The Annals of Statistics
2020, Vol. 48, No. 6, 3417–3441
https://doi.org/10.1214/19-AOS1936

ROBUST MULTIVARIATE NONPARAMETRIC TESTS VIA PROJECTION AVERAGING

BY ILMUN KIM*, SIVARAMAN BALAKRISHNAN† AND LARRY WASSERMAN‡

Department of Statistics & Data Science, Carnegie Mellon University, *ilmunk@stat.cmu.edu; †siva@stat.cmu.edu; ‡larry@stat.cmu.edu

In this work, we generalize the Cramér–von Mises statistic via projection averaging to obtain a robust test for the multivariate two-sample problem. The proposed test is consistent against all fixed alternatives, robust to heavy-tailed data and minimax rate optimal against a certain class of alternatives. Our test statistic is completely free of tuning parameters and is computationally efficient even in high dimensions. When the dimension tends to infinity, the proposed test is shown to have comparable power to the existing high-dimensional mean tests under certain location models. As a by-product of our approach, we introduce a new metric called *the angular distance* which can be thought of as a robust alternative to the Euclidean distance. Using the angular distance, we connect the proposed method to the reproducing kernel Hilbert space approach. In addition to the Cramér–von Mises statistic, we demonstrate that the projection-averaging technique can be used to define robust multivariate tests in many other problems.

1. Introduction. Let X and Y be random vectors defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with distributions P_X and P_Y, respectively. Given two mutually independent samples $\mathcal{X}_m = \{X_1, \dots, X_m\}$ and $\mathcal{Y}_n = \{Y_1, \dots, Y_n\}$ from P_X and P_Y, we want to test

$$H_0 : P_X = P_Y \quad \text{versus} \quad H_1 : P_X \neq P_Y. \tag{1.1}$$

This fundamental problem has received considerable attention in statistics with a wide range of applications (see, e.g., Thas (2010), for a review). A common statistic for the univariate two-sample testing is the Cramér–von Mises (CvM) statistic (Anderson (1962)):

$$\frac{mn}{m+n} \int_{\infty}^{\infty} \big(\widehat{F}_X(t) - \widehat{F}_Y(t)\big)^2 \, d\widehat{H}(t),$$

where $\widehat{F}_X(t)$ and $\widehat{F}_Y(t)$ are the empirical distribution functions of $\mathcal{X}_m$ and $\mathcal{Y}_n$, respectively, and $(m+n)\widehat{H}(t) = m\widehat{F}_X(t) + n\widehat{F}_Y(t)$. Another approach is based on the energy statistic, which is an estimate of the squared energy distance (Székely and Rizzo (2013)):

$$E^2 = 2\mathbb{E}\big[|X_1 - Y_1|\big] - \mathbb{E}\big[|X_1 - X_2|\big] - \mathbb{E}\big[|Y_1 - Y_2|\big],$$

where $|x|$ is the absolute value of $x \in \mathbb{R}$. The energy distance is well defined assuming a finite first moment and it can be written in a form that is similar to Cramér's distance (Cramér (1928)), namely,

$$E^2 = 2\int_{-\infty}^{\infty} \big(F_X(t) - F_Y(t)\big)^2 \, dt,$$

where $F_X(t)$ and $F_Y(t)$ are the distribution functions of X and Y, respectively.

Received May 2019; revised December 2019.
MSC2020 subject classifications. Primary 62G10, 62H15, 62H20; secondary 62G35.
Key words and phrases. Energy statistic, high dimension and low sample size, independence testing, maximum mean discrepancy, permutation tests, U-statistic.

The CvM-statistic has several advantages over the energy statistic for univariate two-sample testing. For instance, the CvM-statistic is distribution-free under H_0 (Anderson (1962)) and its population counterpart is well defined without any moment assumptions. It also has an intuitive probabilistic interpretation in terms of probabilities of concordance and discordance of four independent random variables (Baringhaus and Henze (2017)). Nevertheless, the CvM-statistic has rarely been studied for multivariate testing. A primary reason is that the CvM-statistic is essentially rank-based, which leads to a challenge to generalize it in a multivariate space. In contrast, the energy statistic can be easily applied in arbitrary dimensions as in Baringhaus and Franz (2004) and Székely and Rizzo (2004). Specifically, they defined the squared multivariate energy distance by

$$E_d^2(P_X, P_Y) = 2\mathbb{E}[\|X_1 - Y_1\|] - \mathbb{E}[\|X_1 - X_2\|] - \mathbb{E}[\|Y_1 - Y_2\|], \tag{1.2}$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$. The multivariate energy distance maintains the characteristic property that it is always nonnegative and equal to zero if and only if $P_X = P_Y$. It can also be viewed as the average of univariate Cramér's distances of projected random variables (Baringhaus and Franz (2004)):

$$E_d^2(P_X, P_Y) = \frac{\sqrt{\pi}(d-1)\Gamma(\frac{d-1}{2})}{\Gamma(\frac{d}{2})} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \big(F_{\beta^\top X}(t) - F_{\beta^\top Y}(t)\big)^2 \, dt \, d\lambda(\beta), \tag{1.3}$$

where λ represents the uniform probability measure on the d-dimensional unit sphere $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$, $\Gamma(\cdot)$ is the gamma function and the symbol $\top$ stands for the transpose operation.

Although the multivariate energy distance can be easily estimated in any dimension, it still requires the finite moment assumption as in the univariate case. When the underlying distributions violate this moment condition with potential outliers, the energy statistic becomes extremely unstable and the resulting test might perform poorly. Given that outlying observations arise frequently in practice with high-dimensional data, there is a need to develop a robust counterpart of the energy distance. The primary goal of this work is to introduce a robust, tuning parameter-free, two-sample testing procedure that is easily applicable in arbitrary dimensions and consistent against all fixed alternatives. Specifically, we modify the univariate CvM-statistic to generalize it to an arbitrary dimension by averaging over all one-dimensional projections. In detail, the proposed test statistic is an unbiased estimate of the squared multivariate CvM-distance defined as follows:

$$W_d^2(P_X, P_Y) = \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}} \big(F_{\beta^\top X}(t) - F_{\beta^\top Y}(t)\big)^2 \, dH_\beta(t) \, d\lambda(\beta), \tag{1.4}$$

where $H_\beta(t) = \vartheta_X F_{\beta^\top X}(t) + \vartheta_Y F_{\beta^\top Y}(t)$ and ϑ_X is a fixed value in $(0, 1)$ and $\vartheta_Y = 1 - \vartheta_X$. For simplicity and when there is no ambiguity, we may omit the dependency on P_X, P_Y and write $W_d(P_X, P_Y)$ as W_d.

Throughout this paper, we refer to the process of averaging over all projections as *projection averaging*.

1.1. *Summary of our results.*

The proposed multivariate CvM-distance shares some appealing properties of the energy distance while being robust to heavy-tailed distributions or outliers. For example, W_d is invariant to orthogonal transformations and satisfies the characteristic property (Lemma 2.1), meaning that W_d is nonnegative and equal to zero if and only if $P_X = P_Y$. More importantly, it is straightforward to estimate W_d without using any tuning parameters (Theorem 2.1). Based on an unbiased estimate of W_d^2, we apply the permutation test procedure to determine a critical value of the test statistic. Although the permutation approach has been standard in practical implementations of two-sample testing, its theoretical properties have been less explored beyond simple cases (e.g., Pesarin (2001)). Indeed, previous studies usually consider asymptotic tests in their theory section whereas their actual tests are calibrated via permutations. We bridge the gap between theory and practice by presenting both theoretical and empirical results on the permutation test under various scenarios. Our main results regarding the CvM-distance are summarized as follows:

- *Closed-form expression* (Section 2): Building on Escanciano (2006) and Zhu et al. (2017), we show that the test statistic has a simple closed-form expression.
- *Asymptotic power* (Section 2): We prove that the permutation test based on the proposed statistic has the same asymptotic power as the oracle and asymptotic tests that assume knowledge of the underlying distributions (Section 2.2) against fixed and contiguous alternatives.
- *Robustness* (Section 3): We show that the permutation test based on the proposed statistic maintains good power in the contamination model, while the energy test becomes completely powerless in this setting.
- *Minimax optimality* (Section 4): We analyze the finite-sample power of the proposed permutation test and prove its minimax rate optimality against a class of alternatives that differ from the null in terms of the CvM-distance. We also show that the energy test is not optimal in our context.
- *HDLSS behavior* (Section 5): We consider a *high-dimension, low-sample size* (HDLSS) regime where the dimension tends to infinity while the sample size is fixed. Under this regime, we identify sufficient conditions under which the power of the proposed test converges to one. In addition, we show that the proposed test has comparable power to the high-dimensional mean tests introduced by Chen and Qin (2010) and Chakraborty and Chaudhuri (2017) under certain location models.
- *Angular distance* (Section 6): We introduce the angular distance between two vectors and use this to show that the multivariate CvM-distance is a special case of the generalized energy distance (Sejdinovic et al. (2013)). Furthermore, the CvM-distance is the maximum mean discrepancy (Gretton et al. (2012)) associated with the angular distance.

Beyond the CvM-statistic, the projection-averaging technique can be widely applicable to other nonparametric statistics. In the second part of this study, we revisit some famous univariate sign- or rank-based statistics and propose their multivariate counterparts via projection averaging. Although there has been much effort to extend univariate sign- or rank-based statistics in a multivariate space (see, e.g., Hettmansperger, Möttönen and Oja (1998), Liu (2006), Oja (2010), Oja and Randles (2004)), they are either computationally expensive to implement or less intuitive to understand. Our projection-averaging approach addresses these issues by providing a tractable calculation form of statistics and by having a direct interpretation in terms of projections. In Section 7 and also Appendix D.8, we demonstrate the generality of the projection-averaging approach by presenting multivariate extensions of several existing univariate statistics.

1.2. *Literature review.* There are a number of multivariate two-sample testing procedures available in the literature. We list some fundamental methods and recent developments. Anderson, Hall and Titterington (1994) proposed the two-sample statistic based on the integrated square distance between two kernel density estimates. The energy statistic was introduced by Baringhaus and Franz (2004) and Székely and Rizzo (2004) independently. Biswas and Ghosh (2014) modified the energy statistic to improve the performance of the previous test for the high-dimensional location-scale and scale problems. Gretton et al. (2012) introduced a class of distances between two probability distributions, called the maximum mean discrepancy (MMD), based on a reproducing kernel Hilbert approach. Sejdinovic et al.

(2013) showed that the energy distance is a special case of the MMD associated with the kernel induced by the Euclidean distance. Recently, Pan et al. (2018) proposed a new metric, named the ball divergence, between two probability distributions and connected it to the MMD. A further review of kernel-based two-sample tests can be found in Harchaoui et al. (2013).

Another line of work is based on graph constructions. Schilling (1986) and Henze (1988) introduced a multivariate two-sample test based on the k nearest neighbor (NN) graph. Mondal, Biswas and Ghosh (2015) pointed out that the previous NN test may suffer from low power for the high-dimensional location-scale problem and provided an alternative that addresses this limitation. Another variant of the NN test, which is tailored to imbalanced samples, can be found in Chen, Dou and Qiao (2013). Friedman and Rafsky (1979) considered minimum spanning tree (MST) to present a generalization of the univariate run test in Wald and Wolfowitz (1940). The MST test proposed by Friedman and Rafsky (1979) has recently been modified by Chen and Friedman (2017) and Chen, Chen and Su (2018) to improve power under scale alternatives and imbalanced samples, respectively. Rosenbaum (2005) proposed a distribution-free test in finite samples based on cross-matches. More recently, Biswas, Mukhopadhyay and Ghosh (2014) introduced another distribution-free test based on the shortest Hamiltonian path. A general theoretical framework for graph-based tests has been established by Bhattacharya (2018), Bhattacharya (2019). Other recent developments include Liu and Modarres (2011), Kanamori, Suzuki and Sugiyama (2012), Bera, Ghosh and Xiao (2013), Lopez-Paz and Oquab (2016), Zhou, Zheng and Zhang (2017), Mukhopadhyay and Wang (2018), among others.

The projection-averaging approach to CvM-type statistics can be found in other statistical problems. For example, Zhu, Fang and Bhatti (1997) and Cui (2002) considered the CvM-statistic using projection averaging to investigate one-sample goodness-of-fit tests for multivariate distributions. Escanciano (2006) proposed the CvM-based goodness-of-fit test for parametric regression models. To the best of our knowledge, however, this is the first study that investigates the CvM-statistic for the multivariate two-sample problem via projection averaging.

Our technique to obtain a closed-form expression for projection-averaging statistics is based on Escanciano (2006). The same principle has been exploited by Zhu et al. (2017) in the context of testing for multivariate independence. We further extend the result of Escanciano (2006) to more general cases and provide an alternative proof using orthant probabilities for normal distributions.

Outline. The rest of this paper is organized as follows. In Section 2, we introduce our test statistic and the permutation test procedure. We then study their limiting behaviors under the conventional fixed dimension asymptotic framework. In Section 3, we compare the power of the CvM test with that of the energy test and highlight the robustness of the CvM test. Section 4 establishes minimax rate optimality of the proposed test against a certain class of alternatives associated with the CvM-distance. In Section 5, we study the asymptotic power of the CvM test in the HDLSS setting. We introduce the angular distance between two vectors in Section 6 to show that the CvM-distance is the generalized energy distance built on the introduced metric. In Section 7, the projection-averaging technique is applied to other sign- or rank-based statistics and this allows us to provide new multivariate extensions. Simulation results are reported in Section 8 to demonstrate the competitive power performance of the proposed approach with finite sample size. All proofs of the main results are deferred to the Supplementary Material (Kim, Balakrishnan and Wasserman (2020)).

Notation. For two nonzero vectors $U_1, U_2 \in \mathbb{R}^d$, we denote the angle between U_1 and U_2 by $\mathsf{Ang}(U_1, U_2) = \arccos\{U_1^\top U_2/(\|U_1\|\|U_2\|)\}$. For $1 \le q \le p$, we let $(p)_q = p(p-1)\cdots(p-$

$q+1)$. Let $\mathbb{P}_0$ and $\mathbb{P}_1$ be the probability measures under H_0 and H_1, respectively. Similarly, $\mathbb{E}_0$ and $\mathbb{E}_1$ stand for the expectations with respect to $\mathbb{P}_0$ and $\mathbb{P}_1$. For any two real sequences $\{a_n\}$ and $\{b_n\}$, we use $a_n \asymp b_n$ if there exist constants $C, C' > 0$ such that $C < |a_n/b_n| < C'$ for each n. We write $a_n = O(b_n)$ if there exists $C > 0$ such that $|a_n| \le C|b_n|$ for each n. We also write $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$. For a sequence of random variables X_n, we use the notation $X_n = O_{\mathbb{P}}(a_n)$ when X_n is bounded in probability (tight). The acronym *i.i.d.* stands for independent and identically distributed and we use the symbol $X_1,\dots,X_n \overset{i.i.d.}{\sim} P$ to represent that $X_1,\dots,X_n$ are *i.i.d.* samples from distribution P. We denote the $d\times d$ identity matrix by I_d. The symbol $\mathbb{1}(\cdot)$ is used for indicator functions. We write summation over the set of all k-tuples drawn without replacement from $\{1,\dots,n\}$ by $\sum_{i_1,\dots,i_k=1}^{n,\neq}$. Throughout this paper, we assume that all vectors are column vectors and $m,n\ge 2$.

2. Projection averaging-type Cramér–von Mises statistics. In this section, we start with the basic properties of the CvM-distance. We then introduce our test statistic and study its limiting behavior. We end this section with a description of the permutation test and its large sample properties. Throughout this section, we consider the conventional asymptotic regime where the dimension is fixed and

$$\frac{m}{m+n} \to \vartheta_X \in (0,1) \quad \text{and} \quad \frac{n}{m+n} \to \vartheta_Y \in (0,1) \quad \text{as } N = m+n \to \infty. \tag{2.1}$$

Let us first establish the characteristic property of the CvM-distance.

LEMMA 2.1 (Characteristic property). *W_d is nonnegative and has the characteristic property*:

$$W_d(P_X, P_Y) = 0 \quad \textit{if and only if} \quad P_X = P_Y.$$

Note that W_d involves integration over the unit sphere. One way to approximate this integral is to consider a subset of $\mathbb{S}^{d-1}$, namely $\{\beta_1,\dots,\beta_k\}$, and then to take the sample mean over k different univariate CvM-statistics (see, e.g., Zhu, Fang and Bhatti (1997)). However, this approach has an unpleasant trade-off between accuracy and computational time depending on the choice of k. The problem becomes even worse in high dimensions where one may need exponentially many projections to achieve a certain accuracy. Our approach, on the other hand, does not suffer from this computational issue by explicitly calculating the integral over $\mathbb{S}^{d-1}$. The explicit form of the integration is mainly due to Escanciano (2006) who provided the following lemma.

LEMMA 2.2 (Escanciano (2006)). *For any nonzero vectors $U_1, U_2 \in \mathbb{R}^d$,*

$$\int_{\mathbb{S}^{d-1}} \mathbb{1}(\beta^\top U_1 \le 0)\mathbb{1}(\beta^\top U_2 \le 0)\, d\lambda(\beta) = \frac{1}{2} - \frac{1}{2\pi}\mathsf{Ang}(U_1, U_2).$$

REMARK 2.1. Escanciano (2006) proved Lemma 2.2 using the volume of a spherical wedge. In the Supplementary Material (Appendix C.2), we provide an alternative proof of this result based on orthant probabilities for normal distributions. We also extend this result to integration involving strictly more than two indicator functions in the Supplementary Material (Lemma B.8 and Lemma D.1). This extension allows us to generalize the univariate τ^* (Bergsma and Dassios (2014)) in Theorem 7.1 and potentially many other univariate statistics (see Appendix D.8).

Based on Lemma 2.2, we give another representation of W_d^2 in terms of the expected angle involving three independent random vectors. Here and hereafter, we assume that

$$\beta^\top X \textit{ and } \beta^\top Y \textit{ have continuous distribution functions for } \lambda\textit{-almost all } \beta \in \mathbb{S}^{d-1}.$$

This continuity assumption greatly simplifies the alternative expression for W_d^2 and avoids the possibility that $\mathsf{Ang}(\cdot,\cdot)$ is not well defined when one of the inputs is a zero vector. This issue may be handled by defining $\mathsf{Ang}(\cdot,\cdot)$ differently for those exceptional cases, but we do not pursue this direction here.

THEOREM 2.1 (Closed-form expression). *Suppose that* $X_1, X_2 \overset{i.i.d.}{\sim} P_X$ *and, independently,* $Y_1, Y_2 \overset{i.i.d.}{\sim} P_Y$. *Then the squared multivariate CvM-distance can be written as*

$$W_d^2(P_X, P_Y) = \frac{1}{3} - \frac{1}{2\pi}\mathbb{E}\big[\mathsf{Ang}(X_1 - Y_1, X_2 - Y_1)\big] - \frac{1}{2\pi}\mathbb{E}\big[\mathsf{Ang}(Y_1 - X_1, Y_2 - X_1)\big].$$

The above result highlights that $W_d(P_X, P_Y)$ is invariant to the choice of ϑ_X and ϑ_Y under the continuity assumption. In the next subsection, we introduce the test statistic and study its limiting behavior.

2.1. *Test statistic and limiting distributions.* Theorem 2.1 leads to a natural empirical estimate of W_d^2 based on a U-statistic. Consider the kernel of order two:

$$h_{\mathrm{CvM}}(x_1, x_2; y_1, y_2) = \frac{1}{3} - \frac{1}{2\pi}\mathsf{Ang}(x_1 - y_1, x_2 - y_1) - \frac{1}{2\pi}\mathsf{Ang}(y_1 - x_1, y_2 - x_1). \tag{2.2}$$

We denote its symmetrized version, which is invariant to the order of the first two arguments as well as the last two arguments, by

$$\widetilde{h}_{\mathrm{CvM}}(x_1, x_2; y_1, y_2) = \frac{1}{2}h_{\mathrm{CvM}}(x_1, x_2; y_1, y_2) + \frac{1}{2}h_{\mathrm{CvM}}(x_2, x_1; y_2, y_1).$$

Then our test statistic is defined as follows:

$$U_{\mathrm{CvM}} = \binom{m}{2}^{-1}\binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le m} \sum_{1 \le j_1 < j_2 \le n} \widetilde{h}_{\mathrm{CvM}}(X_{i_1}, X_{i_2}; Y_{j_1}, Y_{j_2}). \tag{2.3}$$

Leveraging the basic theory of U-statistics (e.g., Lee (1990)), it is clear that U_{CvM} is an unbiased estimator of W_d^2. Additionally, U_{CvM} is a degenerate U-statistic under the null hypothesis as we prove in the Supplementary Material (Appendix C.5). Hence we can apply the asymptotic theory for a degenerate two-sample U-statistic (Chapter 3 of Bhat (1995)) to obtain the following result.

THEOREM 2.2 (Asymptotic null distribution of U_{CvM}). *For each* $k = 1, 2, \ldots$, *let* λ_k *be the eigenvalue with the corresponding eigenfunction* ϕ_k *satisfying the integral equation*

$$\mathbb{E}\big[\mathbb{E}\{\widetilde{h}_{\mathrm{CvM}}(x_1, X_2; Y_1, Y_2) \mid X_2\}\phi_k(X_2)\big] = \lambda_k \phi_k(x_1). \tag{2.4}$$

Then U_{CvM} *has the limiting null distribution under the limiting regime* (2.1) *given by*

$$N U_{\mathrm{CvM}} \xrightarrow{d} \vartheta_X^{-1}\vartheta_Y^{-1}\sum_{k=1}^{\infty}\lambda_k(\xi_k^2 - 1),$$

where $\xi_k \overset{i.i.d.}{\sim} N(0,1)$ *and* $\xrightarrow{d}$ *stands for convergence in distribution.*

Under a fixed alternative hypothesis where P_X and P_Y do not change with m and n, the proposed test statistic converges weakly to a normal distribution. We build on Hoeffding's decomposition of a two-sample U-statistic (e.g., page 40 of Lee (1990)) to prove the following result.

THEOREM 2.3 (Asymptotic distribution of U_{CvM} under fixed alternatives). *Let us define*

$$\sigma_{h_X}^2 = \mathbb{V}[\mathbb{E}\{\widetilde{h}_{\mathrm{CvM}}(X_1, X_2; Y_1, Y_2) \mid X_1\}] \quad \textit{and}$$
$$\sigma_{h_Y}^2 = \mathbb{V}[\mathbb{E}\{\widetilde{h}_{\mathrm{CvM}}(X_1, X_2; Y_1, Y_2) \mid Y_1\}],$$

where $\mathbb{V}(\cdot)$ is the variance operator. Then under the limiting regime (2.1) and fixed alternative $P_X \neq P_Y$, we have

$$\sqrt{N}(U_{\mathrm{CvM}} - W_d^2) \xrightarrow{d} N(0, 4\vartheta_X^{-1}\sigma_{h_X}^2 + 4\vartheta_Y^{-1}\sigma_{h_Y}^2).$$

From the previous two theorems, it is clear to see that NU_{CvM} is stochastically bounded under the null hypothesis whereas it diverges to infinity under fixed alternatives. Thus one can expect that any reasonable test based on the proposed test statistic is consistent (meaning that the power converges to one as $N \to \infty$) against all fixed alternatives. In fact, the problem of distinguishing two fixed distributions is too easy in large sample situations and many of nonparametric tests are known to be consistent in this asymptotic regime. We therefore turn now to a more challenging scenario where a distance between P_X and P_Y diminishes as the sample size increases. To this end, we make a standard assumption that the underlying distributions belong to quadratic mean differentiable (QMD) families (e.g., Bhattacharya (2019)).

DEFINITION 2.1 (Quadratic mean differentiable families, page 484 of Lehmann and Romano (2005)). Let $\{P_\theta, \theta \in \Omega\}$ be a family of probability distributions on $(\mathbb{R}^d, \mathcal{B})$ where $\mathcal{B}$ is the Borel σ-field associated with $\mathbb{R}^d$ and Ω is an open subset of $\mathbb{R}^p$. Assume each P_θ is absolutely continuous with respect to Lebesgue measure and set $p_\theta(t) = dP_\theta(t)/dt$. The family $\{P_\theta, \theta \in \Omega\}$ is quadratic mean differentiable at θ_0 if there exists a vector of real-valued functions $\eta(\cdot, \theta_0) = (\eta_1(\cdot, \theta_0), \dots, \eta_p(\cdot, \theta_0))^\top$ such that

$$\int_{\mathbb{R}^d} \left[\sqrt{p_{\theta_0 + b}(t)} - \sqrt{p_{\theta_0}(t)} - b^\top \eta(t, \theta_0)\right]^2 dt = o(\|b\|^2) \quad \text{as } \|b\| \to 0.$$

The QMD families include a broad class of parametric distributions such as exponential families in natural form. By focusing on the QMD families, we are particularly interested in asymptotically nondegenerate situations where the limiting sum of the type I and type II errors of the optimal test is nontrivial, that is, bounded by the nominal level α and one. It has been shown that when P_{θ_0} and P_{θ_N} belong to the QMD families, this nondegenerate situation occurs when $\|\theta_0 - \theta_N\| \asymp N^{-1/2}$ (Chapter 13.1 of Lehmann and Romano (2005)). Hence we consider a sequence of contiguous alternatives where $\theta_N = \theta_0 + bN^{-1/2}$ for some $b \in \mathbb{R}^p$ and establish the asymptotic behavior of U_{CvM} under the given scenario. Our result builds on the prior work by Chikkagoudar and Bhat (2014) and extends it to multivariate cases.

THEOREM 2.4 (Asymptotic distribution of U_{CvM} under contiguous alternatives). *Assume $\{P_\theta, \theta \in \Omega\}$ is quadratic mean differentiable at θ_0 with derivative $\eta(\cdot, \theta_0)$ and Ω is an open subset of $\mathbb{R}^p$. Define the Fisher information matrix to be the matrix $I(\theta)$ with (i, j) entry*

$$I_{i,j}(\theta) = 4\int_{\mathbb{R}^d} \eta_i(t, \theta)\eta_j(t, \theta)\, dt,$$

and assume that $I(\theta_0)$ *is nonsingular. Suppose we observe* $\mathcal{X}_m \overset{i.i.d.}{\sim} P_{\theta_0}$ *and* $\mathcal{Y}_n \overset{i.i.d.}{\sim} P_{\theta_0+bN^{-1/2}}$ *for* $b \in \mathbb{R}^p$. *Then under the limiting regime* (2.1),

$$NU_{\mathrm{CvM}} \overset{d}{\longrightarrow} \vartheta_X^{-1}\vartheta_Y^{-1}\sum_{k=1}^{\infty}\lambda_k\{(\xi_k + \vartheta_X^{1/2}a_k)^2 - 1\},$$

where $a_k = \int_{\mathbb{R}^d} 2\{b^\top \eta(x,\theta_0)\}p_{\theta_0}^{-1/2}(x)\phi_k(x)\,dP_{\theta_0}(x)$.

The above theorem implies that if there exists $k \geq 1$ such that $a_k \neq 0$ and $\lambda_k > 0$, a test based on U_{CvM} can have asymptotic power greater than α (see page 615 of Lehmann and Romano (2005)). This is in contrast to the NN test which has a slower consistency rate given by $N^{-1/4}$ when $d \leq 8$ under some regularity conditions (see Bhattacharya (2018)). In the Supplementary Material, we consider low-dimensional Gaussian location models and illustrate that the proposed CvM test dominates the NN test via simulations. In fact, the former tends to have very close power to Hotelling's T^2 test, which is known to be optimal under the low-dimensional Gaussian scenarios (Anderson (2003)).

2.2. *Critical value and permutation test.* So far, we have investigated the limiting behaviors of the test statistic under the null and (fixed and contiguous) alternative hypotheses. It is important to note that the performance of a test depends not only on its test statistic but also crucially on its critical value. A common approach to determining the critical value is based on the limiting null distribution of the test statistic. Since we are dealing with a general composite null, one can define this limiting null distribution in various ways. Two natural candidates are described as follows:

- $P_{\mathrm{CvM}}^{(\mathrm{single})}$: the limiting distribution of NU_{CvM} based on i.i.d. samples from the single distribution P_X.
- $P_{\mathrm{CvM}}^{(\mathrm{mix})}$: the limiting distribution of NU_{CvM} based on i.i.d. samples from the mixture distribution $\vartheta_X P_X + \vartheta_Y P_Y$.

These two limiting distributions $P_{\mathrm{CvM}}^{(\mathrm{single})}$ and $P_{\mathrm{CvM}}^{(\mathrm{mix})}$ coincide when $P_X = P_Y$ but they are different in general if $P_X \neq P_Y$. By invoking Theorem 2.2, we can conclude that $P_{\mathrm{CvM}}^{(\mathrm{single})}$ and $P_{\mathrm{CvM}}^{(\mathrm{mix})}$ are Gaussian chaos distributions in the low-dimensional setting. The asymptotic tests then reject the null hypothesis when NU_{CvM} is greater than the upper $1-\alpha$ quantile of $P_{\mathrm{CvM}}^{(\mathrm{single})}$ or $P_{\mathrm{CvM}}^{(\mathrm{mix})}$, denoted by $q_{\alpha,\mathrm{CvM}}^{(\mathrm{single})}$ and $q_{\alpha,\mathrm{CvM}}^{(\mathrm{mix})}$, respectively.

Unfortunately, this asymptotic approach is infeasible as the limiting distributions involve quantities that depend on the underlying distributions and that cannot be easily estimated. Even if either $P_{\mathrm{CvM}}^{(\mathrm{single})}$ or $P_{\mathrm{CvM}}^{(\mathrm{mix})}$ is known exactly, the resulting asymptotic test does not have finite-sample guarantees on the type I error control. For this reason, we advocate for using the permutation procedure that resolves the issues of the asymptotic approach. More importantly, as shown in Theorem 2.6, the power of the permutation test is asymptotically the same as that of the asymptotic tests under the conventional asymptotic regime.

Before we describe the permutation procedure, let us introduce the oracle test that serves as a benchmark for the permutation test. Let $T_{m,n}$ be a generic two-sample test statistic. Then the critical value of the oracle test based on $T_{m,n}$ can be determined as follows:

- *Oracle test.*

1. Consider new *i.i.d.* samples $\{\widetilde{Z}_1,\dots,\widetilde{Z}_N\}$ from the mixture $\vartheta_X P_X + \vartheta_Y P_Y$.
2. Let $T_{m,n}(\widetilde{Z})$ be the test statistic of interest calculated based on $\widetilde{\mathcal{X}}_m = \{\widetilde{Z}_1,\dots,\widetilde{Z}_m\}$ and $\widetilde{\mathcal{Y}}_n = \{\widetilde{Z}_{m+1},\dots,\widetilde{Z}_N\}$.

3. Given a significance level $0 < \alpha < 1$, return the critical value $c^*_{\alpha,m,n}$ defined by

$$c^*_{\alpha,m,n} := \inf\{t \in \mathbb{R} : 1 - \alpha \le \mathbb{P}(T_{m,n}(\widetilde{Z}) \le t)\}. \tag{2.5}$$

It is worth pointing out that the oracle statistic $T_{m,n}(\widetilde{Z})$ has the same distribution as the test statistic based on the original samples under H_0, but not necessarily under H_1. Hence the oracle test based on $c^*_{\alpha,m,n}$ is exact under H_0 and can be powerful under H_1. However, $c^*_{\alpha,m,n}$ relies on the unknown mixture distribution $\vartheta_X P_X + \vartheta_Y P_Y$, which makes the oracle test impractical. In sharp contrast, the critical value of the permutation test can be obtained without knowledge of the mixture distribution as follows:

- *Permutation test.*

1. Let $\{Z_1, \dots, Z_N\} = \{X_1, \dots, X_m, Y_1, \dots, Y_n\}$ be the pooled samples and $Z_\varpi = \{Z_{\varpi(1)}, \dots, Z_{\varpi(N)}\}$ where $\varpi = \{\varpi(1), \dots, \varpi(N)\}$ is a permutation of $\{1, \dots, N\}$.
2. Let $T_{m,n}(Z_\varpi)$ be the test statistic of interest calculated based on $\mathcal{X}^\varpi_m = \{Z_{\varpi(1)}, \dots, Z_{\varpi(m)}\}$ and $\mathcal{Y}^\varpi_n = \{Z_{\varpi(m+1)}, \dots, Z_{\varpi(N)}\}$.
3. Given a significance level $0 < \alpha < 1$, return the critical value $c_{\alpha,m,n}$ defined by

$$c_{\alpha,m,n} := \inf\left\{t \in \mathbb{R} : 1 - \alpha \le \frac{1}{N!} \sum_{\varpi \in \mathcal{S}_N} \mathbb{1}(T_{m,n}(Z_\varpi) \le t)\right\}, \tag{2.6}$$

where $\mathcal{S}_N$ is the set of all permutations of $\{1, \dots, N\}$.

In Theorem 2.5, we show that the difference between $c^*_{\alpha,m,n}$ and $c_{\alpha,m,n}$ for the proposed statistic is asymptotically negligible under both the null and alternative hypotheses. This connection in turn implies that the permutation critical value converges to $q^{(\mathrm{mix})}_{\alpha,\mathrm{CvM}}$, which is the limit of the oracle critical value by construction. Moreover, under the contiguous alternative, we also establish that $q^{(\mathrm{single})}_{\alpha,\mathrm{CvM}}$ is the same as $q^{(\mathrm{mix})}_{\alpha,\mathrm{CvM}}$. Building on this observation, we formally prove that (i) the permutation test, (ii) the oracle test and (iii) the asymptotic tests based on $P^{(\mathrm{single})}_{\mathrm{CvM}}$ and $P^{(\mathrm{mix})}_{\mathrm{CvM}}$ have the same asymptotic power against both contiguous and fixed alternatives in Theorem 2.6. In doing so, we develop a general asymptotic theory for the permutation distribution of a two-sample degenerate U-statistic under H_0. This general result is established based on Hoeffding's conditions (Hoeffding (1952)) and extended to H_1 via the coupling argument (Chung and Romano (2013)). The details can be found in Appendix A.

Let us denote by $c^*_{\alpha,\mathrm{CvM}}$ and $c_{\alpha,\mathrm{CvM}}$ the critical values of the oracle test and the permutation test based on the scaled CvM-statistic, that is, NU_{CvM}, as described in the procedures (2.5) and (2.6), respectively. Then our result on the critical values is stated as follows.

THEOREM 2.5 (Asymptotic behavior of the critical values). *Consider the conventional limiting regime in* (2.1) *with the additional assumption that $m/N - \vartheta_X = O(N^{-1/2})$. Then under both the null and* (*fixed or contiguous*) *alternative hypotheses*,

$$c_{\alpha,\mathrm{CvM}} \xrightarrow{p} q^{(\mathrm{mix})}_{\alpha,\mathrm{CvM}} \quad \text{and} \quad c^*_{\alpha,\mathrm{CvM}} \xrightarrow{p} q^{(\mathrm{mix})}_{\alpha,\mathrm{CvM}},$$

where $\xrightarrow{p}$ stands for convergence in probability. Moreover, under the null or contiguous alternative, we further have that $q^{(\mathrm{mix})}_{\alpha,\mathrm{CvM}} = q^{(\mathrm{single})}_{\alpha,\mathrm{CvM}}$.

Leveraging the previous result combined with Slutsky's theorem, we next prove that the asymptotic power of the oracle test, the permutation test and the asymptotic tests are identical against any fixed and contiguous alternatives. This clearly highlights an advantage of the permutation test as it is exact under H_0 and asymptotically as powerful as the oracle and asymptotic tests under H_1. More importantly, the permutation test does not require any prior information on the underlying distributions.

THEOREM 2.6 (Asymptotic equivalence of power). *The oracle test and the permutation test control the type I error under the null hypothesis as*

$$\mathbb{P}_0\big(NU_{\mathrm{CvM}} > c^*_{\alpha,\mathrm{CvM}}\big) \le \alpha \quad \textit{and} \quad \mathbb{P}_0(NU_{\mathrm{CvM}} > c_{\alpha,\mathrm{CvM}}) \le \alpha.$$

On the other hand, under the fixed or contiguous alternative hypotheses considered in Theorem 2.3 and Theorem 2.4 with the additional assumption that $m/N - \vartheta_X = O(N^{-1/2})$, we have

$$\begin{aligned}\mathbb{P}_1(NU_{\mathrm{CvM}} > c_{\alpha,\mathrm{CvM}}) &= \mathbb{P}_1\big(NU_{\mathrm{CvM}} > c^*_{\alpha,\mathrm{CvM}}\big) + o(1)\\ &= \mathbb{P}_1\big(NU_{\mathrm{CvM}} > q^{(\mathrm{single})}_{\alpha,\mathrm{CvM}}\big) + o(1) = \mathbb{P}_1\big(NU_{\mathrm{CvM}} > q^{(\mathrm{mix})}_{\alpha,\mathrm{CvM}}\big) + o(1).\end{aligned}$$

REMARK 2.2. It is worth pointing out that due to the symmetry of the kernel $\widetilde{h}_{\mathrm{CvM}}$, it is enough to consider $\binom{N}{m}$ permutations to obtain the critical value $c_{\alpha,\mathrm{CvM}}$ for the CvM test. Nevertheless, except for small sample sizes, the exact permutation procedure is too expensive to implement in practical applications. A common approach to alleviate this computational issue is to use Monte Carlo sampling of random permutations and approximate the exact permutation p-value. In more detail, note first that the permutation test function can be written as $\mathbb{1}(\widehat{p}_{\mathrm{CvM}} \le \alpha)$ where $\widehat{p}_{\mathrm{CvM}}$ is the permutation p-value given by

$$\widehat{p}_{\mathrm{CvM}} = \frac{1}{N!}\sum_{\varpi \in \mathcal{S}_N} \mathbb{1}\big\{U_{\mathrm{CvM}}(Z_\varpi) \ge U_{\mathrm{CvM}}\big\}.$$

Let $\varpi^{(1)}, \dots, \varpi^{(B)}$ be independent and uniformly distributed on $\mathcal{S}_N$. Then the Monte Carlo version of the permutation p-value is computed by

$$\widehat{p}^{(B)}_{\mathrm{CvM}} = \frac{1}{B+1}\Bigg[\sum_{i=1}^{B} \mathbb{1}\{U_{\mathrm{CvM}}(Z_{\varpi^{(i)}}) \ge U_{\mathrm{CvM}}\} + 1\Bigg].$$

It is well known that $\mathbb{1}(\widehat{p}^{(B)}_{\mathrm{CvM}} \le \alpha)$ is also a valid level α test for any finite sample size and $\widehat{p}_{\mathrm{CvM}} - \widehat{p}^{(B)}_{\mathrm{CvM}} \xrightarrow{p} 0$ as $B \to \infty$ (e.g., page 636 of Lehmann and Romano (2005)). Throughout this paper, we also adopt this approach for our simulation studies.

3. Robustness. Recall that the energy distance and the CvM-distance can be represented by integrals of the L_2^2-type difference between two distribution functions. In view of this, the main difference between the energy distance and the CvM-distance is in their weight function. More precisely, the energy distance is defined with dt, which gives a uniform weight to the whole real line. On the other hand, the CvM-distance is defined with $dH_\beta(t)$, which gives the most weight on high-density regions. As a result, the test based on the CvM-distance is more robust to extreme observations than the one based on the energy distance. It is also important to note that the CvM-distance is well defined without any moment conditions, whereas the energy distance is only well defined assuming a finite first moment. When the moment condition is violated or there exist extreme observations, the test based on the energy distance may perform poorly. The purpose of this section is to demonstrate this point both theoretically and empirically by using contaminated distribution models.

3.1. *Theoretical analysis.* Suppose that we observe samples from an ϵ-contamination model:

$$\begin{aligned} X &\sim P_{X,N} := (1-\epsilon)Q_X + \epsilon G_N \quad \text{and}\\ Y &\sim P_{Y,N} := (1-\epsilon)Q_Y + \epsilon G_N, \end{aligned} \tag{3.1}$$

where G_N can change arbitrarily with N and $\epsilon \in (0, 1)$. Suppose that Q_X and Q_Y are different so that a given test has high power to distinguish between Q_X and Q_Y without contaminations. Then it is natural to expect that the power of the same test would not decrease much for the contamination model when ϵ is close to zero. In other words, an ideal test would maintain robust power against any choice of G_N as long as Q_X and Q_Y are different and ϵ is small. Unfortunately, this is not the case for the energy test. As we shall see, for any arbitrary small (but fixed) ϵ, there exists a contamination G_N such that the energy test becomes asymptotically powerless under mild moment conditions for Q_X and Q_Y. On the other hand, the CvM test is uniformly powerful over any choice of G_N as sample size tends to infinity.

REMARK 3.1. We mainly focus on statistical power to study robustness because one can always employ the permutation procedure to control the type I error under $H_0 : P_{X,N} = P_{Y,N}$.

Let us consider the energy statistic based on a U-statistic:

$$U_{\text{Energy}} = \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\|X_i - Y_j\| - \frac{1}{(m)_2}\sum_{i_1,i_2=1}^{m,\neq}\|X_{i_1} - X_{i_2}\| - \frac{1}{(n)_2}\sum_{j_1,j_2=1}^{n,\neq}\|Y_{j_1} - Y_{j_2}\|. \tag{3.2}$$

Then the main result of this subsection is stated as follows.

THEOREM 3.1 (Robustness under contaminations). *Suppose we observe samples $\mathcal{X}_m$ and $\mathcal{Y}_n$ from the contaminated model in (3.1) with an arbitrary small but fixed contamination ratio ϵ. Assume that Q_X and Q_Y are fixed but $Q_X \neq Q_Y$ while N changes. In addition, assume that Q_X and Q_Y have their finite second moments. Consider the tests based on U_{CvM} and U_{Energy} given by*

$$\phi_{\text{CvM}} := \mathbb{1}(U_{\text{CvM}} > c_{\alpha,\text{CvM}}) \quad \textit{and} \quad \phi_{\text{Energy}} := \mathbb{1}(U_{\text{Energy}} > c_{\alpha,\text{Eng}}),$$

where $c_{\alpha,\text{CvM}}$ and $c_{\alpha,\text{Eng}}$ are α level permutation critical values of U_{CvM} and U_{Energy}, respectively. Then for any (Q_X, Q_Y), there exists a certain G_N such that the energy test becomes asymptotically powerless under the asymptotic regime in (2.1), whereas the CvM test is asymptotically powerful uniformly over all possible G_N. More precisely,

$$\lim_{m,n\to\infty}\inf_{G_N}\mathbb{E}_1[\phi_{\text{Energy}}] \leq \alpha \quad \textit{and} \quad \lim_{m,n\to\infty}\inf_{G_N}\mathbb{E}_1[\phi_{\text{CvM}}] = 1. \tag{3.3}$$

REMARK 3.2. In Theorem 3.1, we made the assumption that Q_X and Q_Y are fixed and have finite second moments. We also assumed the asymptotic regime in (2.1). These assumptions are mainly for the energy test and are not necessary for the CvM test. In fact, the same result can be derived for the CvM test given that there is a positive sequence $b_{m,n} \to \infty$ increasing arbitrary slowly with m, n such that $W_d(Q_X, Q_Y) \geq b_{m,n}(1/\sqrt{m} + 1/\sqrt{n})$ (see Theorem 4.2).

3.2. *Empirical analysis*. To illustrate Theorem 3.1 with finite sample size, we carried out simulation studies using the contamination model in (3.1). In our simulation, we take Q_X and Q_Y to have multivariate normal distributions with different location parameters or different scale parameters. In both examples, we take G_N to have a multivariate normal distribution given by

$$G_N := N\big((0, \dots, 0)^\top, \sigma^2 I_d\big),$$

where σ controls the scale of the contamination G_N.

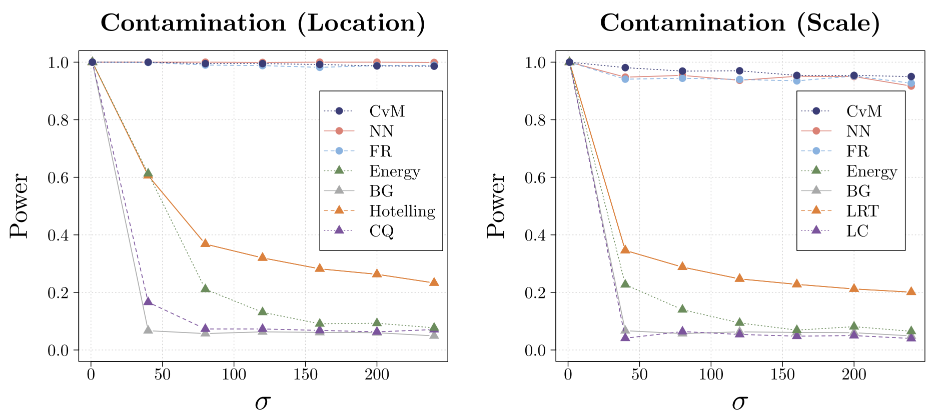

FIG. 1. *Empirical power of NN, FR, Energy, BG, Hotelling, CQ, LRT, LC and CvM tests under the contamination models with* $\epsilon = 0.05$. *See Examples* 3.1 *and* 3.2 *for details.*

EXAMPLE 3.1 (Location difference). For the location alternative, we compare two multivariate normal distributions, where the means are different but the covariance matrices are identical. Specifically, we set

$$Q_X = N\big((-0.5, \dots, -0.5)^\top, I_d\big), \quad \text{and} \quad Q_Y = N\big((0.5, \dots, 0.5)^\top, I_d\big),$$

with $\epsilon = 0.05$. We then change $\sigma = 1, 40, 80, 120, 160, 200$ and 240 to investigate the robustness of the tests against contamination with large scale parameter σ.

EXAMPLE 3.2 (Scale difference). Similar to the location alternative, we again choose multivariate normal distributions which differ in their scale but not in their location parameters. In detail, we have

$$Q_X = N\big((0, \dots, 0)^\top, 0.1^2 \times I_d\big) \quad \text{and} \quad Q_Y = N\big((0, \dots, 0)^\top, I_d\big),$$

with $\epsilon = 0.05$. Again, we change $\sigma = 1, 40, 80, 120, 160, 200$ and 240 to assess the effect of contamination with large scale parameter σ.

In addition to the energy test, we further considered three nonparametric tests in our simulation studies, namely, the k-nearest neighbor test by Schilling (1986) with $k = 3$, the MST test proposed by Friedman and Rafsky (1979) and the interpoint distance test by Biswas and Ghosh (2014). For future reference, we refer to them as the NN test, the FR test and the BG test, respectively. We also added the high-dimensional mean test by Chen and Qin (2010) and Hotelling's T^2 test (e.g., page 188 of Anderson (2003)) for the location alternative and the high-dimensional covariance test by Li and Chen (2012) and the conventional likelihood ratio test (e.g., page 412 of Anderson (2003)) for the scale alternative. We refer to them as the CQ test, Hotelling's test, the LC test and the LRT test, respectively.

Experiments were run 1000 times to estimate the power of different tests with $m = n = 40$ and $d = 10$ at significance level $\alpha = 0.05$. The p-value of each test was computed using 500 permutations as in Remark 2.2. As can be seen from Figure 1, the power of the CvM test is consistently robust to the value of σ, which supports our theoretical result. The power of the energy test, on the other hand, drops down significantly as σ increases for both location and scale differences. As explained in the proof of Theorem 3.1, this poor performance was attributed to the fact that the energy statistic is very much dominated by extreme observations from G_N when σ is large. The graph-based tests, that is, the NN and FR tests, also show a robust power performance against the contamination models. Intuitively speaking, they perform robustly under the given scenarios as their test statistics, which count the number of edges in a graph, do not vary a lot even in the presence of outliers; but as far as we know, there is no theoretical support for this result in the current literature. Moreover, these graph-based tests typically exhibit poorer consistency rates (Bhattacharya (2018)) compared to the proposed CvM test. The other four tests (Hotelling's test, the LRT test, the LC test and the CQ test) perform poorly for large σ, which may be explained similarly as to why the energy test has low power in these examples.

4. Minimax optimality. Although our choice of the U-statistic was a natural one to estimate W_d^2, it remains unclear whether one can come up with a better test statistic for testing whether $H_0 : W_d = 0$ or $H_0 : W_d > 0$. One might also wonder whether there exists a testing procedure that leads to significantly higher power than the permutation test while controlling the type I error. In this section, we shall show that the answer is negative from a minimax point of view. In particular, we prove that the permutation test based on U_{CvM} is minimax rate optimal against a class of alternatives associated with the CvM-distance.

To formulate the minimax problem, let us define the set of two multivariate distributions which are at least ϵ far apart in terms of the CvM-distance, that is,

$$\mathcal{F}(\epsilon) := \{(P_X, P_Y) : W_d(P_X, P_Y) \geq \epsilon\}.$$

For a given significance level $\alpha \in (0, 1)$, let $\mathbb{T}_{m,n}(\alpha)$ be the set of measurable functions $\phi : \{\mathcal{X}_m, \mathcal{Y}_n\} \mapsto \{0, 1\}$ such that

$$\mathbb{T}_{m,n}(\alpha) = \{\phi : \mathbb{P}_0(\phi = 1) \leq \alpha\}.$$

We then define the minimax type II error as follows:

$$1 - \beta_{m,n}(\epsilon) = \inf_{\phi \in \mathbb{T}_{m,n}(\alpha)} \sup_{P_X, P_Y \in \mathcal{F}(\epsilon)} \mathbb{P}_1(\phi = 0). \tag{4.1}$$

Our primary interest is in finding the minimax separation $\epsilon_{m,n}$ satisfying

$$\epsilon_{m,n} = \inf\{\epsilon : 1 - \beta_{m,n}(\epsilon) \leq \zeta\},$$

for some $0 < \zeta < 1 - \alpha$. We start by establishing a lower bound for the minimax separation $\epsilon_{m,n}$ based on Neyman–Pearson lemma.

THEOREM 4.1 (Lower bound). *For $0 < \zeta < 1 - \alpha$, there exists some constant $b = b(\alpha, \zeta)$ independent of the dimension such that $\epsilon_{m,n} = b(m^{-1/2} + n^{-1/2})$ and the minimax type II error is lower bounded by ζ, that is,*

$$1 - \beta_{m,n}(\epsilon_{m,n}) \geq \zeta.$$

The above result shows that if $\epsilon_{m,n}$ is of lower order than $m^{-1/2} + n^{-1/2}$, then no test has the type II error that is uniformly smaller than the nominal level α. We now prove that this lower bound is tight by establishing a matching upper bound. In particular, the upper bound is obtained by the permutation test based on U_{CvM}, highlighting that the proposed approach is minimax rate optimal.

THEOREM 4.2 (Upper bound). *Recall the CvM test ϕ_{CvM} given in Theorem 3.1. For a sufficiently large $c > 0$, let $\epsilon^\star_{m,n}$ be the radius of interest defined by*

$$\epsilon^\star_{m,n} := c\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right). \tag{4.2}$$

Then there exists $\zeta \in (0, 1-\alpha)$ *such that the type II error of* ϕ_{CvM} *is uniformly bounded by* ζ, *that is*,

$$\sup_{P_X, P_Y \in \mathcal{F}(\epsilon^\star_{m,n})} \mathbb{P}_1(\phi_{\mathrm{CvM}} = 0) < \zeta.$$

REMARK 4.1. We would like to emphasize that no assumption has been made in Theorem 4.2 regarding the ratio of the sample sizes. This implies that the proposed test can be consistent against general alternatives even when the two sample sizes are highly unbalanced as $m/n \to 0$ or $m/n \to \infty$.

As a straightforward consequence of Theorem 3.1, we also show that the energy test, which is our main competitor, is not minimax rate optimal in our context.

PROPOSITION 4.1 (Nonoptimality of the energy test). *Recall the energy test* ϕ_{Energy} *given in Theorem* 3.1. *Then there exists a pair of distributions that belongs to* $\mathcal{F}(\epsilon^\star_{m,n})$ *such that the energy test becomes asymptotically powerless*, *that is*,

$$\lim_{m,n\to\infty} \inf_{P_X, P_Y \in \mathcal{F}(\epsilon^\star_{m,n})} \mathbb{P}_1(\phi_{\mathrm{Energy}} = 1) \le \alpha.$$

In the next section, we turn our attention to the asymptotic regime where the sample size is fixed and the dimension tends to infinity and study the limiting behavior of the CvM test.

5. High dimension, low sample size analysis. The high dimension, low sample size (HDLSS) regime has received increasing attention in recent years and has been frequently employed to give statistical insights into high-dimensional two-sample testing (e.g., Biswas and Ghosh (2014), Biswas, Mukhopadhyay and Ghosh (2014), Chakraborty and Chaudhuri (2017), Mondal, Biswas and Ghosh (2015)). Focusing on this HDLSS regime, the goal of this section is twofold: First, we provide sufficient conditions under which the proposed test is consistent in HDLSS situations (Section 5.1). Second, we show that U_{CvM} has the same asymptotic behavior as the high-dimensional mean test statistics proposed by Chen and Qin (2010) and Chakraborty and Chaudhuri (2017) under certain location models (Section 5.2). Along with these mean test statistics, we further establish the equivalence among U_{CvM}, the energy statistic and the MMD statistic with the Gaussian kernel. The latter connection was motivated by Ramdas et al. (2015) who showed that the energy statistic, the MMD statistic and the mean test statistic by Chen and Qin (2010) are asymptotically equivalent in different scenarios.

5.1. *HDLSS consistency.* Let us denote $\mathbb{E}(X) = \mu_X$, $\mathbb{E}(Y) = \mu_Y$, $\mathbb{V}(X) = \Sigma_X$ and $\mathbb{V}(Y) = \Sigma_Y$ where Σ_X and Σ_Y are positive definite matrices. Before presenting the main results, we state the two assumptions.

(A1). $\mathbb{V}(\|Z_1^* - Z_2^*\|^2) = O(d)$, and $\mathbb{V}\{(Z_1^* - Z_3^*)^\top (Z_2^* - Z_3^*)\} = O(d)$,

where Z_1^*, Z_2^*, Z_3^* are independent and each Z_i^* follows either P_X or P_Y.

(A2). $d^{-1}\operatorname{tr}(\Sigma_X) \to \overline{\sigma}_X^2$, $d^{-1}\operatorname{tr}(\Sigma_Y) \to \overline{\sigma}_Y^2$, $d^{-1}\|\mu_X - \mu_Y\|_2^2 \to \overline{\delta}_{XY}^2$,

where $0 < \overline{\sigma}_X^2, \overline{\sigma}_Y^2 < \infty$ and $0 \le \overline{\delta}_{XY}^2 < \infty$.

Assumption **(A1)** implies that component variables are weakly dependent. Under the distributional assumptions (including multivariate normal distributions) made in Bai and Saranadasa (1996) and Chen and Qin (2010), Assumption **(A1)** is satisfied when

$$(5.1)\quad \begin{aligned} &(\mu_X - \mu_Y)^\top(\Sigma_X + \Sigma_Y)(\mu_X - \mu_Y) = O(d) \quad \text{and} \\ &\operatorname{tr}\{(\Sigma_X + \Sigma_Y)^2\} = O(d). \end{aligned}$$

The details of this derivation can be found in Appendix D.1. Assumption **(A2)** is common in the HDLSS literature (e.g., Hall, Marron and Neeman (2005)) and facilitates the analysis. Under these two conditions, the following theorem establishes the HDLSS consistency of the proposed test where we assume that the nominal level satisfies $\alpha > 1/\{(m+n)!/(m!n!)\}$ for $m \neq n$ and $\alpha > 2/\{(m+n)!/(m!n!)\}$ for $m = n$.

THEOREM 5.1 (HDLSS consistency). *Suppose* **(A1)** *and* **(A2)** *hold. Assume that* $\overline{\sigma}_X^2 \neq \overline{\sigma}_Y^2$ *or* $\overline{\delta}_{XY}^2 > 0$. *Then the permutation test based on* U_{CvM} *is consistent under the HDLSS regime, that is,* $\lim_{d\to\infty} \mathbb{E}_1[\phi_{\mathrm{CvM}}] = 1$.

5.2. *HDLSS asymptotic equivalence of CvM-statistic and others.* Next, we focus on mean difference alternatives with equal covariance matrices. There are many types of high-dimensional mean inference procedures in the literature (see Hu and Bai (2016), for a recent review). For example, Chen and Qin (2010) suggest a test statistic based on an unbiased estimator of $\|\mu_X - \mu_Y\|^2$. Specifically, their test statistic is given by

$$U_{\mathrm{CQ}} = \frac{1}{(m)_2(n)_2} \sum_{i_1,i_2=1}^{m,\neq} \sum_{j_1,j_2=1}^{n,\neq} (X_{i_1} - Y_{j_1})^\top (X_{i_2} - Y_{j_2}).$$

More recently, Chakraborty and Chaudhuri (2017) define a test statistic based on spatial ranks as

$$U_{\mathrm{WMW}} = \frac{1}{(m)_2(n)_2} \sum_{i_1,i_2=1}^{m,\neq} \sum_{j_1,j_2=1}^{n,\neq} \frac{(X_{i_1} - Y_{j_1})}{\|X_{i_1} - Y_{j_1}\|}^\top \frac{(X_{i_2} - Y_{j_2})}{\|X_{i_2} - Y_{j_2}\|}.$$

They proved that U_{CQ} and U_{WMW} are asymptotically equivalent under a certain HDLSS setting. Independently, the equivalence between U_{CQ}, U_{Energy} and the MMD statistic with the Gaussian kernel was established by Ramdas et al. (2015) under different settings. Let us denote the MMD statistic with the Gaussian kernel by

$$\begin{aligned} U_{\mathrm{MMD}} = {} & \frac{1}{(m)_2} \sum_{i_1,i_2=1}^{m,\neq} \exp\biggl(-\frac{1}{2\varsigma_d^2}\|X_{i_1} - X_{i_2}\|^2\biggr) \\ & + \frac{1}{(n)_2} \sum_{j_1,j_2=1}^{n,\neq} \exp\biggl(-\frac{1}{2\varsigma_d^2}\|Y_{j_1} - Y_{j_2}\|^2\biggr) - \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} \exp\biggl(-\frac{1}{2\varsigma_d^2}\|X_i - Y_j\|^2\biggr), \end{aligned}$$

where ς_d^2 is the bandwidth parameter. Here, we combine and further extend these results by presenting sufficient conditions under which U_{CvM}, U_{Energy}, U_{MMD}, U_{CQ} and U_{WMW} are asymptotically equivalent. To establish the result, we need two more assumptions:

(A3). $\mathbb{V}\{(Z_1^* - Z_2^*)^\top (Z_3^* - Z_4^*)\} = O(d)$ where

$Z_1^*, Z_2^*, Z_3^*, Z_4^*$ are independent and each Z_i^* follows either P_X or P_Y.

(A4). $\Sigma_X = \Sigma_Y$ and $\|\mu_X - \mu_Y\|^2 = O(\sqrt{d})$.

Assumption **(A3)** is required for studying U_{CQ} and U_{WMW}. Like Assumption **(A1)**, Assumption **(A3)** is also satisfied under condition (5.1). Notice that U_{CQ} and U_{WMW} are only sensitive to location parameters whereas U_{CvM}, U_{Energy} and U_{MMD} are sensitive to both location and scale parameters. This suggests that the equal covariance assumption in **(A4)** is crucial for our result and cannot be dropped. The condition $\|\mu_X - \mu_Y\|^2 = O(\sqrt{d})$ is also important for our analysis and was also considered in Chakraborty and Chaudhuri (2017). Under the given assumptions, we make repeated use of Taylor expansions to establish the equivalence among the test statistics stated as follows.

THEOREM 5.2 (HDLSS equivalence). *Suppose* **(A1)**, **(A2)**, **(A3)** *and* **(A4)** *hold. Let* ϖ *be an arbitrary permutation of* $\{1,\dots,N\}$ *and* $\overline{\sigma}_d^2 = d^{-1}\operatorname{tr}(\Sigma_X)$. *We denote by* $U_{\mathrm{CvM}}^{\varpi}$, $U_{\mathrm{Energy}}^{\varpi}$, $U_{\mathrm{MMD}}^{\varpi}$, U_{CQ}^{ϖ} *and* $U_{\mathrm{WMW}}^{\varpi}$, *the CvM, Energy, MMD, CQ and WMW test statistics, respectively, calculated based on* $\mathcal{X}_m^{\varpi} = \{Z_{\varpi(1)},\dots,Z_{\varpi(m)}\}$ *and* $\mathcal{Y}_n^{\varpi} = \{Z_{\varpi(m+1)},\dots,Z_{\varpi(N)}\}$. *Assume that the bandwidth parameter of the Gaussian kernel satisfies* $\varsigma_d^2 \asymp d$. *Then under the HDLSS asymptotics, we have that*

$$
\begin{aligned}
\sqrt{d}U_{\mathrm{CvM}}^{\varpi} &= \frac{1}{2\pi\sqrt{3d\overline{\sigma}_d^2}}U_{\mathrm{CQ}}^{\varpi} + O_{\mathbb{P}}(d^{-1/2}),\\
U_{\mathrm{Energy}}^{\varpi} &= \frac{1}{\sqrt{2d}\overline{\sigma}_d}U_{\mathrm{CQ}}^{\varpi} + O_{\mathbb{P}}(d^{-1/2}),\\
\sqrt{d}U_{\mathrm{WMW}}^{\varpi} &= \frac{1}{\sqrt{d\overline{\sigma}_d^2}}U_{\mathrm{CQ}}^{\varpi} + O_{\mathbb{P}}(d^{-1/2}),\\
\sqrt{d}U_{\mathrm{MMD}}^{\varpi} &= \frac{\sqrt{d}}{\varsigma_d^2}e^{-d\overline{\sigma}_d^2/\varsigma_d^2}U_{\mathrm{CQ}}^{\varpi} + O_{\mathbb{P}}(d^{-1/2}).
\end{aligned}
\tag{5.2}
$$

Note that the asymptotic equivalence established in (5.2) holds for any permutation. Leveraging this result, we show that the permutation critical values of the test statistics are asymptotically the same as well.

COROLLARY 5.1 (Permutation critical values). *Consider the same assumptions made in Theorem 5.2. Let* $c_{\alpha,\mathrm{CvM}}$, $c_{\alpha,\mathrm{Eng}}$, $c_{\alpha,\mathrm{MMD}}$, $c_{\alpha,\mathrm{CQ}}$ *and* $c_{\alpha,\mathrm{WMW}}$ *be the* $1-\alpha$ *quantile of the permutation distribution of* $2\pi\sqrt{3d\overline{\sigma}_d^2}U_{\mathrm{CvM}}$, $\sqrt{2}\overline{\sigma}_d U_{\mathrm{Energy}}$, $\varsigma_d^2 e^{-d\overline{\sigma}_d^2/\varsigma_d^2}U_{\mathrm{MMD}}/\sqrt{d}$, $U_{\mathrm{CQ}}/\sqrt{d}$ *and* $\sqrt{d\overline{\sigma}_d^2}U_{\mathrm{WMW}}$, *respectively. Then*

$$
\begin{aligned}
c_{\alpha,\mathrm{CvM}} &= c_{\alpha,\mathrm{Eng}} + O_{\mathbb{P}}(d^{-1/2}) = c_{\alpha,\mathrm{MMD}} + O_{\mathbb{P}}(d^{-1/2})\\
&= c_{\alpha,\mathrm{CQ}} + O_{\mathbb{P}}(d^{-1/2}) = c_{\alpha,\mathrm{WMW}} + O_{\mathbb{P}}(d^{-1/2}).
\end{aligned}
$$

From the previous results, we expect that the considered permutation tests have comparable power in the limit as further illustrated by our simulation results in Section 8. We also refer the reader to Appendix D.5 where we present an explicit expression for the limiting power function of the asymptotic tests with extra assumptions. We would like to emphasize, however, that when the moment assumption is violated, the power of these tests can be entirely different. For instance, our simulation results in Section 8 demonstrate that the CQ, energy and MMD tests perform poorly when X and Y have Cauchy distributions with different location parameters. In contrast, the CvM and WMW tests maintain robust power against the same Cauchy alternative, which highlights a benefit of the current approach in high dimensions.

6. Connection to the generalized energy distance and MMD. Recall that the energy distance is defined with the Euclidean distance under the finite first moment condition. By considering a semimetric space $(\mathbb{Z}, \rho)$ of negative type, Sejdinovic et al. (2013) generalized the energy distance by

$$E_\rho^2 = 2\mathbb{E}[\rho(X_1, Y_1)] - \mathbb{E}[\rho(X_1, X_2)] - \mathbb{E}[\rho(Y_1, Y_2)].$$

They further established the equivalence between the generalized energy distance and the MMD with a kernel induced by $\rho(\cdot,\cdot)$. Given a distance-induced kernel $k(\cdot,\cdot)$, the squared MMD is given by

$$\mathrm{MMD}_k^2 = \mathbb{E}[k(X_1, X_2)] + \mathbb{E}[k(Y_1, Y_2)] - 2\mathbb{E}[k(X_1, Y_1)].$$

In this section, we show that the multivariate CvM-distance is a member of the generalized energy distance by the use of the angular distance, and thus also a member of the MMD. Let $\mathcal{M}_X$ and $\mathcal{M}_Y$ be the support of X and Y, respectively, and let $\mathcal{M} = \mathcal{M}_X \cup \mathcal{M}_Y \subseteq \mathbb{R}^d$. Then we define the *angular distance* as follows.

DEFINITION 6.1 (Angular distance). Let Z^* be a random vector having mixture distribution $(1/2)P_X + (1/2)P_Y$. For $z, z' \in \mathcal{M}$, denote the scaled angle between $z - Z^*$ and $z' - Z^*$ by

$$\rho_{\mathrm{Angle}}(z, z'; Z^*) = \frac{1}{\pi}\mathsf{Ang}(z - Z^*, z' - Z^*).$$

The angular distance is defined as the expected value of the scaled angle:

$$\rho_{\mathrm{Angle}}(z, z') = \mathbb{E}[\rho_{\mathrm{Angle}}(z, z'; Z^*)]. \tag{6.1}$$

As shown in Appendix D.6, ρ_{Angle} is a metric of negative type defined on $\mathcal{M}$. With this key property and the identity given in the next proposition, we may conclude that the multivariate CvM-distance is a special case of the generalized energy distance based on the angular distance.

PROPOSITION 6.1 (Another view of the CvM-distance). *Let us consider the angular distance defined in* (6.1). *Then*

$$2W_d^2 = 2\mathbb{E}[\rho_{\mathrm{Angle}}(X_1, Y_1)] - \mathbb{E}[\rho_{\mathrm{Angle}}(X_1, X_2)] - \mathbb{E}[\rho_{\mathrm{Angle}}(Y_1, Y_2)].$$

REMARK 6.1. The angular distance can be generalized by taking the expectation with respect to a different measure. For instance, when the expectation is taken with respect to Lebesgue measure, the generalized angular distance is proportional to the Euclidean distance (see Appendix D.7). The main difference between the Euclidean distance and the proposed angular distance is that the latter takes into account information from the underlying distribution and is less sensitive to outliers. In this respect, the introduced angular distance can be viewed as a robust alternative for the Euclidean distance.

REMARK 6.2. As one of the reviewers pointed out, there might be several ways to enhance the power of the proposed test by modifying the multivariate CvM-distance. For instance, by the characteristic property of W_d^2, it can be seen that H_0 holds if and only if $\mathsf{T}_{xy} = \mathsf{T}_{xx}$ and $\mathsf{T}_{xy} = \mathsf{T}_{yy}$ where $\mathsf{T}_{xy} = \mathbb{E}[\rho_{\mathrm{Angle}}(X_1, Y_1)]$, $\mathsf{T}_{xx} = \mathbb{E}[\rho_{\mathrm{Angle}}(X_1, X_2)]$ and $\mathsf{T}_{yy} = \mathbb{E}[\rho_{\mathrm{Angle}}(Y_1, Y_2)]$. Motivated by this observation, another test statistic can be introduced based on an estimate of $(\mathsf{T}_{xy} - \mathsf{T}_{xx})^2 + (\mathsf{T}_{xy} - \mathsf{T}_{yy})^2$ (and other variants are possible, see Appendix D.4). As demonstrated in Appendix E, the test based on this new statistic tends to be more sensitive to scale differences than the CvM test.

7. Other multivariate extensions via projection averaging. The projection-averaging approach used for the multivariate CvM-statistic is general and can be applied to many other univariate robust statistics. In this section and also Appendix D.8, we illustrate the utility of the projection-averaging approach by considering several examples including Kendall's tau, Spearman's rho and the sign covariance (Bergsma and Dassios (2014)). Let us begin with one-sample and two-sample robust statistics. Given a pair of random variables (X, Y), denote the difference between two random variables by $Z = X - Y$. The univariate sign test statistic is an estimate of $\mathsf{T}_{\text{sign}} := \mathbb{P}(Z > 0) - 1/2$ and it is used to test whether

$$H_0 : \mathbb{P}(Z > 0) = 1/2 \quad \text{versus} \quad H_1 : \mathbb{P}(Z > 0) \neq 1/2.$$

The projection-averaging technique extends T_{sign} to a multivariate case as follows.

PROPOSITION 7.1 (One-sample sign test statistic). *For i.i.d. random vectors Z_1, Z_2 from a multivariate distribution P_Z where $Z \in \mathbb{R}^d$, the projection-averaging approach generalizes T_{sign} as*

$$\int_{\mathbb{S}^{d-1}} \left(\mathbb{P}(\beta^\top Z_1 > 0) - \frac{1}{2}\right)^2 d\lambda(\beta) = \frac{1}{4} - \frac{1}{2\pi}\mathbb{E}[\mathsf{Ang}(Z_1, Z_2)]. \tag{7.1}$$

Given univariate two samples $\mathcal{X}_m = \{X_1, \dots, X_m\}$ and $\mathcal{Y}_n = \{Y_1, \dots, Y_n\}$, the Wilcoxon–Mann–Whitney test is designed for testing whether

$$H_0 : \mathbb{P}(X > Y) = 1/2 \quad \text{versus} \quad H_1 : \mathbb{P}(X > Y) \neq 1/2,$$

based on an estimate of $\mathsf{T}_{\text{WMW}} := \mathbb{P}(X > Y) - 1/2$. The next proposition extends T_{WMW} to a multivariate case via projection averaging.

PROPOSITION 7.2 (Two-sample Wilcoxon–Mann–Whitney test statistic). *Let $X_1, X_2 \overset{i.i.d.}{\sim} P_X$ and, independently, $Y_1, Y_2 \overset{i.i.d.}{\sim} P_Y$ where $X_1, Y_1 \in \mathbb{R}^d$. The projection-averaging approach generalizes T_{WMW} as*

$$\int_{\mathbb{S}^{d-1}} \left(\mathbb{P}(\beta^\top X_1 > \beta^\top Y_1) - \frac{1}{2}\right)^2 d\lambda(\beta) = \frac{1}{4} - \frac{1}{2\pi}\mathbb{E}[\mathsf{Ang}(X_1 - Y_1, X_2 - Y_2)]. \tag{7.2}$$

REMARK 7.1. The first-order Taylor approximation of the inverse cosine function shows that the representations given on the right-hand side of (7.1) and (7.2) are related to the spatial sign-statistics introduced by Wang, Peng and Li (2015) and Chakraborty and Chaudhuri (2017), respectively. In fact, when U-statistics are used to estimate (7.1) and (7.2), the projection-averaging statistics and the spatial sign-statistics are asymptotically equivalent under some regularity conditions (see Appendix D.3 for details).

The same technique can be further applied to some robust statistics for independence testing. To test for independence between two random variables, Kendall's tau statistic is given as an estimate of $\tau := 4\mathbb{P}(X_1 < X_2, Y_1 < Y_2) - 1$. We present a multivariate extension of τ as follows.

THEOREM 7.1 (Kendall's tau). *For i.i.d. pairs of random vectors $(X_1, Y_1), \dots, (X_4, Y_4)$ from a joint distribution P_{XY} where $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$, the multivariate extension of τ via projection averaging is given by*

$$\begin{aligned}
&\int_{\mathbb{S}^{p-1}} \int_{\mathbb{S}^{q-1}} [4\mathbb{P}(\alpha^\top (X_1 - X_2) < 0, \beta^\top (Y_1 - Y_2) < 0) - 1]^2 \, d\lambda(\alpha) \, d\lambda(\beta) \\
&\quad = \mathbb{E}\left[\left(2 - \frac{2}{\pi}\mathsf{Ang}(X_1 - X_2, X_3 - X_4)\right) \cdot \left(2 - \frac{2}{\pi}\mathsf{Ang}(Y_1 - Y_2, Y_3 - Y_4)\right)\right] - 1.
\end{aligned}$$

Recently, Bergsma and Dassios (2014) introduced a modification of Kendall's tau, which is zero if and only if random variables are independent under some mild conditions. Let us denote the univariate Bergsma–Dassios sign covariance by

$$\tau^* = \mathbb{E}\big[a_{\text{sign}}(X_1, X_2, X_3, X_4) \cdot a_{\text{sign}}(Y_1, Y_2, Y_3, Y_4)\big], \tag{7.3}$$

with $a_{\text{sign}}(z_1, z_2, z_3, z_4) = \text{sign}(|z_1 - z_2| + |z_3 - z_4| - |z_1 - z_3| - |z_2 - z_4|)$. Motivated by the projection-averaging approach, we propose the multivariate τ^* as follows.

DEFINITION 7.1 (Multivariate τ^*). Suppose $(X_1, Y_1), \dots, (X_4, Y_4)$ are i.i.d. random vectors from a joint distribution P_{XY} where $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$. We define the multivariate τ^* by

$$\begin{aligned}\tau^*_{p,q} = \int_{\mathbb{S}^{p-1}} \int_{\mathbb{S}^{q-1}} &\mathbb{E}\big[a_{\text{sign}}(\alpha^\top X_1, \alpha^\top X_2, \alpha^\top X_3, \alpha^\top X_4) \\ &\times a_{\text{sign}}\big(\beta^\top Y_1, \beta^\top Y_2, \beta^\top Y_3, \beta^\top Y_4\big)\big]\, d\lambda(\alpha)\, d\lambda(\beta).\end{aligned}$$

Since the kernel of τ^* is sign-invariant, that is, $a_{\text{sign}}(z_1, z_2, z_3, z_4) = a_{\text{sign}}(-z_1, -z_2, -z_3, -z_4)$, it is easy to see that $\tau^*_{p,q}$ becomes the univariate τ^* when $p = q = 1$. Also note that since X and Y are independent if and only if $\alpha^\top X$ and $\beta^\top Y$ are independent for all $\alpha \in \mathbb{S}^{p-1}$ and $\beta \in \mathbb{S}^{q-1}$, the characteristic property of $\tau^*_{p,q}$ follows by that of the univariate τ^*.

Next, we present a closed-form expression for $\tau^*_{p,q}$. For nonzero $U_1, U_2, U_3 \in \mathbb{R}^d$, let us define $g_d(U_1, U_2, U_3)$ and $h_d(Z_1, Z_2, Z_3, Z_4)$ by

$$g_d(U_1, U_2, U_3) = \frac{1}{2} - \frac{1}{4\pi}\big[\mathsf{Ang}(U_1, U_2) + \mathsf{Ang}(U_1, U_3) + \mathsf{Ang}(U_2, U_3)\big]$$

and

$$\begin{aligned}&h_d(Z_1, Z_2, Z_3, Z_4) \\ &\quad = g_d(Z_1 - Z_2, Z_2 - Z_3, Z_3 - Z_4) + g_d(Z_2 - Z_1, Z_1 - Z_3, Z_3 - Z_4) \\ &\qquad + g_d(Z_1 - Z_2, Z_2 - Z_4, Z_4 - Z_3) + g_d(Z_2 - Z_1, Z_1 - Z_4, Z_4 - Z_3).\end{aligned}$$

We note that, in contrast to other applications, Lemma 2.2 is not enough to have an expression for $\tau^*_{p,q}$ without involving integrations over the unit sphere. To this end, we generalize Lemma 2.2 with three indicator functions (see Lemma B.8) and present an alternative expression for $\tau^*_{p,q}$ as follows.

THEOREM 7.2 (Closed-form expression for $\tau^*_{p,q}$). *For i.i.d. random vectors* $(X_1, Y_1), \dots, (X_4, Y_4)$ *from a joint distribution* P_{XY} *where* $X \in \mathbb{R}^p$ *and* $Y \in \mathbb{R}^q$, $\tau^*_{p,q}$ *can be written as*

$$\begin{aligned}\tau^*_{p,q} &= \mathbb{E}\big[h_p(X_1, X_2, X_3, X_4) \cdot h_q(Y_1, Y_2, Y_3, Y_4)\big] \\ &\quad + \mathbb{E}\big[h_p(X_1, X_2, X_3, X_4) \cdot h_q(Y_3, Y_4, Y_1, Y_2)\big] \\ &\quad - 2\mathbb{E}\big[h_p(X_1, X_2, X_3, X_4) \cdot h_q(Y_1, Y_3, Y_2, Y_4)\big].\end{aligned}$$

Theorem 7.2 leads to a straightforward empirical estimate of $\tau^*_{p,q}$ based on a U-statistic. This is also true for the other multivariate generalizations introduced in this section and the Supplementary Material (Appendix D.8). Using these estimates, some theoretical and empirical properties of the proposed measures can be further investigated. These topics are reserved for future work.

8. Simulations. In this section, we report numerical results to support the argument in Section 5 as well as to compare the performance of the CvM test with other competing nonparametric tests against heavy-tailed alternatives. Along with the energy, MMD, NN, FR and BG tests described before, we consider the cross-match test (Rosenbaum (2005)), the multivariate run test (Biswas, Mukhopadhyay and Ghosh (2014)), the modified k-NN test (Mondal, Biswas and Ghosh (2015)) and the ball divergence test (Pan et al. (2018)) for comparison. We refer to them as the CM test, run test, MBG test and ball test, respectively. In our simulations, we used the Gaussian kernel with the median heuristic (Gretton et al. (2012)) for the MMD test and we set the number of nearest neighbors as $k = 3$ for both NN test and MBG test. Since finding the shortest Hamiltonian path for the run test is NP-complete, we employed Kruskal's algorithm (Kruskal (1956)) as suggested by Biswas, Mukhopadhyay and Ghosh (2014).

Throughout our experiments, the significance level was set at 0.05 and the permutation procedure was used to determine the p-value of each test with 200 permutations as in Remark 2.2. The simulations were repeated 500 times to approximate the power of different tests. We set the sample size and the dimension by $m, n = 20$ and $d = 200$ for the balanced cases and by $m = 35$, $n = 5$ and $d = 200$ for the imbalanced cases.

First, we consider several examples where the powers of the five tests (CvM, energy, MMD, CQ and WMW tests) in Section 5 are approximately equivalent to each other. Specifically, we use multivariate normal distributions with different means

$$\mu^{(0)} = (0, \dots, 0)^\top, \qquad \mu^{(1)} = (0.15, \dots, 0.15)^\top \quad \text{and}$$

$$\mu^{(2)} = \sqrt{0.045}(\underbrace{1, \dots, 1}_{d/2 \text{ elements}}, \underbrace{0, \dots, 0}_{d/2 \text{ elements}})^\top$$

and covariance matrices:

1. Identity matrix (denoted by I) where $\sigma_{i,i} = 1$ and $\sigma_{i,j} = 0$ for $i \neq j$.
2. Banded matrix (denoted by Σ_{Band}) where $\sigma_{i,i} = 1$, $\sigma_{i,j} = 0.6$ for $|i - j| = 1$, $\sigma_{i,j} = 0.3$ for $|i - j| = 2$ and $\sigma_{i,j} = 0$ otherwise.
3. Autocorrelation matrix (denoted by Σ_{Auto}) where $\sigma_{i,i} = 1$ and $\sigma_{i,j} = 0.2^{|i-j|}$ when $i \neq j$.
4. Block diagonal matrix (denoted by Σ_{Block}) where the 5×5 main diagonal blocks $\mathbf{A}$ are defined by $a_{i,i} = 1$ and $a_{i,j} = 0.2$ when $i \neq j$, and the off-diagonal blocks are zeros.

Then we generate random samples from $X \sim N(\mu^{(0)}, \Sigma)$ and either $Y \sim N(\mu^{(1)}, \Sigma)$ or $Y \sim N(\mu^{(2)}, \Sigma)$. The results are summarized in Table 1. As can be seen from the table, the empirical powers of the considered tests are very close under the given setting, which supports our theoretical results in Section 5. We also observe that the other nonparametric tests, not considered in Section 5, are significantly less powerful than the proposed test in all normal location alternatives.

In our second experiment, we consider several examples where the moment conditions are not satisfied. We focus on random samples generated from multivariate Cauchy distributions. Let Cauchy(γ, s) refer to the univariate Cauchy distribution where γ, s are the location parameter and the scale parameter, respectively. Let $X = (X^{(1)}, \dots, X^{(d)})$ and $Y = (Y^{(1)}, \dots, Y^{(d)})$ be random vectors where $X^{(i)} \overset{i.i.d.}{\sim}$ Cauchy$(0, 1)$ and $Y^{(i)} \overset{i.i.d.}{\sim}$ Cauchy(γ, s) for $i = 1, \dots, d$. We first consider location differences where γ is not zero but the scale parameters are identical, that is, $s = 1$. Similarly, we consider scale differences where the scale parameter s changes, but the location parameters are identical, that is, $\gamma = 0$.

From the results presented in Table 2 and Table 3, it is seen that, unlike the multivariate normal cases, there are significant differences between power performance among CvM,

TABLE 1
Empirical power of the considered tests against the normal location models at $\alpha = 0.05$

	I_d		Σ_{Band}		Σ_{Block}		Σ_{Auto}	
$m = 20, n = 20$	$\mu^{(1)}$	$\mu^{(2)}$	$\mu^{(1)}$	$\mu^{(2)}$	$\mu^{(1)}$	$\mu^{(2)}$	$\mu^{(1)}$	$\mu^{(2)}$
CvM	**0.662**	**0.646**	**0.418**	**0.406**	**0.572**	**0.584**	**0.452**	**0.442**
Energy	**0.656**	**0.650**	**0.420**	**0.408**	**0.576**	**0.584**	**0.452**	**0.444**
MMD	**0.658**	**0.638**	**0.412**	**0.398**	**0.568**	**0.570**	**0.458**	**0.444**
CQ	**0.656**	**0.650**	**0.416**	**0.412**	**0.578**	**0.580**	**0.454**	**0.448**
WMW	**0.668**	**0.646**	**0.420**	**0.402**	**0.568**	**0.580**	**0.458**	**0.444**
NN	0.288	0.288	0.164	0.154	0.242	0.238	0.176	0.174
FR	0.168	0.170	0.090	0.084	0.158	0.116	0.112	0.088
MBG	0.050	0.050	0.050	0.052	0.048	0.044	0.060	0.046
Ball	0.240	0.254	0.186	0.198	0.262	0.250	0.216	0.226
CM	0.042	0.054	0.028	0.040	0.052	0.050	0.038	0.034
BG	0.070	0.060	0.074	0.074	0.074	0.078	0.084	0.078
Run	0.160	0.153	0.101	0.105	0.146	0.128	0.110	0.102

energy, MMD, CQ and WMW tests. In particular, the tests based on the energy, MMD and CQ statistics have relatively low power against the heavy-tail location alternatives, whereas the tests based on the CvM and WMW statistics show better performance than the others. Turning to the scale problems, it can be seen that the CQ and WMW tests are not sensitive to detect scale differences, which makes sense because they are specifically designed for location problems. On the other hand, the CvM, energy and MMD tests perform reasonably well in these alternatives. Among the omnibus nonparametric tests, the MMD, energy and ball tests have competitive power against the scale differences, but not against the location differences in general. The MBG test is only powerful against the scale differences where the sample sizes are balanced. The CM and run tests are uniformly outperformed by the CvM test under all scenarios. The NN and FR tests perform strongly against the location alternatives especially for the balanced case, but not against the scale alternatives. When the sample

TABLE 2
Empirical power of the considered tests against multivariate Cauchy distributions with $m = n = 20$ at $\alpha = 0.05$ where γ, s represent the location and scale parameter, respectively. The three highest power estimates in each column are highlighted in boldface

	Location				Scale			
$m = 20, n = 20$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$	$s = 2$	$s = 3$	$s = 4$	$s = 5$
CvM	0.124	0.252	0.596	0.842	**0.560**	**0.926**	**0.988**	**1.000**
Energy	0.060	0.066	0.102	0.134	0.316	0.602	0.766	0.866
MMD	0.056	0.064	0.110	0.162	0.448	0.772	0.890	0.970
CQ	0.138	0.268	0.360	0.456	0.046	0.070	0.042	0.068
WMW	**0.324**	**0.698**	**0.912**	**0.988**	0.052	0.064	0.062	0.056
NN	**0.288**	**0.662**	**0.884**	**0.976**	0.214	0.194	0.256	0.224
FR	**0.178**	**0.462**	**0.706**	**0.888**	0.028	0.034	0.048	0.036
MBG	0.060	0.044	0.050	0.074	**0.564**	**0.904**	**0.964**	**0.992**
Ball	0.064	0.064	0.076	0.098	**0.606**	**0.936**	**0.994**	**1.000**
CM	0.030	0.078	0.128	0.226	0.056	0.170	0.334	0.490
BG	0.048	0.038	0.048	0.040	0.238	0.394	0.560	0.632
Run	0.059	0.129	0.274	0.422	0.220	0.506	0.767	0.864

TABLE 3
Empirical power of the considered tests against multivariate Cauchy distributions with $m = 35$ and $n = 5$ at $\alpha = 0.05$ where γ, s represent the location and scale parameter, respectively. The three highest power estimates in each column are highlighted in boldface

	Location				Scale			
$m = 35, n = 5$	$\gamma = 5$	$\gamma = 6$	$\gamma = 7$	$\gamma = 8$	$s = 3$	$s = 4$	$s = 5$	$s = 6$
CvM	**0.340**	**0.498**	**0.652**	**0.758**	**0.570**	**0.806**	**0.928**	**0.952**
Energy	0.110	0.146	0.212	0.262	**0.436**	**0.632**	**0.794**	**0.858**
MMD	0.108	0.148	0.192	0.240	**0.552**	**0.808**	**0.926**	**0.968**
CQ	**0.284**	**0.380**	0.454	0.544	0.178	0.210	0.262	0.290
WMW	**0.796**	**0.890**	**0.942**	**0.960**	0.110	0.126	0.134	0.148
NN	0.144	0.294	0.376	0.558	0.118	0.150	0.154	0.182
FR	0.226	0.360	**0.464**	**0.588**	0.078	0.092	0.104	0.112
MBG	0.010	0.000	0.008	0.000	0.092	0.130	0.176	0.214
Ball	0.072	0.088	0.098	0.122	0.238	0.406	0.594	0.762
CM	0.082	0.176	0.190	0.262	0.030	0.080	0.092	0.126
BG	0.058	0.052	0.058	0.052	0.320	0.386	0.506	0.514
Run	0.088	0.150	0.198	0.228	0.106	0.174	0.248	0.326

sizes are unbalanced, the performance of the NN and FR tests are degraded a little bit, which can be explained by Chen, Dou and Qiao (2013) and Chen, Chen and Su (2018). The CvM test, on the other hand, performs consistently well against the heavy-tail location and scale alternatives and its performance appears immune to the sample proportion.

In summary, the proposed test has almost identical power as the high-dimensional mean tests against the light-tail location alternatives, whereas it outperforms many popular non-parametric competitors under the heavy-tail location and scale alternatives. More simulation results in both high and low dimensions can be found in Appendix E of the Supplementary Material.

9. Concluding remarks. In this work, we extended the univariate Cramér–von Mises statistic for two-sample testing to the multivariate case using projection averaging and demonstrated its robustness, minimax rate optimality and high-dimensional power properties. We applied the same projection technique to other robust statistics and presented their multivariate extensions. We also introduced the angular distance that is closely connected to the Euclidean distance (Remark 6.1) but is more robust to outliers by incorporating information from the underlying distribution. Given that the use of distances is of fundamental importance in many statistical applications (including clustering, classification and regression), we hope that this work will stimulate further study of the angular distance applied to other statistical problems serving as a robust alternative for the Euclidean distance.

Acknowledgments. The authors would like to thank the Associate Editor and two referees for their valuable comments which helped to improve the manuscript.

SUPPLEMENTARY MATERIAL

Supplement to "Robust multivariate nonparametric tests via projection averaging" (DOI: 10.1214/19-AOS1936SUPP; .pdf). This supplemental file includes the technical proofs omitted in the main text, and some additional results.

REFERENCES

ANDERSON, T. W. (1962). On the distribution of the two-sample Cramér–von Mises criterion. *Ann. Math. Stat.* **33** 1148–1159. MR0145607 https://doi.org/10.1214/aoms/1177704477

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley-Interscience, Hoboken, NJ. MR1990662

ANDERSON, N. H., HALL, P. and TITTERINGTON, D. M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Multivariate Anal.* **50** 41–54. MR1292607 https://doi.org/10.1006/jmva.1994.1033

BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

BARINGHAUS, L. and FRANZ, C. (2004). On a new multivariate two-sample test. *J. Multivariate Anal.* **88** 190–206. MR2021870 https://doi.org/10.1016/S0047-259X(03)00079-4

BARINGHAUS, L. and HENZE, N. (2017). Cramér–von Mises distance: Probabilistic interpretation, confidence intervals, and neighbourhood-of-model validation. *J. Nonparametr. Stat.* **29** 167–188. MR3635009 https://doi.org/10.1080/10485252.2017.1285029

BERA, A. K., GHOSH, A. and XIAO, Z. (2013). A smooth test for the equality of distributions. *Econometric Theory* **29** 419–446. MR3042761 https://doi.org/10.1017/S0266466612000370

BERGSMA, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **20** 1006–1028. MR3178526 https://doi.org/10.3150/13-BEJ514

BHAT, B. V. (1995). Theory of U-statistics and its applications. Ph.D. thesis, Karnatak Univ.

BHATTACHARYA, B. B. (2018). Two-sample tests based on geometric graphs: Asymptotic distribution and detection thresholds. Preprint. Available at arXiv:1512.00384v3.

BHATTACHARYA, B. B. (2019). A general asymptotic framework for distribution-free graph-based two-sample tests. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 575–602. MR3961499

BISWAS, M. and GHOSH, A. K. (2014). A nonparametric two-sample test applicable to high dimensional data. *J. Multivariate Anal.* **123** 160–171. MR3130427 https://doi.org/10.1016/j.jmva.2013.09.004

BISWAS, M., MUKHOPADHYAY, M. and GHOSH, A. K. (2014). A distribution-free two-sample run test applicable to high-dimensional data. *Biometrika* **101** 913–926. MR3286925 https://doi.org/10.1093/biomet/asu045

CHAKRABORTY, A. and CHAUDHURI, P. (2017). Tests for high-dimensional data based on means, spatial signs and spatial ranks. *Ann. Statist.* **45** 771–799. MR3650400 https://doi.org/10.1214/16-AOS1467

CHEN, H., CHEN, X. and SU, Y. (2018). A weighted edge-count two-sample test for multivariate and object data. *J. Amer. Statist. Assoc.* **113** 1146–1155. MR3862346 https://doi.org/10.1080/01621459.2017.1307757

CHEN, L., DOU, W. W. and QIAO, Z. (2013). Ensemble subsampling for imbalanced multivariate two-sample tests. *J. Amer. Statist. Assoc.* **108** 1308–1323. MR3174710 https://doi.org/10.1080/01621459.2013.800763

CHEN, H. and FRIEDMAN, J. H. (2017). A new graph-based two-sample test for multivariate and object data. *J. Amer. Statist. Assoc.* **112** 397–409. MR3646580 https://doi.org/10.1080/01621459.2016.1147356

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697 https://doi.org/10.1214/09-AOS716

CHIKKAGOUDAR, M. S. and BHAT, B. V. (2014). Limiting distribution of two-sample degenerate U-statistic under contiguous alternatives and applications. *J. Appl. Statist. Sci.* **22** 127–139. MR3616873

CHUNG, E. and ROMANO, J. P. (2013). Exact and asymptotically robust permutation tests. *Ann. Statist.* **41** 484–507. MR3099111 https://doi.org/10.1214/13-AOS1090

CRAMÉR, H. (1928). On the composition of elementary errors. *Skand. Aktuarietidskr.* **11** 141–180.

CUI, H. (2002). Average projection type weighted Cramér–von Mises statistics for testing some distributions. *Sci. China Ser. A* **45** 562–577. MR1911172

ESCANCIANO, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22** 1030–1051. MR2328527 https://doi.org/10.1017/S0266466606060506

FRIEDMAN, J. H. and RAFSKY, L. C. (1979). Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests. *Ann. Statist.* **7** 697–717. MR0532236

GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716

HALL, P., MARRON, J. S. and NEEMAN, A. (2005). Geometric representation of high dimension, low sample size data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 427–444. MR2155347 https://doi.org/10.1111/j.1467-9868.2005.00510.x

HARCHAOUI, Z., BACH, F., CAPPE, O. and MOULINES, E. (2013). Kernel-based methods for hypothesis testing: A unified view. *IEEE Signal Process. Mag.* **30** 87–97.

HENZE, N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *Ann. Statist.* **16** 772–783. MR0947577 https://doi.org/10.1214/aos/1176350835

HETTMANSPERGER, T. P., MÖTTÖNEN, J. and OJA, H. (1998). Affine invariant multivariate rank tests for several samples. *Statist*. *Sinica* **8** 785–800. MR1651508

HOEFFDING, W. (1952). The large-sample power of tests based on permutations of observations. *Ann*. *Math*. *Stat*. **23** 169–192. MR0057521 https://doi.org/10.1214/aoms/1177729436

HU, J. and BAI, Z. (2016). A review of 20 years of naive tests of significance for high-dimensional mean vectors and covariance matrices. *Sci*. *China Math*. **59** 2281–2300. MR3578957 https://doi.org/10.1007/s11425-016-0131-0

KANAMORI, T., SUZUKI, T. and SUGIYAMA, M. (2012). f-divergence estimation and two-sample homogeneity test under semiparametric density-ratio models. *IEEE Trans*. *Inform*. *Theory* **58** 708–720. MR2917977 https://doi.org/10.1109/TIT.2011.2163380

KIM, I., BALAKRISHNAN, S. and WASSERMAN, L. (2020). Supplement to "Robust multivariate nonparametric tests via projection averaging." https://doi.org/10.1214/19-AOS1936SUPP.

KRUSKAL, J. B. JR. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc*. *Amer. Math*. *Soc*. **7** 48–50. MR0078686 https://doi.org/10.2307/2033241

LEE, A. J. (1990). *U-Statistics*: *Theory and Practice*. *Statistics*: *Textbooks and Monographs* **110**. Dekker, New York. MR1075417

LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927

LI, J. and CHEN, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann*. *Statist*. **40** 908–940. MR2985938 https://doi.org/10.1214/12-AOS993

LIU, R. Y. (2006). *Data Depth*: *Robust Multivariate Analysis*, *Computational Geometry*, *and Applications* **72**. Amer. Math. Soc., Providence.

LIU, Z. and MODARRES, R. (2011). A triangle test for equality of distribution functions in high dimensions. *J. Nonparametr. Stat*. **23** 605–615. MR2836279 https://doi.org/10.1080/10485252.2010.485644

LOPEZ-PAZ, D. and OQUAB, M. (2016). Revisiting classifier two-sample tests. Preprint. Available at arXiv:1610.06545.

MONDAL, P. K., BISWAS, M. and GHOSH, A. K. (2015). On high dimensional two-sample tests based on nearest neighbors. *J. Multivariate Anal*. **141** 168–178. MR3390065 https://doi.org/10.1016/j.jmva.2015.07.002

MUKHOPADHYAY, S. and WANG, K. (2018). A nonparametric approach to high-dimensional K-sample comparison problem. Preprint. Available at arXiv:1810.01724.

OJA, H. (2010). *Multivariate Nonparametric Methods with R*: *An Approach Based on Spatial Signs and Ranks*. *Lecture Notes in Statistics* **199**. Springer, New York. MR2598854 https://doi.org/10.1007/978-1-4419-0468-3

OJA, H. and RANDLES, R. H. (2004). Multivariate nonparametric tests. *Statist*. *Sci*. **19** 598–605. MR2185581 https://doi.org/10.1214/088342304000000558

PAN, W., TIAN, Y., WANG, X. and ZHANG, H. (2018). Ball divergence: Nonparametric two sample test. *Ann*. *Statist*. **46** 1109–1137. MR3797998 https://doi.org/10.1214/17-AOS1579

PESARIN, F. (2001). *Multivariate Permutation Tests*: *With Applications in Biostatistics*. Wiley, Chichester. MR1855501

RAMDAS, A., REDDI, S. J., POCZOS, B., SINGH, A. and WASSERMAN, L. (2015). Adaptivity and computation-statistics tradeoffs for kernel and distance based high dimensional two sample testing. Preprint. Available at arXiv:1508.00655.

ROSENBAUM, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *J. R. Stat*. *Soc*. *Ser. B. Stat*. *Methodol*. **67** 515–530. MR2168202 https://doi.org/10.1111/j.1467-9868.2005.00513.x

SCHILLING, M. F. (1986). Multivariate two-sample tests based on nearest neighbors. *J. Amer. Statist*. *Assoc*. **81** 799–806. MR0860514

SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann*. *Statist*. **41** 2263–2291. MR3127866 https://doi.org/10.1214/13-AOS1140

SZÉKELY, G. J. and RIZZO, M. L. (2004). Testing for equal distributions in high dimension. *Interstate* **5**.

SZÉKELY, G. J. and RIZZO, M. L. (2013). Energy statistics: A class of statistics based on distances. *J. Statist*. *Plann*. *Inference* **143** 1249–1272. MR3055745 https://doi.org/10.1016/j.jspi.2013.03.018

THAS, O. (2010). *Comparing Distributions*. *Springer Series in Statistics*. Springer, New York. MR2547894 https://doi.org/10.1007/b99044

WALD, A. and WOLFOWITZ, J. (1940). On a test whether two samples are from the same population. *Ann*. *Math*. *Stat*. **11** 147–162. MR0002083 https://doi.org/10.1214/aoms/1177731909

WANG, L., PENG, B. and LI, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J. Amer. Statist*. *Assoc*. **110** 1658–1669. MR3449062 https://doi.org/10.1080/01621459.2014.988215

ZHOU, W.-X., ZHENG, C. and ZHANG, Z. (2017). Two-sample smooth tests for the equality of distributions. *Bernoulli* **23** 951–989. MR3606756 https://doi.org/10.3150/15-BEJ766

ZHU, L.-X., FANG, K.-T. and BHATTI, M. I. (1997). On estimated projection pursuit-type Crámer–von Mises statistics. *J. Multivariate Anal*. **63** 1–14. MR1491563 https://doi.org/10.1006/jmva.1997.1673

ZHU, L., XU, K., LI, R. and ZHONG, W. (2017). Projection correlation between two random vectors. *Biometrika* **104** 829–843. MR3737307 https://doi.org/10.1093/biomet/asx043