \frac{2}{n}\sum_{i=1}^n \Sigma_i + \sum_{k=1}^{K+1}\sum_{l\neq k, l\leq K+1}\frac{n_l n_k}{n(n-1)}(\mu^{(k)} - \mu^{(l)})(\mu^{(k)} - \mu^{(l)})^\top,
\end{aligned}$$

where n_k is the segment length between two consecutive changes. When

$n_k/n \to c_k > 0$, for $k = 1, 2, \cdots, K+1$, we have

$$\begin{aligned} E(M_n) &\to 2\sum_{k=1}^{K+1} c_k \Sigma^{(k)} + \sum_{k=1}^{K+1} \sum_{l\neq k, l\leq K+1} c_k c_l (\mu^{(k)} - \mu^{(l)})(\mu^{(k)} - \mu^{(l)})^{\top} \\ &=: 2\Sigma_{pooled} + \Delta = M. \end{aligned} \quad (2.3)$$

Theorem 2.2. *Under the model (2.1), we have* $\mathrm{Span}(\Delta) = \mathrm{S}_{\{E(X_i)\}_{i=1}^n}$. *Furthermore,* $\mathrm{Span}(B) = \mathrm{S}_{\{E(X_i)\}_{i=1}^n}$, *where* $B = (v_1, \cdots, v_q)$ *denotes the matrix consisting of the eigenvectors of* Δ *associated with the nonzero eigenvalues of* Δ.

To efficiently estimate Δ and then the subspace $S_{\{E(X_i)\}_{i=1}^n}$, we need to have a good estimator of the pooled covariance matrix Σ_{pooled}. As the locations of changes are unknown, we suggest a "divide-and-conquer" strategy to estimate this matrix involving the different means $\mu^{(k)}$, for $k = 1, \cdots, K+1$. Let $\tilde{K} = \lfloor n/\beta_n \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor operation and β_n is a tunning parameter depending on n. Divide the data into $\tilde{K}$ segments as $\mathcal{S}_m = \{(m-1)\beta_n + 1, \cdots, m\beta_n\}$, for $m = 1, 2 \cdots, \tilde{K} - 1$ and $\mathcal{S}_{\tilde{K}} = \{(\tilde{K}-1)\beta_n + 1, \cdots, n\}$. Compute the covariance matrices for all segments and then average them to get the final estimator $\Sigma_{pooled,n}$ of

Σ_{pooled} as:

$$\Sigma_{pooled,n} = \frac{1}{\tilde{K}} \sum_{m=1}^{\tilde{K}} \hat{\Sigma}_m \text{ with } \hat{\Sigma}_m = \frac{1}{\#\{\mathcal{S}_m\} - 1} \sum_{k \in \mathcal{S}_m} (X_k - \bar{X}_m)(X_k - \bar{X}_m)^\top \quad (2.4)$$

where $\bar{X}_m = \frac{1}{\#\{\mathcal{S}_m\}} \sum_{k \in \mathcal{S}_m} X_k$ with $\#\{\mathcal{S}_m\}$ being the cardinality of the sets $\mathcal{S}_m$'s. Together with the formula in (2.2) and (2.4), Δ can be estimated as:

$$\Delta_n = M_n - 2\Sigma_{pooled,n}.$$

Then an estimator B_n of the basis matrix B consists of the eigenvectors associated with the largest q eigenvalues of Δ_n.

Theorem 2.3. *Under the model (2.1), assume that $X_i - E(X_i)$ are independent random variables, and Assumptions S3.1, S3.2, S3.3 and S3.4 in Supplementary Materials hold. Then,*

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

where $||\cdot||_F$ denotes the Frobenius norm of a matrix. Furthermore, when q is given,

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

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

Remark 2.1. The above results indicate that when $\beta_n = O(n^m)$ with $0 \leq m \leq 1/2$, including the case where β_n is fixed, the convergence rate of $||B_n - B||_F$ is $O_p\left(\sqrt{p/n}\right)$. The estimation consistency can hold as long as $p = o(n)$. In other words, the convergence rate is identical in a large range of β_n. Further, we note that when there is no change point, the estimator of Σ_{pooled} is unbiased, and the variance of every element is of the order $1/n$ in theory. This reminds us that the tuning parameter β_n intrinsically differs from the bandwidth in a nonparametric estimation, which can be selected through a balance between the bias and variance. Thus, in general, choosing a β_n that could minimize the error, say MSE, seems not possible unless we would have another criterion for such a selection. In practice, if β_n is too small, the invalid estimate of the covariance for each segment maybe lead to a lousy estimator of the pooled covariance matrix Σ_{pooled}. When β_n is too large, each segment may contain multiple distributions, which also leads to a lousy estimator. As a compromise, we recommend $\beta_n = \lfloor\sqrt{n}\rfloor$ by the rule of thumbs in Section 4.

2.2 The structural dimension determination

As the structural dimension q is usually unknown, which is related to the number of change points K, determining q plays a crucial role in efficiently identifying this subspace. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_q > \lambda_{q+1} = \ldots = \lambda_p = 0$ denote the eigenvalues of the $p \times p$ positive semi-definite matrix Δ. As is well known, all the eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_p$ of the estimated target matrix Δ_n are usually non-zero.

Inspired by the method proposed in Zhu et al. [2020a,b], we suggest a thresholding ridge ratio (TRR) criterion to estimate the structural dimension q by:

$$\hat{q} := \max_{1 \leq k \leq p-1} \left\{ k : \ \hat{r}_k = \frac{\hat{\lambda}_{k+1} + c_n}{\hat{\lambda}_k + c_n} \leq \tau \right\}, \tag{2.5}$$

where the ridge value c_n tends to zero at a certain rate of convergence and the thresholding value τ satisfies $0 < \tau < 1$. According to the plug-in principle in Zhu et al. [2020a], choosing $\tau = 0.5$ is reasonable to avoid in general overestimation with large τ and underestimation with small τ. Further, as the target matrix involved herewith is different from those in Zhu et al. [2020a], we then recommend the ridge value to be $c_n = 0.5 \log(\log(n))\sqrt{p/n}$ chosen by the rule of thumb as there is no theoretical result for optimal selection.

The consistency of $\hat{q}$ is stated in the following theorem.

Theorem 2.4. *Let* $\tilde{\eta}_n = \max\left\{\sqrt{\frac{p}{n}}, \frac{\sqrt{p}\beta_n}{n}\right\}$. *Under the same conditions in*

Theorem 2.3, if c_n satisfies $c_n \to 0$, $\tilde{\eta}_n \to 0$, $c_n/\tilde{\eta}_n \to \infty$ as $n \to \infty$, then $P(\hat{q} = q) \to 1$.

3. An iterative algorithm for subspace identification in cluster analysis

Suppose the observations $X_i = (X_{i1}, \cdots, X_{ip})^\top \in \mathbb{R}^p$ for $i = 1, \cdots, n$ are independent. Cluster information may not only be limited to the mean; higher moment clustering as a more general approach could also be of interest. Thus, we define the new high-dimensional variables Z_i based on X_i as:

$$\begin{aligned} Z_i &= (X_{i1}, ..., X_{ip}, X_{i1}^2, X_{i1}X_{i2}, ..., X_{i1}X_{ip}, X_{i2}^2, X_{i2}X_{i3}, ..., X_{i2}X_{ip}, \\ &\quad \cdots X_{i1}^\kappa, X_{i1}^{\kappa-1}X_{i2} \cdots, X_{ip}^\kappa)^\top, \end{aligned} \tag{3.1}$$

where κ denotes some positive integer. As $\kappa = 2$ covers the information of the mean and covariance, this may be used frequently in practice.

Assume that $\{X_i\}_{i=1}^n$ belong to a union of d categories $\{\mathcal{C}_k\}_{k=1}^d$ which satisfy that if both X_i and X_j are in the same $\mathcal{C}_k$ for $k = 1, \cdots, d$, then $E(Z_i) = E(Z_j)$ holds. Each category $\mathcal{C}_k$ contains n_k datum points with $\sum_{k=1}^d n_k = n$. Similarly, when $X_j \in \mathcal{C}_k$, let $E(Z_j) = \mu_Z^{(k)}$ and $\Sigma_Z^{(k)} =$

$\text{Cov}(Z_j)$ for $k = 1, \cdots, d$.

Definition 3.1. $\text{Span}\{\mu_Z^{(k)} - \mu_Z^{(l)},\ \text{for}\ k, l = 1, \cdots, d\}$ is called the central $\kappa-$th moment deviation subspace of the sequence $\{X_i\}_{i=1}^n$ and is written as $S^{\kappa}_{\{X_i\}_{i=1}^n}$. Further, $q_\kappa = \dim\{S^{\kappa}_{\{X_i\}_{i=1}^n}\}$ is called the structural dimension of $S^{\kappa}_{\{X_i\}_{i=1}^n}$.

Consider the following Mahalanobis matrix of the sequence $\{Z_i\}_{i=1}^n$ as:

$$M_{Z,n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j} (Z_i - Z_j)(Z_i - Z_j)^\top. \tag{3.2}$$

We follow the similar arguments as proving the formula in (2.3), we have that as $n_k/n \to c_k > 0$, for $k = 1, \cdots, d$, and $\sum_{k=1}^d c_k = 1$,

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

Define the central κ-th moment deviation subspace $S^{\kappa}_{\{X_i\}_{i=1}^n} = \text{Span}\{\mu_Z^{(k)} - \mu_Z^{(l)},\ \text{for}\ k, l = 1, \cdots, d\}$. Here the dimension q_κ of $S^{\kappa}_{\{X_i\}_{i=1}^n}$ is less than or equal to $\min\{p_Z, d-1\}$. Then the following theorem offers a way to construct a new algorithm to cluster the lower-dimensional data.

Theorem 3.1. *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 $\{B^\top Z_i\}_{i=1}^n$ and $\{Z_i\}_{i=1}^n$ have the same clustering results. Furthermore, we have $\mathrm{Span}(B) = \mathrm{S}^{\kappa}_{\{X_i\}_{i=1}^n}$, where $B = (v_1, \cdots, v_{q_\kappa})$ denotes the eigenvectors of Δ_Z associated with the nonzero eigenvalues of Δ_Z.*

As commented in the Introduction, the subscript of the sequence $\{Z_i\}_{i=1}^n$ can not provide any information such that we can not directly estimate the pooled covariance matrix Σ^Z_{pooled}. We suggest the following iterative subspace clustering procedure.

Initial value choice. Motivated from Xiang et al. [2008], get an initial basis matrix B_n via optimizing the following objective function as:

$$B_n = \arg \max_{B \in \mathcal{R}^{p_Z \times q_\kappa}} \frac{1}{n(n-1)} \sum_{i \neq j} ||B^\top M_{Z,n} B|| \quad \text{s.t} \quad B^\top B = I_{q_\kappa}. \tag{3.3}$$

This is equivalent to learning the central κ-th moment deviation subspace when $\kappa = 1$ and $\mathrm{Cov}(Z_i) = \sigma I_{p_Z \times p_Z}$ for $i = 1, \cdots, n$. See Supplementary Materials. As q_κ of $S^{\kappa}_{\{X_i\}_{i=1}^n}$ is smaller than or equal to $\min\{p_Z, d-1\}$, it is reasonable to learn the basis matrix B_n by (3.3) as an initial value with $\hat{q}_\kappa = d-1$ in the first step.

Clustering step. In this paper, we choose the classical method such as K-means to cluster $\{B_n^\top Z_i\}_{i=1}^n$ to get $\{\hat{\mathcal{C}}_i\}_{i=1}^d$ with the pre-specified number d of categories.

Dimension reduction step. Calculate the covariance for each category and then have a weighted average of them to get an estimator of the pooled covariance matrix Σ_{pooled}^Z as:

$$\Sigma_{pooled,n}^Z = \sum_{k=1}^{d} \frac{\#\{\hat{\mathcal{C}}_k\} - 1}{n-d} \hat{\Sigma}_{Zk}, \tag{3.4}$$

where $\hat{\Sigma}_{Zk} = \frac{1}{\#\{\hat{\mathcal{C}}_k\}-1} \sum_{j \in \hat{\mathcal{C}}_k} (Z_j - \bar{Z}_k)(Z_j - \bar{Z}_k)^\top$ with $\bar{Z}_k = \frac{1}{\#\{\hat{\mathcal{C}}_k\}} \sum_{j \in \hat{\mathcal{C}}_k} Z_j$ and $\#\{\hat{\mathcal{C}}_k\}$ denotes the cardinality of the set $\hat{\mathcal{C}}_k$. Combining the formula (3.2) and (3.4), the estimated target matrix is defined as:

$$\Delta_{Z,n} = M_{Z,n} - 2\Sigma_{pooled,n}^Z. \tag{3.5}$$

Similarly, we can determine the dimension q_κ by TRR defined in (2.5). Then an estimator B_n of the basis matrix B consists of the eigenvectors associated with the largest $\hat{q}_\kappa$ eigenvalues of Δ_n.

Iteration step. Iterate the dimension reduction and clustering steps based on the lower-dimensional data with some stopping criterion. Here we adopt the Rand index (RI) [Rand, 1971] as the stopping criterion as the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

RI describes the similarity between two adjacent clustering results. If two clusters of n observations are given by U and V, the RI is defined as:

$$RI = \frac{a+b}{\binom{n}{2}},$$

where a denotes the number of the point pairs in the same class under U and in the same class under V, b presents the number of the point pairs in the different classes under U and in the different classes under V. The maximum of the RI is 1. A good algorithm performs well with a large RI. The above procedures can be summarized below in Algorithm 1.

Algorithm 1 Iterative Subspace Cluster Algorithm.

Require: $X \in \mathcal{R}^{n\times p}$, $\tau = 0.5$, $c_n = 0.5\log(\log(n))\sqrt{p/n}$;
1: Calculate the $M_{Z,n}$ in (3.2) and set $\hat{q}_\kappa = d-1$, then learn the basis matrix B_n estimated by (3.3);
2: Choose a classical clustering algorithm such as K-means to cluster the lowered data $\{B_n^\top Z_i\}_{i=1}^n$, then get $\hat{\mathcal{C}}_k$ and calculate the pooled covariance matrix $\Sigma^Z_{pooled,n}$ by (3.4);
3: Update the target matrix $\Delta_{Z,n}$ in (3.5) and make the eigen-decomposition: the eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_{p_Z}$ and the eigenvectors $\hat{\nu}_1, \cdots, \hat{\nu}_{p_Z}$;
4: Determine the dimension q_κ based on TRR in (2.5) and then have the matrix $B_n = (\hat{\nu}_1, \cdots, \hat{\nu}_{\hat{q}_\kappa})$;
5: Repeat step 2 and then calculate the RI between the clustering result and the last clustering result.
6: Repeat steps 3-5 until the RI is greater than 0.99;
Ensure: $\{\hat{\mathcal{C}}_1, \cdots, \hat{\mathcal{C}}_d\}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

4. Numerical experiments

In this section, we conduct several experiments on synthetic data and real data examples to examine the finite sample performances of the proposed methods. Throughout the simulations, each experiment is repeated 1000 times.

4.1 Experiments on change point detection

We compare five popularly change-point detection methods with their dimension reduction versions: the E-Divisive method [Matteson and James, 2014], the change-point detection tests using rank statistics [Lung-Yut-Fong et al., 2015], the sparsified binary segmentation (SBS) method [Cho and Fryzlewicz, 2015], the change point procedure via pruned objectives by Kolmogorov-Smirnov statistic [Zhang et al., 2017] and the kernel change-point algorithm [Arlot et al., 2019], which are written as E-Divisive, Multirank, SBS, ks-cp3o and KCP, respectively. Their dimension reduction-based versions are written as E-Divisive$_{dr}$, Multirank$_{dr}$, SBS$_{dr}$, ks-cp3o$_{dr}$ and KCP$_{dr}$, respectively. Because the SBS method is applied to multivariate data, if the dimension q is determined to be 1, it reduces to wild binary segmentation method (WBS) [Fryzlewicz, 2014], which is a univariate change point method. We also compare with the change point detection

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

methods proposed by Wang and Samworth [2018], Cho [2016] and Grundy et al. [2020], which are abbreviated as Inspect, DCBS and GeomCP. The comparison is still sensible as they can also be used in non-sparse scenarios. This section only considers SBS, DCBS, GeomCP, Inspect and Multirank for mean change detection.

To evaluate the performances of different methods for estimating the number of change points, we calculate the average of $\hat{K}$ and the mean squared error (MSE) of $\hat{K}$, and also the RI as an evaluation index for the estimated locations of changes. The E-Divisive and ks-cp3o methods are implemented in the R package: *ecp*. The Multirank method is implemented by the Python code from the author of Lung-Yut-Fong et al. [2015]. The SBS and DCBS methods are implemented in the R package: *hdbinseg*. The GeomCP method is implemented in the R package: *changepoint.geo*. The WBS method is implemented in the R package: *wbs*. The Inspect method is implemented in the R package: $InspectChangepoint$.

Consider the following three situations: (1) changes in mean, (2) changes in covariance matrix, and (3) changes in distribution. To save space of the main context, we put the numerical results with changes in the covariance matrix in Supplementary Materials. The sample size is $n = 500$, and the number of change points is $K = 4$ and 9.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

Experiment 1: Changes in mean with $K = 4$**.** The data are generated from the multivariate normal distributions G_0, G_1, G_2, G_3 and G_4 as $G_i = N(u_i, \mathrm{I}_{p\times p})$, where $\mathrm{I}_{p\times p}$ denotes the identify matrix with $p = 100, 200$. The change points are located at $100i$ for $i = 1, 2, 3, 4$, respectively. We consider the following cases:

Case 1: $u_0 = u_2 = u_4 = -u$ and $u_1 = u_3 = u$, where the first 10 elements of the vector u are equal to $0.1, 0.2$, and the others equal to 0;

Case 2: $u_0 = u_2 = u_4 = -u$ and $u_1 = u_3 = u$, where all the elements of the vector u are equal to $0.1, 0.2$.

Table 1: The RI of different β_n in ***Experiment 1*** with Case 1

	$p = 100$, $u = 0.2$						$p = 200$, $u = 0.2$					
Method	2	5	$\lfloor\sqrt{n}/2\rfloor$	$\lfloor\sqrt{n}\rfloor$	$\lfloor 2\sqrt{n}\rfloor$	$\lfloor n/3\rfloor$	2	5	$\lfloor\sqrt{n}/2\rfloor$	$\lfloor\sqrt{n}\rfloor$	$\lfloor 2\sqrt{n}\rfloor$	$\lfloor n/3\rfloor$
E-Divisive	0.789	0.958	0.964	0.965	0.965	0.737	0.683	0.916	0.939	0.946	0.944	0.796
Multirank	0.198	0.198	0.970	0.973	0.965	0.602	0.198	0.198	0.946	0.940	0.934	0.623
SBS	0.726	0.958	0.968	0.970	0.971	0.730	0.590	0.912	0.932	0.947	0.948	0.795
KCP	0.664	0.933	0.952	0.959	0.953	0.754	0.385	0.834	0.928	0.938	0.936	0.794
ks-cp3o	0.864	0.948	0.965	0.966	0.959	0.820	0.817	0.904	0.939	0.948	0.944	0.818
	$p = 100$, $u = 0.1$						$p = 200$, $u = 0.1$					
Method	2	5	$\lfloor\sqrt{n}/2\rfloor$	$\lfloor\sqrt{n}\rfloor$	$\lfloor 2\sqrt{n}\rfloor$	$\lfloor n/3\rfloor$	2	5	$\lfloor\sqrt{n}/2\rfloor$	$\lfloor\sqrt{n}\rfloor$	$\lfloor 2\sqrt{n}\rfloor$	$\lfloor n/3\rfloor$
E-Divisive	0.433	0.719	0.838	0.866	0.875	0.727	0.446	0.779	0.860	0.872	0.872	0.796
Multirank	0.198	0.198	0.605	0.640	0.641	0.616	0.198	0.198	0.659	0.717	0.734	0.618
SBS	0.339	0.670	0.836	0.855	0.864	0.724	0.371	0.785	0.862	0.861	0.863	0.797
KCP	0.266	0.446	0.614	0.764	0.847	0.749	0.243	0.405	0.641	0.793	0.860	0.796
ks-cp3o	0.775	0.799	0.816	0.848	0.864	0.814	0.775	0.796	0.789	0.831	0.858	0.819

The mean changes are sparse in Case 1 and dense in Case 2. Different values of u can be viewed as the representatives of weak and strong signals. To evaluate the impact of β_n on our method, we compare the performance of the above five methods in Case 1 when β_n takes values of $2, 5, \lfloor\sqrt{n}/2\rfloor, \lfloor\sqrt{n}\rfloor, \lfloor 2\sqrt{n}\rfloor, \lfloor n/3\rfloor$. The results, presented in Table 1, indi-

Table 2: Changes in the mean in ***Experiment 1*** with Case 1

p	u	Method	$\hat{k}$	MSE	RI	u	Method	$\hat{k}$	MSE	RI
100	0.2	E-Divisive$_{dr}$	4.565	0.928	0.967	0.1	E-Divisive$_{dr}$	5.889	6.099	0.871
		E-Divisive	3.406	1.796	0.872		E-Divisive	0.231	14.551	0.260
		Multirank$_{dr}$	4.057	0.206	0.966		Multirank$_{dr}$	3.363	8.049	0.613
		Multirank	0.055	15.895	0.202		Multirank	0.004	15.978	0.199
		SBS$_{dr}$	4.448	0.889	0.972		SBS$_{dr}$	6.192	9.117	0.860
		SBS	0.020	15.862	0.204		SBS	0.003	15.979	0.199
		KCP$_{dr}$	5.541	6.094	0.957		KCP$_{dr}$	4.801	9.481	0.761
		KCP	0.000	16.000	0.198		KCP	0.000	16.000	0.198
		ks-cp3o$_{dr}$	4.235	0.807	0.961		ks-cp3o$_{dr}$	5.995	8.063	0.834
		ks-cp3o	6.304	10.082	0.832		ks-cp3o	6.271	9.989	0.784
		GeomCP	0.005	15.966	0.200		GeomCP	0.014	15.906	0.200
		DCBS	0.088	15.422	0.223		DCBS	0.004	15.975	0.200
		Inspect	0.344	14.076	0.282		Inspect	0.029	15.799	0.208
200	0.2	E-Divisive$_{dr}$	5.635	4.354	0.941	0.1	E-Divisive$_{dr}$	7.596	15.955	0.874
		E-Divisive	2.001	6.513	0.633		E-Divisive	0.133	15.151	0.235
		Multirank$_{dr}$	4.169	0.774	0.931		Multirank$_{dr}$	4.503	7.817	0.704
		Multirank	0.000	16.000	0.198		Multirank	0.000	16.000	0.198
		SBS$_{dr}$	5.912	6.796	0.943		SBS$_{dr}$	8.886	27.745	0.868
		SBS	0.027	15.813	0.207		SBS	0.016	15.888	0.204
		KCP$_{dr}$	5.551	5.534	0.940		KCP$_{dr}$	6.064	13.815	0.792
		KCP	0.000	16.000	0.198		KCP	0.000	16.000	0.198
		ks-cp3o$_{dr}$	4.449	1.483	0.951		ks-cp3o$_{dr}$	6.177	8.818	0.831
		ks-cp3o	6.388	10.372	0.834		ks-cp3o	6.279	10.089	0.787
		GeomCP	0.004	15.975	0.198		GeomCP	0.011	15.928	0.198
		DCBS	0.004	15.975	0.199		DCBS	0.000	16.000	0.198
		Inspect	0.100	15.402	0.224		Inspect	0.027	15.817	0.207

cate that the performances associated with $\beta_n = \lfloor\sqrt{n}/2\rfloor, \lfloor\sqrt{n}\rfloor, \lfloor 2\sqrt{n}\rfloor$ outperform those of $2, 5, \lfloor n/3\rfloor$. This is consistent with the claim in **Remark 2.1**. Notably, the best choice of β_n varies in different scenarios, but overall $\beta_n = \lfloor\sqrt{n}\rfloor$ makes the estimation most robust in terms of performance. Hence, we recommend this value of β_n in the subsequent simulations.

The results are reported in Tables 2 and 3. The findings are as follows. When the magnitudes of changes becomes large, the performances of most

Table 3: Changes in the mean in ***Experiment 1*** with Case 2

p	u	Method	$\hat{k}$	MSE	RI	u	Method	$\hat{k}$	MSE	RI
		E-Divisive$_{dr}$	4.063	0.079	0.992		E-Divisive$_{dr}$	4.187	0.231	0.988
		E-Divisive	4.043	0.049	0.993		E-Divisive	4.048	0.054	0.988
		Multirank$_{dr}$	4.000	0.000	0.993		Multirank$_{dr}$	4.000	0.000	0.991
		Multirank	2.231	8.491	0.420		Multirank	0.297	15.237	0.221
		SBS$_{dr}$	4.065	0.081	0.999		SBS$_{dr}$	4.169	0.249	0.992
		SBS	0.432	13.364	0.311		SBS	0.020	15.864	0.204
100	0.2	KCP$_{dr}$	4.014	0.016	0.994	0.1	KCP$_{dr}$	4.130	0.316	0.989
		KCP	4.000	0.000	0.993		KCP	0.000	16.000	0.198
		ks-cp3o$_{dr}$	4.000	0.000	0.994		ks-cp3o$_{dr}$	4.013	0.015	0.989
		ks-cp3o	6.387	10.627	0.830		ks-cp3o	6.391	10.379	0.789
		GeomCP	4.031	0.034	0.994		GeomCP	4.041	0.052	0.988
		DCBS	4.002	0.002	0.994		DCBS	2.524	4.186	0.752
		Inspect	4.218	0.330	0.993		Inspect	1.890	7.708	0.583
		E-Divisive$_{dr}$	4.074	0.082	0.993		E-Divisive$_{dr}$	4.218	0.272	0.989
		E-Divisive	4.049	0.049	0.993		E-Divisive	4.054	0.054	0.992
		Multirank$_{dr}$	4.000	0.000	0.994		Multirank$_{dr}$	4.000	0.000	0.991
		Multirank	0.051	15.855	0.203		Multirank	0.297	15.237	0.221
		SBS$_{dr}$	4.086	0.104	0.999		SBS$_{dr}$	4.229	0.369	0.996
		SBS	1.101	9.959	0.469		SBS	0.040	15.734	0.210
200	0.2	KCP$_{dr}$	4.001	0.001	0.994	0.1	KCP$_{dr}$	4.091	0.166	0.992
		KCP	4.000	0.000	0.993		KCP	0.000	16.000	0.198
		ks-cp3o$_{dr}$	4.000	0.000	0.994		ks-cp3o$_{dr}$	4.001	0.001	0.993
		ks-cp3o	6.310	10.132	0.836		ks-cp3o	6.164	9.138	0.782
		GeomCP	4.014	0.014	0.994		GeomCP	4.025	0.032	0.989
		DCBS	4.000	0.000	0.995		DCBS	3.380	1.413	0.907
		Inspect	4.162	0.236	0.995		Inspect	3.156	4.404	0.773

competitors are comparable. SBS and Multirank perform worse than the others, and Multirank$_{dr}$ is the best. In the weak signal scenarios, E-Divisive, Multirank, KCP, Inspect, and SBS tend to underestimate the number of change points. Particularly, in the sparse change point settings in Case 1, all five methods fail to work, while their dimension reduction versions still work well. Overall, the dimension reduction strategy greatly improves the performances of the original methods.

In ***Experiment 2***, we design more complicated scenarios to illustrate the impact of various factors, including the structural dimension q, the number of change points K, outliers, and imbalanced data.

Experiment 2: Changes in mean with $K = 9$***.*** The data are generated from the multivariate normal distributions G_0, G_1, G_2,..., G_9 as $G_i = N(u_i, \mathrm{I}_{p\times p})$ with $p = 100, 200$. Consider the following cases:

Case 1: $u_0 = u_2 = u_4 = u_6 = u_8 = 0$, $u_1 = u_5 = u_9 = (a_1, a_2, ...a_5, 0, ..., 0)^\top$, $u_3 = u_7 = (b_1, b_2, ...b_5, 0, ..., 0)^\top$, $a_i = i/v$, $b_i = 1 - i/v$, $v = 5$, 10, the locations of change points are at 30, 95, 140, 175, 245, 295, 360, 390, 450;

Case 2: The settings of change points and u_i are the same as Case 1, but it includes 5% outliers from $N(u_i + w_i, \mathrm{I}_{p\times p})$ between each z_i and z_{i+1}. Here w_i's are p-dimensional vectors. To check the sensitivity of the methods against outliers, for each i, we randomly select 5% of its elements to take values 5, and the other elements are 0;

Case 3: $u_0 = u_2 = u_4 = u_6 = u_8 = 0$, $u_1 = u_9 = (a_1, a_2, ...a_5, 0, ..., 0)^\top$, $u_3 = (b_1, b_2, ...b_5, 0, ..., 0)^\top$, $a_i = i/10$, $b_i = 1 - i/10$, $u_5 = (u\mathrm{I}_{1\times 5}, \frac{u}{2}\mathrm{I}_{1\times 5}, 0, ..., 0)^\top$, $u_7 = (\frac{u}{2}\mathrm{I}_{1\times 5}, u\mathrm{I}_{1\times 5}, 0, ..., 0)^\top$, $u = 0.5$, 1. The settings of change points are the same as Case 1.

In this experiment, we set the structural dimension q to be 2 in Cases 1–2, and 4 in Case 3. All the cases consist of imbalanced data with a mixture of weak and strong signals. We consider outliers in Case 2 to assess the sensitivity of our proposed method. Tables 4–6 report the results. Specifically, E-Divisive$_{dr}$ performs better than the other methods, and SBS$_{dr}$ also shows promising results. Moreover, all five methods yield significant improvements through dimension reduction. Comparing the results of ***Experiment 1*** and ***Experiment 2***, we find that the dimension reduction-based methods are robust against the structural dimension q and the number of change points K. Furthermore, the results of this experiment also suggest that the dimension reduction-based methods are relatively robust against imbalanced data and data with outliers.

To check the sensibility of the strategy based on dimension reduction against the different distributions, we design ***Experiment 3***.

Experiment 3: Changes in distribution. The data are generated in the following settings:

Case 1: $G_0 = G_2 = G_4 = N(0_p, a\mathrm{I}_{p\times p})$ with $a = 0.6, 0.8$, $G_1 = G_3$ are the p-dimensional uniform distribution on the p-dimensional cube $[-1, 1]^p$, and the change points are located at $100i$th for $i = 1, 2, 3, 4$;

Case 2: $G_0 = G_2 = G_4 = N(0_p, \mathrm{I}_{p\times p})$ and $G_1 = G_3 = t(df, \Sigma)$ are the p-dimensional t-distribution with $df = 4$ and $\Sigma = (\sigma_{ij})$, where $\sigma_{ij} = I(i = j) + aI(i \neq j)$ with $a = 0.3, 0.5$, the locations of change points are set to be

Table 4: Changes in the mean in ***Experiment 2*** with Case 1

p	v	Method	$\hat{k}$	MSE	RI	v	Method	$\hat{k}$	MSE	RI
100	10	E-Divisive$_{dr}$	7.612	4.335	0.920	5	E-Divisive$_{dr}$	6.703	7.580	0.910
		E-Divisive	1.638	56.702	0.449		E-Divisive	3.245	37.372	0.641
		Multirank$_{dr}$	2.558	51.070	0.423		Multirank$_{dr}$	5.061	16.539	0.852
		Multirank	0.000	81.000	0.106		Multirank	0.077	80.233	0.110
		SBS$_{dr}$	7.473	5.697	0.901		SBS$_{dr}$	6.070	10.580	0.892
		SBS	0.000	81.000	0.106		SBS	0.032	80.468	0.112
		KCP$_{dr}$	6.686	16.463	0.797		KCP$_{dr}$	7.234	7.032	0.897
		KCP	0.000	81.000	0.106		KCP	0.021	80.601	0.105
		ks-cp3o$_{dr}$	6.633	9.644	0.836		ks-cp3o$_{dr}$	6.636	9.294	0.869
		ks-cp3o	6.351	11.340	0.804		ks-cp3o	6.356	11.346	0.799
		GeomCP	0.005	80.910	0.106		GeomCP	0.000	81.000	0.106
		DCBS	0.000	81.000	0.106		DCBS	0.021	80.638	0.110
		Inspect	0.021	80.638	0.114		Inspect	0.394	74.723	0.176
200	10	E-Divisive$_{dr}$	8.449	2.465	0.922	5	E-Divisive$_{dr}$	7.601	4.293	0.921
		E-Divisive	0.548	72.463	0.237		E-Divisive	1.473	59.718	0.387
		Multirank$_{dr}$	3.824	37.276	0.565		Multirank$_{dr}$	5.273	15.874	0.839
		Multirank	0.000	81.000	0.106		Multirank	0.000	81.000	0.106
		SBS$_{dr}$	9.240	4.773	0.911		SBS$_{dr}$	7.420	5.612	0.906
		SBS	0.005	80.910	0.108		SBS	0.059	80.027	0.120
		KCP$_{dr}$	7.213	11.501	0.844		KCP$_{dr}$	7.941	5.665	0.904
		KCP	0.000	81.000	0.106		KCP	0.000	81.000	0.106
		ks-cp3o$_{dr}$	6.527	10.537	0.833		ks-cp3o$_{dr}$	6.840	8.053	0.878
		ks-cp3o	6.213	12.011	0.801		ks-cp3o	6.399	10.356	0.803
		GeomCP	0.000	81.000	0.106		GeomCP	0.000	81.000	0.106
		DCBS	0.000	81.000	0.106		DCBS	0.000	81.000	0.106
		Inspect	0.016	80.729	0.113		Inspect	0.170	78.149	0.140

the same as Case 1;

Case 3: The settings of G_i are the same as Case 2, except that the locations of change points are 90, 250, 390, 450.

E-Divisive$_{dr}$ performs the best among the competitors, E-Divisive and KCP perform the worst, whereas KCP$_{dr}$ works much better than the original KCP. The results are reported in Table 7. ks-cp3o has a slight overestimation for the number of change points, but ks-cp3o$_{dr}$ significantly im-

Table 5: Changes in the mean in ***Experiment 2*** with Case 2

p	v	Method	$\hat{k}$	MSE	RI	v	Method	$\hat{k}$	MSE	RI
		E-Divisive$_{dr}$	7.489	5.121	0.913		E-Divisive$_{dr}$	6.745	7.268	0.913
		E-Divisive	1.472	59.277	0.415		E-Divisive	3.364	35.333	0.683
		Multirank$_{dr}$	2.059	57.065	0.363		Multirank$_{dr}$	4.819	20.255	0.805
		Multirank	0.003	80.947	0.106		Multirank	0.093	79.922	0.110
		SBS$_{dr}$	8.121	6.567	0.894		SBS$_{dr}$	6.623	8.810	0.893
		SBS	0.004	80.926	0.108		SBS	0.030	80.485	0.114
100	10	KCP$_{dr}$	4.879	30.268	0.663	5	KCP$_{dr}$	6.792	10.537	0.864
		KCP	0.000	81.000	0.106		KCP	0.000	81.000	0.106
		ks-cp3o$_{dr}$	6.701	9.476	0.843		ks-cp3o$_{dr}$	6.658	9.762	0.866
		ks-cp3o	6.117	12.693	0.794		ks-cp3o	6.307	11.108	0.795
		GeomCP	0.009	80.853	0.106		GeomCP	0.009	80.853	0.107
		DCBS	0.000	81.000	0.106		DCBS	0.000	81.000	0.106
		Inspect	0.169	78.364	0.144		Inspect	1.017	65.801	0.273
		E-Divisive$_{dr}$	7.918	3.680	0.911		E-Divisive$_{dr}$	7.506	4.619	0.917
		E-Divisive	0.874	67.329	0.327		E-Divisive	1.792	54.680	0.460
		Multirank$_{dr}$	2.137	55.505	0.380		Multirank$_{dr}$	3.607	36.371	0.604
		Multirank	0.000	81.000	0.106		Multirank	0.000	81.000	0.106
		SBS$_{dr}$	10.485	10.039	0.889		SBS$_{dr}$	9.143	7.104	0.898
		SBS	0.017	80.706	0.113		SBS	0.039	80.338	0.116
200	10	KCP$_{dr}$	3.814	41.662	0.541	5	KCP$_{dr}$	5.394	25.658	0.700
		KCP	0.000	81.000	0.106		KCP	0.000	81.000	0.106
		ks-cp3o$_{dr}$	6.113	12.175	0.807		ks-cp3o$_{dr}$	6.545	9.719	0.860
		ks-cp3o	6.294	11.649	0.805		ks-cp3o	6.186	11.801	0.799
		GeomCP	0.078	79.745	0.114		GeomCP	0.048	80.216	0.112
		DCBS	0.000	81.000	0.106		DCBS	0.000	81.000	0.106
		Inspect	0.325	76.017	0.173		Inspect	0.610	72.442	0.199

proves. It suggests that the dimension reduction-based methods are much more robust against different distributions and imbalanced data than their original counterparts.

To further reveal the reasons for the above phenomena, we draw the scatter plots of the first variable of the original data, namely $\{X_{i1}\}_{i=1}^n$, and of $\{B_{1n}^\top X_i\}_{i=1}^n$ or $\{B_{1n}^\top Z_i\}_{i=1}^n$ with B_{1n} being the 1 column vector of B_n in Figure 1. It is observed that the changes of $\{B_{1n}^\top X_i\}_{i=1}^n$ or $\{B_{1n}^\top Z_i\}_{i=1}^n$ at the

Table 6: Changes in the mean in ***Experiment 2*** with Case 3

p	u	Method	$\hat{k}$	MSE	RI	u	Method	$\hat{k}$	MSE	RI
		E-Divisive$_{dr}$	7.037	6.080	0.909		E-Divisive$_{dr}$	7.032	6.883	0.906
		E-Divisive	4.872	17.383	0.807		E-Divisive	1.697	54.622	0.454
		Multirank$_{dr}$	4.057	29.244	0.702		Multirank$_{dr}$	3.115	44.022	0.506
		Multirank	0.047	80.408	0.110		Multirank	0.000	81.000	0.106
		SBS$_{dr}$	6.239	9.771	0.870		SBS$_{dr}$	6.277	11.234	0.866
		SBS	0.346	75.303	0.209		SBS	0.447	73.404	0.188
100	1	KCP$_{dr}$	7.654	6.420	0.890	0.5	KCP$_{dr}$	6.702	11.936	0.834
		KCP	0.000	81.000	0.106		KCP	0.000	81.000	0.106
		ks-cp3o$_{dr}$	5.707	11.761	0.851		ks-cp3o$_{dr}$	6.670	8.766	0.855
		ks-cp3o	6.186	11.314	0.836		ks-cp3o	6.122	13.431	0.799
		GeomCP	0.011	80.830	0.109		GeomCP	0.032	80.479	0.110
		DCBS	0.734	69.415	0.303		DCBS	0.250	76.750	0.153
		Inspect	2.356	46.665	0.580		Inspect	0.814	67.282	0.256
		E-Divisive$_{dr}$	7.548	4.218	0.922		E-Divisive$_{dr}$	7.793	4.335	0.909
		E-Divisive	4.516	20.644	0.797		E-Divisive	1.011	64.745	0.320
		Multirank$_{dr}$	4.987	18.864	0.811		Multirank$_{dr}$	3.946	34.630	0.600
		Multirank	0.000	81.000	0.106		Multirank	0.000	81.000	0.106
		SBS$_{dr}$	7.277	6.223	0.896		SBS$_{dr}$	7.931	6.112	0.888
		SBS	0.255	76.734	0.185		SBS	0.431	73.676	0.186
200	1	KCP$_{dr}$	8.340	5.351	0.893	0.5	KCP$_{dr}$	7.165	9.144	0.853
		KCP	0.000	81.000	0.106		KCP	0.000	81.000	0.106
		ks-cp3o$_{dr}$	5.872	10.681	0.856		ks-cp3o$_{dr}$	6.601	9.516	0.846
		ks-cp3o	6.213	10.755	0.841		ks-cp3o	6.090	12.771	0.807
		GeomCP	0.000	81.000	0.106		GeomCP	0.000	81.000	0.106
		DCBS	0.005	80.910	0.109		DCBS	0.000	81.000	0.106
		Inspect	1.048	65.314	0.340		Inspect	0.426	73.787	0.181

change points become obviously larger than that of $\{X_{i1}\}_{i=1}^n$. This would explain why the dimension reduction versions work well.

In conclusion, the dimension reduction strategy could significantly improve the performances of the original methods. The more numerical studies with the central κ-th moment deviation subspace are put in Supplementary Materials.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

Table 7: Change in both the distribution and the covariance matrix in ***Experiment 3***

Case	p_z	a	Method	$\hat{k}$	MSE	RI	p_z	a	Method	$\hat{k}$	MSE	RI
1	65	0.6	E-Divisive$_{dr}$	4.264	0.325	0.976	20	0.6	E-Divisive$_{dr}$	4.109	0.323	0.951
			E-Divisive	0.555	12.761	0.337			E-Divisive	0.177	14.879	0.246
			ks-cp3o$_{dr}$	4.097	0.234	0.979			ks-cp3o$_{dr}$	4.712	2.364	0.948
			ks-cp3o	6.250	9.740	0.769			ks-cp3o	6.224	9.890	0.764
			KCP$_{dr}$	5.720	11.760	0.937			KCP$_{dr}$	5.740	12.220	0.903
			KCP	1.020	11.660	0.398			KCP	4.260	38.740	0.486
		0.8	E-Divisive$_{dr}$	4.154	0.181	0.988		0.8	E-Divisive$_{dr}$	4.092	0.111	0.984
			E-Divisive	4.023	0.083	0.980			E-Divisive	1.314	9.370	0.495
			ks-cp3o$_{dr}$	4.007	0.016	0.991			ks-cp3o$_{dr}$	4.043	0.077	0.985
			ks-cp3o	6.243	9.887	0.771			ks-cp3o	6.212	9.982	0.768
			KCP$_{dr}$	4.420	0.940	0.982			KCP$_{dr}$	4.600	1.800	0.985
			KCP	4.400	2.520	0.919			KCP	7.680	30.920	0.958
2	65	0.3	E-Divisive$_{dr}$	4.031	0.434	0.935	20	0.3	E-Divisive$_{dr}$	3.655	1.020	0.871
			E-Divisive	0.608	12.645	0.340			E-Divisive	0.313	14.100	0.281
			ks-cp3o$_{dr}$	4.719	3.047	0.928			ks-cp3o$_{dr}$	5.940	8.727	0.852
			ks-cp3o	5.958	9.503	0.754			ks-cp3o	6.510	11.697	0.779
			KCP$_{dr}$	5.807	11.367	0.905			KCP$_{dr}$	2.756	9.797	0.580
			KCP	0.000	16.000	0.198			KCP	0.092	15.513	0.217
		0.5	E-Divisive$_{dr}$	4.055	0.221	0.962		0.5	E-Divisive$_{dr}$	3.908	0.554	0.924
			E-Divisive	0.927	10.982	0.403			E-Divisive	0.497	13.135	0.316
			ks-cp3o$_{dr}$	4.308	1.096	0.958			ks-cp3o$_{dr}$	5.251	5.619	0.902
			ks-cp3o	6.453	10.889	0.775			ks-cp3o	6.020	8.745	0.761
			KCP$_{dr}$	6.372	15.325	0.950			KCP$_{dr}$	4.782	12.542	0.790
			KCP	0.039	15.799	0.206			KCP	0.334	15.463	0.245
3	65	0.3	E-Divisive$_{dr}$	3.530	1.537	0.897	20	0.3	E-Divisive$_{dr}$	3.086	2.246	0.831
			E-Divisive	0.570	12.643	0.365			E-Divisive	0.350	13.918	0.314
			ks-cp3o$_{dr}$	5.036	6.324	0.860			ks-cp3o$_{dr}$	5.911	8.515	0.787
			ks-cp3o	6.101	9.143	0.732			ks-cp3o	6.399	10.511	0.743
			KCP$_{dr}$	5.222	10.028	0.878			KCP$_{dr}$	2.985	10.128	0.622
			KCP	0.000	16.000	0.236			KCP	0.247	15.540	0.272
		0.5	E-Divisive$_{dr}$	3.827	0.480	0.956		0.5	E-Divisive$_{dr}$	3.394	1.380	0.900
			E-Divisive	0.750	11.679	0.399			E-Divisive	0.483	13.138	0.340
			ks-cp3o$_{dr}$	4.301	2.047	0.932			ks-cp3o$_{dr}$	5.367	7.347	0.840
			ks-cp3o	6.574	11.032	0.751			ks-cp3o	6.222	9.470	0.731
			KCP$_{dr}$	6.055	13.001	0.946			KCP$_{dr}$	4.833	13.104	0.807
			KCP	0.033	15.809	0.245			KCP	0.554	13.914	0.323

4.2 Experiment on clustering

Consider the data with clusters and compare two popularly used clustering methods: the K-means method (K-means) and the density-based spatial clustering with noise method (DBSCAN) with their Iterative Subspace Clustering (ISC) algorithms proposed in this paper. Their ISC versions are

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

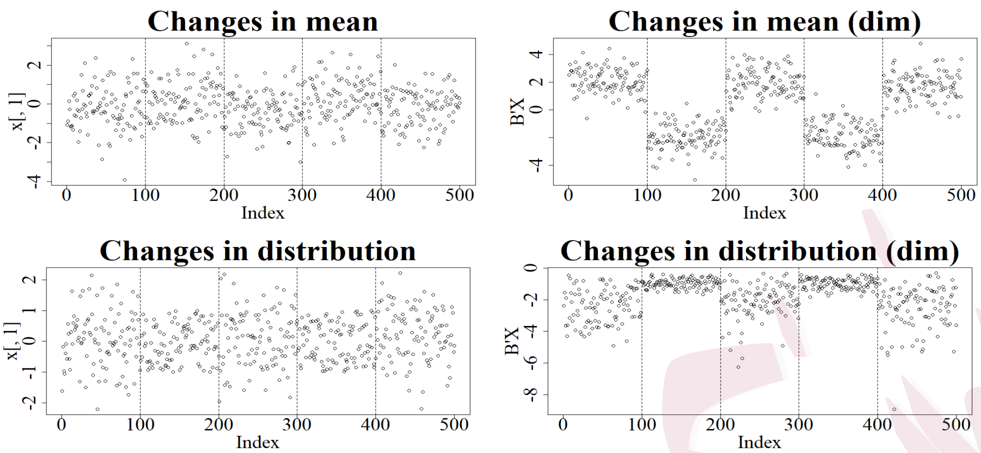

Figure 1: Scatter plots before and after dimension reduction, the above two figures correspond to the dense mean change points with $p = 100$ and $u = 0.2$ in ***Experiment 1*** and the below two figures to the Case 3 with $p = 10$, $p_Z = 65$ and $a = 0.8$ in ***Experiment 3***.

written as ISC$_{K-means}$ and ISC$_{DBSCAN}$, respectively. By optimizing the objective function in (3.3), we can have the lower-dimensional data $\{B_n^\top Z_i\}_{i=1}^n$. The corresponding methods are written as K-means$_{dr}$ and DBSCAN$_{dr}$. We still adopt the RI to measure the similarity between the underlying clusters and estimated clusters to evaluate the performances. We conduct experiments on both balanced and imbalanced datasets with three categories: (1) the balanced dataset has the same sample sizes $n_1 = n_2 = n_3 = n/3$; (2) the imbalanced dataset has sample sizes $n_1 = 300$, $n_2 = 200$, and $n_3 = 100$. The data are generated from the following settings:

- Case 1: (Distance-based example) The kth category is from the multi-dimensional normal distribution $N(a_k \mathrm{I}_p, \sigma^2 \mathrm{I}_{p\times p})$ with $\sigma = 0.5$, $a_k = k$, for $k = 1, 2, 3$, where I_p denotes is an all-one vector with dimension $p = 50, 100$.

- Case 2: (Bull's eye example) The kth category contains $\{X_{k,i}\}_{i=1}^{n_k}$ with $X_{k,i} = \sigma_{k,i} w_{k,i}$, for $k = 1, 2, 3$ and $i = 1, 2, \cdots, n_k$, where $\sigma_{k,i}$ is from the uniform distribution on the regions $[2k-2, 2k-1]$ and $w_{k,i}$ is from the uniform distribution on the unit sphere $\mathbb{S}^p$. Here $p = 5, 10$ corresponding to $p_Z = 20, 65$, respectively.

To make the comparison fairly, we also transform the original data in the Bull's eye example based on the formula in (3.1) and then adapt the methods to cluster the data $\{Z_i\}_{i=1}^n$, which are written as K-means(z) and DBSCAN(z), respectively. The mean and *sd* denote the mean and standard deviation of the RI, respectively. From the results reported in Tables 8 and 9, we can observe that the dimension reduction-based versions significantly outperform their original versions of the methods. Three clustering methods for the original data perform the worst in these examples. Further, the iterative algorithms enhance their performances.

To show the results visually, we plot the first two dimensions of the data in the distance-based example with $p = 50$ and the bull's eye example with $p_Z = 20$. Figure 2 shows the scatter plots of $\{\tilde{B}_n^\top Z_i\}_{i=1}^n$ with $\tilde{B}_n$ being the eigenvectors associated with the largest two eigenvalues of $\Delta_{Z,n}$.

It is observed that the three categories of $\{\tilde{B}_n^\top Z_i\}_{i=1}^n$ can be clearly distinguished. This reveals the reason why our proposed iterative subspace clustering method performs much better than the original K-means.

To check whether the algorithm converges empirically, we present the convergence of our algorithm based on synthetic data. Based on ISC$_{K-means}$, we compute the $||M_n^{(k+1)} - M_n^{(k)}||_F$ at each iteration step, and exhibit the plots of $||M_n^{(k+1)} - M_n^{(k)}||_F$ in the Distance-based example with $p = 50$ and Bull's eye example with $p_Z = 20$ in Figure 3. From Figure 3, we observe that $||M_n^{(k+1)} - M_n^{(k)}||_F$ suggests a downward trend and quickly goes to 0 by less than 5 iterations. Therefore, the iterative algorithm could converge.

4.3 Real data examples

In this subsection, we illustrate the applications of the proposed methods to three real data sets. To save space, the analysis of Genetics data is put in Supplementary Materials.

4.3.1 Financial data with mean and variance changes

Consider the data set on the log-returns of the daily closing price of all constituent stocks of the Standard and Poor's 100 (S&P100) index. This data set is from Yahoo Finance, covering the period from July 1st, 2019, to

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

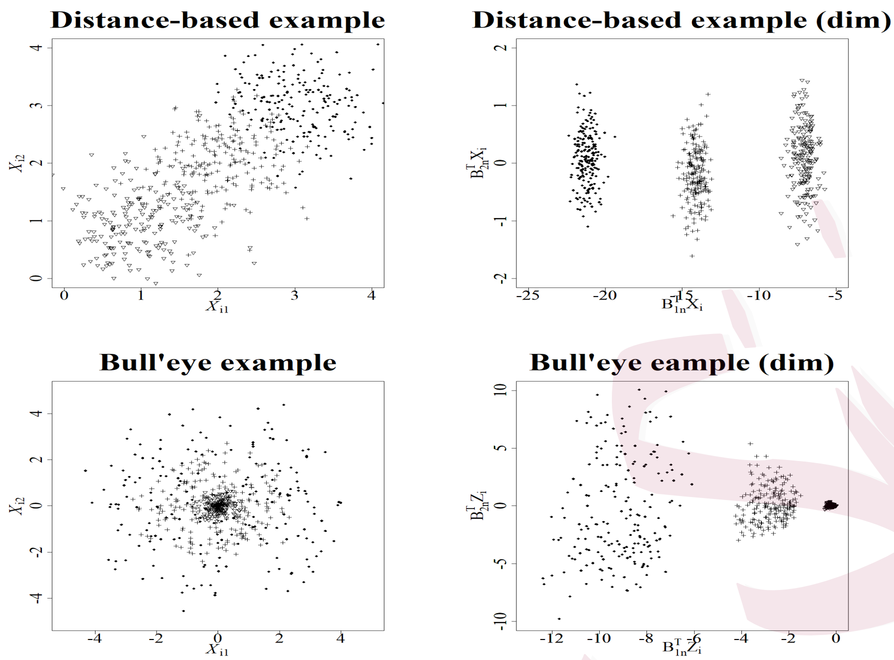

Figure 2: Scatter plots before and after dimension reduction, the two left figures correspond to the first two dimensions of the distance-based example with $p = 50$ and the bull's eye example with $p_Z = 20$, respectively. The right two figures correspond to the first two dimensions of the $\{B_n^\top X_i\}_{i=1}^n$ and $\{B_n^\top Z_i\}_{i=1}^n$ under the distance-based example with $p = 50$ and the bull's eye example with $p_Z = 20$, respectively.

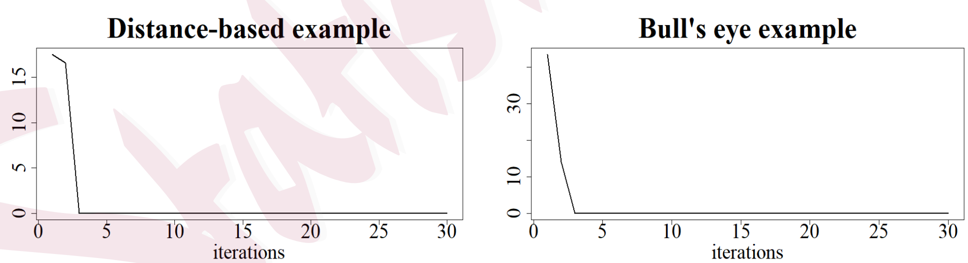

Figure 3: $||M_n^{(k+1)} - M_n^{(k)}||_F$ at each iteration

July 1st, 2020. After cleaning the stocks with missing values, there are 80 constituent stocks, namely $p = 80$, with the sample size $n = 254$. We first

Table 8: The clustering results of balanced data with $n_1 = n_2 = n_3 = 200$

Distance-based example							
p	Method	mean	sd	p	Method	mean	sd
50	ISC_{DBSCAN}	0.994	0.003	100	ISC_{DBSCAN}	0.994	0.003
	DBSCAN	0.332	0.000		DBSCAN	0.332	0.000
	DBSCAN_{dr}	0.839	0.015		DBSCAN_{dr}	0.819	0.017
	$\text{ISC}_{K-means}$	0.976	0.078		$\text{ISC}_{K-means}$	0.971	0.085
	K-means	0.929	0.121		K-means	0.923	0.125
	K-means_{dr}	0.916	0.127		K-means_{dr}	0.916	0.128
Bull's eye example							
p_Z	Method	mean	sd	p_Z	Method	mean	sd
20	ISC_{DBSCAN}	0.967	0.067	65	ISC_{DBSCAN}	0.993	0.005
	DBSCAN	0.428	0.014		DBSCAN	0.401	0.011
	DBSCAN_{dr}	0.801	0.009		DBSCAN_{dr}	0.897	0.017
	DBSCAN(z)	0.421	0.013		DBSCAN(z)	0.399	0.011
	$\text{ISC}_{K-means}$	0.849	0.136		$\text{ISC}_{K-means}$	0.898	0.131
	K-means	0.581	0.011		K-means	0.564	0.012
	K-means_{dr}	0.795	0.012		K-means_{dr}	0.841	0.133
	K-means(z)	0.729	0.136		K-means(z)	0.733	0.009

detect mean changes in the data structure.

As, on the whole, E-Divisive_{dr} and SBS_{dr} perform better than the others in the previous simulation studies, we then adopt the two methods. The dimension q is determined to be 1 using the TRR criterion in (2.5). E-Divisive_{dr} detects a change at the location $t = 164$ on February 20, 2020. This identification seems reasonable as the outbreak of the COVID-19 epidemic led to a serious economic downturn after February 2020. For comparison, E-Divisive detects two change points at $t = 164, 194$, but no other economic events appear to be occurring around $t = 194$. SBS_{dr} identifies

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

Table 9: The clustering results of imbalanced data with $n_1 = 300$, $n_2 = 200$,$n_3 = 100$

Distance-based example							
p	Method	mean	sd	p	Method	mean	sd
50	ISC_{DBSCAN}	0.993	0.003	100	ISC_{DBSCAN}	0.994	0.003
	DBSCAN	0.388	0.000		DBSCAN	0.388	0.000
	DBSCAN_{dr}	0.842	0.017		DBSCAN_{dr}	0.823	0.018
	$\text{ISC}_{K-means}$	0.953	0.097		$\text{ISC}_{K-means}$	0.950	0.097
	K-means	0.919	0.112		K-means	0.917	0.117
	K-means_{dr}	0.905	0.119		K-means_{dr}	0.903	0.121
Bull's eye example							
p_Z	Method	mean	sd	p_Z	Method	mean	sd
20	ISC_{DBSCAN}	0.982	0.006	65	ISC_{DBSCAN}	0.991	0.005
	DBSCAN	0.450	0.011		DBSCAN	0.422	0.007
	DBSCAN_{dr}	0.867	0.004		DBSCAN_{dr}	0.927	0.012
	DBSCAN(z)	0.442	0.010		DBSCAN(z)	0.420	0.006
	$\text{ISC}_{K-means}$	0.880	0.163		$\text{ISC}_{K-means}$	0.938	0.132
	K-means	0.645	0.016		K-means	0.624	0.016
	K-means_{dr}	0.807	0.170		K-means_{dr}	0.865	0.167
	K-means(z)	0.654	0.014		K-means(z)	0.655	0.003

two change points at $t = 171, 181$. Because the time points $t = 171, 181$ are close, both could be viewed as the same change attributed to the COVID-19 epidemic. SBS does not detect any change points.

We further detect changes in the contemporary mean and second-order moment structures. Hence we set $\kappa = 2$. To apply our method efficiently, we choose ten stocks with relatively large changes from the original data. Then $p = 10$ and $p_Z = 65$. We also, via the TRR criterion, found $\hat{q}_\kappa = 1$. The change is also at $t = 164$, the same location detected in the mean structure by E-Divisive$_{dr}$. To further visualize the change at this date,

Figure 4 presents the scatter plots of the lower-dimensional data $\{B_n^\top X_i\}_{i=1}^n$ and $\{B_n^\top Z_i\}_{i=1}^n$. It is observed that both the contemporaneous mean and second-order moment structures should have changed at $t = 164$. In the second-order moment structures, E-Divisive detects two change points at $t = 164, 194$.

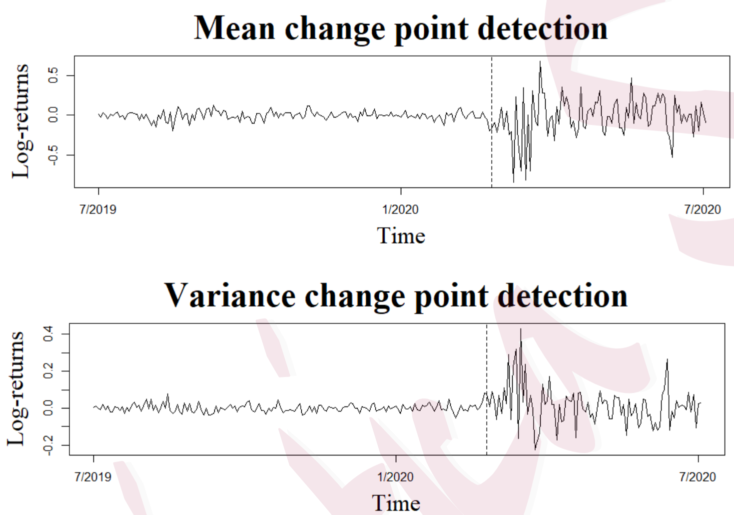

Figure 4: Change point detection after dimension reduction for S&P100 data, the top figure describes detecting the changes in the mean, and the bottom figure presents breaks in the contemporaneous mean and second-order moment structures.

4.3.2 Iris data with clusters

Consider this classical dataset for clustering using the proposed iterative algorithm; see the UCI database. The Iris dataset consists of $n = 150$ samples with $p = 4$ attributes, including sepal length, sepal width, and

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

Table 10: The results of clustering data

Method	RI	Method	RI
ISC$_{DBSCAN}$	0.702	ISC$_{K-means}$	0.887
DBSCAN	0.431	K-means	0.815
DBSCAN$_{dr}$	0.475	K-means$_{dr}$	0.831

petal width. The dataset contains three species of Iris, which are Setosa, Versicolour, and Virginia, respectively. Thus, we cluster the real dataset into three categories. Wang [2010] also analyzed it for clustering.

Table 10 reports the RI and the accuracy of three estimations. Since the results of K-means depend on the selection of initial value points, the result of each experiment may be different; we then repeat the experiment 50 times to have an average. It is easy to observe that ISC$_{Kmeans}$ performs the best. K-means$_{dr}$ can also improve the K-means' accuracy. DBSCAN exhibits improvement after employing the ISC method.

5. Conclusion

In this paper, we propose the notion of moment deviation subspaces and analyze the estimation for the subspaces. This can reduce the dimension of high-dimensional data such that we can efficiently work on them in the lower-dimensional spaces without losing any information. We developed a novel method combining the Mahalanobis matrix and the covariance matrix

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0384

to identify the effective dimension reduction spaces for unsupervised dimension reduction. We then apply this new strategy to changes and clustering in the data structure.

This generic method could apply to other types of high-dimensional data, such as panel data (see, e.g., Düker [2022]) and tensor data (see, e.g., Huang et al. [2022]). In addition, our approach could also be extended to deal with more general models than moment changes. For example, it might detect change points in the more general class of parameters (see, e.g., Dette and Gösmann [2020]) such as parametric distribution, parametric, and semiparametric regression models. Under certain regularity conditions, this might also be used to handle the change point detection problem in ultra-high-dimensional data when sparsity exists in the data structure, as Wang and Samworth [2018] considered. But this may need to combine some penalization approaches in the dimension reduction procedure. The research is ongoing. Another issue is extending the method to change point detection of online data. The current approach has a limitation: only the offline data can be handled. For more general paradigms, it deserves further study.

Supplementary material

In the online supplementary material, we discuss the situation of changes in the covariance matrix. This supplementary material also contains part of numerical studies and all proofs of the theoretical results.

Acknowledgement

The authors thank the editor, the associate editor and two referees for their constructive suggestions that significantly improved an early manuscript. 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 while preparing this research.

References

S. Arlot, A. Celisse, and Z. Harchaoui. A kernel multiple change-point algorithm via model selection. *Journal of Machine Learning Research*, 20:1–56, 2019.

H. Cho. Change-point detection in panel data via double cusum statistic. *Electronic Journal of Statistics*, 10(2):2000–2038, 2016.

H. Cho and P. Fryzlewicz. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):475–507, 2015.

R. D. Cook and S. Weisberg. Discussion of sliced inverse regression for dimension reduction, by K. C. Li. *Journal of the American Statistical Association*, 86(414):328–332, 1991.

H. Dette and J. Gösmann. A likelihood ratio approach to sequential change point detection for a general class of parameters. *Journal of the American Statistical Association*, 115(531): 1361–1377, 2020.

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.

M.-C. Düker. Detection of multiple change-points in high-dimensional panel data with cross-sectional and temporal dependence. 2022.

F. Enikeeva and Z. Harchaoui. High-dimensional change-point detection under sparse alternatives. *Annals of Statistics*, 47(4):2051–2079, 2019.

P. Fryzlewicz. Wild binary segmentation for multiple change-point detection. *Annals of Statistics*, 42(6):2243–2281, 2014.

T. Grundy, R. Killick, and G. Mihaylov. High-dimensional changepoint detection via a geometrically inspired mapping. *Statistics and Computing*, 30(4):1155–1166, 2020.

J. Huang, J. Wang, X. Zhu, and L. Zhu. Two ridge ratio criteria for multiple change point detection in tensors. 2022.

M. Jirak. Uniform change point tests in high dimension. *Annals of Statistics*, 43(6):2451–2483, 2015.

K. C. Li. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327, 1991.

Q. Lin, Z. Zhao, and J. S. Liu. Sparse sliced inverse regression via lasso. *Journal of the American Statistical Association*, 114(528):1726–1739, 2019.

A. Lung-Yut-Fong, C. Lévy-Leduc, and O. Cappé. Homogeneity and change-point detection tests for multivariate data using rank statistics. *Journal de la Société Française de Statistique*, 154(4):133–162, 2015.

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

W. Qian, S. Ding, and R. D. Cook. Sparse minimum discrepancy approach to sufficient dimension reduction with simultaneous variable selection in ultrahigh dimension. *Journal of the American Statistical Association*, 114(527):1277–1290, 2019.

W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

Junhui Wang. Consistent selection of the number of clusters via crossvalidation. *Biometrika*, 97(4):893–904, 2010.

R. Wang, C. Zhu, S. Volgushev, and X. Shao. Inference for change points in high-dimensional data via selfnormalization. *The Annals of Statistics*, 50(2):781–806, 2022.

T. Wang and R. J. Samworth. High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(1):57–83, 2018.

T. Wang, M. Chen, H. Zhao, and L. Zhu. Estimating a sparse reduction for general regression in high dimensions. *Statistics and Computing*, 28(1):33–46, 2018.

Y. Xia, H. Tong, W. Li, and L. Zhu. An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):363–410, 2002.

S. Xiang, F. Nie, and C. Zhang. Learning a mahalanobis distance metric for data clustering and classification. *Pattern Recognition*, 41(12):3600–3612, 2008.

W. Zhang, N. A. James, and D. S. Matteson. Pruning and nonparametric multiple change point detection. *IEEE International Conference on Data Mining Workshops*, pages 288–295, 2017.

L. Zhu, L. Zhu, T. Wang, and L. Ferré. Sufficient dimension reduction through discretization-expectation estimation. *Biometrika*, 97(2):295–304, 2010.

X. Zhu, X. Guo, T. Wang, and L. Zhu. Dimensionality determination: A thresholding double ridge ratio approach. *Computational Statistics and Data Analysis*, 146:106910, 2020a.

Xuehu Zhu, Yue Kang, and Junmin Liu. Estimation of the number of endmembers via thresholding ridge ratio criterion. *IEEE Transactions on Geoscience and Remote Sensing*, 58(1): 637–649, 2020b.