Journal of Machine Learning Research 23 (2022) 1-48 Submitted 10/24; Revised 5/22; Published 9/22

Two-Sample Robust Nonparametric Tests for Matrix-Valued Data: a Double Projection-Averaging Method

Jicai Liu LIUJICAI1234@126.COM
School of Statistics and Mathematics
Shanghai Lixin University of Accounting and Finance
Shanghai, 201209, China

Jinhong Li 52214404007@STU.ECNU.EDU.CN
Department of Statistics
East China Normal University,
Shanghai, 200062, China

Jinhong You JOHNYOU07@163.COM
School of Statistics and Management
Shanghai University of Finance and Economics
Shanghai, 200433, China

Riquan Zhang ZHANGRIQUAN@163.COM
School of Statistics and Information
Shanghai University of International Business and Economics
Shanghai, 201620, China

Editor:

Abstract

Matrix-valued data are becoming increasingly commom, and in this paper three robust nonparametric two-sample testing approaches are proposed for unstructured, symmetric, and symmetric positive-definite (SPD) matrix-valued data, respectively. In particular, the novel double projection-averaging characteristic-function-based distance (DPCD) is developed to test the equality of two unstructured matrix-variate distributions. To take advantage of the matrix structure, the DPCD is refined by introducing (i) the symmetric projection-averaging characteristic-function-based distance (SPCD) for symmetric matrix-variate data by a symmetric projection and (ii) SPCD_{Log} for SPD matrix-variate data by a matrix logarithm transformation.The asymptotic behaviours and robustness of the proposed test statistics are studied, and extensive simulation results show that they are highly competitive with existing tests in a wide range of settings. The proposed approaches are illustrated further by empirical analyses of two real datasets.

Keywords: matrix-valued data, nonparametric two-sample test, inverse multiquadric kernel, maximum mean discrepancy, projection averaging

1. Introduction

Matrix-variate observations are commonly encountered in a wide variety of important applications, including time series (Chen et al., 2020; Chen and Fan, 2021), spatiotemporality (Mardia and Goodall, 1993), genetics (Allen and Tibshirani, 2012; Ning and Liu, 2013), and brain imaging (Poldrack et al., 2011; Mori, 2016), to name but a few. In the analysis of matrix-variate observations, it is often interesting to investigate the characteristics of their distribution. For example, in finance research, we wish to explore whether the Shenzhen and Shanghai stock exchanges of China have the same pattern during a specific period of time. To the end, we collect the 1-min intraday stock data of the top 50 stocks in each stock market, calculate the daily-return covariance matrices for each stock and test whether the two groups of covariance data have same distributions. Obviously, such matrices are symmetric and usually positive-definite, it is very challenging to investigate their distribution characteristics. Another example is from an electroencephalogram (EEG) study of healthy adolescents and adolescents with symptoms of schizophrenia (Borisov et al., 2005). In that study, the EEG signals for each subject were recorded at a sampling rate of 128 Hz from 16 electrodes, and the measurement took 60 s, so the EEG signals for each subject can be represented as a 7680×16 matrix. It is scientifically interesting to study whether the EEG signals are homogeneous between the two groups of adolescents, i.e. whether the EEG signals have different distributions.

Let $\mathbf{X} \in \mathbb{R}^{p\times q}$ and $\mathbf{Y} \in \mathbb{R}^{p\times q}$ be two random matrices defined on a common probability space $(\Omega, \mathcal{A}, P)$ with distributions $P_{\mathbf{X}}$ and $P_{\mathbf{Y}}$, respectively, where $\mathbb{R}^{p\times q}$ denotes the set of all $p \times q$ real matrices. We aim to test the equality of the two distributions, i.e.

$$H_0 : P_{\mathbf{X}} = P_{\mathbf{Y}} \quad \text{versus} \quad H_1 : P_{\mathbf{X}} \neq P_{\mathbf{Y}}. \tag{1}$$

For $p = 1$ or $q = 1$, problem (1) is the classical vector-valued two-sample problem whose long history dates back to the Kolmogorov–Smirnov test (Kolmogorov, 1933; Smirnov, 1948). Earlier work for univariate two-sample tests includes the Anderson–Darling test (Anderson and Darling, 1952) and the Cramér–von Mises criterion (Anderson, 1962). It is non-trivial to extend these tests to multivariate data, but much progress has been made in this direction in the past decade. For example, Gretton et al. (2012) proposed the maximum mean discrepancy (MMD) tests based on the kernel method, Székely and Rizzo (2013) introduced the energy distance test statistics by the Euclidean distance, and Kim et al. (2020) and Li and Zhang (2020) generalized the Cramér–von Mises statistic through the projection-averaging method (Baringhaus and Franz, 2004; Escanciano, 2006).

Testing problem (1) has received limited attention in the literature. In the paper, we propose the double projection-averaging characteristic-function-based distance (DPCD) to measure the difference between two matrix-variate distributions. Key to developing the DPCD is using a double projection technique that projects a matrix onto the row and column directions, and transforms matrix-valued data into scalar-valued data. Note that our double projection-averaging method for the DPCD is different from the existing projection-averaging (Baringhaus and Franz, 2004; Escanciano, 2006). The latter involves only one-direction projection, whereas the former performs both row and column projections and so must handle more-complicated integration. Moreover, it is more challenging to choose suitable weight functions to ensure a tractable calculation form for the DPCD. Our method has several appealing properties: (i) it is applicable for arbitrary matrix-valued data, including scalar and vector-valued data; (ii) it has a closed-form expression that is easily estimated from the data; (iii) it requires no assumption about the matrix-variate distribution and is robust to heavy-tailed data.

Symmetric or symmetric positive-definite (SPD) matrix-valued data arise in many applications; for example, in network analysis, a network with undirected edges can be represented by a symmetric matrix (Newman, 2010), and in diffusion tensor imaging, the diffusion tensor is a 3×3 covariance matrix that is estimated at each voxel in the brain (Schwartzman et al., 2008). Another goal of this paper is to test problem (1) for symmetric or SPD matrix-valued data. To take advantage of the symmetry of the matrix structure, we improve the double projection by its symmetric version and introduce the symmetric projection-averaging characteristic-function-based distance (SPCD) for symmetric matrix-valued data. For SPD matrix-valued data, we first transform the SPD matrices into real symmetric ones by the matrix logarithm transformation (Arsigny et al., 2007), and we propose the $\mathrm{SPCD}_{\mathrm{Log}}$ distance. We show that the SPCD and $\mathrm{SPCD}_{\mathrm{Log}}$ inherit all the desirable properties of the DPCD.

We summarize the main contributions of this paper as follows.

(a) For unstructured, symmetric, and SPD matrix-valued data, we propose three novel two-sample testing approaches, i.e. the DPCD, SPCD, and $\mathrm{SPCD}_{\mathrm{Log}}$ test methods, respectively. These new test methods require no assumption about the matrix-variate distribution and are also adapted to heavy-tailed data.

(b) We establish the asymptotic behaviours of the proposed test statistics, including their limiting distributions under the null and alternative hypotheses.

(c) Because the limiting null distributions of the DPCD, SPCD, and $\mathrm{SPCD}_{\mathrm{Log}}$ test statistics are intractable, we further approximate the asymptotic null distribution through the random-permutation test procedure and establish the consistency.

(d) We show that the proposed test methods are robust to the contamination model. The results are consistent with those proposed by Kim et al. (2020).

(e) Based on the three test methods, we propose three new kernel functions as useful tools for handling statistical learning problems involving matrix-variate data.

The rest of this paper is organized as follows. In Section 2, we introduce the DPCD, SPCD and SPCDLog metrics for problem (1). In Section 3, we study the asymptotic behaviours of the three metrics under the null and alternative hypotheses, obtain the critical value via permutation, and establish its theoretical properties. In Section 4, we establish some connection with the vector-valued methods. We report our numerical results in Sections 5 and 6, and we provide a brief discussion in Section 7. All the technical proofs are deferred to the Appendices.

2. Methodology

In the section, we propose the DPCD, SPCD and $\mathrm{SPCD}_{\mathrm{Log}}$, to test problem (1) for unstructured, symmetric, and symmetric positive-definite matrix-valued data, respectively. Meanwhile, we provide some important properties of the three metrics and their empirical estimates.

2.1 DPCD for matrix-valued data

Here, we begin by proposing the DPCD via double projection-averaging, which is introduced in the following lemma.

Lemma 1 *Let $\boldsymbol{X} \in \mathbb{R}^{p\times q}$ and $\boldsymbol{Y} \in \mathbb{R}^{p\times q}$ be two random matrices with distributions $P_{\boldsymbol{X}}$ and $P_{\boldsymbol{Y}}$, respectively. Then, we have that*

$$P_{\boldsymbol{X}} = P_{\boldsymbol{Y}} \Longleftrightarrow P_{\alpha^T \mathbf{X}\beta} = P_{\alpha^T \mathbf{Y}\beta} \text{ for any } \alpha \in \mathbb{R}^p, \beta \in \mathbb{R}^q, \tag{2}$$

where $\Longleftrightarrow$ stands for 'equivalent to'.

Lemma 1 suggests that testing problem (1) amounts to testing whether the univariate random variables $\alpha^T\mathbf{X}\beta$ and $\alpha^T\mathbf{Y}\beta$ are homogeneous for any $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$. Because the right-hand side of (2) projects the matrices onto the row and column directions α and β, we call it the double-projection approach. Given α and β, $P_{\alpha^T\mathbf{X}\beta} = P_{\alpha^T\mathbf{Y}\beta}$ is equivalent to the relationship

$$E\{\exp\{it\alpha^T\mathbf{X}\beta\}\} - E\{\exp\{it\alpha^T\mathbf{Y}\beta\}\} = 0 \text{ for any } t \in \mathbb{R},$$

where $i = \sqrt{-1}$ is the imaginary unit. Together with Lemma 1, this indicates that testing problem (1) is equivalent to testing whether

$$\int_{\mathbb{R}^q}\int_{\mathbb{R}^p} \|E\{\exp\{i\alpha^T\mathbf{X}\beta\}\} - E\{\exp\{i\alpha^T\mathbf{Y}\beta\}\}\|^2 dw_1(\alpha)dw_2(\beta) = 0, \tag{3}$$

where $\|f\|^2 = f\bar{f}$, $\bar{f}$ is the complex conjugate of f, and $w_1(\alpha)$ and $w_2(\beta)$ are two nonnegative weight functions.

The double integral in (3) can be viewed as averaging over all the two-direction projections, and so it is an extension of the one-direction projection-averaging method for random vectors; see Baringhaus and Franz (2004), Escanciano (2006), and Kim et al. (2020). Herein, we refer to the double integral in (3) as the double projection-averaging approach. Another important difference between the two methods is that the former uses the distance between two characteristic functions, whereas the latter depends on the distance between two cumulative distribution functions (CDFs).

Note that the double integral in (3) may suffer from computational issues due to intractable integration. To overcome these, we can choose suitable weight functions $w_1(\alpha)$ and $w_2(\beta)$ and ensure a closed-form expression; therefore, how to choose the weight functions is crucial. Fortunately, we find that Gaussian distribution functions are the desirable weight functions.

Definition 1 *Let $\boldsymbol{X} \in \mathbb{R}^{p\times q}$ and $\boldsymbol{Y} \in \mathbb{R}^{p\times q}$ be two random matrices. The DPCD between $\mathbf{X}$ and $\mathbf{Y}$ is defined by*

$$\mathrm{DPCD}(\mathbf{X}, \mathbf{Y}) = \int_{\mathbb{R}^q}\int_{\mathbb{R}^p} \|E\{\exp\{i\alpha^T\mathbf{X}\beta\}\} - E\{\exp\{i\alpha^T\mathbf{Y}\beta\}\}\|^2 dG_1(\alpha)dG_2(\beta),$$

where $G_1(\alpha)$ is the CDF of $N_p(0, \gamma\mathbf{I}_p)$ with $\gamma > 0$, and $G_2(\beta)$ is the CDF of $N_q(0, \mathbf{I}_q)$.

The following theorem shows that the Gaussian-distribution weight functions are reasonable, thereby ensuring that $\mathrm{DPCD}(\mathbf{X}, \mathbf{Y})$ has a closed form.

Theorem 1 **(i)** *Suppose that* $\mathbf{X}_1, \mathbf{X}_2 \overset{i.i.d}{\sim} P_{\mathbf{X}}$ *and, independently,* $\mathbf{Y}_1, \mathbf{Y}_2 \overset{i.i.d}{\sim} P_{\mathbf{Y}}$. *Then,* $\mathrm{DPCD}(\mathbf{X}, \mathbf{Y})$ *can be written as*

$$\begin{aligned} \mathrm{DPCD}(\mathbf{X}, \mathbf{Y}) &= E\{\det(\gamma(\mathbf{X}_1 - \mathbf{X}_2)^T(\mathbf{X}_1 - \mathbf{X}_2) + \boldsymbol{I}_q)^{-1/2}\} \\ &\quad -2E\{\det(\gamma(\mathbf{X}_1 - \mathbf{Y}_2)^T(\mathbf{X}_1 - \mathbf{Y}_2) + \boldsymbol{I}_q)^{-1/2}\} \\ &\quad +E\{\det(\gamma(\mathbf{Y}_1 - \mathbf{Y}_2)^T(\mathbf{Y}_1 - \mathbf{Y}_2) + \boldsymbol{I}_q)^{-1/2}\}, \end{aligned}$$

where $\det(\mathbf{A})$ *is the determinant of any square matrix* $\mathbf{A}$*;*

(ii) $\mathrm{DPCD}(\mathbf{X}, \mathbf{Y}) \geq 0$ *and* $\mathrm{DPCD}(\mathbf{X}, \mathbf{Y}) = 0$ *if and only if* $\mathbf{X}$ *and* $\mathbf{Y}$ *are identically distributed.*

Theorem 1(i) provides a tractable calculation form for $\mathrm{DPCD}(\mathbf{X}, \mathbf{Y})$, and it is easy to see that $\mathrm{DPCD}(\mathbf{X}, \mathbf{Y})$ is also equal to

$$\begin{aligned} \mathrm{DPCD}(\mathbf{X}, \mathbf{Y}) &= E\{\det(\gamma(\mathbf{X}_1 - \mathbf{X}_2)(\mathbf{X}_1 - \mathbf{X}_2)^T + \mathbf{I}_p)^{-1/2}\} \\ &\quad -2E\{\det(\gamma(\mathbf{X}_1 - \mathbf{Y}_2)(\mathbf{X}_1 - \mathbf{Y}_2)^T + \mathbf{I}_p)^{-1/2}\} \\ &\quad +E\{\det(\gamma(\mathbf{Y}_1 - \mathbf{Y}_2)(\mathbf{Y}_1 - \mathbf{Y}_2)^T + \mathbf{I}_p)^{-1/2}\}. \end{aligned}$$

Theorem 1(ii) implies that $\mathrm{DPCD}(\mathbf{X}, \mathbf{Y})$ is generally applicable as an index for the difference between two matrix-variate distributions.

Suppose that $\{\mathbf{X}_i, i = 1, \ldots, n\}$ and $\{\mathbf{Y}_i, i = 1, \ldots, m\}$ are two mutually independent random samples drawn from $P_{\mathbf{X}}$ and $P_{\mathbf{Y}}$, respectively. By Theorem 1(i), we provide an unbiased empirical estimate of $\mathrm{DPCD}(\mathbf{X}, \mathbf{Y})$ based on a U-statistic, given by

$$\begin{aligned} \widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y}) &= \frac{1}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \det(\gamma(\mathbf{X}_i - \mathbf{X}_j)^T(\mathbf{X}_i - \mathbf{X}_j) + \mathbf{I}_q)^{-1/2} \\ &\quad - \frac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \det(\gamma(\mathbf{X}_i - \mathbf{Y}_j)^T(\mathbf{X}_i - \mathbf{Y}_j) + \mathbf{I}_q)^{-1/2} \\ &\quad + \frac{1}{m(m-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{m} \det(\gamma(\mathbf{Y}_i - \mathbf{Y}_j)^T(\mathbf{Y}_i - \mathbf{Y}_j) + \mathbf{I}_q)^{-1/2}. \end{aligned}$$

In practice, $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$ can be easily computed by the following matrix form. Let

$$\mathbf{L} = \begin{bmatrix} \frac{1}{n(n-1)} \mathbf{1}_n \mathbf{1}_n^T & -\frac{1}{mn} \mathbf{1}_n \mathbf{1}_m^T \\ -\frac{1}{mn} \mathbf{1}_m \mathbf{1}_n^T & \frac{1}{m(m-1)} \mathbf{1}_m \mathbf{1}_m^T \end{bmatrix} \quad \text{and} \quad \mathbf{K} = \begin{bmatrix} \mathbf{K}_{\mathbf{X},\mathbf{X}} & \mathbf{K}_{\mathbf{X},\mathbf{Y}} \\ \mathbf{K}_{\mathbf{X},\mathbf{Y}}^T & \mathbf{K}_{\mathbf{Y},\mathbf{Y}} \end{bmatrix}, \tag{4}$$

where $\mathbf{1}_n$ and $\mathbf{1}_m$ are $n \times 1$ and $m \times 1$ vectors of ones,

$$(\mathbf{K}_{\mathbf{X},\mathbf{X}})_{ij} = \begin{cases} \det(\gamma(\mathbf{X}_i - \mathbf{X}_j)^T(\mathbf{X}_i - \mathbf{X}_j) + \mathbf{I}_q)^{-1/2}, & i \neq j \\ 0, & i = j, \end{cases}$$

$$(\mathbf{K}_{\mathbf{X},\mathbf{Y}})_{ij} = \det(\gamma(\mathbf{X}_i - \mathbf{Y}_j)^T(\mathbf{X}_i - \mathbf{Y}_j) + \mathbf{I}_q)^{-1/2},$$

and $(\mathbf{K}_{\mathbf{Y},\mathbf{Y}})_{ij}$ can be similarly defined as $(\mathbf{K}_{\mathbf{X},\mathbf{X}})_{ij}$. Henceforth, $(\mathbf{A})_{jk}$ is denoted as the (j, k)-element of any matrix $\mathbf{A}$. It is easy to see that $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$ can be written as $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y}) = \mathrm{tr}(\mathbf{KL})$.

2.2 SPCD for symmetric matrix-valued data

We consider the two-sample problem for symmetric matrix-valued data. Let $\mathbb{S}_p$ be the real-valued symmetric $p \times p$ matrix space, and assume that $\widetilde{\mathbf{X}} \in \mathbb{S}_p$ and $\widetilde{\mathbf{Y}} \in \mathbb{S}_p$ are two random matrices with distributions $P_{\widetilde{\mathbf{X}}}$ and $P_{\widetilde{\mathbf{Y}}}$, respectively. Our goal is to test

$$H_0 : P_{\widetilde{\mathbf{X}}} = P_{\widetilde{\mathbf{Y}}} \quad \text{versus} \quad H_1 : P_{\widetilde{\mathbf{X}}} \neq P_{\widetilde{\mathbf{Y}}}. \tag{5}$$

Note that a $p \times p$ symmetric random matrix includes only $p(p+1)/2$ different random variables, so testing problem (5) is very different from testing problem (1). To deal with the symmetric structure, we propose a novel symmetric version of the double-projection technique in Lemma 1, which is stated as follows.

Lemma 2 *Let $\widetilde{\boldsymbol{X}} \in \mathbb{S}_p$ and $\widetilde{\boldsymbol{Y}} \in \mathbb{S}_p$ be two random matrices. Then, we have that $P_{\widetilde{\boldsymbol{X}}} = P_{\widetilde{\boldsymbol{Y}}}$ is equivalent to $P_{\alpha^T \widetilde{\mathbf{X}} \alpha} = P_{\alpha^T \widetilde{\mathbf{Y}} \alpha}$ for any $\alpha \in \mathbb{R}^p$.*

Unlike the double projection, Lemma 2 indicates that the symmetric version projects both the rows and columns of the matrix onto a common direction, so we call it the symmetric projection approach. By Lemma 2, we obtain

$$\begin{aligned} P_{\widetilde{\mathbf{X}}} = P_{\widetilde{\mathbf{Y}}} \iff & P_{\alpha^T \widetilde{\mathbf{X}} \alpha} = P_{\alpha^T \widetilde{\mathbf{Y}} \alpha} \text{ for any } \alpha \in \mathbb{R}^p \\ \iff & E\{\exp\{it\alpha^T \widetilde{\mathbf{X}} \alpha\}\} = E\{\exp\{it\alpha^T \widetilde{\mathbf{Y}} \alpha\}\} \text{ for any } t \in \mathbb{R}, \alpha \in \mathbb{R}^p \\ \iff & E\{\exp\{i\alpha^T \widetilde{\mathbf{X}} \alpha\}\} = E\{\exp\{i\alpha^T \widetilde{\mathbf{Y}} \alpha\}\} \\ & \text{ or } E\{\exp\{-i\alpha^T \widetilde{\mathbf{X}} \alpha\}\} = E\{\exp\{-i\alpha^T \widetilde{\mathbf{Y}} \alpha\}\} \text{ for any } \alpha \in \mathbb{R}^p, \end{aligned}$$

which suggests that $P_{\widetilde{\mathbf{X}}} = P_{\widetilde{\mathbf{Y}}}$ is equivalent to

$$\int_{\mathbb{R}^p} \| E\{\exp\{i\alpha^T \widetilde{\mathbf{X}} \alpha\}\} - E\{\exp\{i\alpha^T \widetilde{\mathbf{Y}} \alpha\}\} \|^2 dw(\alpha) = 0, \tag{6}$$

where $w(\alpha)$ is a nonnegative weight function.

Similar to (3), we refer to the process in the left-hand side of (6) as the symmetric projection-averaging approach. The integration in (6) may also suffer from computational issues, but through some careful analysis, we find that a Gaussian distribution function can ensure that the integration has a closed-form expression.

Definition 2 *Let $\widetilde{\boldsymbol{X}} \in \mathbb{S}_p$ and $\widetilde{\boldsymbol{Y}} \in \mathbb{S}_p$ be two symmetric random matrices. The SPCD between $\widetilde{\boldsymbol{X}}$ and $\widetilde{\boldsymbol{Y}}$ is defined by*

$$\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) = \int_{\mathbb{R}^p} \| E\{\exp\{i\alpha^T \widetilde{\mathbf{X}} \alpha\}\} - E\{\exp\{i\alpha^T \widetilde{\mathbf{Y}} \alpha\}\} \|^2 dG_1(\alpha),$$

where $G_1(\alpha)$ is the CDF of $N_p(0, \gamma \mathbf{I}_p)$ with $\gamma > 0$.

The following theorem provides some important properties of $\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})$.

Theorem 2 **(i)** *Suppose that* $\widetilde{\mathbf{X}}_1, \widetilde{\mathbf{X}}_2 \overset{i.i.d}{\sim} P_{\widetilde{\mathbf{X}}}$ *and, independently,* $\widetilde{\mathbf{Y}}_1, \widetilde{\mathbf{Y}}_2 \overset{i.i.d}{\sim} P_{\widetilde{\mathbf{Y}}}$. *Then,* $\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})$ *can be written as*

$$\begin{aligned}\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) = & E\{\det(\boldsymbol{I}_p - 2i\gamma(\widetilde{\mathbf{X}}_1 - \widetilde{\mathbf{X}}_2))^{-1/2}\} - 2E\{\det(\boldsymbol{I}_p - 2i\gamma(\widetilde{\mathbf{X}}_1 - \widetilde{\mathbf{Y}}_2))^{-1/2}\} \\ & + E\{\det(\boldsymbol{I}_p - 2i\gamma(\widetilde{\mathbf{Y}}_1 - \widetilde{\mathbf{Y}}_2))^{-1/2}\};\end{aligned}$$

(ii) $\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) \geq 0$ *and* $\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) = 0$ *if and only if* $\widetilde{\mathbf{X}}$ *and* $\widetilde{\mathbf{Y}}$ *are identically distributed.*

Note that $\det(\mathbf{I}_p - 2i\gamma\widetilde{\mathbf{z}})^{-1/2}$ is actually the characteristic function of the central Wishart distribution $\mathbf{W}_p(df = 1, \gamma\mathbf{I}_p)$ (Anderson, 2003, p. 259), which is well defined for any $\gamma > 0$ and $\widetilde{\mathbf{z}} \in \mathbb{S}_p$. More important properties of $\det(\mathbf{I}_p - 2i\gamma\widetilde{\mathbf{z}})^{-1/2}$ are introduced in Section 4.2. Additionally, although $\det(\mathbf{I}_p - 2i\gamma\widetilde{\mathbf{z}})^{-1/2}$ is an imaginary number, $\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})$ is still a real number. Thus, the SPCD-based test does not cause any theoretical and computational issue.

Next, we develop the empirical estimate of $\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})$ by two mutually independent random samples $\{\widetilde{\mathbf{X}}_1, \ldots, \widetilde{\mathbf{X}}_n\}$ and $\{\widetilde{\mathbf{Y}}_1, \ldots, \widetilde{\mathbf{Y}}_m\}$. By Theorem 2(i), an unbiased estimator of $\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})$ can be given by

$$\begin{aligned}\widehat{\mathrm{SPCD}}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) \;=\; & \frac{1}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \det(\mathbf{I}_p - 2i\gamma(\widetilde{\mathbf{X}}_i - \widetilde{\mathbf{X}}_j))^{-1/2} \\ & - \frac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \det(\mathbf{I}_p - 2i\gamma(\widetilde{\mathbf{X}}_i - \widetilde{\mathbf{Y}}_j))^{-1/2} \\ & + \frac{1}{m(m-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{m} \det(\mathbf{I}_p - 2i\gamma(\widetilde{\mathbf{Y}}_i - \widetilde{\mathbf{Y}}_j))^{-1/2}.\end{aligned}$$

Moreover, we can obtain a matrix form of $\widehat{\mathrm{SPCD}}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})$ given by

$$\widehat{\mathrm{SPCD}}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) = \mathrm{tr}(\widetilde{\mathbf{K}}\mathbf{L}),$$

where $\mathbf{L}$ is defined in (4),

$$\widetilde{\mathbf{K}} = \begin{bmatrix} \widetilde{\mathbf{K}}_{\widetilde{\mathbf{X}},\widetilde{\mathbf{X}}} & \widetilde{\mathbf{K}}_{\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}}} \\ \widetilde{\mathbf{K}}_{\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}}}^{T} & \widetilde{\mathbf{K}}_{\widetilde{\mathbf{Y}},\widetilde{\mathbf{Y}}} \end{bmatrix}, \quad \left(\widetilde{\mathbf{K}}_{\widetilde{\mathbf{X}},\widetilde{\mathbf{X}}}\right)_{ij} = \begin{cases} \det(\mathbf{I}_p - 2i\gamma(\widetilde{\mathbf{X}}_i - \widetilde{\mathbf{X}}_j))^{-1/2}, & i \neq j \\ 0, & i = j, \end{cases}$$

$(\widetilde{\mathbf{K}}_{\widetilde{\mathbf{Y}},\widetilde{\mathbf{Y}}})_{ij}$ can be similarly defined, and $(\widetilde{\mathbf{K}}_{\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}}})_{ij} = \det(\mathbf{I}_p - 2i\gamma(\widetilde{\mathbf{X}}_i - \widetilde{\mathbf{Y}}_j))^{-1/2}$.

2.3 $\mathrm{SPCD}_{\mathrm{Log}}$ for SPD matrix-valued data

In the section, we focus on testing the equality of two SPD matrix-variate distributions. Let $\mathbb{S}_p^+$ be the $p \times p$ SPD matrix space, and assume that $\underline{X} \in \mathbb{S}_p^+$ and $\underline{Y} \in \mathbb{S}_p^+$ are two random matrices with distributions $P_{\underline{\mathbf{X}}}$ and $P_{\underline{\mathbf{Y}}}$, respectively. We aim to test

$$H_0 : P_{\underline{\mathbf{X}}} = P_{\underline{\mathbf{Y}}} \quad \text{versus} \quad H_1 : P_{\underline{\mathbf{X}}} \neq P_{\underline{\mathbf{Y}}}. \tag{7}$$

Table 1: Comparison of the four methods for symmetric positive-definite (SPD) matrices.

	$\mathrm{SPCD}_{\mathrm{Log}}$	$\mathrm{DPCD}_{\mathrm{Log}}$	SPCD	DPCD
Positive-definiteness	Yes	Yes	No	No
Symmetry	Yes	No	Yes	No

The matrix logarithm transformation is a commonly used approach in the literature on the statistical analysis of SPD matrices; see Schwartzman (2006) and Arsigny et al. (2007). Specifically, for any $\mathbf{S} \in \mathbb{S}_p^+$ and its spectral decomposition $\mathbf{S} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$, the logarithm of $\mathbf{S}$ is defined by $\mathrm{Log}(\mathbf{S}) = \mathbf{V}\log(\mathbf{D})\mathbf{V}^{-1}$, where $[\log(D)]_{ii} = \log([D]_{ii})$ and $[\log(D)]_{ij} = 0$ for $i \neq j$. From the definition, we can see that the matrix logarithmic map transforms SPD matrices into real symmetric ones. By this fact, we can apply the logarithmic map and propose the following measure to test (7), given by

$$\mathrm{SPCD}_{\mathrm{Log}}(\underline{\mathbf{X}}, \underline{\mathbf{Y}}) = \mathrm{SPCD}(\mathrm{Log}(\underline{\mathbf{X}}), \mathrm{Log}(\underline{\mathbf{Y}})). \tag{8}$$

By the definition, we can obtain an empirical estimator of $\mathrm{SPCD}_{\mathrm{Log}}(\underline{\mathbf{X}}, \underline{\mathbf{Y}})$ by $\widehat{\mathrm{SPCD}}_{\mathrm{Log}}(\underline{\mathbf{X}}, \underline{\mathbf{Y}}) = \widehat{\mathrm{SPCD}}(\mathrm{Log}(\underline{\mathbf{X}}), \mathrm{Log}(\underline{\mathbf{Y}}))$

Note that any SPD matrix and its log-transformation are two special matrices on $\mathbb{R}^{p\times q}$ and $\mathbb{S}_p$. Thus, we can propose the following measures to test (7), given by

$$\text{(i) } \mathrm{DPCD}_{\mathrm{Log}}(\underline{\mathbf{X}}, \underline{\mathbf{Y}}) = \mathrm{DPCD}(\mathrm{Log}(\underline{\mathbf{X}}), \mathrm{Log}(\underline{\mathbf{Y}}));\ \text{(ii) } \mathrm{DPCD}(\underline{\mathbf{X}}, \underline{\mathbf{Y}});\ \text{(iii) } \mathrm{SPCD}(\underline{\mathbf{X}}, \underline{\mathbf{Y}}),$$

although they may suffer from power loss due to ignoring some structure information. Table 1 describes the abilities of the four methods in capturing the specific matrix structure. Specifically, the $\mathrm{SPCD}_{\mathrm{Log}}$ uses positive-definiteness and symmetry, the $\mathrm{DPCD}_{\mathrm{Log}}$ uses only positive-definiteness, the SPCD uses only symmetry, and the DPCD uses neither.

3. Asymptotic properties

In the section, we study the limiting distributions, permutation test and robustness of $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$. Such theoretical results for $\widehat{\mathrm{SPCD}}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})$ and $\widehat{\mathrm{SPCD}}_{\mathrm{Log}}(\underline{\mathbf{X}}, \underline{\mathbf{Y}})$ can be obtained in similar ways and so are omitted to conserve space.

3.1 Limiting distributions

Here, we study the limiting distributions under the null and alternative hypotheses of $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$ by means of the theory of two-sample U-statistics (Lee, 1990). Define the two-sample symmetric kernel function

$$\begin{aligned} h(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}_1, \mathbf{y}_2) \;=\;& \mathrm{K}_{\det}(\mathbf{x}_1, \mathbf{x}_2) + \mathrm{K}_{\det}(\mathbf{y}_1, \mathbf{y}_2) - \frac{1}{2}\mathrm{K}_{\det}(\mathbf{x}_1, \mathbf{y}_1) \\ & -\frac{1}{2}\mathrm{K}_{\det}(\mathbf{x}_1, \mathbf{y}_2) - \frac{1}{2}\mathrm{K}_{\det}(\mathbf{x}_2, \mathbf{y}_1) - \frac{1}{2}\mathrm{K}_{\det}(\mathbf{x}_2, \mathbf{y}_2), \end{aligned} \tag{9}$$

where $\mathrm{K}_{\det}(\mathbf{z}, \mathbf{z}') = \det(\gamma(\mathbf{z}-\mathbf{z}')^T(\mathbf{z}-\mathbf{z}') + \mathbf{I}_q)^{-1/2}$. Then, $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$ can be rewritten as

$$\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y}) = \frac{1}{n(n-1)}\frac{1}{m(m-1)}\sum_{\substack{i_1,i_2=1\\ i_1\neq i_2}}^{n}\sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{m} h\left(\mathbf{X}_{i_1}, \mathbf{X}_{i_2}; \mathbf{Y}_{j_1}, \mathbf{Y}_{j_2}\right).$$

We first study the rate of convergence of $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$ regardless of whether $P_{\mathbf{X}} = P_{\mathbf{Y}}$. By the bounded differences inequality (Wainwright, 2019, Corollary 2.21), we obtain the following result.

Theorem 3 *For any $\alpha \in (0, 1)$, we have that*

$$|\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y}) - \mathrm{DPCD}(\mathbf{X}, \mathbf{Y})| \leq 8\sqrt{2}\sqrt{\log(2/\alpha)}(n+m)^{1/2}(n^{-1} + m^{-1}),$$

with probability at least $1-\alpha$.

Theorem 3 suggests that $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$ converges in probability to $\mathrm{DPCD}(\mathbf{X}, \mathbf{Y})$ at the rate of $(n+m)^{1/2}(n^{-1}+m^{-1})$, without requiring any moment condition on the data. If we further use the conventional assumption (Anderson, 1962; van der Vaart and Wellner, 1996) given by

$$\lim_{n,m\to\infty}\frac{n}{n+m} = \pi_X \in (0,1) \quad \text{and} \quad \lim_{n,m\to\infty}\frac{m}{n+m} = \pi_Y \in (0,1), \tag{10}$$

then we can obtain

$$|\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y}) - \mathrm{DPCD}(\mathbf{X}, \mathbf{Y})| \leq 8\sqrt{2}\sqrt{\log(2/\alpha)}\pi_X^{-1}\pi_Y^{-1}(n+m)^{-1/2}.$$

This indicates that under condition (10), $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$ converges at the rate of $(n+m)^{-1/2}$. Note that the rate $(n+m)^{-1/2}$ is obtained without requiring the null hypothesis H_0 to be true. The following theorem shows that under H_0, we can get a faster convergence rate.

Theorem 4 *Assume that condition (10) holds. Then, under the null hypothesis H_0, we have that as $n, m \to \infty$,*

$$(m+n)\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y}) \xrightarrow{d} \frac{1}{\pi_X\pi_Y}\sum_{k=1}^{\infty}\lambda_k\left(\xi_k^2 - 1\right), \tag{11}$$

where $\xi_k \overset{i.i.d}{\sim} N(0,1)$ and λ_k are the eigenvalues of the integral equation

$$E\{h_{2,0}(\boldsymbol{x}_1, \boldsymbol{X}_2)\phi_k(\boldsymbol{X}_2)\} = \lambda_k\phi_k(\boldsymbol{x}_1),$$

where $h_{2,0}(\boldsymbol{x}_1, \boldsymbol{x}_2) = E\{h(\boldsymbol{x}_1, \boldsymbol{x}_2;\, \boldsymbol{Y}_1,\, \boldsymbol{Y}_2)\}$, and ϕ_k are the resulting eigenfunctions.

We further establish the asymptotic distribution of $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$ under the alternative hypothesis $H_1 : P_{\mathbf{X}} \neq P_{\mathbf{Y}}$ as follows.

Theorem 5 *Assume that condition* (10) *holds. Then, under the alternative hypothesis* H_1, *we have that as* $n, m \to \infty$,

$$(n+m)^{1/2}\{\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y}) - \mathrm{DPCD}(\mathbf{X},\mathbf{Y})\} \xrightarrow{d} N\left(0, 4\pi_X^{-1}\sigma_{1,0}^2 + 4\pi_Y^{-1}\sigma_{0,1}^2\right),$$

where

$$\sigma_{1,0}^2 = \mathrm{Var}\{h_{1,0}(\boldsymbol{X}_1)\} \text{ and } \sigma_{0,1}^2 = \mathrm{Var}\{h_{0,1}(\boldsymbol{Y}_1)\},$$

with $h_{1,0}(\boldsymbol{x}_1) = E\{h(\boldsymbol{x}_1, \boldsymbol{X}_2; \boldsymbol{Y}_1, \boldsymbol{Y}_2)\}$ *and* $h_{0,1}(\boldsymbol{y}_1) = E\{h(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{y}_1, \boldsymbol{Y}_2)\}$.

Note that Theorem 4 suggests that under H_0, $(n+m)\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})$ converges in distribution to $\sum_{k=1}^{\infty} \lambda_k(\xi_k^2 - 1)$, which is stochastically bound. On other hand, under $H_1 : P_{\mathbf{X}} \neq P_{\mathbf{Y}}$, Theorem 5 implies that $(n+m)\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})$ diverges to infinity. Thus, we can use the statistic $(n+m)\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})$ to test problem (1).

To establish the asymptotic properties of $\widehat{\mathrm{SPCD}}(\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}})$, we define the following symmetric kernel function

$$\begin{aligned}\widetilde{h}(\widetilde{\mathbf{x}}_1,\widetilde{\mathbf{x}}_2;\widetilde{\mathbf{y}}_1,\widetilde{\mathbf{y}}_2) &= \widetilde{\mathrm{K}}_{\det}(\widetilde{\mathbf{x}}_1,\widetilde{\mathbf{x}}_2) + \widetilde{\mathrm{K}}_{\det}(\widetilde{\mathbf{y}}_1,\widetilde{\mathbf{y}}_2) - \frac{1}{2}\widetilde{\mathrm{K}}_{\det}(\widetilde{\mathbf{x}}_1,\widetilde{\mathbf{y}}_1) \\ &\quad -\frac{1}{2}\widetilde{\mathrm{K}}_{\det}(\widetilde{\mathbf{x}}_1,\widetilde{\mathbf{y}}_2) - \frac{1}{2}\widetilde{\mathrm{K}}_{\det}(\widetilde{\mathbf{x}}_2,\widetilde{\mathbf{y}}_1) - \frac{1}{2}\widetilde{\mathrm{K}}_{\det}(\widetilde{\mathbf{x}_2},\widetilde{\mathbf{y}}_2),\end{aligned}$$

where $\widetilde{\mathrm{K}}_{\det}(\widetilde{\mathbf{z}},\widetilde{\mathbf{z}}') = \det(\mathbf{I}_p - 2i\gamma(\widetilde{\mathbf{z}} - \widetilde{\mathbf{z}}'))^{-1/2}$. Then, $\widehat{\mathrm{SPCD}}(\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}})$ can be written in the form of a two-sample U-statistic

$$\widehat{\mathrm{SPCD}}(\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}}) = \frac{1}{n(n-1)}\frac{1}{m(m-1)}\sum_{\substack{i_1,i_2=1\\ i_1\neq i_2}}^{n}\sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{m}\widetilde{h}(\widetilde{\mathbf{X}}_{i_1},\widetilde{\mathbf{X}}_{i_2};\widetilde{\mathbf{Y}}_{j_1},\widetilde{\mathbf{Y}}_{j_2}). \tag{12}$$

Using (12), we can obtain asymptotic results for $\widehat{\mathrm{SPCD}}(\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}})$, which are basically similar to those in Theorems 3–5 and so are omitted. In fact, when we establish the properties of $\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})$, a key is the boundedness of $|h(\mathbf{x}_1,\mathbf{x}_2;\mathbf{y}_1,\mathbf{y}_2)| \leq 4$ [see (26)], and the same results can be obtained for $\widetilde{h}(\widetilde{\mathbf{x}}_1,\widetilde{\mathbf{x}}_2;\widetilde{\mathbf{y}}_1,\widetilde{\mathbf{y}}_2)$ by the following Proposition 3.

3.2 Permutation test

Note that the limiting null distribution of $(n+m)\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})$ in Theorem 4 is intractable because the eigenvalues λ_k are unknown and not easy to estimate. To implement the DPCD test, we can approximate the asymptotic null distribution through the random-permutation test procedure as follows.

Step 1. Let $\{\mathbf{Z}_1,\cdots,\mathbf{Z}_{n+m}\} = \{\mathbf{X}_1,\cdots,\mathbf{X}_n,\mathbf{Y}_1,\cdots,\mathbf{Y}_m\}$ be the pooled samples and $\mathbf{Z}_R = \{\mathbf{Z}_{R_1},\cdots,\mathbf{Z}_{R_{n+m}}\}$, where $R = (R_1,\cdots,R_{n+m})$ is a permutation of $\{1,2,\ldots,n+m\}$.

Step 2. Using each permutation sample $\mathbf{Z}_R = \{\mathbf{Z}_{R_1},\cdots,\mathbf{Z}_{R_{n+m}}\}$, calculate the resulting test statistic $\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) = \widehat{\mathrm{DPCD}}(\mathbf{X}_R,\mathbf{Y}_R)$, where $\mathbf{X}_R = \{\mathbf{Z}_{R_1},\cdots,\mathbf{Z}_{R_n}\}$ and $\mathbf{Y}_R = \{\mathbf{Z}_{R_{n+1}},\cdots,\mathbf{Z}_{R_{n+m}}\}$.

Step 3. The exact permutation distribution of $(n+m)\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})$ can be given by

$$P_R\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \le t \mid \mathcal{D}_{n,m}\} = \frac{1}{(n+m)!}\sum_{R\in\mathcal{G}_{n+m}} I\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \le t\},$$

where $\mathcal{G}_{n+m}$ is the set of all permutations among $\{1,2,\ldots,n+m\}$, $I\{\cdot\}$ is the indicator function, and $\mathcal{D}_{n,m} = \{\mathbf{X}_i, \mathbf{Y}_j, i=1,\ldots,n, j=1,\ldots,m\}$.

Step 4. The critical value at the significance level α is calculated by

$$q_{\alpha,m,n} = \inf\Big\{t : P_R\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) > t \mid \mathcal{D}_{n,m}\} \le \alpha\Big\}. \tag{13}$$

Theorem 6 states that the exact permutation distribution provides a reasonable approximation to the null distribution.

Theorem 6 *Under condition* (10)*, it holds that*

$$\sup_{t\ge 0}\Big|P_R\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \le t \mid \mathcal{D}_{n,m}\} - P\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y}) \le t|H_0\}\Big| \xrightarrow{P} 0,$$

as $n,m\to\infty$.

Theorem 6 suggests that the exact permutation distribution is asymptotically equivalent to the limiting null distribution of $\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})$. However, all possible permutations must be exhausted, which may impose a high computational burden. To remedy this problem, a random sampling scheme can be used. Specifically, let $R^{(1)},\ldots,R^{(B)}$ be independent and uniformly distributed on $\mathcal{G}_{n+m}$. The empirical version of the exact permutation distribution is computed by

$$\frac{1}{B}\sum_{b=1}^{B} I\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_{R^{(b)}}) \le t\}, \tag{14}$$

and the following proposition shows that (14) is asymptotically valid.

Proposition 1 *Given the data* $\mathcal{D}_{n,m}$*, it holds that*

$$\sup_{t\ge 0}\Big|\frac{1}{B}\sum_{b=1}^{B} I\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_{R^{(b)}}) \le t\} - P\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \le t \mid \mathcal{D}_{n,m}\}\Big| \xrightarrow{P} 0$$

as $B\to\infty$.

In practice, we can perform the above permutation test by computing the following empirical permutation p-value, given by

$$\widehat{p}^{(B)}_{\mathrm{DPCD}} = \frac{1}{B+1}\Big\{1 + \sum_{b=1}^{B} I\{\widehat{\mathrm{DPCD}}(\mathbf{Z}_{R^{(b)}}) \ge \widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})\}\Big\},$$

and its asymptotic consistency is guaranteed by Theorem 6 and Proposition 1.

3.3 Robustness

We investigate the robustness of the DPCD test, given by

$$\phi_{\text{DPCD}} = I\{(n+m)\widehat{\text{DPCD}}(\mathbf{X},\mathbf{Y}) \geq q_{\alpha,m,n}\},$$

where $q_{\alpha,m,n}$ is the critical value defined in (13). Using the method proposed by Kim et al. (2020), we evaluate the robustness of ϕ_{DPCD} by its power performance in the ϵ-contamination model

$$X \sim P_{\mathbf{X}} = (1-\epsilon)Q_{\mathbf{X}} + \epsilon H_{n,m} \quad \text{and} \quad Y \sim P_{\mathbf{Y}} = (1-\epsilon)Q_{\mathbf{Y}} + \epsilon H_{n,m}, \tag{15}$$

where $Q_{\mathbf{X}}$, $Q_{\mathbf{Y}}$, and $H_{n,m}$ are matrix-variate distributions on $\mathbb{R}^{p\times q}$, $H_{n,m}$ may depend on (n,m), and $\epsilon \in (0,1)$.

In the ϵ-contamination model, for any fixed $(Q_{\mathbf{X}}, Q_{\mathbf{Y}})$ with $Q_{\mathbf{X}} \neq Q_{\mathbf{Y}}$, we can see that the difference between $P_{\mathbf{X}}$ and $P_{\mathbf{Y}}$ increases as ϵ decreases. Thus, a test is robust if it does not lose power against any $H_{n,m}$ when ϵ is small and $Q_{\mathbf{X}} \neq Q_{\mathbf{Y}}$.

Theorem 7 *Suppose that $\{\boldsymbol{X}_i, i=1,\ldots,n\}$ and $\{\boldsymbol{Y}_i, i=1,\ldots,m\}$ are generated independently from the contaminated model in* (15) *with $\epsilon \leq c(n^{-1/2}+m^{-1/2})$, where c is a small positive constant not depending on n,m. Under condition* (10)*, we have that*

$$\lim_{m,n\to\infty} \inf_{H_{n,m}} P\{\phi_{\text{DPCD}} = 1 | H_1\} = 1$$

for any fixed $(Q_{\mathbf{X}}, Q_{\mathbf{Y}})$ with $Q_{\mathbf{X}} \neq Q_{\mathbf{Y}}$.

Theorem 7 suggests that the DPCD test is asymptotically powerful uniformly over all possible $H_{n,m}$, which implies that the DPCD test is insensitive to outliers.

4. Related Works

In this section, we present some connection between our proposed tests and the two commonly used vector-valued tests, the maximum mean discrepancy (Gretton et al., 2012, 'MMD' for short) and the energy distance (Székely and Rizzo, 2013, 'Energy' for short). Recall the MMD and Energy between $\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{Y} \in \mathbb{R}^p$, given by

$$\text{MMD}(\mathbf{X},\mathbf{Y}) = E\{K(\mathbf{X}_1,\mathbf{X}_2)\} - 2E\{K(\mathbf{X}_1,\mathbf{Y}_2)\} + E\{K(\mathbf{Y}_1,\mathbf{Y}_2)\}, \tag{16}$$

$$\text{Energy}(\mathbf{X},\mathbf{Y}) = -E\{\|\mathbf{X}_1-\mathbf{X}_2\|\} + 2E\{\|\mathbf{X}_1-\mathbf{Y}_2\|\} - E\{\|\mathbf{Y}_1-\mathbf{Y}_2\|\}, \tag{17}$$

where $K(\cdot,\cdot) : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ is a specified kernel function, such as the Gaussian kernel and Laplacian kernel.

From Theorem 1(i) and Theorem 2(i), it can be seen that the DPCD and SPCD have similar form to the MMD and Energy, except for using different kernel functions. The functions $\det(\gamma(\mathbf{z}-\mathbf{z}')^T(\mathbf{z}-\mathbf{z}')+\mathbf{I}_q)^{-1/2}$ and $\det(\mathbf{I}_p - 2i\gamma(\widetilde{\mathbf{z}}-\widetilde{\mathbf{z}}'))^{-1/2}$ involved in the DPCD and SPCD have not been proposed in existing literature. In the following sections, we investigate some properties of these functions.

4.1 Inverse determinant quadric kernel

Proposition 2 *Define* $\mathrm{K}_{\det}(\mathbf{z},\mathbf{z}') = \det(\gamma(\mathbf{z}-\mathbf{z}')^T(\mathbf{z}-\mathbf{z}')+\boldsymbol{I}_q)^{-1/2}, \gamma > 0,$ *for* $\mathbf{z},\mathbf{z}' \in \mathbb{R}^{p\times q}$. *Then, we have*

(i) $\mathrm{K}_{\det}(\mathbf{z},\mathbf{z}')$ *is a positive-definite function, i.e.* $\forall n \geq 2, \mathbf{z}_1,\ldots,\mathbf{z}_n \in \mathbb{R}^{p\times q}$ *and* $c_1,\ldots,c_n \in \mathbb{R}$, *we have*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \mathrm{K}_{\det}(\mathbf{z}_i,\mathbf{z}_j) \geq 0;$$

(ii) $\mathrm{K}_{\det}(\mathbf{z},\mathbf{z}') = \prod_{k=1}^{r}(\gamma\lambda_k+1)^{-1/2}$, *where* λ_k*s are the nonzero eigenvalues of* $(\mathbf{z}-\mathbf{z}')^T(\mathbf{z}-\mathbf{z}')$ *and* $r = \mathrm{rank}((\mathbf{z}-\mathbf{z}')^T(\mathbf{z}-\mathbf{z}'))$.

When $q = 1$, we have $\mathrm{K}_{\det}(\mathbf{z},\mathbf{z}') = (\gamma\|\mathbf{z}-\mathbf{z}'\|^2+1)^{-1/2}$, $\mathbf{z},\mathbf{z}' \in \mathbb{R}^p$, which is proportional to the classical inverse multiquadric kernel $k(\mathbf{z},\mathbf{z}') = (\|\mathbf{z}-\mathbf{z}'\|^2+c^2)^{-1/2}$; see Micchelli (1986). Thus, $\mathrm{K}_{\det}(\mathbf{z},\mathbf{z}')$ can be viewed as an extension of the inverse multiquadric kernel from $\mathbb{R}^p$ to $\mathbb{R}^{p\times q}$.

Proposition 2(i) suggests that $\mathrm{K}_{\det}(\mathbf{z},\mathbf{z}')$ is a Mercer kernel. By Mercer's theorem, there exist a real-valued Hilbert space $\mathcal{H}_R$ and a map $\boldsymbol{\phi}: \mathbb{R}^{p\times q} \to \mathcal{H}_R$ such that

$$\mathrm{K}_{\det}(\mathbf{z},\mathbf{z}') = \langle \boldsymbol{\phi}(\mathbf{z}), \boldsymbol{\phi}(\mathbf{z}')\rangle_{\mathcal{H}_R},$$

where $\langle\cdot,\cdot\rangle_{\mathcal{H}_R}$ is an inner product. Because $\mathrm{K}_{\det}(\mathbf{z},\mathbf{z}')$ is an extension of the classical inverse multiquadric kernel, we refer to it as the inverse determinant quadric kernel. Note that the Energy in (17) requires the moment conditions $E\{\|\mathbf{X}\|\} < \infty$ and $E\{\|\mathbf{Y}\|\} < \infty$ to ensure that $\mathrm{Energy}(\mathbf{X},\mathbf{Y})$ is well defined. By contrast, Proposition 2(ii) suggests that $0 \leq \mathrm{K}_{\det}(\mathbf{z},\mathbf{z}') \leq 1$, and thus $\mathrm{DPCD}(\mathbf{X},\mathbf{Y})$ requires no assumption about the underlying distributions.

Although the MMD and Energy were originally designed for vector-valued data, they can be applied directly to matrix-valued data by vectorizing a matrix into a vector. For any $\mathbf{z},\mathbf{z}' \in \mathbb{R}^{p\times q}$, the vectorizing method yields

$$\|\mathrm{vec}(\mathbf{z}) - \mathrm{vec}(\mathbf{z}')\| = \{\mathrm{tr}((\mathbf{z}-\mathbf{z}')^T(\mathbf{z}-\mathbf{z}'))\}^{1/2} = \Big\{\sum_{k=1}^{r}\lambda_k\Big\}^{1/2}, \tag{18}$$

where λ_k are the nonzero eigenvalues of $(\mathbf{z}-\mathbf{z}')^T(\mathbf{z}-\mathbf{z}')$ and $r = \mathrm{rank}((\mathbf{z}-\mathbf{z}')^T(\mathbf{z}-\mathbf{z}'))$. By (18), the Gaussian and Laplacian kernels for the MMD are equal to

$$\mathrm{K}_{\mathrm{gaus}}(\mathrm{vec}(\mathbf{z})-\mathrm{vec}(\mathbf{z}')) = \exp\Big\{-\gamma^{-2}\sum_{k=1}^{r}\lambda_k\Big\},\ \mathrm{K}_{\mathrm{lap}}(\mathrm{vec}(\mathbf{z})-\mathrm{vec}(\mathbf{z}')) = \exp\Big\{-\gamma^{-1}\{\sum_{k=1}^{r}\lambda_k\}^{1/2}\Big\}. \tag{19}$$

For the Energy, the L_2 distance is equal to (18). On the other hand, for the DPCD, Proposition 2(ii) suggests that

$$\mathrm{K}_{\det}(\mathbf{z},\mathbf{z}') = \prod_{k=1}^{r}(\gamma\lambda_k+1)^{-1/2} = \exp\Big\{-\frac{1}{2}\sum_{k=1}^{r}\log\{\gamma\lambda_k+1\}\Big\}. \tag{20}$$

From (18)–(20), we obtain the following interesting findings.

(i) All three methods extract the matrix structure by using their eigenvalues. Moreover, all the methods can identify the difference between two matrix-variate distributions.

(ii) However, the DPCD captures the matrix structure in a very different way. Specifically, the MMD and Energy extract the matrix structure by a linear combination $\sum_{k=1}^{r} \lambda_k$, whereas the DPCD works by a nonlinear combination $\sum_{k=1}^{r} \log\{\gamma\lambda_k + 1\}$.

(iii) $\mathrm{K}_{\text{det}}(\mathbf{z}, \mathbf{z}')$ further works by the exponential transformation, i.e. $\exp\{-\frac{1}{2}\sum_{k=1}^{r} \log\{\gamma\lambda_k + 1\}\}$, which is similar to that for the Gaussian and Laplacian kernels.

From the above observations, it can be seen that the DPCD uses a two-layer nonlinear transformation of eigenvalues to extract the matrix information. To the best of our knowledge, there seem to be few similar methods in the literature of statistics and machine learning. By our extensive simulation studies (see Section 5), we find that the two-layer nonlinear transformation works well. Moreover, as demonstrated in Section 5, the DPCD is comparable to the MMD and Energy when only location parameters are different, and the DPCD significantly outperforms the others when the difference between two distributions is determined by high-order moments.

4.2 Complex inverse determinant quadric kernel

Let $\mathbb{C}$ be the complex domain. By Theorem 2(i), $\mathrm{SPCD}(\mathbf{X}, \mathbf{Y})$ involves the following complex-valued function.

Proposition 3 *Define* $\widetilde{\mathrm{K}}_{\text{det}}(\widetilde{\mathbf{z}}, \widetilde{\mathbf{z}}') = \det(\boldsymbol{I}_p - 2i\gamma(\widetilde{\mathbf{z}} - \widetilde{\mathbf{z}}'))^{-1/2}, \gamma > 0$, *for* $\widetilde{\mathbf{z}}, \widetilde{\mathbf{z}}' \in \mathbb{S}_p$. *Then, we have*

(i) $\widetilde{\mathrm{K}}_{\text{det}}(\widetilde{\mathbf{z}}, \widetilde{\mathbf{z}}')$ *is a positive-definite function, i.e.* $\forall, \widetilde{\mathbf{z}}_1, \ldots, \widetilde{\mathbf{z}}_n \in \mathbb{S}_p$ *and* $c_1, \ldots, c_n \in \mathbb{C}$, *satisfying*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i \bar{c}_j \widetilde{\mathrm{K}}_{\text{det}}(\widetilde{\mathbf{z}}_i, \widetilde{\mathbf{z}}_j) \geq 0;$$

(ii) $\widetilde{\mathrm{K}}_{\text{det}}(\widetilde{\mathbf{z}}, \widetilde{\mathbf{z}}') = \prod_{k=1}^{r}(1 - 2i\gamma\lambda_k)^{-1/2}$, *and thus* $\|\widetilde{\mathrm{K}}_{\text{det}}(\widetilde{\mathbf{z}}, \widetilde{\mathbf{z}}')\| = \prod_{k=1}^{r}(1 + 4\gamma^2\lambda_k^2)^{-1/4}$, *where* λ_k *are the nonzero eigenvalues of* $\widetilde{\mathbf{z}} - \widetilde{\mathbf{z}}'$ *and* $r = \mathrm{rank}(\widetilde{\mathbf{z}} - \widetilde{\mathbf{z}}')$.

According to the theory of reproducing kernel Hilbert spaces (RKHSs) (Steinwart and Christmann, 2008), Proposition 3(i) suggests that there exist a complex-valued Hilbert space $\mathcal{H}_S$ and a feature map $\boldsymbol{\phi} : \mathbb{S}_p \to \mathcal{H}_S$ such that

$$\widetilde{\mathrm{K}}_{\text{det}}(\mathbf{z}, \mathbf{z}') = \langle \boldsymbol{\phi}(\mathbf{z}'), \boldsymbol{\phi}(\mathbf{z}) \rangle_{\mathcal{H}_S},$$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}_S}$ is an inner product. We refer to $\widetilde{\mathrm{K}}_{\text{det}}(\widetilde{\mathbf{z}}, \widetilde{\mathbf{z}}')$ as the complex inverse determinant quadric kernel. By the classical kernel trick, we may consider some statistical learning problems on $\mathbb{S}_p$ based on $\widetilde{\mathrm{K}}_{\text{det}}(\mathbf{z}, \mathbf{z}')$ and the RKHS $\mathcal{H}_S$.

4.3 Log inverse determinant quadric kernel

By the definitions of $\mathrm{DPCD}_{\mathrm{Log}}$ and $\mathrm{SPCD}_{\mathrm{Log}}$ in Section 2.3, the two metrics involve the following functions.

Definition 3 *Define the functions* $\mathrm{K}^{+}_{\det} : \mathbb{S}^{+}_{p} \times \mathbb{S}^{+}_{p} \to \mathbb{R}$ *and* $\widetilde{\mathrm{K}}^{+}_{\det} : \mathbb{S}^{+}_{p} \times \mathbb{S}^{+}_{p} \to \mathbb{C}$ *as*

$$\mathrm{K}^{+}_{\det}(\underline{\mathbf{z}}, \underline{\mathbf{z}}') = \det(\gamma[\mathrm{Log}(\underline{z}) - \mathrm{Log}(\underline{z}')]^{T}[\mathrm{Log}(\underline{z}) - \mathrm{Log}(\underline{z}')] + \boldsymbol{I}_{p})^{-1/2},$$

$$\widetilde{\mathrm{K}}^{+}_{\det}(\underline{\mathbf{z}}, \underline{\mathbf{z}}') = \det(\boldsymbol{I}_{p} - 2i\gamma(\mathrm{Log}(\underline{z}) - \mathrm{Log}(\underline{z}')))^{-1/2}.$$

By Propositions 2 and 3, we immediately obtain the following results.

Proposition 4 **(i)** $\mathrm{K}^{+}_{\det}(\underline{\mathbf{z}}, \underline{\mathbf{z}}')$ *and* $\widetilde{\mathrm{K}}^{+}_{\det}(\underline{\mathbf{z}}, \underline{\mathbf{z}}')$ *are positive-definite functions.*

(ii) *Let* $\{\lambda_k\}_{k=1}^{r}$ *be the nonzero eigenvalues of* $\mathrm{Log}(\underline{z}) - \mathrm{Log}(\underline{z}')$ *and* $r = \mathrm{rank}(\mathrm{Log}(\underline{z}) - \mathrm{Log}(\underline{z}'))$. *Then, we have that* $\mathrm{K}^{+}_{\det}(\underline{\mathbf{z}}, \underline{\mathbf{z}}') = \prod_{k=1}^{r}(1 + \gamma\lambda_k^2)^{-1/2}$ *and* $\widetilde{\mathrm{K}}^{+}_{\det}(\underline{\mathbf{z}}, \underline{\mathbf{z}}') = \prod_{k=1}^{r}(1 - 2i\gamma\lambda_k)^{-1/2}$.

By Proposition 4 and the theory of RKHSs, $\mathrm{K}^{+}_{\det}(\underline{\mathbf{z}}, \underline{\mathbf{z}}')$ and $\widetilde{\mathrm{K}}^{+}_{\det}(\underline{\mathbf{z}}, \underline{\mathbf{z}}')$ can be used as kernel functions. Using these kernels, we can construct a real-valued RKHS $\mathcal{H}^{+}_{R}$ and a complex-valued RKHS $\mathcal{H}^{+}_{S}$. In summary, we have proposed the four new kernel functions $\mathrm{K}_{\det}(\cdot,\cdot)$, $\widetilde{\mathrm{K}}_{\det}(\cdot,\cdot)$, $\mathrm{K}^{+}_{\det}(\cdot,\cdot)$, and $\widetilde{\mathrm{K}}^{+}_{\det}(\cdot,\cdot)$ on $\mathbb{R}^{p\times q}$, $\mathbb{S}_p$, and $\mathbb{S}^{+}_{p}$.

5. Monte Carlo simulations

In the section, we conduct Monte Carlo simulations to assess the finite-sample performances of the proposed DPCD, SPCD, and $\mathrm{SPCD}_{\mathrm{Log}}$. We also compare the performances of our methods with those of the MMD and Energy via vectorization. For the MMD, we consider two kernel functions: the Gaussian kernel ($\mathrm{MMD}_{\mathrm{gaus}}$) and the Laplace kernel ($\mathrm{MMD}_{\mathrm{lap}}$). Throughout our experiments, the p-value of each test was obtained by $B = 999$ permutations. We repeat each experiment 1000 times, and we report the empirical power or type-I error rate of each test.

For the MMD and our methods, the choice of the parameter γ is important. In all the experiments, we use the median-distance heuristic (Gretton et al., 2012) $\gamma_{\mathrm{med}} = (\mathrm{median}\,\{\|\mathrm{vec}(\mathbf{Z}_i - \mathbf{Z}_j)\|^2 : i \neq j\}/2)^{1/2}$, to select γ for $\mathrm{MMD}_{\mathrm{gaus}}$ and $\mathrm{MMD}_{\mathrm{lap}}$. Accordingly, we recommend using $\gamma = \gamma_{\mathrm{med}}^{-1}$ for the DPCD and $\mathrm{DPCD}_{\mathrm{Log}}$, and $\gamma = \gamma_{\mathrm{med}}^{-1/2}$ for the SPCD and $\mathrm{SPCD}_{\mathrm{Log}}$. In the following experiments, we find that the recommended parameters performed well across all the scenarios.

5.1 DPCD for matrix-valued data

We illustrate the finite-sample performance of the DPCD for arbitrary $p \times q$-dimensional matrix-valued data through the following four examples.

Example 1 *In this example, we consider the special case of* $q = 1$, *i.e.* $\mathbf{X}$ *and* $\mathbf{Y}$ *are random vectors generated by the following two cases:*

Table 2: Empirical type-I error rate with $p = 25$ for Example 1.

Settings	$n_1 = n_2$	Method	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Example 1 (Case 1)	20	DPCD	0.007	0.050	0.102
		Energy	0.007	0.048	0.104
		MMD_{gaus}	0.005	0.050	0.099
		MMD_{lap}	0.006	0.051	0.103
Example 1 (Case 2, $df = 3$)	20	DPCD	0.008	0.051	0.106
		Energy	0.014	0.059	0.112
		MMD_{gaus}	0.005	0.056	0.104
		MMD_{lap}	0.009	0.053	0.107
Example 1 (Case 2, $df = 1$)	20	DPCD	0.015	0.057	0.100
		Energy	0.015	0.060	0.107
		MMD_{gaus}	0.014	0.055	0.090
		MMD_{lap}	0.017	0.052	0.095

Case 1: $\mathbf{X} \in \mathbb{R}^p \sim \boldsymbol{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ *and* $\mathbf{Y} \in \mathbb{R}^p \sim \boldsymbol{N}_p(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma})$*;*

Case 2: $\mathbf{X} \in \mathbb{R}^p \sim \boldsymbol{t}_p(\boldsymbol{0}, \boldsymbol{\Sigma}, df)$ *and* $\mathbf{Y} \in \mathbb{R}^p \sim \boldsymbol{t}_p(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}, df)$ *with* $df = 3$ *or 1,*

where $\boldsymbol{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *is a multivariate normal distribution, and* $\boldsymbol{t}_p(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}, df)$ *is a multivariate t-distribution with the degree of freedom df. Here, we set* $\boldsymbol{\Sigma} = (0.5^{|j-k|})$ *and* $\boldsymbol{\mu}_Y = (\mu, \cdots, \mu)^T$*.*

Table 2 reports the empirical type-I error rate of each test at significance levels $\alpha = 0.01$, 0.05, and 0.1 with $p = 25$. From Table 2, it can be seen that each test achieves approximately the three nominal significance levels under the null hypotheses $\mathbf{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ in Case 1 and $\mathbf{t}_p(\mathbf{0}, \boldsymbol{\Sigma}, df)$ in Case 2.

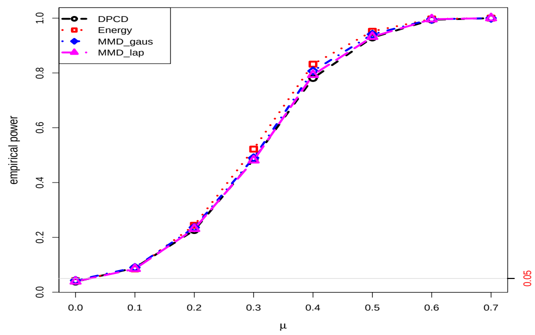
The empirical power comparison at $\alpha = 0.05$ is summarized in Figure 1. Figure 1(a), (c), and (e) show plots of the power against μ at $p = 25$, and Figure 1(b), (d), and (f) show plots of the power against p at $\mu = 0.5$. The results from Figure 1(a)–(d) suggest that the DPCD, Energy, MMD_{gaus}, and MMD_{lap} tests perform similarly. This is consistent with our findings presented in Section 4, because $\mathrm{K}_{\det}(\mathbf{z}, \mathbf{z}')$ is the classical inverse multiquadric kernel in the setting of $q = 1$.

As is well known, the Laplace kernel is robust to outliers, and thus MMD_{lap} is similarly robust. However, somewhat surprisingly, Figure 1(e)–(f) suggest that the DPCD performs more powerfully and robustly than the MMD_{lap} for the multivariate t-distributions with $df = 1$ in Case 2.

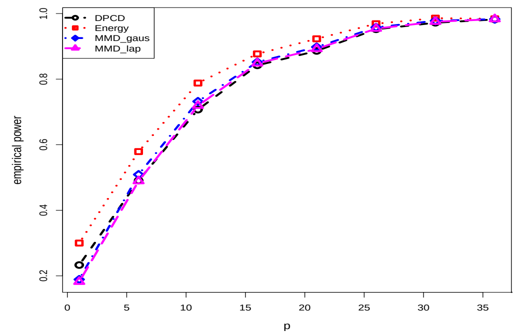
From Figure 1(e)–(f), we can further see that the Energy has lower power than the others; this may be because the moment restrictions $E\{\|\mathbf{X}\|\} < \infty$ and $E\{\|\mathbf{Y}\|\} < \infty$ are violated. In summary, Figure 1(a)–(f) suggest that the DPCD is comparable to the MMD and Energy when $\mathbf{X}$ and $\mathbf{Y}$ are random vectors.

Example 2 *This example examines the general case of* $p, q \neq 1$*. Consider the following two cases:*

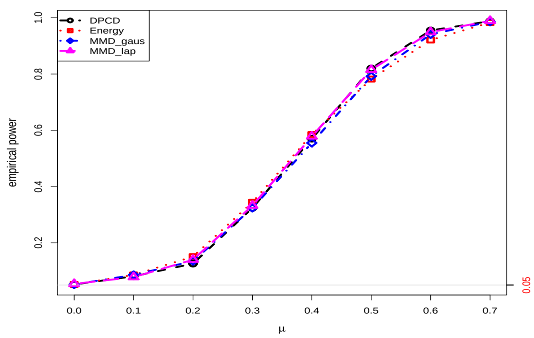
Case 1: $\mathbf{X} \in \mathbb{R}^{p \times q} \sim \boldsymbol{MN}_{p \times q}(\mathbf{0}, \mathbf{U}_1, \mathbf{V}_1)$ *and* $\mathbf{Y} \in \mathbb{R}^{p \times q} \sim \boldsymbol{MN}_{p \times q}(\mathbf{M}_2, \mathbf{U}_2, \mathbf{V}_2)$*;*



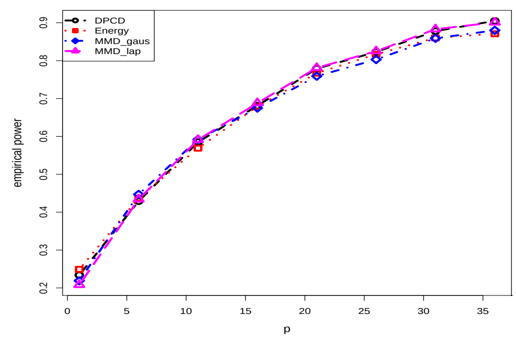
(a) Case 1: Multivariate normal distribution



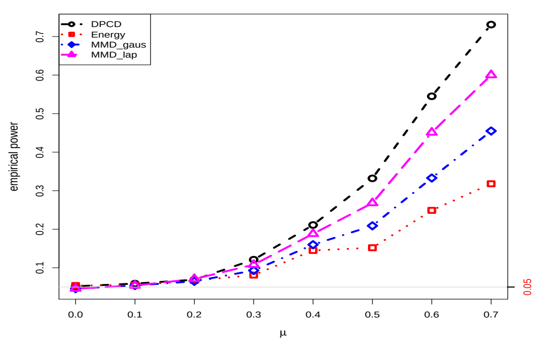
(b) Case 1: Multivariate normal distribution



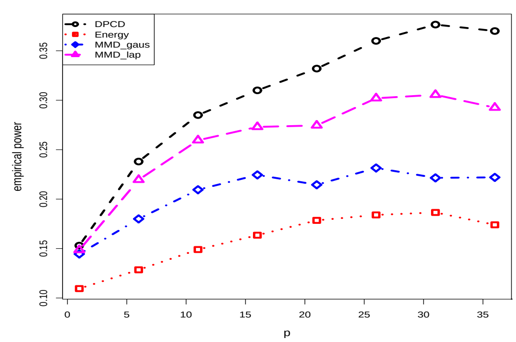
(c) Case 2: t-distribution with $df = 3$



(d) Case 2: t-distribution with $df = 3$



(e) Case 2: t-distribution with $df = 1$



(f) Case 2: t-distribution with $df = 1$

Figure 1: Empirical power at $\alpha = 0.05$ for Example 1 with $n_1 = n_2 = 20$. **(a)**, **(c)**, and **(e)**: $\boldsymbol{\mu}_Y = (\mu, \cdots, \mu)^T$ varies with μ at $p = 25$; **(b)**, **(d)**, and **(f)**: p varies at $\mu = 0.5$.

Table 3: Empirical type-I error rate with $(p, q) = (25, 15)$ for Example 2.

Settings	$n_1 = n_2$	Method	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Example 2 (Case 1)	20	DPCD	0.010	0.063	0.105
		Energy	0.008	0.056	0.110
		MMD_{gaus}	0.008	0.058	0.112
		MMD_{lap}	0.010	0.058	0.110
Example 2 (Case 2, $df = 3$)	20	DPCD	0.006	0.051	0.095
		Energy	0.009	0.048	0.114
		MMD_{gaus}	0.014	0.054	0.108
		MMD_{lap}	0.010	0.055	0.106
Example 2 (Case 2, $df = 1$)	20	DPCD	0.007	0.047	0.092
		Energy	0.008	0.050	0.114
		MMD_{gaus}	0.010	0.050	0.101
		MMD_{lap}	0.010	0.051	0.104

Case 2: $\mathbf{X} \in \mathbb{R}^{p\times q} \sim \boldsymbol{T}_{p\times q}(df, \mathbf{0}, \mathbf{U}_1, \mathbf{V}_1)$ *and* $\mathbf{Y} \in \mathbb{R}^{p\times q} \sim \boldsymbol{T}_{p\times q}(df, \mathbf{M}_2, \mathbf{U}_2, \mathbf{V}_2)$ *with* $df = 3$ *or 1,*

where $\boldsymbol{MN}_{p\times q}(\mathbf{M}, \mathbf{U}, \mathbf{V})$ *is the* $p \times q$*-matrix normal distribution, and* $\boldsymbol{T}_{p\times q}(df, \mathbf{M}, \mathbf{U}, \mathbf{V})$ *is the* $p \times q$*-matrix* t *distribution with the degree of freedom* df*, where* $\mathbf{M} \in \mathbb{R}^{p\times q}$*,* $\mathbf{U}(> 0) \in \mathbb{R}^{p\times p}$*, and* $\mathbf{V}(> 0) \in \mathbb{R}^{q\times q}$*; see Gupta and Nagar (2000). Here, we set* $(\boldsymbol{U}_1)_{jk} = \mathrm{u}_1^{|j-k|}$*,* $(\boldsymbol{V}_1)_{jk} = \mathrm{v}_1^{|j-k|}$*,* $(\boldsymbol{U}_2)_{jk} = \mathrm{u}_2^{|j-k|}$*,* $(\boldsymbol{V}_2)_{jk} = \mathrm{v}_2^{|j-k|}$*, and* $(\boldsymbol{M}_2)_{jk} = \mu$ *for some positive constants* u_1*,* v_1*,* u_2*, and* v_2*.*

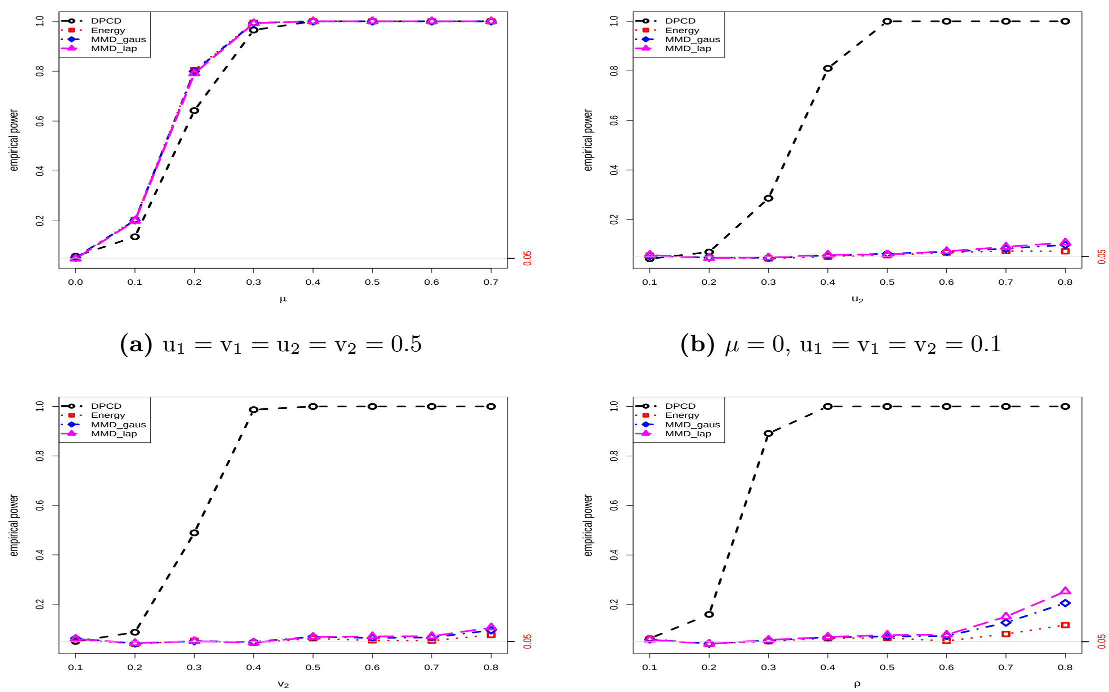
The empirical type-I error rates for Example 2 with $(p, q) = (25, 15)$ are summarized in Table 3. The results indicate that the empirical type-I error rates of all the methods are under reasonable control.

When the location parameter μ varies, the empirical power comparison is summarized in Figures 2(a), 3(a), and 4(a) at $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{v}_2 = 0.1$. From Figure 2(a), we can see that the DPCD has slightly lower power compared to the Energy, MMD_{gaus}, and MMD_{lap} in Case 1. However, Figures 3(a) and 4(a) show that the DPCD is more powerful than the others in Case 2 with $df = 1$ and 3.

When the scale parameters $\mathrm{u}_2, \mathrm{v}_2$ vary, the empirical power comparison is summarized in Figures 2–4(b)–(d) (the results for $\mathrm{u}_1, \mathrm{v}_1$ are similar and so are not reported here). Specifically, Figure 2(b)–(d) show plots of the power against u_2, v_2, and both of them, where the other parameters (such as $\mu, \mathrm{u}_1, \mathrm{v}_1$) are shown in each figure. Figures 3–4(b)–(d) are similarly set.

Figures 2–4(b)–(d) show that the DPCD outperforms substantially the Energy, MMD_{gaus}, and MMD_{lap} in each setting. It is surprising to see that the Energy and MMD suffer from very low power in Case 1. Although the Energy, MMD_{gaus}, and MMD_{lap} perform well when detecting the location shift [see Figures 2(a) and 3(a)], all of them exhibit a weaker ability to capture the two and higher-order moment information. This may be mainly because the Energy and MMD use only the linear combination of eigenvalues $\sum_{k=1}^{r} \lambda_k$ to extract the matrix information; see (18)–(19).

From Figures 2–4(b)–(d), the DPCD is sensitive to the difference between high-order moments, and it has superior performance in all cases. This may be mainly because it



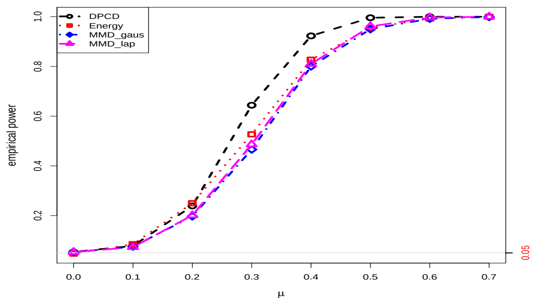
(a) $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{u}_2 = \mathrm{v}_2 = 0.5$

(b) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{v}_2 = 0.1$

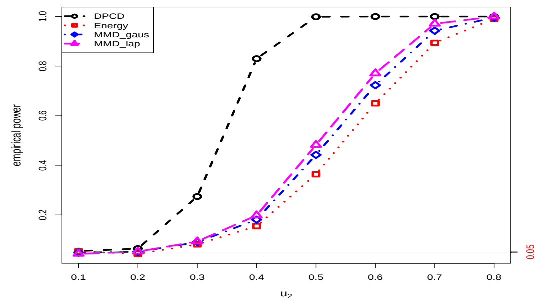
(c) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{u}_2 = 0.1$

(d) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = 0.1$, $\mathrm{u}_2 = \mathrm{v}_2 = \rho$

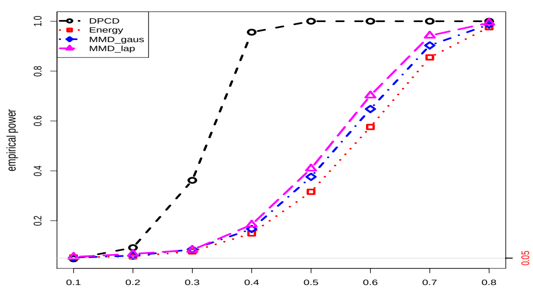
Figure 2: Empirical power at $\alpha = 0.05$ for Case 1 of Example 2 with $n_1 = n_2 = 20$. (a): $\mathbf{M}_2$ varies with μ; (b): $\mathbf{U}_2$ varies with u_2; (c): $\mathbf{V}_2$ varies with v_2; (d): $\mathbf{U}_2$ and $\mathbf{V}_2$ (letting $\mathbf{U}_2 = \mathbf{V}_2$) vary with ρ.



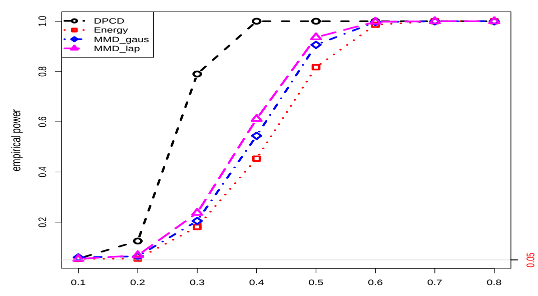
(a) $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{u}_2 = \mathrm{v}_2 = 0.5$



(b) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{v}_2 = 0.1$

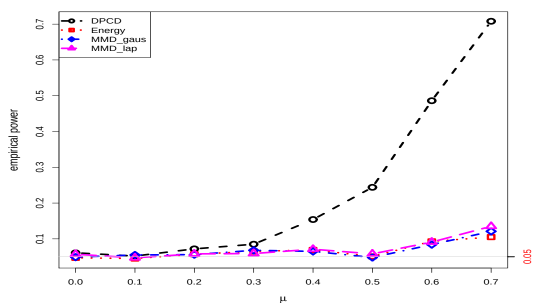


(c) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{u}_2 = 0.1$

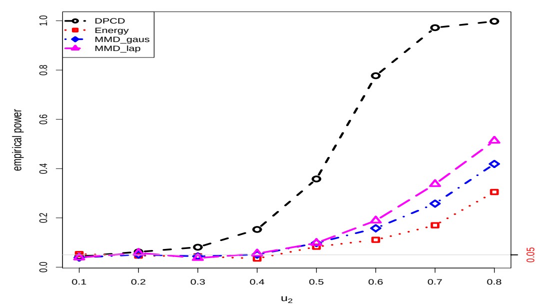


(d) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = 0.1$, $\mathrm{u}_2 = \mathrm{v}_2 = \rho$

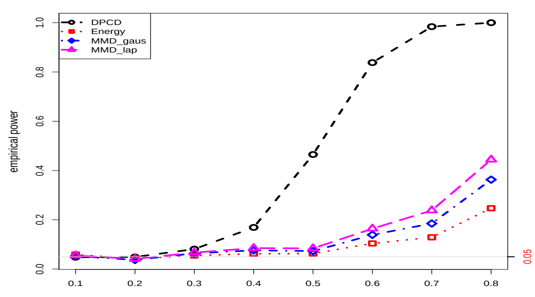
Figure 3: Empirical power at $\alpha = 0.05$ for Case 2 of Example 2 with $n_1 = n_2 = 20$ and $df = 3$. (a): $\mathbf{M}_2$ varies with μ; (b): $\mathbf{U}_2$ varies with u_2; (c): $\mathbf{V}_2$ varies with v_2; (d): $\mathbf{U}_2$ and $\mathbf{V}_2$ (letting $\mathbf{U}_2 = \mathbf{V}_2$) vary with ρ.



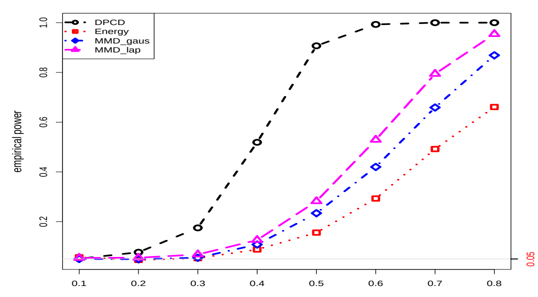
(a) $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{u}_2 = \mathrm{v}_2 = 0.5$



(b) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{v}_2 = 0.1$



(c) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{u}_2 = 0.1$



(d) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = 0.1$, $\mathrm{u}_2 = \mathrm{v}_2 = \rho$

Figure 4: Empirical power at $\alpha = 0.05$ for Case 2 of Example 2 with $n_1 = n_2 = 20$ and $df = 1$. (a): $\mathbf{M}_2$ varies with μ; (b): $\mathbf{U}_2$ varies with u_2; (c): $\mathbf{V}_2$ varies with v_2; (d): $\mathbf{U}_2$ and $\mathbf{V}_2$ (letting $\mathbf{U}_2 = \mathbf{V}_2$) vary with ρ.

Table 4: Empirical type-I error rate with $p = 25$ for Example 3.

Settings	$n_1 = n_2$	Method	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Example 3	20	SPCD	0.008	0.044	0.099
(Case 1)		DPCD	0.006	0.054	0.108
		Energy	0.008	0.052	0.110
		MMD_{gaus}	0.008	0.052	0.110
		MMD_{lap}	0.010	0.050	0.110
Example 3	20	SPCD	0.005	0.044	0.095
(Case 2, $df = 3$)		DPCD	0.007	0.050	0.094
		Energy	0.008	0.043	0.087
		MMD_{gaus}	0.006	0.044	0.094
		MMD_{lap}	0.007	0.042	0.092
Example 3	20	SPCD	0.013	0.054	0.106
(Case 2, $df = 1$)		DPCD	0.010	0.058	0.105
		Energy	0.016	0.062	0.106
		MMD_{gaus}	0.010	0.059	0.114
		MMD_{lap}	0.012	0.060	0.114

uses the two-layer nonlinear transformations of the eigenvalues in (20). In summary, all the results suggest that the DPCD is powerful and robust for detecting the differences of the location and high-order moments.

We provide two additional examples in Appendix. Specifically, in Example C.1, we investigate the effect of the dimensions p, q on the power of the DPCD; In Example C.2, we consider the performance of the DPCD to detect the difference between two general distributions, not just matrix-variate normal and t distributions. The simulation results also indicate that the DPCP outperforms the others across all scenarios.

5.2 SPCD/SPCD$_{\text{Log}}$ for symmetric/SPD matrix data

We study the finite-sample performances of the SPCD for symmetric matrix-valued data in Example 3 and the SPCD$_{\text{Log}}$ for SPD matrix-valued data in Example 4.

Example 3 *In this example, we consider the two-sample test for symmetric matrix-valued data, generated as*

$$\widetilde{\mathbf{X}} = \text{triu}(\mathbf{A}) + (\text{triu}(\mathbf{A}))^T - \text{diag}(\mathbf{A}) \quad \textit{and} \quad \widetilde{\mathbf{Y}} = \text{triu}(\mathbf{B}) + (\text{triu}(\mathbf{B}))^T - \text{diag}(\mathbf{B}),$$

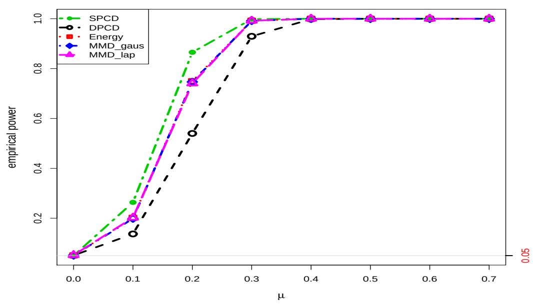
where $\text{triu}(\mathbf{A})$ *is the operator for extracting the upper triangular part of* $\mathbf{A} \in \mathbb{R}^{p\times p}$. *Here, we set the random matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *as follows:*

Case 1: $\mathbf{A} \sim \boldsymbol{MN}_{p\times p}(0, \mathbf{U}_1, \mathbf{V}_1)$ *and* $\mathbf{B} \sim \boldsymbol{MN}_{p\times p}(\mathbf{M}_2, \mathbf{U}_2, \mathbf{V}_2)$*;*

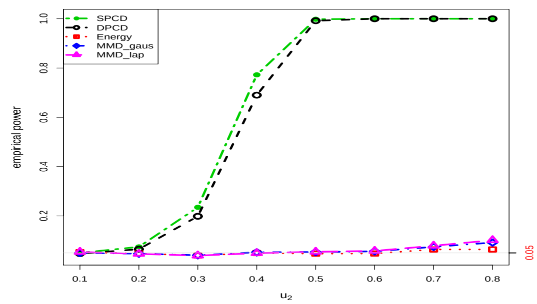
Case 2: $\mathbf{A} \sim \boldsymbol{T}_{p\times p}(df, \mathbf{0}, \mathbf{U}_1, \mathbf{V}_1)$ *and* $\mathbf{B} \sim \boldsymbol{T}_{p\times p}(df, \mathbf{M}_2, \mathbf{U}_2, \mathbf{V}_2)$ *with* $df = 3$ *or 1,*

where $\mathbf{M}_2, \mathbf{U}_1, \mathbf{U}_2, \mathbf{V}_1, \mathbf{V}_2$ *are set as those in Example 2. Here, we consider* $p = 25$.

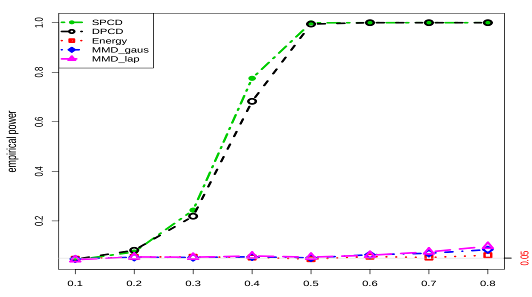
Table 4 summarizes the empirical type-I error rates for Example 3, from which we can see that the empirical sizes of the SPCD, DPCD, Energy, MMD_{gaus}, and MMD_{lap} tests are very close to the significance levels.



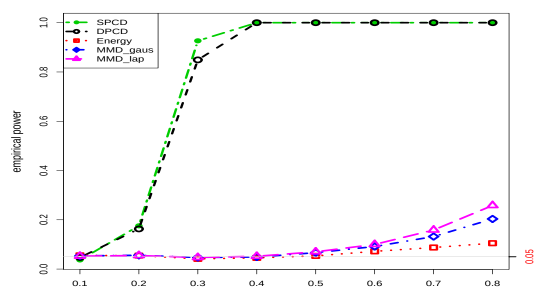
(a) $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{u}_2 = \mathrm{v}_2 = 0.5$



(b) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{v}_2 = 0.1$

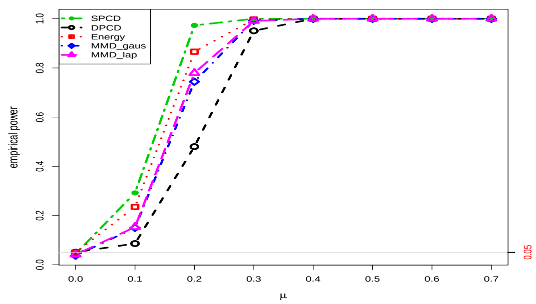


(c) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{u}_2 = 0.1$

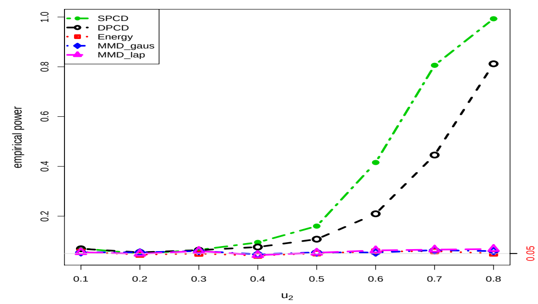


(d) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = 0.1$, $\mathrm{u}_2 = \mathrm{v}_2 = \rho$

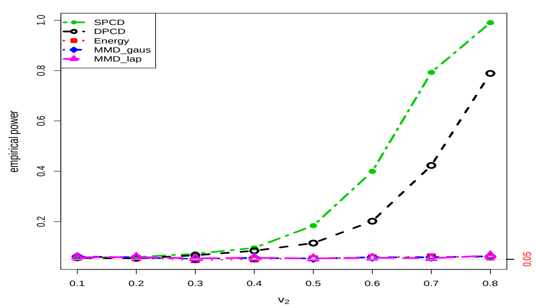
Figure 5: Empirical power at $\alpha = 0.05$ for Case 1 of Example 3 with $n_1 = n_2 = 20$. (a): $\mathbf{M}_2$ varies with μ; (b): $\mathbf{U}_2$ varies with u_2; (c): $\mathbf{V}_2$ varies with v_2; (d): $\mathbf{U}_2$ and $\mathbf{V}_2$ (letting $\mathbf{U}_2 = \mathbf{V}_2$) vary with ρ.



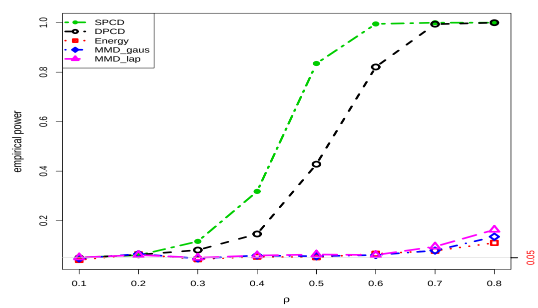
(a) $u_1 = v_1 = u_2 = v_2 = 0.5$



(b) $\mu = 0$, $u_1 = v_1 = v_2 = 0.1$

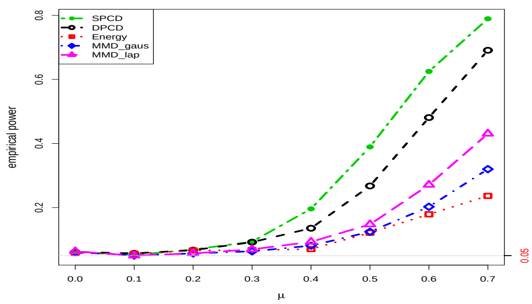


(c) $\mu = 0$, $u_1 = v_1 = u_2 = 0.1$

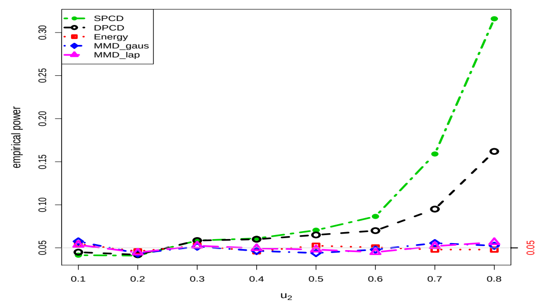


(d) $\mu = 0$, $u_1 = v_1 = 0.1$, $u_2 = v_2 = \rho$

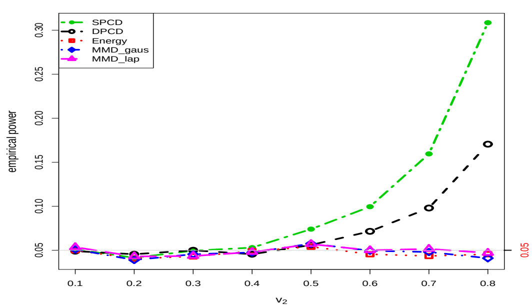
Figure 6: Empirical power at $\alpha = 0.05$ for Case 2 of Example 3 with $n_1 = n_2 = 20$ and $df = 3$. (a): $\mathbf{M}_2$ varies with μ; (b): $\mathbf{U}_2$ varies with u_2; (c): $\mathbf{V}_2$ varies with v_2; (d): $\mathbf{U}_2$ and $\mathbf{V}_2$ (letting $\mathbf{U}_2 = \mathbf{V}_2$) vary with ρ.



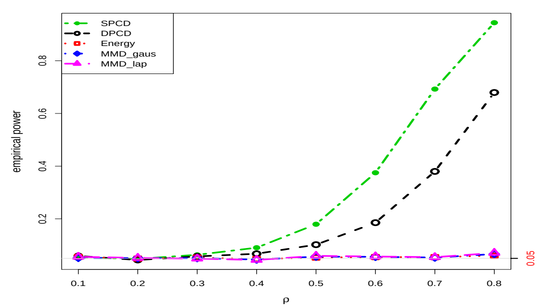
(a) $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{u}_2 = \mathrm{v}_2 = 0.5$



(b) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{v}_2 = 0.1$



(c) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = \mathrm{u}_2 = 0.1$



(d) $\mu = 0$, $\mathrm{u}_1 = \mathrm{v}_1 = 0.1$, $\mathrm{u}_2 = \mathrm{v}_2 = \rho$

Figure 7: Empirical power at $\alpha = 0.05$ for Case 2 of Example 3 with $n_1 = n_2 = 20$ and $df = 1$. (a): $\mathbf{M}_2$ varies with μ; (b): $\mathbf{U}_2$ varies with u_2; (c): $\mathbf{V}_2$ varies with v_2; (d) $\mathbf{U}_2$ and $\mathbf{V}_2$ (letting $\mathbf{U}_2 = \mathbf{V}_2$) vary with ρ.

By settings similar to those in Figures 2–4(a)–(d), Figures 5–7 show plots of the power against μ, u_2, and v_2. From Figures 5–7, we can see that our SPCD performs the best, followed by the DPCD and then Energy, $\mathrm{MMD_{gaus}}$, and $\mathrm{MMD_{lap}}$. We also see that the SPCD inherits the desirable properties of the DPCD, e.g. it is robust to heavy-tailed data and requires no assumption about the distribution.

In Figures 5–7, the SPCD always outperforms the DPCD. It benefits from the double-projection technique in Lemma 2, which makes full use of the symmetry structure. By contrast, from Figures 5(a) and 6(a), we can see that the DPCD has lower power than those of the Energy, $\mathrm{MMD_{gaus}}$, and $\mathrm{MMD_{lap}}$. This may be mainly because the DPCD ignores the prior knowledge of the symmetry.

Example 4 *This example examines the two-sample test for the SPD matrix-variate $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. We generate the data from the noncentral Wishart distribution, denoted by $\boldsymbol{W}_p(df, \boldsymbol{\Sigma}, \boldsymbol{M})$, where df is the degree of freedom and $\boldsymbol{\Sigma}$ and $\boldsymbol{M}$ are parameters. Consider the following cases:*

Case 1: $\boldsymbol{S}_1 \sim \boldsymbol{W}_p(df, \boldsymbol{\Sigma}, \boldsymbol{0})$ *and* $\boldsymbol{S}_2 \sim \boldsymbol{W}_p(df, \boldsymbol{\Sigma}, \boldsymbol{M}_2)$, *with* $\boldsymbol{M}_2 = \mathrm{diag}\{\delta, \cdots, \delta\}$ *and* $(\boldsymbol{\Sigma})_{j,k} = 0.5^{|j-k|}$;

Case 2: $\boldsymbol{S}_1 \sim \boldsymbol{W}_p(df, \boldsymbol{\Sigma}_1, \boldsymbol{0})$ *and* $\boldsymbol{S}_2 \sim \boldsymbol{W}_p(df, \boldsymbol{\Sigma}_2, \boldsymbol{0})$, *with* $(\boldsymbol{\Sigma}_1)_{j,k} = 0.1^{|j-k|}$ *and* $(\boldsymbol{\Sigma}_2)_{j,k} = \delta^{|j-k|}$;

Case 3: $\boldsymbol{S}_1 \sim \boldsymbol{W}_p(df, \boldsymbol{\Sigma}, \boldsymbol{0})$ *and* $\boldsymbol{S}_2 \sim \boldsymbol{W}_p(\delta, \boldsymbol{\Sigma}, \boldsymbol{0})$, *where* $(\boldsymbol{\Sigma})_{j,k} = 0.5^{|j-k|}$;

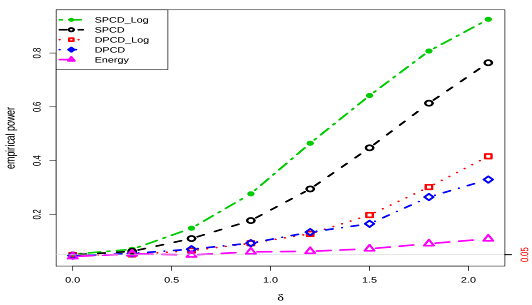
Case 4: $\boldsymbol{S}_1 \sim \boldsymbol{W}_p(df, \boldsymbol{\Sigma}, \boldsymbol{0})$ *and* $\boldsymbol{S}_2 \sim (1-\delta)\, \boldsymbol{W}_p(df, \boldsymbol{\Sigma}, \boldsymbol{0}) + \delta\, \boldsymbol{W}_p(df, \boldsymbol{\Sigma}, \boldsymbol{M}_2)$, *where* $(\boldsymbol{\Sigma})_{j,k} = 0.5^{|j-k|}$ *and* $(\boldsymbol{M}_2)_{j,k} = 0.5$.

Here, we consider $p = 10$ *and* $df = 30$.

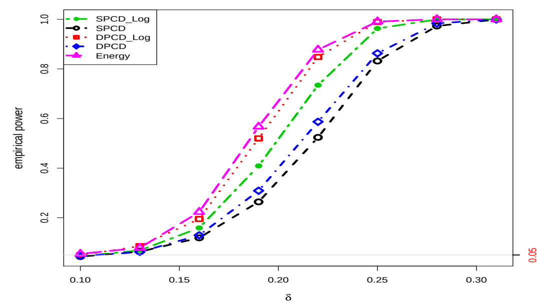
Figure 8 shows plots of the power against δ for Example 4. From Figure 8(a), (c), and (d), we can see that $\mathrm{SPCD_{Log}}$ outperforms the others. These results suggest that $\mathrm{SPCD_{Log}}$ (or $\mathrm{DPCD_{Log}}$) makes effective use of the prior knowledge of positive-definiteness to improve the power of the SPCD (or DPCD), which is consistent with Table 1.

Note that the mean of $\mathbf{W}_p(df, \boldsymbol{\Sigma}, \mathbf{M})$ is $df\boldsymbol{\Sigma}$. Thus, in Case 2, the difference between $\mathbf{S}_1$ and $\mathbf{S}_2$ depends only on their one-order moments $df\boldsymbol{\Sigma}_1$ and $df\boldsymbol{\Sigma}_2$. As demonstrated in the above examples, the Energy and MMD perform well when capturing the one-order moment difference between two distributions. Thus, in Figure 8(b), it is unsurprising that the Energy is slightly better than the $\mathrm{SPCD_{Log}}$ and $\mathrm{DPCD_{Log}}$.

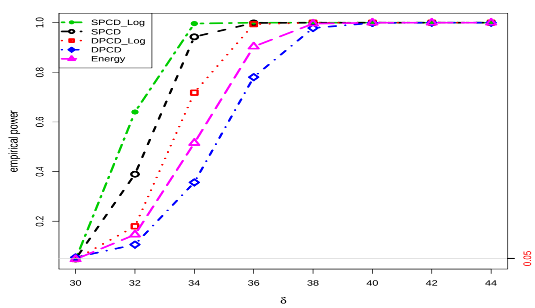
The above numerical results show that the finite-sample performances of the proposed tests are quite encouraging. To conclude, we make the following observations. 1) The proposed DPCD, SPCD, and $\mathrm{SPCD_{Log}}$ are nonparametric and robust, relying on no assumption about the underlying distributions and being robust to heavy-tailed data. 2) By vectorizing, the MMD and Energy capture the one-order moment information effectively but often perform poorly at detecting the high-order moment information. 3) When the difference between two distributions is determined by one-order moments, our methods are mostly comparable to the MMD and Energy; when the difference is determined by high-order moments, our methods always outperform the MMD and Energy significantly. These



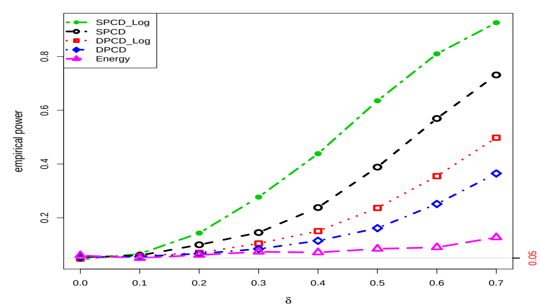
(a) Case 1



(b) Case 2



(c) Case 3



(d) Case 4

Figure 8: Empirical power at $\alpha = 0.05$ for Example 4 with $n_1 = n_2 = 20$.

benefit from the nonlinear way in (20) of extracting the matrix structure information. 4) As expected, the SPCD (resp. SPCD_{Log}) improves the performance of the DPCD (resp. SPCD) by utilizing prior information about the symmetry (resp. positive-definiteness). 5) The recommended parameters of γ_{med}^{-1} for the DPCD and DPCD_{Log} and $\gamma_{\text{med}}^{-1/2}$ for the SPCD and SPCD_{Log} perform well across all scenarios.

6. Analysis of real data

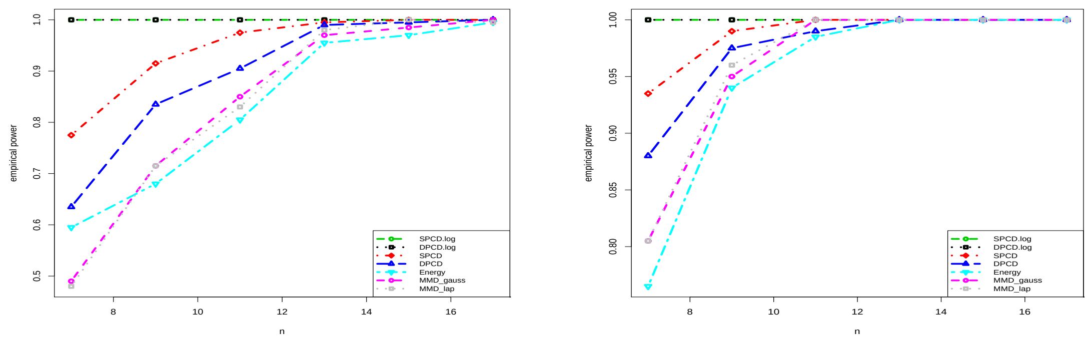
We illustrate the proposed procedures by two real data examples mentioned in Section 1, in which $B = 999$ permutation replicates are carried out to estimate the p-values for each test.

Example 5 (Stock covariance matrix data) *In the example, we would detect whether the Shenzhen and Shanghai stock exchanges of China have the same pattern during a specific time period. As an illustration, we collect the 1-min intraday stock data for the top 50 stocks of the Shenzhen and Shanghai stock exchanges from June 22, 2020 to October 15, 2020, which includes a total of 76 trading days. For each stock, we collected 240-min daily returns during the periods 9:30–11:30 and 13:30–15:30. From the 1-min intraday stock data, we can construct the daily-return covariance matrices* $\underline{\Sigma}_1 \in \mathbb{S}_{50}^{+}$ *for the top 50 stocks in the Shenzhen stock exchange and* $\underline{\Sigma}_2 \in \mathbb{S}_{50}^{+}$ *for the top 50 stocks in the Shanghai stock exchange. Note that* $\underline{\Sigma}_1$ *and* $\underline{\Sigma}_2$ *can be viewed as two SPD matrix-variate populations. Here, our goal is to test whether* $\underline{\Sigma}_1$ *and* $\underline{\Sigma}_2$ *are equally distributed using the data* $\{\underline{\Sigma}_{1,i}, \underline{\Sigma}_{2,i}\}_{i=1}^{76}$*, which can be estimated by the 75-trading-day data.*

To identify the difference between the distributions $P_{\underline{\Sigma}_1}$ and $P_{\underline{\Sigma}_2}$, we use the SPCD_{Log}, DPCD_{Log}, SPCD, DPCD, Energy, MMD_{gaus}, and MMD_{lap}. Our results indicate that the p-values for the results of these seven methods are all less than 0.001 and thus the distributions of $\underline{\Sigma}_1$ and $\underline{\Sigma}_2$ are significantly different. We further apply the seven methods on subsets of the entire data to provide power comparisons. For the resampling, (i) for a given subsample size, we pick randomly and evenly from $\{\underline{\Sigma}_{1,i}, \underline{\Sigma}_{2,i}\}_{i=1}^{76}$ to form the subsets, and (ii) we repeat each resampling 200 times and calculate the empirical power of each test method.

Figure 9 reports the empirical powers at significance levels of 0.01 and 0.05 with the subsample size n ranging from 1 to 17. From Figure 9, we can see that the proposed procedures significantly outperform the Energy, MMD_{gaus}, and MMD_{lap}. The results further suggest that the SPCD_{Log} and DPCD_{Log} make effective use of the prior knowledge of positive-definiteness to improve the power of the SPCD and DPCD, which is consistent with our findings in the previous sections.

Example 6 (EEG data) *In this example, we use the DPCD test method to analyze the EEG data. The data were collected from 45 patients with schizophrenia (aged between 10 and 14 years) and 39 age-matched healthy adolescents. The EEG signals for each subject were recorded at a sampling rate of 128 Hz from 16 channels of electrodes, so in 60 s, 7680 EEG signals were recorded for each channel. Thus, we can obtain the two-sample matrix-valued data given by* $\{\mathbf{X}_i\}_{i=1}^{45}$ *and* $\{\mathbf{Y}_i\}_{i=1}^{39}$ *with* $\mathbf{X}_i, \mathbf{Y}_i \in \mathbb{R}^{7680\times 16}$*. The data set is available at* `http://brain.bio.msu.ru/eeg_schizophrenia.htm`.



$\alpha = 0.01$ $\alpha = 0.05$

Figure 9: Empirical power comparisons for stock covariance matrix data.

Table 5: Analysis of electroencephalogram (EEG) data.

Method	DPCD	Energy	MMD_{gaus}	MMD_{lap}
p-value	**0.0699**	0.1828	0.2460	0.1440

We use the DPCD, Energy, MMD_{gaus}, and MMD_{lap} methods to test whether the EEG recordings are equally distributed between the patients and healthy adolescents. In Table 5, we summarize the resulting p-values based on 999 permutations. From Table 5, we can see that the DPCD test detects significant differences between the two groups because of the smaller p-value. However, the Energy, MMD_{gaus}, and MMD_{lap} have larger p-values; this may be because the vectorizing operator used by the Energy, MMD_{gaus}, and MMD_{lap} turns a 7680×16 matrix into a 122 880-dimensional vector, which generates unmanageably high dimensionality and thus may suffer from power loss.

7. Discussion

Herein, we have studied systematically the two-sample test problem for various matrix-valued data. Specifically, using the double projection-averaging technique, we proposed the DPCD test for arbitrary-dimensional matrix-valued data, which is applicable for any scalar, vector, and matrix-valued data. Using the symmetric projection-averaging technique, we developed the SPCD test for symmetric matrix-valued data, and using the matrix logarithm transformation, we proposed the SPCD_{Log} test for SPD matrix-valued data. Unlike the existing vector-valued tests such as the MMD and Energy, we showed that our methods capture the information of matrix-valued data through a two-layer nonlinear transformation of eigenvalues. Our extensive numerical studies confirmed that our methods based on the nonlinear transformation are very powerful for testing the equality of two matrix-variate distributions.

For the DPCD and SPCD, we chose the multivariate normal CDFs as the weight functions to obtain the closed-form expressions presented in Theorem 1(i) and Theorem 2(i). In fact, any weight functions that guarantee the validity of Theorem 1(ii) and Theorem 2(ii) are desirable. Therefore, we can choose any continuous CDF whose density function has the support $\mathbb{R}^p$ or $\mathbb{R}^q$ as the weight function. However, such weight functions may cause the DPCD and SPCD to lose their closed form. In this case, we can calculate the integrals involved in Definitions 1 and 2 through random sampling at a higher computational cost. Also, although the recommended parameters γ_{med}^{-1} and $\gamma_{\text{med}}^{-1/2}$ performed consistently well in the numerical studies, choosing optimal parameters would be an interesting but challenging topic for future research.

Multi-dimensional array (tensor) data are becoming increasingly common in various fields such as economics, finance, and medicine, among others, and it would be interesting to extend our methods to the two-sample test problem for such data. Generally, the projection averaging methods might be feasible, but it would be difficult to choose appropriate weights to obtain computationally tractable test statistics.

A.

In this appendix, we prove the lemmas, propositions and theorems from the above sections.
Proof of Lemma 1. The '$\Rightarrow$' part is straightforward, and here we prove the '$\Leftarrow$' part. Following Gupta and Nagar (2000), the characteristic functions of the random matrices $\mathbf{X}$ and $\mathbf{Y}$ can be defined as

$$\varphi_{\mathbf{X}}(\mathbf{T}) = E\{\exp\{i\,\mathrm{tr}(\mathbf{TX})\}\} \quad \text{and} \quad \varphi_{\mathbf{Y}}(\mathbf{T}) = E\{\exp\{i\,\mathrm{tr}(\mathbf{TY})\}\} \text{ for any } \mathbf{T} \in \mathbb{R}^{q\times p}.$$

By singular value decomposition (SVD), we obtain

$$\mathbf{T} = \sum_{k=1}^{r} \lambda_k \beta_k \alpha_k^T, \quad \alpha_k \in \mathbb{R}^p,\ \beta_k \in \mathbb{R}^q,$$

where $r = \mathrm{rank}(\mathbf{T})$. Then, we have that

$$\begin{aligned}
\mathrm{tr}(\mathbf{XT}) &= \mathrm{tr}(\mathbf{X}\sum_{k=1}^{r} \lambda_k \beta_k \alpha_k^T) = \sum_{k=1}^{r} \lambda_k\, \mathrm{tr}(\mathbf{X}\beta_k\alpha_k^T) = \sum_{k=1}^{r} \lambda_k \alpha_k^T \mathbf{X}\beta_k,\\
\mathrm{tr}(\mathbf{YT}) &= \mathrm{tr}(\mathbf{Y}\sum_{k=1}^{r} \lambda_k \beta_k \alpha_k^T) = \sum_{k=1}^{r} \lambda_k\, \mathrm{tr}(\mathbf{Y}\beta_k\alpha_k^T) = \sum_{k=1}^{r} \lambda_k \alpha_k^T \mathbf{Y}\beta_k,
\end{aligned}$$

which indicates that

$$\begin{aligned}
\varphi_{\mathbf{X}}(\mathbf{T}) - \varphi_{\mathbf{Y}}(\mathbf{T}) &= E\{\exp\{i\,\mathrm{tr}(\mathbf{TX})\} - E\{\exp\{i\,\mathrm{tr}(\mathbf{TY})\}\}\\
&= E\Big\{\exp\{i\sum_{k=1}^{r} \lambda_k \alpha_k^T \mathbf{X}\beta_k\}\Big\} - E\Big\{\exp\{i\sum_{k=1}^{r} \lambda_k \alpha_k^T \mathbf{Y}\beta_k\}\Big\}. \qquad (21)
\end{aligned}$$

Thus, if $P_{\alpha^T\mathbf{X}\beta} = P_{\alpha^T\mathbf{Y}\beta}$ holds for any $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$, then we obtain

$$\sum_{k=1}^{p} \alpha_k^T \mathbf{X}\beta_k \stackrel{d}{=} \sum_{k=1}^{p} \alpha_k^T \mathbf{Y}\beta_k,$$

where '$\stackrel{d}{=}$' means 'is identically distributed to'. This, together with (21), yields

$$\varphi_{\mathbf{X}}(\mathbf{T}) = \varphi_{\mathbf{Y}}(\mathbf{T}).$$

Thus, if $P_{\alpha^T\mathbf{X}\beta} = P_{\alpha^T\mathbf{Y}\beta}$ holds for any $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$, then we conclude that $P_{\mathbf{X}} = P_{\mathbf{Y}}$. ■

Proof of Theorem 1. (i) For any given $\alpha \in \mathbb{R}^p$ and $\beta \in \mathbb{R}^q$, we have that

$$\begin{aligned}
&\|E\{\exp\{i\alpha^T\mathbf{X}\beta\}\} - E\{\exp\{i\alpha^T\mathbf{Y}\beta\}\}\|^2\\
=\ & \Big[E\{\exp\{i\alpha^T\mathbf{X}\beta\}\} - E\{\exp\{i\alpha^T\mathbf{Y}\beta\}\}\Big]\Big[E\{\exp\{-i\alpha^T\mathbf{X}\beta\}\} - E\{\exp\{-i\alpha^T\mathbf{Y}\beta\}\}\Big]\\
=\ & E\{e^{-i\alpha^T(\mathbf{X}_1-\mathbf{X}_2)\beta}\} - 2E\{e^{-i\alpha^T(\mathbf{X}_1-\mathbf{Y}_2)\beta}\} + E\{e^{-i\alpha^T(\mathbf{Y}_1-\mathbf{Y}_2)\beta}\}.
\end{aligned}$$

Because the characteristic function of $N_p(0,\gamma\mathbf{I}_p)$ is $K_G(\mathbf{t})=\exp\{-\frac{\gamma}{2}\mathbf{t}^T\mathbf{t}\}$ for any $\mathbf{t}\in\mathbb{R}^p$, we have that

$$\begin{aligned}
&\int \left\|E\left\{\exp\left\{i\alpha^T\mathbf{X}\beta\right\}\right\}-E\left\{\exp\left\{i\alpha^T\mathbf{Y}\beta\right\}\right\}\right\|^2 dG_1(\alpha)\\
=&\int\left\{E\left\{e^{-i\alpha^T(\mathbf{X}_1-\mathbf{X}_2)\beta}\right\}-2E\left\{e^{-i\alpha^T(\mathbf{X}_1-\mathbf{Y}_2)\beta}\right\}+E\left\{e^{-i\alpha^T(\mathbf{Y}_1-\mathbf{Y}_2)\beta}\right\}\right\}dG_1(\alpha) \qquad (22)\\
=&E\left\{K_G\left(\left(\mathbf{X}_1-\mathbf{X}_2\right)\beta\right)-2K_G\left(\left(\mathbf{X}_1-\mathbf{Y}_2\right)\beta\right)+K_G\left(\left(\mathbf{Y}_1-\mathbf{Y}_2\right)\beta\right)\right\}.
\end{aligned}$$

When $G_2(\beta)$ is the q-multivariate standard Gaussian distribution, we obtain

$$\begin{aligned}
&\int K_G((\mathbf{z}-\mathbf{z}')\beta)dG_2(\beta)\\
=&\int\exp\Big\{-\frac{\gamma}{2}\beta^T(\mathbf{z}-\mathbf{z}')^T(\mathbf{z}-\mathbf{z}')\beta\Big\}(2\pi)^{-q/2}\exp\Big\{-\frac{\beta^T\beta}{2}\Big\}d\beta\\
=&(2\pi)^{-q/2}\int\exp\Big\{-\frac{1}{2}\beta^T\Big[\gamma(\mathbf{z}-\mathbf{z}')^T(\mathbf{z}-\mathbf{z}')+\mathbf{I}_q\Big]\beta\Big\}d\beta\\
=&\det(\gamma(\mathbf{z}-\mathbf{z}')^T(\mathbf{z}-\mathbf{z}')+\mathbf{I}_q)^{-1/2}
\end{aligned}$$

for any $\mathbf{z},\mathbf{z}'\in\mathbb{R}^{p\times q}$. This, together with (22), yields

$$\begin{aligned}
&\mathrm{DPCD}(\mathbf{X},\mathbf{Y})\\
=&\iint\|E\{\exp\{i\alpha^T\mathbf{X}\beta\}\}-E\{\exp\{i\alpha^T\mathbf{Y}\beta\}\}\|^2dG_1(\alpha)dG_2(\beta)\\
=&E\Big\{\int[K_G((\mathbf{X}_1-\mathbf{X}_2)\beta)-2K_G((\mathbf{X}_1-\mathbf{Y}_2)\beta)+K_G((\mathbf{Y}_1-\mathbf{Y}_2)\beta)]dG_2(\beta)\Big\}\\
=&E\{\det(\gamma(\mathbf{X}_1-\mathbf{X}_2)^T(\mathbf{X}_1-\mathbf{X}_2)+\mathbf{I}_q)^{-1/2}\}-2E\{\det(\gamma(\mathbf{X}_1-\mathbf{Y}_2)^T(\mathbf{X}_1-\mathbf{Y}_2)+\mathbf{I}_q)^{-1/2}\}\\
&+E\{\det(\gamma(\mathbf{Y}_1-\mathbf{Y}_2)^T(\mathbf{Y}_1-\mathbf{Y}_2)+\mathbf{I}_q)^{-1/2}\}.
\end{aligned}$$

(ii) The case of $\mathrm{DPCD}(\mathbf{X},\mathbf{Y})\geq 0$ is straightforward. Moreover, by the definition of $\mathrm{DPCD}(\mathbf{X},\mathbf{Y})$, we have that

$$\begin{aligned}
\mathrm{DPCD}(\mathbf{X},\mathbf{Y})=0 \iff& E\{\exp\{i\alpha^T\mathbf{X}\beta\}\}=E\{\exp\{i\alpha^T\mathbf{Y}\beta\}\},\alpha\in\mathbb{R}^p,\beta\in\mathbb{R}^q\\
\iff& E\{\exp\{it\alpha^T\mathbf{X}\beta\}\}=E\{\exp\{it\alpha^T\mathbf{Y}\beta\}\},t\in\mathbb{R},\alpha\in\mathbb{R}^p,\beta\in\mathbb{R}^q\\
\iff& P_{\alpha^T\mathbf{X}\beta}=P_{\alpha^T\mathbf{Y}\beta},\ \text{for any } \alpha\in\mathbb{R}^p,\beta\in\mathbb{R}^q.
\end{aligned}$$

This, together with Lemma 1, indicates that $\mathrm{DPCD}(\mathbf{X},\mathbf{Y})=0$ if and only if $\mathbf{X}$ and $\mathbf{Y}$ are identically distributed.

■

Proof of Lemma 2. The '⇒' part is immediate by the elementary properties of independence; here, we prove the reverse. Note that $\widetilde{\mathbf{X}}=\widetilde{\mathbf{X}}^T\in\mathbb{R}^{p\times p}$. Thus, $\widetilde{\mathbf{X}}$ has only $p(p+1)/2$ different random variables, and its characteristic functions can be defined as

$$\varphi_{\mathbf{X}}(\theta)=E\{\exp\{i\sum_{j\leq k}\widetilde{\mathbf{X}}_{jk}\theta_{jk}\}\},$$

where θ has $p(p+1)/2$ different elements denoted by $\{\theta_{jk}, 1 \le j \le k \le p\}$. Define

$$T_{jj} = \theta_{jj}, T_{jk} = T_{kj} = \frac{1}{2}\theta_{jk}, 1 \le j \le k \le p.$$

Then, we have that

$$\varphi_{\widetilde{\mathbf{X}}}(\theta) = E\{\exp\{i \sum_{j \le k} \widetilde{\mathbf{X}}_{jk}\theta_{jk}\}\} = E\{\exp\{i \sum_{j,k=1}^{p} \widetilde{\mathbf{X}}_{jk}T_{jk}\}\} = E\{\exp\{i\,\mathrm{tr}(\widetilde{\mathbf{X}}\mathbf{T})\}\},$$

where $T = (T_{jk})$ is any symmetric matrix. Similarly, we define the characteristic function of $\widetilde{\mathbf{Y}}$ by

$$\varphi_{\widetilde{\mathbf{Y}}}(\theta) = E\{\exp\{i \sum_{j \le k} \widetilde{\mathbf{Y}}_{jk}\theta_{jk}\}\} = E\{\exp\{i \sum_{j,k} \widetilde{\mathbf{Y}}_{jk}T_{jk}\}\} = E\{\exp\{i\,\mathrm{tr}(\widetilde{\mathbf{Y}}\mathbf{T})\}\}.$$

By SVD, we have that $\mathbf{T} = \sum_{k=1}^{r} \lambda_k \alpha_k \alpha_k^T$ with $r = \mathrm{rank}(\mathbf{T})$. Thus, we obtain

$$\mathrm{tr}(\widetilde{\mathbf{X}}\mathbf{T}) = \mathrm{tr}(\widetilde{\mathbf{X}} \sum_{k=1}^{r} \lambda_k \alpha_k \alpha_k^T) = \sum_{k=1}^{r} \lambda_k\, \mathrm{tr}(\widetilde{\mathbf{X}}\alpha_k\alpha_k^T) = \sum_{k=1}^{r} \lambda_k \alpha_k^T \widetilde{\mathbf{X}}\alpha_k,$$
$$\mathrm{tr}(\widetilde{\mathbf{Y}}\mathbf{T}) = \mathrm{tr}(\widetilde{\mathbf{Y}} \sum_{k=1}^{r} \lambda_k \alpha_k \alpha_k^T) = \sum_{k=1}^{r} \lambda_k\, \mathrm{tr}(\widetilde{\mathbf{Y}}\alpha_k\alpha_k^T) = \sum_{k=1}^{r} \lambda_k \alpha_k^T \widetilde{\mathbf{Y}}\alpha_k.$$

Then, we have that

$$\begin{aligned} \varphi_{\widetilde{\mathbf{X}}}(\theta) - \varphi_{\widetilde{\mathbf{Y}}}(\theta) &= E\{\exp\{i\,\mathrm{tr}(\widetilde{\mathbf{X}}\mathbf{T})\}\} - E\{\exp\{i\,\mathrm{tr}(\widetilde{\mathbf{Y}}\mathbf{T})\}\}\} \\ &= E\{\exp\{i \sum_{k=1}^{r} \lambda_k \alpha_k^T \widetilde{\mathbf{X}}\alpha_k\}\} - E\{\exp\{i \sum_{k=1}^{r} \lambda_k \alpha_k^T \widetilde{\mathbf{Y}}\alpha_k\}\}. \end{aligned} \tag{23}$$

Thus, if $P_{\alpha^T\widetilde{\mathbf{X}}\alpha} = P_{\alpha^T\widetilde{\mathbf{Y}}\alpha}$ holds for $\alpha \in \mathbb{R}^p$, then we obtain

$$\sum_{k=1}^{r} \lambda_k \alpha_k^T \widetilde{\mathbf{X}}\alpha_k \overset{d}{=} \sum_{k=1}^{r} \lambda_k \alpha_k^T \widetilde{\mathbf{Y}}\alpha_k.$$

This, together with (23), yields that $\varphi_{\widetilde{\mathbf{X}}}(\theta) = \varphi_{\widetilde{\mathbf{Y}}}(\theta)$. Thus, if $P_{\alpha^T\widetilde{\mathbf{X}}\alpha} = P_{\alpha^T\widetilde{\mathbf{Y}}\alpha}$ holds for $\alpha \in \mathbb{R}^p$, then we have that $P_{\widetilde{\mathbf{X}}} = P_{\widetilde{\mathbf{Y}}}$. ■

Proof of Theorem 2. (i) We need the following results (Anderson, 2003, p. 259). When α is a p-dimensional distribution of $N_p(0, \gamma\mathbf{I}_p)$ with $\gamma > 0$, we have that

$$E\{\exp\{i\alpha^T\mathbf{z}\alpha\}\} = \int_{\mathbb{R}^p} \exp\{i\alpha^T\mathbf{z}\alpha\} dG_1(\alpha) = \det(\mathbf{I}_p - 2i\gamma\mathbf{z})^{-1/2} \tag{24}$$

for any given symmetric matrix $\mathbf{z} \in \mathbb{R}^{p\times p}$.

From (36), we have that

$$
\begin{aligned}
&\int_{\mathbb{R}^p} \|E\{\exp\{i\alpha^T\widetilde{\mathbf{X}}\alpha\}] - E[\exp\{i\alpha^T\widetilde{\mathbf{Y}}\alpha\}\}\|^2 dG_1(\alpha) \\
= &\int \Big\{E\{e^{i\alpha^T(\widetilde{\mathbf{X}}_1-\widetilde{\mathbf{X}}_2)\alpha}\} - 2E\{e^{i\alpha^T(\widetilde{\mathbf{X}}_1-\widetilde{\mathbf{Y}}_2)\alpha}\} + E\{e^{i\alpha^T(\widetilde{\mathbf{Y}}_1-\widetilde{\mathbf{Y}}_2)\alpha}\}\Big\} dG_1(\alpha) \\
= &E\{\det(\mathbf{I}_p - 2i\gamma(\widetilde{\mathbf{X}}_1 - \widetilde{\mathbf{X}}_2))^{-1/2}\} - 2E\{\det(\mathbf{I}_p - 2i\gamma(\widetilde{\mathbf{X}}_1 - \widetilde{\mathbf{Y}}_2))^{-1/2}\} \\
&+E\{\det(\mathbf{I}_p - 2i\gamma(\widetilde{\mathbf{Y}}_1 - \widetilde{\mathbf{Y}}_2))^{-1/2}\}.
\end{aligned}
$$

(ii) The case of $\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) \geq 0$ is straightforward. In addition, by the definition of $\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})$, we have that

$$
\begin{aligned}
\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) = 0 \quad &\Longleftrightarrow \quad E\{\exp\{i\alpha^T\widetilde{\mathbf{X}}\alpha\}\} = E\{\exp\{i\alpha^T\widetilde{\mathbf{Y}}\alpha\}\} \text{ for any } \alpha \in \mathbb{R}^p \\
&\Longleftrightarrow \quad E\{\exp\{it\alpha^T\widetilde{\mathbf{X}}\alpha\}\} = E\{\exp\{it\alpha^T\widetilde{\mathbf{Y}}\alpha\}\} \text{ for any } t \in \mathbb{R}, \alpha \in \mathbb{R}^p \\
&\Longleftrightarrow \quad P_{\alpha^T\widetilde{\mathbf{X}}\alpha} = P_{\alpha^T\widetilde{\mathbf{Y}}\alpha}.
\end{aligned}
$$

This, together with Lemma 2, yields that $\mathrm{SPCD}(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}) = 0$ if and only if $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$ are identically distributed. ■

Proof of Theorem 3. By the definition of $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$, we have that

$$
E\{\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})\} = E\{h(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2)\} = \mathrm{DPCD}(\mathbf{X}, \mathbf{Y}). \tag{25}
$$

In addition, note that

$$
\begin{aligned}
|h(\mathbf{x}_1, \mathbf{x}_2; \mathbf{y}_1, \mathbf{y}_2)| &\leq 4 \sup_{\mathbf{z}, \mathbf{z}' \in \mathbb{R}^{p\times q}} \left|\det\left(\gamma\left(\mathbf{z} - \mathbf{z}'\right)^T\left(\mathbf{z} - \mathbf{z}'\right) + \mathbf{I}_q\right)^{-1/2}\right| \\
&= 4 \sup_{\lambda_k \geq 0} \prod_{k=1}^{q} (\gamma\lambda_k + 1)^{-1/2} \\
&\leq 4,
\end{aligned} \tag{26}
$$

where λ_k are the eigenvalues of $(\mathbf{z} - \mathbf{z}')^T(\mathbf{z} - \mathbf{z}')$.

Let $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_{n+m}) = (\mathbf{X}_1, \ldots, \mathbf{X}_n, \mathbf{Y}_1, \ldots, \mathbf{Y}_m)$. For ease of notation, we view $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$ as a function of $\mathbf{Z}$ and thus denote $f(\mathbf{Z}) = \widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$. By (26), for all $\mathbf{z}_1, \cdots, \mathbf{z}_{n+m}$, $\mathbf{z}'_k \in \mathbb{R}^{p\times q}$, we have that if $k \leq n$, then

$$
\begin{aligned}
&|f(\mathbf{z}_1, \ldots, \mathbf{z}_k, \ldots, \mathbf{z}_{n+m}) - f(\mathbf{z}_1, \ldots, \mathbf{z}'_k, \ldots, \mathbf{z}_{n+m})| \\
\leq \quad &\frac{1}{n(n-1)} \frac{1}{m(m-1)} \sum_{\substack{i_1=1 \\ i_1 \neq k}}^{n} \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^{m} |h(\mathbf{x}_{i_1}, \mathbf{x}_k; \mathbf{y}_{j_1}, \mathbf{y}_{j_2}) - h(\mathbf{x}_{i_1}, \mathbf{x}'_k; \mathbf{y}_{j_1}, \mathbf{y}_{j_2})| \leq \frac{16}{n}.
\end{aligned}
$$

Similarly, if $n + 1 \leq k \leq n + m$, then we can obtain

$$
\begin{aligned}
&|f(\mathbf{z}_1, \ldots, \mathbf{z}_k, \ldots, \mathbf{z}_{n+m}) - f(\mathbf{z}_1, \ldots, \mathbf{z}'_k, \ldots, \mathbf{z}_{n+m})| \\
\leq \quad &\frac{1}{n(n-1)} \frac{1}{m(m-1)} \sum_{\substack{i_1, i_2=1 \\ i_1 \neq i_2}}^{n} \sum_{\substack{j_1=1 \\ j_1 \neq k}} |h(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}; \mathbf{y}_{j_1}, \mathbf{y}_k) - h(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}; \mathbf{y}_{j_1}, \mathbf{y}'_k)| \leq \frac{16}{m}.
\end{aligned}
$$

Using these two inequalities, we show that

$$|f(\mathbf{z}_1,\ldots,\mathbf{z}_k,\ldots,\mathbf{z}_{n+m}) - f(\mathbf{z}_1,\ldots,\mathbf{z}'_k,\ldots,\mathbf{z}_{n+m})| \leq 16\Big(\frac{1}{n}+\frac{1}{m}\Big).$$

By (25) and the bounded differences inequality (Wainwright, 2019, Corollary 2.21), taking $L_k = 16(\frac{1}{n}+\frac{1}{m})$ for each $k \in \{1,\cdots,n+m\}$, we have that

$$\begin{aligned} P\Big\{|\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y}) - \mathrm{DPCD}(\mathbf{X},\mathbf{Y})| \geq t\Big\} &\leq 2\exp\Big\{-\frac{2t^2}{\sum_{k=1}^{n+m} L_k^2}\Big\} \\ &\leq 2\exp\Big\{-\frac{2t^2}{16^2(n+m)(\frac{1}{n}+\frac{1}{m})^2}\Big\} \end{aligned}$$

for all $t \geq 0$. Thus, letting $t = 8\sqrt{2}\sqrt{\log(2/\alpha)}\sqrt{n+m}(\frac{1}{n}+\frac{1}{m})$, we have that

$$P\Big\{|\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y}) - \mathrm{DPCD}(\mathbf{X},\mathbf{Y})| \leq 8\sqrt{2}\sqrt{\log(2/\alpha)}\sqrt{n+m}(n^{-1}+m^{-1})\Big\} \geq 1-\alpha$$

for any $\alpha \in (0,1)$. ■

Proof of Theorem 4. We can use the asymptotic theory for a degenerate two-sample U-statistic in Theorem 3.5.1 of Bhat (1995, p. 89) to prove the theorem. First, we show that $h(\mathbf{x}_1,\mathbf{x}_2;\mathbf{y}_1,\mathbf{y}_2)$ in (9) is degenerate under the null hypothesis H_0. To this end, we define

$$\begin{aligned} h_{1,0}(\mathbf{x}_1) = E\{h(\mathbf{x}_1,\mathbf{X}_2;\mathbf{Y}_1,\mathbf{Y}_2)\}, \quad h_{0,1}(\mathbf{y}_1) = E\{h(\mathbf{X}_1,\mathbf{X}_2;\mathbf{y}_1,\mathbf{Y}_2)\}, \\ h_{0,2}(\mathbf{y}_1,\mathbf{y}_2) = E\{h(\mathbf{X}_1,\mathbf{X}_2;\mathbf{y}_1,\mathbf{y}_2)\}, \quad h_{1,1}(\mathbf{x}_1,\mathbf{y}_1) = E\{h(\mathbf{x}_1,\mathbf{X}_2;\mathbf{y}_1,\mathbf{Y}_2)\} \end{aligned} \tag{27}$$

for any $\mathbf{x}_1,\mathbf{y}_1,\mathbf{y}_2 \in \mathbb{R}^{p\times q}$. Under $H_0 : P_{\mathbf{X}} = P_{\mathbf{Y}}$, it is easy to see that

$$\begin{aligned} h_{1,0}(\mathbf{x}_1) &= E\{\mathrm{K}_{\mathrm{det}}(\mathbf{x}_1,\mathbf{X}_2) + \mathrm{K}_{\mathrm{det}}(\mathbf{Y}_1,\mathbf{Y}_2) - \frac{1}{2}\mathrm{K}_{\mathrm{det}}(\mathbf{x}_1,\mathbf{Y}_1) - \frac{1}{2}\mathrm{K}_{\mathrm{det}}(\mathbf{x}_1,\mathbf{Y}_2) \\ &\quad -\frac{1}{2}\mathrm{K}_{\mathrm{det}}(\mathbf{X}_2,\mathbf{Y}_1) - \frac{1}{2}\mathrm{K}_{\mathrm{det}}(\mathbf{X}_2,\mathbf{Y}_2)\} = 0 \end{aligned} \tag{28}$$

and

$$\begin{aligned} h_{0,1}(\mathbf{y}_1) &= E\{\mathrm{K}_{\mathrm{det}}(\mathbf{X}_1,\mathbf{X}_2) + \mathrm{K}_{\mathrm{det}}(\mathbf{y}_1,\mathbf{Y}_2) - \frac{1}{2}\mathrm{K}_{\mathrm{det}}(\mathbf{X}_1,\mathbf{y}_1) - \frac{1}{2}\mathrm{K}_{\mathrm{det}}(\mathbf{X}_1,\mathbf{Y}_2) \\ &\quad -\frac{1}{2}\mathrm{K}_{\mathrm{det}}(\mathbf{X}_2,\mathbf{y}_1) - \frac{1}{2}\mathrm{K}_{\mathrm{det}}(\mathbf{X}_2,\mathbf{Y}_2)\} = 0. \end{aligned} \tag{29}$$

In addition, we have that

$$\begin{aligned} h_{2,0}(\mathbf{x}_1,\mathbf{x}_2) &= E\{\mathrm{K}_{\mathrm{det}}(\mathbf{x}_1,\mathbf{x}_2) + \mathrm{K}_{\mathrm{det}}(\mathbf{Y}_1,\mathbf{Y}_2) - \frac{1}{2}\mathrm{K}_{\mathrm{det}}(\mathbf{x}_1,\mathbf{Y}_1) - \frac{1}{2}\mathrm{K}_{\mathrm{det}}(\mathbf{x}_1,\mathbf{Y}_2) \\ &\quad -\frac{1}{2}\mathrm{K}_{\mathrm{det}}(\mathbf{x}_2,\mathbf{Y}_1) - \frac{1}{2}\mathrm{K}_{\mathrm{det}}(\mathbf{x}_2,\mathbf{Y}_2)\} \\ &= \mathrm{K}_{\mathrm{det}}(\mathbf{x}_1,\mathbf{x}_2) + E\{\mathrm{K}_{\mathrm{det}}(\mathbf{Y}_1,\mathbf{Y}_2)\} - E\{\mathrm{K}_{\mathrm{det}}(\mathbf{x}_1,\mathbf{Y}_1)\} - E\{\mathrm{K}_{\mathrm{det}}(\mathbf{x}_2,\mathbf{Y}_1)\}, \end{aligned}$$

$$\begin{aligned} h_{0,2}(\mathbf{y}_1,\mathbf{y}_2) &= E\{\mathrm{K}_{\det}(\mathbf{X}_1,\mathbf{X}_2)+\mathrm{K}_{\det}(\mathbf{y}_1,\mathbf{y}_2)-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{X}_1,\mathbf{y}_1)-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{X}_1,\mathbf{y}_2)\\ &\quad -\frac{1}{2}\mathrm{K}_{\det}(\mathbf{X}_2,\mathbf{y}_1)-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{X}_2,\mathbf{y}_2)\}\\ &= E\{\mathrm{K}_{\det}(\mathbf{X}_1,\mathbf{X}_2)\}+\mathrm{K}_{\det}(\mathbf{y}_1,\mathbf{y}_2)-E\{\mathrm{K}_{\det}(\mathbf{X}_1,\mathbf{y}_1)\}-E\{\mathrm{K}_{\det}(\mathbf{X}_1,\mathbf{y}_2)\}, \end{aligned}$$

and

$$\begin{aligned} h_{1,1}(\mathbf{x}_1,\mathbf{y}_1) &= E\{\mathrm{K}_{\det}(\mathbf{x}_1,\mathbf{X}_2)+\mathrm{K}_{\det}(\mathbf{y}_1,\mathbf{Y}_2)-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{x}_1,\mathbf{y}_1)-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{x}_1,\mathbf{Y}_2)\\ &\quad -\frac{1}{2}\mathrm{K}_{\det}(\mathbf{X}_2,\mathbf{y}_1)-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{X}_2,\mathbf{Y}_2)\}\\ &= E\{\frac{1}{2}\mathrm{K}_{\det}(\mathbf{x}_1,\mathbf{X}_1)+\frac{1}{2}\mathrm{K}_{\det}(\mathbf{X}_1,\mathbf{y}_1)-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{x}_1,\mathbf{y}_1)-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{X}_1,\mathbf{Y}_1)\}. \end{aligned}$$

Under H_0, the above three equations suggest that

$$h_{0,2}(\mathbf{z}_1,\mathbf{z}_2)=h_{2,0}(\mathbf{z}_1,\mathbf{z}_2),\quad h_{1,1}(\mathbf{z}_1,\mathbf{z}_2)=-\frac{1}{2}h_{2,0}(\mathbf{z}_1,\mathbf{z}_2) \tag{30}$$

for any $\mathbf{z}_1,\mathbf{z}_2\in\mathbb{R}^{p\times q}$.

Note that $h_{2,0}(\mathbf{z}_1,\mathbf{z}_2)$ is not constant for almost all $\mathbf{z}_1,\mathbf{z}_2\in\mathbb{R}^{p\times q}$. In fact, taking $\mathbf{z}_2=\mathbf{z}_1$, we have that

$$\begin{aligned} h_{2,0}(\mathbf{z}_1,\mathbf{z}_1) &= \mathrm{K}_{\det}(\mathbf{z}_1,\mathbf{z}_1)+E\{\mathrm{K}_{\det}(\mathbf{Y}_1,\mathbf{Y}_2)\}-2E\{\mathrm{K}_{\det}(\mathbf{z}_1,\mathbf{Y}_1)\}\\ &= 1+E\{\mathrm{K}_{\det}(\mathbf{Y}_1,\mathbf{Y}_2)\}-2E\{\det(\gamma(\mathbf{z}_1-\mathbf{Y}_1)^T(\mathbf{z}_1-\mathbf{Y}_1)+\mathbf{I}_q)^{-1/2}\}. \end{aligned}$$

We can see that $h_{2,0}(\mathbf{z}_1,\mathbf{z}_1)$ is continuous and not a constant on $\mathbb{R}^{p\times q}$. Thus, $h_{2,0}(\mathbf{z}_1,\mathbf{z}_2)$ is not a constant in some nonzero measure set. This, together with (28)-(29), yields that $h(\mathbf{x}_1,\mathbf{x}_2;\mathbf{y}_1,\mathbf{y}_2)$ has degeneracy of order 1 under H_0.

By (28)–(30) and Theorem 3.5.1 in Bhat (1995), we have that

$$\begin{aligned} (m+n)\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y}) &\xrightarrow{d} \frac{1}{\pi_X}\sum_{k=1}^{\infty}\lambda_k\left(\xi_k^2-1\right)+\frac{1}{\pi_Y}\sum_{k=1}^{\infty}\lambda_k\left(\xi_k'^2-1\right)-\frac{2}{\sqrt{\pi_X\pi_Y}}\sum_{k=1}^{\infty}\lambda_k\xi_k\xi_k'\\ &= \frac{1}{\pi_X\pi_Y}\sum_{k=1}^{\infty}\lambda_k\left[\left(\sqrt{\pi_Y}\xi_k-\sqrt{\pi_X}\xi_k'\right)^2-1\right], \end{aligned}$$

where $\xi_k,\xi_k'\overset{i.i.d}{\sim}N(0,1)$ and λ_k are the eigenvalues of the integral equation

$$E\{h_{2,0}(\mathbf{x}_1,\mathbf{X}_2)\phi_k(\mathbf{X}_2)\}=\lambda_k\phi_k(\mathbf{x}_1)\text{ and }h_{2,0}(\mathbf{x}_1,\mathbf{x}_2)=E\{h(\mathbf{x}_1,\mathbf{x}_2;\mathbf{Y}_1,\mathbf{Y}_2)\}.$$

Note that $\sqrt{\pi_Y}\xi_k-\sqrt{\pi_X}\xi_k'\sim N(0,1)$. Then, we have that

$$(m+n)\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})\xrightarrow{d}\frac{1}{\pi_X\pi_Y}\sum_{k=1}^{\infty}\lambda_k(\xi_k^2-1).$$

■

Proof of Theorem 5. Note that $E\{\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})\} = \mathrm{DPCD}(\mathbf{X},\mathbf{Y}) \stackrel{\triangle}{=} \theta$. By Hoeffing's decomposition of a two-sample U-statistic (see Theorem 3 in Section 2 of Lee (1990)), $\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})$ can be represented by

$$\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y}) - \theta = \frac{2}{n}\sum_{i=1}^{n}\{h_{1,0}(\mathbf{X}_i) - \theta)\} + \frac{2}{m}\sum_{i=1}^{m}\{h_{0,1}(\mathbf{Y}_i) - \theta\} + O_p(R_n), \quad (31)$$

where

$$R_n = \frac{2}{n(n-1)}\sum_{i_1<i_2}^{m} \tilde{h}_{2,0}(\mathbf{X}_{i_1},\mathbf{X}_{i_2}) + \frac{2}{m(m-1)}\sum_{j_1<j_2}^{n} \tilde{h}_{0,2}(\mathbf{Y}_{j_1},\mathbf{Y}_{j_2}) + \frac{4}{nm}\sum_{i_1=1}^{n}\sum_{j_1=1}^{m} \tilde{h}_{1,1}(\mathbf{X}_{i_1};\mathbf{Y}_{j_1})$$

and

$$\begin{aligned}
\tilde{h}_{2,0}(\mathbf{x}_1,\mathbf{x}_2) &= E\{h(\mathbf{x}_1,\mathbf{x}_2;\mathbf{Y}_1,\mathbf{Y}_2)\} - [\tilde{h}_{1,0}(\mathbf{x}_1) + \tilde{h}_{1,0}(\mathbf{x}_2)] - \theta,\\
\tilde{h}_{0,2}(\mathbf{y}_1,\mathbf{y}_2) &= E\{h(\mathbf{X}_1,\mathbf{X}_2;\mathbf{y}_1,\mathbf{y}_2)\} - [\tilde{h}_{0,1}(\mathbf{y}_2) + \tilde{h}_{0,1}(\mathbf{y}_2)] - \theta,\\
\tilde{h}_{1,1}(\mathbf{x}_1;\mathbf{y}_1) &= E\{h(\mathbf{x}_1,\mathbf{X}_2;\mathbf{y}_1,\mathbf{Y}_2)\} - [\tilde{h}_{1,0}(\mathbf{x}_1) + \tilde{h}_{0,1}(\mathbf{y}_1)] - \theta,
\end{aligned}$$

with $\tilde{h}_{1,0}(\mathbf{x}_1) = h_{1,0}(\mathbf{x}_1) - \theta$ and $\tilde{h}_{0,1}(\mathbf{y}_1) = h_{0,1}(\mathbf{y}_1) - \theta$.

By the properties of the above $\tilde{h}_{c,d}$ $(c,d = 0,1,2)$ (see p. 41 of Lee (1990)), we have that $E\{R_n\} = 0$ and

$$\mathrm{Var}\{R_n\} = \frac{2}{n(n-1)}\mathrm{Var}\{\tilde{h}_{2,0}(\mathbf{X}_1,\mathbf{X}_2)\} + \frac{2}{m(m-1)}\mathrm{Var}\{\tilde{h}_{0,2}(\mathbf{Y}_1,\mathbf{Y}_2)\} + \frac{16}{nm}\mathrm{Var}\{\tilde{h}_{1,1}(\mathbf{X}_1;\mathbf{Y}_1)\},$$

which implies that $R_n = O_p((n+m)^{-1})$ by the boundedness of $h(\mathbf{x}_1,\mathbf{x}_2;\mathbf{y}_1,\mathbf{y}_2)$ in (26). This, together with (31) and the condition (10), yields

$$\begin{aligned}
&(n+m)^{1/2}\{\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y}) - \mathrm{DPCD}(\mathbf{X},\mathbf{Y})\}\\
=\ & \sqrt{\frac{n+m}{n}}\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\{h_{1,0}(\mathbf{X}_i) - \theta\} + \sqrt{\frac{n+m}{m}}\frac{2}{\sqrt{m}}\sum_{i=1}^{m}\{h_{0,1}(\mathbf{Y}_i) - \theta\} + o_p(1)\\
\xrightarrow{d}\ & N\left(0, 4\pi_X^{-1}\sigma_{1,0}^2 + 4\pi_Y^{-1}\sigma_{0,1}^2\right).
\end{aligned}$$

■

Proof of Theorem 6. By Theorem A.1 of Kim et al. (2020), we only need to verify the following conditions:

(i) $h_{2,0}(\mathbf{z}_1,\mathbf{z}_1) = h_{0,2}(\mathbf{z}_1,\mathbf{z}_1)$ and $h_{1,1}(\mathbf{z}_1,\mathbf{z}_1) = -\frac{1}{2}h_{0,2}(\mathbf{z}_1,\mathbf{z}_1)$, where $h_{c,d}(\cdot,\cdot)$ are defined in (27);

(ii) $\mathrm{Var}\{h_{1,0}(\mathbf{X}_1)\} = \mathrm{Var}\{h_{0,1}(\mathbf{Y}_1)\} = 0$ and $\mathrm{Var}\{h_{2,0}(\mathbf{X}_1,\mathbf{X}_2)\} > 0$, $\mathrm{Var}\{h_{0,2}(\mathbf{Y}_1,\mathbf{Y}_2)\} > 0$, $\mathrm{Var}\{h_{1,1}(\mathbf{X}_1,\mathbf{Y}_1)\} > 0$.

These conditions are satisfied as in the proof of Theorem 4. ■

Proof of Proposition 1. Let $\{R^{(b)}\}_{b=1}^{B}$ be uniformly distributed random variables on $\mathcal{G}_{n+m} = \{R^{(1)}, \cdots, R^{((n+m)!)}\}$, which is the set of all permutations among $\{1, \ldots, n+m\}$. For any fixed $R^{(k)} \in \mathcal{G}_{n+m}$, define $E_{bk} = I\{R^{(b)} = R^{(k)}\}$. Then, E_{bk} has a Bernoulli distribution with parameter $(n+m)!$. Thus, we have that

$$\begin{aligned}\frac{1}{B}\sum_{b=1}^{B} I\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_{R^{(b)}}) \le t\} &= \frac{1}{B}\sum_{b=1}^{B}\sum_{k=1}^{(n+m)!} E_{bk} I\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_{R^{(k)}}) \le t\} \\ &= \sum_{k=1}^{(n+m)!} I\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_{R^{(k)}}) \le t\} \frac{1}{B}\sum_{b=1}^{B} E_{bk}.\end{aligned}$$

Then, we obtain

$$\begin{aligned}&\sup_{t\ge 0}\Big|\frac{1}{B}\sum_{b=1}^{B} I\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_{R^{(b)}}) \le t\} - P_R\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \le t \mid \mathcal{D}_{n,m}\}\Big| \\ &\le \sup_{t\ge 0}\Big|\sum_{R\in\mathcal{G}_{n+m}} I\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \le t\}\Big| \sup_{1\le k\le (n+m)!}\Big|\frac{1}{B}\sum_{b=1}^{B} E_{bk} - \frac{1}{(n+m)!}\Big|. \end{aligned} \tag{32}$$

By the bounded differences inequality (Wainwright, 2019, Corollary 2.21), we have that for any $\varepsilon \ge 0$,

$$\begin{aligned}P\Big\{\sup_{1\le k\le (n+m)!}\Big|\frac{1}{B}\sum_{b=1}^{B} E_{bk} - \frac{1}{(n+m)!}\Big| \ge \varepsilon\Big\} &\le 2(n+m)!\exp\Big\{-2B\varepsilon^2\Big\} \\ &= 2\exp\Big\{-2B\varepsilon^2 + \sum_{i=1}^{n+m}\log(i)\Big\},\end{aligned}$$

which implies that

$$\sup_{1\le k\le (n+m)!}\Big|\frac{1}{B}\sum_{b=1}^{B} E_{bk} - \frac{1}{(n+m)!}\Big| \le \varepsilon$$

with probability larger than $1 - 2\exp\{-2B\varepsilon^2 + \sum_{i=1}^{n+m}\log(i)\}$. Given the data $\mathcal{D}_{n,m}$, we have that $\sup_{t\ge 0}|\sum_{R\in\mathcal{G}_{n+m}} I\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \le t\}| \le (n+m)!$ is free of any random permutation sample. These, together with (32) and taking sufficiently large B, complete the proof of the proposition. ■

Proof of Theorem 7. First, we bound the critical value of the permutation test $q_{\alpha,m,n}$ in (13). By Chebyshev's inequality and Lemma 3, we have that

$$P_R\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \ge t \mid \mathcal{D}_{n,m}\} \le \frac{(n+m)^2\mathrm{Var}\{\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \mid \mathcal{D}_{n,m}\}}{t^2} \le \frac{C_0}{t^2}.$$

By the definition of $q_{\alpha,m,n}$, we show that $q_{\alpha,m,n}$ is uniformly bounded by

$$q_{\alpha,m,n} \leq \sqrt{\frac{C_0}{\alpha}}. \tag{33}$$

By Theorem 1(i), model (15) suggests that

$$\mathrm{DPCD}(\mathbf{X},\mathbf{Y}) = (1-\epsilon)\mathrm{DPCD}_{Q_{\mathbf{X}},Q_{\mathbf{Y}}}(\mathbf{X},\mathbf{Y}),$$

where $\mathrm{DPCD}_{Q_{\mathbf{X}},Q_{\mathbf{Y}}}(\mathbf{X},\mathbf{Y})$ denotes the DPCD between $\mathbf{X} \sim Q_{\mathbf{X}}$ and $\mathbf{Y} \sim Q_{\mathbf{Y}}$. Because $Q_{\mathbf{X}} \neq Q_{\mathbf{Y}}$, we have that $\delta \triangleq \mathrm{DPCD}_{Q_{\mathbf{X}},Q_{\mathbf{Y}}}(\mathbf{X},\mathbf{Y}) > 0$. Thus, using $\epsilon = c(n^{-1/2} + m^{-1/2})$, we have that

$$\mathrm{DPCD}(\mathbf{X},\mathbf{Y}) = (1-\epsilon)\delta > \delta/2$$

for sufficiently large n, m. This, together with (33), indicates that there exists sufficiently large N_0 such that for all $n + m \geq N_0$,

$$(n+m)^{-1} q_{\alpha,m,n} < \delta/2 = \mathrm{DPCD}(\mathbf{X},\mathbf{Y})/2. \tag{34}$$

By Lemma 4, we obtain

$$\mathrm{Var}\{\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})|H_1\} = O((m+n)^{-1})$$

uniformly. Note that $E\{\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})\} = \mathrm{DPCD}(\mathbf{X},\mathbf{Y})$. By Chebyshev's inequality and (34), we have that

$$\begin{aligned}
& \lim_{m,n\to\infty} \sup_{H_{n,m}} P\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y}) \leq q_{\alpha,m,n}|H_1\} \\
= & \lim_{m,n\to\infty} \sup_{H_{n,m}} P\Big\{\mathrm{DPCD}(\mathbf{X},\mathbf{Y}) - \widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y}) \geq \mathrm{DPCD}(\mathbf{X},\mathbf{Y}) - (n+m)^{-1} q_{\alpha,m,n}|H_1\Big\} \\
\leq & \lim_{m,n\to\infty} \sup_{H_{n,m}} P\Big\{\mathrm{DPCD}(\mathbf{X},\mathbf{Y}) - \widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y}) \geq \frac{1}{2}\mathrm{DPCD}(\mathbf{X},\mathbf{Y})|H_1\Big\} \\
\leq & \lim_{m,n\to\infty} \sup_{H_{n,m}} \frac{4\mathrm{Var}\{\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})|H_1\}}{\mathrm{DPCD}^2(\mathbf{X},\mathbf{Y})} \\
\leq & \lim_{m,n\to\infty} \frac{4}{(m+n)\delta^2}.
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
\lim_{m,n\to\infty} \inf_{H_{n,m}} P\{\phi_{\mathrm{DPCD}} = 1|H_1\} &= 1 - \lim_{m,n\to\infty} \sup_{H_{n,m}} P\{(n+m)\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y}) \leq q_{\alpha,m,n}|H_1\} \\
&\geq 1 - \lim_{m,n\to\infty} \frac{4}{(m+n)\delta^2} = 1.
\end{aligned}$$

■

Proof of Proposition 2. (i) For any $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^p$, define the Gaussian kernel function as $K_G(\mathbf{v}_1, \mathbf{v}_2) = \exp\{-\frac{\gamma}{2}(\mathbf{v}_1 - \mathbf{v}_2)^T(\mathbf{v}_1 - \mathbf{v}_2)\}$. Note that it holds that

$$
\begin{aligned}
& \int \exp\Big\{ -\frac{\gamma}{2}\alpha^T(\mathbf{z}_i - \mathbf{z}_j)^T(\mathbf{z}_i - \mathbf{z}_j)\alpha\Big\}(2\pi)^{-q/2}\exp\Big\{ -\frac{\alpha^T\alpha}{2}\Big\}d\alpha \\
= \;& (2\pi)^{-q/2}\int \exp\Big\{ -\frac{1}{2}\alpha^T\Big[\gamma(\mathbf{z}_i - \mathbf{z}_j)^T(\mathbf{z}_i - \mathbf{z}_j) + \mathbf{I}_q\Big]\alpha\Big\}d\alpha \\
= \;& \det(\gamma(\mathbf{z}_i - \mathbf{z}_j)^T(\mathbf{z}_i - \mathbf{z}_j) + \mathbf{I}_q)^{-1/2} = \mathrm{K}_{\det}(\mathbf{z}_i, \mathbf{z}_j),
\end{aligned}
$$

for any $\mathbf{z}_i, \mathbf{z}_j \in \mathbb{R}^{p\times q}$.

For any $\mathbf{z}_1, \dots, \mathbf{z}_n \in \mathbb{R}^{p\times q}$ and $c_1, \dots, c_n \in \mathbb{R}$, $(n \geq 2)$, we have that

$$
\begin{aligned}
& \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K_{\det}(\mathbf{z}_i, \mathbf{z}_j) \\
= & \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \int \exp\left\{-\frac{\gamma}{2}(\mathbf{z}_i\alpha - \mathbf{z}_j\alpha)^T(\mathbf{z}_i\alpha - \mathbf{z}_j\alpha)\right\}(2\pi)^{-q/2}\exp\left\{-\frac{\alpha^T\alpha}{2}\right\}d\alpha \qquad (35) \\
= & \int \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j K_G(\mathbf{z}_i\alpha, \mathbf{z}_j\alpha)(2\pi)^{-q/2}\exp\left\{-\frac{\alpha^T\alpha}{2}\right\}d\alpha.
\end{aligned}
$$

Because the Gaussian kernel function $K_G : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ is positive-definite, we have that $\sum_{j=1}^{n} c_i c_j K_G(\mathbf{z}_i\alpha, \mathbf{z}_j\alpha) \geq 0$. This, together with (35), implies that

$$
\sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \mathrm{K}_{\det}(\mathbf{z}_i, \mathbf{z}_j) \geq 0.
$$

(ii) Recall that if η_k $(k = 1, 2, \cdots, q)$ are the eigenvalues of any square matrix $\mathbf{A} \in \mathbb{R}^{q\times q}$, then $(\gamma\eta_k + 1)$ are the eigenvalues of $\gamma\mathbf{A} + \mathbf{I}_q$. Using the fact, we have that

$$
\det(\gamma(\mathbf{z} - \mathbf{z}')^T(\mathbf{z} - \mathbf{z}') + \mathbf{I}_q)^{-1/2} = \prod_{k=1}^{r}(\gamma\lambda_k + 1)^{-1/2},
$$

where λ_k are the nonzero eigenvalues of $(\mathbf{z} - \mathbf{z}')^T(\mathbf{z} - \mathbf{z}')$.

■

Proof of Proposition 3. (i) Note that

$$
\int_{\mathbb{R}^p} \exp\{i\alpha^T\widetilde{\mathbf{z}}\alpha\}dG_1(\alpha) = \det(\mathbf{I}_p - 2i\gamma\widetilde{\mathbf{z}})^{-1/2}, \qquad (36)
$$

where $G_1(\alpha)$ is the p-dimensional distribution of $N_p(0, \gamma\mathbf{I}_p)$ with $\gamma > 0$. Thus, for any $\widetilde{\mathbf{z}}_1, \ldots, \widetilde{\mathbf{z}}_n \in \mathbb{S}_p$ and $c_1, \ldots, c_n \in \mathbb{C}$ ($n \geq 2$), we have that

$$\begin{aligned}
\sum_{k=1}^{n}\sum_{j=1}^{n} c_k \bar{c}_j \widetilde{\mathrm{K}}_{\det}(\widetilde{\mathbf{z}}_k, \widetilde{\mathbf{z}}_j) &= \sum_{k=1}^{n}\sum_{j=1}^{n} c_k \bar{c}_j \int_{\mathbb{R}^p} \exp\{i\alpha^T(\widetilde{\mathbf{z}}_k - \widetilde{\mathbf{z}}_j)\alpha\} dG_1(\alpha) \\
&= \int \sum_{k=1}^{n}\sum_{j=1}^{n} c_k \bar{c}_j \exp\{i\alpha^T(\widetilde{\mathbf{z}}_k - \widetilde{\mathbf{z}}_j)\alpha\} dG_1(\alpha) \\
&= \int \Big(\sum_{k=1}^{n} c_k \exp\{i\alpha^T \widetilde{\mathbf{z}}_k \alpha\}\Big)\Big(\sum_{j=1}^{n} \bar{c}_j \exp\{-i\alpha^T \widetilde{\mathbf{z}}_j \alpha\}\Big) dG_1(\alpha) \\
&= \int \Big\| \sum_{k=1}^{n} c_k \exp\{i\alpha^T \widetilde{\mathbf{z}}_k \alpha\} \Big\|^2 dG_1(\alpha) \geq 0.
\end{aligned}$$

In addition, it is easy to see that $\Big(\widetilde{\mathrm{K}}_{\det}(\widetilde{\mathbf{z}}_k, \widetilde{\mathbf{z}}_j)\Big)$ is a Hermitian matrix.

(ii) By SVD, we have that $\widetilde{\mathbf{z}} - \widetilde{\mathbf{z}}' = \sum_{k=1}^{r} \lambda_k \eta_k \eta_k^T$, where λ_k are the nonzero eigenvalues of $\widetilde{\mathbf{z}} - \widetilde{\mathbf{z}}'$ and $r = \mathrm{rank}(\widetilde{\mathbf{z}} - \widetilde{\mathbf{z}}')$. Thus, we have that

$$\det(\mathbf{I}_p - 2i\gamma(\widetilde{\mathbf{z}} - \widetilde{\mathbf{z}}'))^{-1/2} = \det(\mathbf{I}_p - 2i\gamma\Gamma \mathrm{diag}\{\lambda_1, \cdots, \lambda_r\}\Gamma^T)^{-1/2} = \prod_{k=1}^{r}(1 - 2i\gamma\lambda_k)^{-1/2},$$

where $\Gamma = (\eta_1, \cdots, \eta_r)$. ∎

Appendix B: Lemmas 3 and 4

The following two lemmas are needed to prove Theorem 7.

Lemma 3 *The mean and variance of* $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$ *under permutations are*

$$E_R\{\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \mid \mathcal{D}_{n,m}\} = 0 \quad \textit{and} \quad \mathrm{Var}\{\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \mid \mathcal{D}_{n,m}\} \leq C\left(\frac{1}{m} + \frac{1}{n}\right)^2,$$

where C is a universal constant.

Proof To simplify the notation, we denote $E_R\{\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \mid \mathcal{D}_{n,m}\} = E_R\{\widehat{\mathrm{DPCD}}(\mathbf{Z}_R)\}$. By the definitions of $\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})$ and $h(\cdot, \cdot\,;\ \cdot, \cdot)$, we have that

$$\begin{aligned}
E_R\{\widehat{\mathrm{DPCD}}(\mathbf{Z}_R)\} &= E_R\{h(\mathbf{Z}_{R_1}, \mathbf{Z}_{R_2}; \mathbf{Z}_{R_{n+1}}, \mathbf{Z}_{R_{n+2}})\} \\
&= E_R\{\mathrm{K}_{\det}(\mathbf{Z}_{R_1}, \mathbf{Z}_{R_2}) + \mathrm{K}_{\det}(\mathbf{Z}_{R_{n+1}}, \mathbf{Z}_{R_{n+2}}) - \frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_1}, \mathbf{Z}_{R_{n+1}}) \\
&\quad -\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_1}, \mathbf{Z}_{R_{n+2}}) - \frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_2}, \mathbf{Z}_{R_{n+1}}) - \frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_2}, \mathbf{Z}_{R_{n+2}})\},
\end{aligned}$$

where $R=(R_1,\cdots,R_n,R_{n+1},\cdots R_{n+m})$. Note that

$$
\begin{aligned}
&E_R\{\mathrm{K}_{\det}(\mathbf{Z}_{R_1},\mathbf{Z}_{R_2})-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_1},\mathbf{Z}_{R_{n+1}})-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_1},\mathbf{Z}_{R_{n+2}})|R_1\}=0,\\
&E_R\{\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_{n+1}},\mathbf{Z}_{R_{n+2}})-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_2},\mathbf{Z}_{R_{n+1}})|R_{n+1}\}=0,\\
&E_R\{\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_{n+1}},\mathbf{Z}_{R_{n+2}})-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_2},\mathbf{Z}_{R_{n+2}})|R_{n+2}\}=0.
\end{aligned}
$$

Using these facts, we have that

$$
\begin{aligned}
&E_{R}\{\widehat{\mathrm{DPCD}}(\mathbf{Z}_R)\}\\
=\;&E_R\Big\{E_R\{\mathrm{K}_{\det}(\mathbf{Z}_{R_1},\mathbf{Z}_{R_2})-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_1},\mathbf{Z}_{R_{n+1}})-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_1},\mathbf{Z}_{R_{n+2}})|R_1\}\Big\}\\
&+E_R\Big\{E_R\{\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_{n+1}},\mathbf{Z}_{R_{n+2}})-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_2},\mathbf{Z}_{R_{n+1}})|R_{n+1}\}\Big\}\\
&+E_R\Big\{E_R\{\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_{n+1}},\mathbf{Z}_{R_{n+2}})-\frac{1}{2}\mathrm{K}_{\det}(\mathbf{Z}_{R_2},\mathbf{Z}_{R_{n+2}})|R_{n+2}\}\Big\}=0.
\end{aligned}
$$

Next, we calculate $\mathrm{Var}\{\widehat{\mathrm{DPCD}}(\mathbf{Z}_R)\mid\mathcal{D}_{n,m}\}$. By the definition of $\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})$, we have that

$$
\begin{aligned}
\widehat{\mathrm{DPCD}}(\mathbf{X},\mathbf{Y})^2 =&\frac{1}{n^2(n-1)^2}\frac{1}{m^2(m-1)^2}\sum_{\substack{i_1,i_2=1\\ i_1\neq i_2}}^{n}\sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{m}\sum_{\substack{i_1',i_2'=1\\ i_1'\neq i_2'}}^{n}\sum_{\substack{j_1',j_2'=1\\ j_1'\neq j_2'}}^{m}h(\mathbf{X}_{i_1},\mathbf{X}_{i_2};\mathbf{Y}_{n+j_1},\mathbf{Y}_{n+j_2})\\
&\times h(\mathbf{X}_{i_1'},\mathbf{X}_{i_2'};\mathbf{Y}_{n+j_1'},\mathbf{Y}_{n+j_2'}).
\end{aligned}
$$

To bound $\mathrm{Var}\{\widehat{\mathrm{DPCD}}(\mathbf{Z}_R)\mid\mathcal{D}_{n,m}\}$, we let

$$
\mathcal{I}_2=\{i_1,i_2\}\cap\left\{i_1',i_2'\right\}\ \text{ and }\ \mathcal{J}_2=\{j_1,j_2\}\cap\left\{j_1',j_2'\right\}.
$$

If $\{(|\mathcal{I}_2|,|\mathcal{J}_2|):|\mathcal{I}_2|+|\mathcal{J}_2|\leq 1\}$, where $|\mathcal{I}_2|,|\mathcal{J}_2|$ denote the cardinalities of $\mathcal{I}_2,\mathcal{J}_2$, respectively, then it follows from arguments similar to those for $E_{R}\{\widehat{\mathrm{DPCD}}(\mathbf{Z}_R)\}=0$ that

$$
\mathrm{ER}\triangleq E_R\{h(\mathbf{Z}_{R_{i_1}},\mathbf{Z}_{R_{i_2}};\mathbf{Z}_{R_{n+j_1}},\mathbf{Z}_{R_{n+j_2}})h(\mathbf{Z}_{R_{i_1'}},\mathbf{Z}_{R_{i_2'}};\mathbf{Z}_{R_{n+j_1'}},\mathbf{Z}_{R_{n+j_2'}})\}=0. \tag{37}
$$

When $\{(|\mathcal{I}_2|,|\mathcal{J}_2|):|\mathcal{I}_2|+|\mathcal{J}_2|>1\}$, we need the fact that

$$
|\ h(\mathbf{Z}_{R_{i_1}},\mathbf{Z}_{R_{i_2}};\mathbf{Z}_{R_{n+j_1}},\mathbf{Z}_{R_{n+j_2}})h(\mathbf{Z}_{R_{i_1'}},\mathbf{Z}_{R_{i_2'}};\mathbf{Z}_{R_{n+j_1'}},\mathbf{Z}_{R_{n+j_2'}})\ |\leq 16, \tag{38}
$$

which follows from (26). We can separate $\{(|\mathcal{I}_2|,|\mathcal{J}_2|):|\mathcal{I}_2|+|\mathcal{J}_2|>1\}$ into the following cases.

Case 1 $\{|\mathcal{I}_2|+|\mathcal{J}_2|=2\}$: There are three cases that satisfy $|\mathcal{I}_2|+|\mathcal{J}_2|=2$, i.e. $(|\mathcal{I}_2|,|\mathcal{J}_2|)=(2,0),(1,1)$ or $(0,2)$. In these three cases, by (38), it is easy to see that $|\mathrm{ER}|\leq Cn^{-4}m^{-4}[n^2m^4+n^3m^3+n^4m^2]$.

Case 2 $\{|\mathcal{I}_2|+|\mathcal{J}_2|=3\}$: Two cases are satisfied, i.e. $(|\mathcal{I}_2|,|\mathcal{J}_2|)=(2,1)$ or $(1,2)$, which implies that $|\mathrm{ER}|\leq Cn^{-4}m^{-4}[n^2m^3+n^3m^2]$.

Case 3 $\{|\mathcal{I}_2| + |\mathcal{J}_2| = 4\}$**:** There exists one case, i.e. $(|\mathcal{I}_2|, |\mathcal{J}_2|) = (2, 2)$, for which we can see that $|\mathrm{ER}| \le Cn^{-4}m^{-4}n^2m^2$.

By Cases 1–3, we have that

$$|\mathrm{ER}| \le C\left(\frac{1}{m} + \frac{1}{n}\right)^2 \text{ for } \{(|\mathcal{I}_2|, |\mathcal{J}_2|) : |\mathcal{I}_2| + |\mathcal{J}_2| > 1\}. \tag{39}$$

Using (37) and (39), we obtain

$$\mathrm{Var}\{\widehat{\mathrm{DPCD}}(\mathbf{Z}_R) \mid \mathcal{D}_{n,m}\} = E_R\{\widehat{\mathrm{DPCD}}^2(\mathbf{Z}_R) \mid \mathcal{D}_{n,m}\} \le C\left(\frac{1}{m} + \frac{1}{n}\right)^2.$$

■

Lemma 4 *There exists a universal constant $C > 0$ such that*

$$\mathrm{Var}\{\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})\} \le C\left(\frac{1}{m} + \frac{1}{n}\right).$$

Proof By Theorem 2 in Chapter 2 of Lee (1990), we have that

$$\mathrm{Var}\{\widehat{\mathrm{DPCD}}(\mathbf{X}, \mathbf{Y})\} = \sum_{c=0}^{2}\sum_{d=0}^{2} \frac{\binom{2}{c}\binom{2}{d}\binom{n-2}{2-c}\binom{m-2}{2-d}}{\binom{n}{2}\binom{m}{2}}\sigma_{c,d}^2, \tag{40}$$

where $\sigma_{c,d}^2 = \mathrm{Var}\{h_{c,d}(\mathbf{X}_1, \ldots, \mathbf{X}_c; Y_1, \ldots, Y_d)\}$ with $h_{c,d}(\cdot, \cdot)$ in (27).

By the definition of $h_{1,0}(\mathbf{x}_1)$, we obtain

$$\begin{aligned}
& h_{1,0}(\mathbf{x}_1) - E\{h_{1,0}(\mathbf{X}_1)\} \\
= \;& E\{\mathrm{K}_{\det}(\mathbf{x}_1, \mathbf{X}_2)\} - \frac{1}{2}E\{\mathrm{K}_{\det}(\mathbf{x}_1, \mathbf{Y}_1)\} - \frac{1}{2}E\{\mathrm{K}_{\det}(\mathbf{x}_1, \mathbf{Y}_2)\} \\
& -\Big\{E\{\mathrm{K}_{\det}(\mathbf{X}_1, \mathbf{X}_2)\} - \frac{1}{2}E\{\mathrm{K}_{\det}(\mathbf{X}_1, \mathbf{Y}_1)\} - \frac{1}{2}E\{\mathrm{K}_{\det}(\mathbf{X}_1, \mathbf{Y}_2)\}\Big\}.
\end{aligned}$$

By the fact that $(a + b + c)^2 \le 3(a^2 + b^2 + c^2)$ for any $a, b, c \in \mathbb{R}$, we have that

$$\begin{aligned}
\sigma_{1,0}^2 = \mathrm{Var}\{h_{1,0}(\mathbf{X}_1)\} \;&\le\; 3\mathrm{Var}\{E\{\mathrm{K}_{\det}(\mathbf{X}_1, \mathbf{X}_2)|\mathbf{X}_1\}\} + \frac{3}{4}\mathrm{Var}\{E\{\mathrm{K}_{\det}(\mathbf{X}_1, \mathbf{Y}_1)|\mathbf{X}_1\}\} \\
&\quad + \frac{3}{4}\mathrm{Var}\{E\{\mathrm{K}_{\det}(\mathbf{X}_1, \mathbf{Y}_2)|\mathbf{X}_1\}\} \\
&\le\; 3E\{\mathrm{K}_{\det}^2(\mathbf{X}_1, \mathbf{X}_2)\} + \frac{6}{4}E\{\mathrm{K}_{\det}^2(\mathbf{X}_1, \mathbf{Y}_2)\} \\
&\le\; 9/2.
\end{aligned}$$

The second inequality holds by $(E\{\mathrm{K}_{\det}(\mathbf{X}_1, \mathbf{X}_2)|\mathbf{X}_1\})^2 \le E\{\mathrm{K}_{\det}^2(\mathbf{X}_1, \mathbf{X}_2)|\mathbf{X}_1\}$, and the last inequality holds by $0 \le \mathrm{K}_{\det}^2(\mathbf{x}_1, \mathbf{x}_2) \le 1$.

Similarly, we have $\sigma_{1,0}^2 \leq 9/2$. Moreover, $\sigma_{2,0}^2$, $\sigma_{1,1}^2$, and $\sigma_{0,2}^2$ can be uniformly bounded because of $0 \leq \mathrm{K}_{\det}^2(\mathbf{x}_1, \mathbf{x}_2) \leq 1$. These, together with (40), complete the proof of the lemma.

■

Appendix C: Additional simulation studies

Example C.1 *In this example, we investigate the effect of the dimensions p, q on the power of the DPCD. To this end, we consider the following settings:*

(1) *for fixed $p = 5$, q varies from 1 to 50;*

(2) *p and q (letting $p = q$) vary from 1 to 50.*

The data are generated by $\mathbf{X} \sim \mathbf{T}_{p\times q}(df, \mathbf{0}, \mathbf{U}, \mathbf{V})$ and $\mathbf{Y} \sim \mathbf{T}_{p\times q}(df, \mathbf{M}, \mathbf{U}, \mathbf{V})$ with $(\mathbf{M})_{j,k} = 0.5$, $(\mathbf{U})_{j,k} = 0.5^{|j-k|}$, $(\mathbf{V})_{j,k} = 0.5^{|j-k|}$, and $df = 3$ or 1.

Figure 10 shows plots of the power against p, q. Figure 10(c) and (d) indicate that the power of the DPCP test is increasingly superior as the dimensions increase, whereas the Energy, $\mathrm{MMD}_{\mathrm{gaus}}$, and $\mathrm{MMD}_{\mathrm{lap}}$ appear to be very ineffective in the setting of $df = 1$. Thus, Figure 10 shows that the DPCP is applicable in arbitrary dimensions.

Example C.2 *In this example, we generate $\mathbf{X}$ and $\mathbf{Y}$ in the following nonlinear ways:*

Case 1: $\mathbf{X} = 0.5\sin(\mathbf{Z}_1) + \boldsymbol{\varepsilon}_1$ *and* $\mathbf{Y} = \delta\sin(\mathbf{Z}_2) + \boldsymbol{\varepsilon}_2$;

Case 2: $\mathbf{X} = 0.5\log(|\mathbf{Z}_1| + 0.1) + \boldsymbol{\varepsilon}_1$ *and* $\mathbf{Y} = \delta\log(|\mathbf{Z}_2| + 0.1) + \boldsymbol{\varepsilon}_2$;

Case 3: $\mathbf{X} = |\mathbf{Z}_1 + 0.1|^{0.5} + \boldsymbol{\varepsilon}_1$ *and* $\mathbf{Y} = |\mathbf{Z}_1 + 0.1|^{\delta} + \boldsymbol{\varepsilon}_2$;

Case 4: $\mathbf{X} = \exp\{0.5 + \mathbf{Z}_1 + 4\boldsymbol{\varepsilon}_1\}$ *and* $\mathbf{Y} = \exp\{\delta + \mathbf{Z}_2 + 4\boldsymbol{\varepsilon}_2\}$,

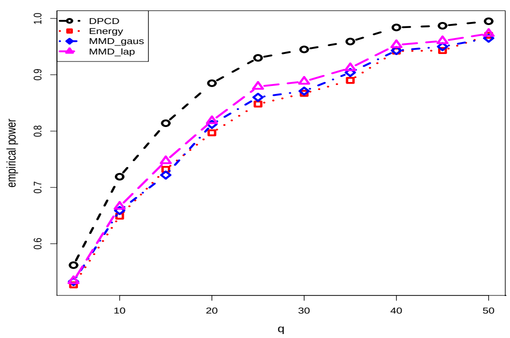
where $\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2} \overset{i.i.d.}{\sim} \boldsymbol{MN}_{p\times q}(\mathbf{0}, \mathbf{I}_p, \mathbf{I}_q)$ and $\mathbf{Z}_1, \mathbf{Z}_2 \overset{i.i.d.}{\sim} \boldsymbol{MN}_{p\times q}(\mathbf{0}, \mathbf{U}, \mathbf{V})$ with $(\mathbf{U})_{j,k} = 0.5^{|j-k|}$ and $(\mathbf{V})_{j,k} = 0.5^{|j-k|}$. The functions in Cases 1–4 are element-wise.

Figure 11 shows plots of the power with respect to δ. The results show that the DPCP outperforms the others in all cases, so it can capture effectively the difference between two general distributions, not just matrix-variate normal and t distributions.

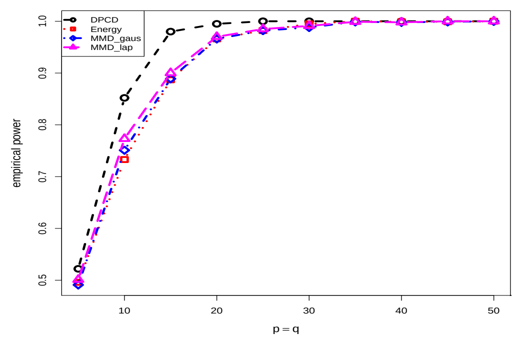
References

Genevera I Allen and Robert Tibshirani. Inference with transposable data: modelling the effects of row and column correlations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(4):721–743, 2012.

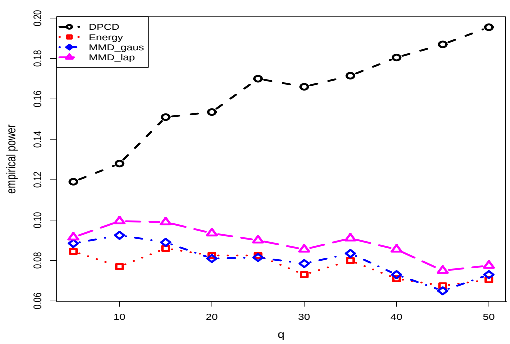
T. W. Anderson. On the distribution of the two-sample Cramér-von Mises criterion. *Annals of Mathematical Statistics*, 33(3):1148–1159, 1962.



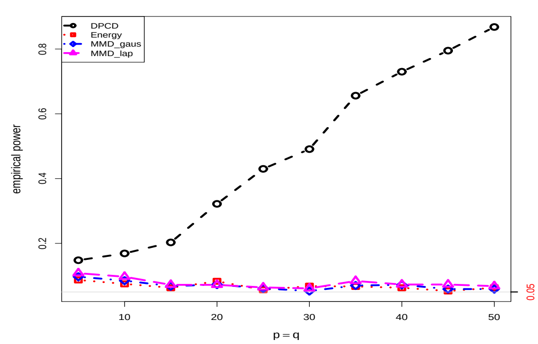
(a) $p = 5$, $df = 3$



(b) $p = q$, $df = 3$

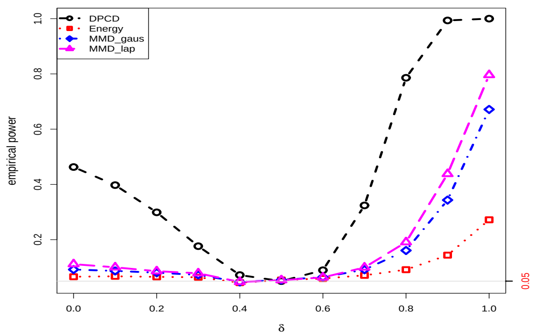


(c) $p = 5$, $df = 1$

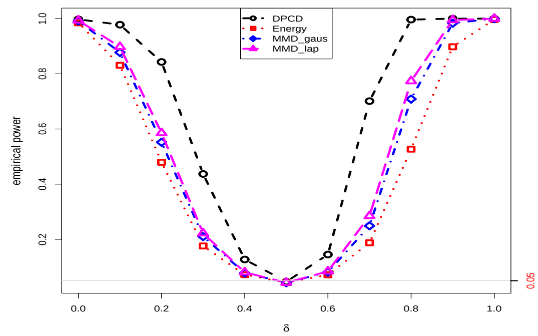


(d) $p = q$, $df = 1$

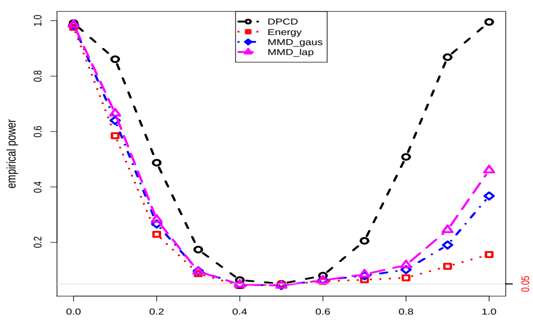
Figure 10: Empirical power at $\alpha = 0.05$ for Example C.1 with $n_1 = n_2 = 20$. (a) and (c): q varies with $p = 5$; (b) and (d): p and q (letting $p = q$) vary.



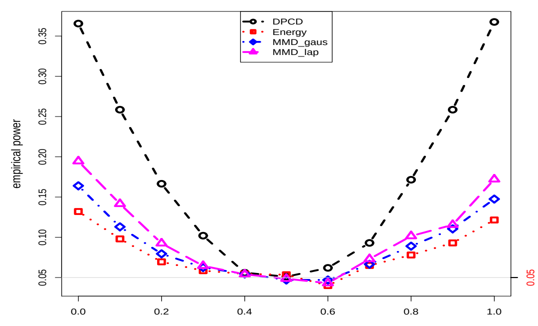
(a) Case 1: $\mathbf{X} = 0.5\sin(\mathbf{Z}_1) + \boldsymbol{\varepsilon}_1$ and $\mathbf{Y} = \delta\sin(\mathbf{Z}_2) + \boldsymbol{\varepsilon}_2$



(b) Case 2: $\mathbf{X} = 0.5\log(|\mathbf{Z}_1| + 0.1) + \boldsymbol{\varepsilon}_1$ and $\mathbf{Y} = \delta\log(|\mathbf{Z}_2| + 0.1) + \boldsymbol{\varepsilon}_2$



(c) Case 3: $\mathbf{X} = |\mathbf{Z}_1 + 0.1|^{0.5} + \boldsymbol{\varepsilon}_1$ and $\mathbf{Y} = |\mathbf{Z}_1 + 0.1|^{\delta} + \boldsymbol{\varepsilon}_2$



(d) Case 4: $\mathbf{X} = \exp\{0.5 + \mathbf{Z}_1 + 4\boldsymbol{\varepsilon}_1\}$ and $\mathbf{Y} = \exp\{\delta + \mathbf{Z}_2 + 4\boldsymbol{\varepsilon}_2\}$

Figure 11: Empirical power at $\alpha = 0.05$ for Example C.2 with $n_1 = n_2 = 20$ and $(p, q) = (25, 15)$.

T. W. Anderson. *An Introduction to Multivariate Statistical Analysis, 3rd Edition.* John Wiley & Sons, 2003.

T. W. Anderson and D. A. Darling. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2):193–212, 1952.

V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis & Applications*, 29(1):328–347, 2007.

L. Baringhaus and C. Franz. On a new multivariate two-sample test. *Journal of Multivariate Analysis*, 88(1):190–206, 2004.

B. V. Bhat. *Theory of U-statistics and its applications.* PhD thesis, Karnatak University, 1995.

S. V. Borisov, A. I. Kaplan, N. L. Gorbachevskaia, and I. A. Kozlova. Analysis of EEG structural synchrony in adolescents suffering from schizophrenic disorders. *Human Physiology*, 31(3):255–261, 2005.

Elynn Y Chen and Jianqing Fan. Statistical inference for high-dimensional matrix-variate factor models. *Journal of the American Statistical Association*, pages 1–18, 2021.

Elynn Y. Chen, Ruey S. Tsay, and Rong Chen. Constrained factor models for high-dimensional matrix-variate time series. *Journal of the American Statistical Association*, 115(530):775–793, 2020.

J Carlos Escanciano. A consistent diagnostic test for regression models using projections. *Econometric Theory*, 22(6):1030–1051, 2006.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(12):723–773, 2012.

A.K. Gupta and D.K Nagar. *Matrix Variate Distributions.* Chapman & Hall/CRC, Boca Raton, 2000.

Ilmun Kim, Sivaraman Balakrishnan, and Larry Wasserman. Robust multivariate nonparametric tests via projection averaging. *Annals of Statistics*, 48(6):3417–3441, 2020.

A. N. Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, (4):83–91, 1933.

Justin Lee. *U-statistics: Theory and Practice.* New York: Marcel Dekker, Inc., 1990.

Zhimei Li and Yaowu Zhang. On a projective ensemble approach to two sample test for equality of distributions. In Hal Daum III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 6020–6027. PMLR, 13–18 Jul 2020.

Kanti V Mardia and Colin R Goodall. Spatial-temporal analysis of multivariate environmental monitoring data. *Multivariate environmental statistics*, 6(76):347–385, 1993.

Charles A. Micchelli. Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constructive Approximation*, 2(1):143–145, 1986.

S. Mori. *Introduction to Diffusion Tensor Imaging*. Springer New York, 2016.

Mark E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

Yang Ning and Han Liu. High-dimensional semiparametric bigraphical models. *Biometrika*, 100(3):655–670, 2013.

R. A. Poldrack, J. A. Mumford, and T. E. Nichols. *Handbook of Functional MRI Data Analysis*. Cambridge University Press, 2011.

Armin Schwartzman. *Random ellipsoids and false discovery rates: statistics for diffusion tensor imaging data*. PhD thesis, Stanford University, 2006.

Armin Schwartzman, Robert Dougherty, and Jonathan Taylor. False discovery rate analysis of brain diffusion direction maps. *Annals of Applied Statistics*, 2(1):153–175, 2008.

N. Smirnov. Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*, 19(2):279–281, 1948.

Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer New York, 2008.

Gábor J Székely and Maria L Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013.

A.W. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes*. New York: Springer, 1996.

Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, 2019. doi: 10.1017/9781108627771.