Submitted to the Annals of Statistics

A NONPARAMETRIC TEST FOR ELLIPTICAL DISTRIBUTION BASED ON KERNEL EMBEDDING OF PROBABILITIES

BY YIN TANG[1,a] AND BING LI[1,b]

[1]Department of Statistics, The Pennsylvania State University,, [a]yqt5219@psu.edu; [b]bxl9@psu.edu

Elliptical distribution is a basic assumption underlying many multivariate statistical methods. For example, in sufficient dimension reduction and statistical graphical models, this assumption is routinely imposed to simplify the data dependence structure. Before applying such methods, we need to decide whether the data are elliptically distributed. Currently existing tests either focus exclusively on spherical distributions, or rely on bootstrap to determine the null distribution, or require specific forms of the alternative distribution. In this paper, we introduce a general nonparametric test for elliptical distribution based on kernel embedding of the probability measure that embodies the two properties that characterize an elliptical distribution: namely, after centering and rescaling, (1) the direction and length of the random vector are independent, and (2) the directional vector is uniformly distributed on the unit sphere. We derive the asymptotic distributions of the test statistic via von-Mises expansion, develop the sample-level procedure to determine the rejection region, and establish the consistency and validity of the proposed test. We also develop the concentration bounds of the test statistic, allowing the dimension to grow with the sample size, and further establish the consistency in this high-dimension setting. We compare our method with several existing methods via simulation studies, and apply our test to a SENIC dataset with and without a transformation aimed to achieve ellipticity.

1. Introduction. Elliptical distribution is a widely-used assumption for many statistical and machine learning methods for multivariate data. For example, in sufficient dimension reduction, the elliptical distribution assumption for the predictor is needed for moment-based methods such as Sliced Inverse Regression (Li, 1991), Ordinary Least Squares (Li and Duan, 1989), and Iterative Hessian Transformation (Cook and Li, 2002). See also Li (2018). When this assumption is violated, one needs either to perform data transformation or to modify the inverse-regression methods as in Li and Dong (2009). Another example is the statistical graphical model, where a class of methods, such as glasso (Yuan and Lin, 2006) and the transelliptical graphical model (Liu, Han and Zhang, 2012), require either a Gaussian or an elliptical distribution. See also Vogel and Fried (2011), which introduces a class of elliptical graphical models as a robust alternative to the Gaussian graphical model.

There are some existing tests on spherical distributions. For example, Baringhaus (1991) proposes a test based on the L^2-distance between the empirical distribution of the data and the distribution function partially specified by the spherical assumption. Liang, Fang and Hickernell (2008) introduces a necessary test by applying the Rosenblatt transformation on each element. The hypothesis in this test is necessary in the sense that it is implied by spherical distribution but does not imply the spherical distribution. Henze, Hlávka and Meintanis (2014) introduces a test by checking whether the characteristic function is constant over surfaces of spheres centered at the origin. Kariya and Eaton (1977) introduces a robust test of

MSC2020 subject classifications: Primary 62G10, 62G20; secondary 62H10.

Keywords and phrases: elliptical distribution, reproducing kernel Hilbert space, kernel embedding of probability, von-Mises expansion.

a spherical distribution centered at the origin against an elliptical or a noncentered spherical distribution. Koltchinskii and Li (1998) uses the multivariate distribution and quantile functions to test spherical distributions with unknown centers. All tests above are for the spherical distributions where the covariance matrix is exactly the identity matrix, and most of them focus on the case when we know the center is zero. These methods do not yield direct extensions for testing elliptical distributions where both the mean vector and covariance matrix are unknown.

There also exist some tests on elliptical distributions. Huffer and Park (2007) gives a test for multivariate normal and elliptical distribution of the data based on a chi-square statistic after slicing the data. For the multivariate normal distribution, they derive the asymptotic null distribution of their test; for the elliptical distribution, they propose a bootstrap test without giving a proof of its consistency and validity. Albisetti, Balabdaoui and Holzmann (2020) introduces a test based on a Kolmogorov-Smirnov type statistic and uses bootstrap to construct the null distribution. Manzotti, Pérez and Quiroz (2002) proposes a test on whether the standardized directional vector is uniformly distributed on the unit sphere, which is, again, only a necessary condition for spherical distribution. Schott (2002) proposes a Wald-type test based on whether the 4th moments are consistent with an elliptical distribution. So its null hypothesis is not exactly the elliptical distribution. Furthermore, since it requires estimation of the 8th moments, it may not be robust. Cassart, Hallin and Paindaveine (2008) proposes a locally and asymptotically optimal Pseudo-Gaussian test for Fechner-type symmetry, which is wider than the class of elliptical distributions. Babić et al. (2021) develops optimal tests for elliptical distributions against some generalized skew-elliptical alternatives. Some of the existing methods are summarized in Babić, Ley and Palangetić (2021).

In this paper, we introduce a nonparametric test for elliptical distributions based on Hilbert-space embedding of a product probability measure that characterizes an elliptical distribution. The basic idea is the following. It is well known that a random vector X follows a spherical distribution centered at 0 if and only if

1. its Euclidean norm $\|X\|$ and its direction $X/\|X\|$ are statistically independent;
2. the direction vector $X/\|X\|$ is uniformly distributed on the unit sphere.

See, for example, Anderson (2003); Paindaveine (2012). This converts testing of sphericity to testing of the two conditions. Since an elliptical distribution can always be transformed into a spherical distribution by a linear map, we can further develop tests of ellipticity by testing the two conditions for the linearly transformed data. However, since the mean vector and covariance matrix need to be estimated, we have to take into account the estimation error in this step when deriving the asymptotic distribution.

More specifically, let $U = \|X\|$ and $V = X/\|X\|$. If X has a spherical distribution, then the distribution $P_{U,V}$ can be expressed by the product measure $P_U \times P_V$, where P_V is a *known* distribution. We embed this distribution into a reproducing kernel Hilbert space as a cross-covariance operator and compare it against the kernel embedding of the fully empirical distribution. The norm of the difference should be small if X has a spherical distribution, and large otherwise. This is the core idea of our method. One side-note is that we may replace V by its polar coordinate representation, which will significantly simplify the computation. This procedure has several appealing features: first, the hypothesis is both necessary and sufficient, that is, X has a spherical distribution if and only if the distance is small; second, it is rather straightforward to go from spherical distribution to elliptical distribution by replacing X with its centered and rescaled version; third, since the test only involves functions of sample moments, its asymptotic null distribution can be relatively easily derived from the infinite-dimensional δ-method, or the von-Mises expansion.

Probability embedding (Sriperumbudur, Fukumizu and Lanckriet, 2010, 2011) is a powerful method that has been used in a variety of settings in statistics and machine learning, such as test of independence and the two-sample problem. See, for example, Gretton et al. (2005), Gretton et al. (2007), Gretton et al. (2008), Gretton et al. (2009), and Gretton et al. (2012). There is another type of tests of independence based on the distance covariance; see Székely, Rizzo and Bakirov (2007) and Székely and Rizzo (2009) among others. Sejdinovic et al. (2013) establishes the relation between these two types of tests of independence. Our test of elliptical distribution goes beyond the test of independence between U and V, as it must also incorporate the fact that the distribution of V is known.

The rest of the paper is organized as follows. In Section 2, we lay out the two characterizing properties of an elliptical distribution and construct the probability embedding into a reproducing kernel Hilbert space that embodies the two characterizing properties. In Sections 3 and 4, we introduce the test statistic based on the probability embedding and implement it numerically through coordinate mapping. In Section 5, for the fixed dimension, we derive the asymptotic null and alternative distributions of the test statistic via von-Mises expansion, and in Section 6, we implement the asymptotic null distribution at the sample level, and establish the validity and consistency of our test. In Section 7, we derive the uniform concentration bounds for our test statistic allowing the dimension to grow with the sample size, and further establish the consistency of our test in this high-dimensional setting. In Section 8, we conduct simulation studies to demonstrate the usage and effectiveness of the proposed test. In Section 9, we apply our test to a data example. Section 10 is devoted to some discussions on the choice of kernel functions.

Due to space limit, some technical lemmas, most of the proofs, and discussions of some theoretical results are placed in Appendix A; additional simulation comparisons are placed in Appendices B and C; further discussions on the choice of kernel functions are placed in Appendix D; and the scatter plot matrices of the dataset in Section 9 are placed in Appendix E. All Appendices are in the online Supplementary Material (Tang and Li (2024)). Our proposed method is implemented in the R package `KEPTED` (Kernel-Embedding-of-Probability Test for Elliptical Distribution).

2. Elliptical distribution and its kernel embedding.

In this section we introduce the definition of the elliptical distribution and develop two equivalent conditions, one at the level of probability measures and the other at the level of linear operators. The sufficient condition at the operator level is the theoretical basis of our test.

2.1. *Spherical and elliptical distributions.*

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $(\Omega_X, \mathcal{F}_X)$ be a measurable space, where Ω_X is a subset of $\mathbb{R}^d$, and $\mathcal{F}_X$ is the Borel σ-field on Ω_X. Let $X : \Omega \to \Omega_X$ be a Borel random vector, and let $P_X = P \circ X^{-1}$ be the distribution of X. Denote λ the Lebesgue measure in $\mathbb{R}^d$. We say that X has a spherical distribution centered at $\mu \in \mathbb{R}^d$ if

1. P_X is dominated by the Lebesgue measure in $\mathbb{R}^d$ with density $f_X = dP_X/d\lambda$;
2. $f_X(x)$ is a function of $(x-\mu)^\intercal(x-\mu)$, that is,

$$f_X(x) = h\left((x-\mu)^\intercal(x-\mu)\right), \tag{2.1}$$

for some nonnegative function h satisfying

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x^\intercal x)dx_1 \ldots dx_d = 1. \tag{2.2}$$

Here, μ is necessarily EX if X is integrable. Throughout, $\|\cdot\|$ will denote the Euclidean norm (or L_2 norm). Let $U = \|X - \mu\|$ and $V = (X-\mu)/U$. By Theorem 1 of Cambanis, Huang and Simons (1981), X has a spherical distribution if and only if $U \perp\!\!\!\perp V$ and V has a uniform distribution on the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$, where $\perp\!\!\!\perp$ refers to independent throughout this paper. See also Eaton (1986) and Schmidt (2002).

More generally, X is said to have an elliptical distribution centered at $\mu \in \mathbb{R}^d$ with a positive definite shape parameter $\Lambda \in \mathbb{R}^{d\times d}$ if

1. P_X is dominated by the Lebesgue measure in $\mathbb{R}^d$ with density $f_X = dP_X/d\lambda$;
2. the density of X is a function of $(x-\mu)^\intercal\Lambda^{-1}(x-\mu)$, that is,
$$f_X(x) = |\Lambda|^{-\frac{1}{2}} h\left((x-\mu)^\intercal\Lambda^{-1}(x-\mu)\right),$$
for some nonnegative function h satisfying (2.2), where $|\Lambda|$ denotes the determinant of Λ.

A direct corollary is that $Y = \Lambda^{-\frac{1}{2}}(X-\mu)$ has a spherical distribution centered at 0. Consequently, if we let

$$U = \sqrt{(X-\mu)^\intercal\Lambda^{-1}(X-\mu)}, \quad V = \Lambda^{-1/2}(X-\mu)/U, \tag{2.3}$$

then X has an elliptical distribution with center μ and shape parameter Λ if and only if $U \perp\!\!\!\perp V$, and V has a uniform distribution on the unit sphere $\mathbb{S}^{d-1}$.

According to Theorem 2.7.2 of Anderson (2003), if the components of X are square-integrable, then

$$EX = \mu, \quad \mathrm{var}(X) = \frac{EU^2}{d}\Lambda.$$

Let $\Sigma = \mathrm{var}(X)$. In this paper, we always assume that X has finite mean and variance. Obviously, neither the dependence between U and V nor the distribution of V will be affected if we replace Λ by Σ in their definitions in (2.3). See Paindaveine (2012) for a detailed discussion. So, for convenience, we reset U and V to be

$$U = \sqrt{(X-\mu)^\intercal\Sigma^{-1}(X-\mu)}, \quad V = \Sigma^{-1/2}(X-\mu)/U, \tag{2.4}$$

for the rest of the paper. The sufficient and necessary condition for elliptical distribution still applies to the redefined U and V, which we record below formally for easy reference.

PROPOSITION 1. *A random vector X has an elliptical distribution if and only if, for U and V defined in (2.4),*

1. *$U \perp\!\!\!\perp V$;*
2. *V is uniformly distributed in $\mathbb{S}^{d-1}$.*

2.2. *Polar coordinate transformation and equivalent condition.* Since the random vector V only takes values in the unit sphere $\mathbb{S}^{d-1}$ in $\mathbb{R}^d$, it is more convenient to transform it into a $d-1$ dimensional vector representing the direction of the unit vector V via the polar coordinate system. Specifically, let $v = (v_1, \ldots, v_d)^\intercal \in \mathbb{S}^{d-1}$ and let

$$\theta = (\theta_1, \ldots, \theta_{d-1})^\intercal \in (-\pi/2, \pi/2] \times \cdots \times (-\pi/2, \pi/2] \times (-\pi, \pi] \equiv \Omega_\Theta.$$

For clarity of the subsequent discussion, we first give a definition of the precise meaning of the arc tangent function in the range of $(-\pi, \pi]$. For a real number $r \in \mathbb{R}$, let $\arctan(r)$ be the unique $\theta \in (-\pi/2, \pi/2)$ such that $r = \tan(\theta)$. Then, for $(x, y) \in \mathbb{R}^2$, let

$$\mathrm{Arctan}(x,y) = \begin{cases} \arctan(y/x), & \text{if } x > 0, \\ \arctan(y/x) + \pi, & \text{if } x < 0, y \geq 0, \\ \arctan(y/x) - \pi, & \text{if } x < 0, y < 0, \\ \pi/2, & \text{if } x = 0,\ y > 0, \\ -\pi/2, & \text{if } x = 0,\ y < 0. \end{cases} \tag{2.5}$$

We have written this function as $\mathrm{Arctan}(x, y)$ instead of $\mathrm{Arctan}(y/x)$ because it is no longer the function of the ratio y/x. Some useful properties of the Arctan function (2.5) are summarized in the Supplementary Materials. The next lemma, whose proof is also placed in the Supplementary Materials, gives the explicit one-to-one correspondence between $v \in \mathbb{S}^{d-1}$ and $\theta \in \Omega_\Theta$. In the following, we use S_j to denote the Euclidean norm of the vector $(v_j, \ldots, v_d)^\intercal$.

LEMMA 1. *The following function from Ω_Θ to $\mathbb{S}^{d-1}$*

$$\begin{aligned} v_1 &= \sin\theta_1, \\ v_2 &= \cos\theta_1 \sin\theta_2, \\ &\vdots \\ v_{d-1} &= \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{d-2} \sin\theta_{d-1}, \\ v_d &= \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{d-2} \cos\theta_{d-1}. \end{aligned} \tag{2.6}$$

is bijective with inverse

$$\theta_j = \begin{cases} \mathrm{Arctan}(S_{j+1}, v_j) & j = 1, \ldots, d-2, \\ \mathrm{Arctan}(v_d, v_{d-1}) & j = d-1. \end{cases} \tag{2.7}$$

We will denote the first function (2.6) as $v = \rho(\theta)$ and the second function (2.7) as $\theta = g(v)$. Then

$$x = uv = u\rho(\theta) \equiv \tau(u, \theta),$$

Evidently τ is an invertible function, and the joint distribution of U, Θ can be written as

$$f_{U\Theta}(u, \theta) = f_X(\tau(u, \theta)) \left| \tfrac{\partial \tau(u,\theta)}{\partial (u, \theta^\intercal)} \right|. \tag{2.8}$$

According to Anderson (2003), the Jacobian on the right-hand side above is

$$\left| \tfrac{\partial \tau(u,\theta)}{\partial (u, \theta^\intercal)} \right| = u^{d-1} \cos^{d-2}\theta_1 \cos^{d-3}\theta_2 \ldots \cos\theta_{d-2}. \tag{2.9}$$

Hence, by (2.1), (2.8) and (2.9), the joint p.d.f. of (U, Θ) is

$$f_{U,\Theta}(u, \theta) = u^{d-1} \cos^{d-2}\theta_1 \cos^{d-3}\theta_2 \ldots \cos\theta_{d-2} h(u^2), \quad (u, \theta) \in \Omega_U \times \Omega_\Theta, \tag{2.10}$$

where $\Omega_U = [0, \infty)$. This relation implies (i) U and Θ are independent, and (ii) Θ has a known distribution. We summarize this result as the following proposition.

PROPOSITION 2. *If $(U, \Theta) = \tau^{-1}(X)$, then X has a spherical distribution centered at 0 if and only if*

1. *$U \perp\!\!\!\perp \Theta$, or equivalently $P_{U,\Theta} = P_U \times P_\Theta$;*
2. *P_Θ has p.d.f. $f_\Theta(\theta) = c \cos^{d-2}(\theta_1) \cos^{d-3}(\theta_2) \cdots \cos(\theta_{d-2})$, where $\theta \in \Omega_\Theta$ and*

$$c = \left(\int_{\Omega_\Theta} \cos^{d-2}(\theta_1) \cos^{d-3}(\theta_2) \cdots \cos(\theta_{d-2})\, d\theta \right)^{-1}.$$

2.3. *Kernel embedding of $P_U \times P_\Theta$.* Let $\kappa_U : \Omega_U \times \Omega_U \to \mathbb{R}$ and $\kappa_\Theta : \Omega_\Theta \times \Omega_\Theta \to \mathbb{R}$ be positive definite kernels, and let $\mathcal{H}_U$ and $\mathcal{H}_\Theta$ be the reproducing kernel Hilbert space (RKHS) generated by κ_U and κ_Θ. Let $\mathcal{B}$ be the collection of linear operators $\{f \otimes g : f \in \mathcal{H}_U, g \in \mathcal{H}_\Theta\}$. Thus, each member of $\mathcal{B}$ is a linear operator mapping from $\mathcal{H}_\Theta$ to $\mathcal{H}_U$ such that, for any

$h \in \mathcal{H}_\Theta$, $(f \otimes g)(h) = f\langle g, h\rangle_{\mathcal{H}_\Theta}$. Let $\mathcal{S}$ be the linear span of $\mathcal{B}$, consisting of finite linear combinations of members of $\mathcal{B}$ with real coefficients. Define in $\mathcal{S}$ the inner product

$$\langle \alpha_1(f_1 \otimes g_1) + \cdots + \alpha_r(f_r \otimes g_r), \tilde{\alpha}_1(\tilde{f}_1 \otimes \tilde{g}_1) + \cdots + \tilde{\alpha}_r(\tilde{f}_s \otimes \tilde{g}_s)\rangle$$
$$= \textstyle\sum_{i=1}^r \sum_{j=1}^s \alpha_i \tilde{\alpha}_j \langle f_i, \tilde{f}_i\rangle_{\mathcal{H}_U} \langle g_j, \tilde{g}_j\rangle_{\mathcal{H}_\Theta}.$$

Endowed with this inner product, $\mathcal{S}$ is an inner product space; its completion as a Hilbert space is the tensor product space $\mathcal{H}_U \otimes \mathcal{H}_\Theta$.

Let $\mathcal{F}_U$ and $\mathcal{F}_\Theta$ be the Borel σ-fields on Ω_U and Ω_Θ, and let $\mathcal{F}_U \times \mathcal{F}_\Theta$ be the product σ-field. Abbreviate $\Omega_U \times \Omega_\Theta$ and $\mathcal{F}_U \times \mathcal{F}_\Theta$ by $\Omega_{U,\Theta}$ and $\mathcal{F}_{U,\Theta}$. Let $\mathcal{M}(\Omega_{U,\Theta}, \mathcal{F}_{U,\Theta})$ denote the class of all probability measures on $(\Omega_{U,\Theta}, \mathcal{F}_{U,\Theta})$. We want to find an injective mapping from $\mathcal{M}(\Omega_{U,\Theta}, \mathcal{F}_{U,\Theta})$ to $\mathcal{H}_U \otimes \mathcal{H}_\Theta$ so that testing equality of two measures in $\mathcal{M}(\Omega_{U,\Theta}, \mathcal{F}_{U,\Theta})$ is equivalent to testing the equality of two operators in $\mathcal{H}_U \otimes \mathcal{H}_\Theta$. Such a mapping is provided by the following theorem. Recall that a kernel κ is characteristic if the mapping $P \mapsto \int \kappa(\cdot, X)dP$ is injective. Here, the dot notation $\int \kappa(\cdot, X)dP$ simply means the function $x \mapsto \int \kappa(x, X)dP$. So κ being characteristic means if $\int \kappa(x, X)dP_1 = \int \kappa(x, X)dP_2$ for all x then $P_1 = P_2$.

THEOREM 1. *If κ_U and κ_Θ are characteristic, then so is tensor product kernel $\kappa_U \otimes \kappa_\Theta$; that is, the mapping*

$$\mathcal{M}(\Omega_{U,\Theta}, \mathcal{F}_{U,\Theta}) \to \mathcal{H}_U \otimes \mathcal{H}_\Theta, \quad P_{U,\Theta} \mapsto \int_{\Omega_{U,\Theta}} \kappa_U(\cdot, U) \otimes \kappa_\Theta(\cdot, \Theta) dP_{U,\Theta}$$

is injective.

The dot notation $\kappa_U(\cdot, U)$ means the function $u \mapsto \kappa_U(u, U)$; the same applies to $\kappa_\Theta(\cdot, \Theta)$. So the operator $\kappa_U(\cdot, U) \otimes \kappa_\Theta(\cdot, \Theta)$ maps a function $f \in \mathcal{H}_\Theta$ to $f(\Theta)\kappa_U(\cdot, U)$, which is a member of $\mathcal{H}_U$. Correspondingly, the operator $\int_{\Omega_{U,\Theta}} \kappa_U(\cdot, U) \otimes \kappa_\Theta(\cdot, \Theta) dP_{U,\Theta}$ maps a function $f \in \mathcal{H}_\Theta$ to $\int_{\Omega_{U,\Theta}} \kappa_U(\cdot, U) f(\Theta) dP_{U,\Theta}$, which is a member of $\mathcal{H}_U$. This type of dot notations will be used throughout the rest of the paper. The proof of Theorem 1 can be found in Theorem 4 of Szabó and Sriperumbudur (2018). As a result, we only need to guarantee that both κ_U and κ_Θ are characteristic kernels. In fact, the Gaussian radial basis kernels, which in our context are

$$\kappa_U(u_1, u_2) = \exp[-\gamma_U(u_1 - u_2)^2], \quad \kappa_\Theta(\theta_1, \theta_2) = \exp(-\gamma_\Theta \|\theta_1 - \theta_2\|^2), \tag{2.11}$$

for some $\gamma_U, \gamma_\Theta > 0$, indeed satisfy the condition in Theorem 1: it is shown in Theorem 3.2 of Guella (2022) that a Gaussian radial basis kernel is integrally strictly positive definite, and it is further shown that an integrally strictly positive definite kernel is characteristic (see, for example, Sriperumbudur, Fukumizu and Lanckriet (2011); Fukumizu et al. (2009); Sriperumbudur et al. (2010)).

Theorem 1 implies the following equivalence which is the basis of our test of ellipticity.

COROLLARY 1. *Suppose*

1. *X is a random vector in $\mathbb{R}^d$ with mean μ and covariance matrix Σ;*
2. *$(U, \Theta) = \tau^{-1}(\Sigma^{-1/2}(X - \mu))$;*
3. *κ_U and κ_Θ are characteristic kernels.*

Then X has an elliptical distribution with parameters μ and Σ if and only if

$$\int_{\Omega_{U,\Theta}} [\kappa_U(\cdot, U) \otimes \kappa_\Theta(\cdot, \Theta)] dP_{U,\Theta} = \int_{\Omega_U} \kappa_U(\cdot, U) dP_U \otimes \int_{\Omega_\Theta} \kappa_\Theta(\cdot, \Theta) dP_0$$

where P_0 is the known true P_Θ with its form given in Proposition 2.

3. Construction of Test Statistic. In this section we construct our test statistic based on Corollary 1. Let $X_1, \ldots, X_n$ be an i.i.d. sample of X. For a function $A(x)$, we use $E_n A(X)$ to denote the sample average of $A(X_1), \ldots, A(X_n)$. Let $\hat{\mu}$ and $\hat{\Sigma}$ denote the sample mean and sample variance; that is, $\hat{\mu} = E_n(X)$ and $\hat{\Sigma} = E_n[(X - \hat{\mu})(X - \hat{\mu})^\intercal]$. For any $x \in \Omega_X$, let

$$\begin{aligned} W(x) &= \Sigma^{-1/2}(x - \mu), & \hat{W}(x) &= \hat{\Sigma}^{-1/2}(x - \hat{\mu}), \\ U(x) &= \|W(x)\|, & \hat{U}(x) &= \|\hat{W}(x)\|, \\ V(x) &= W(x)/U(x), & \hat{V}(x) &= \hat{W}(x)/\hat{U}(x), \\ \Theta(x) &= g(V(x)), & \hat{\Theta}(x) &= g(\hat{V}(x)), \end{aligned}$$

where $g(\cdot)$ is the polar coordinate transformation as defined in (2.6). When no ambiguity is likely, we abbreviate $W(X)$ by W, $W(X_i)$ by W_i, and $\hat{W}(X_i)$ by $\hat{W}_i$. The same applies to U, V and Θ. Let

$$\Sigma_{U\Theta} = E[\kappa_U(\cdot, U) \otimes \kappa_\Theta(\cdot, \Theta)] - E[\kappa_U(\cdot, U)] \otimes E_0[\kappa_\Theta(\cdot, \Theta)] \tag{3.1}$$

where E_0 refers to the expectation with respect to P_0. Note that $\Sigma_{U\Theta}$ is different from a usual cross-covariance operator between two random variables, because P_Θ is given the specific form P_0. By Corollary 1, X has an elliptical distribution if and only if $\Sigma_{U\Theta} = 0$, which implies $U \perp\!\!\!\perp \Theta$ and P_Θ has the form given in Proposition 2. Thus, our goal is to test the hypothesis

$$H_0 : \Sigma_{U\Theta} = 0. \tag{3.2}$$

Note that this is not merely a test of independence between U and Θ, because P_Θ has a known, specific form.

Let

$$\breve{\Sigma}_{U\Theta} = E_n[\kappa_U(\cdot, \hat{U}) \otimes \kappa_\Theta(\cdot, \hat{\Theta})] - E_n[\kappa_U(\cdot, \hat{U})] \otimes E_0[\kappa_\Theta(\cdot, \Theta)], \tag{3.3}$$

where, for example, the first term on the right is the sample average of

$$\{\kappa_U(\cdot, \hat{U}(X_i)) \otimes \kappa_\Theta(\cdot, \hat{\Theta}(X_i)) : i = 1, \ldots, n\}.$$

Since the last term is an integral with respect to the known distribution P_0, this operator is *not* the usual sample estimate of the cross-covariance operator $\Sigma_{U\Theta}$. That is why we denote it by $\breve{\Sigma}_{U\Theta}$ instead of $\hat{\Sigma}_{U\Theta}$. The operator $\breve{\Sigma}_{U\Theta}$ is a mapping from $\mathcal{H}_\Theta$ to $\mathcal{H}_U$: for a given $f \in \mathcal{H}_\Theta$, $\breve{\Sigma}_{U\Theta} f$ is the function

$$E_n[\kappa_U(\cdot, \hat{U}) f(\hat{\Theta})] - E_n[\kappa_U(\cdot, \hat{U})] E_0[f(\Theta)],$$

which is a member of $\mathcal{H}_U$.

For convenience, let $\tilde{\kappa}_\Theta(\cdot, \theta) = \kappa_\Theta(\cdot, \theta) - \int \kappa_\Theta(\cdot, \theta) dP_0(\theta)$ denote the centered kernel function in $\mathcal{H}_\Theta$. Then, (3.1) and (3.3) can be simplified as

$$\Sigma_{U\Theta} = E[\kappa_U(\cdot, U) \otimes \tilde{\kappa}_\Theta(\cdot, \Theta)], \quad \breve{\Sigma}_{U\Theta} = E_n[\kappa_U(\cdot, \hat{U}) \otimes \tilde{\kappa}_\Theta(\cdot, \hat{\Theta})].$$

We use

$$T_n = n \|\breve{\Sigma}_{U\Theta}\|^2_{\mathrm{HS}}$$

as our test statistic for the hypothesis (3.2). The squared Hilbert-Schmidt norm $\|\breve{\Sigma}_{U\Theta}\|^2_{\mathrm{HS}}$ is the sum of eigenvalues of the operator $\breve{\Sigma}_{U\Theta}\breve{\Sigma}^*_{U\Theta}$, where $\breve{\Sigma}^*_{U\Theta}$ is the adjoint operator of $\breve{\Sigma}_{U\Theta}$. This is analogous to the Frobenius norm of a matrix.

By our definition, T_n/n is close to $\|\Sigma_{U\Theta}\|_{\text{HS}}^2$. An equivalent definition of $\|\Sigma_{U\Theta}\|_{\text{HS}}$ is

$$\begin{aligned}\|\Sigma_{U\Theta}\|_{\text{HS}} = & \quad \|E[\kappa_U(\cdot,U)\otimes\kappa_\Theta(\cdot,\Theta)] - E[\kappa_U(\cdot,U)]\otimes E_0[\kappa_\Theta(\cdot,\Theta)]\|_{\text{HS}} \\ = & \sup_{f_1\otimes f_2\in\mathcal{F}} |E_{P_{U\Theta}}[f_1(U)f_2(\Theta)] - E_{P_U}[f_1(U)]E_{P_0}[f_2(\Theta)]|\end{aligned}$$

where $\mathcal{F} = \{f_1\otimes f_2 \in \mathcal{H}_U\otimes\mathcal{H}_\Theta : \|f_1\otimes f_2\| \le 1\}$, and P_0 is the true distribution of Θ as defined in Proposition 2. In this way, we can interpret $\|\Sigma_{U\Theta}\|_{\text{HS}}$ as a "distance" between $P_{U\Theta}$ and the "closest" elliptically symmetric distribution $P_U \times P_0$. Here, "closest" can be interpreted as keeping the marginal distribution P_U unchanged and using it to construct an elliptical distribution. Therefore, intuitively, T_n/n is small if X has an elliptical distribution. On the other hand, if X does not have an elliptical distribution, then either U and Θ are not independent, or the marginal distribution of Θ is not the one given by Proposition 2. In either case T_n/n will not be small.

As we will see in the next section, T_n/n can be re-expressed as

$$n^{-2}\textstyle\sum_{i,j}\kappa_U(\hat U_i,\hat U_j)\langle\tilde\kappa_\Theta(\cdot,\hat\Theta_i),\tilde\kappa_\Theta(\cdot,\hat\Theta_j)\rangle_{\mathcal{H}_\Theta}.$$

If U and Θ are independent, then the above is approximately

$$\left(n^{-2}\textstyle\sum_{i,j}\kappa_U(\hat U_i,\hat U_j)\right)\left(n^{-2}\textstyle\sum_{i,j}\langle\tilde\kappa_\Theta(\cdot,\hat\Theta_i),\tilde\kappa_\Theta(\cdot,\hat\Theta_j)\rangle_{\mathcal{H}_\Theta}\right).$$

If, furthermore, Θ is distributed as P_0, then the second term will be near 0. Hence T_n/n will be near 0 if and only if both $\Theta\perp\!\!\!\perp U$ and $\Theta\sim P_0$ hold; that is, X has an elliptical distribution.

4. Computing the test statistic.

4.1. *Coordinate mapping.* The implementation of the test at the sample level relies on coordinate representation of linear operators. Let $\mathcal{H}$ be an r-dimensional space with a basis $\mathcal{B} = \{b_1,\dots,b_r\}$. Then every member f of $\mathcal{H}$ can be represented as $c_1b_1+\cdots+c_rb_r$. The mapping from $\mathcal{H}$ to $\mathbb{R}^r$ defined by $C(f) = (c_1,\dots,c_r)^\intercal$ is called the coordinate mapping. Let G be the Gram matrix $\{\langle b_i,b_j\rangle_{\mathcal{H}}\}_{i,j=1}^r$. Then, for any $f,g\in\mathcal{H}$,

$$\langle f,g\rangle_{\mathcal{H}} = \textstyle\sum_{i=1}^r\sum_{j=1}^r C(f)_iC(g)_j\langle b_i,b_j\rangle_{\mathcal{H}} = C(f)^\intercal GC(g).$$

In other words, if we let $\mathbb{R}^r(G)$ represent the Hilbert space consisting of the vector space $\mathbb{R}^r$ along with the inner product $\langle a,b\rangle = a^\intercal Gb$, then $C:\mathcal{H}\to\mathbb{R}^r(G)$ is an isomorphism. Let $A:\mathcal{H}\to\mathcal{H}$ be a self-adjoint operator. An eigenvalue of A is defined by the following relations

$$Af = \lambda f, \quad \langle f,f\rangle_{\mathcal{H}} = 1,$$

or equivalently,

$$CAC^*Cf = \lambda Cf, \quad (Cf)^\intercal GCf = 1,$$

where, C^* stands for the adjoint operator of C. Note, again, that $*$ and $\star$ denote different concepts. Letting $v = G^{1/2}Cf$, the above can be restated as

$$G^{1/2}CAC^*G^{-1/2}v = \lambda v, \quad v^\intercal v = 1.$$

In other words, a number is an eigenvalue of the operator A if and only if it is an eigenvalue of the matrix $G^{1/2}CAC^*G^{-1/2}$, which can be shown to be a symmetric matrix. In particular,

$$\text{tra}(A) = \text{tra}(G^{1/2}CAC^*G^{-1/2}) = \text{tra}(CAC^*),$$

where tra on the left represents the trace of a linear operator, in the middle and on the right represent that of a matrix. This identity allows us to express the Hilbert-Schmidt norm of an operator as the trace of a matrix.

Let e_i represent the ith column of the identity matrix I_r. Since $e_i = Cb_i$, CAC^* is simply the $n \times n$ matrix

$$CAC^*(e_1, \ldots, e_p) = CAC^*(Cb_1, \ldots, Cb_p) = (CAb_1, \ldots, CAb_r). \tag{4.1}$$

4.2. *Test statistic.* At the sample level, $\mathcal{H}_U$ and $\mathcal{H}_\Theta$ are spaces spanned by the two bases

$$\mathcal{B}_U = \{\kappa_U(\cdot, \hat{U}_i) : i = 1, \ldots, n\}, \quad \mathcal{B}_\Theta = \{\kappa_\Theta(\cdot, \hat{\Theta}_i) : i = 1, \ldots, n\}$$

respectively. Let K_U and $\tilde{K}_\Theta$ be the $n \times n$ matrices whose (i, j)th entries are

$$(K_U)_{ij} = \kappa_U(\hat{U}_i, \hat{U}_j), \quad (\tilde{K}_\Theta)_{ij} = \langle \tilde{\kappa}_\Theta(\cdot, \hat{\Theta}_i), \tilde{\kappa}_\Theta(\cdot, \hat{\Theta}_j)\rangle_{\mathcal{H}_\Theta} \tag{4.2}$$

respectively. Note that K_U is simply the Gram matrix of $\mathcal{B}_U$, and $(\tilde{K}_\Theta)_{ij}$ can be expanded as

$$\kappa_\Theta(\hat{\Theta}_i, \hat{\Theta}_j) - \int \kappa_\Theta(\hat{\Theta}_i, \theta)dP_0(\theta) - \int \kappa_\Theta(\hat{\Theta}_j, \theta)dP_0(\theta) + \int\int \kappa_\Theta(\theta, \theta')d(P_0 \times P_0)(\theta, \theta').$$

Let $C : \mathcal{H}_U \to \mathbb{R}^n$ be the coordinate mapping. Our goal is to compute

$$T_n = n\|\breve{\Sigma}_{U\Theta}\|^2_{\mathrm{HS}} = n\,\mathrm{tra}(\breve{\Sigma}_{U\Theta}\breve{\Sigma}^*_{U\Theta}).$$

By the discussion in Section 4.1, we have

$$\mathrm{tra}(\breve{\Sigma}_{U\Theta}\breve{\Sigma}^*_{U\Theta}) = \mathrm{tra}(C\breve{\Sigma}_{U\Theta}\breve{\Sigma}^*_{U\Theta}C^*).$$

The next proposition gives the coordinate of $\breve{\Sigma}_{U\Theta}\breve{\Sigma}^*_{U\Theta}$.

PROPOSITION 3. *If K_U and $\tilde{K}_\Theta$ are the matrices defined in (4.2), then*

$$C\breve{\Sigma}_{U\Theta}\breve{\Sigma}^*_{U\Theta}C^* = \frac{1}{n^2}\tilde{K}_\Theta K_U.$$

Let $\odot$ denote Hadamard product between matrices, and let 1_n be the n-dimensional vector with all entries equal to 1, then we have the following alternative expression for $\|\breve{\Sigma}_{U\Theta}\|^2_{\mathrm{HS}}$:

$$\|\breve{\Sigma}_{U\Theta}\|^2_{\mathrm{HS}} = \frac{1}{n^2}1_n^{\intercal}(K_U \odot \tilde{K}_\Theta)1_n.$$

The computation of K_U is straightforward. However, for computing $\tilde{K}_\Theta$, we need

$$\int \kappa_\Theta(\hat{\Theta}_i, \theta)dP_0(\theta), \quad \int\int \kappa_\Theta(\theta, \theta')d(P_0 \times P_0)(\theta, \theta').$$

We propose to compute these by numerical integration. By Proposition 2, the $d-1$ components of Θ, namely, $\Theta_1, \ldots, \Theta_{d-1}$, are independent with densities

$$f_{\Theta_j}(\theta_j) = \begin{cases} \cos^{d-1-j}(\theta_j)/\int_{-\pi/2}^{\pi/2}\cos^{d-1-j}(\theta_j)d\theta_j, & \theta_j \in (-\pi/2, \pi/2],\ j = 1, \ldots, d-2 \\ 1/(2\pi), & \theta_j \in (-\pi, \pi],\ j = d-1. \end{cases}$$

Let

$$\Omega_{\Theta_j} = \begin{cases} (-\pi/2, \pi/2], & j = 1, \ldots, d-2, \\ (-\pi, \pi], & j = d-1. \end{cases}$$

If we choose κ_Θ to be the product kernel,

$$\kappa_\Theta(\theta, \theta') = \prod_{j=1}^{d-1}\kappa_{\Theta j}(\theta_j, \theta'_j), \tag{4.3}$$

of which a typical example is the Gaussian radial basis kernel as in (2.11), then we have

$$\int_{\Omega_\Theta} \kappa_\Theta(\hat{\Theta}_i, \theta) dP_0(\theta) = \prod_{j=1}^{d-1} \int_{\Omega_{\Theta_j}} \kappa_{\Theta_j}(\hat{\Theta}_{ij}, \theta_j) f_{\Theta_j}(\theta_j) d\theta_j, \tag{4.4}$$

$$\int_{\Omega_\Theta \times \Omega_\Theta} \kappa_\Theta(\theta, \theta') d(P_0 \times P_0)(\theta, \theta') = \prod_{j=1}^{d-2} \int_{\Omega_{\Theta_j} \times \Omega_{\Theta_j}} \kappa_{\Theta_j}(\theta_j, \theta'_j) f_{\Theta_j}(\theta_j) f_{\Theta_j}(\theta'_j) d\theta_j d\theta'_j, \tag{4.5}$$

where, in (4.4), $\hat{\Theta}_{ij}$ means the j-th element of $\hat{\Theta}_i$.

The quantities in (4.4) can be computed by the function `integrate` in R, and the quantity in (4.5) can be computed by the function `cubintegrate` in the R package `cubature` (Narasimhan et al. (2023)).

5. Asymptotic distribution. In this section we derive the asymptotic distribution of T_n under the null hypothesis (3.2). We use the von-Mises expansion to achieve this purpose; see, for example, van der Vaart (1998); Fernholz (1983); Li (2018). We first outline the key steps and notations for the von-Mises expansion, tailored for our current application.

5.1. *von-Mises expansion.* Let $\mathfrak{F}$ denote the class of all distributions on $(\Omega_X, \mathcal{F}_X)$. Let $\mathcal{H}$ be a generic Hilbert space. Let $T: \mathfrak{F} \to \mathcal{H}$ be a mapping — such mappings are known as statistical functionals. Endow $\mathfrak{F}$ with the uniform metric, and $\mathcal{H}$ with the metric induced by its inner product. Let F_0 be a member of $\mathfrak{F}$. Then T is Frechet differentiable at F_0 if there is a linear operator $A: \mathfrak{F} \to \mathcal{H}$ such that

$$\lim_{\|F-F_0\|_{\mathfrak{F}} \to 0} \frac{\|T(F) - T(F_0) - A(F - F_0)\|_{\mathcal{H}}}{\|F - F_0\|_{\mathfrak{F}}} = 0.$$

Under the Frechet differentiability, the linear operator A can be calculated using Gateaux derivative: for any $F \in \mathfrak{F}$, $A(F)$ is simply

$$\lim_{\epsilon \to 0} \frac{T((1-\epsilon)F_0 + \epsilon F) - T(F_0)}{\epsilon}.$$

Let x be a member of Ω_X, and δ_x be the Dirac measure at x. Then the mapping $x \mapsto A(\delta_x)$ from Ω_X to $\mathcal{H}$ is called influence function of T. We write $A(\delta_x)$ as $T^\star(x)$. Note that we use $\star$ to indicate influence function and $*$ to indicate the adjoint operator. Both notations will be used heavily in our exposition. The key result that we use is this: if $X_1, \ldots, X_n$ are i.i.d. from F_0 and if $T: \mathfrak{F} \to \mathcal{H}$ is Frechet differentiable at F_0, then

$$\sqrt{n}[T(F_n) - T(F_0)] \xrightarrow{\mathcal{D}} N(0, \Gamma), \tag{5.1}$$

where $\Gamma \in \mathcal{H} \otimes \mathcal{H}$ is the linear operator

$$E(T^\star(X) \otimes T^\star(X)) - E(T^\star(X)) \otimes E(T^\star(X)) = E(T^\star(X) \otimes T^\star(X)).$$

This fact is known as the δ-method for statistical functionals. In our case, $\mathcal{H}$ will be the tensor product space $\mathcal{H}_U \otimes \mathcal{H}_\Theta$ introduced earlier. From the above discussion we see that the key to computing the asymptotic normal distribution of $\sqrt{n}(T(F_n) - T(F_0))$ is to compute the influence function $T^\star(x)$. Next, we define our statistical functional.

5.2. *Statistical functional for testing elliptical distributions.* Now let us consider the statistical functional $T(F)$ in our setting. Since it is a simple function of $\breve{\Sigma}_{U\Theta}$, let us first figure out the statistical functional corresponding to this linear operator. For any $F \in \mathfrak{F}$, let

$$\mu(F) = \int x dF(x), \quad \Sigma(F) = \int (x - \mu(F))(x - \mu(F))^\intercal dF(x).$$

Clearly, $\mu(F_0) = \mu$, $\mu(F_n) = \hat{\mu}$, $\Sigma(F_0) = \Sigma$, and $\Sigma(F_n) = \hat{\Sigma}$. For each $x \in \Omega_X$, let

$$\begin{aligned} W(x, F) &= \Sigma(F)^{-1/2}(x - \mu(F)), \\ U(x, F) &= \|W(x, F)\|, \\ V(x, F) &= W(x, F)/U(x, F), \\ \Theta(x, F) &= g(V(x, F)). \end{aligned}$$

It is easy to see that, when evaluated at $F = F_0$, $U(x, F)$, $V(x, F)$, and $\Theta(x, F)$ reduce to $U(x)$, $V(x)$, and $\Theta(x)$, and when evaluated at $F = F_n$, they reduce to $\hat{U}(x)$, $\hat{V}(x)$ and $\hat{\Theta}(x)$. Our statistical functional of interest is the Hilbert-Schmidt norm of the operator

$$\begin{aligned} \Sigma_{U\Theta}(F) = &\textstyle\int \kappa_U(\cdot, U(x, F)) \otimes \kappa_\Theta(\cdot, \Theta(x, F)) dF(x) \\ &- \textstyle\int \kappa_U(\cdot, U(x, F)) dF(x) \otimes \int \kappa_\Theta(\cdot, \theta) dP_0(\theta). \end{aligned} \tag{5.2}$$

It is important to note that the last term on the right, $\int \kappa_\Theta(\cdot, \theta) dP_0(\theta)$, does not involve the unknown distribution F. This is the true expectation determined by the known distribution P_0 in Proposition 2. The operator in (5.2) can be re-expressed via the centered kernel as

$$\Sigma_{U\Theta}(F) = \textstyle\int \kappa_U(\cdot, U(x, F)) \otimes \tilde{\kappa}_\Theta(\cdot, \Theta(x, F)) dF(x). \tag{5.3}$$

Clearly, when evaluated at $F = F_0$, $\Sigma_{U\Theta}(F)$ reduces to $\Sigma_{U\Theta}$, and when evaluated at $F = F_n$, $\Sigma_{U\Theta}(F)$ reduces to $\breve{\Sigma}_{U\Theta}$. Our statistical functional of interest is then defined as

$$\mathfrak{F} \to \mathbb{R}, \quad F \mapsto \|\breve{\Sigma}_{U\Theta}(F)\|^2_{\mathrm{HS}}.$$

5.3. *Derivations of influence functions.* In this subsection we derive the influence functions of statistical functionals involved in $\|\breve{\Sigma}_{U\Theta}(F)\|^2_{\mathrm{HS}}$. Some of these functionals are of the form $F \mapsto G(x, F)$, which already depends on x. To make a distinction with this x and the argument x in $T^\star(x) = A(\delta_x)$, we denote the argument in any influence function by z. Thus, we denote the influence function of the statistical functional $F \mapsto G(x, F)$ as $G^\star(x, z)$. That is,

$$G^\star(x, z) = [\partial G(x, (1 - \epsilon)F_0 + \epsilon\delta_z)/\partial\epsilon]_{\epsilon=0}.$$

We will refer to the process of deriving $A^\star(z)$ from $A(F)$ as the $\star$-operation. The basic rules for the $\star$-operation are given in Proposition 9.2 of Li (2018). We start with the influence functions about $\mu(F)$, $\Sigma(F)$. The results are given by Lemma 9.1 of Li (2018), and we reproduce them here for later references.

LEMMA 2. *If X is integrable, then the influence function of μ is*

$$\mu^\star(z) = z - \mu.$$

Furthermore, if X is square integrable, then

$$\Sigma^\star(z) = (z - \mu)(z - \mu)^\intercal - \Sigma,$$

$$(\Sigma^{-1})^\star(z) = -\Sigma^{-1}\Sigma^\star(z)\Sigma^{-1}, \tag{5.4}$$

$$\mathrm{vec}[(\Sigma^{-1/2})^\star(z)] = -(\Sigma^{1/2} \otimes \Sigma + \Sigma \otimes \Sigma^{1/2})^{-1}\mathrm{vec}[\Sigma^\star(z)], \tag{5.5}$$

where, for a matrix A with columns $a_1, \ldots, a_m$, $\mathrm{vec}(A)$ denotes the vector $(a_1^\intercal, \ldots, a_m^\intercal)^\intercal$.

We next derive the influence functions for $U(x, F)$, $V(x, F)$, and $\Theta(x, F)$. For deriving the influence function of $\Theta(x, F)$, we need the derivative of the polar coordinate transformation, which is given in the next lemma.

LEMMA 3. *Let $v=(v_1,\dots,v_d)^\intercal\in\mathbb{S}^{d-1}$ and $\theta=g(v)=(\theta_1,\dots,\theta_{d-1})^\intercal\in\Omega_\Theta$. Let S_i be the Euclidean norm of the vector $(v_j,\dots,v_d)^\intercal$ for $i=1,\dots,d$. Then*

$$\frac{\partial g(v)}{\partial v^\intercal}=\begin{pmatrix} \frac{S_2}{S_1^2} & -\frac{v_1v_2}{S_1^2S_2} & -\frac{v_1v_3}{S_1^2S_2} & \dots & -\frac{v_1v_{d-2}}{S_1^2S_2} & -\frac{v_1v_{d-1}}{S_1^2S_2} & -\frac{v_1v_d}{S_1^2S_2} \\ 0 & \frac{S_3}{S_2^2} & -\frac{v_2v_3}{S_2^2S_3} & \dots & -\frac{v_2v_{d-2}}{S_2^2S_3} & -\frac{v_2v_{d-1}}{S_2^2S_3} & -\frac{v_2v_d}{S_2^2S_3} \\ 0 & 0 & \frac{S_4}{S_3^2} & \dots & -\frac{v_3v_{d-1}}{S_3^2S_4} & -\frac{v_3v_{d-2}}{S_3^2S_4} & -\frac{v_3v_d}{S_3^2S_4} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{S_{d-1}}{S_{d-2}^2} & -\frac{v_{d-2}v_{d-1}}{S_{d-2}^2S_{d-1}} & -\frac{v_{d-2}v_d}{S_{d-2}^2S_{d-1}} \\ 0 & 0 & 0 & \dots & 0 & \frac{v_d}{S_{d-1}^2} & -\frac{v_{d-1}}{S_{d-1}^2} \end{pmatrix}.$$

Based on Lemma 2, we next derive the influence functions for the statistical functionals

$$F\mapsto U(x,F),\quad F\mapsto V(x,F),\quad F\mapsto\Theta(x,F).$$

LEMMA 4. *Suppose that $F\mapsto U(x,F)$, $F\mapsto V(x,F)$ and $F\mapsto\Theta(x,F)$ are Frechet differentiable at F_0. Let*

$$\begin{aligned} A_1(x)&=-\frac{(x-\mu)^\intercal\Sigma^{-1}}{[(x-\mu)^\intercal\Sigma^{-1}(x-\mu)]^{1/2}},\\ A_2(x)&=-\frac{[(x-\mu)^\intercal\Sigma^{-1}]\otimes[(x-\mu)^\intercal\Sigma^{-1}]}{2[(x-\mu)^\intercal\Sigma^{-1}(x-\mu)]^{1/2}},\\ B_1(x)&=-\frac{\Sigma^{-1/2}(x-\mu)A_1(x)}{(x-\mu)^\intercal\Sigma^{-1}(x-\mu)}-\frac{\Sigma^{-1/2}}{[(x-\mu)^\intercal\Sigma^{-1}(x-\mu)]^{1/2}},\\ B_2(x)&=-\frac{\Sigma^{-1/2}(x-\mu)A_2(x)}{(x-\mu)^\intercal\Sigma^{-1}(x-\mu)}-\frac{[(x-\mu)^\intercal\otimes I_d]\,(\Sigma^{1/2}\otimes\Sigma+\Sigma\otimes\Sigma^{1/2})^{-1}}{[(x-\mu)^\intercal\Sigma^{-1}(x-\mu)]^{1/2}},\\ C_i(x)&=[\partial g(V(x))/\partial v^\intercal]B_i(x),\quad i=1,2, \end{aligned}\tag{5.6}$$

where $\partial g(V(x))/\partial v^\intercal$ is the matrix given by Lemma 3. Then, the influence functions of $U(x,F)$, $V(x,F)$ and $\Theta(x,F)$ are

$$U^\star(x,z)=A_1(x)\mu^\star(z)+A_2(x)\mathrm{vec}[\Sigma^\star(z)],\tag{5.7}$$

$$V^\star(x,z)=B_1(x)\mu^\star(z)+B_2(x)\mathrm{vec}[\Sigma^\star(z)],\tag{5.8}$$

$$\Theta^\star(x,z)=C_1(x)\mu^\star(z)+C_2(x)\mathrm{vec}[\Sigma^\star(z)].\tag{5.9}$$

The next lemma gives the influence function of $\Sigma_{U\Theta}(F)$. Henceforth, for a kernel function $\kappa(\cdot,t)$, we use $\dot\kappa(\cdot,t)$ to denote the partial derivative with respect to the second argument, $\partial\kappa(\cdot,t)/\partial t$.

LEMMA 5. *Suppose $F\mapsto\Sigma_{U\Theta}(F)$ is Frechet differentiable at F_0. Then*

$$\begin{aligned} \Sigma_{U\Theta}^\star(z)=\;&\kappa_U(\cdot,U(z))\otimes\tilde\kappa_\Theta(\cdot,\Theta(z))-\Sigma_{U\Theta}\\ &+E\{[\dot\kappa_U(\cdot,U(X))U^\star(X,z)]\otimes\tilde\kappa_\Theta(\cdot,\Theta(X))\}\\ &+E\{\kappa_U(\cdot,U(X))\otimes\dot\kappa_\Theta(\cdot,\Theta(X))^\intercal\Theta^\star(X,z)\}, \end{aligned}\tag{5.10}$$

where $U^\star(x,z)$ and $\Theta^\star(x,z)$ are given by Lemma 4.

Even though $\Sigma_{U\Theta}$ is always a member of $\mathcal{H}_U \otimes \mathcal{H}_\Theta$, it does not follow that $\Sigma^\star_{U\Theta}(z)$ must also be a member of $\mathcal{H}_U \otimes \mathcal{H}_\Theta$. However, to facilitate computation of (5.10) at the sample level, we need $\Sigma^\star_{U\Theta}(z)$ to be a member of $\mathcal{H}_U \otimes \mathcal{H}_\Theta$. Fortunately, this can be ensured if

$$\kappa_U \in C^2(\Omega_U \times \Omega_U), \quad \kappa_\Theta \in C^2(\Omega_\Theta \times \Omega_\Theta), \tag{5.11}$$

where, for a set $A \subseteq \mathbb{R}^m$, $C^2(A \times A)$ denotes the set of all real-valued functions on $A \times A$ that are twice differentiable with a bounded Hessian matrix. The establishment is placed in the Supplementary Materials, which requires a special case of Theorem 1 of Zhou (2008).

5.4. *Asymptotic distribution of the test statistic.* Based on the influence function of $\Sigma_{U\Theta}(F)$ computed in Lemma 5 and the functional Delta method expressed in (5.1), we can directly write down the asymptotic distribution of $\breve{\Sigma}_{U\Theta}$.

THEOREM 2. *If the statistical functional $F \mapsto \Sigma_{U\Theta}(F)$ is Frechet differentiable at F_0 with respect to the uniform metric in $\mathfrak{F}$ and conditions in (5.11) are satisfied, then*

$$\sqrt{n}(\breve{\Sigma}_{U\Theta} - \Sigma_{U\Theta}) \xrightarrow{\mathcal{D}} N(0, \Gamma),$$

where $\Gamma : \mathcal{H}_U \otimes \mathcal{H}_\Theta \to \mathcal{H}_U \otimes \mathcal{H}_\Theta$ is the operator

$$\Gamma = E\left[\Sigma^\star_{U\Theta}(X) \otimes \Sigma^\star_{U\Theta}(X)\right],$$

and $\Sigma^\star_{U\Theta}(z)$ is given by (5.10).

Note that the assertion that Γ is an operator from $\mathcal{H}_U \otimes \mathcal{H}_\Theta$ to $\mathcal{H}_U \otimes \mathcal{H}_\Theta$ is a consequence of (5.11), which guarantees that $\Sigma^\star_{U\Theta}(z) \in \mathcal{H}_U \otimes \mathcal{H}_\Theta$. Since, under the null hypothesis, $\Sigma_{U\Theta} = 0$, we have the following corollary of Theorem 2, which will be important for sample-level implementation.

COROLLARY 2. *Under the null hypotheses $H_0 : \Sigma_{U\Theta} = 0$ and the conditions in Theorem 2, we have*

$$\sqrt{n}\breve{\Sigma}_{U\Theta} \xrightarrow{\mathcal{D}} N(0, \Gamma). \tag{5.12}$$

where $\Gamma = E\left[\Sigma^\star_{U\Theta}(X) \otimes \Sigma^\star_{U\Theta}(X)\right]$ and $\Sigma^\star_{U\Theta}(z)$ as given in (5.10) but without the $\Sigma_{U\Theta}$ term.

Applying continuous mapping theorem to (5.12) by taking the squared Hilbert-Schmidt norm, we have the following corollary.

COROLLARY 3. *Under the null hypotheses $H_0 : \Sigma_{U\Theta} = 0$ and the conditions in Theorem 2, we have*

$$n\|\breve{\Sigma}_{U\Theta}\|^2_{\mathrm{HS}} \xrightarrow{\mathcal{D}} \textstyle\sum_{i=1}^\infty \lambda_i Z_i^2, \tag{5.13}$$

where $Z_1, Z_2, \ldots$ are i.i.d. standard normal random variables, and $\lambda_1, \lambda_2, \ldots$ are eigenvalues of Γ in Corollary 2.

Note that the null distribution in (5.13) only depends on the eigenvalues of Γ, which gives us the chance of not having to save the whole Γ on sample-level implement. Based on Theorem 2, we can also derive the asymptotic distribution of $\sqrt{n}(\|\breve{\Sigma}_{U\Theta}\|^2_{\mathrm{HS}} - \|\Sigma_{U\Theta}\|^2_{\mathrm{HS}})$ under the alternative hypothesis.

COROLLARY 4. *Suppose that the conditions of Theorem 2 hold. Then, under the alternative hypothesis, i.e.,* $\Sigma_{U\Theta} \neq 0$*, we have*

$$\sqrt{n}(\|\breve{\Sigma}_{U\Theta}\|^2_{\mathrm{HS}} - \|\Sigma_{U\Theta}\|^2_{\mathrm{HS}}) \xrightarrow{\mathcal{D}} N(0, 4E(\langle \Sigma^{\star}_{U\Theta}(X), \Sigma_{U\Theta}\rangle^2_{\mathrm{HS}})).$$

Corollary 4 also implies that $\sqrt{n}\|\breve{\Sigma}_{U\Theta}\|^2_{\mathrm{HS}} \xrightarrow{P} \infty$ under the alternative hypothesis, which verifies the consistency of our test.

5.5. *Local power analysis.* In this subsection, we derive the local alternative distribution, which can be used to compute the local power. The idea of the proof is similar to Theorem 13 of Gretton et al. (2012).

THEOREM 3. *Suppose*

1. Γ *has spectral decomposition* $\sum_{j=1}^{\infty}\lambda_j(v_j \otimes v_j)$*, where* $v_1, v_2, \ldots$ *is an orthonormal basis in* $\mathcal{H}_U \otimes \mathcal{H}_\Theta$*;*
2. Σ_1 *is a fixed linear operator in* $\mathcal{H}_U \otimes \mathcal{H}_\Theta$ *with expansion* $\sum_{j=1}^{\infty}\sigma_j v_j$ *and* $\|\Sigma_1\|_{\mathrm{HS}} = c > 0$.

Then, under the local alternative hypothesis $H_1^{(\mathrm{n})} : \Sigma_{U\Theta} = n^{-1/2}\Sigma_1$*, we have*

$$n\|\breve{\Sigma}_{U\Theta}\|^2_{\mathrm{HS}} \xrightarrow{\mathcal{D}} \sum_{j=1}^{\infty}\lambda_j \tilde{Z}_j^2$$

where $\tilde{Z}_j$ *are independent* $N(\sigma_j/\sqrt{\lambda_j}, 1)$ *random variables.*

Using this theorem, we calculate the local power of our test as

$$P\left(n\|\breve{\Sigma}_{U\Theta}\|^2_{\mathrm{HS}} > s\right) \to P\left(\sum_{j=1}^{\infty}\lambda_j \tilde{Z}_j^2 > s\right).$$

6. Approximating the asymptotic null distribution.

6.1. *Outline of the problem and notations.* In this section we approximate the asymptotic distribution of T_n, which is $\sum_{i=1}^{\infty}\lambda_i Z_i^2$, where $\lambda_1, \lambda_2, \ldots$ are the eigenvalues of Γ, and $Z_1, Z_2, \ldots$ are i.i.d. $N(0,1)$. The operator Γ is estimated by substituting, wherever possible, the expectation E by the sample average E_n in the expression $E(\Sigma^{\star}_{U\Theta} \otimes \Sigma^{\star}_{U\Theta})$. Denote the estimate of Γ as $\hat{\Gamma}$. We use the eigenvalues $\hat{\lambda}_i$ of $\hat{\Gamma}$ to estimate λ_i in the asymptotic distribution $\sum_{i=1}^{\infty}\lambda_i Z_i^2$. We then use the plug-in estimate $\sum_{i=1}^{\infty}\hat{\lambda}_i Z_i^2$ to approximate the asymptotic distribution of T_n.

In the following, for an integer m, we use $[m]$ to represent the set $\{1, \ldots, m\}$. If A is a set and $a : [m] \to A$ is a function, we use $a_{[m]}$ to represent the vector $(a_1, \ldots, a_m)^{\intercal}$. Furthermore, for a function f defined on A, and a function $a : [m] \to A$, we use $f(a_{[m]})$ to denote the vector $(f(a_1), \ldots, f(a_m))^{\intercal}$. This notation also extends to functions involving other variables. For example, $\kappa(\cdot, a_{[m]})$ represents the vector of functions $(\kappa(\cdot, a_1), \ldots, \kappa(\cdot, a_m))^{\intercal}$; and $f(x, y, z_{[m]})$ represents the vector $(f(x, y, z_1), \ldots, f(x, y, z_m))^{\intercal}$.

One of the advantages of the RKHS is that we can use the representer theorem (see, for example, Schölkopf, Herbrich and Smola (2001)) to turn an infinite dimensional problem into a finite dimensional problem. Let $\mathcal{H}$ be the RKHS generated by a kernel $\kappa : A \times A \to \mathbb{R}$. Suppose our statistical procedure relies on functions in $\mathcal{H}$ only through $f(a_1), \ldots, f(a_m)$ for $a_1, \ldots, a_m \in A$. Then we can, without loss of generality, restrict our attention to $\hat{\mathcal{H}} \subseteq \mathcal{H}$, where $\hat{\mathcal{H}}$ is spanned by $\kappa(\cdot, a_1), \ldots, \kappa(\cdot, a_m)$. This is because there is always a function $\hat{f} \in \hat{\mathcal{H}}$ such that $f(a_i) = \hat{f}(a_i)$ for $i = 1, \ldots, m$. In fact, let

$$\hat{f} = t_1\kappa(\cdot, a_1) + \cdots + t_m\kappa(\cdot, a_m) = t_{[m]}^{\intercal}\kappa(\cdot, a_{[m]}).$$

Up on solving the equation $\hat{f}(a_{[m]}) = f(a_{[m]})$, we have $t_{[m]} = K^{-1} f(a_{[m]})$, where K is the Gram matrix $\{\kappa(a_i, a_j)\}_{i,j=1}^m$. For this reason, in this section, we will reset $\mathcal{H}_U$ and $\mathcal{H}_\Theta$ to be the RKHS spanned by

$$\mathcal{B}_U = \{\kappa_U(\cdot, \hat{U}_i) : i = 1, \ldots, n\}, \quad \mathcal{B}_\Theta = \{\kappa_\Theta(\cdot, \hat{\Theta}_i) : i = 1, \ldots, n\}.$$

A basis for $\mathcal{H}_U \otimes \mathcal{H}_\Theta$ is

$$\{\kappa_U(\cdot, \hat{U}_i) \otimes \kappa_\Theta(\cdot, \hat{\Theta}_j) : i, j = 1, \ldots, n\}.$$

Let $C_{U\Theta} : \mathcal{H}_U \otimes \mathcal{H}_\Theta \to \mathbb{R}^{n^2}$ be the coordinate mapping that takes $\kappa_U(\cdot, \hat{U}_i) \otimes \kappa_\Theta(\cdot, \hat{\Theta}_j)$ to $e_i \otimes e_j$ in $\mathbb{R}^{n^2}$. Then, it is easy to see that $C_{U\Theta} = C_U \otimes C_\Theta$, where C_U takes $\kappa_U(\cdot, \hat{U}_i)$ to e_i and C_Θ takes $\kappa_\Theta(\cdot, \hat{\Theta}_j)$ to e_j. Let $K_{U\Theta}$ be the $\mathbb{R}^{n^2 \times n^2}$ Gram matrix whose $((i,j),(i',j'))$th entry is

$$\begin{aligned}(K_{U\Theta})_{(i,j),(i',j')} &= \langle \kappa_U(\cdot, \hat{U}_i) \otimes \kappa_\Theta(\cdot, \hat{\Theta}_j), \kappa_U(\cdot, \hat{U}_i') \otimes \kappa_\Theta(\cdot, \hat{\Theta}_j') \rangle_{\mathcal{H}_U \otimes \mathcal{H}_\Theta} \\ &= \kappa_U(\hat{U}_i, \hat{U}_i') \kappa_\Theta(\hat{\Theta}_j, \hat{\Theta}_j') = (K_U)_{ii'} (K_\Theta)_{jj'}.\end{aligned}$$

Thus, in matrix notation,

$$K_{U\Theta} = K_U \otimes K_\Theta,$$

where $\otimes$ is the Kronecker product between matrices. As discussed earlier, the eigenvalues of $\hat{\Gamma}$ are the same as the eigenvalues of

$$K_{U\Theta}^{1/2} C_{U\Theta} \hat{\Gamma} C_{U\Theta}^* K_{U\Theta}^{-1/2} = (K_U \otimes K_\Theta)^{1/2} (C_U \otimes C_\Theta) \hat{\Gamma} (C_U \otimes C_\Theta)^* (K_U \otimes K_\Theta)^{-1/2}. \tag{6.1}$$

So it all boils down to computing the coordinate $(C_U \otimes C_\Theta)\hat{\Gamma}(C_U \otimes C_\Theta)^*$.

6.2. *Estimation of* Γ. By Corollary 2, under $U \perp\!\!\!\perp \Theta$, Γ is of the form

$$\begin{aligned}\Sigma_{U\Theta}^\star(z) &= \kappa_U(\cdot, U(z)) \otimes \tilde{\kappa}_\Theta(\cdot, \Theta(z)) \\ &\quad + \int \kappa_U^\star(x, z) \otimes \tilde{\kappa}_\Theta(\cdot, \Theta(x)) dF_0(x) + \int \kappa_U(\cdot, U(x)) \otimes \kappa_\Theta^\star(x, z) dF_0(x) \\ &\equiv \Sigma_{U\Theta,1}^\star(z) + \Sigma_{U\Theta,2}^\star(z) + \Sigma_{U\Theta,3}^\star(z).\end{aligned} \tag{6.2}$$

Let $\hat{A}_1(x)$ and $\hat{A}_2(x)$ be as defined in (5.6) with μ and Σ replaced by $\hat{\mu}$ and $\hat{\Sigma}$; let $\hat{B}_1(x)$ and $\hat{B}_2(x)$ be as defined in (5.6) with $\mu(x)$, $\Sigma(x)$, $A_1(x)$, and $A_2(x)$ replaced by $\hat{\mu}(x)$, $\hat{\Sigma}(x)$, $\hat{A}_1(x)$, and $\hat{A}_2(x)$; let $\hat{C}_1(x)$ and $\hat{C}_2(x)$ be as defined in (5.6) with $B_1(x)$ and $B_2(x)$ replaced by $\hat{B}_1(x)$ and $\hat{B}_2(x)$. Let $\hat{\mu}^\star(z)$ and $\hat{\Sigma}^\star(z)$ be as defined in Lemma 2 with μ and Σ replaced by $\hat{\mu}$ and $\hat{\Sigma}$. Let $\hat{U}^\star(x,z)$ and $\hat{\Theta}^\star(x,z)$ be as defined in (5.7) and (5.9) with $A_1(x)$, $A_2(x)$, $C_1(x)$, $C_2(x)$, $\mu^\star(x)$, and $\Sigma^\star(x)$ replaced by $\hat{A}_1(x)$, $\hat{A}_2(x)$, $\hat{C}_1(x)$, $\hat{C}_2(x)$, $\hat{\mu}^\star(x)$, and $\hat{\Sigma}^\star(x)$.

To approximate $\Sigma_{U\Theta,1}^\star(z)$ in (6.2), we replace $U(z)$ and $\Theta(z)$ by $\hat{U}(z)$ and $\hat{\Theta}(z)$, and replace $\tilde{\kappa}_\Theta(\cdot, \theta)$ by

$$\tilde{\kappa}_\Theta^{(e)}(\cdot, \theta) = \kappa_\Theta(\cdot, \theta) - n^{-1} \textstyle\sum_{i=1}^n \kappa_\Theta(\cdot, \hat{\Theta}_i),$$

where the superscript in $\tilde{\kappa}^{(e)}$ indicates the word "empirical", as we use the sample average $n^{-1} \sum_{i=1}^n \kappa_\Theta(\cdot, \hat{\Theta}_i)$ instead of the population average $E\kappa_\Theta(\cdot, \Theta)$ to center the kernel κ_Θ. So $\Sigma_{U\Theta,1}^\star(z)$ is approximated by

$$\hat{\Sigma}_{U\Theta,1}^\star(z) = \kappa_U(\cdot, \hat{U}(z)) \otimes \tilde{\kappa}_\Theta^{(e)}(\cdot, \hat{\Theta}(z)).$$

To approximate $\Sigma_{U\Theta,2}^\star(z)$ and $\Sigma_{U\Theta,3}^\star(z)$ in (6.2), we replace $\kappa_U^\star(\cdot, x, z)$ and $\kappa_\Theta^\star(\cdot, x, z)$ by

$$\hat{\kappa}_U^\star(\cdot, x, z) = \dot{\kappa}_U(\cdot, \hat{U}(x)) \hat{U}^\star(x, z), \quad \hat{\kappa}_\Theta^\star(\cdot, x, z) = \dot{\kappa}_\Theta(\cdot, \hat{\Theta}(x))^\intercal \hat{\Theta}^\star(x, z).$$

We replace $\int \cdots dF_0(x)$ by $E_n(\cdots)$. So $\Sigma^\star_{U\Theta,2}(z)$ is approximated by

$$\hat{\Sigma}^\star_{U\Theta,2}(z) = n^{-1}\textstyle\sum_{i=1}^n \hat{\kappa}^\star_U(\cdot, X_i, z) \otimes \tilde{\kappa}^{(e)}_\Theta(\cdot, \hat{\Theta}_i).$$

Similarly, $\Sigma^\star_{U\Theta,3}(z)$ is approximated by

$$\hat{\Sigma}^\star_{U\Theta,3}(z) = n^{-1}\textstyle\sum_{i=1}^n \kappa_U(\cdot, \hat{U}_i) \otimes \hat{\kappa}^\star_\Theta(\cdot, X_i, z).$$

Combining the above results, we approximate $\Sigma^\star_{U\Theta}(z)$ by

$$\begin{aligned}\hat{\Sigma}^\star_{U\Theta}(z) &= \hat{\Sigma}^\star_{U\Theta,1}(z) + \hat{\Sigma}^\star_{U\Theta,2}(z) + \hat{\Sigma}^\star_{U\Theta,3}(z)\\ &= \kappa_U(\cdot, \hat{U}(z)) \otimes \tilde{\kappa}^{(e)}_\Theta(\cdot, \hat{\Theta}(z)) + n^{-1}\textstyle\sum_{i=1}^n \hat{\kappa}^\star_U(\cdot, X_i, z) \otimes \tilde{\kappa}^{(e)}_\Theta(\cdot, \hat{\Theta}_i)\\ &\quad + n^{-1}\textstyle\sum_{i=1}^n \kappa_U(\cdot, \hat{U}_i) \otimes \hat{\kappa}^\star_\Theta(\cdot, X_i, z).\end{aligned} \tag{6.3}$$

To approximate Γ, we replace $\Sigma^\star_{U\Theta}(z)$ by $\hat{\Sigma}^\star_{U\Theta}(z)$ and $\int \cdots dF_0(x)$ by $E_n(\cdots)$, as follows

$$\hat{\Gamma} = n^{-1}\textstyle\sum_{i=1}^n [\hat{\Sigma}^\star_{U\Theta}(X_i) \otimes \hat{\Sigma}^\star_{U\Theta}(X_i)].$$

6.3. *Approximating the asymptotic distribution of* $n\|\breve{\Sigma}_{U\Theta}\|^2_{\mathrm{HS}}$. Our goal is to find the eigenvalues of $\hat{\Gamma}$ so as to form the approximation of $\sum \lambda_i Z_i^2$. Recall that Γ is a self adjoint linear operator from $\mathcal{H}_U \otimes \mathcal{H}_\Theta$ to $\mathcal{H}_U \otimes \mathcal{H}_\Theta$, where, as mentioned earlier, $\mathcal{H}_U$ and $\mathcal{H}_\Theta$ can be regarded as the finite-dimensional RKHS's spanned by $\mathcal{B}_U$ and $\mathcal{B}_\Theta$. As mentioned at the end of Section 6.1, we need to compute the matrix (6.1).

We need to review a coordinate mapping rule in a tensor product space. Suppose $\mathcal{H}_1$ and $\mathcal{H}_2$ are finite-dimensional Hilbert spaces of dimension m_1 and m_2, with bases $\mathcal{B}_1 = \{b_1^{(1)}, \ldots, b_{m_1}^{(1)}\}$ and $\mathcal{B}_2 = \{b_1^{(2)}, \ldots, b_{m_2}^{(2)}\}$, respectively. Let $C_1 : \mathcal{H}_1 \to \mathbb{R}^m$ and $C_2 : \mathcal{H}_2 \to \mathbb{R}^m$ be coordinate mappings with respect to $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively. Let G_1 and G_2 be the Gram matrices for $\mathcal{B}_1$ and $\mathcal{B}_2$. A basis of the tensor product space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is

$$\mathcal{B}_{12} = \{b_{ij} = b_i^{(1)} \otimes b_j^{(2)} : i = 1, \ldots, m_1, j = 1, \ldots, m_2\}.$$

For $f_1 \in \mathcal{H}_1$ and $f_2 \in \mathcal{H}_2$, the tensor product $f_1 \otimes f_2$ can be viewed in two ways: as a member of $\mathcal{H}_1 \otimes \mathcal{H}_2$ or as an operator from $\mathcal{H}_2$ to $\mathcal{H}_1$. As a member of $\mathcal{H}_1 \otimes \mathcal{H}_2$, the coordinate of $f_1 \otimes f_2$ with respect to $\mathcal{B}_{12}$ is

$$(C_1 \otimes C_2)(f_1 \otimes f_2) = (C_1 f_1) \otimes (C_2 f_2),$$

where $\otimes$ on the right is the Kronecker product between matrices. As a linear operator from $\mathcal{H}_2$ to $\mathcal{H}_2$, the coordinate of $f_1 \otimes f_2$ with respect to the bases $\{\mathcal{B}_1, \mathcal{B}_2\}$ is

$$C_1(f_1 \otimes f_2)C_2^*.$$

By Li and Solea (2018), we have

$$C_1(f_1 \otimes f_2)C_2^* = (C_1 f_1)(C_2 f_2)^\intercal G_2. \tag{6.4}$$

Applying (6.4) to $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}_U \otimes \mathcal{H}_\Theta$, we have

$$\begin{aligned}&(C_U \otimes C_\Theta)[\hat{\Sigma}^\star_{U\Theta}(X_i) \otimes \hat{\Sigma}^\star_{U\Theta}(X_i)](C_U \otimes C_\Theta)^*\\ &\qquad = [(C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta}(X_i)][(C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta}(X_i)]^\intercal (K_U \otimes K_\Theta).\end{aligned}$$

Therefore, we can reexpress (6.1) as

$$\begin{aligned}&(K_U \otimes K_\Theta)^{1/2}(C_U \otimes C_\Theta)\hat{\Gamma}(C_U \otimes C_\Theta)^*(K_U \otimes K_\Theta)^{-1/2}\\ &= n^{-1}\textstyle\sum_{i=1}^n [(K_U \otimes K_\Theta)^{1/2}(C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta}(X_i)][(K_U \otimes K_\Theta)^{1/2}(C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta}(X_i)]^\intercal.\end{aligned} \tag{6.5}$$

Let

$$\hat{m}(X_i) = n^{-1/2}(K_U \otimes K_\Theta)^{1/2}(C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta}(X_i), \;\; \hat{M} = (\hat{m}(X_1), \ldots, \hat{m}(X_n)).$$

Then the matrix in (6.5) can be rewritten as

$$(K_U \otimes K_\Theta)^{1/2}(C_U \otimes C_\Theta)\hat{\Gamma}(C_U \otimes C_\Theta)^*(K_U \otimes K_\Theta)^{-1/2} = \hat{M}\hat{M}^\intercal.$$

It remains to calculate $\hat{m}(X_i)$, where the key is to calculate $(C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta}(X_i)$. By (6.3), this is

$$(C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta,1}(X_i) + (C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta,2}(X_i) + (C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta,3}(X_i).$$

Reading off $\hat{\Sigma}^\star_{U\Theta,1}(z)$ from (6.3), we have

$$\begin{aligned}(C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta,1}(X_i) &= (C_U \otimes C_\Theta)[\kappa_U(\cdot, \hat{U}_i) \otimes \tilde{\kappa}^{(e)}_\Theta(\cdot, \hat{\Theta}_i)] \\ &= [C_U\kappa_U(\cdot, \hat{U}_i)] \otimes [C_\Theta\tilde{\kappa}^{(e)}_\Theta(\cdot, \hat{\Theta}_i)],\end{aligned}$$

where

$$C_U\kappa_U(\cdot, \hat{U}_i) = e_i, \quad C_\Theta\tilde{\kappa}^{(e)}_\Theta(\cdot, \hat{\Theta}_i) = e_i - 1_n/n.$$

Reading off $\hat{\Sigma}^\star_{U\Theta,2}(z)$ from (6.3), we have

$$\begin{aligned}(C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta,2}(X_i) &= n^{-1}\textstyle\sum_{j=1}^n (C_U \otimes C_\Theta)[\hat{\kappa}^\star_U(\cdot, X_j, X_i) \otimes \tilde{\kappa}^{(e)}_\Theta(\cdot, \hat{\Theta}_j)] \\ &= n^{-1}\textstyle\sum_{j=1}^n [C_U\hat{\kappa}^\star_U(\cdot, X_j, X_i)] \otimes [C_\Theta\tilde{\kappa}^{(e)}_\Theta(\cdot, \hat{\Theta}_j)].\end{aligned} \tag{6.6}$$

To calculate $C_U\hat{\kappa}^\star_U(\cdot, X_j, X_i)$, we now give an expression for $C_U(f)$ for any $f \in \mathcal{H}_U$.

LEMMA 6. *Suppose $\mathcal{H}$ is a finite-dimensional RKHS generated by a kernel κ, with basis $\kappa(\cdot, a_a), \ldots, \kappa(\cdot, a_m)$, f is a member of $\mathcal{H}$, and C is the coordinate mapping. Then*

$$C(f) = K^{-1}f(a_{[k]}),$$

where $f(a_{[k]})$ is the vector $(f(a_1), \ldots, f(a_k))^\intercal$, and K is the Gram matrix $\{\kappa(a_i, a_j)\}_{i,j=1}^m$.

PROOF. Since $f \in \mathcal{H}$, we have $f = C(f)^\intercal\kappa(\cdot, a_{[m]})$ where $\kappa(\cdot, a_{[m]})$ represents the vector of functions $(\kappa(\cdot, a_1), \ldots, \kappa(\cdot, a_m))^\intercal$. Evaluate this equation at $a_1, \ldots, a_m$, we have

$$f(a_{[k]}) = KC(f).$$

Solving this equation, we have the desired result. □

Applying Lemma 6 to (6.6), we have

$$\begin{aligned}&(C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta,2}(X_i) = n^{-1}\textstyle\sum_{j=1}^n [K_U^{-1}\dot{\kappa}_U(\hat{U}_{[n]}, \hat{U}_j)\hat{U}^\star(X_j, X_i)] \otimes (e_j - 1_n/n) \\ &= n^{-1}\textstyle\sum_{j=1}^n \{K_U^{-1}\dot{\kappa}_U(\hat{U}_{[n]}, \hat{U}_j)[\hat{A}_1(X_j)\hat{\mu}^\star(X_i) + \hat{A}_2(X_j)\mathrm{vec}(\hat{\Sigma}^\star(X_i))]\} \otimes (e_j - 1_n/n).\end{aligned}$$

Reading off $\hat{\Sigma}^\star_{U\Theta,3}(z)$ from (6.3), we have

$$\begin{aligned}(C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta,3}(X_i) &= n^{-1}\textstyle\sum_{j=1}^n [C_U\kappa_U(\cdot, \hat{U}_j)] \otimes [C_\Theta\hat{\kappa}^\star_\Theta(\cdot, X_j, X_i)] \\ &= n^{-1}\textstyle\sum_{j=1}^n e_j \otimes [C_\Theta\hat{\kappa}^\star_\Theta(\cdot, X_j, X_i)].\end{aligned} \tag{6.7}$$

Applying Lemma 6 again to (6.7), we have

$$\begin{aligned}&(C_U \otimes C_\Theta)\hat{\Sigma}^\star_{U\Theta,3}(X_i) = n^{-1}\textstyle\sum_{j=1}^n e_j \otimes [K_\Theta^{-1}\dot{\kappa}_\Theta(\hat{\Theta}_{[n]}, \hat{\Theta}_j)\hat{\Theta}^\star(X_j, X_i)] \\ &= n^{-1}\textstyle\sum_{j=1}^n e_j \otimes \{K_\Theta^{-1}\dot{\kappa}_\Theta(\hat{\Theta}_{[n]}, \hat{\Theta}_j)[\hat{C}_1(X_j)\hat{\mu}^\star(X_i) + \hat{C}_2(X_j)\mathrm{vec}(\hat{\Sigma}^\star(X_i))]\}.\end{aligned}$$

To summarize, we have

$$\begin{aligned}\hat{m}(X_i) = &\, n^{-1/2}(K_U \otimes K_\Theta)^{1/2}[e_j \otimes (e_j - 1_n/n)] \\ &+ n^{-1/2}(K_U \otimes K_\Theta)^{1/2}\left\{n^{-1}\textstyle\sum_{j=1}^n [K_U^{-1}\dot{\kappa}_U(\hat{U}_{[n]}, \hat{U}_j)\hat{U}^\star(X_j, X_i)] \otimes (e_j - 1_n/n)\right\} \\ &+ n^{-1/2}(K_U \otimes K_\Theta)^{1/2}\left\{n^{-1}\textstyle\sum_{j=1}^n e_j \otimes [K_\Theta^{-1}\dot{\kappa}_\Theta(\hat{\Theta}_{[n]}, \hat{\Theta}_j)\hat{\Theta}^\star(X_j, X_i)]\right\}.\end{aligned}$$

We need to calculate the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$ of the matrix $\hat{M}\hat{M}^\intercal$. Since this is $n^2 \times n^2$, its eigenvalues are expensive to compute if computed directly. However, it is equivalent to calculate the eigenvalues of $\hat{M}^\intercal\hat{M}$, which is an $n \times n$ matrix. We then use $\sum_{i=1}^n \hat{\lambda}_i Z_i^2$ to approximate the asymptotic null distribution in (5.13). We apply the function `imhof` in the R package `CompQuadForm` (Duchesne and de Micheaux (2010)) to compute the p-value of the distribution of $\sum_{i=1}^n \hat{\lambda}_i Z_i^2$. If the p-value is smaller than some prespecified significance level α, then we reject H_0.

6.4. *Validity and consistency.* Since we have plugged the estimated eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$ into the asymptotic distribution of $n\|\breve{\Sigma}_{U\Theta}\|_{HS}^2$, we want to show that this step preserves the validity and consistency of the test.

Notice that $\hat{\Gamma}$ is a consistent estimator of Γ. Suppose that $\sum_{j=1}^\infty \lambda_j^{1/2} < \infty$, then, by Theorem 1 of Gretton et al. (2009), we have

$$\textstyle\sum_{j=1}^n \hat{\lambda}_j Z_j^2 \xrightarrow{\mathcal{D}} \sum_{j=1}^\infty \lambda_j Z_j^2,$$

which indicates the validity of the test. The consistency is straightforward since, under the alternative hypothesis, $\breve{\Sigma}_{U\Theta} \xrightarrow{P} \Sigma_{U\Theta}$ which is nonzero, indicating that $n\|\breve{\Sigma}_{U\Theta}\|_{HS}^2 \xrightarrow{P} \infty$ (see, for example, Section 12.4 of Kokoszka and Reimherr (2017)).

7. Uniform concentration bounds. In this section, we develop the concentration bounds for $\big|\|\breve{\Sigma}_{U\Theta}\|_{\mathrm{HS}} - \|\Sigma_{U\Theta}\|_{\mathrm{HS}}\big|$ in two cases: we first considered the simple case where μ and Σ are known, and then considered the general case where μ and Σ are unknown. These bounds also allow us to establish the consistency of our method when the dimension d goes to infinity with the sample size n.

7.1. *Case for known μ and Σ.* We first consider the case where μ and Σ are known, which would be true, for example, for testing a spherical distribution, where $\mu = 0$ and $\Sigma = I_p$. In this case $\hat{U}$ and $\hat{\Theta}$ are replaced by U and Θ, and our test statistic reduces to

$$\breve{\Sigma}_{U\Theta} = E_n[\kappa_U(\cdot, U) \otimes \kappa_\Theta(\cdot, \Theta)] - E_n[\kappa_U(\cdot, U)] \otimes E[\kappa_\Theta(\cdot, \Theta)]. \tag{7.1}$$

The next theorem gives the concentration bound for $\|\breve{\Sigma}_{U\Theta}\|_{\mathrm{HS}}$, which is similar to Theorem 7 in Gretton et al. (2012).

THEOREM 4. *Suppose that μ and Σ are known and $\breve{\Sigma}_{U\Theta}$ is defined as (7.1). Furthermore, suppose the kernels κ_U and κ_Θ are bounded: $0 \le \kappa_U(u, u') \le M_U$ and $0 \le \kappa_\Theta(\theta, \theta') \le M_\Theta$ for all u, u', θ, θ'. Then,*

$$P\left(\big|\|\breve{\Sigma}_{U\Theta}\|_{\mathrm{HS}} - \|\Sigma_{U\Theta}\|_{\mathrm{HS}}\big| \ge t + 4(M_U M_\Theta/n)^{1/2}\right) \le \exp\left(-\frac{t^2 n}{10 M_U M_\Theta}\right). \tag{7.2}$$

The proof of Theorem 4 is placed in the Supplementary Materials. Let $u = \frac{nt^2}{10M_U M_\Theta}$. Then, (7.2) is equivalent to

$$P\left(\Big|\|\breve{\Sigma}_{U\Theta}\|_{\mathrm{HS}} - \|\Sigma_{U\Theta}\|_{\mathrm{HS}}\Big| \ge (10 M_U M_\Theta u/n)^{1/2} + 4(M_U M_\Theta/n)^{1/2}\right) \le e^{-u}. \tag{7.3}$$

7.2. *Case for unknown μ and Σ.* Since we would like our concentration bounds to reflect the behavior with respect to both the sample size n and the dimension d, we need to separate out d and n from any constants in our derivations. To make this explicit, we call a constant that doesn't depend on d or n an *absolute constant*. We first make the following assumptions.

ASSUMPTION 1. There exist absolute constants $c_1 > 0$ and $c_2 > 0$ such that

$$c_1 \le \lambda_{\min}(\Sigma) \le \lambda_{\max}(\Sigma) \le c_2.$$

ASSUMPTION 2. The random vector X has a uniform sub-Gaussian distribution in the sense that, there is a constant σ^2 that is independent of d such that, for all $v \in \mathbb{S}$,

$$E(e^{\lambda\langle v, X-\mu\rangle}) \le e^{\sigma^2\lambda^2/2}.$$

ASSUMPTION 3. The density of W is bounded by c^d for some absolute constant c.

ASSUMPTION 4. The kernel functions κ_U and κ_Θ are bounded and Lipschitz continuous:

$$0 \le \kappa_U(u,u') \le M_U, \quad \forall u,u', \quad 0 \le \kappa_\Theta(\theta,\theta') \le M_\Theta, \quad \forall \theta,\theta',$$

and

$$\|\kappa_U(\cdot,u) - \kappa_U(\cdot,u')\| \le L_U|u-u'|, \quad \forall u,u', \tag{7.4}$$

$$\|\kappa_\Theta(\cdot,g(v)) - \kappa_\Theta(\cdot,g(v'))\| \le d^{-1}L_V\|v-v'\|, \quad \forall v,v'. \tag{7.5}$$

In the subsequent discussions we will carefully track the indices of the absolute constants as they will eventually appear in the same expression. Let

$$\begin{aligned}
f_1(n,d,u) &= c_1\sqrt{\frac{d[\log(2d)+u]}{n}},\\
f_2(n,d,u) &= c_2 d\max\left\{\left(\frac{d+u}{n}\right)^{1/4}, \left(\frac{d+u}{n}\right)^{1/2}\right\} + c_3\sqrt{d},\\
f_3(n,d,u) &= c_4 d\sqrt{\frac{d+u}{n}},\\
f_4(n,u) &= c_9\sqrt{\frac{u}{n}} + c_{10}\sqrt{\frac{1}{n}},
\end{aligned} \tag{7.6}$$

where $c_1, c_2, c_3, c_4, c_9, c_{10}$ are some positive absolute constants.

THEOREM 5. *Suppose X satisfies Assumption 2, and $X_1,\dots,X_n$ are i.i.d samples of X. Further suppose Σ satisfies Assumption 1, the density of W satisfies Assumption 3, and the kernel functions κ_U and κ_Θ satisfy Assumption 4. Then, for any $\epsilon > 0$, we have*

$$\begin{aligned}
&P\Big(\big|\|\breve{\Sigma}_{U\Theta}\|_{\mathrm{HS}} - \|\Sigma_{U\Theta}\|_{\mathrm{HS}}\big| \ge [c_7 + 2c_8/(\epsilon d)] \times\\
&[f_3(n,d,u)f_2(n,d,u) + c_5 f_1(n,d,u)] + f_4(n,u)\Big) \le n(c_6\epsilon)^d + 7e^{-u},
\end{aligned}$$

where f_1, f_2, f_3, f_4 are as defined in (7.6), and $c_1,\dots,c_{10}$ are some positive absolute constants.

The tail bounds in Theorem 5 allows us to establish the consistency of the test even when d goes to infinity with n, as shown in the next theorem.

THEOREM 6. *Suppose all conditions in Theorem 5 are satisfied. If $\log n \prec d \prec n^{1/4}$, then*

$$\|\breve{\Sigma}_{U\Theta}\|_{\mathrm{HS}} \xrightarrow{P} \|\Sigma_{U\Theta}\|_{\mathrm{HS}}.$$

8. Simulations. In this section we present some simulation results of using T_n to test ellipticity under both the null distribution and the alternative distribution. In all the simulations, we use Gaussian radial basis kernels for both U and Θ as given in (2.11). To select the tuning parameters, we use the criterion in Section 6.4 of Li and Solea (2018):

$$\frac{1}{\sqrt{\gamma_U}} = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |\hat{U}_i - \hat{U}_j|, \qquad \frac{1}{\sqrt{\gamma_\Theta}} = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \|\hat{\Theta}_i - \hat{\Theta}_j\|.$$

8.1. *Results under the null distribution.* In this subsection we perform simulations under the null distribution. We consider scenarios consisting of different sample sizes and dimensions:

$$(n, d) \in \{500, 1000\} \times \{3, 4, 5, 6, 10, 15, 20\}. \tag{8.1}$$

For each (n, d), we generate $T = 100$ datasets as follows. We first generate the mean vector μ from $N(0, \sigma_\mu^2 I_d)$, where $\sigma_\mu^2 = 100$, and the covariance matrix Σ using R function `genPositiveDefMat` in the package `clusterGeneration` (Qiu and Joe. (2020)) under its default settings. We then simulate $X_1, \ldots, X_n$ as i.i.d. samples from $N(\mu, \Sigma)$.

Based on the samples $X_1, \ldots, X_n$ we compute the test statistic $n\|\breve{\Sigma}_{U\Theta}\|_{\mathrm{HS}}^2$ and compute the p-value based on our test. One side-note is that we add a small number, $\epsilon = 10^{-6}$, to the diagonal of a matrix whenever we compute its inverse, square-root or eigenvalues. This is done to avoid numerical instability that can happen in some extreme samples. Figure 1 shows the boxplots of p-values under different combinations of (n, d) in (8.1), and the empirical type-I errors among 100 experiments at $\alpha = 0.1$ are summarized in Table 1.

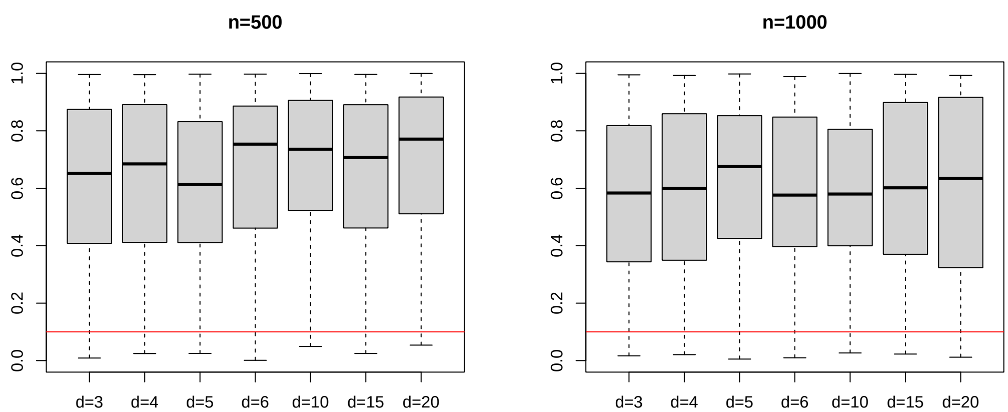

FIG 1. *Boxplots of p-values under the null hypothesis. The red line represents 0.1.*

TABLE 1
Empirical type-I errors (in percentage) at $\alpha = 0.1$.

n	$d=3$	$d=4$	$d=5$	$d=6$	$d=10$	$d=15$	$d=20$
500	6	4	5	4	5	1	2
1000	7	6	3	8	2	7	7

As we can see from Figure 1, the bulk of the p-values are quite large for all combinations of n and d, indicating the elliptical distribution hypothesis is not rejected in the great majority of cases. In Table 1, the empirical type-I errors are all below the significance level, which is consistent with the validity of our test.

8.2. *Results under alternative distributions.* We consider alternative distributions with different degrees of departure from the elliptical distribution. We first generate Z_k independently from $N(0,4)$, and set $Z = (Z_1, \dots, Z_d)^\intercal$. We then randomly select a subset J of $\{1, \dots, d\}$ of cardinality $\lceil d/3 \rceil$ and, for each $k \in J$, we replace Z_k by $W_k - df$, with W_k generated from $\chi^2(df)$. We denote the resulting random vector by $\tilde{Z}$. We then construct X by

$$X = \mu + \Sigma^{1/2}\tilde{Z},$$

where Σ is generated in the same way as it was in the null distribution case, and $\mu = (\mu_1, \dots, \mu_d)^\intercal$ is generated from

$$\mu_k = 40\beta_k - 20, \quad k = 1, \dots, d,$$

where each β_k is independently generated from $\mathrm{Beta}(0.5, 0.5)$. That is, each μ_k is a rescaled and centered Beta variable. We take the degrees of freedom df to be 2 or 4, with $df = 2$ representing stronger departure from ellipticity.

We then perform our proposed test on the i.i.d. samples $X_1, \dots, X_n$ from X. The boxplots in Figure 2 show p-values for $df = 4$ with (n, d) in the range (8.1); those in Figure 3 show p-values for $df = 2$ with (n, d) in the same range. The numerical values of the empirical powers among 100 experiments at $\alpha = 0.1$ are summarized in Table 2.

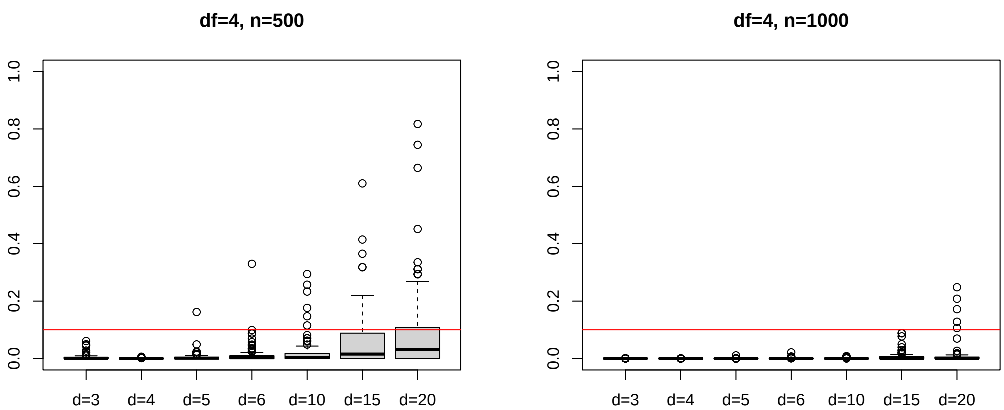

FIG 2. *Boxplots of p-values under the alternative hypothesis with $df = 4$. The red line represents 0.1.*

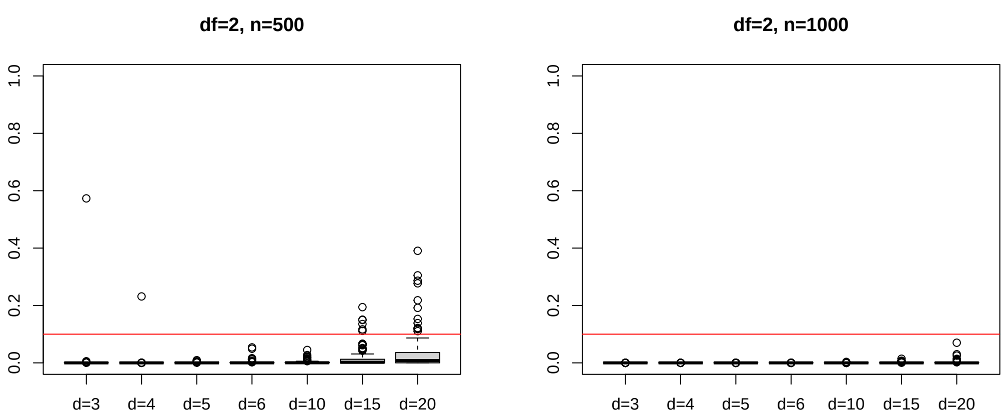

FIG 3. *Boxplots of p-values under the alternative hypothesis with $df = 2$. The red line represents 0.1.*

TABLE 2
Results under the alternative hypothesis

n	df	$d=3$	$d=4$	$d=5$	$d=6$	$d=10$	$d=15$	$d=20$
500	4	100	100	99	99	94	78	73
500	2	99	99	100	100	100	94	88
1000	4	100	100	100	100	100	100	95
1000	2	100	100	100	100	100	100	100

According to Figure 2 and Figure 3, the bulk of the p-values are smaller than 0.1, indicating strong evidence that X does not have an elliptical distribution. This, together with the empirical powers in Table 2, implies the consistency of our test. More specifically, in each figure, when the sample size n is fixed, the boxplots becomes slightly higher as d increases, and the empirical power is lower when d is high, indicating it is more difficult to detect non-ellipticity in higher dimensions. This is reasonable because the skewness might be masked by higher dimensions. If we fix the dimension d and the skewness as represented by df, an increase of sample size n from 500 to 1000 makes the p-values more concentrated around 0. Furthermore, comparing Figure 3 with Figure 2, we see that the p-values for $df = 2$ are smaller than those for $df = 4$, indicating that an increase of skewness makes non-ellipticity more detectable.

9. Application. In this section we apply our test to a dataset concerning a Study on the Efficacy of Nosocomial Infection Control (SENIC Project), which is used in Haley et al. (1980). We download the dataset from https://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/KutnerData/Appendix%20C%20Data%20Sets/APPENC01.txt. This is one of the datasets from the book Kutner et al. (2005).

According to Kutner et al. (2005), the data of 113 hospitals during the 1975-76 study period are sampled from the original 338 hospitals surveyed. For a single hospital, there are 11 variables:

> length of stay, age, infection risk, routine culturing, routine chest x-ray, number of beds, medical school affiliation, region, average daily census, number of nurses, available facilities.

For detailed information about these variables, see Kutner et al. (2005).

We treat the variable "length of stay" as the response variable, and check whether other predictor variables are elliptically distributed. We remove the two categorical variables, "medical school affiliation" and "region". We carry out our test on the dataset where $n = 113$ and $d = 8$ at a significance level $\alpha = 0.05$, and obtained the p-value 0.0014, which leads to rejection of the elliptical distribution hypothesis. The original data is as shown in Figure 1 in Appendix E of the online Supplementary Material (Tang and Li (2024)).

We then perform the Box-Cox transformation on the dataset. Using the method presented in Chapter 7 of Li (2018), we find the optimal λ for the Box-Cox transformation as

$$1.158, 0.737, -0.105, 0.947, -0.316, -0.105, -0.316, 0.947.$$

The transformed variables are shown in Figure 2 in Appendix E of the online Supplementary Material (Tang and Li (2024)). We carry out our test on the transformed data, and obtain the p-value 0.2454. This is much larger than 0.05, leading us to accept the elliptical distribution hypothesis after the Box-Cox transformation.

10. Discussion. For our method, the choice of kernels are indeed important. In general, the kernel functions κ_U and κ_Θ should satisfy the following three conditions.

1. Both κ_U and κ_Θ should be characteristic kernels. This is to guarantee that $\Sigma_{U\Theta} = 0$ if and only if X follows an elliptical distribution. If either of them fails to be characteristic, then there will exist cases when X does not follow an elliptical distribution and yet $\Sigma_{U\Theta} = 0$ is satisfied, which leads to non-consistency of the test.
2. Both κ_U and κ_Θ should be C^2-smooth kernel functions, i.e., the conditions in (5.11) should be satisfied. This is to guarantee that $\Sigma^{\star}_{U\Theta}(z)$ is a member of $\mathcal{H}_U \otimes \mathcal{H}_\Theta$. Also, we need the derivatives $\dot{\kappa}_U(\hat{U}_i, \hat{U}_j)$ and $\dot{\kappa}_\Theta(\hat{\Theta}_i, \hat{\Theta}_j)$ in the implementation, which requires differentiability in κ_U and κ_Θ.
3. From a computational perspective, a product-type kernel for κ_Θ is preferred. This is because, otherwise, we will need to compute n numerical integrals of dimension $(d-1)$ in equation (4.4) and one numerical integral of dimension $(2d-2)$ in equation (4.5). The product-type kernels allow us to replace these high-dimension numerical integration by 1- or 2-dimensional integration.

Clearly, the Gaussian kernel satisfies all three conditions above, so it is preferred. It also works well in our simulation studies. However, many other kernels also satisfy the three conditions. We can construct a broad class of product kernels that are computationally feasible for our method.

Specifically, for any $d-1$ kernel functions on $\mathbb{R} \times \mathbb{R}$, say $\kappa_{\Theta_1}, \dots, \kappa_{\Theta_{d-1}}$, their product given by (4.3) is still a reproducing kernel. Furthermore, by Theorem 1, the product of characteristic kernels is still characteristic. For example, we can use the product-type inverse-quadratic (PIQ) kernels defined as follows:

$$\kappa_U(u, u') = \frac{1}{1 + \gamma_U (u - u')^2}, \quad \kappa_\Theta(\theta, \theta') = \prod_{j=1}^{d-1} \frac{1}{1 + \gamma_\Theta (\theta_j - \theta'_j)^2}, \tag{10.1}$$

which also satisfy the above three requirements. In principle, all such kernels can be used for our purpose. This gives us broad choices for kernels.Further discussions on the choice of kernels are given in the Supplementary Materials.

Acknowledgments. The authors would like to thank two referees and an Associate Editor for their insightful comments and suggestions, which helped us greatly in improving this work.

Funding. The research of Bing Li is supported in part by the U.S. National Science Foundation (NSF) Grant DMS-2210775 and the U.S. National Institutes of Health (NIH) grant 1 R01 GM152812-01.

SUPPLEMENTARY MATERIAL

Supplementary Material for "A nonparametric test for elliptical distribution based on kernel embedding of probabilities"
Most of the proofs, additional simulation studies, further discussions, and the scatter plot matrices for Section 9 can be found in the Supplementary Material.

REFERENCES

ALBISETTI, I., BALABDAOUI, F. and HOLZMANN, H. (2020). Testing for spherical and elliptical symmetry. *Journal of Multivariate Analysis* **180** 104667. https://doi.org/10.1016/j.jmva.2020.104667

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Inc. Hoboken, New Jersey.

BABIĆ, S., LEY, C. and PALANGETIĆ, M. (2021). The R Journal: Elliptical Symmetry Tests in R. *The R Journal* **13** 661-672. https://doi.org/10.32614/RJ-2021-078. https://doi.org/10.32614/RJ-2021-078

BABIĆ, S., GELBGRAS, L., HALLIN, M. and LEY, C. (2021). Optimal tests for elliptical symmetry: Specified and unspecified location. *Bernoulli* **27** 2189 – 2216. https://doi.org/10.3150/20-BEJ1305

BARINGHAUS, L. (1991). Testing for Spherical Symmetry of a Multivariate Distribution. *The Annals of Statistics* **19** 899–917.

CAMBANIS, S., HUANG, S. and SIMONS, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis* **11** 368-385. https://doi.org/10.1016/0047-259X(81)90082-8

CASSART, D., HALLIN, M. and PAINDAVEINE, D. (2008). Optimal detection of Fechner-asymmetry. *Journal of Statistical Planning and Inference* **138** 2499-2525. https://doi.org/10.1016/j.jspi.2007.10.011

COOK, R. D. and LI, B. (2002). Dimension Reduction for Conditional Mean in Regression. *The Annals of Statistics* **30** 455–474.

DUCHESNE, P. and DE MICHEAUX, P. L. (2010). Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. *Computational Statistics and Data Analysis* **54** 858-862.

EATON, M. L. (1986). A characterization of spherical distributions. *Journal of Multivariate Analysis* **20** 272–276.

FERNHOLZ, L. T. (1983). *von Mises Calculus For Statistical Functionals*. *Lecture Notes in Statistics*. Springer New York.

FUKUMIZU, K., GRETTON, A., LANCKRIET, G., SCHÖLKOPF, B. and SRIPERUMBUDUR, B. K. (2009). Kernel Choice and Classifiability for RKHS Embeddings of Probability Distributions. In *Advances in Neural Information Processing Systems* (Y. BENGIO, D. SCHUURMANS, J. LAFFERTY, C. WILLIAMS and A. CULOTTA, eds.) **22**. Curran Associates, Inc.

GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. (2005). Measuring Statistical Dependence with Hilbert-Schmidt Norms. In *Algorithmic Learning Theory* (S. JAIN, H. U. SIMON and E. TOMITA, eds.) 63–77. Springer Berlin Heidelberg, Berlin, Heidelberg.

GRETTON, A., BORGWARDT, K. M., RASCH, M., SCHÖLKOPF, B. and SMOLA, A. J. (2007). A Kernel Method for the Two-Sample-Problem. In *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference* The MIT Press. https://doi.org/10.7551/mitpress/7503.003.0069

GRETTON, A., FUKUMIZU, K., TEO, C., SONG, L., SCHÖLKOPF, B. and SMOLA, A. (2008). A Kernel Statistical Test of Independence. In *Advances in Neural Information Processing Systems* (J. PLATT, D. KOLLER, Y. SINGER and S. ROWEIS, eds.) **20**. Curran Associates, Inc.

GRETTON, A., FUKUMIZU, K., HARCHAOUI, Z. and SRIPERUMBUDUR, B. K. (2009). A Fast, Consistent Kernel Two-Sample Test. In *Advances in Neural Information Processing Systems* (Y. BENGIO, D. SCHUURMANS, J. LAFFERTY, C. WILLIAMS and A. CULOTTA, eds.) **22**. Curran Associates, Inc.

GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A Kernel Two-Sample Test. *Journal of Machine Learning Research* **13** 723-773.

GUELLA, J. C. (2022). On Gaussian kernels on Hilbert spaces and kernels on hyperbolic spaces. *Journal of Approximation Theory* **279** 105765. https://doi.org/10.1016/j.jat.2022.105765

HALEY, R. W., QUADE, D., FREEMAN, H. E. and BENNETT, J. V. (1980). The SENIC Project. Study on the efficacy of nosocomial infection control (SENIC Project). Summary of study design. *American Journal of Epidemiology* **111** 472-485. https://doi.org/10.1093/oxfordjournals.aje.a112928

HENZE, N., HLÁVKA, Z. and MEINTANIS, S. G. (2014). Testing for spherical symmetry via the empirical characteristic function. *Statistics* **48** 1282-1296. https://doi.org/10.1080/02331888.2013.832764

HUFFER, F. W. and PARK, C. (2007). A test for elliptical symmetry. *Journal of Multivariate Analysis* **98** 256-281. https://doi.org/10.1016/j.jmva.2005.09.011

KARIYA, T. and EATON, M. L. (1977). Robust Tests for Spherical Symmetry. *The Annals of Statistics* **5** 206–215.

KOKOSZKA, P. and REIMHERR, M. (2017). *Introduction to Functional Data Analysis*. *Chapman & Hall/CRC Texts in Statistical Science*. CRC Press.

KOLTCHINSKII, V. I. and LI, L. (1998). Testing for Spherical Symmetry of a Multivariate Distribution. *Journal of Multivariate Analysis* **65** 228-244. https://doi.org/10.1006/jmva.1998.1743

KUTNER, M., NACHTSHEIM, C., NETER, J. and LI, W. (2005). *Applied Linear Statistical Models*. *McGrwa-Hill international edition*. McGraw-Hill Irwin.

LI, K.-C. (1991). Sliced Inverse Regression for Dimension Reduction. *Journal of the American Statistical Association* **86** 316–327.

LI, B. (2018). *Sufficient Dimension Reduction: Methods and Applications with R*. *Chapman & Hall/CRC Monographs on Statistics and Applied Probability*. CRC Press.

LI, B. and DONG, Y. (2009). Dimension reduction for nonelliptically distributed predictors. *The Annals of Statistics* **37** 1272 – 1298. https://doi.org/10.1214/08-AOS598

LI, K.-C. and DUAN, N. (1989). Regression Analysis Under Link Violation. *The Annals of Statistics* **17** 1009 – 1052. https://doi.org/10.1214/aos/1176347254

LI, B. and SOLEA, E. (2018). A Nonparametric Graphical Model for Functional Data With Application to Brain Networks Based on fMRI. *Journal of the American Statistical Association* **113** 1637-1655. https://doi.org/10.1080/01621459.2017.1356726

LIANG, J., FANG, K.-T. and HICKERNELL, F. J. (2008). Some necessary uniform tests for spherical symmetry. *Annals of the Institute of Statistical Mathematics* **60** 679–696. https://doi.org/10.1007/s10463-007-0121-9

LIU, H., HAN, F. and ZHANG, C.-H. (2012). Transelliptical Graphical Models. In *Advances in Neural Information Processing Systems* (F. PEREIRA, C. J. BURGES, L. BOTTOU and K. Q. WEINBERGER, eds.) **25**. Curran Associates, Inc.

MANZOTTI, A., PÉREZ, F. J. and QUIROZ, A. J. (2002). A Statistic for Testing the Null Hypothesis of Elliptical Symmetry. *Journal of Multivariate Analysis* **81** 274-285. https://doi.org/10.1006/jmva.2001.2007

NARASIMHAN, B., JOHNSON, S. G., HAHN, T., BOUVIER, A. and KIÊU, K. (2023). cubature: Adaptive Multivariate Integration over Hypercubes R package version 2.0.4.6.

PAINDAVEINE, D. (2012). Elliptical Symmetry. *Encyclopedia of Environmetrics* 802–807.

QIU, W. and JOE., H. (2020). clusterGeneration: Random Cluster Generation (with Specified Degree of Separation) R package version 1.3.7.

SCHMIDT, R. (2002). Tail dependence for elliptically contoured distributions. *Mathematical Methods of Operations Research* **55** 301-327. https://doi.org/10.1007/s001860200191

SCHÖLKOPF, B., HERBRICH, R. and SMOLA, A. J. (2001). A Generalized Representer Theorem. In *Computational Learning Theory* (D. HELMBOLD and B. WILLIAMSON, eds.) 416–426. Springer Berlin Heidelberg, Berlin, Heidelberg.

SCHOTT, J. R. (2002). Testing for elliptical symmetry in covariance-matrix-based analyses. *Statistics & Probability Letters* **60** 395-404. https://doi.org/10.1016/S0167-7152(02)00306-1

SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of Distance-Based and RKHS-Based Statistics in Hypothesis Testing. *The Annals of Statistics* **41** 2263–2291.

SRIPERUMBUDUR, B., FUKUMIZU, K. and LANCKRIET, G. (2010). On the relation between universality, characteristic kernels and RKHS embedding of measures. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (Y. W. TEH and M. TITTERINGTON, eds.). *Proceedings of Machine Learning Research* **9** 773–780. PMLR, Chia Laguna Resort, Sardinia, Italy.

SRIPERUMBUDUR, B. K., FUKUMIZU, K. and LANCKRIET, G. R. G. (2011). Universality, Characteristic Kernels and RKHS Embedding of Measures. *Journal of Machine Learning Research* **12** 2389–2410.

SRIPERUMBUDUR, B., GRETTON, A., FUKUMIZU, K., SCHÖLKOPF, B. and LANCKRIET, G. (2010). Hilbert Space Embeddings and Metrics on Probability Measures. *Journal of Machine Learning Research* **11** 1517-1561.

SZABÓ, Z. and SRIPERUMBUDUR, B. K. (2018). Characteristic and Universal Tensor Product Kernels. *Journal of Machine Learning Research* **18** 1–29.

SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics* **35** 2769 – 2794. https://doi.org/10.1214/009053607000000505

SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *The Annals of Applied Statistics* **3** 1236 – 1265. https://doi.org/10.1214/09-AOAS312

TANG, Y. and LI, B. (2024). Supplementary material for "A nonparametric test for elliptical distribution based on kernel embedding of probabilities".

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511802256

VOGEL, D. and FRIED, R. (2011). Elliptical graphical modelling. *Biometrika* **98** 935–951.

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68** 49–67.

ZHOU, D.-X. (2008). Derivative reproducing properties for kernel methods in learning theory. *Journal of Computational and Applied Mathematics* **220** 456–463.