Computational Statistics and Data Analysis 178 (2023) 107619

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

Novel multiplier bootstrap tests for high-dimensional data with applications to MANOVA

Nilanjan Chakraborty, Lyudmila Sakhanenko *

Department of Statistics and Probability, Michigan State University, 619 Red Cedar Rd, East Lansing, 48824, USA

ARTICLE INFO

Article history:
Received 20 December 2021
Received in revised form 6 September 2022
Accepted 8 September 2022
Available online 14 September 2022

Keywords:
Bootstrap
MANOVA
GLHT

ABSTRACT

New bootstrap tests are proposed for linear hypotheses testing of high-dimensional means. In particular, they handle multiple-sample one- and two-way MANOVA tests with unequal cell sizes and unequal unknown cell covariances, as well as contrast tests in elegant and unified way. New tests are compared theoretically and on simulations studies with existing popular contemporary tests. They enjoy consistency, computational efficiency, very mild moment/tail conditions. They avoid the estimation of correlation or precision matrices, and allow the dimension to grow with sample size exponentially. Additionally, they allow the number of groups and the sparsity to grow with the sample size exponentially, thus broadening their applicability.

1. Introduction

There are K independent groups of random vectors-columns $V_{k,i} \in \mathbb{R}^p, i = 1, \dots, n, k = 1, \dots, K$, drawn from K populations with means $\mu_1, \dots, \mu_K \in \mathbb{R}^p$. We are interested in generalized linear hypothesis testing (GLHT) for the group means, which includes MANOVA and contrast tests as particular cases.

These vectors do not have to come from the same distribution. Also we can let the dimension p grow with n or the number of groups K grow with n or both. This is quite rare in the literature but it is a very useful setup in practice. Santo and Zhong (2021) argued that functional data can be viewed this way. We adopt this viewpoint in the real data example later in this paper. This setup was recently used in Lin et al. (2021).

Let $[V_{k,i}]_q$ denote the q-th entry of the i-th observation from the k-th group, where $q = 1, \dots, p; i = 1, \dots, n; k = 1, \dots, K$. Given a class of non-singular matrices $A^{(l)} \in \mathbb{R}^{Kp \times Kp}, l = 1, \dots, L$ with components $A^{(l)}_{jj'}, j, j' = 1, \dots, Kp$, we propose the test statistics of the form

$$T_n = \max_{l=1,\dots,L} \max_{j=1,\dots,Kp} n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{q=1}^{p} A^{(l)}_{j(q+(k-1)p)} [V_{k,i}]_q.$$

Intuitively, in T_n short p-column-vectors $V_{1,i}, \dots, V_{K,i}$ are stacked into long Kp-vectors that are then multiplied by matrices $A^{(l)}$, averaged over i, normalized and finally maximized over the components of the resulting long vector and over the class of matrices (over l).

* Corresponding author.
E-mail addresses: chakra46@msu.edu (N. Chakraborty), sakhanen@msu.edu (L. Sakhanenko).

https://doi.org/10.1016/j.csda.2022.107619
0167-9473/

Next, introduce the multiplier bootstrapped form of the test statistics

$$T_n^e = \max_{l=1,\dots,L} \max_{j=1,\dots,Kp} n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{q=1}^{p} A^{(l)}_{j(q+(k-1)p)} e_i [V_{k,i} - \bar{V}_k]_q,$$

where a vector $e = (e_1, \dots, e_n)$ of i.i.d. $N(0,1)$ random variables is independent of all $V_{k,i}$'s and $\bar{V}_k = \frac{1}{n}\sum_{i=1}^{n} V_{k,i}, k = 1, \dots, K$, are the groupwise averages. Intuitively, the symmetrized sums would have distribution that approximates the distribution of the original test statistics under the null hypothesis that all the group means are equal (MANOVA) or are related linearly (GLHT).

The proposed bootstrap test at the significance level $\alpha \in (0,1)$ rejects if

$$T_n \geq Q_\alpha := \inf\{u \in \mathbb{R} : P_e(T_n^e \leq u) \geq 1 - \alpha\},$$

where P_e stands for the probability with respect to the Gaussian vector e only.

We show that these bootstrap tests have excellent level and power performance for MANOVA and GLHT in high dimensional setup including contrasts tests. Our approach produces a whole class of tests that are flexible and can be tuned to many different scenarios. Indeed, the class of matrices $A^{(l)}, l = 1, \dots, L$, serves as a tuning mechanism, where this class can be chosen to maximize components of $A^{(l)}[\mu_1, \dots, \mu_K]$ under alternative in order to gain good empirical power. Our tests do not need sparsity assumptions on the means and/or covariances. We require the moment and tail assumptions on the distributions that are commonly used and at times less stringent than those in the existing literature. Our approach allows data from multivariate sub-exponential distributions, some heavy-tailed distributions, skewed distributions, and thus, it broadens the regime of its practical use.

The high dimensional MANOVA problem for testing $H_0 : \mu_1 = \cdots = \mu_K$ for $K > 2$ has been a focus of many recent works due to its growing importance in genomics, econometrics, and neuroimaging among many other fields of science. For example, Fujikoshi et al. (2004) considered the ratio of the traces of between-sample covariance and within-sample covariance. Meanwhile, Schott (2007) proposed a test based on the difference of those two traces. Srivastava (2007) used the Moore-Penrose inverse of the within-sample covariance matrix to construct a test. Cai and Xia (2014) proposed a test based on the maximum-norm of the squared differences between K groups. All mentioned tests either have been formulated under the assumption that the data is generated from a multivariate normal population or under some stringent distributional or sparsity assumptions. Moreover, all these tests assume equal covariance structure among all the groups. Recently Chen et al. (2019) proposed a thresholded L_2-norm-type statistics assuming sparsity in means, mixing, and multivariate sub-Gaussianity. They consider different sparse covariance matrices across different groups. The sparsity assumptions on the means and covariances were very important and crucial in their work. In this work, we eliminate the need for these assumptions.

From a different point of view, this work also extends the recent work by Xue and Yao (2020) for $K = 2$ to the case $K > 2$. This extension is elegant and less technical than the direct reproving of the results in Xue and Yao (2020), where one would have to tackle intricate block-type dependency structures and work with U-statistics similar to what was done in Chen (2018). The class of our tests enjoys all the good properties of the tests in Xue and Yao (2020). In particular, our tests are computationally fast and simple, since they do not require the estimation of covariance and/or precision matrices. Our tests do not need for the sample size ratios between groups to converge, they can merely stay bounded in an interval. Our tests have one extra advantage of being versatile, so they can be just as easily adopted to solve MANOVA or to test for a linear structure on the means such as contrasts. Even for a scenario with sparse means, our tests have comparable power performance than more complicated and computationally slower tests by Chen et al. (2019) and Cai and Xia (2014), which are specifically designed for this scenario.

The technical basis for our tests lies in the introduction of the class of sparse convex sets which are intersections of a finite number of half-spaces in the context of investigation of quality of Gaussian approximations for sums. This class generalizes hyper-rectangles and half-spaces that were proposed by Chernozhukov et al. (2017). Unlike previous works our tests have tractable dependence on sparsity specifier. The details are given in the Appendix.

The rest of the paper is organized as follows. Section 2 contains main theoretical results devoted to the statistical applications: one- and two-way MANOVA and GLHT in high-dimensional setup. Section 3 establishes the connection with other recent tests that are related to our approach. In section 4 we perform detailed high-dimensional simulation study where we compare our method with two existing competitors by Chen et al. (2019) and Cai and Xia (2014). We consider various distributions and sparsity scenarios. We illustrate the practicality of our test on a real data example about fish shape comparisons in section 5, which is followed by a discussion with a conclusion in section 6. Technical conditions, details, and proofs are gathered in the Appendix.

2. Main results

We obtain different rates of approximation of test statistic T_n by its bootstrapped counterpart T_n^e under various moment and tail conditions on Vs and conditions on K and L. Those approximations allow us to study theoretically the properties of our tests. These conditions are quite technical and given in the Appendix.

Roughly speaking for a particular case of i.i.d. $V_{k,i}, i = 1, \dots, n$, for each group $k = 1, \dots, K$, these conditions allow $\log(LKp)$ to grow with the sample size n as

- $o(n^{1/3}(\log n)^{-2/3})$ for bounded random variables;
- $o(n^{1/4})$ for some sub-Gaussian random variables with finite 4-th moments;
- $o(n^{1/4}(\log n)^{-1/2})$ when both $V_{k,i}$ and $\max_{k,j}[V_{k,i}]_j$ have finite 4-th moments;
- $o(n^{1/3}(\log n)^{-2/3} \vee n^{1/2-1/q}(\log n)^{-1/2})$ when $V_{k,i}$ have finite 4-th moments and $\max_{k,j}[V_{k,i}]_j$ have finite q-th moments for some $q > 4$;
- $o(n^{1/5})$ for some random variables with exponential tails and finite 4-th moments;
- $o(n^{1/5} \vee n^{(q-2)/(3q-2)})$ when $V_{k,i}$ have finite 4-th moments and $\max_{k,j}[V_{k,i}]_j$ have finite q-th moments for some $q > 2$.

The case of non-identical $V_{k,i}$ is also allowed. In this case the average 4-th moments are allowed to grow as n^{α} for some $\alpha < 1/3$ or $1/4$ or $1/5$, but then the rate of growth for $\log(LKp)$ would need to grow slower than the rates listed above for i.i.d. cases. The details are explained in the Appendix.

Meanwhile, we also impose conditions on the class of matrices:

(A1) $\sum_{m=1}^{d} |A^{(l)}_{jm}| \le s \ \forall j = 1, \dots, d, \ \forall l = 1, \dots, L$;

(A2) $\sum_{k=1}^{K} A^{(l)}_{j(q+(k-1)p)} = 0 \ \forall j = 1, \dots, d \ \forall q = 1, \dots, p \ \forall l = 1, \dots, L$;

(A3) $\min_{1 \le l \le L} \min_{1 \le j \le d} [A^{(l)} \bar{\Sigma} (A^{(l)})^T]_{jj} \ge b > 0$ for some fixed constant b, where $\bar{\Sigma}$ is defined as a $Kp \times Kp$ block matrix of the covariances

$$\frac{1}{n} \sum_{i=1}^{n} Cov(V_{k_1,i}, V_{k_2,i}),$$

stacked by varying $k_1, k_2 = 1, \dots, K$.

Condition (A1) can be viewed as a sparsity control on matrices, where parameter s characterizes the amount of sparsity. We allow sparse regimes of just a few non-zero components and dense regimes of many near-zero components for matrices, while the underlying distribution does not need any sparsity assumptions. Note that if all covariance matrices are the same and $A^{(l)}$ is an identity matrix we recover condition (M.1) in Chernozhukov et al. (2017).

Condition (A2) comes from the null hypothesis discussed in the next subsection. Meanwhile, condition (A3) ensures that the smallest eigenvalues of $A^{(l)} \bar{\Sigma} (A^{(l)})^T$ are bounded away from 0. Note that these matrices are the covariance matrices of the averaged long Kp-vectors that serve as building blocks in the test statistics T_n.

We consider several linear hypotheses: MANOVA (balanced case), MANOVA (unbalanced case), two-way MANOVA, and contrasts.

2.1. Balanced MANOVA

We are interested in the hypotheses

$$H_0 : \mu_1 = \dots = \mu_K \text{ vs } H_A : \text{ otherwise.}$$

For the i-th vector in the k-th group $V_{k,i}$ we denote its components by $[V_{k,i}]_q, q = 1, \dots, p$. Recall that we propose the test statistic

$$T_n = \max_{l=1,\dots,L} \max_{j=1,\dots,Kp} n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{q=1}^{p} A^{(l)}_{j(q+(k-1)p)} [V_{k,i}]_q,$$

where each matrix $A^{(l)}$ with components $A^{(l)}_{jj'}$, $j, j' = 1, \dots, Kp$, satisfies conditions (A1)-(A3). Recall that the bootstrapped version of the test statistics is

$$T^e_n = \max_{l=1,\dots,L} \max_{j=1,\dots,Kp} n^{-1/2} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{q=1}^{p} A^{(l)}_{j(q+(k-1)p)} e_i [V_{k,i} - \bar{V}_k]_q.$$

Also recall that

$$Q_\alpha = \inf\{u \in \mathbb{R} : P_e(T^e_n \le u) \ge 1 - \alpha\}, \ \alpha \in (0, 1).$$

Under certain moment and tail conditions on V_is, say (C), given in the Appendix we obtain consistency of the tests. To formulate it, we stack group mean vectors $\mu_k \in \mathbb{R}^p, k = 1, \dots, K$, into a long vector $\mu \in \mathbb{R}^{Kp}$.

Theorem 1. *Under (A1)-(A3) and (C) we have as* $n \to \infty$

$$P(T_n \geq Q_\alpha | H_0) \to \alpha$$

and

$$P(T_n \geq Q_\alpha | H_A) \to 1,$$

provided that there exist $j = 1, \ldots, Kp$ *and* $l = 1, \ldots, L$ *such that the* j*-th component of* $A^{(l)}\mu$ *is non-zero:* $[A^{(l)}\mu]_j \neq 0$.

We also remark that p-vectors $V_{k,i}, i = 1, \ldots, n$, can come from different distributions with the same mean vector μ_k. Our test can accommodate the non-identically-distributed scenario which is rare in the existing literature.

Next, in order to assess the local power of the test, we consider a class of contiguous alternatives which converge to the null hypothesis as $n \to \infty$. Let

$$H_A^{(n)} : \mu_1, \ldots, \mu_K \in \mathbb{R}^p \ : \min_{l=1,\ldots,L} \min_{j=1,\ldots,Kp} [A^{(l)}\mu]_j \geq c_n n^{-1/2},$$

where the sequence $c_n \to \infty$ diverges slowly as $n \to \infty$. Recall that Kp is allowed to grow with the sample size n.

Corollary 1. *Suppose conditions of Theorem 1 are satisfied. With probability tending to 1*

$$P(T_n \geq Q_\alpha | H_A^{(n)}) \to 1 \text{ as } n \to \infty.$$

This corollary indicates that the proposed test would successfully reject alternatives that are quite close to the null hypothesis by as little as almost a factor of $n^{-1/2}$, which is almost a parametric rate (up to c_n).

2.2. Unbalanced MANOVA

Suppose that now samples have different sample sizes $n_k, k = 1, \ldots, K$. Introduce a modified test statistic

$$\widetilde{T}_n = \max_{l=1,\ldots,L} \max_{j=1,\ldots,Kp} \sum_{t=1}^{K} \sum_{i=1}^{n_t} \sum_{q=1}^{p} n_t^{-1/2} A^{(l)}_{j(q+(t-1)p)} [V_{t,i}]_q.$$

Then introduce its bootstrapped version and quantile as

$$\widetilde{T}_n^e = \max_{l=1,\ldots,L} \max_{j=1,\ldots,Kp} \sum_{t=1}^{K} \sum_{i=1}^{n_t} \sum_{q=1}^{p} n_t^{-1/2} A^{(l)}_{j(q+(t-1)p)} e_i [V_{t,i} - \bar{V}_t]_q$$

and

$$\widetilde{Q_\alpha} = \inf\{u \in \mathbb{R} : P_e(\widetilde{T}_n^e \leq u) \geq 1 - \alpha\}, \ \alpha \in (0, 1).$$

With a slight abuse of notation denote $n = \min_{1 \leq k \leq K} n_k$, which is heuristically the effective sample size. Renumber groups so that the first group has the smallest sample size. Assume

$$(D) \quad \frac{n_k}{n} = \lambda_{k,n} \in [1, \infty), \ k = 2, \ldots, K,$$

where ratios $\lambda_{k,n}$ do not have to converge as $n \to \infty$ but should remain bounded.

We decompose $\widetilde{T}_n$ into a version of T_n by grouping terms into blocks of the same size n as follows

$$\begin{aligned}
\widetilde{T}_n &= \max_{l=1,\ldots,L} \max_{j=1,\ldots,Kp} \sum_{t=1}^{K} \sum_{q=1}^{p} \Bigg[\sum_{i=1}^{n} n^{-1/2} \lambda_{t,n}^{-1/2} A^{(l)}_{j(q+(t-1)p)} [V_{t,i}]_q \\
&\quad + n^{-1/2} \lambda_{t,n}^{-1/2} \sum_{i=n+1}^{\lambda_{t,n} n} A^{(l)}_{j(q+(t-1)p)} [V_{t,i}]_q \Bigg] \\
&\approx \max_{l=1,\ldots,L} \max_{j=1,\ldots,Kp} \sum_{t=1}^{K} \sum_{q=1}^{p} \Bigg[\sum_{i=1}^{n} n^{-1/2} \lambda_{t,n}^{-1/2} A^{(l)}_{j(q+(t-1)p)} [V_{t,i}]_q \\
&\quad + n^{-1/2} [\lambda_{t,n}]^{-1/2} \sum_{m=1}^{n} \sum_{r=1}^{[\lambda_{t,n}]-1} A^{(l)}_{j(q+(t-1)p)} [V_{t,n+(m-1)([\lambda_{t,n}]-1)+r}]_q \Bigg]
\end{aligned}$$

$$\approx \max_{l=1,\dots,L} \max_{j=1,\dots,Kp} \sum_{t=1}^{K}\sum_{q=1}^{p} n^{-1/2} \sum_{i=1}^{n} A^{(l)}_{j(q+(t-1)p)} \lambda_{t,n}^{1/2}$$
$$\times [\lambda_{t,n}]^{-1}\left[V_{t,i} + \sum_{r=1}^{[\lambda_{t,n}]-1} V_{t,n+(m-1)([\lambda_{t,n}]-1)+r}\right]_q,$$

where $\gamma_n \approx \delta_n$ means $\lim_{n\to\infty} \frac{\gamma_n}{\delta_n} = 1$ with probability 1. The approximation appeared due to use of integer parts of $\lambda_{t,n}, t = 1,\dots,K$. Define a new matrix $\widetilde{A^{(l)}}$ with components

$$\widetilde{A^{(l)}_{j(q+(t-1)p)}} = A^{(l)}_{j(q+(t-1)p)} \lambda_{t,n}^{1/2}.$$

Finally, define new variables as

$$\widetilde{V_{t,i}} = [\lambda_{t,n}]^{-1}\left[V_{t,i} + \sum_{r=1}^{[\lambda_{t,n}]-1} V_{t,n+(m-1)([\lambda_{t,n}]-1)+r}\right],\ t = 1,\dots,K, i = 1,\dots,n.$$

Note that $\mathbb{E}\widetilde{V_{t,i}} = \mu_t$ and the samples of these new variables are independent. Also note that $\widetilde{T_n}$ is an analogue of T_n with V replaced by $\widetilde{V}$. We remark that $\widetilde{V}_i, i = 1,\dots,n$, are independent but they will not be identically distributed even if the original observations V_i's are.

Corollary 2. *Suppose the conditions of Theorem 1 are satisfied for $\widetilde{V}_{t,i}, i = 1,\dots,n, t = 1,\dots,K$, and $\widetilde{A^{(l)}}, l = 1,\dots,L$. Moreover, assume condition (D) holds for $n_k, k = 2,\dots,K$. We have as $n \to \infty$*

$$P(\widetilde{T_n} \geq \widetilde{Q_\alpha} | H_0) \to \alpha$$

and

$$P(\widetilde{T_n} \geq \widetilde{Q_\alpha} | H_A) \to 1 \text{ provided } \exists j \exists l : [\widetilde{A^{(l)}}\mu]_j \neq 0.$$

This corollary establishes the consistency of the proposed test in the unbalanced MANOVA setup.

2.3. Two-way MANOVA

Consider the setup in Watanabe et al. (2020)

$$Y_{ijk} = \mu_0 + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, k \in \{1,\dots,N_{ij}\}, i = 1,\dots,I, j = 1,\dots,J,$$

where $\mu_0, \alpha_i, \beta_j, \gamma_{ij}$ are unknown $p \times 1$ vectors of parameters, while ε_{ijk} are mean zero $p \times 1$ random vectors with unknown covariances Σ_{ij}. For identifiability we are given a sequence of positive weights $w_{ij}, i = 1,\dots,I, j = 1,\dots,J$, so that $\sum_{i=1}^{I} w_{i\cdot}\alpha_i = 0$, $\sum_{j=1}^{J} w_{\cdot j}\beta_j = 0$, $\sum_{i=1}^{I} w_{ij}\gamma_{ij} = 0$, $\sum_{j=1}^{J} w_{ij}\gamma_{ij} = 0$, and $\sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}\gamma_{ij} = 0$, where $w_{i\cdot} = \sum_{j=1}^{J} w_{ij}$ and $w_{\cdot j} = \sum_{i=1}^{I} w_{ij}$.

Consider the null hypothesis

$$H_0 : \alpha_1 = \cdots = \alpha_I = 0.$$

Then look at the setup in the previous subsection with $K = I$ groups and the comparison of means $\mu_k = \mu_0 + \alpha_k$ when there are $n_k = \sum_{j=1}^{J} N_{kj}$ non-identically distributed observations $(V_{k,1},\dots,V_{k,n_k}) = (Y_{k11},\dots,Y_{k1N_{k1}}, Y_{k21},\dots, Y_{k2N_{k2}},\dots,Y_{kJ1},\dots,Y_{kJN_{kJ}})$. Then the test statistic $\widetilde{T_n}$ would be well-defined and the test rejects H_0 if $\widetilde{T_n} > \widetilde{Q_\alpha}$. Analogously, for the null hypothesis

$$H_0 : \beta_1 = \cdots = \beta_J = 0$$

consider $K = J$ groups and the comparison of means $\mu_k = \mu_0 + \beta_k$ when there are $n_k = \sum_{i=1}^{I} N_{ik}$ non-identically distributed observations $(V_{k,1},\dots,V_{k,n_k}) = (Y_{1k1},\dots,Y_{1kN_{1k}}, Y_{2k1},\dots,Y_{2kN_{2k}},\dots,Y_{Ik1},\dots,Y_{IkN_{Ik}})$. Then the test statistic $\widetilde{T_n}$ would be well-defined and the test rejects H_0 if $\widetilde{T_n} > \widetilde{Q_\alpha}$. Finally, consider the null hypothesis

$$H_0 : \gamma_{11} = \cdots = \gamma_{1J} = \cdots = \gamma_{IJ}.$$

This time one needs to look at $K = IJ$ groups and the comparison of means $\mu_0 + \alpha_i + \beta_j + \gamma_{ij}$ when there are n_{ij} observations $V_{k,m} = Y_{ijm}, k = i + (j-1)I, i = 1,\dots,I, j = 1,\dots,J$. Then the test statistic $\widetilde{T_n}$ would be well-defined and the test rejects H_0 if $\widetilde{T_n} > \widetilde{Q_\alpha}$.

Unlike L_2-based tests proposed by Watanabe et al. (2020), our tests do not need to estimate the unknown unequal covariances Σ_{ij} and do not need to compute computationally expensive estimator of the standard deviation of the test statistic. Thus, our tests are more computationally efficient than theirs.

Finally, we remark that it is straightforward to extend our framework to the 3-way, 4-way and so on, multi-level MANOVA setup especially given that we can let the number of groups K grow exponentially in n.

2.4. GLHT

Similar to Zhang et al. (2017), we are now interested in the hypotheses

$$H_0 : G\mu = 0 \text{ vs } H_A : \ G\mu \neq 0,$$

where G is a known $q \times Kp$ matrix of the rank $q < Kp$. This setup includes contrast tests and MANOVA. Note that $(GG^T)^{-1}$ exists.

Consider $Kp \times Kp$ matrices $A^{(l)}$ such that $A^{(l)} = M^{(l)}[G^T(GG^T)^{-1}G]^T$ for some non-singular $Kp \times Kp$ matrices $M^{(l)}$ for all $l = 1, \dots, L$. The expression in squared brackets is related to Moore-Penrose matrix inverse of G. This condition on $A^{(l)}$ replaces condition (A2). Under H_0 we have $A^{(l)}\mu = 0$, while under any alternative $A^{(l)}\mu \neq 0$. Then the test described above works for this setup, and Theorem 1 holds with (A2) replaced by this new structure of matrices $A^{(l)}$.

Let $(c_1, \dots, c_K)$ be a non-zero vector of constants in $\mathbb{R}^K$. As an illustrative example consider the following hypotheses

$$H_0 : c_1\mu_1 + \cdots c_K\mu_K = 0 \text{ vs } H_A : \ c_1\mu_1 + \cdots c_K\mu_K \neq 0.$$

In this case $G = (c_1\mathbb{I}_p, \dots, c_K\mathbb{I}_p)$ with $\mathbb{I}_p$ being a $p \times p$ identity matrix and let

$$A^{(l)} = \frac{M^{(l)}}{\|c\|_2^2} \begin{bmatrix} c_1^2\mathbb{I}_p & c_1c_2\mathbb{I}_p & \cdots & c_1c_K\mathbb{I}_p \\ c_1c_2\mathbb{I}_p & c_2^2\mathbb{I}_p & \cdots & c_2c_K\mathbb{I}_p \\ \cdots & \cdots & \cdots & \cdots \\ c_1c_K\mathbb{I}_p & c_2c_K\mathbb{I}_p & \cdots & c_K^2\mathbb{I}_p \end{bmatrix},$$

where $\|c\|_2^2 = \sum_{k=1}^K c_k^2$ and $M^{(l)}$ is an arbitrary non-degenerate $Kp \times Kp$ matrix. As in balanced MANOVA define test statistics as

$$\widetilde{T}_n = \max_{l=1,\dots,L} \max_{j=1,\dots,Kp} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{q=1}^p n_k^{-1/2} A^{(l)}_{j(q+(k-1)p)}[V_{k,i}]_q.$$

Then the bootstrapped test statistics are

$$\widetilde{T}_n^e = \max_{l=1,\dots,L} \max_{j=1,\dots,Kp} \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{q=1}^p n_k^{-1/2} A^{(l)}_{j(q+(k-1)p)} e_i[V_{k,i} - \bar{V}_k]_q.$$

Define quantiles of the bootstrapped distribution as

$$\widetilde{Q_\alpha} = \inf\{u \in \mathbb{R} : P_e(\widetilde{T}_n^e \leq u) \geq 1 - \alpha\}, \ \alpha \in (0, 1).$$

We reject H_0 if

$$\widetilde{T}_n \geq \widetilde{Q_\alpha}.$$

Since $A^{(l)}\mu = 0$ holds if and only if H_0 holds, then $P(\widetilde{T}_n \geq \widetilde{Q_\alpha}|H_0) \to \alpha$ and $P(\widetilde{T}_n \geq \widetilde{Q_\alpha}|H_A) \to 1$ provided that conditions of Theorem 1 are satisfied. This test allows to treat high-dimensional MANOVA and contrast tests in the unified framework unlike L_2-norm-based tests by Cai and Xia (2014) and Chen et al. (2019).

3. Connection with other tests

In this section we show that the tests introduced in Xue and Yao (2020) and those in Lin et al. (2021) fit into our framework, which increases its versatility.

3.1. 2-sample test by Xue and Yao (2020)

Consider the setup in Xue and Yao (2020). There are $K=2$ independent groups of random vectors $V_{i1}, V_{i2}, i=1,\dots,n$, drawn from 2 populations in $\mathbb{R}^{2p}$ with means $\mu_1, \mu_2 \in \mathbb{R}^p$. Their test statistics are the maximums of AS_n^X, where the matrix $A \in \mathbb{R}^{2p\times 2p}$ is tri-diagonal. It has 1 on the main diagonal and -1 on the diagonals that start from $a_{1(p+1)}$ and $a_{(p+1)1}$. Indeed, $X=(V_{i1}, V_{i2})^T \in \mathbb{R}^{2p}$ and

$$AX_i = ([V_{i1}-V_{i2}]_1, \dots, [V_{i1}-V_{i2}]_p, [V_{i2}-V_{i1}]_1, \dots, [V_{i2}-V_{i1}]_p)^T.$$

Then

$$AS_n^X = (S_n^{V_1} - S_n^{V_2}, S_n^{V_2} - S_n^{V_1})^T.$$

Therefore, their test statistic is

$$\|S_n^{V_1} - S_n^{V_2}\|_\infty = \max_{q=1,\dots,p} |[S_n^{V_1} - S_n^{V_2}]_q| = \max_{j=1,\dots,2p} [AS_n^X]_j.$$

Note that condition (A2) holds for this A. Then $L=1$ and $\mathcal{A}$ consists of just A, while $s=2$. Their results can be viewed as the special case of our results. Note that their proofs are specially designed for $K=2$ and the direct generalization would be difficult. It would require intricate work with U-statistics in the same spirit as what is done in Chen (2018).

3.2. MANOVA for functional data by Lin et al. (2021)

Consider the setup in Lin et al. (2021). There are K independent groups of random vectors drawn from K populations with means $\mu_1, \dots, \mu_K \in \mathbb{R}^p$. The k-th group after centering consists of $Z_{k,1}, \dots, Z_{k,n}$ independent observations in $\mathbb{R}^p$. Then stack vectors $Z_{1,1}, \dots, Z_{K,1}$ into X_1, and so on, finally stack $Z_{1,n}, \dots, Z_{K,n}$ into X_n. Then $X_i \in \mathbb{R}^{Kp}$.

The test statistics in Lin et al. (2021) are the maximums of components of AS_n^X, where the rows of matrix A are $(0,\dots,0, \frac{1}{\sqrt{2}\sigma^\tau_{k,l,j}}, 0,\dots,0, -\frac{1}{\sqrt{2}\sigma^\tau_{k,l,j}}, 0,\dots,0)$, where $\tau \in [0,1)$ and the non-zero entries are in the positions $(k-1)p+j, (l-1)p+j$ for $1 \le j \le p$. Here

$$\sigma^2_{k,l,j} = 0.5Var(X_{1,(k-1)p+j}) + 0.5Var(X_{1,(l-1)p+j}), \tau \in [0,1).$$

Thus, A satisfies (A2). From the setup in Lin et al. (2021), the indices (k,l) belong to the set

$$(k,l) \in \mathcal{P} \subset \{(i_1,i_2) : 1 \le i_1 < i_2 \le K\}.$$

Their test statistic is of the form of our statistic with $\mathcal{A}$ that consists of A and $-A$. Then $L=2$ while $s = \frac{\sqrt{2}}{\min_{k,l,j}\sigma^\tau_{k,l,j}}$.

Note that our theorems provide at best the rates for Kolmogorov distance between the test statistic and its bootstrapped version of the order $n^{-1/2+\delta}\log n$ with respect to n with $\log(LKp) = o(n^\delta(\log n)^{-2/3})$ for $\delta \in (0, 1/6)$, while Lin et al. (2021) get rates of the order $n^{-1/2+\delta}$ for an arbitrary $\delta > 0$ with $Kp \le pe^{\sqrt{\log n}}$. Lin et al. (2021) attain this rate under a stringent requirement of a special structure of the matrices $\Sigma^{(i)}$ with many restrictions that are difficult to check in practice. Their conditions essentially reduce the high dimensional problem to a problem, where $p \approx n^{1/(\log n)^a} \vee (\log n)^3$ with $a \in (0, 0.5)$, whereas our tests remain valid even when p is much larger.

4. Simulation studies

The purpose of this section is to compare with existing methods and to investigate the effect of the tuning specifiers $A^{(l)}, l=1,\dots,L$. We compare against 2 methods: Cai and Xia (2014) and Chen et al. (2019). However, both of these methods have limitations that our method does not have. Cai and Xia (2014) require complicated sparsity and boundedness conditions on the covariance and precision matrices, which are difficult to verify in practice. In particular, they assume that the sparsity of precision matrix is $o(n^{(1-q)/2}(\log p)^{-(3-q)/2})$ for some $q \in [0,1)$ in their Theorem 4, which addresses the case of unknown covariance. Their algorithm requires estimation of the precision matrix in the high-dimensional setup. They also assume that group covariances are the same. Chen et al. (2019) require sparsity, α-mixing of components of the observations, and $\log p = o(n^{1/3})$ when the precision matrix is unknown. Their algorithm also requires the estimation of precision matrices in the high-dimensional setup. Both methods by Cai and Xia (2014) and by Chen et al. (2019) are less efficient computationally than our tests largely due to precision/correlation matrices estimation.

We also stress that all compared tests have rather varied empirical levels that are quite far from the nominal level of 0.05. Therefore we employed the two methods described in Lloyd (2005) to adjust the powers for sizes for a fair comparison.

4.1. Four group MANOVA: tuning A

The goal of the simulations in this subsection is to tune the class of matrices $A^{(l)}$ and to demonstrate stability of the test for different classes of matrices. To this end, we are using a combination of setups in Zhang et al. (2017) and Cai and Xia (2014). We consider MANOVA for 4 groups for $V_{t,i} = \mu_t + \Gamma Z_{t,i}, t = 1, 2, 3, 4; i = 1, \dots, n_t$, where $Z_{t,i}$ are generated from one of 3 models such that the p components are i.i.d. standard normal $N(0, 1)$, standardized t-distributed with 4 degrees of freedom t_4, or normalized chi-squared of degree 1 χ^2_1, so they have mean 0 and variance 1. Three sets of samples sizes and 3 sets of means (for power comparisons) are considered. They are $\mathbf{n_1} = (25, 30, 40, 50)$, $\mathbf{n_2} = (50, 60, 80, 100)$, and $\mathbf{n_3} = (100, 120, 160, 200)$. Meanwhile, $\mu_1 = 0, \mu_2 = 1.5\delta\mathbf{h}$, $\mu_3 = \delta\mathbf{h}$, and $\mu_4 = 2\delta\mathbf{h}$, where a p-dimensional vector is $\mathbf{h} = (1, \dots, p)^T/\|(1, \dots, p)\|$ and a number δ varies between 0.4 and 2.6 depending on the set of the sample sizes $\mathbf{n}$ and the dimension p. Then as in Zhang et al. (2017) $\delta(\mathbf{n}, p) = [0.8, 1.9, 2.6; 0.5, 1.4, 1.8; 0.4, 0.9, 1.3]$. The larger δ represents the larger separation between the null hypothesis and the alternative. Following Zhang et al. (2017) we consider three choices for dimension p as 50, 500, and 1000.

Finally, the covariance matrix is $\Gamma\Gamma^T = (1-\rho)\mathbb{I}_p + \rho\mathbb{J}_p$, where $\mathbb{I}_p$ stands for the identity $p \times p$ matrix and $\mathbb{J}_p$ denotes $p \times p$ matrix of ones. The parameter ρ takes values 0.1, 0.5, 0.9. We also consider the covariance of the form $0.6^{|i-j|}, i, j, = 1, \dots, p$ as in Model 4 in Cai and Xia (2014); this case we denote by $\rho =$ NA. We use 10K bootstrap samples to obtain Q_α with $\alpha = 0.05$.

The choice of matrices $A^{(l)}, l = 1, \dots, L$, affects the performance of our test statistic. In the first simulation study we consider 5-diagonal matrices. Since the problem is invariant with respect to order of the datasets and with respect to the coordinate system, we choose matrices that preserve these invariance properties. Define the matrix A_1 that has the following block structure

$$(2\mathbb{I}_p, -\mathbb{I}_p, 0_p, -\mathbb{I}_p; -\mathbb{I}_p, 2\mathbb{I}_p, -\mathbb{I}_p, 0_p; 0_p, -\mathbb{I}_p, 2\mathbb{I}_p, -\mathbb{I}_p; -\mathbb{I}_p, 0_p, -\mathbb{I}_p, 2\mathbb{I}_p),$$

where $\mathbb{I}_p$ stands for the identity $p \times p$ matrix. The first test statistic $T_n^{(1)}$ is based on $\{A_1\}$. Note that $L = 1, s = 4$ for it.

Next, we consider 5-diagonal matrix A_2 with the following block structure

$$(2\mathbb{D}_p, -\mathbb{D}, 0_p, -\mathbb{D}_p; -\mathbb{D}_p, 2\mathbb{D}_p, -\mathbb{D}_p, 0_p\ ; 0_p, -\mathbb{D}_p, 2\mathbb{D}_p, -\mathbb{D}_p; \\ -\mathbb{D}_p, 0_p, -\mathbb{D}_p, 2\mathbb{D}_p),$$

where $\mathbb{D}_p$ is a diagonal matrix with diagonal entries $d_k = \log(2k), k = 1, \dots, p$. The second test statistic $T_n^{(2)}$ is based on $\{A_2\}$. Note that $L = 2, s = 2\log(2p)$ for it.

The level and power results are summarized in Table 1. The numbers for $T_n^{(2)}$ are in brackets. Across the table $T_n^{(2)}$ performs better than $T_n^{(1)}$. The powers increase when sample sizes increase. The power decreases with the increase in dimension. The power decreases as the covariance structure changes from nearly diagonal ($\rho = 0.1$) to nearly singular $\rho = 0.9$. Both tests have mediocre performance for t_4 distribution with covariance structures $\rho = 0.5, NA$. The performance is excellent for normal and χ^2_1 distributions.

For the second simulation, we consider 15-diagonal matrices. Define matrix A_3 with the following block structure

$$(3\mathbb{B}_p, -\mathbb{B}_p, -\mathbb{B}_p, -\mathbb{B}_p;\ -\mathbb{B}_p, 3\mathbb{B}_p, -\mathbb{B}_p, -\mathbb{B}_p\ ; -\mathbb{B}_p, -\mathbb{B}_p, 3\mathbb{B}_p, -\mathbb{B}_p; \\ -\mathbb{B}_p, -\mathbb{B}_p, -\mathbb{B}_p, 3\mathbb{B}_p),$$

where $\mathbb{B}_p$ has ones on the main and above-main diagonals, it has zeros everywhere else. The third test statistic $T_n^{(3)}$ is based on $\{A_3\}$. Note that $L = 1, s = 6$ for it.

Next, consider A_4 with the block structure

$$(9\mathbb{D}_p, -\mathbb{D}_p, -2\mathbb{D}_p, -6\mathbb{D}_p;\ -6\mathbb{D}_p, 9\mathbb{D}_p, -\mathbb{D}_p, -2\mathbb{D}_p\ ; -2\mathbb{D}_p, -6\mathbb{D}_p, 9\mathbb{D}_p, -\mathbb{D}_p; \\ -\mathbb{D}_p, -2\mathbb{D}_p, -6\mathbb{D}_p, 9\mathbb{D}_p),$$

where $\mathbb{D}_p$ is defined in $T_n^{(2)}$. The fourth test statistic $T_n^{(4)}$ is based on matrices obtained from A_4 by permuting the blocks in a circular fashion among indices (1234), (2341), (3412), (4123). Note that $L = 4, s = 18\log(2p)$ for it.

The level and power results are summarized in Table 2. The numbers for $T_n^{(4)}$ are in brackets. The two tests have comparable level performance. As expected, the powers increase when sample sizes increase. The power decreases with the dimension increase. The power decreases as the covariance structure changes from nearly diagonal ($\rho = 0.1$) to nearly singular $\rho = 0.9$. The performance is excellent for normal and χ^2_1 distributions. As before t_4 distribution presents the most difficult challenge for the tests. The level is poor for $p = 1000, \mathbf{n_1}, \rho = NA$, but it dramatically improves with sample size increase. For other cases under t_4 model, the tests perform good.

Upon examination of both Tables 1 and 2, across all scenarios $T_n^{(4)}$ has the best power performance than other 3 tests $T_n^{(m)}, m = 1, 2, 3$. This can be explained through the structure of the underlying class of convex sets. This class is richer for

Table 1
Level and adjusted power performance of $T_n^{(1)}(T_n^{(2)})$. We used 5-diagonal matrices, 10K bootstrap repetitions, and 1K empirical repetitions for each entry.

ρ	p	**n**	Size N	Size t_4	Size χ_1^2	Power N	Power t_4	Power χ_1^2
0.1	50	$\mathbf{n_1}$	3.9(4.1)	2.2(2.1)	3.7(3.6)	29.7(48.9)	24.0(41.8)	19.9(34.8)
		$\mathbf{n_2}$	5.2(3.5)	2.6(2.2)	4.9(3.7)	30.4(46.7)	23.5(39.1)	19.4(36.4)
		$\mathbf{n_3}$	4.4(5.4)	3.4(3.7)	5.0(5.1)	39.4(48.6)	35.9(44.8)	30.0(43.5)
0.1	500	$\mathbf{n_1}$	2.5(2.5)	0.2(0.1)	1.4(1.8)	25.9(41.0)	43.2(27.2)	13.7(19.9)
		$\mathbf{n_2}$	3.3(3.3)	0.8(0.4)	3.0(3.3)	31.0(45.1)	11.3(34.5)	17.7(29.5)
		$\mathbf{n_3}$	4.0(3.7)	0.8(1.0)	5.5(6.1)	24.0(48.8)	21.1(34.0)	19.7(25.3)
0.1	1000	$\mathbf{n_1}$	2.3(2.0)	0.3(0.1)	1.0(0.9)	23.7(40.3)	7.8(14.3)	9.8(21.2)
		$\mathbf{n_2}$	3.1(2.7)	0.3(0.1)	2.7(2.9)	27.3(42.0)	11.1(30.9)	14.3(23.4)
		$\mathbf{n_3}$	4.3(4.5)	0.3(0.6)	5.4(3.4)	26.6(38.2)	21.0(30.5)	16.5(35.6)
0.5	50	$\mathbf{n_1}$	4.2(5.3)	2.4(2.7)	4.6(4.8)	25.8(36.0)	24.7(36.2)	23.0(30.9)
		$\mathbf{n_2}$	3.6(4.1)	3.7(3.2)	5.0(4.3)	29.3(33.3)	20.0(31.0)	22.0(26.9)
		$\mathbf{n_3}$	5.8(5.5)	5.0(3.6)	5.4(4.9)	24.3(35.4)	25.3(36.0)	22.9(38.6)
0.5	500	$\mathbf{n_1}$	3.2(3.0)	1.8(1.3)	1.7(2.9)	22.5(28.3)	9.3(16.3)	20.5(19.6)
		$\mathbf{n_2}$	5.1(4.6)	3.1(1.7)	3.9(5.4)	19.2(23.6)	11.1(26.8)	15.7(19.6)
		$\mathbf{n_3}$	4.6(4.4)	3.7(3.9)	5.4(5.1)	26.2(24.8)	12.4(18.5)	13.8(21.4)
0.5	1000	$\mathbf{n_1}$	3.5(4.2)	0.6(0.2)	2.2(2.7)	17.1(24.1)	12.2(26.8)	11.8(19.3)
		$\mathbf{n_2}$	5.3(4.8)	0.7(1.2)	4.5(3.8)	13.8(20.9)	17.1(18.8)	12.7(21.5)
		$\mathbf{n_3}$	4.7(4.6)	1.7(2.7)	4.8(5.3)	18.6(25.2)	14.9(18.3)	15.5(21.1)
0.9	50	$\mathbf{n_1}$	5.5(6.3)	5.0(3.6)	5.1(4.8)	20.1(27.1)	23.0(35.2)	18.4(31.0)
		$\mathbf{n_2}$	5.0(5.7)	4.6(4.5)	4.5(5.1)	22.3(23.2)	24.4(26.0)	21.4(25.3)
		$\mathbf{n_3}$	5.3(5.2)	5.1(4.8)	5.3(4.7)	28.5(32.3)	24.3(29.3)	23.9(32.4)
0.9	500	$\mathbf{n_1}$	5.1(3.5)	3.4(5.1)	3.6(4.6)	15.0(23.9)	15.6(15.3)	19.4(19.0)
		$\mathbf{n_2}$	5.2(5.0)	4.2(4.3)	5.0(5.7)	19.8(22.1)	17.4(18.4)	16.3(18.1)
		$\mathbf{n_3}$	5.1(4.6)	4.7(4.8)	5.2(5.4)	21.1(23.7)	17.0(17.4)	16.7(19.3)
0.9	1000	$\mathbf{n_1}$	5.7(4.8)	2.7(3.4)	4.7(4.3)	13.6(18.9)	13.1(18.8)	13.4(16.3)
		$\mathbf{n_2}$	5.3(4.8)	3.9(3.7)	3.9(4.4)	15.7(19.1)	14.2(15.8)	15.9(20.7)
		$\mathbf{n_3}$	5.5(4.5)	4.0(4.4)	4.7(5.3)	15.9(21.3)	15.7(17.5)	15.0(18.7)
NA	50	$\mathbf{n_1}$	2.4(4.4)	2.9(2.3)	5.5(4.7)	36.4(45.3)	26.5(38.6)	20.5(33.3)
		$\mathbf{n_2}$	4.0(4.5)	3.4(3.1)	5.3(4.9)	31.6(46.1)	26.4(44.7)	23.0(40.2)
		$\mathbf{n_3}$	5.0(3.5)	4.0(4.5)	5.5(6.3)	34.5(53.0)	30.5(40.0)	33.1(42.4)
NA	500	$\mathbf{n_1}$	2.5(2.4)	0.4(0.2)	1.5(2.0)	24.7(40.9)	9.4(33.3)	15.8(23.2)
		$\mathbf{n_2}$	3.2(3.8)	0.4(0.7)	2.9(2.9)	32.8(44.2)	27.4(33.6)	20.9(32.6)
		$\mathbf{n_3}$	4.3(4.9)	2.8(1.7)	5.9(4.4)	31.3(36.8)	16.8(29.5)	20.8(35.6)
NA	1000	$\mathbf{n_1}$	2.3(2.1)	0.1(0.2)	1.0(1.2)	29.1(41.7)	7.6(12.9)	15.3(22.9)
		$\mathbf{n_2}$	2.3(2.9)	0.1(0.4)	2.6(2.8)	32.3(47.3)	18.9(21.8)	17.1(28.7)
		$\mathbf{n_3}$	2.7(4.2)	0.7(1.1)	5.5(4.1)	37.6(42.5)	19.7(28.4)	19.8(36.0)

$T_n^{(4)}$ than the rest of the tests. On a rare occasion, test $T_n^{(2)}$ has comparable or slightly better power than $T_n^{(4)}$, for instance when $p = 50, \mathbf{n_3}, \rho = 0.5, NA$ for χ_1^2 model. In general, one needs to select a finite number L of sparse matrices $A^{(l)}$ such that $A^{(l)}\mu$ is maximized in some sense under the fixed alternative.

Finally, we remark that the case of t distribution with 4 degrees of freedom seems to be the most challenging distribution for our tests. The reason is that we perform a relatively small sample size comparison under i.i.d. setup, when B_n is a constant. However, this constant is 4.312 times higher for t with 4 degrees of freedom than that constant for the standard normal distribution. This impacts the performance when n is relatively small compared to p. The power picks up significantly if we increase the minimum sample size from 100 in $\mathbf{n}_3$ to 200.

In comparison to the various tests in Zhang et al. (2017) our tests have good level performance but tend to be a bit conservative, while Zhang et al. (2017) tests tend to overshoot the nominal level, see their Table 1. Looking at power performance in their Table 2, their tests do good for normal distribution with $\rho = 0.1$ but our tests do better for all other cases. We also remark that case $\rho = NA$ was not studied in Zhang et al. (2017). In particular, we refer to Table 3 that contains the level and power information for T_{L2D} test in Zhang et al. (2017) in comparison to our $T_n^{(4)}$.

4.2. Comparison with Cai and Xia (2014) approach

In the third simulation we compare our approach with that of Cai and Xia (2014). We consider 5-diagonal matrix A_5 with the following block structure

Table 2
Level and adjusted power performance of $T_n^{(3)}$ ($T_n^{(4)}$). We used 15-diagonal matrices, 10K bootstrap repetitions, and 1K empirical repetitions for each entry.

ρ	p	**n**	Size N	Size t_4	Size χ^2_1	Power N	Power t_4	Power χ^2_1
0.1	50	$\mathbf{n_1}$	3.1(2.7)	3.2(2.3)	4.5(3.9)	40.7(71.7)	33.1(60.9)	29.9(54.9)
		$\mathbf{n_2}$	4.5(4.2)	3.5(3.5)	4.5(5.8)	42.9(67.6)	36.2(57.4)	34.4(49.8)
		$\mathbf{n_3}$	5.1(4.4)	4.4(4.5)	5.4(5.2)	46.0(68.3)	39.5(62.4)	40.6(59.2)
0.1	500	$\mathbf{n_1}$	3.0(2.3)	0.6(0.7)	3.1(2.8)	31.9(53.8)	24.1(36.2)	21.9(33.0)
		$\mathbf{n_2}$	2.7(3.1)	1.1(1.6)	4.8(3.9)	44.1(55.5)	31.0(45.6)	27.8(43.7)
		$\mathbf{n_3}$	5.2(4.9)	1.4(2.5)	6.4(4.9)	36.6(43.4)	34.7(36.4)	24.6(37.4)
0.1	1000	$\mathbf{n_1}$	1.5(2.1)	0.1(0.4)	3.1(1.5)	39.2(56.9)	31.5(35.0)	17.9(39.6)
		$\mathbf{n_2}$	3.3(2.9)	0.6(0.6)	4.9(4.0)	32.5(49.6)	27.6(45.8)	20.9(36.4)
		$\mathbf{n_3}$	4.1(4.3)	1.4(2.5)	5.4(4.2)	33.8(50.3)	32.8(40.6)	26.5(47.7)
0.5	50	$\mathbf{n_1}$	5.1(4.0)	3.2(3.4)	5.3(5.3)	18.3(38.6)	20.7(35.2)	18.9(29.1)
		$\mathbf{n_2}$	3.9(4.0)	5.5(4.0)	6.9(5.3)	25.3(39.0)	20.0(33.2)	16.7(31.2)
		$\mathbf{n_3}$	5.5(4.2)	4.7(4.3)	5.4(5.1)	22.9(38.7)	24.7(32.5)	22.4(32.8)
0.5	500	$\mathbf{n_1}$	5.4(3.9)	2.3(2.7)	3.6(4.5)	13.9(22.1)	13.0(18.9)	15.0(19.8)
		$\mathbf{n_2}$	4.7(4.7)	3.3(2.9)	4.7(6.8)	18.3(26.5)	16.3(26.7)	18.4(17.3)
		$\mathbf{n_3}$	4.0(5.2)	3.7(2.9)	5.7(5.5)	21.0(19.9)	16.1(23.4)	16.4(21.5)
0.5	1000	$\mathbf{n_1}$	4.0(3.5)	1.5(2.0)	4.3(3.6)	14.8(24.1)	12.9(18.7)	11.9(18.5)
		$\mathbf{n_2}$	5.0(4.4)	2.8(2.4)	4.4(5.3)	15.4(23.3)	15.2(19.0)	18.9(17.2)
		$\mathbf{n_3}$	3.8(5.1)	3.2(3.4)	3.3(4.2)	20.9(20.4)	15.0(20.2)	23.6(22.9)
0.9	50	$\mathbf{n_1}$	6.0(5.1)	4.8(6.2)	5.9(5.6)	18.4(28.7)	21.2(26.5)	18.7(27.9)
		$\mathbf{n_2}$	4.8(4.2)	4.9(4.4)	5.2(4.8)	31.5(32.6)	27.1(33.5)	37.7(30.9)
		$\mathbf{n_3}$	3.8(5.5)	3.3(5.1)	5.1(6.3)	49.3(30.9)	46.4(33.6)	43.1(27.8)
0.9	500	$\mathbf{n_1}$	5.3(5.1)	4.8(3.7)	3.6(6.1)	13.4(18.0)	10.9(20.0)	19.1(14.4)
		$\mathbf{n_2}$	5.2(3.9)	4.6(5.4)	5.2(6.9)	15.3(25.4)	13.7(18.5)	15.5(15.7)
		$\mathbf{n_3}$	4.8(5.0)	4.1(4.3)	7.1(5.2)	16.5(22.0)	15.8(22.0)	12.7(20.4)
0.9	1000	$\mathbf{n_1}$	5.4(6.3)	5.7(4.1)	5.8(4.5)	14.4(14.9)	8.9(17.9)	11.3(18.3)
		$\mathbf{n_2}$	4.1(4.8)	6.2(4.2)	6.1(5.5)	17.4(17.7)	10.5(16.6)	11.2(15.0)
		$\mathbf{n_3}$	4.0(4.6)	4.2(4.2)	5.4(6.0)	17.9(20.9)	14.7(17.5)	13.4(15.3)
NA	50	$\mathbf{n_1}$	4.0(4.0)	2.0(2.3)	4.2(4.2)	31.0(42.9)	30.0(42.5)	24.4(32.0)
		$\mathbf{n_2}$	2.8(3.0)	3.6(3.0)	6.9(5.9)	38.6(48.3)	23.1(43.2)	18.3(33.0)
		$\mathbf{n_3}$	5.3(4.7)	3.7(4.2)	5.0(5.6)	30.2(45.0)	30.6(40.4)	26.9(40.6)
NA	500	$\mathbf{n_1}$	1.9(1.7)	0.7(0.6)	3.0(3.2)	28.5(45.4)	23.9(29.3)	15.1(18.2)
		$\mathbf{n_2}$	2.9(3.9)	1.0(1.2)	3.9(4.1)	34.4(40.7)	22.8(38.2)	22.1(32.0)
		$\mathbf{n_3}$	3.1(3.2)	1.2(2.4)	4.9(5.2)	33.9(45.7)	21.8(29.3)	20.8(28.8)
NA	1000	$\mathbf{n_1}$	1.3(1.5)	0.3(0.4)	3.2(1.7)	29.5(44.7)	14.3(19.5)	12.5(22.3)
		$\mathbf{n_2}$	3.2(3.8)	0.9(0.8)	2.9(4.6)	43.2(59.4)	17.5(27.8)	22.4(25.9)
		$\mathbf{n_3}$	3.4(4.1)	1.7(1.1)	5.7(4.9)	33.3(43.1)	21.2(37.1)	19.9(33.7)

$$(\mathbb{D}_p, -\mathbb{D}_p/3, 0_p, -2\mathbb{D}_p/3; -2\mathbb{D}_p/3, 2\mathbb{D}_p, -\mathbb{D}_p/3, 0_p\ ; 0_p, -2\mathbb{D}_p/3, 2\mathbb{D}_p, -\mathbb{D}_p/3; \\ -\mathbb{D}_p/3, 0_p, -2\mathbb{D}_p/3, 2\mathbb{D}_p),$$

where $\mathbb{D}_p$ is a diagonal matrix with diagonal entries $d_k = \log(2k), k = 1, \ldots, p$. The fifth test statistic $T_n^{(5)}$ is based on $\{A_5, A_5'\}$, where A_5' is similar to A_5 with weights 1/3 and 2/3 switched. Note that $L = 2, s = 2\log(2p)$ for it.

For Cai and Xia (2014) approach we use their tests $\Phi_\alpha(\hat{\Omega}), \Psi_\alpha(\hat{\Omega})$ with the CLIME estimator $\hat{\Omega}$ of Cai et al. (2011) for the unknown precision matrix $\Omega = \Sigma^{-1}$. As recommended in Cai and Xia (2014) we used intermediate correction $\Psi_\alpha(\hat{\Omega})$, which helps with level performance, but it lowers power. We notice that when $\rho = 0.5$ or NA the correction does not change the test $\Phi_\alpha(\hat{\Omega})$ into $\Psi_\alpha(\hat{\Omega})$. Also for these two cases the power gets worse with changes in sample size, the power performance of $\Psi_\alpha(\hat{\Omega})$ is unstable and poor. This can be explained by the accumulated errors from CLIME estimation of the precision matrix.

The results are summarized in Table 4. Our test $T_n^{(5)}$ outperforms $\Psi_\alpha(\hat{\Omega})$ when $\rho = 0.1$ (almost a diagonal covariance matrix) for all dimensions and all distributions except t_4, for which the results are mixed. It outperforms $\Psi_\alpha(\hat{\Omega})$ when $\rho = 0.5$ everywhere except for t_4 and χ^2_1 distributions with $p = 500, \mathbf{n_1}$ and $p = 1000, \mathbf{n_2}$.

However, for $\rho = 0.9$ (almost singular covariance matrix) $\Psi_\alpha(\hat{\Omega})$ has exceptionally high power but poor level while our test has good level and solid power performance. After singular value decomposition such covariance is quite sparse, and that is when test $\Psi_\alpha(\hat{\Omega})$ really shines.

For the case of $\rho =$ NA, our test outperforms $\Psi_\alpha(\hat{\Omega})$ for all dimensions and all distributions. We conjecture that the power loss happens due to numerical errors accumulated in the precision matrix estimation process where we used CLIME.

Table 3
Level and adjusted power performance of $T_n^{(4)}$ and T_{L2D} in Zhang et al. (2017) in brackets.

ρ	p	**n**	Size N	Size t_4	Size χ_1^2	Power N	Power t_4	Power χ_1^2
0.1	50	$\mathbf{n_1}$	2.7(6.1)	2.3(5.1)	3.9(4.9)	71.7(52.0)	60.9(54.0)	54.9(54.8)
		$\mathbf{n_2}$	4.2(5.5)	3.5(5.1)	5.8(5.4)	67.6(42.5)	57.4(42.3)	49.8(42.3)
		$\mathbf{n_3}$	4.4(6.0)	4.5(5.4)	5.2(4.9)	68.3(52.8)	62.4(54.7)	59.2(55.8)
0.1	500	$\mathbf{n_1}$	2.3(6.8)	0.7(6.1)	2.8(6.6)	53.8(42.5)	36.2(43.4)	33.0(41.7)
		$\mathbf{n_2}$	3.1(6.4)	1.6(6.4)	3.9(6.2)	55.5(46.2)	45.6(46.0)	43.7(47.7)
		$\mathbf{n_3}$	4.9(6.3)	2.5(6.0)	4.9(6.3)	43.4(38.8)	36.4(39.2)	37.4(38.6)
0.1	1000	$\mathbf{n_1}$	2.1(6.5)	0.4(6.9)	1.5(6.8)	56.9(42.2)	35.0(40.0)	39.6(40.8)
		$\mathbf{n_2}$	2.9(6.3)	0.6(6.6)	4.0(6.9)	49.6(40.4)	45.8(38.9)	36.4(38.4)
		$\mathbf{n_3}$	4.3(6.5)	2.5(6.8)	4.2(6.2)	50.3(41.5)	40.6(40.6)	47.7(43.1)
0.5	50	$\mathbf{n_1}$	4.0(6.3)	3.4(6.1)	5.3(5.8)	38.6(15.3)	35.2(16.6)	29.1(16.4)
		$\mathbf{n_2}$	4.0(5.8)	4.0(5.7)	5.3(5.6)	39.0(13.8)	33.2(13.4)	31.2(14.3)
		$\mathbf{n_3}$	4.2(5.4)	4.3(5.3)	5.1(6.0)	38.7(17.5)	32.5(17.2)	32.8(16.3)
0.5	500	$\mathbf{n_1}$	3.9(6.2)	2.7(6.3)	4.5(6.0)	22.1(10.6)	18.9(10.6)	19.8(10.9)
		$\mathbf{n_2}$	4.7(5.6)	2.9(6.0)	6.8(5.9)	26.5(11.9)	26.7(11.9)	17.3(11.5)
		$\mathbf{n_3}$	5.2(5.7)	2.9(5.8)	5.5(5.7)	19.9(10.4)	23.4(10.1)	21.5(10.4)
0.5	1000	$\mathbf{n_1}$	3.5(6.3)	2.0(6.2)	3.6(6.1)	24.1(10.0)	18.7(10.7)	18.5(10.9)
		$\mathbf{n_2}$	4.4(5.9)	2.4(5.8)	5.3(6.0)	23.3(10.9)	19.0(10.6)	17.2(10.7)
		$\mathbf{n_3}$	5.1(5.5)	3.4(6.0)	4.2(6.2)	20.4(10.8)	20.2(9.8)	22.9(10.1)
0.9	50	$\mathbf{n_1}$	5.1(6.1)	6.2(5.9)	5.6(6.0)	28.7(10.3)	26.5(10.3)	27.9(10.3)
		$\mathbf{n_2}$	4.2(5.5)	4.4(5.3)	4.8(5.0)	32.6(9.6)	33.5(9.2)	30.9(9.7)
		$\mathbf{n_3}$	5.5(5.6)	5.1(5.2)	6.3(5.5)	30.9(10.8)	33.6(10.9)	27.8(10.4)
0.9	500	$\mathbf{n_1}$	5.1(5.5)	3.7(5.5)	6.1(5.7)	18.0(8.5)	20.0(8.0)	14.4(8.4)
		$\mathbf{n_2}$	3.9(5.7)	5.4(5.5)	6.9(5.4)	25.4(7.9)	18.5(8.6)	15.7(8.5)
		$\mathbf{n_3}$	5.0(5.3)	4.3(5.0)	5.2(5.2)	22.0(7.7)	22.0(8.3)	20.4(7.6)
0.9	1000	$\mathbf{n_1}$	6.3(6.1)	4.1(5.5)	4.5(5.7)	14.9(7.3)	17.9(8.1)	18.3(7.8)
		$\mathbf{n_2}$	4.8(5.8)	4.2(5.4)	5.5(5.0)	17.7(7.1)	16.6(7.7)	15.0(8.2)
		$\mathbf{n_3}$	4.6(5.3)	4.2(5.7)	6.0(5.7)	20.9(7.6)	17.5(7.6)	15.3(7.3)

It is possible that with better and faster precision matrix estimators $\Psi_\alpha(\hat{\Omega})$ could have better performance. We also remark that for several scenarios test $\Psi_\alpha(\hat{\Omega})$ struggles to differentiate between null and alternative hypotheses, since the level and power values are close.

We performed all the simulations in Matlab. This computational work was partially supported by Michigan State University High Performance Computing Center through computational resources provided by the Institute for Cyber–Enabled Research. We would like to mention the computational cost in this simulation study. The need for precision matrix estimation and lack of sparsity in means and covariances lead to severe increase in computational time for the test from Cai and Xia (2014) in comparison to our test. For example, for $p = 500, K = 4, \rho = NA$ our test did 1000 empirical iterations in 45 minutes while 1 iteration of Cai and Xia's test took 72 minutes. For $p = 1000, K = 4, \rho = NA$ our test did 1000 empirical iterations in 6 hours while 1 iteration of test from Cai and Xia (2014) took 15 hours. This comparison is based on running both codes on 1 node Intel(R) Xeon(R) CPU E5-2680 v4 2.40 GHz with 492 GB memory and 190 GB disk size. Usage of the fastclime approach from Pang et al. (2014) had the modest effect for the datasets that had non-sparse precision matrices.

4.3. Comparison with Chen et al. (2019) approach

For the fourth simulation study we consider the setup in Chen et al. (2019). Under normal distribution model they consider $K = 3$ groups with different covariance matrices of the form $\lambda^{|i-j|}, i, j, = 1, \ldots, p$, with $\lambda = 0.4, 0.5, 0.6$ for group 1, 2, 3, respectively. Under the null hypothesis H_0 all means μ_1, μ_2, μ_3 are zero. Under their sparse alternative H_A the first group has $\mu_1 = 0$, the other two means combined form a sparse vector of length $2p$ with only $[(2p)^{0.4}]$ non-zero entries of magnitude $(2r\log(2p)/n)^{1/2}$ uniformly distributed among $2p$ components. Parameter r controls the strength of the signal and varies between 0.1, 0.2, and 0.4. Chen et al. (2019) proposed multi-thresholding method for MANOVA problem. Mult-A1 and Mult-A2 correspond to their tests without and with data transformation, the latter requires precision matrix estimation for unknown $\Omega_i, i = 1, 2, 3$.

We consider 5-diagonal matrix A_6 with the block structure

$$(\mathbb{D}_p, -\mathbb{D}_p/3, -2\mathbb{D}_p/3; -\mathbb{D}_p/3, \mathbb{D}_p, -2\mathbb{D}_p/3; -2\mathbb{D}_p/3, \mathbb{D}_p, -\mathbb{D}_p/3),$$

Table 4
Comparison of our method vs Cai and Xia (2014) method whose numbers are given in brackets. Powers are adjusted for sizes.

ρ	p	$\mathbf{n}$	Size N	Size t_4	Size χ^2_1	Power N	Power t_4	Power χ^2_1
0.1	50	$\mathbf{n_1}$	5.7(13.1)	2.4(14.8)	4.3(14.8)	55.6(15.1)	50.7(15.6)	38.3(12.2)
		$\mathbf{n_2}$	6.1(9.0)	3.5(8.7)	5.5(9.3)	55.5(10.0)	49.9(11.6)	37.5(10.9)
		$\mathbf{n_3}$	7.0(5.9)	4.9(5.9)	8.7(6.9)	55.0(13.9)	50.4(13.6)	44.2(10.3)
0.1	500	$\mathbf{n_1}$	3.9(9.0)	0.4(3.8)	1.8(9.1)	44.7(11.0)	22.9(27.6)	26.6(10.2)
		$\mathbf{n_2}$	5.4(7.6)	1.6(3.1)	4.5(7.2)	46.0(9.4)	23.5(26.6)	29.3(8.9)
		$\mathbf{n_3}$	7.9(6.5)	1.6(3.9)	6.3(3.9)	32.5(9.1)	30.7(18.8)	33.5(12.3)
0.1	1000	$\mathbf{n_1}$	3.3(15.9)	0.1(22.1)	1.1(32.4)	44.3(24.2)	15.6(19.8)	24.4(7.6)
		$\mathbf{n_2}$	4.5(15.8)	0.3(18.9)	4.0(12.6)	42.0(9.7)	23.1(9.0)	22.9(9.8)
		$\mathbf{n_3}$	6.5(7.1)	1.0(9.0)	6.0(12.1)	45.9(14.3)	32.0(12.6)	31.2(9.0)
0.5	50	$\mathbf{n_1}$	5.3(7.5)	3.2(8.5)	5.9(6.5)	42.7(16.8)	44.1(17.3)	30.5(20.0)
		$\mathbf{n_2}$	6.7(6.6)	4.2(6.5)	6.0(5.1)	39.1(15.4)	41.0(13.0)	31.0(22.0)
		$\mathbf{n_3}$	8.9(5.1)	4.5(5.9)	8.3(4.7)	33.7(17.6)	43.1(13.5)	31.5(21.9)
0.5	500	$\mathbf{n_1}$	5.0(9.4)	1.8(3.9)	4.1(7.8)	27.5(16.5)	20.5(31.7)	18.6(20.1)
		$\mathbf{n_2}$	8.0(8.8)	2.2(4.9)	7.0(7.6)	23.3(9.0)	28.3(13.9)	19.6(10.8)
		$\mathbf{n_3}$	7.1(5.6)	3.4(6.1)	7.1(7.0)	24.3(9.7)	23.7(8.6)	20.3(9.6)
0.5	1000	$\mathbf{n_1}$	6.0(11.8)	1.0(12.9)	3.3(9.6)	20.5(10.9)	17.5(15.2)	22.1(14.4)
		$\mathbf{n_2}$	6.7(2.8)	2.0(1.9)	6.4(7.1)	21.5(21.8)	19.6(33.5)	18.3(10.5)
		$\mathbf{n_3}$	6.5(2.7)	3.6(2.9)	6.3(6.2)	25.5(17.8)	20.5(7.1)	24.0(11.2)
0.9	50	$\mathbf{n_1}$	6.9(5.6)	8.2(4.3)	7.9(5.0)	29.1(42.3)	29.5(50.2)	26.2(49.4)
		$\mathbf{n_2}$	6.7(5.1)	6.2(5.5)	7.0(4.7)	30.5(48.1)	35.8(47.2)	31.1(59.2)
		$\mathbf{n_3}$	6.2(5.0)	5.3(4.9)	6.8(5.4)	34.1(59.6)	39.2(59.7)	33.6(60.6)
0.9	500	$\mathbf{n_1}$	7.6(7.0)	4.7(15.6)	7.2(11.3)	20.7(61.6)	19.4(46.8)	16.4(51.8)
		$\mathbf{n_2}$	8.8(7.3)	5.9(14.7)	6.7(14.6)	20.9(63.9)	20.3(56.0)	21.8(51.5)
		$\mathbf{n_3}$	7.8(7.4)	6.8(13.6)	6.7(15.7)	24.7(73.1)	19.9(60.3)	22.7(64.9)
0.9	1000	$\mathbf{n_1}$	6.9(10.1)	3.8(12.9)	6.0(14.9)	18.0(44.8)	18.7(42.6)	17.5(35.4)
		$\mathbf{n_2}$	7.6(5.0)	5.0(9.6)	6.2(5.9)	18.1(60.1)	18.1(56.1)	18.9(57.6)
		$\mathbf{n_3}$	8.4 (6.8)	5.5(3.9)	7.8(4.1)	17.8(58.0)	17.6(75.4)	16.5(69.4)
NA	50	$\mathbf{n_1}$	5.5(12.0)	4.3(12.3)	5.9(11.3)	51.8(6.3)	41.4(5.6)	35.6(5.5)
		$\mathbf{n_2}$	6.0(7.1)	5.5(7.6)	6.5(6.0)	50.9(8.3)	40.6(5.1)	41.1(6.5)
		$\mathbf{n_3}$	7.3(5.8)	6.4(6.0)	6.8(5.6)	48.3(3.9)	43.4(6.2)	49.5(5.1)
NA	500	$\mathbf{n_1}$	4.5(18.1)	0.4(17.1)	2.2(21.4)	43.6(10.4)	30.3(13.0)	27.9(13.8)
		$\mathbf{n_2}$	5.5(17.2)	1.6(13.7)	4.4(19.2)	46.0(5.5)	31.3(12.7)	36.0(7.3)
		$\mathbf{n_3}$	5.8(7.5)	2.4(7.1)	6.7(13.1)	54.3(8.6)	34.3(9.5)	35.6(6.6)
NA	1000	$\mathbf{n_1}$	2.6(15.8)	0.3(20.9)	2.0(24.1)	52.7(21.1)	21.0(15.0)	22.6(9.4)
		$\mathbf{n_2}$	5.2(14.5)	0.8(18.0)	4.8(17.1)	49.2(9.2)	25.2(7.8)	30.6(9.1)
		$\mathbf{n_3}$	7.9(12.1)	1.8(8.8)	6.6(16.4)	46.8(7.6)	32.3(11.5)	39.0(5.6)

where $\mathbb{D}_p$ is a diagonal matrix with diagonal entries $d_k = \log(2k), k = 1, \ldots, p$, as in A_5. The sixth test statistic $T_n^{(6)}$ is based on $\{A_6, A_6'\}$, where A_6' is similar to A_6 with weights 1/3 and 2/3 switched. Note that $L = 2, s = 2\log(2p)$ for it. We use 10K bootstrap samples to obtain Q_α with $\alpha = 0.05$.

The comparison results are summarized in Table 5. The level performance of our test is better than both tests from Chen et al. (2019) especially for higher dimensional case $p = 400$. With respect to power, our test is comparable to Mult-A1, which is applied to the untransformed data. Of course Mult-A2 is better than our test because it is adjusted to the covariance structure but it requires precision matrix estimation, which makes it rather slow. Both Mult-A1 and Mult-A2 are designed for sparse means (controlled by r in simulation) and sub-Gaussian distributions. They are not applicable under scenarios considered in subsections 4.1 and 4.2. Overall, the applicability of both Mult-A1 and Mult-A2 tests is more narrow than the applicability of our test.

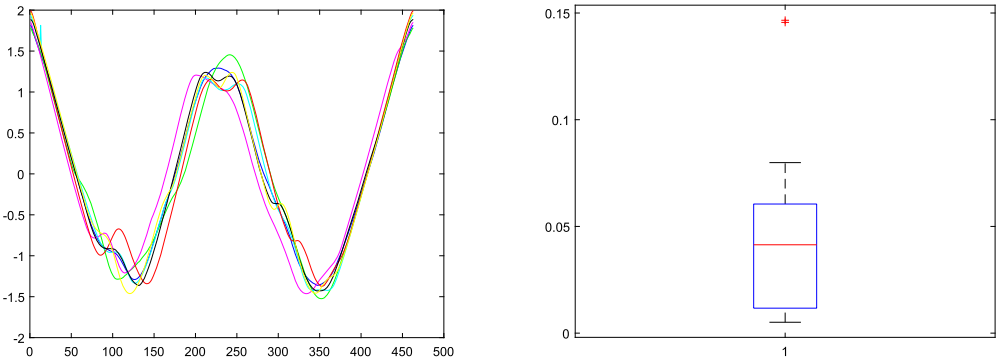
5. Real data example

We consider the second example in Zhang and Sakhanenko (2019). It is based on the dataset from Lee et al. (2008) work on fish species' recognition and migration monitoring. For each fish, the shape pattern as a vector of dimension 463 is obtained. They consider 7 fish species that have similar shape characteristics. For each species 50 fish are sampled. The dataset can be obtained from the UCR Time Series Classification and Clustering archive

`http://www.cs.ucr.edu/~eamonn/time_series_data`

Table 5
Comparison of our method $T_n^{(6)}$ with Chen et al. (2019) multi-thresholding methods for sparse mean alternative. Powers are adjusted for sizes, r controls the strength of the signal in the alternatives.

(p, n_1, n_2, n_3)	Method	Size	$r = 0.1$	$r = 0.2$	$r = 0.4$
(100, 40, 40, 40)	$T_n^{(6)}$	4.7	8.8	12.8	26.2
	Mult-A1	2.5	9.8	16.6	33.7
	Mult-A2	4.3	16.8	30.3	89.6
(100, 80, 80, 80)	$T_n^{(6)}$	4.2	14.7	16.9	30.0
	Mult-A1	3.4	8.3	16.2	39.4
	Mult-A2	4.0	26.6	59.3	98.4
(100, 100, 100, 100)	$T_n^{(6)}$	4.8	11.5	18.4	29.0
	Mult-A1	5.0	8.3	12.4	35.7
	Mult-A2	4.9	20.1	48.7	99.1
(200, 40, 40, 40)	$T_n^{(6)}$	5.8	8.2	10.7	23.3
	Mult-A1	2.2	8.7	14.7	37.9
	Mult-A2	2.0	24.1	69.9	98.7
(200, 80, 80, 80)	$T_n^{(6)}$	6.0	10.3	13.3	27.5
	Mult-A1	4.1	6.9	14.6	44.6
	Mult-A2	5.0	27.8	66.9	99.0
(200, 100, 100, 100)	$T_n^{(6)}$	5.8	12.3	14.8	26.8
	Mult-A1	3.7	9.1	17.1	48.5
	Mult-A2	5.0	21.2	75.2	99.5
(400, 40, 40, 40)	$T_n^{(6)}$	6.2	8.3	11.1	23.0
	Mult-A1	2.0	9.4	12.2	39.7
	Mult-A2	2.7	19.5	51.4	97.9
(400, 80, 80, 80)	$T_n^{(6)}$	6.0	9.2	13.4	29.7
	Mult-A1	4.7	6.3	13.0	46.3
	Mult-A2	4.1	28.6	73.2	99.4
(400, 100, 100, 100)	$T_n^{(6)}$	5.3	14.4	15.8	33.8
	Mult-A1	3.3	10.8	17.9	78.7
	Mult-A2	4.1	30.9	52.9	100



(a) group means (b) group differences

Fig. 1. Shape characteristics for 7 different fish species are summarized.

Initial visual assessment is presented in Fig. 1. There are $K = 7, p = 463, n = 50$. In particular, we computed $\frac{1}{p}\|\bar{V}_i - \bar{V}_j\|_2^2$ for all pairs $1 \le i < j \le K$, and summarized them in the boxplot in Fig. 1. We also graphed $\bar{V}_i, i = 1, \ldots, K$ for reference. We can speculate that differences between species 2 and species 3, and species 3 and species 5 are more pronounced than the rest of the species. Note that the data is not sparse and the pairwise differences are quite close.

We apply our method to this problem. We consider the class of matrices that are obtained by permutations of group indices with base matrix A that consists of 7-by-7 blocks as follows

$$(21B, -B, -2B, -3B, -4B, -5B, -6B; -6B, 21B, -B, -2B, -3B, -4B,$$
$$-5B; -5B, -6B, 21B, -B, -2B, -3B, -4B; -4B, -5B, -6B, 21B, -B,$$

$$-2B, -3B; -3B, -4B, -5B, -6B, 21B, -B, -2B; -2B, -3B, -4B,$$

$$-5B, -6B, 21B, -B; -B, -2B, -3B, -4B, -5B, -6B, 21B),$$

where $B = \mathbb{B}_{50}$ is introduced in subsection 4.1. Note that the resulting matrix has 30 diagonals. The test statistic is $T_{50} = 9225.5$, while the bootstrapped quantile is $Q = 3237.7$ for $\alpha = 0.05$ with 10K bootstrap repetitions. Thus, there are statistically significant differences between means for the shapes of fish species.

6. Discussion and conclusion

In this paper we have introduced a framework of bootstrap tests that can address several different testing problems in high-dimensional setup ($n << p$) in a unified fashion. This is done by considering tests statistics that are maximums of sums over sparse classes of convex sets of a novel type. These classes serve the role of a tuning mechanism, which can be chosen based on the particular problem. Basically for a hypothesis about means, one needs to select a finite number of sparse matrices $A^{(l)}, l = 1, \dots, L$, such that $A^{(l)}\mu = 0$ under the null hypothesis. To get a test with high power under a specific alternative, one needs to select these matrices so that components of $A^{(l)}\mu$ are maximized. For instance, in case of very sparse alternative, one needs several more dense matrices, however, that is controlled by condition (A1) and one can only ask for $s = O((\log(LKp))^a)$ for power $a = 3/2$ or $5/2$ depending on moment conditions. On the other hand, for dense means μ_k one can use a single matrix (so $L = 1$) with just a few non-zero diagonals, so in this case s is a finite number. In practice, one can estimate μ by stacking groupwise sample means, say $\hat{\mu}_n$ and then choose a relatively small number of relatively sparse matrices $A^{(l)}$ that maximize some norm of $A^{(l)}\hat{\mu}_n$. This would lead to good empirical power.

Intuitively speaking, we control the sparsity of the tuning mechanism instead of sparsity assumptions on the underlying data.

The resulting bootstrap tests have many advantages. In particular, they are consistent against any fixed alternative; they attain good level and power for large p and small n. They are distribution and correlation free. They are computationally fast. In particular, they are faster by as much as p times in comparison to methods that require precision matrix estimation. Even for sparse alternatives, our tests have comparable performance to that of the existing specialized tests. We only require mild moment and tail assumptions on the distributions. We do not require that the ratios of sample sizes converge to a specific limit. Unlike current tests in literature, we do not require the samples to come from the same distribution, the tail and moment conditions have B_n that can grow with n in the non-identical case.

The only drawback is that our methods have the rate of convergence of the distance between the test statistic and its bootstrapped version of at best $(\log(LKp))^{3/2}n^{-1/2}\log n$ for bounded distributions, and at worst this rate is $(\log(LKp))^{5/6}n^{-1/6}$ for variables with finite 4-th moments and control of the maximum of the variables, which is quite far from a parametric rate. Our framework can accommodate tests of the type as in Lin et al. (2021), where they are exploiting the covariance structure. In particular, the decay of covariance components with the dimension p leads to specialized tests with nice near-parametric rates as in Lin et al. (2021). However, this covariance structure has to be verified in practice, which could be difficult in general apart from functional and sparse count data. Intuitively, such covariance structures give a dimension reduction mechanism and the effective dimension becomes of the same order as n.

Our work is different from Zhang et al. (2018) who proposed MANOVA test that adapts to the sparsity of the alternative based on the data. Our work comes with exact rates for the tests, see Corollary 2 unlike their tests. Moreover, our methods allow for growing $K = K_n$ as long as $\log(LKp) = o(n^\delta)$ for $\delta \in (0, 1/3)$ for bounded random variables and for $\delta \in (0, 1/5)$ for random variables with finite 4-th moments. The computation is also much easier as we do not require estimation of individual elements of the covariances. We can allow all the quantities $s = s_n, B_n, L = L_n, K = K_n$, and $p = p_n$ to grow with n. We can consider extreme particular cases. For example, if only s is allowed to grow then it can be as large as $s = o(n^{1/2})$. If only p is allowed to grow, it can be as large as $\log p = o(n^{1/3}(\log n)^{-2/3})$. The same argument holds for K and L. One can re-balance different growth assumptions put on s, K, L, and p. These growth assumptions are better than those in Zhang et al. (2018).

From a probabilistic point of view, this work also introduces a novel class of convex classes, for which the Berry-Esseen type results are obtained. This new class of convex sets generalizes the classes of half-spaces and hyper-rectangles to classes of "hyper-polygons", which are linear transformations of intersections of a finite number of hyper-spaces. Such sets are sparse in the sense of Chernozhukov et al. (2017). We manage to explicitly track the effect of the sparsity parameter s on the rates. Under certain tail and moment assumptions on the distribution, this effect is linear. Chernozhukov et al. (2017) did not establish this dependency explicitly, since it was buried in the complicated and intricate implicit geometric structures.

Finally, we remark that the multipliers $e_1, \dots, e_n$ can be chosen mean-zero unit-variance random variables from a sub-Gaussian distribution. One needs to refine the theoretical derivations as done in the recent work of Chernozhukov et al. (2019). For instance, Rademacher multipliers are quite popular in literature. Our test will work with such multipliers and it will have all the properties listed in this paper.

Acknowledgements

The authors acknowledge the support by Michigan State University High Performance Computing Center through computational resources provided by the Institute for Cyber-Enabled Research. The research was supported in part by NSF Grants DMS-1612867 and DMS-2111251.

The authors would like to thank the anonymous referees, an Associate Editor, and the Editor for their constructive comments that improved the quality of this paper.

Appendix A. Technical conditions and the main theoretical result

The k-th group after centering consists of $V_{k,1}-\mu_k,\dots,V_{k,n}-\mu_k$ independent observations in $\mathbb{R}^p$. Now stack those vectors $V_{1,1}-\mu_1,\dots,V_{K,1}-\mu_K$ into X_1, stack vectors $V_{1,2}-\mu_1,\dots,V_{K,2}-\mu_K$ into X_2, and so on, stack $V_{1,n}-\mu_1,\dots,V_{K,n}-\mu_K$ into X_n. We obtain centered long vectors $X_i\in\mathbb{R}^d$ with $d=Kp$. Note that the test statistics T_n are functionals of normalized sums $S^X=\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$. Note that using long stacked vectors the test statistics can be rewritten as

$$T_n=\max_{l=1,\dots,L}\max_{j=1,\dots,d}\left([A^{(l)}S_n^X]_j+n^{1/2}[A^{(l)}\mu]_j\right),$$

where $\mu\in\mathbb{R}^d$ consists of vectors $\mu_k\in\mathbb{R}^p, k=1,\dots,K$, stacked into one long vector. Under H_0 test statistics become

$$T_n=\max_{l=1,\dots,L}\max_{j=1,\dots,d}[A^{(l)}S_n^X]_j,$$

due to condition (A2). Also note that the bootstrapped statistics can be rewritten as $T_n^e=\max_{l=1,\dots,L}\max_{j=1,\dots,d}[A^{(l)}S_n^{eX}]_j$.

Introduce the following moment and tail conditions collectively denoted by (C). Let $\{B_n, D_n, n=1,2,\dots\}$ be sequences of positive constants, possibly growing to infinity. Let $\sigma_j^2=\mathbb{E}(S_j^X)^2$.

(S) $\min_{1\le j\le d}\sigma_j\ge b>0$ for some fixed constant b.

(E1) $|X_{ij}|\le B_n\sigma_j\ \forall i=1,\dots,n\ \forall j=1,\dots,d$ a.s.

(Eψ) For Orlicz norms based on function ψ, we have $\|X_{ij}\|_\psi\le B_n\ \forall i=1,\dots,n\ \forall j=1,\dots,d$.

(E$\psi\sigma$) For Orlicz norms based on function ψ, we have $\|X_{ij}/\sigma_j\|_\psi\le B_n\ \forall i=1,\dots,n\ \forall j=1,\dots,d$.

Recall $\|Y\|_{L_q}^q:=\mathbb{E}\sum_{j=1,\dots,d}|Y_j|^q$ for a random vector $Y\in\mathbb{R}^d$.

(E3) $\|\max_{1\le j\le d}|X_{ij}|\|_{L_q}\le D_n\ \forall i=1,\dots,n$.

(E3σ) $\|\max_{1\le j\le d}|X_{ij}/\sigma_j|\|_{L_q}\le B_n\ \forall i=1,\dots,n$.

(M) $n^{-1}\sum_{i=1}^n\mathbb{E}X_{ij}^4\le B_n^2\ \forall j=1,\dots,d$.

(Mσ) $n^{-1}\sum_{i=1}^n\mathbb{E}X_{ij}^4\le B_n^2\sigma_j^4\ \forall j=1,\dots,d$.

For completeness recall the conditions on the class of matrices:

(A1) $\sum_{m=1}^d|A_{jm}^{(l)}|\le s\ \forall j=1,\dots,d,\ \forall l=1,\dots,L$;

(A2) $\sum_{k=1}^K A_{j(q+(k-1)p)}^{(l)}=0\ \forall j=1,\dots,d\ \forall q=1,\dots,p\ \forall l=1,\dots,L$;

(A3) $\min_{1\le l\le L}\min_{1\le j\le d}[A^{(l)}\bar{\Sigma}(A^{(l)})^T]_{jj}\ge b>0$ for some fixed constant b.

We utilize multiplier bootstrap, thus, introduce

$$S_n^{eX}=\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(X_i-\bar{X}),\ \bar{X}=\frac{1}{n}\sum_{i=1}^n X_i,$$

where a vector $e=(e_1,\dots,e_n)$ of i.i.d. $N(0,1)$ random variables is independent of $X_i, i=1,\dots,n$. Consider the Kolmogorov distance defined as

$$KD=\sup_{u\in\mathbb{R}}\left|P(\max_{l=1,\dots,L}\max_{j=1,\dots,d}[A^{(l)}S_n^X]_j\le u)-P_e(\max_{l=1,\dots,L}\max_{j=1,\dots,d}[A^{(l)}S_n^{eX}]_j\le u)\right|,$$

where P_e stands for the probability with respect to the Gaussian vector e only. Recall that $\psi_p(u)=e^{u^p}-1$ for $p\ge1$.

Theorem 2. *Suppose conditions (A1)–(A3) hold. Under condition (E1)*

$$KD\le\frac{CsB_n(\log(Ld))^{3/2}\log n}{\sqrt{n}};$$

under conditions (Mσ) and (E$\psi_2\sigma$)

$$KD\le C\left(\frac{sB_n(\log(Ld))^{3/2}\log n}{\sqrt{n}}+\frac{s^2B_n(\log(Ld))^2}{\sqrt{n}}\right);$$

under conditions (Mσ) and (E3σ) for some $q \geq 4$

$$KD \leq C\left(\frac{sB_n(\log(Ld))^{3/2}\log n}{\sqrt{n}} + \frac{s^4 B_n^2(\log(Ld))^2\log n}{n^{1-2/q}} + \left(\frac{s^q B_n^q(\log(Ld))^{3q/2-4}(\log n)(\log(Ldn))}{n^{q/2-1}}\right)^{1/(q-2)}\right);$$

under conditions (M) and (Eψ_1)

$$KD \leq C\left(\frac{s^{2/3}B_n^{1/3}(\log(Ld))^{5/6}}{n^{1/6}} + \frac{s^{4/3}B_n^{2/3}(\log(Ld))^{2/3}(\log n)^{2/3}}{n^{1/3}}\right);$$

under conditions (M) and (E3) for some $q > 2$

$$KD \leq C\left(\frac{s^{2/3}B_n^{1/3}(\log(Ld))^{5/6}}{n^{1/6}} + \frac{s^{2/3}D_n^{2/3}(\log(Ld))^{1-2/(3q)}}{n^{1/3-2/(3q)}}\right);$$

where the constant $C > 0$ depends on b from (A3) in all cases and it also depends on q when q is used.

We remark that Kolmogorov distance KD goes to 0 as n grows as long as B_n, D_n, L, d, and s grow not too fast. In particular, under condition (E1) we need $sB_n(\log(Ld))^{3/2} = o(\sqrt{n}/\log n)$, which would be the same under (Mσ) and (Eψ_2) if $s(\log(Ld))^{1/2} = O(\log n)$ and otherwise under (Mσ) and (Eψ_2) we need $s^2 B_n(\log(Ld))^2 = o(\sqrt{n})$. Meanwhile, under (M$\sigma$) and (E3$\sigma$) with $q = 4$ we need $s^4 B_n^2(\log(Ld))^2 = o(\sqrt{n}/\log n)$. However, for $q > 4$ three conditions should be satisfied without a clear winner: $sB_n(\log(Ld))^{3/2} = o(\sqrt{n}/\log n)$, $s^4 B_n^2(\log(Ld))^2 = o(n^{1-2/q}/\log n)$, and the most intricate condition of the three $s^{1-2/(q-2)}B_n^{1-2/(q-2)}(\log(Ld))^{3/2} = o(\sqrt{n}/\log n)$. Under (M) and (E$\psi_1$) we have a special case when $\log(Ld)$ grows faster or the same as $(\log n)^{2/3}$ and $s^2 B_n^2(\log(Ld))^{5/2} = o(\sqrt{n})$, otherwise we need both $s^2 B_n^2(\log(Ld))^{5/2} = o(\sqrt{n})$ and $s^2 B_n \log(Ld) = o(\sqrt{n}/\log n)$. Finally, under (M) and (E3) we need both $s^2 B_n(\log(Ld))^{5/2} = o(\sqrt{n})$ and $s^2 D_n^2(\log(Ld))^{3-2/q} = o(n^{1-2/q})$.

Note that the proof of Theorem 1 follows from the proof of Theorem 2 immediately.

We also remark that one can choose a sequence of $\alpha_n \downarrow 0$ such that $\sum_{n=1}^{\infty}\alpha_n < \infty$ and consider events in Theorem 1 and Corollary 1 in the spirit of Proposition 4.1 and remark 4.2 in Chernozhukov et al. (2017). Then apply Borel-Cantelli lemma to obtain the results with probability tending to 1.

Appendix B. Proofs

Here we provide the outline of the proofs. Theorem 2 follows from Corollary 2.1 and Corollary 3.1 in Chernozhukov et al. (2021) and Corollary 3.1 in Koike (2020), which are improvements of the Key Lemma 5.1 in Chernozhukov et al. (2017).

The main trick is to stack $(A^{(1)}X_i, \ldots, A^{(L)}X_i)$ into Ld-dimensional vectors $\tilde{X}_i$ for $i = 1, \ldots, n$. Then for any $u = (u_1, \ldots, u_L) \in \mathbb{R}^{Ld}$

$$P(S_n^X \in C) = P\left(n^{-1/2}\sum_{i=1}^{n} A^{(l)}X_i \leq u_l,\ l = 1, \ldots, L\right) = P(S_n^{\tilde{X}} \leq u).$$

Next, one just needs to check that conditions of Corollary 2.1 in Chernozhukov et al. (2021) and Corollary 3.1 in Koike (2020) for $\tilde{X}_i$ follow from conditions in Theorem 2 on X_i.

Proof of Theorem 2. We start by observing that for $j = 1, \ldots, LKp$

$$\tilde{\sigma}_j^2 = \mathbb{E}\left(\frac{1}{n}(\sum_{i=1}^{n}\tilde{X}_{ij})^2\right) = [A^{(l)}\bar{\Sigma}(A^{(l)})^T]_{j_0 j_0}$$

for some $l = 1, \ldots, L$, $j_0 = 1, \ldots, Kp$, where $\bar{\Sigma}$ from condition (A3) can be written as $\bar{\Sigma}_{km} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_{ik}X_{im})$. We remark that conditions (A3) and (A1) imply condition (S) for $\tilde{\sigma_j}$ with a different b that is related to b and the eigenvalues of $A^{(l)}$.

Now, condition (E1) $|X_{ij}| \leq B_n\sigma_j$ implies after (A1)

$$|\tilde{X_{ij}}| = |\sum_{k=1}^{Kp} A_{j_0 k}^{(l)}X_{ik}| \leq \max_{1\leq k\leq Kp}|X_{ik}|\sum_{k=1}^{Kp}|A_{j_0 k}^{(l)}| \leq s\max_{1\leq k\leq Kp}\sigma_k B_n \leq \kappa s B_n\tilde{\sigma}_j,$$

where κ is such a number that

$$\max_{1\leq k\leq Kp}\frac{\sigma_k^2}{\tilde{\sigma_j}^2} \leq \kappa^2\ \forall j = 1, \ldots, LKp.$$

Existence of κ follows from conditions (A1) and (A3), it is related to the ratio of smallest and largest eigenvalues of $\bar{\Sigma}$ and $A^{(l)}\bar{\Sigma}(A^{(l)})^T$. Thus, $\tilde{X}_i, i=1,\dots,n$, satisfy condition (E1) with $\kappa s B_n$ in place of B_n.

Next, consider condition (E$\psi\sigma$) for $\tilde{X}$. Since ψ is a convex increasing positive function, we have

$$\begin{aligned}
&\mathbb{E}\psi\left(\left|\frac{\tilde{X}_{ij}}{\tilde{c}\tilde{\sigma}_j}\right|\right) = \mathbb{E}\psi\left(\frac{1}{\tilde{c}\tilde{\sigma}_j}\left|\sum_{k=1}^{Kp}\frac{A^{(l)}_{j0k}}{\sum_{m=1}^{Kp}|A^{(l)}_{j_0m}|}X_{ik}\right|\sum_{m=1}^{Kp}|A^{(l)}_{j_0m}|\right)\\
&\le \mathbb{E}\psi\left(\frac{s}{\tilde{c}\tilde{\sigma}_j}\sum_{k=1}^{Kp}\frac{|A^{(l)}_{j0k}|}{\sum_{m=1}^{Kp}|A^{(l)}_{j_0m}|}|X_{ik}|\right) \le \sum_{k=1}^{Kp}\frac{|A^{(l)}_{j0k}|}{\sum_{m=1}^{Kp}|A^{(l)}_{j_0m}|}\mathbb{E}\psi\left(\frac{s}{\tilde{c}\tilde{\sigma}_j}|X_{ik}|\right)\\
&\le \sum_{k=1}^{Kp}\frac{|A^{(l)}_{j0k}|}{\sum_{m=1}^{Kp}|A^{(l)}_{j_0m}|}\mathbb{E}\psi\left(\frac{s\kappa}{\tilde{c}}\frac{|X_{ik}|}{\sigma_k}\right),
\end{aligned}$$

where we used definition of κ. Now, using the definition of Orlicz norm we have

$$\begin{aligned}
&\left\|\frac{\tilde{X}_{ij}}{\tilde{\sigma}_j}\right\|_\psi := \inf_{\tilde{c}>0}\left\{\mathbb{E}\psi\left(\frac{|\tilde{X}_{ij}|}{\tilde{c}\tilde{\sigma}_j}\le 1\right)\right\}\\
&\le \inf_{\tilde{c}>0}\left\{\sum_{k=1}^{Kp}\frac{|A^{(l)}_{j0k}|}{\sum_{m=1}^{Kp}|A^{(l)}_{j_0m}|}\mathbb{E}\psi\left(\frac{s\kappa}{\tilde{c}}\frac{|X_{ik}|}{\sigma_k}\right)\le 1\right\}\\
&\le s\kappa\max_{1\le k\le Kp}\inf_{c>0}\{\mathbb{E}\psi\left(\frac{1}{c}\frac{|X_{ik}|}{\sigma_k}\right)\le 1\} \le s\kappa B_n,
\end{aligned}$$

where $\tilde{c} = cs\kappa$. Thus, $\tilde{X}_i, i=1,\dots,n$, satisfy condition (E$\psi\sigma$) with $\kappa s B_n$ in place of B_n. A similar argument works for condition (Eψ) with sB_n in place of B_n. Also note that (E3σ) and (E3) are proved by the similar argument with $\phi(u) = \|u\|^q_{L_q}$ with sD_n in place of D_n.

Finally, consider condition (Mσ) for $\tilde{X}_i$. Indeed, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\frac{\tilde{X}_{ij}}{\tilde{\sigma}_j}\right)^4 = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\frac{1}{\tilde{\sigma}_j}\sum_{k=1}^{Kp}\frac{A^{(l)}_{j0k}}{\sum_{m=1}^{Kp}|A^{(l)}_{j_0m}|}X_{ik}\right)^4\left(\sum_{m=1}^{Kp}|A^{(l)}_{j_0m}|\right)^4\\
&\le \frac{s^4}{n\tilde{\sigma}_j{}^4}\sum_{i=1}^{n}\left\|\sum_{k=1}^{Kp}\frac{A^{(l)}_{j0k}}{\sum_{m=1}^{Kp}|A^{(l)}_{j_0m}|}X_{ik}\right\|^4_{L_4}\\
&\le \frac{s^4}{n\tilde{\sigma}_j{}^4}\sum_{i=1}^{n}\left(\sum_{k=1}^{Kp}\frac{A^{(l)}_{j0k}}{\sum_{m=1}^{Kp}|A^{(l)}_{j_0m}|}\|X_{ik}\|_{L_4}\right)^4\\
&\le \frac{s^4}{n\tilde{\sigma}_j{}^4}\sum_{i=1}^{n}\sum_{k=1}^{Kp}\frac{|A^{(l)}_{j0k}|}{\sum_{m=1}^{Kp}|A^{(l)}_{j_0m}|}\|X_{ik}\|^4_{L_4} \le \kappa^4 s^4 B_n^2,
\end{aligned}$$

where we used convexity property of $\|u\|_{L_4}$ and u^4. Thus, $\tilde{X}_i, i=1,\dots,n$, satisfy condition (Mσ) with $\kappa^2 s^2 B_n$ in place of B_n. A similar argument works for condition (M).

Now the statement of Theorem 2 follows from Corollary 2.1 and Corollary 3.1 in Chernozhukov et al. (2021) and Corollary 3.1 in Koike (2020) applied to $\tilde{X}_i, i=1,\dots,n$. □

Corollary 1 is proved similarly to establishing consistency of the test.

Proof of Corollary 1. Note that

$$\begin{aligned}
&P(T_n \ge Q_\alpha | H_A^{(n)}) = P\Bigg(\max_{l=1,\dots,L}\max_{j=1,\dots,d}\Bigg[n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{q=1}^{p}A^{(l)}_{j(q+(k-1)p)}[V_{k,i}]_q\\
&-\sqrt{n}[A^{(l)}\mu_A^{(n)}]_j + \sqrt{n}[A^{(l)}\mu_A^{(n)}]_j\Bigg] \ge Q_\alpha\Bigg)\\
&\ge P\Bigg(\max_{l=1,\dots,L}\max_{j=1,\dots,d}\Bigg[n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{q=1}^{p}A^{(l)}_{j(q+(k-1)p)}[V_{k,i}]_q - \sqrt{n}[A^{(l)}\mu_A^{(n)}]_j
\end{aligned}$$

$$\geq Q_\alpha - c_n, \min_{l=1,\dots,L} \min_{j=1,\dots,d} [A^{(l)}\mu_A^{(n)}]_j \geq c_n n^{-1/2}\Big) - P_e(T_n^e \geq Q_\alpha - c_n | H_A^{(n)})$$
$$+P_e(T_n^e \geq Q_\alpha - c_n | H_A^{(n)}) \geq P_e(T_n^e \geq Q_\alpha - c_n | H_A^{(n)}) - KD \to 1$$

as $n \to \infty$. □

In order to establish Corollary 2 we need to check that conditions of Theorem 2 are satisfied for $\widetilde{X}_i$ obtained by stacking the variables $\widetilde{V_{t,i}}, t = 1, \dots, K$.

Proof of Corollary 2. Indeed, condition (A1) becomes

$$\sum_{m=1}^{d} |\widetilde{A_{jm}^{(l)}}| \leq \max_{k=1,\dots K} \lambda_k^{1/2} \sum_{m=1}^{d} |A_{jm}^{(l)}| \leq s \max_{k=1,\dots K} \lambda_{k,n}^{1/2} = \tilde{s}.$$

Due to independence we have

$$\mathbb{E}(\widetilde{X_{i(q_1+(t_1-1)p)}} \widetilde{X_{i(q_2+(t_2-1)p)}}) = \lambda^{-1}_{(q_1+(t_1-1)p)} \lambda^{-1}_{(q_2+(t_2-1)p),n} \Sigma^{(i)}_{(q_1+(t_1-1)p)(q_2+(t_2-1)p)}$$
$$+I(t_1 = t_2)\lambda^{-1}_{(q_1+(t_1-1)p),n} \lambda^{-1}_{(q_2+(t_1-1)p)} \sum_{r_1=1}^{\lambda_{t_1}-1} \Sigma^{(n+(i-1)(\lambda_{t_1}-1)+r_1)}_{(q_1+(t_1-1)p)(q_2+(t_2-1)p)} I(r_1 = r_2).$$

Condition (A3) becomes

$$\sum_{k=1}^{d} \sum_{m=1}^{d} \widetilde{A_{ik}^{(l)}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(\widetilde{X_{ik}} \widetilde{X_{im}}) \widetilde{A_{jm}^{(l)}} \geq \tilde{b}.$$

Note that new $\tilde{b}$ is proportional to old b. All the moment and tail conditions will work for $B_n(\min_{t=2,\dots K} \lambda_{t,n})^{-2}$ in place of B_n. Then we apply Theorem 2 to new variables and get the desired result. □

References

Cai, T., Liu, W., Luo, X., 2011. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. J. Am. Stat. Assoc. 106 (494), 594–607.

Cai, T., Xia, Y., 2014. High-dimensional sparse MANOVA. J. Multivar. Anal. 131, 174–196.

Chen, X., 2018. Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. Ann. Stat. 46 (2), 642–678.

Chen, S.X., Li, J., Zhong, P-S., 2019. Two-sample and ANOVA tests for high dimensional means. Ann. Stat. 47 (3), 1443–1474.

Chernozhukov, V., Chetverikov, D., Koike, Y., 2021. Nearly optimal central limit theorem and bootstrap approximations in high dimensions. Preprint.

Chernozhukov, V., Chetverikov, D., Kato, K., Koike, Y. Improved central limit theorem and bootstrap approximations in high dimensions. Preprint.

Chernozhukov, V., Chetverikov, D., Kato, K., 2017. Central limit theorems and bootstrap in high dimensions. Ann. Probab. 45 (4), 2309–2352.

Fujikoshi, Y., Himeno, T., Wakaki, H., 2004. Asymptotic results of a high dimensional MANOVA test and power comparisons when the dimension is large compared to the sample size. J. Japan Statist. Soc. 34, 19–26.

Koike, Y., 2020. Notes of the dimension dependence in high-dimensional central limit theorems for hyperrectangles. Jpn. J. Stat. Data Sci. 4 (1), 257–297.

Lee, D.-J., Archibald, J.K., Schoenberger, R.B., Dennis, A.W., Shiozawa, D.K., 2008. Contour matching for fish species recognition and migration monitoring. In: Applications of Computational Intelligence in Biology. Springer, pp. 183–207.

Lin, Z., Lopes, M., Müller, H.-G., 2021. High-dimensional MANOVA via bootstrapping and its application to functional and sparse count data. J. Am. Stat. Assoc. https://doi.org/10.1080/01621459.2021.1920959.

Lloyd, C., 2005. Estimating test power adjusted for size. J. Stat. Comput. Simul. 75 (11), 921–933.

Pang, H., Liu, H., Vanderbei, R., 2014. The fastclime package for linear programming and large scale precision matrix estimation in R. J. Mach. Learn. Res. 15 (1), 489–493.

Santo, S., Zhong, P-S., Homogeneity tests of covariance and change-points identification for high-dimensional functional data. 2021. Manuscript.

Schott, J.R., 2007. Some high-dimensional tests for a one-way MANOVA. J. Multivar. Anal. 98, 1825–1839.

Srivastava, M., 2007. Multivariate theory for analyzing high dimensional data. J. Japan Statist. Soc. 37, 53–86.

Watanabe, H., Hyodo, M., Nakagawa, S., 2020. Two-way MANOVA with unequal cell sizes and unequal cell covariance matrices in high-dimensional settings. J. Multivar. Anal. 179, 1–15.

Xue, K., Yao, F., 2020. Distribution and correlation free two-sample test of high-dimensional means. Ann. Stat. 48 (3), 1304–1328.

Zhang, J-T., Guo, J., Zhou, B., 2017. Linear hypothesis testing in high-dimensional one-way MANOVA. J. Multivar. Anal. 155, 200–216.

Zhang, M., Zhou, C., He, Y., Liu, B., 2018. Data-adaptive test for high-dimensional multivariate analysis of variance problem. Aust. N. Z. J. Stat. 60 (4), 447–470.

Zhang, Y-C., Sakhanenko, L., 2019. The Naive Bayes classifier for functional data. Stat. Probab. Lett. 152, 137–146.