IMS Collections
High Dimensional Probability V: The Luminy Volume
Vol. 5 (2009) 153–178

DOI: 10.1214/09-IMSCOLL511

On Stein's method for multivariate normal approximation

Elizabeth Meckes[1,*]

Case Western Reserve University

Abstract: The purpose of this paper is to synthesize the approaches taken by Chatterjee-Meckes and Reinert-Röllin in adapting Stein's method of exchangeable pairs for multivariate normal approximation. The more general linear regression condition of Reinert-Röllin allows for wider applicability of the method, while the method of bounding the solution of the Stein equation due to Chatterjee-Meckes allows for improved convergence rates. Two abstract normal approximation theorems are proved, one for use when the underlying symmetries of the random variables are discrete, and one for use in contexts in which continuous symmetry groups are present. A first application is presented to projections of exchangeable random vectors in $\mathbb{R}^n$ onto one-dimensional subspaces. The application to runs on the line from Reinert-Röllin is reworked to demonstrate the improvement in convergence rates, and a new application to joint value distributions of eigenfunctions of the Laplace-Beltrami operator on a compact Riemannian manifold is presented.

1. Introduction

In 1972, Charles Stein [28] introduced a powerful new method for estimating the distance from a probability distribution on $\mathbb{R}$ to a Gaussian distribution. Central to the method was the notion of a characterizing operator: Stein observed that the standard normal distribution was the unique probability distribution μ with the property that

$$\int \big[f'(x) - xf(x)\big]\mu(dx) = 0$$

for all f for which the left-hand side exists and is finite. The operator T_o defined on C^1 functions by

$$T_o f(x) = f'(x) - xf(x)$$

is called the characterizing operator of the standard normal distribution. The left-inverse to T_o, denoted U_o, is defined by the equation

$$T_o(U_o f)(x) = f(x) - \mathbb{E}f(Z),$$

where Z is a standard normal random variable; the boundedness properties of U_o are an essential ingredient of Stein's method.

Stein and many other authors continued to develop this method; in 1986, Stein published the book [29], which laid out his approach to the method, called the

---

*Research supported by an American Institute of Mathematics Five-year Fellowship.

[1]Department of Mathematics, Case Western Reserve University, 10900 Euclid Ave., Cleveland, OH 44106.

AMS 2000 subject classifications: Primary 60F05; secondary 60D05.

Keywords and phrases: multivariate analysis, normal approximation, Stein's method, eigenfunctions, Laplacian.

method of exchangeable pairs, in detail. Stein's method has proved very useful in situations in which local dependence or weak global dependence are present. One of the chief advantages of the method is that it is specifically a method for bounding the distance from a fixed distribution to Gaussian, and thus automatically produces concrete error bounds in limit theorems. The method is most naturally formulated by viewing probability measures as dual to various classes of functions, so that the notions of distance that arise are those which can be expressed as differences of expectations of test functions (e.g., the total variation distance, Wasserstein distance, or bounded Lipschitz distance). Several authors (particularly Bolthausen [2], Götze [8], Rinott and Rotar [24], and Shao and Su [27]) have extended the method to non-smooth test functions, such as indicator functions of intervals in $\mathbb{R}$ and indicator functions of convex sets in $\mathbb{R}^k$.

Heuristically, the univariate method of exchangeable pairs goes as follows. Let W be a random variable conjectured to be approximately Gaussian; assume that $\mathbb{E}W = 0$ and $\mathbb{E}W^2 = 1$. From W, construct a new random variable W' such that the pair (W, W') has the same distribution as (W', W). This is usually done by making a "small random change" in W, so that W and W' are close. Let $\Delta = W' - W$. If it can be verified that there is a $\lambda > 0$ such that

$$\mathbb{E}\left[\Delta\middle|W\right] = -\lambda W + E_1, \tag{1}$$

$$\mathbb{E}\left[\Delta^2\middle|W\right] = 2\lambda + E_2, \tag{2}$$

$$\mathbb{E}|\Delta|^3 = E_3, \tag{3}$$

with the random quantities E_1, E_2 and the deterministic quantity E_3 being small compared to λ, then W is indeed approximately Gaussian, and its distance to Gaussian (in some metric) can be bounded in terms of the E_i and λ.

While there had been successful uses of multivariate versions of Stein's method for normal approximation in the years following the introduction of the univariate method (e.g., by Götze [8], Rinott and Rotar [24], [25], and Raič [21]), there had not until recently been a version of the method of exchangeable pairs for use in a multivariate setting. This was first addressed in joint work by the author with S. Chatterjee [3], where several abstract normal approximation theorems, for approximating by standard Gaussian random vectors, were proved. The theorems were applied to estimate the rate of convergence in the multivariate central limit theorem and to show that rank k projections of Haar measure on the orthogonal group $\mathcal{O}_n$ and the unitary group $\mathcal{U}_n$ are close to Gaussian measure on $\mathbb{R}^k$ (respectively $\mathbb{C}^k$), when $k = o(n)$. The condition in the theorems of [3] corresponding to condition (1) above was that, for an exchangeable pair of random vectors (X, X'),

$$\mathbb{E}\left[X' - X\middle|X\right] = -\lambda X. \tag{4}$$

The addition of a random error to this equation was not needed in the applications in [3], but is a straightforward modification of the theorems proved there.

After the initial draft of [3] appeared on the ArXiv, a preprint was posted by Reinert and Röllin [23] which generalized one of the abstract normal approximation theorems of [3]. Instead of condition (4) above, they required

$$\mathbb{E}\left[X' - X\middle|X\right] = -\Lambda X + E, \tag{5}$$

where Λ is a positive definite matrix and E is a random error. This more general condition allowed their abstract approximation theorem to be used directly to estimate the distance to Gaussian random vectors with non-identity (even singular) covariance matrices. They then used what they call "the embedding method" for approximating real random variables by the normal distribution, by observing that in many cases in which the condition (1) does not hold, the random variable in question can be viewed as one component of a random vector which satisfies condition (5) with a non-diagonal Λ. Many examples are given, both of the embedding method and the multivariate normal approximation theorem directly, including applications to runs on the line, statistics of Bernoulli random graphs, U-statistics, and doubly-indexed permutation statistics. The embedding approach used in [23] is similar to the approach taken by Janson in a continuous time setting in [11], where results about statistics of random graphs originally proved by Ruciński [26], Barbour, Karoński, and Ruciński [1] and Janson and Nowicki [12] were reproved by viewing the random graphs as particular time instants of a stochastic process.

After [23] was posted, [3] underwent significant revisions, largely to change the metrics which were used on the space of probability measures on $\mathbb{R}^k$ and $\mathbb{C}^k$. As mentioned above, Stein's method works most naturally to compare measures by using (usually smooth) classes of test functions. The smoothness conditions used by Reinert and Röllin, and those initially used in [3], are to assume bounds on the quantities

$$|h|_r := \sup_{1 \le i_1, \ldots, i_r \le k} \left\| \frac{\partial^r h}{\partial x_{i_1} \cdots \partial x_{i_r}} \right\|_\infty .$$

The approach taken in the published version of [3] is to give smoothness conditions instead by requiring bounds on the quantities

$$M_r(h) := \sup_{x \in \mathbb{R}^k} \|D^r h(x)\|_{op},$$

where $\|D^r h(x)\|_{op}$ is the operator norm of the r-th derivative of h, as an r-linear form. These smoothness conditions seem preferable for several reasons. Firstly, they are more geometrically natural, as they are coordinate-free; they depend only on distances and not on the choice of orthonormal basis of $\mathbb{R}^k$. Particularly when approximating by the standard Gaussian distribution on $\mathbb{R}^k$, which is of course rotationally invariant, it seems desirable to have a notion of distance which is also rotationally invariant. In more practical terms, considering classes of functions defined in terms of bounds on the quantities M_r and modifying the proofs of the abstract theorems accordingly allows for improved error bounds. The original bound on the Wasserstein distance from a k-dimensional projection of Haar measure on $\mathcal{O}_n$ to standard Gauss measure from the first version of [3] was $c\frac{k^{3/2}}{n}$, while the coordinate-free viewpoint allowed the bound to be improved to $c\frac{k}{n}$ (in the same metric). In Section 3 below, the example of runs on the line from [23] is reworked with this viewpoint, with essentially the same ingredients, to demonstrate that the rates of convergence obtained are improved. Finally, most of the bounds in [3] and below, and those from the main theorem in [23] require two or three derivatives, so that an additional smoothing argument is needed to move to one of the more usual metrics on probability measures (e.g. Wasserstein distance, total variation distance, or bounded Lipschitz distance). Starting from bounds in terms of the $M_r(h)$ instead of the $|h|_r$ typically produces better results in the final metric; compare, e.g., Proposition 3.2 of the original ArXiv version of the paper [20] of M. Meckes with Corollary 3.5 of the published version, in which one of the abstract approximation theorems of [3] was applied to the study of the distribution of marginals of the uniform measure on high-dimensional convex bodies.

The purpose of this paper is to synthesize the approaches taken by the author and Chatterjee in [3] and Reinert and Röllin in [23]. In Section 2, two preliminary lemmas are proved, identifying a characterizing operator for the Gaussian distribution on $\mathbb{R}^k$ with covariance matrix Σ and bounding the derivatives of its left-inverse in terms of the quantities M_r. Then, two abstract normal approximation theorems are proved. The first is a synthesis of Theorem 2.3 of [3] and Theorem 2.1 of [23], in which the distance from X to a Gaussian random variable with mean zero and covariance Σ is bounded, for X the first member of an exchangeable pair (X, X') satisfying condition (5) above. The second approximation theorem is analogous to Theorem 2.4 of [3], and is for situations in which the underlying random variable possesses "continuous symmetries." A condition similar to (5) is used in that theorem as well. Finally, in Section 3, three applications are carried out. The first is an illustration of the use of the method, which shows that projections of exchangeable random vectors in $\mathbb{R}^n$ onto k-dimensional subspaces are usually close to Gaussian if the dependence among the components is not too great; "usually" refers to which subspaces one can project onto, and the error is given in terms of ℓ_3 norms of a spanning set of vectors for the subspace. The second application is simply a reworking of the runs on the line example of [23], making use of their analysis together with Theorem 3 below to obtain a better rate of convergence. Lastly, an application is given to the joint value distribution of a finite sequence of orthonormal eigenfunctions of the Laplace-Beltrami operator on a compact Riemannian manifold. This is a multivariate version of the main theorem of [17]. As an example, the error bound of this theorem is computed explicitly for a certain class of flat tori.

1.1. Notation and conventions

The Wasserstein distance $d_W(X, Y)$ between the random variables X and Y is defined by

$$d_W(X, Y) = \sup_{M_1(g) \le 1} \left|\mathbb{E}g(X) - \mathbb{E}g(Y)\right|,$$

where $M_1(g) = \sup_{x \neq y} \frac{|g(x)-g(y)|}{|x-y|}$ is the Lipschitz constant of g. On the space of probability distributions with finite absolute first moment, Wasserstein distance induces a stronger topology than the usual one described by weak convergence, but not as strong as the topology induced by the total variation distance. See [5] for detailed discussion of the various notions of distance between probability distributions.

We will use $\mathfrak{N}(\mu, \Sigma)$ to denote the normal distribution on $\mathbb{R}^k$ with mean μ and covariance matrix Σ; unless otherwise stated, the random variable $Z = (Z_1, \ldots, Z_k)$ is understood to be a standard Gaussian random vector on $\mathbb{R}^k$.

In $\mathbb{R}^n$, the Euclidean inner product is denoted $\langle \cdot, \cdot \rangle$ and the Euclidean norm is denoted $|\cdot|$. On the space of real $n \times n$ matrices, the Hilbert-Schmidt inner product is defined by

$$\langle A, B \rangle_{H.S.} = \mathrm{Tr}\,(AB^T),$$

with corresponding norm

$$\|A\|_{H.S.} = \sqrt{\mathrm{Tr}\,(AA^T)}.$$

The operator norm of a matrix A over $\mathbb{R}$ is defined by

$$\|A\|_{op} = \sup_{|v|=1,|w|=1} |\langle Av, w\rangle|.$$

More generally, if A is a k-linear form on $\mathbb{R}^n$, the operator norm of A is defined to be

$$\|A\|_{op} = \sup\{|A(u_1,\ldots,u_k)| : |u_1| = \cdots = |u_n| = 1\}.$$

The $n \times n$ identity matrix is denoted I_n and the $n \times n$ matrix of all zeros is denoted 0_n.

For Ω a domain in $\mathbb{R}^n$, the notation $C^k(\Omega)$ will be used for the space of k-times continuously differentiable real-valued functions on Ω, and $C_o^k(\Omega) \subseteq C^k(\Omega)$ are those C^k functions on Ω with compact support. The k-th derivative $D^k f(x)$ of a function $f \in C^k(\mathbb{R}^n)$ is a k-linear form on $\mathbb{R}^n$, given in coordinates by

$$\left\langle D^k f(x), (u_1,\ldots,u_k)\right\rangle = \sum_{i_1,\ldots,i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_1}\cdots\partial x_{i_k}}(x)(u_1)_{i_1}\cdots(u_k)_{i_k},$$

where $(u_i)_j$ denotes the j-th component of the vector u_i. For an intrinsic, coordinate-free developement, see Federer [6]. For $f : \mathbb{R}^n \to \mathbb{R}$, sufficiently smooth, let

$$M_k(f) := \sup_{x\in\mathbb{R}^n} \|D^k f(x)\|_{op}. \tag{6}$$

In the case $k = 2$, define

$$\widetilde{M}_2(f) := \sup_{x\in\mathbb{R}^n} \|\mathrm{Hess}\, f(x)\|_{H.S.}. \tag{7}$$

Note also that

$$M_k(f) = \sup_{x\neq y} \frac{\|D^{k-1}f(x) - D^{k-1}f(y)\|_{op}}{|x-y|};$$

that is, $M_k(f)$ is the Lipschitz constant of the $k-1$-st derivative of f.

This general definition of M_k is a departure from what was done by Raič in [22]; there, smoothness conditions on functions are also given in coordinate-independent ways, and M_1 and M_2 are defined as they are here, but in case $k = 3$, the quantity M_3 is defined as the Lipschitz constant of the Hessian with respect to the Hilbert-Schmidt norm as opposed to the operator norm.

2. Abstract Approximation Theorems

This section contains the basic lemmas giving the Stein characterization of the multivariate Gaussian distribution and bounds to the solution of the Stein equation, together with two multivariate abstract normal approximation theorems and their proofs. The first theorem is a reworking of the theorem of Reinert and Röllin on multivariate normal approximation with the method of exchangeable pairs for vectors with non-identity covariance. The second is an analogous result in the context of "continuous symmetries" of the underlying random variable, as has been previously studied by the author in [18], [17], and (jointly with S. Chatterjee) in [3].

The following lemma gives a second-order characterizing operator for the Gaussian distribution with mean 0 and covariance Σ on $\mathbb{R}^d$. The characterizing operator for this distribution is already well-known. The proofs available in the literature generally rely on viewing the Stein equation in terms of the generator of the Ornstein-Uhlenbeck semi-group; the proof given here is direct.

Lemma 1. *Let $Z \in \mathbb{R}^d$ be a random vector with $\{Z_i\}_{i=1}^d$ independent, identically distributed standard Gaussian random variables, and let $Z_\Sigma = \Sigma^{1/2} Z$ for a symmetric, non-negative definite matrix Σ.*

1. *If $f : \mathbb{R}^d \to \mathbb{R}$ is two times continuously differentiable and compactly supported, then*
$$\mathbb{E}\big[\langle \operatorname{Hess} f(Z_\Sigma), \Sigma \rangle_{H.S.} - \langle Z_\Sigma, \nabla f(Z_\Sigma) \rangle \big] = 0.$$
2. *If $Y \in \mathbb{R}^d$ is a random vector such that*
$$\mathbb{E}\big[\langle \operatorname{Hess} f(Y), \Sigma \rangle_{H.S.} - \langle Y, \nabla f(Y) \rangle \big] = 0$$
for every $f \in C^2(\mathbb{R}^d)$ with $\mathbb{E}|\langle \operatorname{Hess} f(Y), \Sigma \rangle_{H.S.} - \langle Y, \nabla f(Y) \rangle| < \infty$, then $\mathcal{L}(Y) = \mathcal{L}(Z_\Sigma)$.
3. *If $g \in C^\infty(\mathbb{R}^d)$, then the function*
$$U_o g(x) := \int_0^1 \frac{1}{2t} \big[\mathbb{E} g(\sqrt{t} x + \sqrt{1-t} Z_\Sigma) - \mathbb{E} g(Z_\Sigma) \big] dt \tag{8}$$
is a solution to the differential equation
$$\langle x, \nabla h(x) \rangle - \langle \operatorname{Hess} h(x), \Sigma \rangle_{H.S.} = g(x) - \mathbb{E} g(Z_\Sigma). \tag{9}$$

Proof. Part (1) follows from integration by parts.

Part (2) follows easily from part (3): note that if $\mathbb{E}[\langle \operatorname{Hess} f(Y), \Sigma \rangle_{H.S.} - \langle Y, \nabla f(Y) \rangle] = 0$ for every $f \in C^2(\mathbb{R}^d)$ with $\mathbb{E}|\langle \operatorname{Hess} f(Y), \Sigma \rangle_{H.S.} - \langle Y, \nabla f(Y) \rangle| < \infty$, then for $g \in C_o^\infty$ given,
$$\mathbb{E} g(Y) - \mathbb{E} g(Z) = \mathbb{E}\big[\langle \operatorname{Hess} (U_o g)(Y), \Sigma \rangle_{H.S.} - \langle Y, \nabla (U_o g)(Y) \rangle \big] = 0,$$
and so $\mathcal{L}(Y) = \mathcal{L}(Z)$ since C^∞ is dense in the class of bounded continuous functions, with respect to the supremum norm.

For part (3), first note that since g is Lipschitz, if $t \in (0, 1)$
$$\begin{aligned}\left| \frac{1}{2t} \big[\mathbb{E} g(\sqrt{t} x + \sqrt{1-t} \Sigma^{1/2} Z) - \mathbb{E} g(\Sigma^{1/2} Z) \big] \right| &\le \frac{L}{2t} \mathbb{E} \left| \sqrt{t} x + (\sqrt{1-t} - 1) \Sigma^{1/2} Z \right| \\ &\le \frac{L}{2t} \left[\sqrt{t} |x| + t \sqrt{\operatorname{Tr}(\Sigma)} \right],\end{aligned}$$
which is integrable on $(0, 1)$, so the integral exists by the dominated convergence theorem.

To show that $U_o g$ is indeed a solution to the differential equation (9), let
$$Z_{x,t} = \sqrt{t} x + \sqrt{1-t} \Sigma^{1/2} Z$$
and observe that
$$\begin{aligned} g(x) - \mathbb{E} g(\Sigma^{1/2} Z) &= \int_0^1 \frac{d}{dt} \mathbb{E} g(Z_{x,t}) dt \\ &= \int_0^1 \frac{1}{2\sqrt{t}} \mathbb{E}(x \cdot \nabla g(Z_t)) dt - \int_0^1 \frac{1}{2\sqrt{1-t}} \mathbb{E} \left\langle \Sigma^{1/2} Z, \nabla g(Z_t) \right\rangle dt \\ &= \int_0^1 \frac{1}{2\sqrt{t}} \mathbb{E}(x \cdot \nabla g(Z_t)) dt - \int_0^1 \frac{1}{2} \mathbb{E} \left\langle \operatorname{Hess} g(Z_t), \Sigma \right\rangle_{H.S.} dt \end{aligned}$$

by integration by parts. Noting that

$$\mathrm{Hess}\,(U_o g)(x) = \int_0^1 \frac{1}{2}\mathbb{E}\big[\mathrm{Hess}\,g(Z_t)\big]\,dt$$

and

$$x \cdot \nabla(U_o g)(x) = \int_0^1 \frac{1}{2\sqrt{t}}\mathbb{E}(x \cdot \nabla g(Z_t))dt$$

completes part 3. $\square$

The next lemma gives useful bounds on $U_o g$ and its derivatives in terms of g and its derivatives. As in [22], bounds are most naturally given in terms of the quantities $M_i(g)$ defined in the introduction.

Lemma 2. *For $g : \mathbb{R}^d \to \mathbb{R}$ given, $U_o g$ satisfies the following bounds:*

1.
$$M_k(U_o g) \le \frac{1}{k} M_k(g) \qquad \forall k \ge 1.$$

2.
$$\widetilde{M}_2(U_o g) \le \frac{1}{2}\widetilde{M}_2(g).$$

If, in addition, Σ is positive definite, then

3.
$$M_1(U_o g) \le M_o(g)\|\Sigma^{-1/2}\|_{op}\sqrt{\frac{\pi}{2}}.$$

4.
$$\widetilde{M}_2(U_o g) \le \sqrt{\frac{2}{\pi}} M_1(g)\|\Sigma^{-1/2}\|_{op}.$$

5.
$$M_3(U_o g) \le \frac{\sqrt{2\pi}}{4} M_2(g)\|\Sigma^{-1/2}\|_{op}.$$

Remark. Bounds (2), (2), and (2) are mainly of use when Σ has a fairly simple form, since they require an estimate for $\|\Sigma^{-1/2}\|_{op}$. They are also of theoretical interest, since they show that if Σ is non-singular, then the operator U_o is smoothing; functions $U_o g$ are typically one order smoother than g. The bounds (1) and (2), while not showing the smoothing behavior of U_o, are useful when Σ is complicated (or singular) and an estimate of $\|\Sigma^{-1/2}\|_{op}$ is infeasible or impossible.

Proof of Lemma 2. Write $h(x) = U_o g(x)$ and $Z_{x,t} = \sqrt{t}x + \sqrt{1-t}\Sigma^{1/2}Z$. Note that by the formula for $U_o g$,

$$\frac{\partial^r h}{\partial x_{i_1}\cdots\partial x_{i_r}}(x) = \int_0^1 (2t)^{-1}t^{r/2}\mathbb{E}\left[\frac{\partial^r g}{\partial x_{i_1}\cdots\partial x_{i_r}}(Z_{x,t})\right]dt. \tag{10}$$

Thus

$$\big\langle D^k(U_o g)(x),(u_1,\dots,u_k)\big\rangle = \int_0^1 \frac{t^{\frac{k}{2}-1}}{2}\mathbb{E}\big[\big\langle D^k g(Z_{x,t}),(u_1,\dots,u_k)\big\rangle\big]dt$$

for unit vectors $u_1,\dots,u_k$, and part (1) follows immediately.

For the second part, note that (10) implies that

$$\operatorname{Hess} h(x) = \frac{1}{2}\int_0^1 \mathbb{E}\left[\operatorname{Hess} g(Z_{x,t})\right] dt.$$

Fix a $d \times d$ matrix A. Then

$$\begin{aligned}\left|\langle \operatorname{Hess} h(x), A\rangle_{H.S.}\right| &\le \frac{1}{2}\int_0^1 \mathbb{E}\left|\langle \operatorname{Hess} g(Z_{x,t}), A\rangle_{H.S.}\right| dt \\ &\le \frac{1}{2}\left(\sup_x \|\operatorname{Hess} g(x)\|_{H.S.}\right)\|A\|_{H.S.},\end{aligned}$$

hence part (2).

For part (2), note that it follows by integration by parts on the Gaussian expectation that

$$\begin{aligned}\frac{\partial h}{\partial x_i}(x) &= \int_0^1 \frac{1}{2\sqrt{t}}\mathbb{E}\left[\frac{\partial g}{\partial x_i}(\sqrt{t}x + \sqrt{1-t}\Sigma^{1/2}Z)\right] dt \\ &= \int_0^1 \frac{1}{2\sqrt{t(1-t)}}\mathbb{E}\left[(\Sigma^{-1/2}Z)_i g(\sqrt{t}x + \sqrt{1-t}\Sigma^{1/2}Z)\right] dt,\end{aligned}$$

thus

$$\nabla h(x) = \int_0^1 \frac{1}{2\sqrt{t(1-t)}}\mathbb{E}\left[g(Z_{x,t})\Sigma^{-1/2}Z\right] dt,$$

and so

$$M_1(h) \le \|g\|_\infty \mathbb{E}\left|\Sigma^{-1/2}Z\right| \int_0^1 \frac{1}{2\sqrt{t(1-t)}}dt.$$

Now, $\mathbb{E}|\Sigma^{-1/2}Z| \le \|\Sigma^{-1/2}\|_{op}\mathbb{E}|Z_1| = \|\Sigma^{-1/2}\|_{op}\sqrt{\frac{2}{\pi}}$, since $\Sigma^{-1/2}Z$ is a univariate Gaussian random variable, and $\int_0^1 \frac{1}{2\sqrt{t(1-t)}} = \frac{\pi}{2}$. This completes part (2).

For part (2), again using integration by parts on the Gaussian expectation,

$$\begin{aligned}\frac{\partial^2 h}{\partial x_i \partial x_j}(x) &= \int_0^1 \frac{1}{2}\mathbb{E}\left[\frac{\partial^2 g}{\partial x_i \partial x_j}(\sqrt{t}x + \sqrt{1-t}\Sigma^{1/2}Z)\right] dt \\ &= \int_0^1 \frac{1}{2\sqrt{1-t}}\mathbb{E}\left[\left[\Sigma^{-1/2}Z\right]_i \frac{\partial g}{\partial x_j}(Z_{x,t})\right] dt,\end{aligned} \tag{11}$$

and so

$$\operatorname{Hess} h(x) = \int_0^1 \frac{1}{2\sqrt{1-t}}\mathbb{E}\left[\Sigma^{-1/2}Z\left(\nabla g(Z_{x,t})\right)^T\right] dt. \tag{12}$$

Fix a $d \times d$ matrix A. Then

$$\langle \operatorname{Hess} h(x), A\rangle_{H.S.} = \int_0^1 \frac{1}{2\sqrt{1-t}}\mathbb{E}\left[\left\langle A^T\Sigma^{-1/2}Z, \nabla g(Z_{x,t})\right\rangle\right] dt,$$

thus

$$\left|\langle \operatorname{Hess} h(x), A\rangle_{H.S.}\right| \le M_1(g)\mathbb{E}|A^T\Sigma^{-1/2}Z| \int_0^1 \frac{1}{2\sqrt{1-t}}dt = M_1(g)\mathbb{E}|A^T\Sigma^{-1/2}Z|.$$

As above,

$$\mathbb{E}|A^T\Sigma^{-1/2}Z| \le \|A^T\Sigma^{-1/2}\|_{op}\sqrt{\frac{2}{\pi}} \le \sqrt{\frac{2}{\pi}}\|\Sigma^{-1/2}\|_{op}\|A\|_{H.S.}.$$

It follows that

$$\|\mathrm{Hess}\, h(x)\|_{H.S.} \le \sqrt{\frac{2}{\pi}}M_1(g)\|\Sigma^{-1/2}\|_{op}$$

for all $x \in \mathbb{R}^d$, hence part (2).

For part (2), let u and v be fixed vectors in $\mathbb{R}^d$ with $|u| = |v| = 1$. Then it follows from (12) that

$$\begin{aligned}&\langle (\mathrm{Hess}\, h(x) - \mathrm{Hess}\, h(y))\, u, v\rangle \\ &\quad = \int_0^1 \frac{1}{2\sqrt{1-t}}\mathbb{E}\left[\left\langle \Sigma^{-1/2}Z, v\right\rangle \langle \nabla g(Z_{x,t}) - \nabla g(Z_{y,t}), u\rangle\right] dt,\end{aligned}$$

and so

$$\begin{aligned}|\langle (\mathrm{Hess}\, h(x) - \mathrm{Hess}\, h(y))u, v\rangle| &\le |x-y|\, M_2(g)\,\mathbb{E}\left|\left\langle Z, \Sigma^{-1/2}v\right\rangle\right| \int_0^1 \frac{\sqrt{t}}{2\sqrt{1-t}}dt \\ &= |x-y|\, M_2(g)\left|\Sigma^{-1/2}v\right|\frac{\sqrt{2\pi}}{4} \\ &\le |x-y|\, M_2(g)\left\|\Sigma^{-1/2}\right\|_{op}\frac{\sqrt{2\pi}}{4}.\end{aligned}$$

□

Theorem 3. *Let (X, X') be an exchangeable pair of random vectors in $\mathbb{R}^d$. Let $\mathcal{F}$ be a σ-algebra with $\sigma(X) \subseteq \mathcal{F}$, and suppose that there is an invertible matrix Λ, a symmetric, non-negative definite matrix Σ, an $\mathcal{F}$-measureable random vector E and an $\mathcal{F}$-measureable random matrix E' such that*

1.
$$\mathbb{E}\left[X' - X\middle|\mathcal{F}\right] = -\Lambda X + E$$

2.
$$\mathbb{E}\left[(X' - X)(X' - X)^T\middle|\mathcal{F}\right] = 2\Lambda\Sigma + E'.$$

Then for $g \in C^3(\mathbb{R}^d)$,

$$\begin{aligned}&\left|\mathbb{E}g(X) - \mathbb{E}g(\Sigma^{1/2}Z)\right| \\ &\quad\le \|\Lambda^{-1}\|_{op}\left[M_1(g)\mathbb{E}|E| + \frac{1}{4}\widetilde{M}_2(g)\mathbb{E}\|E'\|_{H.S.} + \frac{1}{9}M_3(g)\mathbb{E}|X'-X|^3\right] \\ &\quad\le \|\Lambda^{-1}\|_{op}\left[M_1(g)\mathbb{E}|E| + \frac{\sqrt{d}}{4}M_2(g)\mathbb{E}\|E'\|_{H.S.} + \frac{1}{9}M_3(g)\mathbb{E}|X'-X|^3\right],\end{aligned} \tag{13}$$

where Z is a standard Gaussian random vector in $\mathbb{R}^d$.

If Σ is non-singular, then for $g \in C^2(\mathbb{R}^d)$,

$$\begin{aligned}\left|\mathbb{E}g(X) - \mathbb{E}g(\Sigma^{1/2}Z)\right| &\le M_1(g)\|\Lambda^{-1}\|_{op}\left[\mathbb{E}|E| + \frac{1}{2}\|\Sigma^{-1/2}\|_{op}\mathbb{E}\|E'\|_{H.S.}\right] \\ &\quad + \frac{\sqrt{2\pi}}{24}M_2(g)\|\Sigma^{-1/2}\|_{op}\|\Lambda^{-1}\|_{op}\mathbb{E}|X'-X|^3.\end{aligned} \tag{14}$$

Proof. Fix g, and let $U_o g$ be as in Lemma 1. Note that it suffices to assume that $g \in C^\infty(\mathbb{R}^d)$: let $h : \mathbb{R}^d \to \mathbb{R}$ be a centered Gaussian density with covariance matrix $\epsilon^2 I_d$. Approximate g by $g * h$; clearly $\|g * h - g\|_\infty \to 0$ as $\epsilon \to 0$, and by Young's inequality, $M_k(g * h) \le M_k(g)$ for all $k \ge 1$.

For notational convenience, let $f = U_o g$. By the exchangeability of (X, X'),

$$\begin{aligned}
0 &= \frac{1}{2}\mathbb{E}\left[\left\langle \Lambda^{-1}(X' - X), \nabla f(X') + \nabla f(X)\right\rangle\right] \\
&= \mathbb{E}\left[\frac{1}{2}[\left\langle \Lambda^{-1}(X' - X), \nabla f(X') - \nabla f(X)\right\rangle + \left\langle \Lambda^{-1}(X' - X), \nabla f(X)\right\rangle\right] \\
&= \mathbb{E}\left[\frac{1}{2}\left\langle \operatorname{Hess} f(X), \Lambda^{-1}(X' - X)(X' - X)^T\right\rangle_{H.S.}\right. \\
&\qquad \left. + \left\langle \Lambda^{-1}(X' - X), \nabla f(X)\right\rangle + \frac{R}{2}\right],
\end{aligned}$$

where R is the error in the Taylor approximation. By conditions (1) and (2), it follows that

$$\begin{aligned}
0 = \mathbb{E}&\left[\langle \operatorname{Hess} f(X), \Sigma\rangle_{H.S.} - \langle X, \nabla f(X)\rangle + \frac{1}{2}\left\langle \operatorname{Hess} f(X), \Lambda^{-1}E'\right\rangle_{H.S.}\right. \\
&\qquad \left. + \left\langle \nabla f(X), \Lambda^{-1}E\right\rangle + \frac{R}{2}\right];
\end{aligned}$$

that is (making use of the definition of f),
(15)

$$\mathbb{E}g(X) - \mathbb{E}g(\Sigma^{1/2}Z) = \mathbb{E}\left[\frac{1}{2}\left\langle \operatorname{Hess} f(X), \Lambda^{-1}E'\right\rangle_{H.S.} + \left\langle \nabla f(X), \Lambda^{-1}E\right\rangle + \frac{R}{2}\right].$$

Next,

$$\begin{aligned}
&\mathbb{E}\left|\frac{1}{2}\left\langle \operatorname{Hess} f(X), \Lambda^{-1}E'\right\rangle_{H.S.}\right| \\
&\qquad \le \frac{1}{2}\left(\sup_{x \in \mathbb{R}^d} \|\operatorname{Hess} f(x)\|_{H.S.}\right)\|\Lambda^{-1}E'\|_{H.S.} \\
&\qquad \le \frac{1}{2}\left(\sup_{x \in \mathbb{R}^d} \|\operatorname{Hess} f(x)\|_{H.S.}\right)\|\Lambda^{-1}\|_{op}\|E'\|_{H.S.} \\
&\qquad \le \frac{1}{2}\|\Lambda^{-1}\|_{op}\|E'\|_{H.S.}\left(\min\left\{\frac{1}{2}\widetilde{M}_2(g), \sqrt{\frac{2}{\pi}}M_1(g)\|\Sigma^{-1/2}\|_{op}\right\}\right),
\end{aligned}$$

where the first line is by the Cauchy-Schwarz inequality, the second is by the standard bound $\|AB\|_{H.S.} \le \|A\|_{op}\|B\|_{H.S.}$, and the third uses the bounds (2) and (2) from Lemma 2.

Similarly,

$$\begin{aligned}
\mathbb{E}\left|\left\langle \nabla f(X), \Lambda^{-1}E\right\rangle\right| &\le M_1(f)\|\Lambda^{-1}\|_{op}\mathbb{E}\,|E| \\
&\le \|\Lambda^{-1}\|_{op}\mathbb{E}\,|E|\left(\min\left\{M_1(g), \sqrt{\frac{\pi}{2}}M_o(g)\|\Sigma^{-1/2}\|_{op}\right\}\right).
\end{aligned}$$

Finally, by Taylor's theorem and Lemma 2,

$$
\begin{aligned}
|R| &\leq \frac{M_3(f)}{3}\big|X' - X\big|^2\big|\Lambda^{-1}(X' - X)\big| \\
&\leq \frac{1}{3}\|\Lambda^{-1}\|_{op}\big|X' - X\big|^3\left(\min\left\{\frac{1}{3}M_3(g), \frac{\sqrt{2\pi}}{4}M_2(g)\|\Sigma^{-1/2}\|_{op}\right\}\right).
\end{aligned}
$$

The first bound of the theorem results from choosing the first term from each minimum; the second bound results from the second terms. □

Theorem 4. *Let X be a random vector in $\mathbb{R}^d$ and, for each $\epsilon \in (0,1)$, suppose that (X, X_ϵ) is an exchangeable pair. Let $\mathcal{F}$ be a σ-algebra such that $\sigma(X) \subseteq \mathcal{F}$ and suppose that there is an invertible matrix Λ, a symmetric, non-negative definite matrix Σ, an $\mathcal{F}$-measureable random vector E, an $\mathcal{F}$-measureable random matrix E', and a deterministic function $s(\epsilon)$ such that*

1.
$$\frac{1}{s(\epsilon)}\mathbb{E}\left[X' - X\big|\mathcal{F}\right] \xrightarrow[\epsilon\to 0]{L_1} -\Lambda X + E$$

2.
$$\frac{1}{s(\epsilon)}\mathbb{E}\left[(X' - X)(X' - X)^T\big|\mathcal{F}\right] \xrightarrow[\epsilon\to 0]{L_1(\|\cdot\|_{H.S.})} 2\Lambda\Sigma + E'.$$

3. *For each $\rho > 0$,*
$$\lim_{\epsilon\to 0}\frac{1}{s(\epsilon)}\mathbb{E}\left[|X_\epsilon - X|^2\mathbb{I}(|X_\epsilon - X|^2 > \rho)\right] = 0.$$

Then for $g \in C^2(\mathbb{R}^d)$,

$$
\begin{aligned}
\big|\mathbb{E}g(X) - \mathbb{E}g(\Sigma^{1/2}Z)\big| &\leq \|\Lambda^{-1}\|_{op}\left[M_1(g)\mathbb{E}|E| + \frac{1}{4}\widetilde{M}_2(g)\mathbb{E}\|E'\|_{H.S.}\right] \\
&\leq \|\Lambda^{-1}\|_{op}\left[M_1(g)\mathbb{E}|E| + \frac{\sqrt{d}}{4}M_2(g)\mathbb{E}\|E'\|_{H.S.}\right],
\end{aligned} \tag{16}
$$

where Z is a standard Gaussian random vector in $\mathbb{R}^d$.

Also, if Σ is non-singular,

$$d_W(X, \Sigma^{1/2}Z) \leq \|\Lambda^{-1}\|_{op}\left[\mathbb{E}|E| + \frac{1}{2}\|\Sigma^{-1/2}\|_{op}\mathbb{E}\|E'\|_{H.S.}\right]. \tag{17}$$

Proof. Fix g, and let $U_o g$ be as in Lemma 1. As in the proof of Theorem 3, it suffices to assume that $g \in C^\infty(\mathbb{R}^d)$.

For notational convenience, let $f = U_o g$. Beginning as before,

$$
\begin{aligned}
0 &= \frac{1}{2s(\epsilon)}\mathbb{E}\left[\big\langle\Lambda^{-1}(X_\epsilon - X), \nabla f(X_\epsilon) + \nabla f(X)\big\rangle\right] \\
&= \frac{1}{s(\epsilon)}\mathbb{E}\left[\frac{1}{2}[\big\langle\Lambda^{-1}(X_\epsilon - X), \nabla f(X_\epsilon) - \nabla f(X)\big\rangle + \big\langle\Lambda^{-1}(X_\epsilon - X), \nabla f(X)\big\rangle\right] \\
&= \frac{1}{s(\epsilon)}E\left[\frac{1}{2}\left\langle \operatorname{Hess} f(X), \Lambda^{-1}(X_\epsilon - X)(X_\epsilon - X)^T\right\rangle_{H.S.}\right. \\
&\quad \left. + \left\langle\Lambda^{-1}(X_\epsilon - X), \nabla f(X)\right\rangle + \frac{R}{2}\right],
\end{aligned} \tag{18}
$$

where R is the error in the Taylor approximation.

Now, by Taylor's theorem, there exists a real number K depending on f, such that

$$\begin{aligned}|R| &\le K \min\left\{|X_\epsilon - X|^2|\Lambda^{-1}(X_\epsilon - X)|, |X_\epsilon - X||\Lambda^{-1}(X_\epsilon - X)|\right\}\\ &\le K\|\Lambda^{-1}\|_{op}\min\left\{|X_\epsilon - X|^3, |X_\epsilon - X|^2\right\}\end{aligned}$$

Breaking up the expectation over the sets on which $|X_\epsilon - X|^2$ is larger and smaller than a fixed $\rho > 0$,

$$\begin{aligned}\frac{1}{s(\epsilon)}\mathbb{E}\big|R\big| &\le \frac{K\|\Lambda^{-1}\|_{op}}{s(\epsilon)}\mathbb{E}\Big[|X_\epsilon - X|^3\mathbb{I}(|X_\epsilon - X| \le \rho) + |X_\epsilon - X|^2\mathbb{I}(|X_\epsilon - X| > \rho)\Big]\\ &\le \frac{K\|\Lambda^{-1}\|_{op}\rho\mathbb{E}\big|X_\epsilon - X\big|^2}{s(\epsilon)} + \frac{K\|\Lambda^{-1}\|_{op}}{s(\epsilon)}\mathbb{E}\Big[|X_\epsilon - X|^2\mathbb{I}(|X' - X| > \rho)\Big].\end{aligned}$$

The second term tends to zero as $\epsilon \to 0$ by condition 3; condition 2 implies that the first is bounded by $CK\|\Lambda^{-1}\|_{op}\rho$ for a constant C depending on the distribution of X. It follows that

$$\lim_{\epsilon\to 0}\frac{1}{s(\epsilon)}\mathbb{E}\big|R\big| = 0.$$

For the rest of (18),

$$\begin{aligned}&\lim_{\epsilon\to 0}\frac{1}{s(\epsilon)}\mathbb{E}\left[\frac{1}{2}\left\langle \operatorname{Hess} f(X), \Lambda^{-1}(X_\epsilon - X)(X_\epsilon - X)^T\right\rangle_{H.S.} + \left\langle\Lambda^{-1}(X_\epsilon - X), \nabla f(X)\right\rangle\right]\\ &\quad = \mathbb{E}\Bigg[\langle \operatorname{Hess} f(X), \Sigma\rangle_{H.S.} - \langle X, \nabla f(X)\rangle\\ &\qquad + \frac{1}{2}\left\langle \operatorname{Hess} f(X), \Lambda^{-1}E'\right\rangle_{H.S.} + \left\langle \nabla f(X), \Lambda^{-1}E\right\rangle\Bigg],\end{aligned}$$

where conditions (1) and (2) together with the boundedness of $\operatorname{Hess} f$ and ∇f have been used. That is (making use of the definition of f),

$$\mathbb{E}g(X) - \mathbb{E}g(\Sigma^{1/2}Z) = \mathbb{E}\left[\frac{1}{2}\left\langle \operatorname{Hess} f(X), \Lambda^{-1}E'\right\rangle_{H.S.} + \left\langle \nabla f(X), \Lambda^{-1}E\right\rangle\right]. \tag{19}$$

As in the proof of Theorem 3,

$$\begin{aligned}&\mathbb{E}\left|\frac{1}{2}\left\langle \operatorname{Hess} f(X), \Lambda^{-1}E'\right\rangle_{H.S.}\right|\\ &\qquad \le \frac{1}{2}\|\Lambda^{-1}\|_{op}\|E'\|_{H.S.}\left(\min\left\{\frac{1}{2}\widetilde{M}_2(g), \sqrt{\frac{2}{\pi}}M_1(g)\|\Sigma^{-1/2}\|_{op}\right\}\right),\end{aligned}$$

and

$$\mathbb{E}\left|\left\langle \nabla f(X), \Lambda^{-1}E\right\rangle\right| \le \|\Lambda^{-1}\|_{op}\mathbb{E}\,|E|\,M_1(g).$$

This completes the proof. □

Remarks.

1. Note that the condition

$$(3') \qquad \lim_{\epsilon\to 0}\tfrac{1}{s(\epsilon)}\mathbb{E}\big|X_\epsilon - X\big|^3 = 0,$$

is stronger than condition (3) of Theorem 4 and may be used instead; this is what is done in the application given in Section 3.

2. In [23], singular covariance matrices are treated by comparing to a nearby non-singular covariance matrix rather than directly. However, this is not necessary as all the proofs except those explicitly involving $\Sigma^{-1/2}$ go through for non-negative definite Σ.

3. Examples

3.1. Projections of exchangeable random vectors

Let $X = (X_1, \ldots, X_n)$ be an exchangeable random vector; that is, a random vector whose distribution is invariant under permutation of the components. Assume that $\mathbb{E}X_1 = 0$ and $\mathbb{E}X_1^2 = 1$. Let $\{\theta_i\}_{i=1}^k$ be a collection of unit vectors in $\mathbb{R}^n$, and consider the random vector W with ith component $W_i := \langle \theta_i, X \rangle$ (the dependence of W on the θ_i is suppressed). In this example we will show that under some conditions on the θ, W is approximately distributed as a Gaussian random vector with covariance matrix Σ, whose entries are given by $\sigma_{ij} := \langle \theta_i, \theta_j \rangle$. A particular case of interest is when X is drawn uniformly from a convex body in $\mathbb{R}^n$ invariant under permutation of the coordinates; this includes for example all so-called 1-symmetric bodies (in particular, the ℓ_p balls) as well as the simplex. It was recently shown by Bo'az Klartag in [13] (and quantitatively improved by Klartag in [14]) that in fact most such projections of a random vector drawn from a convex body are close to Gaussian, without any symmetry assumptions. The exchangeability assumed here allows for a fairly straightforward proof of this special case.

Before proceeding with the example, we make one simplifying assumption: write $\theta_i = (\theta_{i1}, \ldots, \theta_{in})$ and assume that for each $i \in \{1, \ldots, k\}$, $\sum_{r=1}^n \theta_{ir} = 0$. This assumption is convenient for technical reasons, but can be removed at the cost of some extra error terms.

First observe that with W defined as above, $\mathbb{E}W = \mathbf{0}$. Also,

$$\mathbb{E}\big[W_i W_j\big] = \sum_{r=1}^n \theta_{ir}\theta_{jr} + \mathbb{E}\big[X_1 X_2\big] \sum_{r \neq s} \theta_{ir}\theta_{js} = \big(1 - \mathbb{E}\big[X_1 X_2\big]\big) \langle \theta_i, \theta_j \rangle ,$$

where the first equality is by the exchangeability of X and the second uses the fact that $\sum_{r=1}^n \theta_{ir} = 0$ for each i. While there is no requirement that $\mathbb{E}[X_1 X_2] = 0$, it will turn out that this quantity must be small for this approach to yield an interesting bound on the distance from W to Gaussian, so the approximation will be by the Gaussian distribution with covariance matrix $\Sigma = [\sigma_{ij}]_{i,j=1}^k$ with $\sigma_{ij} = \langle \theta_i, \theta_j \rangle$.

To use Theorem 3 to show that W as defined above is approximately Gaussian, the first step is to construct an exchangeable pair (W, W'); as is frequently done in applying the method of exchangeable pairs, this is done by first constructing an exchangeable pair at the level of X. Let $\tau = (I\, J)$ be uniformly distributed among the transpositions on S_n, independent of X, and define X' by

$$X' := \big(X_{\tau(1)}, \ldots, X_{\tau(n)}\big) .$$

Then (X, X') is exchangeable, and so if $W' := W(X')$, then (W, W') is exchangeable. Observe that $W_i' = \langle \theta^i, X' \rangle = W_i + \theta_{iI}X_J - \theta_{iI}X_I + \theta_{iJ}X_I - \theta_{iJ}X_J$, that is,

$$W_i' - W_i = (\theta_{iI} - \theta_{iJ})(X_J - X_I). \tag{20}$$

Let $\sum'$ denote summing over distinct indices. Then

$$\mathbb{E}\left[W_i' - W_i \middle| X\right] = \frac{1}{n(n-1)} \sum_{r,s}{}' \left(\theta_{ir} X_s - \theta_{ir} X_r + \theta_{is} X_r - \theta_{is} X_s\right) = -\frac{2}{n-1} W_i,$$

again using that $\sum_r \theta_{ir} = 0$. Condition (1) of Theorem 3 is thus satisfied with $\Lambda = \frac{2}{n-1} I_k$, with I_k denoting the $k \times k$ identity matrix, and $E = \mathbf{0}$. The factor $\|\Lambda^{-1}\|_{op}$ that appears in the statement of the theorem can thus be immediately identified: $\|\Lambda^{-1}\|_{op} = \frac{n-1}{2}$.

Next, the random matrix E' of condition (2) of Theorem 3 must be identified. Letting $\mathbf{1}$ stand for the vector in $\mathbb{R}^n$ with 1 as each entry,

$$\begin{aligned}
&\mathbb{E}\left[(W_i' - W_i)(W_j' - W_j) \middle| X\right] \\
&\quad = \frac{1}{n(n-1)} \sum_{r,s}{}' (\theta_{ir} - \theta_{is})(\theta_{jr} - \theta_{js})(X_s - X_r)^2 \\
&\quad = \frac{4}{n-1} \langle \theta_i, \theta_j \rangle - \frac{2}{n(n-1)} \Bigg[\langle \theta_i, \theta_j \rangle (|X|^2 - n) + n \left(\sum_{r=1}^n \theta_{ir}\theta_{jr} X_r^2 - \langle \theta_i, \theta_j \rangle \right) \\
&\qquad\qquad -2 \langle X, \mathbf{1} \rangle \sum_{r=1}^n \theta_{ir}\theta_{jr} X_r + W_i W_j \Bigg],
\end{aligned}$$

where the second equality follows from straightforward manipulation after expanding the squares in the line above. It thus follows that

$$\begin{aligned}
E_{ij}' = -\frac{2}{n(n-1)} \Bigg[&\langle \theta_i, \theta_j \rangle (|X|^2 - n) + n \left(\sum_{r=1}^n \theta_{ir}\theta_{jr} X_r^2 - \langle \theta_i, \theta_j \rangle \right) \\
&-2 \langle X, \mathbf{1} \rangle \sum_{r=1}^n \theta_{ir}\theta_{jr} X_r + W_i W_j \Bigg].
\end{aligned}$$

To apply Theorem 3, a bound on $\mathbb{E}\|E'\|_{HS}$ is required. Using the $\|\cdot\|_{HS}$-triangle inequality and considering each of the four terms separately, the first term yields

$$\begin{aligned}
\|\Sigma\|_{HS} \mathbb{E} \left| |X|^2 - n \right| &\le \|\Sigma\|_{HS} \sqrt{\mathbb{E} \sum_{i,j} (X_i^2 - 1)(X_j^2 - 1)} \\
&\le \|\Sigma\|_{HS} \left[\sqrt{n \mathbb{E} X_1^4} + \sqrt{n(n-1) \mathbb{E}(X_1^2 - 1)(X_2^2 - 1)} \right].
\end{aligned}$$

The second term yields a similar contribution:

$$\begin{aligned}
n \left\| \left[\sum_{r=1}^n \theta_{ir}\theta_{jr} X_r^2 - \langle \theta_i, \theta_j \rangle \right]_{ij} \right\|_{HS} &\le n \sqrt{\mathbb{E} \sum_{ij} \left(\langle \theta_i, \theta_j \rangle - \sum_r \theta_{ir}\theta_{jr} X_r^2 \right)^2} \\
&= n \sqrt{\mathbb{E}(X_1^2 - 1)(X_2^2 - 1)} \|\Sigma\|_{HS}.
\end{aligned}$$

For the third term,

$$\begin{aligned}
&\mathbb{E}\left\|\langle X,\mathbf{1}\rangle\left[\sum_{r=1}^{n}\theta_{ir}\theta_{jr}X_r\right]_{ij}\right\|_{HS}\\
&\quad\le\sqrt{\mathbb{E}\left\langle X,\mathbf{1}\right\rangle^2\mathbb{E}\left(\sum_{i,j}\sum_{r,s}\theta_{ir}\theta_{jr}\theta_{is}\theta_{js}X_rX_s\right)}\\
&\quad=\sqrt{(n+n(n-1)\mathbb{E}X_1X_2)\left[(1-\mathbb{E}X_1X_2)\left(\sum_{i,j}\sum_{r}\theta_{ir}^2\theta_{jr}^2\right)+\|\Sigma\|_{HS}^2\mathbb{E}X_1X_2\right]}\\
&\quad\le\sqrt{n+n^2|\mathbb{E}X_1X_2|}\left[2\sum_i\|\theta_i\|_4^2+\|\Sigma\|_{HS}\sqrt{|\mathbb{E}X_1X_2|}\right],
\end{aligned}$$

using the Cauchy-Schwarz and Hölder inequalities to get the last bound.

For the last term, define the matrix $C=[c_{ij}]_{i,j=1}^k$ by $c_{ij}:=\sum_{r=1}^n\theta_{ir}^2\theta_{jr}^2$. Some tedious but straightforward manipulation yields

$$\begin{aligned}
\mathbb{E}W_i^2W_j^2&=c_{ij}\left[\mathbb{E}X_1^4-3\mathbb{E}X_1^2X_2^2-4\mathbb{E}X_1^3X_2+12\mathbb{E}X_1^2X_2X_3-6\mathbb{E}X_1X_2X_3X_4\right]\\
&\quad+\sigma_{ij}^2\left[2\mathbb{E}X_1^2X_2^2-4\mathbb{E}X_1^2X_2X_3+2\mathbb{E}X_1X_2X_3X_4\right]\\
&\quad+\left[\mathbb{E}X_1^2X_2^2-2\mathbb{E}X_1^2X_2X_3+\mathbb{E}X_1X_2X_3X_4\right].
\end{aligned}\tag{21}$$

Making use of the Cauchy-Schwarz inequality on the inner sums and the fact that $\|\theta_i\|_4\le\|\theta_i\|_2=1$ for all i,

$$\|C\|_{HS}=\sqrt{\sum_{i,j=1}^{k}\sum_{r,s=1}^{n}\theta_{ir}^2\theta_{jr}^2\theta_{is}^2\theta_{js}^2}\le\sqrt{\sum_{i,j}\|\theta_i\|_4^4\|\theta_j\|_4^4}\le\sum_i\|\theta_i\|_4^4\le k.$$

Recall that $|\sigma_{ij}|=|\langle\theta_i,\theta_j\rangle|\le 1$ as well, thus

$$\left\|\left[\sigma_{ij}^2\right]_{i,j=1}^k\right\|_{HS}=\sqrt{\sum_{i,j}\sigma_{ij}^4}\le k.$$

Trivially, the $k\times k$ matrix with every entry equal to 1 also has Hilbert-Schmidt norm equal to k. Furthermore, each of the expectations of products of the X_i that appears in (21) is bounded in absolute value by $\mathbb{E}X_1^4$, thus

$$\mathbb{E}\left\|WW^T\right\|_{HS}\le 38k\mathbb{E}X_1^4.$$

Collecting terms and making some simplifications, it has been shown that

$$\begin{aligned}
\mathbb{E}\|E'\|_{HS}\le\frac{2}{n(n-1)}\Bigg[\|\Sigma\|_{HS}\Bigg(&2n\sqrt{\mathbb{E}(X_1^2-1)(X_2^2-1)}\\
&+\sqrt{(n+n^2|\mathbb{E}X_1X_2|)|\mathbb{E}X_1X_2|}\Bigg)\\
&+\sqrt{(n+n^2|\mathbb{E}X_1X_2|)}\sum_i\|\theta_i\|_4^2+38k\mathbb{E}X_1^4\Bigg].
\end{aligned}\tag{22}$$

It remains to bound $\mathbb{E}|W'-W|^3$. First averaging over τ, and then making use of Hölder's inequality and repeated use of the L_3-triangle inequality,

$$\begin{aligned}
\mathbb{E}|W'-W|^3 &\le \frac{\sqrt{k}}{n(n-1)}\sum_{r,s=1}^{n}\sum_{i=1}^{k}|\theta_{ir}-\theta_{is}|^3\mathbb{E}|X_s-X_r|^3 \\
&\le \frac{8\sqrt{k}\mathbb{E}|X_1|^3}{n(n-1)}\sum_{r,s=1}^{n}\sum_{i=1}^{k}|\theta_{ir}-\theta_{is}|^3 \\
&\le \frac{64\sqrt{k}\mathbb{E}|X_1|^3}{n-1}\sum_{i=1}^{k}\|\theta_i\|_3^3.
\end{aligned}$$

Collecting all of the estimates and applying Theorem 3 now yields the following.

Theorem 5. *Let X be an exchangeable random vector in $\mathbb{R}^n$ and $\{\theta_i\}_{i=1}^k$ unit vectors in $\mathbb{R}^n$ with $\sum_{r=1}^n \theta_{ir} = 0$ for each i. Define the random vector W with ith component $W_i := \langle \theta_i, X\rangle$, and define the matrix Σ by $\sigma_{ij} = \langle \theta_i, \theta_j\rangle$. Then if $g \in C^3(\mathbb{R}^k)$ and Z is a standard Gaussian random vector in $\mathbb{R}^k$,*

(23)
$$\begin{aligned}
&\left|\mathbb{E}g(W) - \mathbb{E}g(\Sigma^{1/2}Z)\right| \\
&\le \frac{\sqrt{k}M_2(g)}{4}\Bigg[\|\Sigma\|_{HS}\left(2\sqrt{\mathbb{E}(X_1^2-1)(X_2^2-1)} + \sqrt{\left(\frac{1}{n}+|\mathbb{E}X_1X_2|\right)|\mathbb{E}X_1X_2|}\right) \\
&\qquad + \sqrt{\left(\frac{1}{n}+|\mathbb{E}X_1X_2|\right)\sum_i\|\theta_i\|_4^2 + \frac{38k}{n}\mathbb{E}X_1^4}\Bigg] \\
&\quad + 4\sqrt{k}M_3(g)\mathbb{E}|X_1|^3\sum_{i=1}^{k}\|\theta_i\|_3^3.
\end{aligned}$$

Remarks.

1. The appearance of the term $\sum_{i=1}^k \|\theta_i\|_3^3$ (or something similar) is expected; see, e.g., [19],[20],[15]. The typical behavior (in a measure theoretic sense) for $\theta \in \mathbb{S}^{n-1}$ is that $\|\theta\|_3^3$ is of order $\frac{1}{\sqrt{n}}$. However, if θ is a standard basis vector then $\|\theta\|_3^3 = 1$, thus the error bound is not small in this case, nor should it be as this corresponds to simply truncating the random vector X.
2. The term $\sum_{i=1}^k \|\theta_i\|_4^2$ could simply be estimated by k; however, for typical choices of the θ_i, the bound will be of order $\frac{k}{\sqrt{n}}$.
3. If the θ_i are orthonormal, then $\Sigma = I_k$ and so the second half of Theorem 3 can be applied. In this case, the bound (23) may be replaced with

   (24)
   $$\begin{aligned}
   &\left|\mathbb{E}g(W) - \mathbb{E}g(Z)\right| \\
   &\le \frac{1}{2}M_1(g)\Bigg[\sqrt{k}\left(2\sqrt{\mathbb{E}(X_1^2-1)(X_2^2-1)} + \sqrt{\left(\frac{1}{n}+|\mathbb{E}X_1X_2|\right)|\mathbb{E}X_1X_2|}\right) \\
   &\qquad + \sqrt{\left(\frac{1}{n}+|\mathbb{E}X_1X_2|\right)\sum_i\|\theta_i\|_4^2 + \frac{38k}{n}\mathbb{E}X_1^4}\Bigg] \\
   &\quad + \frac{4\sqrt{2\pi k}M_2(g)\mathbb{E}|X_1|^3}{3}\sum_{i=1}^{k}\|\theta_i\|_3^3.
   \end{aligned}$$

4. In the case that X is drawn uniformly from a suitably translated and rescaled simplex, expressions for the moments appearing in the bound (23) are available (see [19]). Making use of these expressions shows that there are absolute constants c, c' such that

$$\begin{aligned} \left|\mathbb{E}g(W) - \mathbb{E}g(\Sigma^{1/2}Z)\right| &\le c\sqrt{\frac{k}{n}}M_2(g)\left[\|\Sigma\|_{HS} + \sum_{i=1}^{k}\|\theta_i\|_4^2\right] \\ &\quad + c'\sqrt{k}M_3(g)\sum_{i=1}^{k}\|\theta_i\|_3^3. \end{aligned} \tag{25}$$

3.2. Runs on the line

The following example was treated by Reinert and Röllin [23] as an example of the embedding method. It should be emphasized that showing that the number of d-runs on the line is asymptotically Gaussian seems infeasible with Stein's original method of exchangeable pairs because of the failure of condition (1) from the introduction, but in [23], the random variable of interest is embedded in a random vector whose components can be shown to be jointly Gaussian by making use of the more general condition (5) of the introduction. The example is reworked here making use of the analysis of [23] together with Theorem 3, yielding an improved rate of convergence.

Let $X_1, \ldots, X_n$ be independent $\{0,1\}$-valued random variables, with $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = 0) = 1 - p$. For $d \ge 1$, define the (centered) number of d-runs as

$$V_d := \sum_{m=1}^{n}(X_m X_{m+1}\cdots X_{m+d-1} - p^d),$$

assuming the torus convention, namely that $X_{n+k} = X_k$ for any k. For this example, we assume that $d < \frac{n}{2}$. To make an exchangeable pair, $d-1$ sequential elements of $X := (X_1, \ldots, X_n)$ are resampled. That is, let I be a uniformly distributed element of $\{1, \ldots, n\}$ and let $X_1', \ldots, X_n'$ be independent copies of the X_i. Let X' be constructed from X by replacing $X_I, \ldots, X_{I+d-2}$ with $X_I', \ldots, X_{I+d-2}'$. Then (X, X') is an exhangeable pair, and, defining $V_i' := V_i(X)$ for $i \ge 1$, it is easy to see that

$$\begin{aligned} V_i' - V_i = &-\sum_{m=I-i+1}^{I+d-2} X_m\cdots X_{m+i-1} + \sum_{m=I+d-i}^{I+d-2} X_m'\cdots X_{I+d-2}'X_{I+d-1}\cdots X_{m+i-1} \\ &+ \sum_{m=I}^{I+d-i-1} X_m'\cdots X_{m+i-1}' + \sum_{m=I-i+1}^{I-1} X_m\cdots X_{I-1}X_I'\cdots X_{m+i-1}', \end{aligned} \tag{26}$$

where sums $\sum_a^b$ are taken to be zero if $a > b$. It follows that

$$\mathbb{E}\left[V_i' - V_i\middle|X\right] = -\frac{1}{n}\left[(d+i-2)V_i - 2\sum_{k=1}^{i-1}p^{i-k}V_k\right].$$

Standard calculations show that, for $1 \le j \le i \le d$,

$$
\begin{aligned}
\mathbb{E}\big[V_i V_j\big] &= n\left[(i-j+1)p^i + 2\sum_{k=1}^{j-1} p^{i+j-k} - (i+j-1)p^{i+j}\right] \\
&= np^i(1-p)\sum_{k=0}^{j-1}(i-j+1+2k)p^k.
\end{aligned}
\tag{27}
$$

In particular, it follows from this expression that $np^i(1-p) \le \mathbb{E}V_i^2 \le np^i(1-p)i^2$, suggesting the renormalized random variables

$$
W_i := \frac{V_i}{\sqrt{np^i(1-p)}}. \tag{28}
$$

It then follows from (27) that, for $1 \le i, j \le d$,

$$
\sigma_{ij} := \mathbb{E}\big[W_i W_j\big] = p^{\frac{|i-j|}{2}} \sum_{k=0}^{i\wedge j-1} (|i-j|+1+2k)p^k, \tag{29}
$$

and from (26) that if $W := (W_1, \ldots, W_d)$, then $\mathbb{E}\left[W' - W \middle| X\right] = \Lambda W$, where

$$
\Lambda = \frac{1}{n}\begin{bmatrix}
d-1 & & & & \\
-2p^{\frac{1}{2}} & d & & 0 & \\
\vdots & & \ddots & & \\
-2p^{\frac{k-1}{2}} & \cdots & -2p^{\frac{1}{2}} & d+k-2 & \\
\vdots & & & \ddots & \\
-2p^{\frac{d-1}{2}} & & \cdots & -2p^{\frac{1}{2}} & 2(d-1)
\end{bmatrix}.
$$

Condition (1) of Theorem 3 thus applies with $E = 0$ and Λ as above.

To apply Theorem 3, an estimate on $\|\Lambda^{-1}\|_{op}$ is needed. Following Reinert and Röllin, we make use of known estimates of condition numbers for triangular matrices (see, e.g., the survey of Higham [10]). First, write $\Lambda =: \Lambda_E \Lambda_D$, where Λ_D is diagonal with the same diagonal entries as Λ and Λ_E is lower triangular with diagonal entries equal to one and $(\Lambda_E)_{ij} = \frac{\Lambda_{ij}}{\Lambda_{jj}}$ for $i > j$. Note that all non-diagonal entries of Λ_E are bounded in absolute value by $\frac{2\sqrt{p}}{d-1}$. From Lemeire [16], this implies the bounds

$$
\|\Lambda_E^{-1}\|_1 \le \left(1 + \frac{2\sqrt{p}}{d-1}\right)^{d-1} \qquad \text{and} \qquad \|\Lambda_E^{-1}\|_\infty \le \left(1 + \frac{2\sqrt{p}}{d-1}\right)^{d-1}.
$$

From Higham, $\|\Lambda_E^{-1}\|_{op} \le \sqrt{\|\Lambda_E^{-1}\|_1 \|\Lambda_E^{-1}\|_\infty}$, thus

$$
\|\Lambda_E^{-1}\|_{op} \le \left(1 + \frac{2\sqrt{p}}{d-1}\right)^{d-1}.
$$

Trivially, $\|\Lambda_D^{-1}\|_{op} = \frac{n}{d-1}$, and thus

$$
\|\Lambda^{-1}\|_{op} \le \frac{n}{d-1}\left(1 + \frac{2\sqrt{p}}{d-1}\right)^{d-1} \le \frac{ne^{2\sqrt{p}}}{d-1} \le \frac{15n}{d}. \tag{30}
$$

Now observe that, if condition (1) of Theorem 3 is satisfied with $E = 0$, then it follows that $\mathbb{E}[(W' - W)(W' - W)^T] = 2\Lambda\Sigma$, and thus we may take

$$E' := \mathbb{E}\left[(W' - W)(W' - W)^T - 2\Lambda\Sigma \middle| W\right].$$

It follows that

$$\mathbb{E}\|E'\|_{H.S.} \le \sqrt{\sum_{i,j} \mathbb{E}(E'_{ij})^2} = \sqrt{\sum_{i,j} \operatorname{Var}\left(\mathbb{E}\left[(W'_i - W_i)(W'_j - W_j)\middle|W\right]\right)}.$$

It was determined by Reinert and Röllin that

$$\operatorname{Var}\left(\mathbb{E}\left[(W'_i - W_i)(W'_j - W_j)\middle|W\right]\right) \le \frac{96d^5}{n^3 p^{2d}(1-p)^2},$$

thus

$$\mathbb{E}\|E'\|_{H.S.} \le \frac{4\sqrt{6}d^{7/2}}{n^{3/2}p^d(1-p)}.$$

Finally, note that

$$\mathbb{E}|W' - W|^3 \le \sqrt{d}\sum_{i=1}^d \mathbb{E}\left|W'_i - W_i\right|^3.$$

Reinert and Röllin showed that

$$\mathbb{E}\left|(W'_i - W_i)(W'_j - W_j)(W'_k - W_k)\right| \le \frac{8d^3}{n^{3/2}p^{3d/2}(1-p)^{3/2}}$$

for all i, j, k, thus

$$\mathbb{E}|W' - W|^3 \le \frac{8d^{9/2}}{n^{3/2}p^{3d/2}(1-p)^{3/2}}.$$

Using these bounds in inequality (13) from Theorem 3 yields the following.

Theorem 6. *For $W = (W_1, \ldots, W_d)$ defined as in (28) with $d < \frac{n}{2}$, $\Sigma = [\sigma_{ij}]_{i,j=1}^d$ given by (29), and $h \in C^3(\mathbb{R}^d)$,*

$$\left|\mathbb{E}h(W) - \mathbb{E}h(\Sigma^{1/2}Z)\right| \le \left[\frac{15\sqrt{6}d^3 M_2(h)}{p^d(1-p)\sqrt{n}} + \frac{40d^{7/2}M_3(h)}{3p^{3d/2}(1-p)^{3/2}\sqrt{n}}\right], \tag{31}$$

where Z is a standard d-dimensional Gaussian random vector.

Remarks. Compare this result to that obtained in [23]:

$$\left|\mathbb{E}h(W) - \mathbb{E}h(\Sigma^{1/2}Z)\right| \le \frac{37d^{7/2}|h|_2}{p^d(1-p)\sqrt{n}} + \frac{10d^5|h|_3}{p^{3d/2}(1-p)^{3/2}\sqrt{n}}, \tag{32}$$

where $|h|_2 = \sup_{i,j}\|\frac{\partial^2 h}{\partial x_i \partial x_j}\|_\infty$ and $|h|_3 = \sup_{i,j,k}\|\frac{\partial^3 h}{\partial x_i \partial x_j \partial x_k}\|_\infty$.

3.3. Eigenfunctions of the Laplacian

Consider a compact Riemannian manifold M with metric g. Integration with respect to the normalized volume measure is denoted $d\overline{\mathrm{vol}}$, thus $\int_M 1 d\overline{\mathrm{vol}} = 1$. For coordinates $\{\frac{\partial}{\partial x_i}\}_{i=1}^n$ on M, define

$$(G(x))_{ij} = g_{ij}(x) = \left\langle \frac{\partial}{\partial x_i}\bigg|_x, \frac{\partial}{\partial x_j}\bigg|_x \right\rangle, \qquad g(x) = \det(G(x)),$$
$$g^{ij}(x) = (G^{-1}(x))_{ij}.$$

Define the gradient ∇f of $f : M \to \mathbb{R}$ and the Laplacian $\Delta_g f$ of f by

$$\nabla f(x) = \sum_{j,k} \frac{\partial f}{\partial x_j} g^{jk} \frac{\partial}{\partial x_k}, \qquad \Delta_g f(x) = \frac{1}{\sqrt{g}} \sum_{j,k} \frac{\partial}{\partial x_j}\left(\sqrt{g} g^{jk} \frac{\partial f}{\partial x_k}\right).$$

The function $f : M \to \mathbb{R}$ is an eigenfunction of Δ with eigenvalue $-\mu$ if $\Delta f(x) = -\mu f(x)$ for all $x \in M$; it is known (see, e.g., [4]) that on a compact Riemannian manifold M, the eigenvalues of Δ form a sequence $0 \geq -\mu_1 \geq -\mu_2 \geq \cdots \searrow -\infty$. Eigenspaces associated to different eigenvalues are orthogonal in $L_2(M)$ and all eigenfunctions of Δ are elements of $C^\infty(M)$.

Let X be a uniformly distributed random point of M. The value distribution of a function f on M is the distribution (on $\mathbb{R}$) of the random variable $f(X)$. In [17], a general bound was given for the total variation distance between the value distribution of an eigenfunction and a Gaussian distribution, in terms of the eigenvalue and the gradient of f. The proof made use of a univariate version of Theorem 4. Essentially the same analysis is used here to prove a multivariate version of that theorem.

Let $f_1, \ldots, f_k$ be a sequence of orthonormal (in L_2) eigenfunctions of Δ with corresponding eigenvalues $-\mu_i$ (some of the μ_i may be the same if the eigenspaces of M have dimension greater than 1). Define the random vector $W \in \mathbb{R}^k$ by $W_i := f_i(X)$. We will apply Theorem 4 to show that W is approximately distributed as a standard Gaussian random vector (i.e., $\Sigma = I_k$).

For $\epsilon > 0$, an exchangeable pair (W, W_ϵ) is constructed from W as follows. Given X, choose an element $V \in S_X M$ (the unit sphere of the tangent space to M at X) according to the uniform measure on $S_X M$, and let $X_\epsilon = \exp_X(\epsilon V)$. That is, pick a direction at random, and move a distance ϵ from X along a geodesic in that direction. It was shown in [17] that this construction produces an exchangeable pair of random points of M; it follows that if $W_\epsilon := (f_1(X_\epsilon), \ldots, f_k(X_\epsilon))$, then (W, W_ϵ) is an exchangeable pair of random vectors in $\mathbb{R}^k$.

In order to identify Λ, E and E' so as to apply Theorem 4, first let $\gamma : [0, \epsilon] \to M$ be a constant-speed geodesic such that $\gamma(0) = X$, $\gamma(\epsilon) = X_\epsilon$, and $\gamma'(0) = V$. Then applying Taylor's theorem on $\mathbb{R}$ to the function $f_i \circ \gamma$ yields

$$\begin{aligned} f_i(X_\epsilon) - f_i(X) &= \epsilon \cdot \frac{d(f_i \circ \gamma)}{dt}\bigg|_{t=0} + \frac{\epsilon^2}{2} \cdot \frac{d^2(f_i \circ \gamma)}{dt^2}\bigg|_{t=0} + O(\epsilon^3) \\ &= \epsilon \cdot d_X f_i(V) + \frac{\epsilon^2}{2} \cdot \frac{d^2(f_i \circ \gamma)}{dt^2}\bigg|_{t=0} + O(\epsilon^3), \end{aligned} \tag{33}$$

where the coefficient implicit in the $O(\epsilon^3)$ depends on f_i and γ and $d_x f_i$ denotes the differential of f_i at x. Recall that $d_x f_i(v) = \langle \nabla f_i(x), v\rangle$ for $v \in T_x M$ and the gradient $\nabla f_i(x)$ defined as above. Now, for X fixed, V is distributed according to normalized Lebesgue measure on $S_X M$ and $d_X f_i$ is a linear functional on $T_X M$. It follows that

$$\mathbb{E}\left[d_X f_i(V)\middle|X\right] = \mathbb{E}\left[d_X f_i(-V)\middle|X\right] = -\mathbb{E}\left[d_X f_i(V)\middle|X\right],$$

thus $\mathbb{E}[d_X f_i(V)|X] = 0$. This implies that

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon^2}\mathbb{E}\big[f_i(X_\epsilon) - f_i(X)\big|X\big]$$

exists and is finite; we will take $s(\epsilon) = \epsilon^2$. Indeed, it is well-known (see, e.g., Theorem 11.12 of [9]) that

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon^2}\mathbb{E}\big[f_i(X_\epsilon) - f_i(X)\big|X\big] = \frac{1}{2n}\Delta_g f_i(X) = \frac{-\mu_i}{2n} f_i(X) \tag{34}$$

for $n = \dim(M)$. It follows that $\Lambda = \frac{1}{2n}\operatorname{diag}(\mu_1, \ldots, \mu_k)$ and $E' = 0$. The expression $\mathbb{E}[W_\epsilon - W|W]$ satisfies the L_1 convergence requirement of Theorem 4, since the f_i are necessarily smooth and M is compact. Furthermore, it is immediate that $\|\Lambda^{-1}\|_{op} = 2n \max_{1\le i\le k}(\frac{1}{\mu_i})$.

For the second condition of Theorem 4, it is necessary to determine

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon^2}\mathbb{E}\big[(W_\epsilon - W)_i(W_\epsilon - W)_j\big|X\big] = \lim_{\epsilon\to 0} \frac{1}{\epsilon^2}\mathbb{E}\big[(f_i(X_\epsilon) - f_i(X))(f_j(X_\epsilon) - f_j(X))\big|X\big].$$

By the expansion (33),

$$\mathbb{E}\left[(f_i(X_\epsilon) - f_i(X))(f_j(X_\epsilon) - f_j(X))\middle|X\right] = \epsilon^2\mathbb{E}\left[(d_X f_i(V))(d_X f_j(V))\middle|X\right] + O(\epsilon^3).$$

Choose coordinates $\{\frac{\partial}{\partial x_i}\}_{i=1}^n$ in a neighborhood of X which are orthonormal at X. Then

$$\nabla f(X) = \sum_i \frac{\partial f}{\partial x_i}\frac{\partial}{\partial x_i},$$

for any function $f \in C^1(M)$, thus

$$\begin{aligned}(d_x f_i(v)) \cdot (d_x f_j(v)) &= \langle \nabla f_i, v\rangle \langle \nabla f_j, v\rangle \\ &= \sum_{r=1}^n \frac{\partial f_i}{\partial x_r}(x)\frac{\partial f_j}{\partial x_r}(x) v_r^2 + \sum_{r\neq s} \frac{\partial f_i}{\partial x_r}(x)\frac{\partial f_j}{\partial x_s}(x) v_r v_s.\end{aligned}$$

Since V is uniformly distributed on a Euclidean sphere, $\mathbb{E}[V_r V_s] = \frac{1}{n}\delta_{rs}$. Making use of this fact yields

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon^2}\mathbb{E}\left[(d_X f_i(V))(d_X f_j(V))\middle|X\right] = \frac{1}{n}\langle \nabla f_i(X), \nabla f_j(X)\rangle,$$

thus condition (2) is satisfied with

$$E' = \frac{1}{n}\Big[\langle \nabla f_i(X), \nabla f_j(X)\rangle\Big]_{i,j=1}^k - 2\Lambda.$$

(As before, the convergence requirement is satisfied since the f_i are smooth and M is compact.)

By Stokes' theorem,

$$\mathbb{E}\langle \nabla f_i(X), \nabla f_j(X)\rangle = -\mathbb{E}\big[f_i(X)\Delta_g f_j(X)\big] = \mu_j\mathbb{E}\big[f_i(X)f_j(X)\big] = \mu_i\delta_{ij},$$

thus

$$\mathbb{E}\|E'\|_{H.S.} = \frac{1}{n}\mathbb{E}\sqrt{\sum_{i,j=1}^{k}\Big[\langle\nabla f_i(X),\nabla f_j(X)\rangle - \mathbb{E}\langle\nabla f_i(X),\nabla f_j(X)\rangle\Big]}$$

Finally, (33) gives immediately that

$$\mathbb{E}\left[|W_\epsilon - W|^3\middle|W\right] = O(\epsilon^3),$$

(where the implicit constants depend on the f_i and on k), thus condition (3) of Theorem 4 is satisfied.

All together, we have proved the following.

Theorem 7. *Let M be a compact Riemannian manifold and $f_1,\dots,f_k$ an orthonormal (in $L_2(M)$) sequence of eigenfunctions of the Laplacian on M, with corresponding eigenvalues $-\mu_i$. Let X be a uniformly distributed random point of M. Then if $W := (f_1(X),\dots,f_k(X))$,*

$$d_W(W,Z) \le \left[\max_{1\le i\le k}\left(\frac{1}{\mu_i}\right)\right]\mathbb{E}\sqrt{\sum_{i,j=1}^{k}\Big[\langle\nabla f_i(X),\nabla f_j(X)\rangle - \mathbb{E}\langle\nabla f_i(X),\nabla f_j(X)\rangle\Big]}.$$

Example: The torus.

In this example, Theorem 7 is applied to the value distributions of eigenfunctions on flat tori. The class of functions considered here are random functions; that is, they are linear combinations of eigenfunctions with random coefficients.

Let (M,g) be the torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$, with the metric given by the symmetric positive-definite bilinear form B:

$$(x,y)_B = \langle Bx, y\rangle\,.$$

With this metric, the Laplacian Δ_B on $\mathbb{T}^n$ is given by

$$\Delta_B f(x) = \sum_{j,k}(B^{-1})_{jk}\frac{\partial^2 f}{\partial x_j\partial x_k}(x).$$

Eigenfunctions of Δ_B are given by the real and imaginary parts of functions of the form

$$f_v(x) = e^{2\pi i\langle v,x\rangle_B} = e^{2\pi i\langle Bv,x\rangle},$$

for vectors $v\in\mathbb{R}^n$ such that Bv has integer components, with corresponding eigenvalue $-\mu_v = -(2\pi\|v\|_B)^2$.

Consider a collection of k random eigenfunctions $\{f_j\}_{j=1}^k$ of Δ_B on the torus which are linear combinations of eigenfunctions with random coefficients:

$$f_j(x) := \Re\left(\sum_{v\in\mathcal{V}_j} a_v e^{2\pi i\langle Bv,x\rangle}\right),$$

where $\mathcal{V}_j$ is a finite collection of vectors v such that Bv has integer components and $\langle v, Bv\rangle = \frac{\mu_j}{(2\pi)^2}$ for each $v\in\mathcal{V}_j$, and $\{\{a_v\}_{v\in\mathcal{V}_j} : 1\le j\le k\}$ are k independent random vectors (indexed by j) on the spheres of radius $\sqrt{2}$ in $\mathbb{R}^{|\mathcal{V}_j|}$. Assume that

$v + w \neq 0$ for $v \in \mathcal{V}_r$ and $w \in \mathcal{V}_s$ (r and s may be equal) and that $\mathcal{V}_r \cap \mathcal{V}_s = \emptyset$ for $r \neq s$; it follows easily that the f_j are orthonormal in $L_2(\mathbb{T}^n)$.

To apply Theorem 7, first note that

$$\begin{aligned}\nabla_B f_r(x) &= \left\{ \Re \left(\sum_{j=1}^{n} \sum_{v \in \mathcal{V}_r} (2\pi i) a_v (Bv)_j (B^{-1})_{j\ell} e^{2\pi i \langle Bv, x\rangle} \right) \right\}_{\ell=1}^{n} \\ &= -\Im \left(\sum_{v \in \mathcal{V}_r} (2\pi) a_v e^{2\pi i \langle Bv, x\rangle} v \right),\end{aligned}$$

using the fact that B is symmetric.

It follows that

$$\begin{aligned}&\langle \nabla_B f_r(x), \nabla_B f_s(x) \rangle_B \\ &\quad = \sum_{j,\ell=1}^{n} B_{j\ell} \Im \left(\sum_{v \in \mathcal{V}_r} (2\pi) a_v e^{2\pi i \langle Bv, x\rangle} v_j \right) \Im \left(\sum_{w \in \mathcal{V}_s} (2\pi) a_w e^{2\pi i \langle Bw, x\rangle} w_\ell \right) \\ &\quad = \frac{1}{2} \Re \left[\sum_{\substack{v \in \mathcal{V}_r \\ w \in \mathcal{V}_s}} 4\pi^2 a_v a_w \langle v, w \rangle_B \left(e^{2\pi i \langle Bv - Bw, x\rangle} - e^{2\pi i \langle Bv + Bw, x\rangle} \right) \right].\end{aligned} \tag{35}$$

Let X be a randomly distributed point on the torus. Let $\mathbb{E}_a$ denote averaging over the coefficients a_v and $\mathbb{E}_X$ denote averaging over the random point X. To estimate $\mathbb{E}_a d_W(W, Z)$ from Theorem 7, first apply the Cauchy-Schwartz inequality and then change the order of integration:

$$\begin{aligned}&\mathbb{E}_a \mathbb{E}_X \sqrt{\sum_{i,j=1}^{k} \left[\langle \nabla f_i(X), \nabla f_j(X) \rangle_B - \mathbb{E}_X \langle \nabla f_i(X), \nabla f_j(X) \rangle_B \right]} \\ &\qquad \leq \sqrt{\sum_{i,j=1}^{k} \mathbb{E}_X \mathbb{E}_a \left[\langle \nabla f_i(X), \nabla f_j(X) \rangle_B - \mathbb{E}_X \langle \nabla f_i(X), \nabla f_j(X) \rangle_B \right]^2}.\end{aligned}$$

Start by computing $\mathbb{E}_X \mathbb{E}_a \langle \nabla_B f_r(X), \nabla_B f_s(X) \rangle_B^2$. From above,

$$\begin{aligned}&\langle \nabla_B f_r(x), \nabla_B f_s(X) \rangle_B^2 \\ &\quad = 2\pi^4 \Re \Bigg[\sum_{\substack{v, v' \in \mathcal{V}_r \\ w, w' \in \mathcal{V}_s}} a_v a_w a_{v'} a_{w'} \langle v, w \rangle_B \langle v', w' \rangle_B \\ &\qquad\qquad \Big[e^{2\pi i \left\langle Bv - Bw - Bv' + Bw', x \right\rangle} - e^{2\pi i \left\langle Bv - Bw - Bv' - Bw', x \right\rangle} \\ &\qquad\qquad + e^{2\pi i \left\langle Bv - Bw + Bv' - Bw', x \right\rangle} - e^{2\pi i \left\langle Bv - Bw + Bv' + Bw', x \right\rangle} \\ &\qquad\qquad - e^{2\pi i \left\langle Bv + Bw - Bv' + Bw', x \right\rangle} + e^{2\pi i \left\langle Bv + Bw - Bv' - Bw', x \right\rangle} \\ &\qquad\qquad - e^{2\pi i \left\langle Bv + Bw + Bv' - Bw', x \right\rangle} + e^{2\pi i \left\langle Bv + Bw + Bv' + Bw', x \right\rangle} \Big] \Bigg].\end{aligned}$$

Averaging over the coefficients $\{a_v\}$ using standard techniques (see Folland [7] for general formulae and [17] for a detailed explanation of the univariate version of this result), and then over the random point $X \in \mathbb{T}^n$, it is not hard to show that

$$\mathbb{E}_X\mathbb{E}_a\|\nabla_B f_r(X)\|_B^4$$
$$= \frac{8\pi^4}{|\mathcal{V}_r|(|\mathcal{V}_r|+2)}\left[3\sum_{v\in\mathcal{V}_r}\|v\|_B^4 + 2\left(\sum_{v\in\mathcal{V}_r}\|v\|_B^2\right)^2 + 4\sum_{v,w\in\mathcal{V}_r}\langle v,w\rangle_B^2\right],$$

and

$$\mathbb{E}_X\mathbb{E}_a\left\langle\nabla_B f_r(X),\nabla_B f_s(X)\right\rangle_B^2 = \frac{4\pi^4}{|\mathcal{V}_r||\mathcal{V}_s|}\sum_{\substack{v\in\mathcal{V}_r\\ w\in\mathcal{V}_s}}\langle v,w\rangle_B^2.$$

Now,

$$\mathbb{E}_a\Big[\mathbb{E}_X\|\nabla_B f_r(X)\|_B^2\Big]^2 = \mathbb{E}_a\left[2\pi^2\sum_{v\in\mathcal{V}_r}a_v^2\|v\|_B^2\right]^2$$
$$= \frac{(2\pi)^4}{|\mathcal{V}_r|(|\mathcal{V}_r|+2)}\left[\left(\sum_{v\in\mathcal{V}_r}\|v\|_B^2\right)^2 + 2\sum_{v\in\mathcal{V}_r}\|v\|_B^4\right],$$

and

$$\mathbb{E}_X\left\langle\nabla_B f_r(X),\nabla_B f_s(X)\right\rangle_B = 0$$

for $r \neq s$. It follows that

$$\mathbb{E}_X\mathbb{E}_a\|\nabla_B f_r(X)\|_B^4 - \mathbb{E}_a\left(\mathbb{E}_X\|\nabla_B f_r(X)\|_B^2\right)^2 \le \frac{2(2\pi)^4}{|\mathcal{V}_r|(|\mathcal{V}_r|+2)}\sum_{v,w\in\mathcal{V}_r}\langle v,w\rangle_B^2,$$

and, applying Theorem 7, we have shown that

Theorem 8. *Let the random orthonormal set of functions $\{f_r\}_{r=1}^k$ be defined on $\mathbb{T}^n$ as above, and let the random vector W be defined by $W_i := f_i(X)$ for X a random point of $\mathbb{T}^n$. Then*

$$\mathbb{E}_a d_W(W,Z) \le \frac{4\pi^2}{\min_r \mu_r}\sqrt{\sum_{r,s=1}^k\left(\frac{2}{|\mathcal{V}_r||\mathcal{V}_s|}\sum_{\substack{v\in\mathcal{V}_r\\ w\in\mathcal{V}_s}}\langle v,w\rangle_B^2\right)}.$$

Remarks. Note that if the elements of $\cup_{r=1}^k\mathcal{V}_r$ are mutually orthogonal, then the right-hand side becomes

$$\frac{4\pi^4}{\min_r\mu_r}\sqrt{\sum_{r=1}^k\frac{2\mu_r}{|\mathcal{V}_r|^2}},$$

thus if it is possible to choose the $\mathcal{V}_r$ such that their sizes are large for large n, and the range of the μ_r is not too big, the error is small. One can thus find vectors of orthonormal eigenfunctions of $\mathbb{T}^n$ which are jointly Gaussian (and independent) in the limit as the dimension tends to infinity, if the matrix B is such that there are large collections of vectors v which are "close to orthogonal" and have the same lengths with respect to $\langle\cdot,\cdot\rangle_B$ and with the vectors Bv having integer components. It is possible to extend the analysis here, in a fairly straightfoward manner, to require rather less of the matrix B (essentially all the conditions here can be allowed to hold only approximately), but for simplicity's sake, we include only this most basic version here. The univariate version of this relaxing of conditions is carried out in detail in [17].

Acknowledgements

The author thanks M. Meckes for many useful discussions.

References

[1] Barbour, A. D., Karoński, M. and Ruciński, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47**(2) 125–145.
[2] Bolthausen, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66**(3) 379–386.
[3] Chatterjee, S. and Meckes, E. (2008). Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 257–283.
[4] Chavel, I. (1984). *Eigenvalues in Riemannian Geometry. Pure and Applied Mathematics* **115**. Academic Press Inc., Orlando, FL. Including a chapter by Burton Randol, With an appendix by Jozef Dodziuk.
[5] Dudley, R. M. (1989). *Real Analysis and Probability. The Wadsworth & Brooks/Cole Mathematics Series*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.
[6] Federer, H. (1969). *Geometric Measure Theory. Die Grundlehren der mathematischen Wissenschaften* **153**. Springer-Verlag, New York.
[7] Folland, G. B. (2001). How to integrate a polynomial over a sphere. *Amer. Math. Monthly* **108**(5) 446–448.
[8] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**(2) 724–739.
[9] Gray, A. (2004). *Tubes. Progress in Mathematics* **221**. Birkhäuser Verlag, Basel, second edition. With a preface by Vicente Miquel.
[10] Higham, N. J. (1987). A survey of condition number estimation for triangular matrices. *SIAM Rev.* **29**(4) 575–596.
[11] Janson, S. (1990). A functional limit theorem for random graphs with applications to subgraph count statistics. *Random Structures Algorithms* **1**(1) 15–37.
[12] Janson, S. and Nowicki, K. (1991). The asymptotic distributions of generalized U-statistics with applications to random graphs. *Probab. Theory Related Fields* **90**(3) 341–375.
[13] Klartag, B. (2007). A central limit theorem for convex sets. *Invent. Math.* **168**(1) 91–131.
[14] Klartag, B. (2007). Power-law estimates for the central limit theorem for convex sets. *J. Funct. Anal.* **245**(1) 284–310.
[15] Klartag, B. (2009). A berry-esseen type inequality for convex bodies with an unconditional basis. *Probab. Theory Related Fields* **45**(1) 1–33.
[16] Lemeire, F. (1975). Bounds for condition numbers of triangular and trapezoid matrices. *Nordisk Tidskr. Informationsbehandling (BIT)* **15**(1) 58–64.
[17] Meckes, E. (2009). On the approximate normality of eigenfunctions of the Laplacian. *Trans. Amer. Math. Soc.* **361**(10) 5377–5399.
[18] Meckes, E. (2008). Linear functions on the classical matrix groups. *Trans. Amer. Math. Soc.* **360**(10) 5355–5366.
[19] Meckes, E. S. and Meckes, M. W. (2007). The central limit problem for random vectors with symmetries. *J. Theoret. Probab.* **20**(4) 697–720.
[20] Meckes, M. (2009). Gaussian marginals of convex bodies with symmetries. *Beiträge Algebra Geom.* **50**(1) 101–118.

[21] Raič, M. (2004). A multivariate CLT for decomposable random vectors with finite second moments. *J. Theoret. Probab.* **17**(3) 573–603.

[22] Raič, M. A multivariate CLT for decomposable random vectors with finite second moments. *J. Theoret. Probab.* **17**(3) 573–603, 2004.

[23] Reinert, G. and Röllin, A. Multivariate normal approimation with Stein's method of exchangeable pairs under a general linearity condition. Available online at `http://arxiv.org/abs/0711.1082`.

[24] Rinott, Y. and Rotar, V. (1996). A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics. *J. Multivariate Anal.* **56**(2) 333–350.

[25] Rinott, Y. and Rotar, V. (1997). On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.* **7**(4) 1080–1105.

[26] Ruciński, A. (1988). When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields* **78**(1) 1–10.

[27] Shao, Q.-M. and Su, Z.-G. (2006). The Berry-Esseen bound for character ratios. *Proc. Amer. Math. Soc.* **134**(7) 2153–2159 (electronic).

[28] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory* 583–602. Univ. California Press, Berkeley, CA.

[29] Stein, C. (1986). *Approximate Computation of Expectations. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **7**. Institute of Mathematical Statistics, Hayward, CA.