An Infinitesimal Version of Stein's Method of Exchangeable Pairs

A dissertation submitted to the Department of Mathematics and the Committee on Graduate Studies of Stanford University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Elizabeth Meckes
June 2006

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Persi Diaconis) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Amir Dembo)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Rafe Mazzeo)

Approved for the University Committee on Graduate Studies.

Abstract

The central theme of this dissertation is an extension of Stein's method of exchangeable pairs for use in proving the approximate normality of random variables which are invariant under a continuous group of symmetries. A key feature of the technique is that, for univariate approximation, it provides convergence rates in the total variation metric as opposed to the weaker notions of distance obtained via the classical versions of Stein's method. This new technique is applied to projections of Haar measure on the classical matrix groups as well as spherically symmetric distributions on Euclidean space. The technique is also used in studying eigenfunctions of the Laplacian on a large class of Riemannian manifolds. A multivariate version of the method is developed and applied to higher dimensional projections of Haar measure on the classical matrix groups and spherically symmetric distributions on Euclidean space.

Acknowledgements

I am indebted to a large group of people for essential help in writing this dissertation.

Many thanks to Leo Rosales, Steve Kerckhoff, Brett Parker, Steve Zelditch, Elon Lindenstrauss, Olivier Glassey, Daniel Ford and Geir Helleloid for helpful conversations and valuable feedback.

Particular thanks to my collaborator Sourav Chatterjee, whose ideas for a multivariate version of the method of exchangeable pairs got the work of chapter 5 started.

Tyler Lawson has been a never-ending source of patient explanations and also a great friend.

Pierre Albin's enthusiasm and willingness to bridge gaps between fields made it far less daunting to try to venture into the world of differential geometry. Thanks to Pierre and Kay Albin for their friendship and support.

I thank Amir Dembo for many helpful conversations; he is an inspiration as a mathematician and an educator.

Rafe Mazzeo has been a steady source of very useful conversations and is also a lot of fun at dinner parties.

I'd like to give special thanks to Charles Stein. His ideas form the basis for everything in this dissertation.

Thanks the Stanford Department of Mathematics for being a great place to be a graduate student; in particular, thanks for all the ballet classes.

Thanks to the ARCS foundation for support during my final year as a graduate student.

Particular thanks to the members of my committees: Hans Andersen, Amir Dembo, Persi Diaconis, Sharad Goel and Rafe Mazzeo.

For his roles as collaborator, sounding board, husband and friend, I cannot thank Mark Meckes enough.

Finally, Persi Diaconis has been a wonderful advisor. It would be impossible to list all of his contributions to my work and my time as a graduate student. I would, however, like to thank him particularly for teaching me to enjoy mathematics; for saying, "It feels good to prove a theorem."

This dissertation is dedicated to the Pink Floyd album *The Wall* and the R.E.M. album *Automatic for the People*, which were my constant companions while studying for quals.

Contents

CHAPTER 1

Introduction

1.1. Background: Stein's method

Charles Stein first published what has come to be known as Stein's method in 1970 in the paper [**78**], as a tool for proving central limit theorems for sums of dependent random variables. The central idea of the method is that the standard Gaussian distribution on $\mathbb{R}$ is characterized by a certain differential operator. This idea is made precise by the following lemma.

LEMMA 1.1 (Stein). *Let Z be a standard Gaussian random variable. Then*

(i) *For all $f \in C_o^1(\mathbb{R})$,*

$$\mathbb{E}\big[f'(Z) - Zf(Z)\big] = 0.$$

(ii) *If Y is a random variable such that*

$$\mathbb{E}\big[f'(Y) - Yf(Y)\big] = 0$$

for all $f \in C^1(\mathbb{R})$, then $\mathcal{L}(Y) = \mathcal{L}(Z)$; i.e., Y is also distributed as a standard Gaussian random variable.

Remark: The differential operator T_o defined by

$$T_o f(x) = f'(x) - xf(x) \tag{1.1}$$

is called the characterizing operator of Z.

The first statement of the lemma follows trivially by integration by parts. To prove the second part, Stein introduced a left inverse to the characterizing operator T_o given by the following formula

$$U_o g(t) = e^{t^2/2} \int_{-\infty}^{t} \big[g(x) - \mathbb{E}g(Z)\big] e^{-x^2/2} dx. \tag{1.2}$$

Integration by parts shows

$$\begin{aligned} T_o U_o g(x) &= (U_o g)'(x) - x(U_o g)(x) \\ &= g(x) - \mathbb{E}g(Z), \\ U_o T_o f(x) &= f(x); \end{aligned}$$

in particular, $U_o g$ is a solution to the differential equation

$$h'(x) - xh(x) = g(x) - \mathbb{E}g(Z).$$

Now to prove (ii), let g be a smooth, compactly supported test function and let $h = U_o g$. Then, since $\mathbb{E}\big[f'(Y) - Yf(Y)\big] = 0$ for all f,

$$\mathbb{E}g(Y) - \mathbb{E}g(Z) = \mathbb{E}\big[h'(Y) - Yh(Y)\big] = 0.$$

Since C_o^∞ is dense (with respect to the supremum norm) in the set of bounded continuous functions vanishing at infinity, this proves that Y and Z have the same distribution.

After observing that the Gaussian distribution is characterized by the differential operator T_o, the next essential observation made by Stein was that if $\mathbb{E}T_o f(Y) = \mathbb{E}\big[f'(Y) - Yf(Y)\big]$ is *small* for all f in some large class of functions, then Y is close to Z in law. As above, if g is a test function and $h = U_o g$, then

$$\big|\mathbb{E}g(Y) - \mathbb{E}g(Z)\big| = \big|\mathbb{E}\big[h'(Y) - Yh(Y)\big]\big|. \tag{1.3}$$

Now,

$$\sup_{\substack{\|g\|_\infty \le 1, \\ g\, continuous}} \big|\mathbb{E}g(Y) - \mathbb{E}g(Z)\big| =: 2d_{TV}(Y, Z)$$

and

$$\sup_{\|g\|_L \le 1} \big|\mathbb{E}g(Y) - \mathbb{E}g(Z)\big| =: d_{L^*}(Y, Z),$$

where $\|g\|_L$ is the Lipschitz norm of g; i.e., the sum of $\|g\|_\infty$ and the Lipschitz constant of g. This means that estimating $\left|\mathbb{E}\big[h'(Y) - Yh(Y)\big]\right|$ over a class of h which includes $U_o g$ for g bounded and continuous (resp. bounded, Lipschitz) gives the distance between the random variables Y and Z in the total variation (resp. dual-Lipschitz) metric. Indeed, one of the advantages of Stein's method is that when used to prove probabilistic limit theorems, it automatically provides a rate of convergence in a metric of this kind.

There are various ways to try to estimate the right-hand side of (1.3). One general technique, developed by Stein in [**79**] is the method of exchangeable pairs. Very roughly, the idea is the following. Given a random variable W conjectured to be approximately normal, make a small change to W to construct a random variable W' such that (W, W') is exchangeable. Use the pair (W, W') to approximate $h'(W)$ when estimating

$$\mathbb{E}\big[h'(W) - Wh(W)\big]$$

and use the exchangeability to evaluate the resulting expression. More formally, the following abstract normal approximation theorem is the basis for the method of exchangeable pairs. The proof is given below.

Theorem 1.2 (Stein). *Let (W, W') be an exchangeable pair of random variables with $\mathbb{E}W^2 = 1$, and let $\Delta = W' - W$. Suppose that*

$$\mathbb{E}\left[\Delta \middle| W\right] = -\lambda W.$$

Then

(i)

$$\left|\mathbb{E}g(W) - \mathbb{E}g(Z)\right| \le \frac{1}{\lambda}\|g - \mathbb{E}g(Z)\|_\infty \sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + \frac{\|g'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3$$

(ii) *for any $t \in \mathbb{R}$,*

$$\left|\mathbb{P}(W \le t) - \Phi(t)\right| \le \frac{1}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[\Delta^2 \middle| W\right]\right)} + (8\pi)^{-1/4}\sqrt{\frac{1}{\lambda}\mathbb{E}|\Delta|^3},$$

where $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$.

This theorem was improved by Rinott and Rotar [**71**] in two main directions. They showed that the condition $\mathbb{E}\left[\Delta \middle| W\right] = -\lambda W$ need only hold approximately and gave a modified form of the bound (i) taking $R := \mathbb{E}\left[\Delta \middle| W\right] + \lambda W$ into account. In the case that Δ is almost surely bounded, they improved the error bound of (ii) significantly. Theorem 1.2 above yields only a rate of $n^{-1/4}$ in the Berry-Esseen theorem, whereas for bounded random variables, Rinott and Rotar's improvement yields the correct rate of $n^{-1/2}$. Further work in replacing the boundedness assumption by moment assumptions and getting sharp bounds has been done by Chatterjee [**20**] and Shao and Su [**76**].

The following lemma is one of the key ingredients of the proof of Theorem 1.2 and will also be used in later sections.

Lemma 1.3 (Stein). *Let U_o be the operator defined in equation (1.2) and let $g : \mathbb{R} \to \mathbb{R}$ be bounded and continuous. Then*

(i) $\|U_o g\|_\infty \leq \sqrt{\frac{\pi}{2}} \|g - \mathbb{E}g(Z)\|_\infty \leq \sqrt{2\pi} \|g\|_\infty$
(ii) $\|(U_o g)'\|_\infty \leq 2\|g - \mathbb{E}g(Z)\|_\infty \leq 4\|g\|_\infty$
(iii) *If g is also differentiable, then* $\|(U_o g)''\|_\infty \leq 2\|g'\|_\infty$

The proof of the lemma is by calculus and can be found in [**79**].

Proof of Theorem 1.2. Let $g : \mathbb{R} \to \mathbb{R}$ be continuous and compactly supported with $\|g\|_L \leq 1$, and let $h = U_o g$ for U_o as in equation (1.2). By the exchangeability of (W, W'),

$$\begin{aligned} 0 &= \mathbb{E}\left[(W' - W)(h(W') + h(W))\right] \\ &= \mathbb{E}\left[(W' - W)(h(W') - h(W)) + 2(W' - W)h(W)\right] \\ &= \mathbb{E}\left[h'(W)\mathbb{E}\left[(W' - W)^2 \middle| W\right] + 2h(W)\mathbb{E}\left[(W' - W)\middle| W\right] + R\right], \end{aligned}$$

where R is the error in the derivative approximation. Applying the condition of the theorem and rearranging gives

$$0 = \mathbb{E}\left[2\lambda h'(W) - 2\lambda W h(W) + h'(W)\left(\mathbb{E}\left[(W' - W)^2 \middle| W\right] - 2\lambda\right) + R\right],$$

thus

$$\left|\mathbb{E}\big[g(W)-g(Z)\big]\right| = \left|\mathbb{E}\left[\frac{h'(W)}{2\lambda}\left(2\lambda - \mathbb{E}\left[(W'-W)^2\middle|W\right]\right) - \frac{R}{2\lambda}\right]\right| \tag{1.4}$$

Note that

$$\mathbb{E}\left(\mathbb{E}\left[(W'-W)^2\middle|W\right]\right) = 1 - 2\mathbb{E}\big[(1-\lambda)W^2\big] + 1 = 2\lambda.$$

It follows by the Cauchy-Schwarz inequality and Lemma 1.3 that

$$\left|\frac{h'(W)}{2\lambda}\left(\mathbb{E}\left[(W'-W)^2\middle|W\right] - 2\lambda\right)\right| \le \frac{\|g-\mathbb{E}g(Z)\|_\infty}{\lambda}\sqrt{\operatorname{Var}\left(\mathbb{E}\left[\Delta^2\middle|W\right]\right)}. \tag{1.5}$$

By Taylor's theorem,

$$\begin{aligned}\frac{1}{2\lambda}\mathbb{E}|R| &\le \frac{\|h''\|_\infty}{4\lambda}\mathbb{E}|W'-W|^3 \\ &\le \frac{\|g'\|_\infty}{2\lambda}\mathbb{E}|\Delta|^3.\end{aligned} \tag{1.6}$$

Applying the triangle inequality to (1.4) and using the bounds (1.5) and (1.6) proves (i).

For (ii), let

$$h_{t,\delta}(x) = \begin{cases} 1 & x \le t \\ 1 - \frac{x-t}{\delta} & t \le x \le t+\delta \\ 0 & x \ge t+\delta. \end{cases}$$

Then

(i) $\mathbb{E}h_{t-\delta,\delta}(W) \le \mathbb{P}(W \le t) \le \mathbb{E}h_{t,\delta}(W)$,

(ii) $|h_{t,\delta}(x) - \mathbb{E}h_{t,\delta}(Z)| \le 1$ for all $x \in \mathbb{R}$, and

(iii) $|h'_{t,\delta}(x)| \le \dfrac{1}{\delta}$ for all $x \in \mathbb{R}$.

Applying the first statement of the theorem to the test function $h_{t,\delta}$ gives

$$\begin{aligned}\mathbb{P}(W \le t) &\le \mathbb{E}h_{t,\delta}(W) \\ &\le \mathbb{E}h_{t,\delta}(Z) + \frac{1}{\lambda}\sqrt{\operatorname{Var}\left(\mathbb{E}\left[(W'-W)^2\middle|W\right]\right)} + \frac{1}{2\lambda\delta}\mathbb{E}|W'-W|^3 \\ &\le \Phi(t) + \frac{1}{\lambda}\sqrt{\operatorname{Var}\left(\mathbb{E}\left[(W'-W)^2\middle|W\right]\right)} + \frac{\delta}{\sqrt{2\pi}} + \frac{1}{2\lambda\delta}\mathbb{E}|W'-W|^3.\end{aligned}$$

The right-hand side is minimized by choosing

$$\delta = \frac{(8\pi)^{1/4}}{2}\sqrt{\frac{1}{\lambda}\mathbb{E}|W' - W|^3};$$

this gives that

$$\mathbb{P}(W \leq t) \leq \Phi(t) + \frac{1}{\lambda}\sqrt{\mathrm{Var}\left(\mathbb{E}\left[(W'-W)^2\middle|W\right]\right)} + (8\pi)^{-1/4}\sqrt{\frac{1}{\lambda}\mathbb{E}|W'-W|^3}.$$

A corresponding lower bound can be proved similarly by using the lower bound from (i) above.

$\square$

Remark: Estimating the remainder in the derivative approximation more carefully can lead to a slight improvement in the coefficient of $\mathbb{E}|\Delta|^3$; see [**79**].

Since the first publication of Stein's method in 1970, a great deal of work has been done in extending and applying the method. As mentioned above, Rinott and Rotar [**71**] made significant improvements to the method of exchangeable pairs. They applied their result to studying the stationary distribution of the antivoter chain introduced by Matloff [**61**] and to weighted U-statistics. The method is extended to a multivariate context in [**22**], which includes the results of chapter 5 of this thesis, as well as an application to linear functions on the permutation group.

Usually when using the method of exchangeable pairs, construction of the exchangeable pair itself is quite straightforward and suggested by the natural symmetries of the problem. In many examples, the random variable W being studied is a function of some other random object X. For example, X may be a sequence of n i.i.d. random variables or a random element of a compact group. In such cases, W' is often constructed by first defining an exchangeable pair (X, X') and then applying the same function to X and X' to get W and W'. For example, a natural choice in the case of a sum of a sequence of random variables $\{X_i\}$ is to choose an index I at random and replace X_I with an independent copy X_I^*. However, it need not be the case that W' is constructed in this way. In studying certain permutation statistics, Jason Fulman [**41**] introduced a clever construction for an exchangeable pair (W, W') where $W = W(\pi)$ with π a random permutation, such that $W' = W(\pi')$ but (π, π') is not exchangeable.

As mentioned above, the method of exchangeable pairs is just one possible implementation of the basic idea of Stein's method, namely that the normal distribution can be characterized by a differential operator. There are other somewhat similar approaches which go under the general name of 'coupling' or 'auxiliary randomization'. In these methods, an auxiliary random variable (like W' above) is constructed from W in some way as a tool for estimating

$$\mathbb{E}\big[h'(W) - Wh(W)\big] \tag{1.7}$$

over a large class of h. One such approach is the zero-bias coupling method of Goldstein and Reinert [**44**]. Motivated by the form of (1.7), they construct a random variable W^* from W such that

$$\mathbb{E}Wh(W) = \mathbb{E}h'(W^*)$$

for all $h \in C^1(\mathbb{R})$ for which the expectations above are defined. Goldstein and Reinert applied this coupling to simple random sampling; Goldstein [**43**] later applied it in proving various combinatorial central limit theorems and, in collaboration with Aihua Xia [**47**], in proving a 'discrete central limit theorem'.

Another coupling approach which has been used frequently, e.g. in [**5**], [**46**], [**68**], is the size bias coupling. This approach has been especially useful for positive random variables such as counts; for example, it was used in [**5**] in studying the number of local maxima of random functions on graphs. Assume $W \geq 0$ and let λ, σ be defined by $\mathbb{E}W = \lambda$ and $\mathbb{E}W^2 = \sigma^2$. The variable W^* is defined by the relation

$$\mathbb{E}Wh(W) = \lambda\mathbb{E}h(W^*).$$

A rather different approach to Stein's method is the method of 'local dependence', introduced for use in Poisson approximation by Chen in [**25**], [**26**], and further developed in [**2**]. Baldi and Rinott [**4**] developed a version for studying the approximate normality of graph-related statistics. The method is useful when there exists a natural dependence structure in the problem, e.g., a random variable $W = \sum_i X_i$ where the X_i depend only on X_j with $|i-j|$ small. Situations of this sort often arise in geometric problems; the local dependence method has recently been of use in the study of random polytopes (see [**69**], [**6**]). A more refined version of local dependency structure was exploited in a multivariate context in [**70**], which contains a fairly general multivariate central limit theorem with applications to graph colorings.

For a very readable introduction to Stein's method for normal approximation via these methods, see Rinott and Rotar [**72**].

Stein's method has also been used in conjunction with induction. Bolthausen [**16**] used this approach in determining a rate of convergence in the combinatorial central limit theorem of Hoeffding (see [**51**]). Stein [**80**] later used this approach to obtain fast rates of convergence to normal for $\mathrm{Tr}\,(M^k)$ for M a random $n \times n$ orthogonal matrix and k fixed.

One of the most celebrated aspects of Stein's method is that it allows the relaxation of independence assumptions. The other main advantage is that the method can be used for many different approximating distributions, not just the normal distribution. Shortly after its introduction, Louis Chen [**25**], [**26**] developed a version for Poisson approximation which he applied to sums of dependent Bernoulli random variables. Stein's method for Poisson approximation was developed further by Arratia, Goldstein, and Gordon in [**1**] and [**2**], including developing a local dependency approach which only requires separating summands into weakly and strongly dependent groups, rather than requiring strict independence. Penrose's book [**66**] has applications of local dependency versions of Stein's method in studying random geometric graphs. The book [**10**] of Barbour, Holst, and Janson on Poisson approximation uses Stein's method in conjunction with size bias coupling and contains applications to a wide variety of problems. The survey paper [**21**] gives a version of the method of exchangeable pairs for Poisson approximation, including many examples.

Stein's method has also been developed in the context of approximation by Γ distributions (see Luk [**60**]), χ^2 distributions (see Pickett [**67**]), the uniform distribution on the discrete circle (see the paper [**33**] of Diaconis in [**37**]), the semi-circle law (see Götze and Tikhomirov [**48**]) the binomial and multinomial distributions (see Holmes [**52**], Loh [**59**]), the hypergeometric distribution (also in [**52**]) and the uniform distribution on S^1 (see Meckes [**62**]).

Barbour made a significant contribution to the development of Stein's method in the papers [**7**] and [**8**], which develop versions of the method in the context of approximation by Poisson processes and Gaussian processes. A different approach to Poisson process approximation was introduced by Arratia, Goldstein and Gordon in [**1**]. The ideas introduced in [**8**] have subsequently been widely exploited; see e.g. [**46**], [**45**], and [**22**].

A testament to the power of Stein's method is the extremely wide variety of problems which have been successfully treated using one or more of the approaches discussed above. As has already been mentioned, many problems involving graph statistics and various types of random graphs have been treated; see [**5**], [**4**], [**12**], [**11**], [**66**], [**70**]. Reitzner [**69**] applied Stein's method in studying random polytopes; see also [**6**]. Many types of sums of dependent random variables, both with local dependence and global dependence, have been treated with various versions of the method; see [**25**], [**26**], [**30**], [**53**], [**71**]. In particular, limiting distributions of linear functions in various contexts ([**16**], [**75**], [**63**], [**64**]) have been treated. Many other applications have been explored, including dissociated statistics [**9**], scan statistics [**30**], permutation statistics [**41**], pattern occurrences in sequences of trials [**27**], the bootstrap [**53**], and stationary distributions of Markov chains [**33**], [**54**], [**71**].

1.2. Summary of results

One of the fundamental results of this thesis is the following extension of Theorem 1.2 to situations in which continuous symmetries are present.

THEOREM 1.4. *Suppose that (W, W_ϵ) is a family of exchangeable pairs defined on a common probability space with $\mathbb{E}W = 0$ and $\mathbb{E}W^2 = \sigma^2$. Suppose there is a random variable $E = E(W)$, deterministic functions $h(\epsilon)$ and $k(\epsilon)$ with*

$$\lim_{\epsilon \to 0} h(\epsilon) = \lim_{\epsilon \to 0} k(\epsilon) = 0,$$

and functions α *and* β *with*

$$\mathbb{E}|\alpha(\sigma^{-1}W)| < \infty, \qquad \mathbb{E}|\beta(\sigma^{-1}W)| < \infty,$$

such that

(i)
$$\frac{1}{\epsilon^2}\mathbb{E}\left[W_\epsilon - W\middle|W\right] = -\lambda W + h(\epsilon)\alpha(W),$$

(ii)
$$\frac{1}{\epsilon^2}\mathbb{E}\left[(W_\epsilon - W)^2\middle|W\right] = 2\lambda\sigma^2 + E\sigma^2 + k(\epsilon)\beta(W),$$

(iii)
$$\frac{1}{\epsilon^2}\mathbb{E}\left|W_\epsilon - W\right|^3 = o(1),$$

where $o(1)$ *refers to the limit as* $\epsilon \to 0$ *with implied constants depending on the distribution of* W*. Then*

$$d_{TV}(W, Z) \le \frac{1}{\lambda}\mathbb{E}\left|E\right|,$$

where $Z \sim \mathfrak{N}(0, \sigma^2)$*.*

In situations in which it is possible to construct (W, W_ϵ) as above, Theorem 1.4 gives a significant improvement to Theorem 1.2. Consider the following example. Let X be a random point of $\sqrt{n}S^{n-1} \subseteq \mathbb{R}^n$. The following theorem, first proved in [**36**], is shown in section 2.1 below to be a consequence of Theorem 1.4.

Theorem 1.5. *Let* X *be as above. For* $\theta \in S^{n-1}$*, define* $W_\theta = \langle X, \theta\rangle$*. Then*

$$d_{TV}\big(W_\theta, Z\big) \le \frac{4}{n-1},$$

where $Z \sim \mathfrak{N}(0, 1)$*.*

One could try to prove the asymptotic normality of W_θ as a consequence of Theorem 1.2. First note that by the symmetry of the sphere, it suffices to prove the theorem for one fixed θ, e.g., the principle diagonal $\theta = \frac{1}{\sqrt{n}}(1, \ldots, 1)$. A natural exchangeable pair is the following. Let X be a random point of $\sqrt{n}S^{n-1}$, and let $I \in \{1, \dots, n\}$ be a random index, independent of X. Let

$$X' = X - 2e_I X_I, \tag{1.8}$$

where e_j is the jth standard basis vector of $\mathbb{R}^n$. Thus X' is constructed from X by choosing a coordinate at random and changing the sign. Let $W = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ and $W' = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i'$; the pair (W, W') is exchangeable by the symmetry of the sphere.

The following computation shows that Theorem 1.2 applies with $\lambda = \frac{2}{n}$. Making use of the independence of X and I,

$$\begin{aligned}
\mathbb{E}\left[W' - W \middle| W\right] &= \mathbb{E}\left[-\frac{2}{\sqrt{n}}X_I \middle| W\right] \\
&= -\frac{2}{n^{3/2}}\sum_{i=1}^n \mathbb{E}\left[X_i \middle| W\right] \\
&= -\frac{2}{n}\mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \middle| W\right] \\
&= -\frac{2}{n}W.
\end{aligned}$$

It remains to bound the error terms provided by Theorem 1.2. The following formula from [**40**] for integrating polynomials over spheres is useful. The theorem in [**40**] is only stated for polynomials with even powers of each coordinate (and with a different normalization of surface measure on S^{n-1}), but the proof of the theorem stated below is identical.

THEOREM 1.6. *Let $P(x) = |x_1|^{\alpha_1}|x_2|^{\alpha_2}\cdots|x_n|^{\alpha_n}$. Then if X is uniformly distributed on $\sqrt{n}S^{n-1}$,*

$$\mathbb{E}\big[P(X)\big] = \frac{\Gamma(\beta_1)\cdots\Gamma(\beta_n)\Gamma(\frac{n}{2})n^{\left(\frac{1}{2}\sum\alpha_i\right)}}{\Gamma(\beta_1 + \cdots + \beta_n)\pi^{n/2}},$$

where $\beta_i = \frac{1}{2}(\alpha_i + 1)$ for $1 \le i \le n$ and

$$\Gamma(t) = \int_0^\infty s^{t-1}e^{-s}ds = 2\int_0^\infty r^{2t-1}e^{-r^2}dr.$$

Now,

$$\begin{aligned}\mathbb{E}\left[(W'-W)^2\middle|W\right] &= \mathbb{E}\left[\frac{4}{n}X_I^2\middle|W\right]\\ &= \frac{4}{n^2}\sum_{i=1}^n \mathbb{E}\left[X_i^2\middle|W\right]\\ &= \frac{4}{n}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i^2\middle|W\right]\\ &= \frac{4}{n}.\end{aligned}$$

Therefore,

$$\mathrm{Var}\left(\mathbb{E}\left[(W'-W)^2\middle|W\right]\right) = 0,$$

and the first half of the error term in either statement of Theorem 1.2 is zero.

For the second term,

$$\begin{aligned}\mathbb{E}\left|W'-W\right|^3 &= \frac{8}{n^{3/2}}\mathbb{E}\left|X_I\right|^3\\ &= \frac{8}{n^{5/2}}\sum_{i=1}^n \mathbb{E}\left|X_i\right|^3 \qquad (1.9)\\ &= \frac{8}{n^{3/2}}\mathbb{E}\left|X_1\right|^3,\end{aligned}$$

where the last line follows again by the symmetries of the sphere. By Theorem 1.6 and the Cauchy-Schwarz inequality,

$$\mathbb{E}\left|X_1\right|^3 \le \left(\mathbb{E}X_1^4\right)^{3/4} = \left(\frac{3n}{n+2}\right)^{3/4} \le 3^{3/4}.$$

Note that the Cauchy-Schwarz inequality also implies that

$$\mathbb{E}\left|X_1\right|^3 \ge \left(\mathbb{E}X_1^2\right)^{3/2} = 1,$$

so the loss in estimating $\mathbb{E}\left|X_1\right|^3$ by $3^{3/4}$ is only a loss in the constant. Combining this estimate with (1.9) yields

$$\frac{1}{\lambda}\mathbb{E}\left|W'-W\right|^3 \le \frac{4\cdot 3^{3/4}}{\sqrt{n}}, \qquad (1.10)$$

thus Theorem 1.2 gives a rate of convergence of W_θ to Gaussian on the order of $n^{-1/2}$ in the dual-Lipschitz distance and shows that the distance between the distribution functions is bounded by a bound of order $n^{-1/4}$. The theorem does not provide a bound on the total variation distance from W_θ to a Gaussian random variable. Theorem 1.4 thus gives a faster rate of convergence in a stronger metric in this problem.

Though its proof is quite simple, Theorem 1.4 above is one of the foundational results of this thesis; all of the results of chapters 3 and 4 make use of it. Chapter 2 contains the proof of the theorem and a first application: under certain conditions, rank one projections of spherically symmetric distributions on $\mathbb{R}^n$ are close to the standard Gaussian distribution in the total variation distance. This is related to work of Diaconis and Freedman [**36**]; some comparison is offered in section 2.1.

Chapter 3 is concerned with applications of Theorem 1.4 to studying the behavior of linear functions on the classical matrix groups. A theme in studying random matrices distributed according to Haar measure on the orthogonal group $\mathcal{O}_n$ and the unitary group $\mathcal{U}_n$ has been that in many ways, Haar measure is 'like' Gaussian measure; see, e.g., [**36**], [**29**] and [**55**]. This is not true in all ways – in particular, [**42**] and [**35**] show that the eigenvalue distributions are very different – but is often true when studying the joint distribution of the entries. One way of defining Gaussian measure on higher dimensional space (such as the space of $n \times n$ matrices) is to require all projections to be Gaussian. Thus a natural comparison between Haar measure and Gaussian measure on $\mathcal{O}_n$ or $\mathcal{U}_n$ is to determine the distributions of linear functions and compare them to the normal distribution on $\mathbb{R}$. This motivates the following theorem of section 3.1.

Theorem 1.7. *Let A be a fixed $n \times n$ matrix over $\mathbb{R}$ such that $\mathrm{Tr}\,(AA^t) = n$, $M \in \mathcal{O}_n$ distributed according to Haar measure, and $W = \mathrm{Tr}\,(AM)$. Let Z be a standard normal random variable. Then for $n > 1$,*

$$d_{TV}(W, Z) \leq \frac{2\sqrt{3}}{n-1}. \tag{1.11}$$

Further motivation and historical background for this theorem are given in chapter 3.

The natural analog of Theorem 1.7 in the unitary case would be a bound on the total variation distance between the complex random variable $\mathrm{Tr}\,(AM)$ and a standard complex normal random variable, where M is a random unitary matrix and A a fixed $n \times n$ matrix over $\mathbb{C}$. The difficulty is that looking for such a bound is a bivariate problem, thus Theorem 1.4 does not apply directly. It can be applied to prove the following univariate theorem.

THEOREM 1.8. *Let M be a random unitary matrix, distributed according to Haar measure, and let A be a fixed $n \times n$ matrix over $\mathbb{C}$ with $\mathrm{Tr}\,(AA^*) = n$. Let $W = \mathrm{Tr}\,(AM)$ and let W_θ be the inner product of W with the unit vector making angle θ with the real axis. Then*
$$d_{TV}\left(W_\theta, \mathfrak{N}\left(0, \frac{1}{2}\right)\right) \leq \frac{c}{n}$$
for a constant c which is independent of θ.

The bivariate approximation problem is treated in chapter 5, which has a multivariate analog of Theorem 1.4.

Chapter 4 contains an abstraction of the results of chapter 3 to a much more general setting than the classical matrix groups. The main result of the chapter is the following.

THEOREM 1.9. *Let M be a compact locally symmetric space without boundary and f an eigenfunction for the Laplacian on M with eigenvalue $\lambda \neq 0$, normalized so that $\int_M f^2 = 1$. Let X be a random (i.e., distributed according to normalized volume measure) point of M. Then*
$$d_{TV}(f(X), Z) \leq \frac{1}{|\lambda|}\sqrt{\mathrm{Var}\big(\|\nabla f(X)\|^2\big)},$$
where Z is a standard Gaussian random variable on $\mathbb{R}$.

In this generality, the theorem is not a limit theorem but a relationship between the distribution of eigenfunctions and how much their gradients vary. One could try to apply Theorem 1.9 to obtain limit theorems in various contexts. One possibility is to consider a sequence of manifolds with dimension tending to infinity, and bound the right-hand side of the expression above by something tending to zero as the dimension tends to infinity. An example of such an application is the following.

THEOREM 1.10. *Let* $g(x) = \sum_{i,j} a_{ij} x_i x_j$*, where* $A = (a_{ij})_{i,j=1}^n$ *is a symmetric matrix with* $\mathrm{Tr}\,(A) = 0$*. Let* $f = Cg$*, where* C *is chosen such that* $\|f\|_2 = 1$ *when* f *is considered as a function on* S^{n-1}*. Let* $W = f(X)$*, where* X *is a random point on* S^{n-1}*. If* d *is the vector in* $\mathbb{R}^n$ *whose entries are the eigenvalues of* A*, then*

$$d_{TV}(W, Z) \leq \sqrt{5} \left(\frac{\|d\|_4}{\|d\|_2} \right)^2,$$

where $\|d\|_p = \left(\sum_i |d_i|^p\right)^{1/p}$.

In certain natural examples, the bound on the right hand side above is of the order $\frac{1}{\sqrt{n}}$, a statement which is discussed in more detail in section 4.1.

Another direction one could try to go with Theorem 1.9 is to consider a fixed manifold and investigate behavior of eigenfunctions as the eigenvalue tends to infinity. This is closely connected with the field of quantum chaos; in particular, Hejhal and various coauthors (see, e.g, [**49**]) have provided numerical evidence that on certain hyperbolic surfaces, eigenfunctions with high eigenvalues are approximately normally distributed.

Chapter 5 deals with multivariate extensions of the results of chapters 2 and 3. The following theorem of section 5.1 gives an abstract framework for approximation by a multivariate Gaussian random variable in situations having continuous symmetries.

THEOREM 1.11. *Let* X *and* X_ϵ *be two random vectors such that* $\mathcal{L}(X) = \mathcal{L}(X_\epsilon)$ *with the property that* $\lim_{\epsilon \to 0} X_\epsilon = X$ *almost surely. Suppose there is a constant* λ*, functions* $E_{ij} = E_{ij}(X)$*,* h *and* k *with*

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon^2} h(\epsilon) = \lim_{\epsilon \to 0} \frac{1}{\epsilon^2} k(\epsilon) = 0,$$

and $\alpha(X)$ *and* $\beta(X)$ *with*

$$\mathbb{E}|\alpha(X)|, \mathbb{E}|\beta(X)| < \infty,$$

such that

(i) $\mathbb{E}\left[(X_\epsilon - X)_i \middle| X\right] = -\lambda \epsilon^2 X_i + h(\epsilon)\alpha(X)$
(ii) $\mathbb{E}\left[(X_\epsilon - X)_i (X_\epsilon - X)_j | X\right] = 2\lambda \epsilon^2 \delta_{ij} + \epsilon^2 E_{ij} + k(\epsilon)\beta(X)$
(iii) $\mathbb{E}\left[|X_\epsilon - X|^3\right] = o(\epsilon^2)$.

Then

$$d_{L^*}(X,Z) \leq \min\left\{\frac{1}{2\lambda}\sum_{i,j}\mathbb{E}|E_{ij}|, \frac{\sqrt{k}}{2\lambda}\mathbb{E}\left(\sum_{i,j}E_{ij}^2\right)^{1/2}\right\}.$$

If furthermore $E_{ij} = E_iF_j + \delta_{ij}R_i$, *then writing* E *for the vector with components* E_i *and* F *for the vector with components* F_j,

$$d_{L^*}(X,Z) \leq \frac{1}{\lambda\sqrt{2\pi}}\mathbb{E}\left(\|E\|_2\|F\|_2\right) + \frac{1}{\lambda}\sum_{i=1}^{k}\mathbb{E}|R_i|.$$

Theorem 1.11 is not analogous to Theorem 1.4 in the strictest sense, since the bounds given are for dual-Lipschitz distance rather than total variation distance. It seems to be unavoidable using the techniques of section 5.1 that rates of convergence will be with respect to a weaker notion of distance between measures than total variation distance.

The rest of chapter 5 contains applications of Theorem 1.11. The first example, treated in section 5.2, is a multivariate analog of the example of section 2.1, showing that rank k projections of spherically symmetric distributions on $\mathbb{R}^n$ are close to normal for $k = o(n)$. As this is an application of Theorem 1.11, 'close' refers here to the dual-Lipschitz distance. This is closely related to the results of the paper [**36**] of Diaconis and Freedman; discussion of the connections and some comparison of results are given in section 5.2.

As alluded to above, chapter 5 contains an extension of Theorem 1.8 to the complex setting; this is the topic of section 5.3. The result is the following.

THEOREM 1.12. *Let* $M \in \mathfrak{U}_n$ *be distributed according to Haar measure, and let* A *be a fixed* $n \times n$ *matrix such that* $\mathrm{Tr}\,(AA^*) = n$. *Let* $W = \mathrm{Tr}\,(AM)$ *and let* Z *be a standard complex normal random variable. Then there is a constant* c *such that*

$$d_{L^*}(W,Z) \leq \frac{c}{n}.$$

Finally, section 5.4 gives a multivariate extension of the main result of section 3.1 by showing that for $k = o(n^{2/3})$, rank k projections of Haar measure are close to

k-dimensional Gaussian random vectors. This is a finer comparison between Haar-distributed orthogonal matrices and Gaussian matrices, since for Gaussian matrices, all projections to any lower dimensional spaces are Gaussian. A somewhat simplified version of the main result of section 5.4 is the following.

THEOREM 1.13. *Let* $A_1, \ldots, A_k$ *be* $n \times n$ *matrices over* $\mathbb{R}$ *satisfying* $\mathrm{Tr}\,(A_i A_j^t) = n\delta_{ij}$*; for* $i \neq j$*,* A_i *and* A_j *are orthogonal with respect to the Hilbert-Schmidt inner product. Let* M *be a random orthogonal matrix, and consider the vector*

$$W = (\mathrm{Tr}\,(A_1 M), \mathrm{Tr}\,(A_2 M), \ldots, \mathrm{Tr}\,(A_k M)) \in \mathbb{R}^k.$$

Let $Z = (Z_1, \ldots, Z_k)$ *be a random vector whose components are independent standard normal random variables. Then for* $f : \mathbb{R}^k \to \mathbb{R}$ *bounded and Lipschitz,*

$$d_{L^*}(W, Z) \leq \frac{2k^{3/2}}{n-1}$$

for $n \geq 3$.

In section 5.4, this result is extended by linear algebra to parameter matrices $\{A_i\}$ which need only be linearly independent.

1.3. Notation and conventions

The total variation distance $d_{TV}(\mu, \nu)$ between the measures μ and ν on $\mathbb{R}$ is defined by

$$d_{TV}(\mu, \nu) = \sup_A |\mu(A) - \nu(A)|,$$

where the supremum is over measurable sets A. This is equivalent to

$$d_{TV}(\mu, \nu) = \frac{1}{2} \sup_f \left| \int f(t) d\mu(t) - \int f(t) d\nu(t) \right|,$$

where the supremum is taken over continuous functions which are bounded by 1 and vanish at infinity; this is the definition most commonly used in what follows. The total variation distance between two random variables X and Y is defined to be the total variation distance between their distributions:

$$d_{TV}(X,Y) = \sup_A \big|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)\big| = \frac{1}{2}\sup_f \big|\mathbb{E}f(X) - \mathbb{E}f(Y)\big|.$$

The Lipschitz norm of a Lipschitz function g on $\mathbb{R}^n$ is defined by

$$\|g\|_L = \|g\|_\infty + M_g,$$

where M_g is the Lipschitz constant of g. The dual-Lipschitz distance $d_{L^*}(X,Y)$ between the random variables X and Y is defined by

$$d_{L^*}(X,Y) = \sup_{\|g\|_L \le 1} \big|\mathbb{E}g(X) - \mathbb{E}g(Y)\big|.$$

It is shown in [**39**] that the dual-Lipschitz distance metrizes the weak-star topology on laws.

We will use $\mathfrak{N}(\mu,\sigma^2)$ to denote the normal distribution on $\mathbb{R}$ with mean μ and variance σ^2; unless otherwise stated, the random variable Z is understood to be a standard Gaussian random variable on the space in question.

The notation $C^k(\Omega)$ will be used for the space of k-times continuously differentiable real-valued functions on Ω, and $C^k_o(\Omega) \subseteq C^k(\Omega)$ are those C^k functions on Ω with compact support.

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 is defined by

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

Finally, I_n denotes the $n \times n$ identity matrix, 0_n the $n \times n$ matrix of all zeros, and $A \oplus B$ the standard block direct sum of A and B.

CHAPTER 2

An abstract normal approximation theorem

In this section, the general framework for normal approximation to random variables with continuous symmetries is developed; in particular, Theorem 2.1 below is one of the basic results of this thesis. The theorem is an abstraction of an argument which first appeared in Stein [**80**], where fast rates of convergence to Gaussian were obtained for $\mathrm{Tr}\,(M^k)$, where $k \in \mathbb{N}$ is fixed and M is a random (Haar distributed) orthogonal matrix.

THEOREM 2.1. *Suppose that* (W, W_ϵ) *is a family of exchangeable pairs defined on a common probability space with* $\mathbb{E}W = 0$ *and* $\mathbb{E}W^2 = \sigma^2$*. Suppose there is a random variable* $E = E(W)$*, deterministic functions* h *and* k *with*

$$\lim_{\epsilon \to 0} h(\epsilon) = \lim_{\epsilon \to 0} k(\epsilon) = 0$$

and functions α *and* β *with*

$$\mathbb{E}|\alpha(\sigma^{-1}W)| < \infty, \qquad \mathbb{E}|\beta(\sigma^{-1}W)| < \infty,$$

such that

(i)
$$\frac{1}{\epsilon^2}\mathbb{E}\left[W_\epsilon - W \middle| W\right] = -\lambda W + h(\epsilon)\alpha(W),$$

(ii)
$$\frac{1}{\epsilon^2}\mathbb{E}\left[(W_\epsilon - W)^2 \middle| W\right] = 2\lambda\sigma^2 + E\sigma^2 + k(\epsilon)\beta(W),$$

(iii)
$$\frac{1}{\epsilon^2}\mathbb{E}\left|W_\epsilon - W\right|^3 = o(1),$$

where $o(1)$ refers to the limit as $\epsilon \to 0$ with implied constants depending on the distribution of W. Then

$$d_{TV}(W, Z) \le \frac{1}{\lambda}\mathbb{E}\,|E|\,,$$

where $Z \sim \mathfrak{N}(0, \sigma^2)$.

Remark: The factor of $\frac{1}{\epsilon^2}$ in each of the three expressions above could be replaced by a general function $f(\epsilon)$. In practice, W_ϵ is typically constructed such that $W_\epsilon - W = O(\epsilon)$. This makes it clear that $f(\epsilon) = \frac{1}{\epsilon^2}$ is the suitable choice for condition (ii). It is less clear that $f(\epsilon) = \frac{1}{\epsilon^2}$ is the suitable choice for condition (i). In the applications given here, while $W_\epsilon - W = O(\epsilon)$, symmetry conditions imply that

$$\mathbb{E}\left[W_\epsilon - W \middle| W\right] = O(\epsilon^2).$$

Proof. First note that by replacing W by $\sigma^{-1}W$ and W_ϵ by $\sigma^{-1}W_\epsilon$, it suffices to consider the case $\sigma = 1$. To prove the theorem, it is enough to show that

$$\left|f(W) - f(Z)\right| \le \frac{2}{\lambda}\mathbb{E}|E|$$

for all smooth, compactly supported f with $\|f\|_\infty \le 1$, since such f are dense in the class of continuous functions vanishing at infinity and bounded by 1, with respect to the supremum norm. Fix an $f : \mathbb{R} \to \mathbb{R}$ in this class and let g be the solution given in equation 1.2 to the differential equation

$$g'(x) - xg(x) = f(x) - \mathbb{E}f(Z).$$

Fix ϵ. By the exchangeability of (W, W_ϵ),

$$\begin{aligned} 0 &= \mathbb{E}\left[(W_\epsilon - W)(g(W_\epsilon) + g(W))\right] \\ &= \mathbb{E}\left[(W_\epsilon - W)(g(W_\epsilon) - g(W)) + 2(W_\epsilon - W)g(W)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[(W_\epsilon - W)^2\middle|W\right]g'(W) + 2\mathbb{E}\left[(W_\epsilon - W)\middle|W\right]g(W) + R\right], \end{aligned} \tag{2.1}$$

where

$$R = (W_\epsilon - W)\left[g(W_\epsilon) - g(W) - (W_\epsilon - W)g'(W)\right]. \tag{2.2}$$

By Taylor's theorem,

$$g(W_\epsilon) = g(W) + (W_\epsilon - W)g'(W) + \frac{1}{2}(W_\epsilon - W)^2 g''(c),$$

where c is a point between W and W_ϵ. Thus

$$g(W_\epsilon) - g(W) - (W_\epsilon - W)g'(W) = \frac{1}{2}(W_\epsilon - W)^2 g''(c). \tag{2.3}$$

Combining (2.2) and (2.3) with Lemma 1.3 (iii), it follows that

$$|R| \leq \frac{\|g''\|_\infty}{2}|W_\epsilon - W|^3 \leq \|f'\|_\infty |W_\epsilon - W|^3,$$

and by the assumption on $|W_\epsilon - W|^3$

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon^2}\mathbb{E}|R| \leq \|f'\|_\infty \lim_{\epsilon \to 0} \frac{1}{\epsilon^2}\mathbb{E}|W_\epsilon - W|^3 = 0. \tag{2.4}$$

Inserting the assumptions of conditions (i) and (ii) of the statement of the theorem into (2.1) and dividing both sides by $2\lambda\epsilon^2$ gives:

$$\begin{aligned}
0 &= \mathbb{E}\left[g'(W)\left(1 + \frac{E}{2\lambda} + \frac{k(\epsilon)}{\epsilon^2}\beta(W)\right) - g(W)\left(W - \frac{h(\epsilon)}{\epsilon^2}\alpha(W)\right) + \frac{R}{2\lambda\epsilon^2}\right] \\
&= \mathbb{E}\left[g'(W) - Wg(W) + g'(W)\left(\frac{E}{2\lambda} + \frac{k(\epsilon)}{\epsilon^2}\beta(W)\right) + g(W)\left(\frac{h(\epsilon)}{\epsilon^2}\alpha(W)\right) + \frac{R}{2\lambda\epsilon^2}\right] \\
&= \mathbb{E}\left[f(W) - f(Z) + g'(W)\left(\frac{E}{2\lambda} + \frac{k(\epsilon)}{\epsilon^2}\beta(W)\right) + g(W)\left(\frac{h(\epsilon)}{\epsilon^2}\alpha(W)\right) + \frac{R}{2\lambda\epsilon^2}\right].
\end{aligned}$$

Taking the limit of both sides as $\epsilon \to 0$ yields

$$\begin{aligned}
0 &= \mathbb{E}\big[f(W) - f(Z)\big] + \mathbb{E}\left[g'(W)\frac{E}{2\lambda}\right] \\
&\quad + \lim_{\epsilon \to 0} \frac{k(\epsilon)}{\epsilon^2}\mathbb{E}\big[g'(W)\beta(W)\big] + \lim_{\epsilon \to 0} \frac{h(\epsilon)}{\epsilon^2}\mathbb{E}\big[g(W)\alpha(W)\big] + \lim_{\epsilon \to 0} \frac{1}{2\lambda\epsilon^2}\mathbb{E}R
\end{aligned} \tag{2.5}$$

We have already seen that

$$\lim_{\epsilon \to 0} \frac{1}{2\lambda\epsilon^2}\mathbb{E}R = 0.$$

Next,

$$\big|\mathbb{E}g'(W)\beta(W)\big| \leq \|g'\|_\infty \mathbb{E}|\beta(W)| < \infty,$$

so

$$\lim_{\epsilon\to 0}\frac{k(\epsilon)}{\epsilon^2}\mathbb{E}\big[g'(W)\beta(W)\big]=0$$

by the assumtption on k. Similarly,

$$\big|\mathbb{E}g(W)\alpha(W)\big|\leq\|g\|_\infty\mathbb{E}|\alpha(W)|<\infty,$$

so

$$\lim_{\epsilon\to 0}\frac{h(\epsilon)}{\epsilon^2}\mathbb{E}\big[g(W)\alpha(W)\big].$$

It thus follows from (2.5) that

$$\big|\mathbb{E}\big[f(W)-f(Z)\big]\big|=\frac{1}{2\lambda}\big|\mathbb{E}\big(g'(W)E\big)\big|.$$

Applying the bound on $\|g'\|_\infty$ from Lemma 1.3 to this last equation completes the proof.

□

2.1. Spherically symmetric distributions on $\mathbb{R}^n$

This section contains a first example of an application of Theorem 2.1.

THEOREM 2.2. *Let X be a random vector with $\mathbb{E}|X|^3<\infty$, and suppose that X has a spherically symmetric distribution. Assume that $\mathbb{E}X_iX_j=\delta_{ij}$. For $\theta\in S^{n-1}$, define $W_\theta=\langle X,\theta\rangle$. Then for $Z\sim\mathfrak{N}(0,1)$,*

$$\begin{aligned}d_{TV}\big(W_\theta,Z\big)&\leq 2\mathbb{E}\big|1-\mathbb{E}[X_2^2|X_1]\big|\\&\leq\frac{2}{n-1}\sqrt{\mathrm{Var}\left(\|X\|_2^2\right)}+\frac{4}{n-1}.\end{aligned}$$

Note that $\mathbb{E}X_i=0$ for all i and $\mathbb{E}X_iX_j=c\delta_{ij}$, where c is independent of i and j, by spherical symmetry. The condition $\mathbb{E}X_iX_j=\delta_{ij}$ is thus only a scaling of X.

This result is closely connected with the following result from [**36**].

THEOREM 2.3 (Diaconis-Freedman). *Let Z be a standard Gaussian random variable and let $\mathbb{P}_\sigma$ be the law of σZ. For a probability μ on $[0,\infty)$, define $\mathbb{P}_\mu$ by*

$$\mathbb{P}_\mu=\int\mathbb{P}_\sigma d\mu(\sigma).$$

Let $X \in \mathbb{R}^n$ be an orthogonally invariant random vector, and let $\mathbb{P}$ be the law of X_1. Then there is a probability μ on $[0, \infty)$ such that

$$d_{TV}(\mathbb{P}, \mathbb{P}_\mu) \leq \frac{4}{n-4}.$$

Theorem 2.3 says that the first component of an orthogonally invariant random vector is close to a mixture of Gaussian random variables. Theorem 2.2 above gives a sufficient condition for when the mixing measure μ can be taken to be a point mass.

We next give some comparison between the two bounds of Theorem 2.2. The first line is sharp enough to reprove the following theorem, the two-dimensional version of which is the classical Herschel-Maxwell theorem (see page 1 of [**19**] for the Herschel-Maxwell theorem and pointers to various related results throughout that book).

Corollary 2.4. *Let X be a random vector with $\mathbb{E}|X|^3 < \infty$ whose distribution is spherically symmetric. Assume that $\mathbb{E}X_i^2 = 1$ for all $1 \leq i \leq n$. If X_i and X_j are independent for $i \neq j$, then X is distributed as a standard Gaussian random vector.*

Proof. If the components of X are independent, then

$$1 - \mathbb{E}\left[X_2^2 \middle| X_1\right] = 0;$$

the result now follows immediately from the first line of the conclusion of Theorem 2.2. □

The proof of Corollary 2.4 shows that the first bound of Theorem 2.2 is sometimes significantly better than the second. However, the following example shows that there may be little loss in using the second line, which is likely to be much easier to compute. Let X be a uniform random point on $\sqrt{n}S^{n-1}$. Then the second statement of the theorem says that

$$d_{TV}(W_\theta, Z) \leq \frac{4}{n-1}.$$

Making use of the first line instead yields the following.

$$\begin{aligned}1 - \mathbb{E}\left[X_2^2\middle|X_1\right] &= 1 - \frac{1}{n-1}\mathbb{E}\left[\sum_{i=2}^n X_i^2\middle| X_1\right]\\ &= 1 - \frac{1}{n-1}\left[n - X_1^2\right]\\ &= \frac{1}{n-1}\big(X_1^2 - 1\big).\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{E}\left|1 - \mathbb{E}\left[X_2^2\middle|X_1\right]\right| &= \frac{1}{n-1}\mathbb{E}\left|X_1^2 - 1\right|\\ &\le \frac{1}{n-1}\sqrt{\mathbb{E}\big[X_1^4 - 2X_1^2 + 1\big]}.\end{aligned} \tag{2.6}$$

By the formula from Theorem 1.6, $\mathbb{E}X_1^4 = \frac{3n}{n+2}$; combining this with (2.6) and Theorem 2.2 gives

$$d_{TV}(W_\theta, Z) \le \frac{2\sqrt{2}}{n-1}.$$

Thus in this example, using the first bound instead of the second resulted only in a minor improvement in the constant.

Proof of Theorem 2.2. By the spherical symmetry of the distribution of X, it suffices to prove the theorem in the case $\theta = e_1$, so $W = X_1$. Note that the normalization is such that $\mathbb{E}W^2 = 1$, thus $\sigma = 1$.

To apply Theorem 2.1, construct a family of random variables $\{W_\epsilon\}$, for $\epsilon \in (0, \frac{1}{2})$ as follows. Let A_ϵ be the $n \times n$ orthogonal matrix

$$A_\epsilon = \begin{pmatrix}\sqrt{1-\epsilon^2} & \epsilon\\ -\epsilon & \sqrt{1-\epsilon^2}\end{pmatrix} \oplus I_{n-2}$$

and let U be a random $n \times n$ orthogonal matrix, chosen independently of X according to Haar measure; $U^T A_\epsilon U$ is a rotation in a random two-dimensional subspace through an angle $\sin^{-1}(\epsilon)$. Define

$$W_\epsilon = \left\langle (U^T A_\epsilon U)X, e_1\right\rangle.$$

By the rotational invariance of X, (W, W_ϵ) is an exchangeable pair for each ϵ.

The conditions of Theorem 2.1 are satisfied as follows. (All assertions are proved below.)

$$\mathbb{E}[W_\epsilon - W|W] = -\big(1 + O(\epsilon^2)\big)\frac{\epsilon^2}{n}W, \tag{2.7}$$

$$\mathbb{E}[(W_\epsilon - W)^2|W] = \big(1 + O(\epsilon)\big)\frac{2\epsilon^2}{n}\mathbb{E}[X_2^2|W] + O(\epsilon^3), \tag{2.8}$$

$$\mathbb{E}|W_\epsilon - W|^3 = O(\epsilon^3). \tag{2.9}$$

Here and throughout this proof the O notation refers to asymptotic behavior as $\epsilon \to 0$, with deterministic implied constants (that may depend on n, f, and the distribution of X).

The first statement says in particular that $\lambda = \frac{1}{n}$. It follows immediately from the second statement that

$$E = \frac{2}{n}\mathbb{E}\left[X_2^2 - 1\middle|W\right].$$

This proves the first bound in the theorem.

For the second bound, observe that by spherical symmetry,

$$\mathbb{E}\big[X_2^2\big|X_1\big] = \frac{1}{n-1}\left(\mathbb{E}\big[\|X\|_2^2\big|X_1\big] - X_1^2\right),$$

and therefore

$$\begin{aligned}\mathbb{E}\big|1 - \mathbb{E}[X_2^2|X_1]\big| &\le \frac{1}{n-1}\Big(\mathbb{E}\big|n - \mathbb{E}\big[\|X\|_2^2\big|X_1\big]\big| + \mathbb{E}\big|X_1^2 - 1\big|\Big)\\ &\le \frac{1}{n-1}\sqrt{\mathrm{Var}\left(\|X\|_2^2\right)} + \frac{2}{n-1}\end{aligned}$$

by the Cauchy-Schwarz inequality and the normalization $\mathbb{E}X_1^2 = 1$.

It remains to show (2.7), (2.8) and (2.9). To prove (2.7), first observe that

$$A_\epsilon = I_n + \left[\epsilon C_2 - \big(1 + O(\epsilon^2)\big)\frac{\epsilon^2}{2}I_2\right] \oplus 0_{n-2},$$

where 0_{n-2} denotes the $(n-2)\times(n-2)$ zero matrix and

$$C_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Denote by K the $2 \times n$ matrix consisting of the first two rows of the random orthogonal matrix U. Then

$$(U^T A_\epsilon U)X = X + K^T \left[\epsilon C_2 - \big(1 + O(\epsilon^2)\big)\frac{\epsilon^2}{2} I_2 \right] KX,$$

and so

$$W - W_\epsilon = -\epsilon \left\langle (K^T C_2 K)X, e_1 \right\rangle + \big(1 + O(\epsilon^2)\big)\frac{\epsilon^2}{2} \left\langle (K^T K)X, e_1 \right\rangle. \tag{2.10}$$

Now if u_{ij} denotes the entries of U, then by expanding in components,

$$\mathbb{E}\left[\left\langle (K^T C_2 K)X, e_1 \right\rangle \middle| X \right] = \sum_{i=2}^n X_i \mathbb{E}(u_{11}u_{2i} - u_{21}u_{1i})$$

and

$$\mathbb{E}\left[\left\langle (K^T K)X, e_1 \right\rangle \middle| X \right] = \sum_{i=1}^n X_i \mathbb{E}(u_{11}u_{1i} + u_{21}u_{2i}).$$

Computing these expectations is not difficult because the distribution of U is unchanged by multiplying any row or column by -1, and any row or column of U is distributed uniformly on S^{n-1}. Therefore

$$\mathbb{E} u_{ij}u_{kl} = \delta_{ik}\delta_{jl}\frac{1}{n}, \tag{2.11}$$

and so

$$\mathbb{E}\left[\left\langle (K^T C_2 K)X, e_1 \right\rangle \middle| X \right] = 0$$

and

$$\mathbb{E}\left[\left\langle (K^T K)X, e_1 \right\rangle \middle| X \right] = \frac{2}{n} W.$$

Putting these together proves (2.7).

The proofs of (2.8) and (2.9) follow similarly from (2.10). For (2.8),

$$\begin{aligned}
\mathbb{E}\left[(W_\epsilon - W)^2 \middle| X \right] &= \epsilon^2 \mathbb{E}\left[\left\langle (K^T C_2 K)X, e_1 \right\rangle^2 \middle| X \right] + O(\epsilon^3) \\
&= \epsilon^2 \mathbb{E}\left[\sum_{i,j=2}^n X_i X_j (u_{11}^2 u_{2i}u_{2j} - u_{11}u_{21}u_{2i}u_{1j} - u_{21}u_{1i}u_{11}u_{2j} + u_{21}^2 u_{1i}u_{1j}) \middle| X \right] \\
&= \frac{2}{n(n-1)} \sum_{i=2}^n X_i^2,
\end{aligned}$$

where the expectations are evaluating using the formulae in section 3.1. Equation 2.8 now follows by noting that since X is spherically symmetric, $\mathbb{E}\left[X_i^2 \middle| X_1\right] = \mathbb{E}\left[X_2^2 \middle| X_1\right]$ for each $i \neq 1$.

Finally, (2.9) follows immediately from (2.10) and the fact that $\mathbb{E}|X|^3 < \infty$.

$\square$

CHAPTER 3

Linear functions on the classical matrix groups

Let $\mathcal{O}_n$ denote the group of $n \times n$ orthogonal matrices, and let M be distributed according to Haar measure on $\mathcal{O}_n$. Let A be a fixed $n \times n$ matrix over $\mathbb{R}$, subject to the condition that $\mathrm{Tr}\,(AA^t) = n$, and let $W = \mathrm{Tr}\,(AM)$. D'Aristotile, Diaconis, and Newman showed in [**29**] that

$$\sup_{\substack{\mathrm{Tr}\,(AA^t)=n \\ -\infty<x<\infty}} |\mathbb{P}(W \leq x) - \Phi(x)| \to 0$$

as $n \to \infty$. Their argument uses classical methods involving sub-subsequences and tightness, and cannot be improved to yield a theorem for finite n. Theorem 3.1 below is an application of Theorem 2.1 which gives an explicit rate of convergence of the law of W to standard normal in the total variation metric on probability measures, specifically,

$$d(\mathcal{L}_W, \mathfrak{N}(0,1))_{TV} \leq \frac{2\sqrt{3}}{n-1} \tag{3.1}$$

for all $n \geq 2$.

The history of this problem begins with the following theorem, first given rigorous proof by Borel in [**17**]: let X be a random vector on the unit sphere S^{n-1}, and let X_1 be the first coordinate of X. Then $\mathbb{P}\big(\sqrt{n}X_1 \leq t\big) \longrightarrow \Phi(t)$ as $n \to \infty$, where $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}}\,dx$. Since the first column of a Haar-distributed orthogonal matrix is uniformly distributed on the unit sphere, Borel's theorem follows from Theorem 3.1 by taking $A = \sqrt{n} \oplus \mathbf{0}$. Borel's theorem was generalized in one direction by Diaconis and Freedman [**36**], who proved the convergence of the first k coordinates of $\sqrt{n}X$ to independent standard normals in total variation distance for $k = o(n)$; [**36**] also contains a detailed history of this problem. This line of research was further developed in [**34**], where a total variation bound was given between an $r \times r$ block of a random orthogonal matrix and an $r \times r$ matrix of independent standard normals, for $r = O\left(n^{1/3}\right)$. This was later improved by Jiang (see [**55**]) to $r = O\left(n^{1/2}\right)$, which he proved was sharp. In the same paper, Jiang also showed that given a sequence of Haar distributed random matrices $\{M_n\}$, there is a sequence of Gaussian matrices $\{Y_n\}$ with Y_j defined on the same probability space as M_j such that if

$$\epsilon_n = \max_{\substack{1\le i\le n \\ 1\le j\le m_n}} \left|\sqrt{n}M_{ij} - Y_{ij}\right|$$

with $m_n \le \frac{n}{\log^2 n}$, then $\epsilon_n \to 0$ in probability as $n \to \infty$. Thus an $n \times \frac{n}{\log^2 n}$ block of a Haar distributed matrix can be approximated by a Gaussian matrix 'in probability'. Theorem 3.1 gives another sense in which a random orthogonal matrix is close to a matrix of independent Gaussians by giving a uniform bound of distance to normal over all linear combinations of entries of M.

Another special case of theorem 3.1 is $A = I$, so that $W = \mathrm{Tr}\,(M)$. Diaconis and Mallows (see [**31**]) first proved that $\mathrm{Tr}\,(M)$ is approximately normal; Stein [**80**] and Johansson [**56**] later independently obtained fast rates of convergence to normal of $\mathrm{Tr}\,(M^k)$ for fixed k, with Johansson's rates an improvement on Stein's. In studying eigenvalues of random orthogonal matrices, Diaconis and Shahshahani [**38**] extended this to show that the joint limiting distribution of $\mathrm{Tr}\,(M), \mathrm{Tr}\,(M^2), \ldots, \mathrm{Tr}\,(M^k)$ converges to that of independent normal variables as $n \to \infty$, for k fixed.

The other source of motivation for theorems like Theorem 3.1 is Hoeffding's combinatorial central limit theorem [**51**], which can be stated as follows. Let $A = (a_{ij})$ be a fixed $n \times n$ matrix over $\mathbb{R}$, normalized to have row and column sums equal to zero and $\frac{1}{n-1}\sum_{i,j} a_{ij}^2 = 1$. Let π be a random permutation in S_n, and let $W(\pi) = \sum_i a_{i\pi(i)}$. Then under certain conditions on A, W is approximately normal. Later, Bolthausen [**16**] proved an explicit rate of convergence via Stein's method, which was extended by Schneller [**75**] to give Edgeworth corrections. Bolthausen's work was extended in another direction by Bloznelis and Götze (see [**15**], e.g), who considered the case of vector-valued $a_{i\pi(i)}$. Zhao, Bai, Chao, and Liang [**84**] also used an approach similar to Bolthausen's in considering statistics of the form

$$\sum_{1 \le i,j \le n} \zeta(i, j, \pi(i), \pi(j)),$$

with applications to limits of Spearman's rho and Kendall's tau. Note that if

$$M_{ij} = \begin{cases} 1 & \pi(j) = i \\ 0 & \text{otherwise} \end{cases}$$

then $W = \mathrm{Tr}\,(AM)$, and so Hoeffding's theorem is really a theorem about the distribution of linear functions on the set of permutation matrices.

The unitary group is another source of many important applications; see, e.g. [**32**]. In section 3.2, the random variable $\mathrm{Tr}\,(AM)$ for A a fixed matrix over $\mathbb{C}$ and M a random unitary matrix distributed according to Haar measure on $\mathcal{U}_n$ is considered. The main theorem of the section, Theorem 3.4 gives a bound on the total variation distance of $Re\big[\,\mathrm{Tr}\,(AM)\big]$ to standard normal analogous to that of Theorem 3.1; this can be viewed as theorem about real-linear functions on $\mathcal{U}_n$. Corollary 3.6 shows that in the limit, the complex random variable $\mathrm{Tr}\,(AM)$ is close to standard complex normal.

A more natural approach might be to consider the distance between $\mathrm{Tr}\,(AM)$ to standard complex normal directly. However, as this is a multivariate problem, Theorem 2.1 cannot be applied. Chapter 5 contains a multivariate version of Theorem 2.1, and section 5.3 applies this theorem to the complex random variable $\mathrm{Tr}\,(AM)$.

3.1. The Orthogonal Group

This section is mainly devoted to the proof of the following theorem.

THEOREM 3.1. *Let A be a fixed $n \times n$ matrix over $\mathbb{R}$ such that $\mathrm{Tr}\,(AA^t) = n$, $M \in \mathcal{O}_n$ distributed according to Haar measure, and $W = \mathrm{Tr}\,(AM)$. Let Z be a standard normal random variable. Then for $n > 1$,*

$$d_{TV}(W, Z) \le \frac{2\sqrt{3}}{n-1}. \tag{3.2}$$

The bound in theorem 3.1 is sharp up to the constant; consider the matrix $A = \sqrt{n}\oplus\mathbf{0}$ where $\mathbf{0}$ is the $n-1\times n-1$ matrix with all zeros. For this A, Theorem 3.1 reproves the following theorem, proved in [**36**] with slightly worse constant

THEOREM 3.2. *Let $\mathbf{x} \in \sqrt{n}S^{n-1}$ be uniformly distributed, and let Z be a standard normal random variable. Then*

$$d_{TV}(x_1, Z) \leq \frac{2\sqrt{3}}{n-1}.$$

It is argued in [**36**], remark (2.12), that the order of this error term is correct.

Before beginning the proof of Theorem 3.1, we give the following lemma which will be used frequently both in this section and in chapter 5.

LEMMA 3.3. *Let $H \in \mathcal{O}_n$ be distributed according to Haar measure. Then $\mathbb{E}\big[h_{ij}h_{kl}h_{mn}h_{pq}\big]$ is only non-zero if there are an even number of entries from each row and each column. Fourth-order mixed moments are as follows.*

(i) $\mathbb{E}\left[h_{11}^2 h_{12}^2\right] = \dfrac{1}{n(n+2)}$,

(ii) $\mathbb{E}\left[h_{11}^2 h_{22}^2\right] = \dfrac{n+1}{(n-1)n(n+2)}$,

(iii) $\mathbb{E}\left[h_{11}h_{12}h_{21}h_{22}\right] = \dfrac{-1}{(n-1)n(n+2)}$, *and*

(iv) $\mathbb{E}\big[(h_{i1}h_{i'2} - h_{i2}h_{i'1})(h_{j1}h_{j'2} - h_{j2}h_{j'1})\big] = \dfrac{2}{n(n-1)}\Big[\delta_{ij}\delta_{i'j'} - \delta_{ij'}\delta_{i'j}\Big]$, *where $i \neq i'$ and $j \neq j'$.*

PROOF. Recall that the first row of H is distributed uniformly on the unit sphere of $\mathbb{R}^n$. Part (i) thus follows from Theorem 1.6, after renormalizing to go from the sphere of radius $\sqrt{n}$ to the sphere of radius one.

Part (ii) is straightforward using the symmetries of Haar measure and the first equation:

$$\begin{aligned}
\mathbb{E}\left[h_{11}^2 h_{22}^2\right] &= \frac{1}{n-1}\mathbb{E}\left[h_{11}^2\left(\sum_{k>1} h_{k2}^2\right)\right] \\
&= \frac{1}{n-1}\mathbb{E}\left[h_{11}^2(1-h_{12}^2)\right] \\
&= \frac{1}{n-1}\left[\frac{1}{n} - \frac{1}{n(n+2)}\right]
\end{aligned}$$

$$= \frac{n+1}{(n-1)n(n+2)}.$$

For part (iii),

$$\begin{aligned}
\mathbb{E}\left[h_{11}h_{12}h_{21}h_{22}\right] &= \frac{1}{n(n-1)}\mathbb{E}\left[\sum_{\substack{i\\ j\neq i}} h_{i1}h_{i2}h_{j1}j_{j2}\right] \\
&= \frac{1}{n(n-1)}\mathbb{E}\left[\left(\sum_i h_{i1}h_{i2}\right)^2 - \sum_i h_{i1}^2 h_{i2}^2\right] \\
&= \frac{1}{n(n-1)}\mathbb{E}\left[(h_{i1})\cdot(h_{i2})\right] - \frac{1}{n-1}\mathbb{E}\left[h_{11}^2 h_{12}^2\right] \\
&= \frac{-1}{(n-1)n(n+2)},
\end{aligned}$$

where the orthogonality of the columns of H has been used to get the last line. Finally, part (iv) follows from parts (ii) and (iii):

$$\begin{aligned}
&\mathbb{E}\big[(h_{i1}h_{i'2} - h_{i2}h_{i'1})(h_{j1}h_{j'2} - h_{j2}h_{j'1})\big] \\
&\quad = \mathbb{E}\big[h_{i1}h_{i'2}h_{j1}h_{j'2} + h_{i2}h_{i'1}h_{j2}h_{j'1} - h_{i2}h_{i'1}h_{j1}h_{j'2} - h_{i1}h_{i'2}h_{j2}h_{j'1}\big] \\
&\quad = 2\left[\delta ij\delta_{i'j'}\left(\frac{n+1}{(n-1)n(n+2)}\right) - \delta_{ij'}\delta_{i'j}\left(\frac{1}{(n-1)n(n+2)}\right)\right] \\
&\qquad - 2\left[\delta ij'\delta_{i'j}\left(\frac{n+1}{(n-1)n(n+2)}\right) - \delta_{ij}\delta_{i'j'}\left(\frac{1}{(n-1)n(n+2)}\right)\right] \\
&\quad = \frac{2}{(n-1)n(n+2)}\left[(n+2)\delta_{ij}\delta_{i'j'} - (n+2)\delta_{ij'}\delta_{i'j}\right] \\
&\quad = \frac{2}{n(n-1)}\left[\delta_{ij}\delta_{i'j'} - \delta_{ij'}\delta_{i'j}\right].
\end{aligned}$$

$\square$

Proof of theorem 3.1. First note that one can assume without loss of generality that A is diagonal: let $A = UDV$ be the singular value decomposition of A. Then $W = \mathrm{Tr}\,(UDVM) = \mathrm{Tr}\,(DVMU)$, and the distribution of VMU is the same as the distribution of M by the translation invariance of Haar measure.

Now define the pair (W, W_ϵ) for each ϵ as follows. Choose $H = (h_{ij}) \in \mathcal{O}(n)$ according to Haar measure, independent of M, and let $M_\epsilon = HA_\epsilon H^t M$, where

$$A_\epsilon = \begin{bmatrix} \sqrt{1-\epsilon^2} & \epsilon & & & \\ -\epsilon & \sqrt{1-\epsilon^2} & & 0 & \\ & & 1 & & \\ & 0 & & \ddots & \\ & & & & 1 \end{bmatrix},$$

thus M_ϵ can be thought of as a small random rotation of M. Let $W_\epsilon = W(M_\epsilon)$; (W, W_ϵ) is an exchangeable pair by construction.

As in the example of chapter 2, M_ϵ can be written as a perturbation of M: let I_2 be the 2×2 identity matrix, K the $n \times 2$ matrix made from the first two columns of H, and let

$$C_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Then

$$M_\epsilon = M + K\big[(\sqrt{1-\epsilon^2} - 1)I_2 + \epsilon C_2\big]K^t M.$$

By Taylor's theorem, $\sqrt{1-\epsilon^2} = 1 - \frac{\epsilon^2}{2} + O(\epsilon^4)$, thus

$$M_\epsilon = M + K\left[\left(-\frac{\epsilon^2}{2} + O(\epsilon^4)\right) I_2 + \epsilon C_2\right] K^t M.$$

This yields

$$\begin{aligned} W_\epsilon - W &= \operatorname{Tr}\left(\left(-\frac{\epsilon^2}{2} + O(\epsilon^4)\right) AKK^tM + \epsilon AKC_2K^tM\right) \\ &= \epsilon\left[\left(-\frac{\epsilon}{2} + O(\epsilon^3)\right) \operatorname{Tr}(AKK^tM) + \operatorname{Tr}(AKC_2K^tM)\right]. \end{aligned} \tag{3.3}$$

It is necessary to expand KK^t and KC_2K^t in components to compute the necessary expectations. If u_{ij} denotes the i-jth entry of U, then

$$\begin{aligned} (KK^t)_{ij} &= u_{i1}u_{j1} + u_{i2}u_{j2} \\ (KC_2K^t)_{ij} &= u_{i1}u_{j2} - u_{i2}u_{j1}. \end{aligned}$$

Recall equation (2.11):

$$\mathbb{E}u_{ij}u_{k\ell} = \frac{1}{n}\delta_{ik}\delta_{j\ell}.$$

It follows that

$$\begin{aligned} \mathbb{E}\left[KK^t\right] &= \frac{2}{n}I_n \\ \mathbb{E}\left[KC_2K^t\right] &= 0_n. \end{aligned} \tag{3.4}$$

Combining these equations with (3.3) yields

$$\begin{aligned} &\frac{n}{\epsilon^2}\mathbb{E}\big[(W_\epsilon - W)\big|W\big] \\ &\qquad = -\frac{n}{2}\mathbb{E}\left[\mathbb{E}\left[\mathrm{Tr}\,(AKK^tM)\big|M\right]\big|W\right] + \frac{n}{\epsilon}\mathbb{E}\left[\mathbb{E}\left[\mathrm{Tr}\,(AKC_2K^tM)\big|M\right]\big|W\right] + O(\epsilon) \\ &\qquad = -\frac{n}{2}\mathbb{E}\left[\mathrm{Tr}\,\left(A\left(\frac{2}{n}I_n\right)M\right)\bigg|W\right] + \frac{n}{\epsilon}\mathbb{E}\left[\mathrm{Tr}\,\big(A\big(0_n\big)M\big)\big|W\right] + O(\epsilon) \\ &\qquad = -\mathbb{E}\left[\mathbb{E}\left[\mathrm{Tr}\,(AM)\big|M\right]\big|W\right] + O(\epsilon) \\ &\qquad = -W + O(\epsilon), \end{aligned}$$

where the independence of M and H has been used to get the second line. In this computation and in what follows, the implied constants in the $O(\epsilon)$ may depend on A and n. It now follows that

$$\lim_{\epsilon\to 0}\frac{n}{\epsilon^2}\mathbb{E}\left[(W_\epsilon - W)\big|W\right] = -W, \tag{3.5}$$

so the first condition of Theorem 2.1 is satisfied with $\lambda = \frac{1}{n}$.

We next determine the quantity E from the second condition of Theorem 2.1. Recall that A is assumed to be diagonal. By 3.3,

$$\begin{aligned} \mathbb{E}\left[(W_\epsilon - W)^2\big|W\right] = \epsilon^2\mathbb{E}&\left[(\mathrm{Tr}\,(AKC_2K^tM))^2\big|W\right] \\ &+ 2\left(-\frac{\epsilon^3}{2} + O(\epsilon^5)\right)\mathbb{E}\left[\mathrm{Tr}\,(AKK^tM)\,\mathrm{Tr}\,(AKC_2K^2M)\big|W\right] \\ &+ \left(-\frac{\epsilon^2}{2} + O(\epsilon^4)\right)^2\mathbb{E}\left[\big(\mathrm{Tr}\,(AKK^tM)\big)^2\big|W\right]. \end{aligned}$$

Only the first term has quadratic factors of ϵ; the other two terms are of higher order in ϵ. This observation means that in the following expression, the other two terms can and have been absorbed into the $O(\epsilon)$.

(3.6)

$$
\begin{aligned}
&\frac{n}{2\epsilon^2}\mathbb{E}\left[(W_\epsilon - W)^2 \middle| W\right] \\
&\quad= \frac{n}{2}\mathbb{E}\left[\mathbb{E}\left[(\mathrm{Tr}\,(AKC_2K^tM))^2\middle|M\right]\middle|W\right] + O(\epsilon) \\
&\quad= \frac{n}{2}\mathbb{E}\left[\mathbb{E}\left[\left(\sum_{i,i'}(AKC_2K^t)_{ii'}m_{i'i}\right)\left(\sum_{j,j'}(AKC_2K^t)_{jj'}m_{j'j}\right)\middle|M\right]\middle|W\right] + O(\epsilon) \\
&\quad= \frac{n}{2}\mathbb{E}\left[\sum_{i,i',j,j'} m_{i'i}m_{j'j}\mathbb{E}\left[(AKC_2K^t)_{ii'}(AKC_2K^t)_{jj'}\middle|M\right]\middle|W\right] + O(\epsilon) \\
&\quad= \frac{n}{2}\mathbb{E}\left[\sum_{i,j}\sum_{\substack{i'\neq i\\ j'\neq j}} m_{i'i}m_{j'j}a_{ii}a_{jj}\mathbb{E}\left[(h_{i1}h_{i'2}-h_{i2}h_{i'1})(h_{j1}h_{j'2}-h_{j2}h_{j'1})\middle|M\right]\middle|W\right] + O(\epsilon),
\end{aligned}
$$

where the conditions on i' and j' are justified as the expression inside the expectation is identically 0 when either $i = i'$ or $j = j'$.

Recall that M and H are independent, thus the expectation above can be evaluated using Lemma 3.3 (iv). This yields

$$
\begin{aligned}
\frac{n}{2\epsilon^2}\mathbb{E}\left[(W_\epsilon - W)^2\middle|W\right] &= \frac{1}{n-1}\sum_{i,j}\sum_{\substack{i'\neq i\\ j'\neq j}} m_{i'i}m_{j'j}a_{ii}a_{jj}\left[\delta_{i'j'}\delta_{ij} - \delta_{ij'}\delta_{ji'}\right] + O(\epsilon) \\
&= \frac{1}{n-1}\left[\sum_{i,i'\neq i} m_{i'i}^2 a_{ii}^2 - \sum_{i,i'\neq i} m_{i'i}m_{ii'}a_{ii}a_{i'i'}\right] + O(\epsilon) \\
&= \frac{1}{n-1}\left[\sum_i a_{ii}^2\left[(M^tM)_{ii} - m_{ii}^2\right] - \sum_{i,i'\neq i}(MA)_{ii'}(MA)_{i'i}\right] + O(\epsilon) \\
&= \frac{1}{n-1}\left[n - \sum_i a_{ii}^2 m_{ii}^2 - \left[\mathrm{Tr}\,((MA)^2) - \sum_i a_{ii}^2 m_{ii}^2\right]\right] + O(\epsilon) \\
&= 1 + \frac{1}{n-1}\left[1 - \mathrm{Tr}\,((AM)^2)\right] + O(\epsilon),
\end{aligned}
$$

where the normalization $\sum_i a_{ii}^2 = n$ has been used to get the fourth line. Thus

$$
\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\left[(W_\epsilon - W)^2\middle|W\right] = \frac{2}{n} + \frac{2}{n(n-1)}\left[1 - \mathrm{Tr}\,((AM)^2)\right] \tag{3.7}
$$

and so the error E of Theorem 2.1 has the form

$$E = \frac{2}{n(n-1)} \left[1 - \operatorname{Tr}\left((AM)^2\right)\right]. \tag{3.8}$$

Finally, (3.3) gives immediately that

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

It remains to bound $n\mathbb{E}|E|$. First,

$$\begin{aligned} \mathbb{E}\left[\operatorname{Tr}\left((AM)^2\right)\right] &= \mathbb{E}\left[\sum_{i,j} a_{ii}a_{jj}m_{ij}m_{ji}\right] \\ &= \frac{1}{n}\sum_i a_{ii}^2 \\ &= 1. \end{aligned} \tag{3.9}$$

The following computation makes use of the formulae of Lemma 3.3 and the normalization condition on A. Let $\sum'$ stand for summing over distinct indices.

$$\begin{aligned} &\mathbb{E}\left[(\operatorname{Tr}\left((AM)^2\right))^2\right] \\ &\quad = \mathbb{E}\left[\left(\sum_{i,j} a_{ii}a_{jj}m_{ij}m_{ji}\right)\left(\sum_{k,l} a_{kk}a_{ll}m_{kl}m_{lk}\right)\right] \\ &\quad = \sum_{i,j,k,l} a_{ii}a_{jj}a_{kk}a_{ll}\left[\frac{n+1}{(n-1)n(n+2)}\left[\delta_{ij}\delta_{kl}\big(1-\delta_{ik}\big) + \delta_{ik}\delta_{jl}\big(1-\delta_{ij}\big) + \right.\right. \\ &\qquad\qquad \left.\left. \delta_{il}\delta_{jk}\big(1-\delta_{ij}\big)\right] + \frac{3}{n(n+2)}\mathbb{I}(i=j=k=l)\right] \\ &\quad = \frac{n+1}{n(n-1)(n+2)}\left({\sum_{i,k}}' a_{ii}^2a_{kk}^2 + {\sum_{i,j}}' a_{ii}^2a_{jj}^2 + {\sum_{i,j}}' a_{ii}^2a_{jj}^2\right) + \frac{3}{n(n+2)}\sum_i a_{ii}^4 \end{aligned}$$

Now,

$${\sum_{i,j}}' a_{ii}^2a_{jj}^2 = \sum_i a_{ii}^2(n - a_{ii}^2) = n^2 - \sum_i a_{ii}^4.$$

Applying this above gives

$$\begin{aligned}
\mathbb{E}\big[(\mathrm{Tr}\,((AM)^2))^2\big] &= \frac{3(n+1)}{(n-1)n(n+2)}\left(n^2 - \sum_i a_{ii}^4\right) + \frac{3}{n(n+2)}\sum_i a_{ii}^4 \\
&= \frac{3(n+1)n^2}{(n-1)n(n+2)} - \frac{3(n+1)}{(n-1)n(n+2)}\sum_i a_{ii}^4 + \frac{3}{n(n+2)}\sum_i a_{ii}^4 \\
&= 3 + \frac{6}{(n-1)(n+2)} - \left(\frac{6}{(n-1)n(n+2)}\right)\sum_i a_{ii}^4 \\
&\leq 3 + \frac{6}{(n-1)(n+2)}.
\end{aligned}$$

Putting this estimate into Theorem 2.1 gives:

$$d_{TV}(W,Z) \leq \frac{2\sqrt{2+\frac{6}{(n-1)(n+2)}}}{(n-1)}. \tag{3.10}$$

Noting that $\frac{6}{(n-1)(n+2)} \leq 1$ for $n \geq 3$ and that the bound in theorem 3.1 is trivially true for $n = 2$ completes the proof. □

3.2. The Unitary Group

Now let $M \in \mathfrak{U}_n$ be distributed according to Haar measure, A be an $n \times n$ matrix over $\mathbb{C}$, and $W = \mathrm{Tr}\,(AM)$. In [**29**] it was shown that if $M = \Gamma + i\Lambda$ and A and B are fixed real diagonal matrices with $\mathrm{Tr}\,(AA^t) = \mathrm{Tr}\,(BB^t) = n$, then $\mathrm{Tr}\,(A\Gamma) + i\,\mathrm{Tr}\,(B\Lambda)$ converges in distribution to standard complex normal. This implies in particular that $\mathrm{Re}(W)$ converges in distribution to $\mathfrak{N}\left(0, \frac{1}{2}\right)$. The main theorem of this section gives a rate of this convergence in total variation distance. As the convergence of the complex random variable $\mathrm{Tr}\,(AM)$ to a standard complex Gaussian random variable is a bivariate problem, it will be treated in chapter 5.

Theorem 3.4. *With M, A, and W as above, let W_θ be the inner product of W with the unit vector making angle θ with the real axis. Then*

$$d_{TV}\left(W_\theta, \mathfrak{N}\left(0, \frac{1}{2}\right)\right) \leq \frac{c}{n} \tag{3.11}$$

for a constant c which is independent of θ.

The constant c is asymptotically equal to $\sqrt{2}$; for $n \geq 8$ it can be taken to be 2.

Before beginning the proof, we give the following complex analog of Lemma 3.3 for use here and in chapter 5.

LEMMA 3.5. *Let $H \in \mathfrak{U}_n$ be distributed according to Haar measure. Then the expected value of a product of entries of H and their conjugates is non-zero only when there are the same number of entries as conjugates of entries from each row and from each column. The following are formulae for some non-zero expectations.*

(i) $\mathbb{E}\big[|h_{ij}|^2\big] = \frac{1}{n}$,

(ii) $\mathbb{E}\big[|h_{ij}|^4\big] = \frac{2}{n(n+1)}$,

(iii) $\mathbb{E}\big[|h_{ij}|^2|h_{ik}|^2\big] = \frac{1}{n(n+1)}$ *for* $j \neq k$,

(iv) $\mathbb{E}\big[|h_{ij}|^2|h_{k\ell}|^2\big] = \frac{1}{(n-1)(n+1)}$ *for* $i \neq k$ *and* $j \neq \ell$,

(v) $\mathbb{E}\big[h_{ij}\overline{h}_{ik}\overline{h}_{\ell j}h_{\ell k}\big] = -\frac{1}{(n-1)n(n+1)}$,

(vi)

$$\mathbb{E}\big[(h_{i1}\overline{h}_{j2} - h_{i2}\overline{h}_{j1})(h_{k1}\overline{h}_{\ell 2} - h_{k2}\overline{h}_{\ell 1})\big] = -\frac{2\delta_{i\ell}\delta_{jk}(1-\delta_{ij})}{(n-1)(n+1)} + \frac{2\delta_{ij}\delta_{k\ell}(1-\delta_{ik})}{(n-1)n(n+1)} - \frac{2\mathbb{I}(i=j=k=\ell)}{n(n+1)},$$

and

(vii)

$$\mathbb{E}\big[(h_{i1}\overline{h}_{j2} - h_{i2}\overline{h}_{j1})(\overline{h}_{k1}h_{\ell 2} - \overline{h}_{k2}h_{\ell 1})\big] = \frac{2\delta_{ik}\delta_{j\ell}(1-\delta_{ij})}{(n-1)(n+1)} - \frac{2\delta_{ij}\delta_{k\ell}(1-\delta_{ik})}{n(n-1)(n+1)} + \frac{2\mathbb{I}(i=j=k=\ell)}{n(n+1)}.$$

PROOF. The requirement that the same number of entries as conjugates of entries from each row and column be present is a simple consequence of the invariance properties of Haar measure; any row or column is invariant under multiplication by a complex number of unit modulus.

(i) Since $\sum_k |h_{ik}|^2 = 1$, this part is immediate by the invariance of Haar measure under permuting the columns of H.

(ii) This is proved in detail in section 4.2, page 139 of [**50**].

(iii) It suffices to assume that $j = 1$ and $k = 2$. By symmetry,

$$\begin{aligned}
\mathbb{E}\left[|h_{i1}|^2|h_{i2}|^2\right] &= \frac{1}{n-1}\sum_{j\neq 1}\mathbb{E}\left[|h_{i1}|^2|h_{ij}|^2\right] \\
&= \frac{1}{n-1}\mathbb{E}\left[|h_{i1}|^2\left(1-|h_{i1}|^2\right)\right] \\
&= \frac{1}{n-1}\mathbb{E}\left[|h_{i1}|^2-|h_{i1}|^4\right] \\
&= \frac{1}{n(n+1)}.
\end{aligned}$$

(iv) This part is the same as the previous part:

$$\begin{aligned}
\mathbb{E}\left[|h_{i1}|^2|h_{j2}|^2\right] &= \frac{1}{n-1}\sum_{k\neq i}\mathbb{E}\left[|h_{i1}|^2|h_{k2}|^2\right] \\
&= \frac{1}{n-1}\mathbb{E}\left[|h_{i1}|^2\left(1-|h_{i2}|^2\right)\right] \\
&= \frac{1}{(n-1)(n+1)}.
\end{aligned}$$

(v) This part makes use of the complex-orthogonality of the columns of H as well as the invariance under permutations of the rows. It is no loss to assume that $i = j = 1$ and $k = \ell = 2$. We have

$$\begin{aligned}
\mathbb{E}\big[h_{11}\overline{h}_{12}\overline{h}_{21}h_{22}\big] &= \frac{1}{n-1}\sum_{k\neq 1}\mathbb{E}\big[h_{11}\overline{h}_{12}\overline{h}_{k1}h_{k2}\big] \\
&= \frac{1}{n-1}\mathbb{E}\left[\big\langle (h_{k1}),\big(\overline{h}_{k2}\big)\big\rangle - |h_{11}|^2|h_{12}|^2\right] \\
&= -\frac{1}{n-1}\mathbb{E}\left[|h_{11}|^2|h_{12}|^2\right] \\
&= -\frac{1}{(n-1)n(n+2)}.
\end{aligned}$$

(vi) This part is just putting together the previous parts. Multiplying out the product gives

$$\begin{aligned}
&\mathbb{E}\big[(h_{i1}\overline{h}_{j2}-h_{i2}\overline{h}_{j1})(h_{k1}\overline{h}_{\ell 2}-h_{k2}\overline{h}_{\ell 1})\big] \\
&\qquad = \mathbb{E}\big[h_{i1}h_{k1}\overline{h}_{j2}\overline{h}_{\ell 2}+h_{i2}h_{k2}\overline{h}_{j1}\overline{h}_{\ell 1}-h_{i2}h_{k1}\overline{h}_{j1}\overline{h}_{\ell 2}-h_{i1}h_{k2}\overline{h}_{j2}\overline{h}_{\ell 1}\big].
\end{aligned}$$

The first two terms integrate to zero, since H is invariant under multiplying the first column by $\sqrt{-1}$. Thus

$$\mathbb{E}\big[(h_{i1}\overline{h}_{j2} - h_{i2}\overline{h}_{j1})(h_{k1}\overline{h}_{\ell 2} - h_{k2}\overline{h}_{\ell 1})\big] = \mathbb{E}\big[- h_{i2}h_{k1}\overline{h}_{j1}\overline{h}_{\ell 2} - h_{i1}h_{k2}\overline{h}_{j2}\overline{h}_{\ell 1}\big].$$

Each of these terms integrates to something non-zero if and only if either $i = \ell$ and $j = k$ or $i = j$ and $k = \ell$. The second term is the obtained from the first by switching the indices 1 and 2, thus they integrate to the same thing by spherical symmetry. It follows that

$$\begin{aligned}
&\mathbb{E}\big[(h_{i1}\overline{h}_{j2} - h_{i2}\overline{h}_{j1})(h_{k1}\overline{h}_{\ell 2} - h_{k2}\overline{h}_{\ell 1})\big] \\
&\qquad = -2\mathbb{E}\big[\delta_{i\ell}\delta_{jk}(1-\delta_{ij})|h_{i2}|^2|h_{j1}|^2 + \delta_{ij}\delta_{k\ell}(1-\delta_{ik})h_{i2}h_{k1}\overline{h}_{i1}\overline{h}_{k2} \\
&\qquad\qquad + \mathbb{I}(i=j=k=\ell)|h_{i1}|^2|h_{i2}|^2\big].
\end{aligned}$$

Part (vi) now follows from parts (iii), (iv) and (v).

(vii) This part is just like the last part. First,

$$\begin{aligned}
&\mathbb{E}\big[(h_{i1}\overline{h}_{j2} - h_{i2}\overline{h}_{j1})(\overline{h}_{k1}h_{\ell 2} - \overline{h}_{k2}h_{\ell 1})\big] \\
&\qquad = \mathbb{E}\big[h_{i1}h_{\ell 2}\overline{h}_{j2}\overline{h}_{k1} + h_{i2}h_{\ell 1}\overline{h}_{j1}\overline{h}_{k2} - h_{i1}h_{\ell 1}\overline{h}_{j2}\overline{h}_{k2} - h_{i2}h_{\ell 2}\overline{h}_{j1}\overline{h}_{k1}\big].
\end{aligned}$$

Here, the last two terms integrate to zero by symmetry. Thus

$$\mathbb{E}\big[(h_{i1}\overline{h}_{j2} - h_{i2}\overline{h}_{j1})(\overline{h}_{k1}h_{\ell 2} - \overline{h}_{k2}h_{\ell 1})\big] = \mathbb{E}\big[h_{i1}h_{\ell 2}\overline{h}_{j2}\overline{h}_{k1} + h_{i2}h_{\ell 1}\overline{h}_{j1}\overline{h}_{k2}\big].$$

Again, the two integrals are the same by switching the indices 1 and 2, and each is non-zero if and only if $i = k$ and $j = \ell$ or $i = j$ and $k = \ell$. It follows that

$$\begin{aligned}
&\mathbb{E}\big[(h_{i1}\overline{h}_{j2} - h_{i2}\overline{h}_{j1})(\overline{h}_{k1}h_{\ell 2} - \overline{h}_{k2}h_{\ell 1})\big] \\
&\qquad = 2\mathbb{E}\big[\delta_{ik}\delta_{j\ell}(1-\delta_{ij})|h_{i1}|^2|h_{j2}|^2 + \delta_{ij}\delta_{k\ell}(1-\delta_{ik})h_{i1}h_{k2}\overline{h}_{i2}\overline{h}_{k1} \\
&\qquad\qquad + \mathbb{I}(i=j=k=\ell)|h_{i1}|^2|h_{i2}|^2\big];
\end{aligned}$$

the formula of part (vii) follows from parts (iv), (v) and (ii).

□

PROOF OF THEOREM 3.4. To prove the theorem, first note that it suffices to consider the case $\theta = 0$, that is, to prove that

$$d_{TV}\left(\mathrm{Re}(W), \mathfrak{N}\left(0, \frac{1}{2}\right)\right) \le \frac{c}{n};$$

the theorem then follows as stated since the distribution of W is invariant under multiplication by any complex number of unit modulus. By the singular value decomposition (see, e.g., [**14**], page 6), there is an $n \times n$ diagonal matrix S with non-negative entries such that $A = USV$ for two unitary matrices U and V. Then $\mathrm{Tr}\,(AM) = \mathrm{Tr}\,(USVM) = \mathrm{Tr}\,(SVMU)$ and VMU has the same distribution as M by the translation invariance of Haar measure. It follows that A can be assumed to be diagonal with non-negative real entries.

The proof is almost identical to the orthogonal case. Let $H \in \mathfrak{U}_n$ be a random unitary matrix, independent of M, and let $M_\epsilon = HA_\epsilon H^* M$, where

$$A_\epsilon = \begin{bmatrix} \sqrt{1-\epsilon^2} & \epsilon & & & \\ -\epsilon & \sqrt{1-\epsilon^2} & & 0 & \\ & & 1 & & \\ & 0 & & \ddots & \\ & & & & 1 \end{bmatrix}.$$

Let $W_\epsilon = W(M_\epsilon)$.

Let I_2 be the 2×2 identity matrix, K the $n \times 2$ matrix made from the first two columns of H, and let

$$C_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Then

$$\begin{aligned} W_\epsilon - W &= \mathrm{Tr}\left(\left(-\frac{\epsilon^2}{2} + O(\epsilon^4)\right) AKK^*M + \epsilon AKC_2K^*M\right) \\ &= \epsilon\left[\left(-\frac{\epsilon}{2} + O(\epsilon^3)\right) \mathrm{Tr}\,(AKK^*M) + \mathrm{Tr}\,(AKC_2K^*M)\right]. \end{aligned} \tag{3.12}$$

Let $W^r = Re(W)$ and $W^r_\epsilon = Re(W_\epsilon)$. To verify the conditions of Theorem 2.1, first observe that by expanding in components and applying Lemma 3.5,

$$\begin{aligned}
KK^*_{ij} &= h_{i1}\overline{h}_{j1} + h_{i2}\overline{h}_{j2} \\
\mathbb{E}\left[KK^*_{ij}\right] &= \frac{2}{n}\delta_{ij} && (3.13)\\
KC_2K^*_{ij} &= h_{i1}\overline{h}_{j2} - h_{i2}\overline{h}_{j1} \\
\mathbb{E}\left[KC_2K^*_{ij}\right] &= 0. && (3.14)
\end{aligned}$$

Line (3.14) implies that

$$\mathbb{E}\left[\mathrm{Tr}\,(AKC_2K^*M)|M\right] = 0,$$

since M and H are independent. It follows by (3.12) that

$$\lim_{\epsilon\to 0}\frac{n}{\epsilon^2}\mathbb{E}\left[W^r_\epsilon - W^r\middle|W\right] = -W^r; \tag{3.15}$$

condition (1) of Theorem 2.1 is satisfied with $\lambda = \frac{1}{n}$. Also by (3.12),

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}[(W^r_\epsilon - W^r)^2\big|W] = \lim_{\epsilon\to 0}\mathbb{E}\left[(Re(\mathrm{Tr}\,(AKC_2K^*M)))^2\,\middle|W\right]$$

since all the other terms are higher order in ϵ. Expanding this expression gives

(3.16)
$$\begin{aligned}
&\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}[(W^r_\epsilon - W^r)^2\big|W] \\
&\quad= \mathbb{E}\left[\sum_{i,j,k,l} Re\left[a_{ii}m_{ji}(h_{i1}\overline{h}_{j2} - h_{i2}\overline{h}_{j1})\right] Re\left[a_{kk}m_{lk}(h_{k1}\overline{h}_{l2} - h_{k2}\overline{h}_{l1})\right]\middle| W\right] \\
&\quad= \frac{1}{2}Re\,\mathbb{E}\left[\sum_{i,j,k,l} a_{ii}m_{ji}a_{kk}m_{lk}(h_{i1}\overline{h}_{j2} - h_{i2}\overline{h}_{j1})(h_{k1}\overline{h}_{l2} - h_{k2}\overline{h}_{l1}) + \right.\\
&\qquad\qquad \left. a_{ii}m_{ji}a_{kk}\overline{m}_{lk}(h_{i1}\overline{h}_{j2} - h_{i2}\overline{h}_{j1})(\overline{h}_{k1}h_{l2} - \overline{h}_{k2}h_{l1})\middle| W\right]
\end{aligned}$$

Let ${\sum_{i,j}}'$ stand for summing over all pairs (i,j) where i and j are distinct. Applying parts (vi) and (vii) of Lemma 3.5 now gives

$$
\begin{aligned}
&\lim_{\epsilon\to 0}\frac{n}{2\epsilon^2}\mathbb{E}\big[\big(W_\epsilon^r - W^r\big)^2\big|W\big]\\
&\quad= \frac{n}{2(n-1)(n+1)} Re\,\mathbb{E}\Bigg[\sum_{i,j,k,\ell} a_{ii}m_{ji}a_{kk}m_{\ell k}\bigg(-\delta_{i\ell}\delta_{jk}(1-\delta_{ij}) + \frac{1}{n}\delta_{ij}\delta_{k\ell}(1-\delta_{ik})\\
&\qquad\qquad - \left(\frac{n-1}{n}\right)\mathbb{I}(i=j=k=\ell)\bigg)\\
&\qquad\qquad + \sum_{i,j,k,\ell} a_{ii}m_{ji}a_{kk}m_{\ell k}\bigg(\delta_{ik}\delta_{j\ell}(1-\delta_{ij}) - \frac{1}{n}\delta_{ij}\delta_{k\ell}(1-\delta_{ik})\\
&\qquad\qquad + \left(\frac{n-1}{n}\right)\mathbb{I}(i=j=k=l)\bigg)\Bigg|W\Bigg]\\
&\quad= \frac{n}{2(n-1)(n+1)} Re\,\mathbb{E}\Bigg[-{\sum_{i,j}}' a_{ii}a_{jj}m_{ij}m_{ji} + \frac{1}{n}{\sum_{i,k}}' a_{ii}a_{kk}m_{ii}m_{kk}\\
&\qquad\qquad - \left(\frac{n-1}{n}\right)\sum_i a_{ii}^2 m_{ii}^2 + {\sum_{i,j}}' a_{ii}^2|m_{ji}|^2\\
&\qquad\qquad - \frac{1}{n}{\sum_{i,k}}' a_{ii}a_{kk}m_{ii}\overline{m}_{kk} + \frac{n-1}{n}\sum_i a_{ii}^2|m_{ii}|^2\Bigg|W\Bigg]\\
&\quad= \frac{n}{2(n-1)(n+1)} Re\,\mathbb{E}\Bigg[-\left(\mathrm{Tr}\left((AM)^2\right) - \sum_i (AM)_{ii}^2\right) + \frac{1}{n}\left(W^2 - \sum_i (AM)_{ii}^2\right)\\
&\qquad\qquad - \left(\frac{n-1}{n}\right)\sum_i (AM)_{ii}^2 + \sum_i a_{ii}^2(1-|m_{ii}|^2)\\
&\qquad\qquad - \frac{1}{n}\left(|W|^2 - \sum_i a_{ii}^2|m_{ii}|^2\right) + \frac{n-1}{n}\sum_i a_{ii}^2|m_{ii}|^2\Bigg|W\Bigg]\\
&\quad= \frac{1}{2} + \frac{1}{2(n-1)(n+1)}\\
&\qquad\qquad + \frac{n}{2(n-1)(n+1)} Re\,\mathbb{E}\left[-\mathrm{Tr}\left((AM)^2\right) + \frac{W^2-|W|^2}{n}\middle| W\right].
\end{aligned}
$$

Condition (2) of Theorem 2.1 is thus satisfied with

(3.17)

$$
nE = \frac{1}{2(n-1)(n+1)} + \frac{n}{2(n-1)(n+1)} Re\,\mathbb{E}\left[-\mathrm{Tr}\left((AM)^2\right) + \frac{W^2-|W|^2}{n}\middle| W\right],
$$

and it remains to estimate $n\mathbb{E}|E|$. First,

$$\begin{aligned}
\mathbb{E}\left|\mathrm{Tr}\left((AM)^2\right)\right| &= \mathbb{E}\sqrt{\sum_{i,j,k,l} a_{ii}a_{jj}m_{ij}m_{ji}a_{kk}a_{ll}\overline{m}_{kl}\overline{m}_{lk}} \\
&\leq \sqrt{\sum_{i,j,k,l} a_{ii}a_{jj}a_{kk}a_{ll}\mathbb{E}\left[m_{ij}m_{ji}\overline{m}_{kl}\overline{m}_{lk}\right]} \\
&= \sqrt{\frac{2n^2}{(n-1)(n+1)} - \frac{2}{(n-1)n(n+1)}\left(\sum_i a_{ii}^4\right)} \\
&\leq \sqrt{2 + \frac{1}{n^2-1}},
\end{aligned}$$

using Lemma 3.5 to evaluate the integrals.

Next,

$$\begin{aligned}
\mathbb{E}|W|^2 &= \mathbb{E}\left[\sum_{i,j} a_{ii}a_{jj}m_{ii}\overline{m}_{jj}\right] \\
&= \frac{1}{n}\sum_i a_{ii}^2 \\
&= 1.
\end{aligned}$$

Putting these estimates into (3.17) proves the theorem.

□

Theorem 3.4 yields the following bivariate corollary, which can also be seen as a corollary of the main unitary lemma of [**29**].

Corollary 3.6. *Let M be a random unitary matrix, A a fixed $n \times n$ matrix over $\mathbb{C}$ with $\mathrm{Tr}\,(AA^*) = n$, and let $W = \mathrm{Tr}\,(AM)$. Then the distribution of W converges to the standard complex normal distribution in the weak-star topology.*

Proof. The result follows immediately from Theorem 3.4 by considering the characteristic function of W. □

The corollary is refined to include a rate of convergence in the dual-Lipschitz distance in section 5.3.

CHAPTER 4

Eigenfunctions of the Laplacian

The results of this chapter are an abstraction of the results of section 2.1 and chapter 3. In those cases, a specific class of functions (linear) were considered on three specific manifolds – the sphere, the orthogonal group, and the unitary group. The exchangeable pair in each case was a small random rotation. Here, this approach is generalized to a large class of Riemannian manifolds, with the linear functions being replaced by eigenfunctions of the Laplacian.

Eigenvalues and eigenfunctions of the Laplacian Δ_g on a Riemannian manifold (M, g) have appeared as important and interesting objects in several branches of analysis. The spectrum of the Laplacian is a central object of study in spectral geometry, which is concerned with relationships between the spectrum of Δ_g and the geometry of M. In particular, spectral geometry deals with inverse problems such as Mark Kac's [**57**] famous question, "Can you hear the shape of a drum?" More specifically, how much information can one recover about the geometry of a manifold from knowledge of the spectrum of its Laplacian? Weyl [**82**] showed that you can determine the volume of the manifold: if $N(\lambda)$ denotes the number of eigenvalues less than λ, he showed that

$$\lim_{\lambda\to\infty} \frac{N(\lambda)}{\lambda} = \frac{\mathrm{vol}(M)}{2\pi}.$$

Milnor [**65**] showed that not all information about the manifold could be recovered. He found two Riemannian flat tori of dimension 16 which were not isometric, yet whose Laplacians had the same sequence of eigenvalues.

Eigenvalues and eigenfunctions of the Laplacian are also of central interest in quantum chaos. An important open question in the field is whether the quantum unique ergodicity conjecture of Rudnick and Sarnak [**73**] holds. The conjecture is the following.

CONJECTURE 4.1 (Rudnick-Sarnak). *Let M be a compact hyperbolic surface, and let $\{\phi_k\}_{k=1}^{\infty}$ be a sequence of linearly independent eigenfunctions of the Laplacian Δ_g on M, normalized to have L^2-norm one. Then the probability measures μ_k defined by*

$$\mu_k(A) = \int_A |\phi_k(x)|^2 \, d\mathrm{vol}(x)$$

converge to the measure vol *in the weak-star topology.*

Šnirel′man [**77**], Colin de Verdière [**28**], and Zelditch [**83**] have shown that the conjecture is true up to the removal of a subsequence of density zero, on manifolds on which the geodesic flow is ergodic. The conjecture above is consistent with the 'random wave model' of eigenstates (see [**13**]), which conjectures that individual eigenfunctions behave like random waves. See Sarnak [**74**] for an introduction to this area of study. In particular, it is discussed in [**74**] that a consequence of the random wave model would be that high eigenfunctions would be approximately normally distributed. In connection with this conjecture, Hejhal and various coauthors (see, e.g., [**49**]) and Aurich and Steiner [**3**] have provided numerical evidence indicating that on certain hyperbolic manifolds, the distributions of eigenfunctions become close to Gaussian as the eigenvalue tends to infinity. Theorem 4.3 below suggests a possible approach to rigorous results in this direction.

In this chapter we will freely use standard facts and notations of Riemannian geometry. For an introduction to the ideas used here, see the book [**24**] of Chavel, particularly chapters 1 and 3.

Let (M, g) be an n-dimensional Riemannian manifold. Integration with respect to the normalized volume measure is denoted $d\overline{\mathrm{vol}}$, thus $\int_M 1 d\overline{\mathrm{vol}} = 1$. For a linearly independent set of vector fields $\left\{\frac{\partial}{\partial x_i}\right\}_{i=1}^{n}$, define

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

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

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

Let $\Phi^t(x, v)$ denote the geodesic flow on the tangent bundle TM and let $\pi : TM \to M$ be the projection map of a tangent vector onto its base point. Recall that the manifold is said to have 'bounded geometry' if there is a non-trivial lower bound on the injectivity radius of the exponential map. We call a Riemannian manifold M *locally weakly symmetric* if M has bounded geometry, and there is an ϵ with $exp_x(t, v)$ injective on the unit tangent bundle SM for $t \in (-\epsilon, \epsilon)$, such that the volume measure on M is invariant under the substitution

$$\pi \circ \Phi^t(x, v) \to \pi(-\Phi^t(x, v))$$

for $t \in (-\epsilon, \epsilon)$. In particular, a compact locally symmetric space is weakly locally symmetric. The class of manifolds considered here is thus quite broad and includes Lie groups, quotients of Lie groups, flat tori and tori of genus at least two with hyperbolic geometry.

In what follows, some basic facts about the Laplacian on a Riemannian manifold are needed. Chavel's book [**23**] is a good source of background; in particular, the following theorem is Theorem 1 of section 1.3 (page 8).

THEOREM 4.2 (see [**23**]). *Let M be a compact, connected Riemannian manifold without boundary. The set of real numbers λ for which there exists a nontrivial solution $\varphi \in C^2(M)$ to*

$$\Delta\varphi + \lambda\varphi = 0$$

consists of a sequence

$$0 \leq \lambda_1 < \lambda_2 \cdots \uparrow +\infty$$

and each associated eigenspace is finite dimensional. Eigenspaces belonging to distinct eigenvalues are orthogonal in $L^2(M)$, and $L^2(M)$ is the direct sum of all the eigenspaces. Furthermore, each eigenfunction is C^∞ on M.

The following theorem is the main result of this section.

THEOREM 4.3. *Let M be a finite-volume locally weakly symmetric space (without boundary), and f an eigenfunction for the Laplacian on M with eigenvalue $\lambda \neq 0$,* *normalized so that* $\int_M f^2 d\overline{\mathrm{vol}} = 1$. *Let* X *be a random (i.e., distributed according to normalized volume measure) point of* M. *Then*

$$d_{TV}(f(X), Z) \leq \frac{1}{|\lambda|}\sqrt{\mathrm{Var}\big(\|\nabla f(X)\|^2\big)},$$

where Z *is a standard Gaussian random variable on* $\mathbb{R}$.

PROOF. Start by constructing a family of exchangeable pairs of points in M parametrized by ϵ. Let X be a random point of M and let $\epsilon > 0$ be smaller than the injectivity radius of the exponential map at X. Since M has bounded geometry, there is a range of ϵ small enough to work at every point. Now, choose a unit vector $V \in T_X M$ at random, independent of X, and define

$$X_\epsilon = exp_X(\epsilon V).$$

The fact that M is a locally weakly symmetric space implies that the pair (X, X_ϵ) is exchangeable. To see this, we show that

$$\int_{M\times M} g(x, x_\epsilon) d\mu = \int_{M\times M} g(x_\epsilon, x) d\mu$$

for all integrable $g : M \times M \to \mathbb{R}$, where μ is the measure defined by the construction of the pair (X, X_ϵ).

Let L denote the normalized Liouville measure on SM, which is locally the product of the normalized volume measure on M and normalized Lebesgue measure on the unit spheres of $T_x M$. The measure L also has the property that it is invariant under the geodesic flow $\Phi^t(x, v)$. See [**24**], section 5.1 for a construction of L and proofs of its key properties. By construction of the measure μ,

$$\begin{aligned}\int_{M\times M} g(x, x_\epsilon) d\mu &= \int_{SM} g(x, \pi \circ \Phi^\epsilon(x, v)) dL(x, v) \\ &= \int_{SM} g(\pi \circ \Phi^{-\epsilon}(y, \eta), y) dL(y, \eta)\end{aligned}$$

by the substitution $(y, \eta) = \Phi^\epsilon(x, v)$, since Liouville measure is invariant under the geodesic flow. Now,

$$\Phi^{-\epsilon}(y, \eta) = -\Phi^\epsilon(y, -\eta),$$

as both correspond to going backwards on the same geodesic. It follows that

$$\begin{aligned}\int_{SM} g(x, x_\epsilon) dL(x, v) &= \int_{SM} g(\pi(-\Phi^\epsilon(y, -\eta)), y) dL(y, \eta) \\ &= \int_{SM} g(\pi(-\Phi^\epsilon(y, \eta)), y) dL(y, \eta) \\ &= \int_{SM} g(\pi(\Phi^\epsilon(y, \eta)), y) dL(y, \eta) \\ &= \int_{SM} g(y_\epsilon, y) d\mu,\end{aligned}$$

where the second line follows from the substitution $-\eta \to \eta$, under which Lebesgue measure on the tangent space is invariant, and the third line follows by the symmetry assumption on M. Thus the pair (X, X_ϵ) is exchangeable as required.

Now, let f be an eigenfunction of the Laplacian on M with eigenvalue $\lambda < 0$ and $\|f\|_2 = 1$. Recall that the eigenspaces of the Laplacian are orthogonal and that a constant function on M is an eigenfunction with eigenvalue 0, thus $\int_M f d\mu = 0$. A consequence of Stokes' theorem is that

$$\int_M f \Delta f d\overline{\text{vol}} = - \int_M \|\nabla f\|^2 d\overline{\text{vol}}.$$

Since f is an eigenfunction with eigenvalue λ and $\|f\|_2 = 1$, this means that

$$\frac{1}{|\lambda|} \int_M \|\nabla f\|^2 d\overline{\text{vol}} = 1.$$

For notational convenience, let $W = f(X)$ and $W_\epsilon = f(X_\epsilon)$. Since (X, X_ϵ) is exchangeable, (W, W_ϵ) is an exchangeable pair as well.

In order to verify the conditions of Theorem 2.1, first let $\gamma : [0, 1] \to M$ be a constant-speed geodesic such that $\gamma(0) = X$, $\gamma(1) = X_\epsilon$, and $\gamma'(0) = \epsilon V$. Then applying Taylor's theorem on $\mathbb{R}$ to the function $f \circ \gamma$ yields

$$\begin{aligned}f(X_\epsilon) - f(X) &= \epsilon \cdot \left.\frac{d(f \circ \gamma)}{dt}\right|_{t=0} + \epsilon^2 \cdot \left.\frac{d^2(f \circ \gamma)}{dt^2}\right|_{t=0} + O(\epsilon^3) \\ &= \epsilon \cdot d_X f(V) + \epsilon^2 \cdot \left.\frac{d^2(f \circ \gamma)}{dt^2}\right|_{t=0} + O(\epsilon^3),\end{aligned} \tag{4.1}$$

where the coefficient implicit in the $O(\epsilon^3)$ depends on f and γ and $d_x f$ denotes the differential of f at x. Recall that $d_x f(v) = \langle \nabla f(x), v \rangle$ for $v \in T_x M$ and the gradient

$\nabla f(x)$ defined as above. Now, for X fixed, V is distributed according to normalized Lebesgue measure on $S_X M$ and $d_X f$ is a linear functional on $T_X M$. It follows that

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

so

$$\mathbb{E}\left[d_X f(V)\big|X\right] = 0.$$

This implies that

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

exists and is finite.

CLAIM 4.4.

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\big[f(X_\epsilon) - f(X)\big|X\big] = \frac{1}{2n}\Delta_g f(X) = \frac{\lambda}{2n} f(X).$$

PROOF. Let S_X^ϵ denote the geodesic sphere of radius ϵ in M centered at X, and let B_X^ϵ be the corresponding geodesic ball. Let A be the area form on S_X^ϵ; that is, dA denotes integration over S_X^ϵ with respect to the volume measure given by the Riemannian structure. Then

$$\begin{aligned}
&\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\left[f(X_\epsilon) - f(X)\big|X\right] \\
&\quad= \lim_{\epsilon\to 0}\frac{1}{\epsilon^2 A(S_X^\epsilon)}\int_{S_X^\epsilon}\big[f(X_\epsilon) - f(X)\big]dA(v) \\
&\quad= \lim_{\epsilon\to 0}\frac{1}{\epsilon A(S_X^\epsilon)}\int_{S_X^\epsilon}\frac{1}{\epsilon}\int_0^1\big\langle\nabla f, \gamma'(t)\big\rangle_{\gamma(t)}\,dt dA(v) \qquad (4.2)\\
&\quad= \lim_{\epsilon\to 0}\frac{1}{\epsilon A(S_X^\epsilon)}\int_{S_X^\epsilon}\frac{1}{\epsilon}\int_0^1\Big[\epsilon\langle v, \nabla f\rangle_X + \epsilon^2 t\langle v, \nabla_v(\nabla f)\rangle_X \\
&\qquad\qquad + \big\langle\gamma'(t), \nabla f\big\rangle_{\gamma(t)} - \langle\epsilon v, \nabla f + t\nabla_{\epsilon v}(\nabla f)\rangle_X\Big]\,dt dA(v),
\end{aligned}$$

by the fundamental theorem of calculus. (Recall that ∇_v denotes the covariant derivative in the direction v.)

Now,

$$\int_{S_X^\epsilon}\langle v, \nabla f\rangle_X\, dA(v) = 0$$

by symmetry as discussed above (the area form on S_X^ϵ pulls back to uniform measure on the sphere in the tangent space), so the first term of (4.2) is zero.

For the last term, note that $\langle \gamma'(t), \nabla f\rangle_{\gamma(t)} - \langle \epsilon v, \nabla f + t\nabla_{\epsilon v}(\nabla f)\rangle_X = 0$ for $t = 0$. The first derivative with respect to t of this expression is

$$\left\langle \gamma'(t), \nabla_{\gamma'(t)}(\nabla f)\right\rangle_{\gamma(t)} - \langle \epsilon v, \nabla_{\epsilon v}(\nabla f)\rangle_X$$

(which is also zero at $t = 0$) and the second derivative is

$$\left\langle \gamma'(t), \nabla_{\gamma'(t)}(\nabla_{\gamma'(t)}(\nabla f))\right\rangle_{\gamma(t)},$$

using the fact that $\nabla_{\gamma'(t)}(\gamma'(t)) = 0$, since γ is a geodesic. This last expression is $O(\epsilon^3)$, since $\|\gamma'(t)\| = \epsilon$ for each t, and f is smooth, so taking the limit gives

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\left[f(X_\epsilon) - f(X)\middle|X\right] &= \lim_{\epsilon\to 0}\frac{1}{2A(S_X^\epsilon)}\int_{S_X^\epsilon}\langle v, \nabla_v(\nabla f)\rangle_X\,dA \\ &= \frac{1}{2\sigma(S^{n-1})}\int_{S^{n-1}\subseteq T_XM}\langle v, \nabla_v(\nabla f)\rangle\,d\sigma(v),\end{aligned} \tag{4.3}$$

where σ is uniform measure on the sphere. This is now an integral over a Euclidean sphere. Choose an orthonormal basis $\left\{\frac{\partial}{\partial x_i}\Big|_X\right\}$ of T_XM which extends to Riemannian coordinates in a neighborhood of X. Then at X,

$$g_{ij} = \delta_{ij}, \qquad g^{ij} = \delta_{ij}, \qquad g = 1,$$

thus

$$\Delta_g f(X) = \sum_i \frac{\partial^2 f}{\partial x_i^2}.$$

Expressing the integrand in these coordinates gives:

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\left[f(X_\epsilon) - f(X)\middle|X\right] &= \frac{1}{2\sigma(S^{n-1})}\int_{S^{n-1}\subseteq T_XM}\sum_{i,j} v_iv_j\left\langle \frac{\partial}{\partial x_i}, \nabla_{\frac{\partial}{\partial x_j}}(\nabla f)\right\rangle d\sigma(v) \\ &= \frac{1}{2}\sum_i \frac{1}{n}\frac{\partial^2 f}{\partial x_i^2} \\ &= \frac{1}{2n}\Delta_g f(X) \\ &= \frac{\lambda}{2n}f(X),\end{aligned}$$

since $\mathbb{E}\big[v_i v_j\big] = \frac{1}{n}\delta_{ij}$ for v uniform on the sphere with $\{v_i\}$ the coordinates of v in an orthonormal basis.

$\square$

Consider next

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

By the expansion (4.1),

$$\begin{aligned}
\mathbb{E}\left[(W_\epsilon - W)^2 \big| X\right] &= \mathbb{E}\left[(f(X_\epsilon) - f(X))^2 \big| X\right] \\
&= \epsilon^2 \mathbb{E}\left[(d_x f(v))^2 \big| X\right] + O(\epsilon^3).
\end{aligned}$$

In the coordinates introduced above,

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

thus

$$\begin{aligned}
[d_x f(v)]^2 &= [\langle \nabla f, v\rangle]^2 \\
&= \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}\right)^2 v_i^2 + \sum_{i \neq j} \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j} v_i v_j.
\end{aligned}$$

Making use again of the independence of v and X and the fact that $\mathbb{E}[v_i v_j] = \frac{1}{n}\delta_{ij}$ yields

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon^2}\mathbb{E}\left[(d_x f(v))^2 \big| X\right] = \frac{1}{n}\|\nabla f\|^2 = -\frac{\lambda}{n} + \frac{1}{n}\left[\|\nabla f\|^2 + \lambda\right],$$

thus condition (2) is satisfied with

$$E = \frac{2}{\lambda}\left[\|\nabla f\|^2 + \lambda\right].$$

Finally, (4.1) gives immediately that

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon^2}\mathbb{E}\left[|W_\epsilon - W|^3 \big| W\right] = 0.$$

Recall that $\mathbb{E}\|\nabla f\|^2 = -\lambda$; the proof is completed by applying the Cauchy-Schwarz inequality to $\mathbb{E}|E|$.

□

4.1. Example: Second order spherical harmonics

In section 2.1, linear functions of a random point on the sphere were shown to be approximately normally distributed. The following theorem is a non-linear example of an application of Theorem 4.3, treating quadratic functions of the form $\sum_{i,j} a_{ij}x_ix_j$ of a random point on the sphere. These are eigenfunctions of the Laplacian on the sphere with eigenvalue $\lambda = -2n$ provided that $\mathrm{Tr}\,(A) = \sum_i a_{ii} = 0$. Some examples and remarks illustrating the theorem appear after the proof.

Theorem 4.5. *Let $g(x) = \sum_{i,j} a_{ij}x_ix_j$, where $A = (a_{ij})_{i,j=1}^n$ is a symmetric matrix with $\mathrm{Tr}\,(A) = 0$. Let $f = Cg$, where C is chosen such that $\|f\|_2 = 1$ when f is considered as a function on S^{n-1}. Let $W = f(X)$, where X is a random point on S^{n-1}. If d is the vector in $\mathbb{R}^n$ whose entries are the eigenvalues of A, then*

$$d_{TV}(W, Z) \leq \sqrt{5}\left(\frac{\|d\|_4}{\|d\|_2}\right)^2,$$

where $\|d\|_p = \left(\sum_i |d_i|^p\right)^{1/p}$.

Proof. To apply Theorem 4.3, it is first necessary to show that f is indeed an eigenfunction of the Laplacian on S^{n-1} and to identify its eigenvalue. It is discussed in section 9.5 of [**81**] that the eigenfunctions of the spherical Laplacian are exactly the restrictions of homogeneous harmonic polynomials on $\mathbb{R}^n$ to the sphere, and that such polynomials of degree l have eigenvalue $-l(l + n - 2)$. It is trivial to verify that since A is traceless, g is harmonic on $\mathbb{R}^n$ and thus is an eigenfunction with eigenvalue $\lambda = -2n$.

Next, note that $g(x) = \langle x, Ax\rangle$, and A is a symmetric matrix, so $A = U^*DU$ for a diagonal matrix D and an orthogonal matrix U. Thus

$$g(x) = \langle x, U^*DUx\rangle = \langle Ux, DUx\rangle.$$

Since the uniform distribution on the sphere is invariant under the action of the orthogonal group, this observation means that it suffices to prove the theorem in the case that $A = diag(a_1, \ldots, a_n)$.

By Theorem 4.3,

$$d_{TV}(W,Z) \leq \frac{1}{|\lambda|}\sqrt{Var\big(\|\nabla f\|^2\big)},$$

and the main task of the proof is to estimate the right-hand side. A consequence of Stokes' theorem is that

$$\int_{S^{n-1}} \|\nabla g\|^2 d\sigma = |\lambda| \int_{S^{n-1}} g^2 d\sigma,$$

so one can calculate $\mathbb{E}\|\nabla g\|^2$ in order to determine C. The spherical gradient ∇g is just the projection onto the hyperplane orthogonal to the radial direction of the usual gradient. Letting $\sum'$ stand for summing over distinct indexes,

$$\begin{aligned}
\mathbb{E}\|\nabla_{S^{n-1}} g(x)\|^2 &= \mathbb{E}\|\nabla_{\mathbb{R}^n} g(x)\|^2 - \mathbb{E}\left(x \cdot \nabla_{\mathbb{R}^n} g(x)\right)^2 \\
&= \mathbb{E}\left[\sum_{i=1}^n 4a_i^2 x_i^2\right] - 4\mathbb{E}\left[\sum_{i=1}^n a_i^2 x_i^4 + {\sum_{i,j}}' a_i a_j x_i^2 x_j^2\right] \\
&= \frac{4}{n}\sum_{i=1}^n a_i^2 - \frac{4}{n(n+2)}\left[3\sum_i a_i^2 + {\sum_{i,j}}' a_i a_j\right] \\
&= \frac{4}{n}\sum_{i=1}^n a_i^2 - \frac{4}{n(n+2)}\left[2\sum_i a_i^2 + \left(\sum_{i=1}^n a_i\right)^2\right] \\
&= \frac{4}{n+2}\sum_{i=1}^n a_i^2,
\end{aligned}$$

where Lemma 1.6 (after renormalization) has been used to get the third line, and the fact that $\mathrm{Tr}\,(A) = 0$ is used to get the last line. From this computation it follows that the constant C in the statement of the theorem should be taken to be $\sqrt{\frac{n(n+2)}{2\|a\|_2^2}}$, where $a = (a_i)_{i=1}^n$.

Now,

$$\mathbb{E}\|\nabla_{S^{n-1}} g\|^4 \leq \mathbb{E}\|\nabla_{\mathbb{R}^n} g\|^4,$$

since $\nabla_{S^{n-1}} g$ is a projection of $\nabla_{\mathbb{R}^n} g$.

$$\mathbb{E}\|\nabla_{\mathbb{R}^n} g\|^4 = 16\,\mathbb{E}\left[\sum_{i=1}^n a_i^4 x_i^4 + {\sum_{i,j}}' a_i^2 a_j^2 x_i^2 x_j^2\right]$$

$$= \frac{16}{n(n+2)}\left[\left(\sum_i a_i^2\right)^2 + 2\sum_i a_i^4\right],$$

and so

$$\begin{aligned}\mathbb{E}\|\nabla_{S^{n-1}}f\|^4 &\le \frac{n^2(n+2)^2}{4\|a\|_2^4}\mathbb{E}\|\nabla_{\mathbb{R}^n}g\|^4 \\ &= \frac{4n(n+2)}{\|a\|_2^4}\left[\left(\sum_i a_i^2\right)^2 + 2\sum_i a_i^4\right] \\ &= 4n(n+2)\left[1 + \frac{2\|a\|_4^4}{\|a\|_2^4}\right].\end{aligned}$$

This gives that

$$\mathrm{Var}\left(\|\nabla f\|^2\right) \le 8n^2\left(\frac{\|a\|_4^4}{\|a\|_2^4}\right) + 8n\left[1 + \frac{2\|a\|_4^4}{\|a\|_2^4}\right],$$

thus

$$d_{TV}(W,Z) \le \sqrt{\left(2+\frac{2}{n}\right)\left(\frac{\|a\|_4}{\|a\|_2}\right)^4 + \frac{2}{n}} \le \sqrt{4+\frac{2}{n}}\left(\frac{\|a\|_4}{\|a\|_2}\right)^2,$$

since $\|a\|_4 \ge n^{-1/4}\|a\|_2$. □

Remark: For n even, consider the quadratic function

$$f(x) = \sqrt{\frac{n+2}{2}}\left[\sum_{i=1}^{\frac{n}{2}} x_i^2 - \sum_{j=\frac{n}{2}+1}^{n} x_j^2\right].$$

This f satisfies the conditions of the theorem and is normalized such that $\mathbb{E}f^2(X) = 1$ for X a random point on S^{n-1}. Theorem 4.5 gives that

$$d_{TV}(f(X), Z) \le \sqrt{\frac{5}{n}},$$

so in this case, Theorem 4.5 gives a rate of convergence to normal.

The cases in which the bound does not tend to zero as $n \to \infty$ are those in which a small number of coordinates of X control the value of $f(X)$; in those situation one would not expect f to be normally distributed. For example, if $f(x) = cx_1^2 - cx_2^2$, where c is the proper normalization constant, then because x_1 and x_2 are asymptotically independent

Gaussian random variables (see section 5.2), f is asymptotically distributed as $c(Z_1^2 - Z_2^2)$ for independent Gaussians Z_1 and Z_2.

CHAPTER 5

Multivariate normal approximation

5.1. A multivariate abstract approximation theorem

This section contains the statement and proof of a multivariate analog to Theorem 2.1; the results of this section are joint work with Sourav Chatterjee.

There has been some previous work in using Stein's method for multivariate normal approximation. Rinott and Rotar used a 'two level' dependency graph approach to prove central limit theorems for dependent random vectors, which they applied to studying statistics on nearest neighbor random graphs. Their approach yields bounds on

$$\big|\mathbb{E}h(W) - \mathbb{E}h(Z)\big|$$

for Z a multivariate Gaussian random vector, W the random vector being studied, and h in a class $\mathcal{H}$ of (not necessarily smooth) functions. They restrict attention to classes of functions $\mathcal{H}$ with the following properties. For $h : \mathbb{R}^n \to \mathbb{R}$, set

$$h_\epsilon^+ = \sup\{h(x+y) : |y| \le \epsilon\}, \qquad h_\epsilon^- = \inf\{h(x+y) : |y| \le \epsilon\},$$
$$\tilde{h}(x;\epsilon) = h_\epsilon^+(x) - h_\epsilon^-(x).$$

Then it is required of $\mathcal{H}$ that

(i) If $h \in \mathcal{H}$, $\|h\|_\infty \le 1$.

(ii) If $h \in \mathcal{H}$, then $h_\epsilon^+ \in \mathcal{H}$ and $h_\epsilon^- \in \mathcal{H}$ for any $\epsilon > 0$.

(iii) For any $n \times n$ matrix A and vector $b \in \mathbb{R}^n$, if $h \in \mathcal{H}$ then $h(Ax+b) \in \mathcal{H}$.

(iv) For all $\epsilon > 0$,
$$\sup\{\mathbb{E}\tilde{h}(Z;\epsilon) : h \in \mathcal{H}\} < a\epsilon,$$
where $Z \sim \mathfrak{N}(0,1)$ and a is a constant depending only on $\mathcal{H}$ and n.

The constant a is sometimes called the isoperimetric constant of $\mathcal{H}$.

The class $\mathcal{H}$ can for example be taken to be the indicator functions of convex sets. However, the restrictions on $\mathcal{H}$ are enough that their approach does not yield a bound in total variation distance.

Goldstein and Rinott [**46**] have also developed a technique for multivariate normal approximation using Stein's method and size-bias coupling, where the dependence in the problem need not be local. They apply their techniques to statistics of random graphs. The notion of distance used is bounding

$$\mathbb{E}h(W) - \mathbb{E}h(Z)$$

over a class of smooth functions h.

More recently, Goldstein and Reinert [**45**] have developed a version of multivariate normal approximation using Stein's method together with zero-bias coupling, with applications to simple random sampling. The also consider a notion of distance which compares expectations of smooth functions with respect to the random variable in question and the normal distribution.

In the univariate case, there are problems which are most easily treated via the method of exchangeable pairs than other approaches to Stein's method, e.g., when the random variable under study depends on an underlying reversible Markov chain. See in particular the analysis of the antivoter chain in [**71**] and the examples in [**33**]. This partly motivated the development of a multivariate version of the method of exchangeable pairs in [**22**]. Another motivation for this development is the fact that the extension of Stein's method to situations with continuous symmetries given in chapter 2 seems very well suited to work with the method of exchangeable pairs. This is still the case in a multivariate context; see Theorem 5.3 below.

The following lemma gives a second-order characterizing operator for the Gaussian distribution on $\mathbb{R}^k$. For $f \in C^1(\mathbb{R}^k)$, define the gradient of f by $\nabla f(x) = (\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_k}(x))^t$. Define the Laplacian of f by $\Delta f(x) = \sum_{i=1}^k \frac{\partial^2 f}{\partial x_i^2}(x)$.

LEMMA 5.1. *Let $Z \in \mathbb{R}^k$ be a random vector with $\{Z_i\}$ i.i.d. standard Gaussians.*

(i) *If* $f : \mathbb{R}^k \to \mathbb{R}$ *is two times continuously differentiable and compactly supported, then*

$$\mathbb{E}\big[\Delta f(Z) - Z \cdot \nabla f(Z)\big] = 0.$$

(ii) *If* $Y \in \mathbb{R}^k$ *is a random vector such that*

$$\mathbb{E}\big[\Delta f(Y) - Y \cdot \nabla f(Y)\big] = 0$$

for every $f \in C^2(\mathbb{R}^k)$*, then* $\mathcal{L}(Y) = \mathcal{L}(Z)$*.*

(iii) *If* $g \in C_o^\infty(\mathbb{R}^k)$*, then the function*

$$U_o g(x) := \int_0^1 \frac{1}{2t}\big[\mathbb{E}g(\sqrt{t}x + \sqrt{1-t}Z) - \mathbb{E}g(Z)\big]dt$$

is a solution to the differential equation

$$\Delta h(x) - x \cdot \nabla h(x) = g(x) - \mathbb{E}g(Z). \tag{5.1}$$

Proof. Part (i) is just integration by parts.

For part (iii), first note that since g is Lipschitz, if $t \in \big(0, \frac{1}{2}\big)$ and L is the Lipschitz constant of g,

$$\begin{aligned}
\left|\frac{1}{2t}\big[\mathbb{E}g(\sqrt{t}x + \sqrt{1-t}Z) - \mathbb{E}g(Z)\big]\right| &\le \frac{L}{2t}\mathbb{E}\left|\sqrt{t}x + (\sqrt{1-t} - 1)Z\right| \\
&\le \frac{L}{2t}\left[\sqrt{t}|x| + \frac{t}{\sqrt{2}}\sqrt{k}\right],
\end{aligned}$$

which is integrable on $\big(0, \frac{1}{2}\big)$, and the integrand is bounded on $\big(\frac{1}{2}, 1\big)$, so the integral exists by the dominated convergence theorem.

To show that it is indeed a solution to the differential equation (5.1), let

$$Z_t = \sqrt{t}x + \sqrt{1-t}Z$$

and observe that

$$\begin{aligned}
g(x) - \mathbb{E}g(Z) &= \int_0^1 \frac{d}{dt}\mathbb{E}g(Z_t)dt \\
&= \int_0^1 \frac{1}{2\sqrt{t}}\mathbb{E}(x \cdot \nabla g(Z_t))dt - \int_0^1 \frac{1}{2\sqrt{1-t}}\mathbb{E}(Z \cdot \nabla g(Z_t))dt
\end{aligned}$$

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

by integration by parts. Noting that

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

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 (iii).

Finally, if

$$\mathbb{E}\big[\Delta f(Y) - Y \cdot \nabla f(Y)\big] = 0$$

for every $f \in C^2(\mathbb{R}^k)$, then for $g \in C_o^\infty$ given,

$$\mathbb{E}g(Y) - \mathbb{E}g(Z) = \mathbb{E}\big[\Delta(U_o g)(Y) - Y \cdot \nabla(U_o g)(Y)\big] = 0,$$

and so $\mathcal{L}(Y) = \mathcal{L}(Z)$ since C_o^∞ is dense in the class of bounded continuous functions vanishing at infinity, with respect to $\|\cdot\|_\infty$.

The next lemma gives various useful bounds on $U_o g$ and its derivatives in terms of g and its derivatives. Below, $|\nabla g|$ is the length of the gradient vector ∇g, thus $\||\nabla g|\|_\infty$ denotes the supremum of $|\nabla g(x)|$ over $x \in \mathbb{R}^k$.

□

LEMMA 5.2. *For $g \in C_o^3(\mathbb{R}^k)$, $U_o g$ satisfies the following bounds:*

(i) *For all $1 \le i \le k$,*

$$\left\| \frac{\partial (U_o g)}{\partial x_i} \right\|_\infty \le \frac{1}{\sqrt{2}} \|g\|_\infty.$$

(ii) *For $1 \le i, j \le k$,*

$$\left\| \frac{\partial^2 (U_o g)}{\partial x_i \partial x_j} \right\|_\infty \le \left\| \frac{\partial g}{\partial x_j} \right\|_\infty .$$

(iii) *If H is the Hessian matrix of $U_o g$, then*

$$\|H\|_{H.S.} = \sqrt{\operatorname{Tr}(HH^t)} \le \sqrt{k} \||\nabla g|^2\|_\infty$$

and

$$\|H\|_{op} = \sup_{|v|=1,|w|=1} |\langle Hv, w\rangle| \leq \sqrt{\frac{2}{\pi}} \||\nabla g|\|_\infty.$$

(iv) *For* $1 \leq i, j, \ell \leq k$,

$$\left\| \frac{\partial^3 (U_o g)}{\partial x_i \partial x_j \partial x_\ell} \right\|_\infty \leq \left\| \frac{\partial^2 g}{\partial x_j \partial x_\ell} \right\|_\infty .$$

Proof. Write $h(x) = U_o g(x)$. Note that by the formula for $U_o g$, if $g \in C_o^r(\mathbb{R}^k)$, then

$$\frac{\partial^r h}{\partial x_{i_1} \cdots \partial x_{i_r}}(x) = \int_0^1 (2t)^{-1} t^{r/2} \mathbb{E}\left[\frac{\partial^r g}{\partial x_{i_1} \cdots \partial x_{i_r}}(Z_t) \right] dt. \tag{5.2}$$

By integration by parts,

$$\mathbb{E}\left[\frac{\partial g}{\partial x_i}(Z_t) \right] = \frac{1}{\sqrt{1-t}} \mathbb{E}\left[Z_i g(Z_t) \right],$$

thus

$$\begin{aligned} \left| \frac{\partial h}{\partial x_i}(x) \right| &= \left| \int_0^1 \frac{1}{2\sqrt{t(1-t)}} \mathbb{E}\left[Z_i g(Z_t) \right] dt \right| \\ &\leq \|g\|_\infty \sqrt{\frac{2}{\pi}} \int_0^1 \frac{1}{2\sqrt{t(1-t)}} dt \\ &= \frac{1}{\sqrt{2}} \|g\|_\infty. \end{aligned}$$

This proves part (i); parts 2 and 4 are essentially the same argument.

For part 3,

$$\begin{aligned} \frac{\partial^2 h}{\partial x_i \partial x_j}(x) &= \int_0^1 \frac{1}{2} \mathbb{E}\left[\frac{\partial^2 g}{\partial x_i \partial x_j}(Z_t) \right] dt \\ &= \int_0^1 \frac{1}{2\sqrt{1-t}} \mathbb{E}\left[Z_i \frac{\partial g}{\partial x_j}(Z_t) \right] dt. \end{aligned} \tag{5.3}$$

Summing in k gives

$$\begin{aligned}\sum_{j=1}^{k}\left(\frac{\partial^2 h}{\partial x_i \partial x_j}(x)\right)^2 &= \sum_{j=1}^{k}\left(\int_0^1 \frac{1}{2\sqrt{1-t}}\mathbb{E}\left[Z_i \frac{\partial g}{\partial x_j}(Z_t)\right]dt\right)^2 \\ &\leq \int_0^1 \frac{1}{2\sqrt{1-t}}\mathbb{E}\left[Z_i^2 \sum_{j=1}^{k}\left(\frac{\partial g}{\partial x_j}(Z_t)\right)^2\right]dt \\ &\leq \||\nabla g|^2\|_\infty,\end{aligned}$$

where the second line follows by two applications of Jensen's inequality, first considering the function $\mathbb{E}\left[Z_i \frac{\partial g}{\partial x_j}(Z_t)\right]$ being integrated with respect to $\frac{1}{2\sqrt{1-t}}dt$ on $(0,1)$, then to the expression $\left(\mathbb{E}\left[Z_i \frac{\partial g}{\partial x_j}(Z_t)\right]\right)^2$. Summing in i now gives that $\|H(x)\|_{H.S.}^2 \leq k\||\nabla g|^2\|_\infty$ for all x.

The bound on $\|H\|_{op}$ is proved similarly. Making use of equation (5.3),

$$\begin{aligned}|\langle Hv, w\rangle| &= \left|\int_0^1 \frac{1}{2\sqrt{1-t}}\mathbb{E}\left[\left(\sum_{i=1}^{k} v_i Z_i\right)\left(\sum_{j=1}^{k} w_j \frac{\partial g}{\partial x_j}(Z_t)\right)\right]dt\right| \\ &= \left|\int_0^1 \frac{1}{2\sqrt{1-t}}\mathbb{E}\left[\langle v, Z\rangle \langle w, \nabla g(Z_t)\rangle\right]dt\right| \\ &\leq \||\nabla g|\|_\infty \int_0^1 \frac{1}{2\sqrt{1-t}}\mathbb{E}\left|\langle v, Z\rangle\right|dt \\ &\leq \||\nabla g|\|_\infty \sqrt{\frac{2}{\pi}},\end{aligned}$$

since $\langle v, Z\rangle$ is a standard Gaussian random variable. □

The following is a multivariate analog of Theorem 2.1. It is not an analog in the strictest sense, as it gives convergence rates only in the dual-Lipschitz distance rather than the stronger total variation distance. This seems to be an unavoidable consequence of the fact that the characterizing operator used here for the multivariate normal distribution is a second-order operator, whereas the characterizing operator for the standard normal distribution on $\mathbb{R}$ used in section 1 is first-order.

Theorem 5.3. *Let X and X_ϵ be two random vectors such that $\mathcal{L}(X) = \mathcal{L}(X_\epsilon)$ with the property that $\lim_{\epsilon\to 0} X_\epsilon = X$ almost surely. Suppose there is a constant λ, functions*

$E_{ij} = E_{ij}(X)$, h and k with

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}h(\epsilon) = \lim_{\epsilon\to 0}\frac{1}{\epsilon^2}k(\epsilon) = 0,$$

and $\alpha(X)$ and $\beta(X)$ with

$$\mathbb{E}|\alpha(X)| < \infty, \qquad \mathbb{E}|\beta(X)| < \infty,$$

such that

(i) $\mathbb{E}\left[(X_\epsilon - X)_i \big| X\right] = -\lambda\epsilon^2 X_i + h(\epsilon)\alpha(X)$

(ii) $\mathbb{E}\left[(X_\epsilon - X)_i(X_\epsilon - X)_j | X\right] = 2\lambda\epsilon^2\delta_{ij} + \epsilon^2 E_{ij} + k(\epsilon)\beta(X)$

(iii) $\mathbb{E}\big[|X_\epsilon - X|^3\big] = o(\epsilon^2)$.

Then

$$d_{L^*}(X,Z) \le \min\left\{\frac{1}{2\lambda}\sum_{i,j}\mathbb{E}|E_{ij}|, \frac{\sqrt{k}}{2\lambda}\mathbb{E}\left(\sum_{i,j}E_{ij}^2\right)^{1/2}\right\}. \tag{5.4}$$

If furthermore $E_{ij} = E_iF_j + \delta_{ij}R_i$, then writing E for the vector with components E_i and F for the vector with components F_j,

$$d_{L^*}(X,Z) \le \frac{1}{\lambda\sqrt{2\pi}}\mathbb{E}\left(\|E\|_2\|F\|_2\right) + \frac{1}{\lambda}\sum_{i=1}^k \mathbb{E}|R_i|.$$

Proof. Fix g with $\|g\|_L \le 1$ and let $U_o g$ be as in Lemma 5.1. As before, write $h(x) = U_o g(x)$ and observe

$$\begin{aligned} 0 &= \frac{1}{\epsilon^2\lambda}\mathbb{E}\left[h(X_\epsilon) - h(X)\right] \\ &= \frac{1}{\epsilon^2\lambda}\mathbb{E}\left[(X_\epsilon - X)\cdot\nabla h(X) + \frac{1}{2}(X_\epsilon - X)^t(\operatorname{Hess} h(X))(X_\epsilon - X) + R\right], \end{aligned} \tag{5.5}$$

where R is the error in the second-order Taylor approximation of $h(X_\epsilon) - h(X)$. By Taylor's theorem,

$$R = O\left(|X_\epsilon - X|^3\right),$$

where the implied constant depends on g. It follows from (iii) that

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon^2\lambda}\mathbb{E}\big[R\big] = 0.$$

For the first two terms of 5.5,

(5.6)

$$\begin{aligned}
&\frac{1}{\epsilon^2\lambda}\mathbb{E}\Big[(X_\epsilon - X)\cdot\nabla h(X) + \frac{1}{2}(X_\epsilon - X)^t(\operatorname{Hess} h(X))(X_\epsilon - X)\Big] \\
&\qquad = \frac{1}{\epsilon^2\lambda}\mathbb{E}\left[\mathbb{E}\left[(X_\epsilon - X)\big|X\right]\cdot\nabla h(X) + \frac{1}{2}\sum_{i,j=1}^{k}\mathbb{E}\left[(X_\epsilon - X)_i(X_\epsilon - X)_j\big|X\right]\frac{\partial^2 h}{\partial x_i \partial x_j}(X)\right] \\
&\qquad = \mathbb{E}\left[-X\cdot\nabla h(X) + \Delta h(X) + \frac{1}{2\lambda}\sum_{i,j}E_{ij}\frac{\partial^2 h}{\partial x_i\partial x_j}(X) + o(1)(\alpha(X)+\beta(X))\right] \\
&\qquad = \mathbb{E}g(X) - \mathbb{E}g(Z) + \mathbb{E}\left[\frac{1}{2\lambda}\sum_{i,j}E_{ij}\frac{\partial^2 h}{\partial x_i\partial x_j}(X)\right] + o(1),
\end{aligned}$$

where (i) and (ii) have been used to get the third line and Lemma 5.1 (iii) and the assumptions on α and β have been used to get the last line. Taking the limit of both sides of (5.5) as $\epsilon \to 0$ now gives

$$\big|\mathbb{E}g(X) - \mathbb{E}g(Z)\big| \leq \frac{1}{2\lambda}\mathbb{E}\left|\sum_{i,j}E_{ij}\frac{\partial^2 h}{\partial x_i\partial x_j}(X)\right| \tag{5.7}$$

The bound

$$d_{L^*}(X,Z) \leq \frac{1}{2\lambda}\sum_{i,j}\mathbb{E}|E_{ij}|$$

is now immediate from the bound on $\frac{\partial^2 h}{\partial x_i \partial x_j}$ of Lemma 5.2. Alternatively, apply the Cauchy-Schwarz inequality to (5.7) to get

$$\begin{aligned}
\big|\mathbb{E}g(X) - \mathbb{E}g(Z)\big| &\leq \frac{1}{2\lambda}\mathbb{E}\left[\left(\sum_{i,j}E_{ij}^2\right)^{1/2}\left(\sum_{i,j}\left(\frac{\partial^2 h}{\partial x_i\partial x_j}(X)\right)^2\right)^{1/2}\right] \\
&\leq \frac{\|H\|_{H.S.}}{2\lambda}\mathbb{E}\left(\sum_{i,j}E_{ij}^2\right)^{1/2} \\
&\leq \frac{\sqrt{k}}{2\lambda}\mathbb{E}\left(\sum_{i,j}E_{ij}^2\right)^{1/2}.
\end{aligned}$$

If the condition $E_{ij} = E_iF_j + \delta_{ij}R_i$ holds, then write

$$\frac{1}{2\lambda}\mathbb{E}\left|\sum_{i,j} E_{ij}\frac{\partial^2 h}{\partial x_i \partial x_j}\right| \leq \frac{1}{2\lambda}\mathbb{E}\left[\left|\sum_{i,j}\frac{\partial^2 h}{\partial x_i \partial x_j}E_iF_j\right| + \left|\sum_i \frac{\partial^2 h}{\partial^2 x_i}R_i\right|\right].$$

The result is now immediate by applying the bound on $\|H\|_{op}$ to the first term and the bounds on the $\frac{\partial^2 h}{\partial^2 x_i}$ to the second term.

□

5.2. Rank k projections of spherically symmetric measures on $\mathbb{R}^n$

Consider a random vector $Y \in \mathbb{R}^n$ whose distribution is spherically symmetric; i.e., if U is a fixed orthogonal matrix, then the distribution of Y is the same as the distribution of UY. Assume that $\mathbb{E}|Y|^2 < \infty$. By the spherical symmetry, $\mathbb{E}Y_i = 0$ for each i and $\mathbb{E}Y_iY_j = c\delta_{ij}$ for some c; assume that Y has been normalized so that $c = 1$. Assume further that there is a constant a so that

$$\mathrm{Var}(|Y|^2) \leq a.$$

For k fixed, let P_k denote the orthogonal projection of $\mathbb{R}^n$ onto the span of the first k standard basis vectors. In this example, Theorem 5.3 is applied to show that $X := P_k(Y)$ is approximately distributed as a k-dimensional Gaussian random vector if $k = o(n)$.

This is a multivariate version of the example of section 2.1, and as discussed in that section, is closely related to the results of Diaconis and Freedman in [**36**]. In particular, they proved the following.

Theorem 5.4 (Diaconis-Freedman). *Let Y be as above and let $\mathbb{P}_k$ be the law of $(Y_1, \ldots, Y_k)$. Let $Z_1, Z_2, \ldots$ be independent standard normal random variables. For $k \in \mathbb{N}$, let $\mathbb{P}_\sigma^k$ be the law of $(\sigma Z_1, \ldots, \sigma Z_k)$ and for μ a probability on $[0, \infty)$, let*

$$\mathbb{P}_{\mu k} = \int P_\sigma^k d\mu(\sigma).$$

Then there is a probability μ on $[0, \infty)$ such that for $1 \leq k \leq n - 4$,

$$d_{TV}(\mathbb{P}_k, \mathbb{P}_{\mu k}) \leq \frac{2(k+3)}{n-k-3}.$$

The mixing measure μ can be taken to be the law of $|Y|^2$.

Theorem 5.4 says that orthogonally invariant vectors have their first k components close (in total variation distance) to a mixture of normals. In some cases, the explicit form given for the mixing measure has allowed the theorem to be used to prove central limit theorems of interest in convex geometry; see [**18**] and [**58**]. Theorem 5.5 below says that the variance bound above is sufficient to show that the mixing measure of Theorem 5.4 can be taken to be a point mass. The rates obtained are of the same order, though the rate of Theorem 5.5 is not in total variation distance, but in the weaker dual-Lipschitz distance.

To apply Theorem 5.3 to projections of the vector Y, construct a family of exchangeable pairs (X, X_ϵ) as follows. For $\epsilon > 0$ fixed, let

$$\begin{aligned} A_\epsilon &= \begin{bmatrix} \sqrt{1-\epsilon^2} & \epsilon \\ -\epsilon & \sqrt{1-\epsilon^2} \end{bmatrix} \oplus I_{n-2} \\ &= I_n + \begin{bmatrix} -\frac{\epsilon^2}{2} + O(\epsilon^4) & \epsilon \\ -\epsilon & -\frac{\epsilon^2}{2} + O(\epsilon^4) \end{bmatrix} \oplus 0_{n-2} \end{aligned}$$

Let U be a random $n \times n$ orthogonal matrix, independent of Y, and let

$$Y_\epsilon = UA_\epsilon U^t Y;$$

Y_ϵ is a small random rotation of Y. Let $X_\epsilon = P_k Y_\epsilon$. Note that

$$X_\epsilon - X = \epsilon\left[-\left(\frac{\epsilon}{2} + O(\epsilon^3)\right) P_k K K^t + P_k K C_2 K^t\right] Y, \tag{5.8}$$

where K is the $k \times 2$ matrix made of the first two columns of U and $C_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

Recall from section 3.1 that $\mathbb{E}\big[KK^t\big] = \frac{2}{n} I$ and $\mathbb{E}\big[KC_2K^t\big] = 0$, and so

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

condition 1 thus holds with $\lambda = \frac{1}{n}$. The E_{ij} are determined as follows.

$$\begin{aligned}
\lim_{\epsilon\to 0}\frac{n}{\epsilon^2}\mathbb{E}\left[(X_\epsilon - X)_i^2\middle|X\right] &= n\mathbb{E}\left[(P_kKC_2K^tY)_i^2\middle|X\right] \\
&= n\mathbb{E}\left[\left(\sum_j \mathbb{I}(i\le k)(KC_2K^t)_{ij}Y_j\right)^2\middle|X\right] \\
&= n\mathbb{I}(i\le k)\mathbb{E}\left[\sum_{j,\ell} Y_jY_\ell(u_{i1}u_{j2}-u_{i2}u_{j1})(u_{i1}u_{\ell 2}-u_{i2}u_{\ell 1})\middle|X\right] \\
&= 2n\mathbb{I}(i\le k)\mathbb{E}\left[\sum_{j,\ell} Y_jY_\ell\left(\frac{1}{n(n-1)}\right)\delta_{j\ell}(1-\delta_{ij})\middle|X\right] \\
&= \frac{2\mathbb{I}(i\le k)}{n-1}\mathbb{E}\left[\sum_{j\ne i} Y_j^2\middle|X\right] \\
&= \frac{2\mathbb{I}(i\le k)}{n-1}\mathbb{E}\left[|Y|^2 - Y_i^2\middle|X\right],
\end{aligned}$$

thus for $1\le i\le k$

$$E_{ii} = \lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\left[(X_\epsilon - X)_i^2\middle|X\right] - \frac{2}{n} = \frac{2}{n(n-1)}\Big[\mathbb{E}\left[|Y|^2 - n\middle|X\right] - (X_i^2-1)\Big].$$

For $i\ne j$,

$$\begin{aligned}
&\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\left[(X_\epsilon - X)_i(X_\epsilon - X)_j\middle|X\right] \\
&\quad= \mathbb{E}\left[(P_kKC_2K^tY)_i(P_kKC_2K^tY)_j\middle|X\right] \\
&\quad= \mathbb{E}\left[\mathbb{I}(i\le k)\mathbb{I}(j\le k)\sum_{\ell,m} Y_\ell Y_m(u_{i1}u_{\ell 2}-u_{i2}u_{\ell 1})(u_{j1}u_{m2}-u_{j2}u_{m1})\middle|X\right] \\
&\quad= 2\mathbb{I}(i\le k)\mathbb{I}(j\le k)\mathbb{E}\left[\sum_{\ell,m} Y_\ell Y_m\left(\frac{-1}{n(n-1)}\right)\delta_{im}\delta_{\ell j}\middle|X\right] \\
&\quad= \frac{-2}{n(n-1)}X_iX_j.
\end{aligned}$$

Thus

$$E_{ij} = \frac{-2}{n(n-1)}X_iX_j + \frac{2\delta_{ij}}{n(n-1)}\mathbb{E}\left[|Y|^2 - (n-1)\middle|X\right];$$

in particular, E_{ij} has the form $E_iF_j + \delta_{ij}R_i$ with

$$\begin{aligned} E_i &= \sqrt{\frac{2}{n(n-1)}}X_i, \\ F_j &= -\sqrt{\frac{2}{n(n-1)}}X_j, \\ R_i &= \frac{2}{n(n-1)}\mathbb{E}\left[|Y|^2 - (n-1)\middle|X\right]. \end{aligned}$$

Now,

$$\mathbb{E}\|X\|_2^2 = k$$

and

$$\mathbb{E}\left|\mathbb{E}\left[|Y|^2 - (n-1)\middle|X\right]\right| \le \sqrt{a} + 1$$

by assumption, so applying Theorem 5.3 gives:

THEOREM 5.5. *With notation as above, there is a constant c (depending on a) such that*

$$d_{L^*}(X, Z) \le \frac{ck}{n}.$$

5.3. Complex-linear functions of random unitary matrices

The following theorem is an application of Theorem 5.3 giving a rate of convergence in Theorem 3.6

THEOREM 5.6. *Let $M \in \mathfrak{U}_n$ be distributed according to Haar measure, and let A be a fixed $n \times n$ matrix such that* $\mathrm{Tr}\,(AA^*) = n$. *Let $W = \mathrm{Tr}\,(AM)$ and let Z be a standard complex normal random variable. Then there is a constant c such that*

$$d_{L^*}(W, Z) \le \frac{c}{n}.$$

Remark: The constant in the statement of the Theorem is asymptotically equal to $3\sqrt{2}$; for $n \ge 6$, c can be taken to be 6.

PROOF. Much, though not all, of the work needed to apply Theorem 5.3 was done in section 3.2. Recall from the beginning of the proof of Theorem 3.4 that by the singular value decomposition and the translation invariance of Haar measure, it suffices to prove the theorem for A diagonal with positive entries. Note that the normalization is such that $\mathrm{Re}\, W$ and $\mathrm{Im}\, W$ each have variance $\frac{1}{2}$, so Theorem 5.3 will actually be applied to the random variable $\sqrt{2}W$.

To apply the theorem, we must first construct a family of pairs (W, W_ϵ). Let $U \in \mathfrak{U}_n$ be a random unitary matrix, independent of M, and let $M_\epsilon = UA_\epsilon U^* M$, where

$$A_\epsilon = \begin{bmatrix} \sqrt{1-\epsilon^2} & \epsilon & & & \\ -\epsilon & \sqrt{1-\epsilon^2} & & 0 & \\ & & 1 & & \\ & 0 & & \ddots & \\ & & & & 1 \end{bmatrix},$$

thus M_ϵ is a small random rotation of M. Let $W_\epsilon = W(M_\epsilon)$; (W, W_ϵ) is exchangeable by construction.

Now, let I_2 be the 2×2 identity matrix, K the $n \times 2$ matrix made from the first two columns of U, and let

$$C_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Then

$$M_\epsilon = M + K\big[(\sqrt{1-\epsilon^2} - 1)I_2 + \epsilon C_2\big]K^* M.$$

By Taylor's theorem, $\sqrt{1-\epsilon^2} = 1 - \frac{\epsilon^2}{2} + O(\epsilon^4)$, thus

$$M_\epsilon = M + K\left[\left(-\frac{\epsilon^2}{2} + O(\epsilon^4)\right) I_2 + \epsilon C_2\right] K^* M,$$

and so

$$\begin{aligned} W_\epsilon - W &= \mathrm{Tr}\left(\left(-\frac{\epsilon^2}{2} + O(\epsilon^4)\right) AKK^*M + \epsilon AKC_2K^*M\right) \\ &= \epsilon\left[\left(-\frac{\epsilon}{2} + O(\epsilon^3)\right) tr(AKK^*M) + tr(AKC_2K^*M)\right]. \end{aligned} \tag{5.9}$$

Expanding in components and using the independence of M and U together with Lemma 3.5 (i) gives

$$\begin{aligned}
(KK^*)_{ij} &= u_{i1}\overline{u}_{j1} + u_{i2}\overline{u}_{j2} \\
\mathbb{E}\left[(KK^*)_{ij}\middle|W\right] &= \frac{2}{n}\delta_{ij} \qquad (5.10)\\
(KC_2K^*)_{ij} &= u_{i1}\overline{u}_{j2} - u_{i2}\overline{u}_{j1} \\
\mathbb{E}\left[(KC_2K^*)_{ij}\middle|W\right] &= 0 \qquad (5.11)
\end{aligned}$$

Thus

$$\begin{aligned}
\frac{n}{\epsilon^2}\mathbb{E}\left[W_\epsilon - W\middle|W\right] &= -\frac{n}{2}\mathbb{E}\left[\mathrm{Tr}\,(AKK^*M)\middle|W\right] + \frac{n}{\epsilon}\mathbb{E}\left[\mathrm{Tr}\,(AKC_2K^*M)\middle|W\right] \\
&= -W,
\end{aligned} \tag{5.12}$$

and the first condition of Theorem 5.3 holds with $\lambda = \frac{1}{n}$. To check the second condition of the theorem and determine the E_{ij}, the following three expressions are needed:

$$\begin{aligned}
\mathbb{E}\left[\big(\mathrm{Re}(W_\epsilon - W)\big)^2\middle|W\right] &= \frac{1}{2}\,\mathrm{Re}\left[(W_\epsilon - W)^2 + |W_\epsilon - W|^2\right], \\
\mathbb{E}\left[\big(\mathrm{Im}(W_\epsilon - W)\big)^2\middle|W\right] &= \frac{1}{2}\,\mathrm{Re}\left[|W_\epsilon - W|^2 - (W_\epsilon - W)^2\right], \\
\mathbb{E}\left[\big(\mathrm{Re}(W_\epsilon - W)\big)\big(\mathrm{Im}(W_\epsilon - W)\big)\middle|W\right] &= \frac{1}{2}\,\mathrm{Im}\left[(W_\epsilon - W)^2\right].
\end{aligned}$$

By (5.9),

$$\begin{aligned}
\lim_{\epsilon\to 0}&\frac{1}{\epsilon^2}\mathbb{E}[(W_\epsilon - W)^2|W] \\
&= \lim_{\epsilon\to 0}\mathbb{E}\left[\left(\mathrm{Tr}\,(AKC_2K^*M)\right)^2\middle|W\right] \\
&= \mathbb{E}\left[\sum_{i,j,k,l} a_{ii}m_{ji}a_{kk}m_{lk}(u_{i1}\overline{u}_{j2} - u_{i2}\overline{u}_{j1})(u_{k1}\overline{u}_{l2} - u_{k2}\overline{u}_{l1})\middle| W\right]
\end{aligned} \tag{5.13}$$

Recall (see Lemma 3.5 (vi)) that

$$\begin{aligned}
\mathbb{E}[(u_{i1}\overline{u}_{j2} - &u_{i2}\overline{u}_{j1})(u_{k1}\overline{u}_{l2} - u_{k2}\overline{u}_{l1})] \\
&= -\frac{2\delta_{il}\delta_{jk}(1-\delta_{ij})}{(n-1)(n+1)} + \frac{2\delta_{ij}\delta_{kl}(1-\delta_{ik})}{(n-1)n(n+1)} - \frac{2\mathbb{I}(i=j=k=l)}{n(n+1)}.
\end{aligned} \tag{5.14}$$

Let $\sum_{i,j}'$ stand for summing over all pairs (i,j) where i and j are distinct. Putting (5.14) into (5.13) and using the independence of M and U gives:

(5.15)

$$\begin{aligned}
&\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\left[(W_\epsilon - W)^2\,\middle|W\right]\\
&\qquad= \frac{2}{(n-1)(n+1)}\mathbb{E}\left[-\sum_{i,j}{}' a_{ii}a_{jj}m_{ij}m_{ji} + \frac{1}{n}\sum_{i,k}{}' a_{ii}a_{kk}m_{ii}m_{kk}\right.\\
&\qquad\qquad\qquad\qquad\left. - \left(\frac{n-1}{n}\right)\sum_i a_{ii}^2 m_{ii}^2\right| W\Bigg]\\
&\qquad= \frac{2}{(n-1)(n+1)}\mathbb{E}\left[-\left(\mathrm{Tr}\,((AM)^2) - \sum_i (AM)_{ii}^2\right)\right.\\
&\qquad\qquad\qquad\left. + \frac{1}{n}\left(W^2 - \sum_i (AM)_{ii}^2\right) - \frac{n-1}{n}\sum_i (AM)_{ii}^2\right| W\Bigg]\\
&\qquad= \frac{2}{(n-1)(n+1)}\mathbb{E}\left[-\,\mathrm{Tr}\,((AM)^2) + \frac{1}{n}W^2\right| W\Bigg].
\end{aligned}$$

Next,

$$\begin{aligned}
&\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}[|W_\epsilon - W|^2 | W]\\
&\qquad = \lim_{\epsilon\to 0}\mathbb{E}\left[|\mathrm{Tr}\,(AKC_2K^*M|)^2\,\middle|W\right]
\end{aligned}$$

$$= \mathbb{E}\left[\sum_{i,j,k,l} a_{ii}m_{ji}a_{kk}\overline{m}_{lk}(u_{i1}\overline{u}_{j2} - u_{i2}\overline{u}_{j1})(\overline{u}_{k1}u_{l2} - \overline{u}_{k2}u_{l1})\right| W\Bigg]. \tag{5.16}$$

From Lemma 3.5 ((vii)),

$$\begin{aligned}
&\mathbb{E}\big[(u_{i1}\overline{u}_{j2} - u_{i2}\overline{u}_{j1})(\overline{u}_{k1}u_{l2} - \overline{u}_{k2}u_{l1})\big]\\
&\qquad = \frac{2\delta_{ik}\delta_{jl}(1-\delta_{ij})}{(n-1)(n+1)} - \frac{2\delta_{ij}\delta_{kl}(1-\delta_{ik})}{n(n-1)(n+1)} + \frac{2\mathbb{I}(i=j=k=l)}{n(n+1)},
\end{aligned} \tag{5.17}$$

thus

(5.18)

$$\begin{aligned}
&\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\left[|W_\epsilon - W|^2\,|W\right]\\
&\qquad= \frac{2}{(n-1)(n+1)}\mathbb{E}\left[\left.{\sum_{i,j}}' a_{ii}^2|m_{ji}|^2 - {\sum_{i,j}}' a_{ii}a_{jj}m_{ii}\overline{m}_{jj} + \frac{n-1}{n}\sum_i a_{ii}^2|m_{ii}|^2\right| W\right]\\
&\qquad= \frac{2}{(n-1)(n+1)}\mathbb{E}\left[\sum_i a_{ii}^2(1-|m_{ii}|^2) - \left|\sum_i a_{ii}m_{ii}\right|^2\right.\\
&\qquad\qquad\left.+\sum_i a_{ii}^2|m_{ii}|^2 + \frac{n-1}{n}\sum_i a_{ii}^2|m_{ii}|^2\right| W\Bigg]\\
&\qquad= \frac{2}{n} + \frac{2(1-|W|^2)}{(n-1)n(n+1)},
\end{aligned}$$

where the normalization $tr(AA^*) = n$ has been used to get the last line.

Using (5.15) and (5.18) now gives

$$\begin{aligned}
&\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\big[\big(\operatorname{Re}(W_\epsilon - W)\big)^2\big|W\big]\\
&\qquad\qquad= \frac{1}{n} + \frac{1}{(n-1)(n+1)}\operatorname{Re}\left[\frac{1-|W|^2}{n} + \frac{W^2}{n} + \mathbb{E}\left[-\operatorname{Tr}\left((AM)^2\right)\big|\,W\right]\right],
\end{aligned}$$

$$\begin{aligned}
&\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\big[\big(\operatorname{Im}(W_\epsilon - W)\big)^2\big|W\big]\\
&\qquad\qquad= \frac{1}{n} + \frac{1}{(n-1)(n+1)}\operatorname{Re}\left[\frac{1-|W|^2}{n} - \frac{W^2}{n} + \mathbb{E}\left[\operatorname{Tr}\left((AM)^2\right)\big|\,W\right]\right],
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\big[\big(\operatorname{Re}(W_\epsilon - W)\big)\big(\operatorname{Im}(W_\epsilon - W)\big)\big|W\big]\\
&\qquad\qquad= \frac{1}{(n-1)(n+1)}\operatorname{Im}\mathbb{E}\left[\left.-\operatorname{Tr}\left((AM)^2\right) + \frac{W^2}{n}\right| W\right].
\end{aligned}$$

This gives, for E_{ij} as in Theorem 5.3 for the random variable $\sqrt{2}W$,

$$\begin{aligned}E_{11} &= \lim_{\epsilon\to 0}\frac{2}{\epsilon^2}\mathbb{E}\left[\big(\operatorname{Re}(W_\epsilon - W)\big)^2\big|W\right] - \frac{2}{n}\\ &= \frac{2}{(n-1)(n+1)}\operatorname{Re}\left[\frac{1-|W|^2}{n} + \frac{W^2}{n} + \mathbb{E}\left[-\operatorname{Tr}\left((AM)^2\right)\big|\,W\right]\right],\end{aligned}$$

$$\begin{aligned}E_{22} &= \lim_{\epsilon\to 0}\frac{2}{\epsilon^2}\mathbb{E}\left[\big(\operatorname{Im}(W_\epsilon - W)\big)^2\big|W\right] - \frac{2}{n}\\ &= \frac{2}{(n-1)(n+1)}\operatorname{Re}\left[\frac{1-|W|^2}{n} - \frac{W^2}{n} + \mathbb{E}\left[\operatorname{Tr}\left((AM)^2\right)\big|\,W\right]\right],\end{aligned}$$

$$\begin{aligned}E_{12} &= \lim_{\epsilon\to 0}\frac{2}{\epsilon^2}\mathbb{E}\left[\big(\operatorname{Re}(W_\epsilon - W)\big)\big(\operatorname{Im}(W_\epsilon - W)\big)\big|W\right]\\ &= \frac{2}{(n-1)(n+1)}\operatorname{Im}\mathbb{E}\left[-\operatorname{Tr}\left((AM)^2\right) + \frac{W^2}{n}\middle|\,W\right].\end{aligned}$$

The normalization is such that $\mathbb{E}|W|^2 = 1$, so the first two terms of both E_{11} and E_{22} each have expectations bounded by $\frac{2}{n(n-1)(n+1)}$. It remains to bound $\mathbb{E}|\operatorname{Tr}\left((AM)^2\right)|$. Making use of Lemma 3.5,

$$\begin{aligned}\mathbb{E}\big|\operatorname{Tr}\left((AM)^2\right)\big|^2 &= \mathbb{E}\big[\operatorname{Tr}\left((AM)^2\right)\overline{\operatorname{Tr}\left((AM)^2\right)}\big]\\ &= \mathbb{E}\left[\sum_{i,j,k,\ell} a_{ii}a_{jj}a_{kk}a_{\ell\ell}m_{ij}m_{ji}\overline{m_{k\ell}m_{\ell k}}\right]\\ &= \frac{1}{(n-1)(n+1)}\sum_{i,j,k,\ell} a_{ii}a_{jj}a_{kk}a_{\ell\ell}\Big[\delta_{ik}\delta_{j\ell}(1-\delta_{ij}) + \delta_{i\ell}\delta_{jk}(1-\delta_{ik})\\ &\quad + \left(\frac{2(n-1)}{n}\right)\mathbb{I}(i=j=k=\ell)\Big]\\ &= \frac{2}{(n-1)(n+1)}\left[{\sum_{i,j}}' a_{ii}^2a_{jj}^2 + \left(\frac{n-1}{n}\right)\sum_i a_{ii}^4\right]\\ &= \frac{2}{(n-1)(n+1)}\left[n^2 - \sum_i a_{ii}^4 + \left(\frac{n-1}{n}\right)\sum_i a_{ii}^4\right]\\ &= 2 + \frac{2}{(n-1)(n+1)} - \frac{2}{(n-1)n(n+1)}\sum_i a_{ii}^4\\ &\le 2 + \frac{2}{(n-1)(n+1)}.\end{aligned}$$

This completes the proof. □

5.4. Rank k projections of Haar measure on $\mathcal{O}_n$

In this section, the result of section 3.1 on the asymptotic normality of rank 1 projections of Haar measure on the orthogonal group is extended to rank k projections for $k = o(n^{2/3})$. This is also a comparison theorem between Gaussian matrices and Haar distributed matrices as discussed in section 3.1, since for Gaussian matrices, all lower dimensional projections are Gaussian.

THEOREM 5.7. *Let $B_1, \ldots, B_k$ be linearly independent $n \times n$ matrices over $\mathbb{R}$ such that $\mathrm{Tr}\,(B_i B_i^t) = n$ for each i. Let $b_{ij} = \mathrm{Tr}\,(B_i B_j^t)$. Let M be a random orthogonal matrix and let*

$$W = (\mathrm{Tr}\,(B_1 M), \mathrm{Tr}\,(B_2 M), \ldots, \mathrm{Tr}\,(B_k M)) \in \mathbb{R}^k.$$

Let $Y = (Y_1, \ldots, Y_k)$ be a random vector whose components are standard normals, with covariance matrix $\frac{1}{n}\,(b_{ij})_{i,j=1}^k$. Then

$$\sup_{\|f\|_L \le 1} \left|\mathbb{E}f(W) - \mathbb{E}f(Y)\right| \le \frac{2\sqrt{\lambda}k^{3/2}}{n-1}$$

for $n \ge 3$, where λ is the largest eigenvalue of $\frac{1}{n}\,(b_{ij})_{i,j=1}^k$ and $\|f\|_L$ is the Lipschitz norm of f.

This theorem follows fairly easily from the following special case.

THEOREM 5.8. *Let $A_1, \ldots, A_k$ be $n \times n$ matrices over $\mathbb{R}$ satisfying $\mathrm{Tr}\,(A_i A_j^t) = n\delta_{ij}$; for $i \neq j$, A_i and A_j are orthogonal with respect to the Hilbert-Schmidt inner product. Let M be a random orthogonal matrix, and consider the vector*

$$W = (\mathrm{Tr}\,(A_1 M), \mathrm{Tr}\,(A_2 M), \ldots, \mathrm{Tr}\,(A_k M)) \in \mathbb{R}^k.$$

Let $Z = (Z_1, \ldots, Z_k)$ be a random vector whose components are independent standard normal random variables. Then for $f : \mathbb{R}^k \to \mathbb{R}$ bounded and Lipschitz,

$$\left|\mathbb{E}f(W) - \mathbb{E}f(Z)\right| \le \frac{2\|f\|_L k^{3/2}}{n-1}$$

for $n \geq 3$.

Remark: The rate of $\frac{k^{3/2}}{n}$ is probably not sharp. If A_i has $(A_i)_{1i} = \sqrt{n}$ and $(A_i)_{jk} = 0$ otherwise, then the random vector W above is just $\sqrt{n}$ times the first k entries of the i-th column of M. But a column of M is distributed as a random point on S^{n-1}, so this reduces to the previous example, where the rate was on the order $\frac{k}{n}$. It is argued on page 406 of [**36**] that the rate $\frac{k}{n}$ (in the total variation distance) is sharp, suggesting that $\frac{k}{n}$ is the best one can hope for. It does not follow formally that $\frac{k}{n}$ is a lower bound, as the dual-Lipschitz distance is weaker than the total variation distance.

Example. Let M be a random $n \times n$ orthogonal matrix, and let

$$0 < a_1 < a_2 < \ldots < a_k = n.$$

For each $1 \leq i \leq n$, let

$$B_i = \sqrt{\frac{n}{a_i}} I_{a_i} \oplus \mathbf{0}_{n-a_i};$$

B_i has $\sqrt{\frac{n}{a_i}}$ in the first a_i diagonal entries and zeros everywhere else. If $i \leq j$, then $\langle B_i, B_j \rangle_{HS} = n\sqrt{\frac{a_i}{a_j}}$; in particular, $\langle B_i, B_i \rangle_{HS} = n$. The B_i are linearly independent w.r.t. the Hilbert-Schmidt inner product since the a_i are all distinct, so to apply Theorem 5.7, we have only to bound the eigenvalues of the matrix $\left(\sqrt{\frac{a_{\min(i,j)}}{a_{\max(i,j)}}}\right)_{i,j=1}^{k}$. But this is easy, since $|\lambda| \leq \sqrt{\sum_{i,j=1}^{k} \frac{a_{\min(i,j)}}{a_{\max(i,j)}}} \leq k$ for all eigenvalues λ (see, e.g., [**14**], page 7). It now follows from Theorem 5.7 that if Y is a vector of standard normals with covariance matrix $\left(\sqrt{\frac{a_{\min(i,j)}}{a_{\max(i,j)}}}\right)_{i,j=1}^{k}$ and $W = (\mathrm{Tr}\,(B_1 M), \ldots, \mathrm{Tr}\,(B_k M))$, then

$$\sup_{\|f\|_L \leq 1} \left|\mathbb{E}f(W) - \mathbb{E}f(Y)\right| \leq \frac{2k^2}{n-1}.$$

Proof of Theorem 5.7 from Theorem 5.8. Perform the Gram-Schmidt algorithm on the matrices $\{B_1, \ldots, B_k\}$ with respect to the Hilbert-Schmidt inner product $\langle C, D \rangle = \mathrm{Tr}\,(CD^t)$ to get matrices $\{A_1, \ldots, A_k\}$ which are mutually orthogonal and have

H-S norm $\sqrt{n}$. Denote the matrix which takes the B's to the A's by C^{-1}; the matrix is invertible since the B's are linearly independent. Now by assumption,

$$\begin{aligned} b_{ij} &= \langle B_i, B_j \rangle \\ &= \left\langle \sum_k c_{ik} A_k, \sum_l c_{jl} A_l \right\rangle \\ &= n \sum_k c_{ik} c_{jk}. \end{aligned}$$

Thus $CC^t = \frac{1}{n} \left(b_{ij} \right)_{i,j=1}^k$.

Now, let $f : \mathbb{R}^k \to \mathbb{R}$ with $\|f\|_L \le 1$. Define $h : \mathbb{R}^k \to \mathbb{R}$ by $h(x) = f(Cx)$. Then $|h|_L \le \sqrt{\lambda}$, since the operator norm of C as a function from $\mathbb{R}^k$ to itself is bounded by the largest singular value of C. By Theorem 5.8,

$$\left| \mathbb{E} h(\mathrm{Tr}\,(A_1 M), \ldots, \mathrm{Tr}\,(A_k M)) - \mathbb{E} h(Z) \right| \le \frac{2\sqrt{\lambda} k^{3/2}}{n-1}$$

for Z a standard Gaussian random vector in $\mathbb{R}^k$. But $C\big(\mathrm{Tr}\,(A_1 M), \ldots, \mathrm{Tr}\,(A_k M)\big) = \big(\mathrm{Tr}\,(B_1 M), \ldots, \mathrm{Tr}\,(B_k M)\big)$ and CZ has standard normal components with covariance matrix $\frac{1}{n} \left(b_{ij} \right)_{i,j=1}^k$. $\square$

Proof of Theorem 5.8. Make an exchangeable pair (M, M_ϵ) as before; $M_\epsilon = U A_\epsilon U^t M$, where U is a random orthogonal matrix independent of M and

$$A_\epsilon = \begin{bmatrix} \sqrt{1-\epsilon^2} & \epsilon \\ -\epsilon & \sqrt{1-\epsilon^2} \end{bmatrix} \oplus \mathbf{0}_{n-2}.$$

Let $W_\epsilon = (\mathrm{Tr}\,(A_1 M_\epsilon), \ldots, \mathrm{Tr}\,(A_k M_\epsilon))$.

Recall from section 3.1 that

$$M_\epsilon - M = \epsilon \left[\left(\frac{-\epsilon}{2} + O(\epsilon^3) \right) K K^t + K C_2 K^2 \right] M, \tag{5.19}$$

where K is the first two columns of U and $C_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. It follows that

$$(W_\epsilon)_i - W_i = \epsilon \left[\left(\frac{-\epsilon}{2} + O(\epsilon^3) \right) \mathrm{Tr}\,(A_i K K^t M) + \mathrm{Tr}\,(A_i K C_2 K^2 M) \right].$$

As in section 3.1, $\mathbb{E}\big[KK^t\big] = \frac{2}{n} I$ and $\mathbb{E}\big[KC_2K^t\big] = 0$, and so

$$\lim_{\epsilon \to 0} \frac{n}{\epsilon^2} \mathbb{E}\left[W_\epsilon - W \middle| W\right] = -W.$$

Condition 1 of Theorem 5.3 is thus satisfied with $\lambda = \frac{1}{n}$. If $\alpha \neq \beta$, $E_{\alpha\beta}$ is computed as follows. For notational convenience, write $A_\alpha = A = (a_{ij})$ and $A_\beta = B = (b_{ij})$, and let $\sum'$ denote summing over distinct indices. By (5.19),

(5.20)
$$\lim_{\epsilon \to 0} \frac{1}{\epsilon^2} \mathbb{E}\left[(W_\epsilon - W)_\alpha (W_\epsilon - W)_\beta \middle| W\right] = \mathbb{E}\left[\mathrm{Tr}\,(AKC_2K^tM)\,\mathrm{Tr}\,(BKC_2K^tM) \middle| W\right],$$

since all of the other terms in the product are of higher order in ϵ. Making use of Lemma 3.3 (iv),

(5.21)

$$\begin{aligned}
&\mathbb{E}\big[\operatorname{Tr}(AKC_2K^tM)\operatorname{Tr}(BKC_2K^tM)\big|W\big] \\
&\quad= \mathbb{E}\left[\sum_{i,j,k,l,p,q} a_{ik}b_{jp}m_{li}m_{qj}(u_{k1}u_{l2}-u_{k2}u_{l1})(u_{p1}u_{q2}-u_{p2}u_{q1})\big|W\right] \\
&\quad= \mathbb{E}\left[\sum_{ij}\sideset{}{'}\sum_{kl}\sideset{}{'}\sum_{p,q} a_{ik}b_{jp}m_{li}m_{qj}\left(\frac{2}{n(n-1)}\right)(\delta_{kp}\delta_{lq}-\delta_{kq}\delta_{lp})\big|W\right] \\
&\quad= \frac{2}{n(n-1)}\left[\sum_{i,j}\sideset{}{'}\sum_{k,l} a_{ik}b_{jk}m_{li}m_{lj} - \sum_{i,j}\sideset{}{'}\sum_{k,l} a_{ik}b_{jl}m_{li}m_{kj}\,\middle|\, W\right] \\
&\quad= \frac{2}{n(n-1)}\mathbb{E}\left[\sum_{i,j}\left(\sum_k a_{ik}b_{jk}\right)\left(\sum_l m_{li}m_{lj}\right) - \sum_{i,j,k} a_{ik}b_{jk}m_{ki}m_{kj}\right. \\
&\qquad\qquad \left. -\sum_{i,j}\left(\sum_k a_{ik}m_{kj}\right)\left(\sum_l b_{jl}m_{li}\right) + \sum_{i,j,k} a_{ik}b_{jk}m_{ki}m_{kj}\,\middle|\, W\right] \\
&\quad= \frac{2}{n(n-1)}\mathbb{E}\left[-\sum_{i,j}(AM)_{ij}(BM)_{ji}\,\middle|\, W\right] \\
&\quad= \frac{-2}{n(n-1)}\mathbb{E}\left[\operatorname{Tr}(AMBM)\big|W\right],
\end{aligned}$$

where the orthogonality of the columns of M was used to get the second-last line. In section 3.1 it was shown that

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon^2}\mathbb{E}\left[(W_\epsilon - W)_i^2\big|W\right] = \frac{2}{n} + \frac{2}{n(n-1)}\mathbb{E}\left[1-\operatorname{Tr}((A_iM)^2)\big|W\right], \tag{5.22}$$

thus

$$E_{ii} = \frac{2}{n(n-1)}\mathbb{E}\left[1-\operatorname{Tr}((A_iM)^2)\big|W\right].$$

It remains to estimate $\mathbb{E}\big(E_{ij}^2\big)$. **Claim:**

(i) If $n \geq 3$, then $\mathbb{E}\left(\operatorname{Tr}((A_iM)^2)-1\right)^2 \leq 4$ for all i

(ii) If $n \geq 2$, then $\mathbb{E}\left(\operatorname{Tr}(A_iMA_jM)\right)^2 \leq 4$ for all i and j with $i \neq j$.

The claim gives that $\mathbb{E}(E_{ij}^2) \leq \frac{16}{n^2(n-1)^2}$ for all i, j, which together with Theorem 5.3 proves Theorem 5.8.

It remains to prove the claim. In fact, part (i) of the claim was proved in section 3.1, as the last step in the proof of Theorem 3.1.

The second part of the claim is messier than the first, since the matrices A_i and A_j cannot be simultaneously diagonalized in general. However, we can assume (making use of the singular value decomposition as always) that one of the two parameter matrices is diagonal. Again write $A_i = A$ and $A_j = B$, and suppose that A is diagonal. The strategy is to write $\operatorname{Tr}(AMBM)$ in components, break up the sums according to the equality structure of the indices, use Lemma 3.3 and collect terms.

$$
\begin{aligned}
\mathbb{E}\big(\operatorname{Tr}(AMBM)\big)^2 &= \mathbb{E}\left[\sum_{i,j,k,\alpha,\beta,\gamma} a_{ii}a_{\alpha\alpha}b_{jk}b_{\beta\gamma}m_{ij}m_{ki}m_{\alpha\beta}m_{\gamma\alpha}\right] \\
&= \sum_i a_{ii}^2 b_{ii}^2 \mathbb{E}\left[m_{ii}^4\right] + {\sum_{i,j}}' a_{ii}^2 b_{ji}^2 \mathbb{E}\left[m_{ii}^2 m_{ij}^2\right] + {\sum_{i,j}}' a_{ii}a_{jj}b_{ii}b_{jj}\mathbb{E}\left[m_{ii}^2 m_{jj}^2\right] \\
&\quad + {\sum_{i,j}}' a_{ii}a_{jj}b_{ij}b_{ji}\mathbb{E}\left[m_{ii}m_{ij}m_{ji}m_{jj}\right] + {\sum_{i,j}}' a_{ii}^2 b_{ij}^2 \mathbb{E}\left[m_{ii}^2 m_{ji}^2\right] \\
&\quad + \sum_i \sum_{\substack{j\neq i\\ k\neq i}} a_{ii^2} b_{jk}^2 \mathbb{E}\left[m_{ij}^2 m_{ki}^2\right] + {\sum_{i,j}}' a_{ii}a_{jj}b_{ij}b_{ji}\mathbb{E}\left[m_{ii}m_{ij}m_{ji}m_{jj}\right] \\
&\quad + {\sum_{i,j}}' a_{ii}a_{jj}b_{ii}b_{jj}\mathbb{E}\left[m_{ij}^2 m_{ji}^2\right] \\
&= \frac{3}{n(n+2)}\sum_i a_{ii}^2 b_{ii}^2 + \frac{1}{n(n+2)}{\sum_{i,j}}' a_{ii}^2 b_{ji}^2 + \frac{n+1}{(n-1)n(n+2)}{\sum_{i,j}}' a_{ii}a_{jj}b_{ii}b_{jj} \\
&\quad - \frac{1}{(n-1)n(n+2)}{\sum_{i,j}}' a_{ii}a_{jj}b_{ij}b_{ji} + \frac{1}{n(n+2)}{\sum_{i,j}}' a_{ii}^2 b_{ij}^2 \\
&\quad + \frac{n+1}{(n-1)n(n+2)}\sum_i \sum_{\substack{j\neq i\\ k\neq i}} a_{ii}^2 b_{jk}^2 - \frac{1}{(n-1)n(n+2)}{\sum_{i,j}}' a_{ii}a_{jj}b_{ij}b_{ji} \\
&\quad + \frac{n+1}{(n-1)n(n+2)}{\sum_{i,j}}' a_{ii}a_{jj}b_{ii}b_{jj} \\
&= \frac{-2}{(n-1)n(n+2)}\left[\sum_{i,j} a_{ii}^2 b_{ij}^2 + \sum_{i,j} a_{ii}^2 b_{ji}^2\right] + \frac{2(n+1)}{(n-1)n(n+2)}\sum_{i,j} a_{ii}a_{jj}b_{ii}b_{jj} \\
&\quad - \frac{2}{(n-1)n(n+1)}\sum_{i,j} a_{ii}a_{jj}b_{ij}b_{ji} + \frac{n(n+1)}{(n-1)(n+2)}.
\end{aligned}
$$

Now,

$$\sum_{i,j} a_{ii}a_{jj}b_{ii}b_{jj} \le \left(\sum_i a_{ii}^2\right)\left(\sum_i b_{ii}^2\right) \le n^2.$$

By the Cauchy-Schwarz inequality,

$$\left|\sum_{i,j} a_{ii}a_{jj}b_{ij}b_{ji}\right| \le \sqrt{\left(\sum_{i,j} a_{ii}^2 a_{jj}^2\right)\left(\sum_{i,j} a_{ij}^2 b_{ji}^2\right)}.$$

Now,

$$\sum_{i,j} a_{ii}^2 a_{jj}^2 = \left(\sum_i a_{ii}^2\right)^2$$
$$= n^2.$$

By repeated application of the Cauchy-Schwarz inequality,

$$\sum_{i,j} b_{ij}^2 b_{ji}^2 \le \sum_{i,j} b_{ij}^4 \le \left(\sum_{i,j} b_{ij}^2\right)^2 = n^2.$$

It follows that

$$\left|\sum_{i,j} a_{ii}a_{jj}b_{ij}b_{ji}\right| \le n^2,$$

and so

$$\mathbb{E}\big(\operatorname{Tr}(AMBM)\big)^2 \le \frac{3n(n+1)}{(n-1)(n+2)} + \frac{2}{(n-1)n(n+2)} = 3 + \frac{2}{(n-1)(n+2)}\left[1+\frac{1}{n}\right] \le 4$$

for $n \ge 2$. This proves the second claim.

$\square$

Bibliography

[1] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.*, 17(1):9–25, 1989.

[2] R. Arratia, L. Goldstein, and L. Gordon. Poisson approximation and the Chen-Stein method. *Statist. Sci.*, 5(4):403–434, 1990. With comments and a rejoinder by the authors.

[3] R. Aurich and F. Steiner. Statistical properties of highly excited quantum eigenstates of a strongly chaotic system. *Phys. D*, 64(1-3):185–214, 1993.

[4] P. Baldi and Y. Rinott. Asymptotic normality of some graph-related statistics. *J. Appl. Probab.*, 26(1):171–175, 1989.

[5] P. Baldi, Y. Rinott, and C. Stein. A normal approximation for the number of local maxima of a random function on a graph. In *Probability, Statistics, and Mathematics*, pages 59–81. Academic Press, Boston, MA, 1989.

[6] I. Bárány and V. Vu. Central limit theorems for Gaussian random polytopes. Preprint, 2006.

[7] A. D. Barbour. Stein's method and Poisson process convergence. *J. Appl. Probab.*, (Special Vol. 25A):175–184, 1988. A celebration of applied probability.

[8] A. D. Barbour. Stein's method for diffusion approximations. *Probab. Theory Related Fields*, 84(3):297–322, 1990.

[9] A. D. Barbour and G. K. Eagleson. Multiple comparisons and sums of dissociated random variables. *Adv. in Appl. Probab.*, 17(1):147–162, 1985.

[10] A. D. Barbour, L. Holst, and S. Janson. *Poisson Approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1992. Oxford Science Publications.

[11] A. D. Barbour, S. Janson, M. Karoński, and A. Ruciński. Small cliques in random graphs. *Random Structures Algorithms*, 1(4):403–434, 1990.

[12] A. D. Barbour, M. Karoński, and A. Ruciński. A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B*, 47(2):125–145, 1989.

[13] M. V. Berry. Regular and irregular semiclassical wavefunctions. *J. Phys. A*, 10(12):2083–2091, 1977.

[14] R. Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[15] M. Bloznelis and F. Götze. An Edgeworth expansion for finite-population U-statistics. *Bernoulli*, 6(4):729–760, 2000.

[16] E. Bolthausen. An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete*, 66(3):379–386, 1984.

[17] E. Borel. Sur les principes de la theorie cinétique des gaz. *Annales de l'ecole normale sup.*, 23:9–32, 1906.

[18] U. Brehm and J. Voigt. Asymptotics of cross sections for convex bodies. *Beiträge Algebra Geom.*, 41(2):437–454, 2000.

[19] W. Bryc. *The Normal Distribution – Characterizations with Applications*, volume 100 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1995.

[20] S. Chatterjee. An abstract berry-esseen theorem. Preprint, 2006.

[21] S. Chatterjee, P. Diaconis, and E. Meckes. Exchangeable pairs and Poisson approximation. *Probab. Surv.*, 2:64–106 (electronic), 2005.

[22] S. Chatterjee and E. Meckes. Multivariate normal approximation using exchangeable pairs. Preprint, 2006.

[23] I. Chavel. *Eigenvalues in Riemannian Geometry*, volume 115 of *Pure and Applied Mathematics*. Academic Press Inc., Orlando, FL, 1984. Including a chapter by Burton Randol, With an appendix by Jozef Dodziuk.

[24] I. Chavel. *Riemannian Geometry—a Modern Introduction*, volume 108 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1993.

[25] L. H. Y. Chen. On the convergence of Poisson binomial to Poisson distributions. *Ann. Probability*, 2(1):178–180, 1974.

[26] L. H. Y. Chen. Poisson approximation for dependent trials. *Ann. Probability*, 3(3):534–545, 1975.

[27] O. Chryssaphinou and S. Papastavridis. A limit theorem for the number of nonoverlapping occurrences of a pattern in a sequence of independent trials. *J. Appl. Probab.*, 25(2):428–431, 1988.

[28] Y. Colin de Verdière. Ergodicité et fonctions propres du laplacien. *Comm. Math. Phys.*, 102(3):497–502, 1985.

[29] A. D'Aristotile, P. Diaconis, and C. Newman. Brownian motion and the classical groups. In *Probability, Statistics and Their Applications: Papers in Honor of Rabi Bhattacharya*, volume 41 of *IMS Lecture Notes Monogr. Ser.*, pages 97–116. Inst. Math. Statist., Beachwood, OH, 2003.

[30] A. Dembo and Y. Rinott. Some examples of normal approximations by Stein's method. In *Random Discrete Structures (Minneapolis, MN, 1993)*, volume 76 of *IMA Vol. Math. Appl.*, pages 25–44. Springer, New York, 1996.

[31] P. Diaconis. Application of the method of moments in probability and statistics. In *Moments in Mathematics (San Antonio, Tex., 1987)*, volume 37 of *Proc. Sympos. Appl. Math.*, pages 125–142. Amer. Math. Soc., Providence, RI, 1987.

[32] P. Diaconis. Patterns in eigenvalues: the 70th Josiah Willard Gibbs lecture. *Bull. Amer. Math. Soc. (N.S.)*, 40(2):155–178 (electronic), 2003.

[33] P. Diaconis. Stein's method for Markov chains: first examples. In *Stein's Method: Expository Lectures and Applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, pages 27–43. Inst. Math. Statist., Beachwood, OH, 2004.

[34] P. Diaconis, M. Eaton, and S. Lauritzen. Finite de Finetti theorems in linear models and multivariate analysis. *Scand. J. Statist.*, 19(4):289–315, 1992.

[35] P. Diaconis and S. Evans. Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.*, 353(7):2615–2633 (electronic), 2001.

[36] P. Diaconis and D. Freedman. A dozen de Finetti-style results in search of a theory. *Ann. Inst. H. Poincaré Probab. Statist.*, 23(2, suppl.):397–423, 1987.

[37] P. Diaconis and S. Holmes, editors. *Stein's Method: Expository Lectures and Applications.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 46. Institute of Mathematical Statistics, Beachwood, OH, 2004. Papers from the Workshop on Stein's Method held at Stanford University, Stanford, CA, 1998.

[38] P. Diaconis and M. Shahshahani. On the eigenvalues of random matrices. *J. Appl. Probab.*, 31A:49–62, 1994. Studies in applied probability.

[39] R. M. Dudley. *Real Analysis and Probability.* Wadsworth & Brooks/Cole, Pacific Grove, California, 1989.

[40] G. Folland. How to integrate a polynomial over a sphere. *Amer. Math. Monthly*, 108(5):446–448, 2001.

[41] J. Fulman. Stein's method and non-reversible Markov chains. In *Stein's Method: Expository Lectures and Applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, pages 69–77. Inst. Math. Statist., Beachwood, OH, 2004.

[42] V. Girko. Strong circular law. *Random Oper. Stochastic Equations*, 5(2):173–196, 1997.

[43] L. Goldstein. Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.*, 42(3):661–683, 2005.

[44] L. Goldstein and G. Reinert. Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.*, 7(4):935–952, 1997.

[45] L. Goldstein and G. Reinert. Zero biasing in one and higher dimensions, and applications. In *Stein's Method and Applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 1–18. Singapore Univ. Press, Singapore, 2005.

[46] L. Goldstein and Y. Rinott. Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.*, 33(1):1–17, 1996.

[47] L. Goldstein and A. Xia. Zero biasing and a discrete central limit theorem. Preprint.

[48] F. Götze and A. N. Tikhomirov. Limit theorems for spectra of random matrices with martingale structure. In *Stein's Method and Applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 181–193. Singapore Univ. Press, Singapore, 2005.

[49] D. Hejhal and B. Rackner. On the topography of Maass waveforms for $\mathrm{PSL}(2,\mathbf{Z})$. *Experiment. Math.*, 1(4):275–305, 1992.

[50] F. Hiai and D. Petz. *The Semicircle Law, Free Random Variables and Entropy*, volume 77 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2000.

[51] W. Hoeffding. A combinatorial central limit theorem. *Ann. Math. Statistics*, 22:558–566, 1951.

[52] S. Holmes. Stein's method for birth and death chains. In *Stein's Method: Expository Lectures and Applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, pages 45–67. Inst. Math. Statist., Beachwood, OH, 2004.

[53] S. Holmes and G. Reinert. Stein's method for the bootstrap. In *Stein's Method: Expository Lectures and Applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, pages 95–136. Inst. Math. Statist., Beachwood, OH, 2004.

[54] M. Huber and G. Reinert. The stationary distribution in the antivoter model: exact sampling and approximations. In *Stein's Method: Expository Lectures and Applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, pages 79–93. Inst. Math. Statist., Beachwood, OH, 2004.

[55] T. Jiang. Maxima of entries of Haar distributed matrices. *Probab. Theory Related Fields*, 131(1):121–144, 2005.

[56] K. Johansson. On random matrices from the compact classical groups. *Ann. of Math. (2)*, 145(3):519–545, 1997.

[57] M. Kac. Can one hear the shape of a drum? *Amer. Math. Monthly*, 73(4, part II):1–23, 1966.

[58] B. Klartag. A central limit theorem for convex sets. Preprint, 2006.

[59] W.-L. Loh. Stein's method and multinomial approximation. *Ann. Appl. Probab.*, 2(3):536–554, 1992.

[60] H. Luk. *Stein's Method for the Gamma Distribution and Related Statistical Applications*. PhD thesis, University of Southern California, 1994.

[61] N. Matloff. Ergodicity conditions for a dissonant voting model. *Ann. Probability*, 5(3):371–386, 1977.

[62] E. Meckes. Unpublished manuscript.

[63] E. Meckes. Linear functions on the classical matrix groups. Preprint, 2006.

[64] E. Meckes and M. Meckes. The central limit problem for random vectors with symmetries, 2005. Preprint.

[65] J. Milnor. Eigenvalues of the Laplace operator on certain manifolds. *Proc. Nat. Acad. Sci. U.S.A.*, 51:542, 1964.

[66] M. Penrose. *Random Geometric Graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.

[67] A. Pickett. *Rates of Convergence of χ^2 Approximations via Stein's Method*. PhD thesis, University of Oxford, 2004.

[68] G. Reinert. Couplings for normal approximations with Stein's method. In *Microsurveys in Discrete Probability (Princeton, NJ, 1997)*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 193–207. Amer. Math. Soc., Providence, RI, 1998.

[69] M. Reitzner. Central limit theorems for random polytopes. *Probab. Theory Related Fields*, 133(4):483–507, 2005.

[70] Y. Rinott and V. Rotar. 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, 1996.

[71] Y. Rinott and V. Rotar. 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, 1997.

[72] Y. Rinott and V. Rotar. Normal approximations by Stein's method. *Decis. Econ. Finance*, 23(1):15–29, 2000.

[73] Z. Rudnick and P. Sarnak. The behaviour of eigenstates of arithmetic hyperbolic manifolds. *Comm. Math. Phys.*, 161(1):195–213, 1994.

[74] P. Sarnak. Arithmetic quantum chaos. In *The Schur Lectures (1992) (Tel Aviv)*, volume 8 of *Israel Math. Conf. Proc.*, pages 183–236. Bar-Ilan Univ., Ramat Gan, 1995.

[75] W. Schneller. Edgeworth expansions for linear rank statistics. *Ann. Statist.*, 17(3):1103–1123, 1989.

[76] Q. Shao and Z. Su. The berry-esseen bound for character ratios. Preprint, 2006.

[77] A. I. Šnirel′man. Ergodic properties of eigenfunctions. *Uspehi Mat. Nauk*, 29(6(180)):181–182, 1974.

[78] C. Stein. 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*, pages 583–602, Berkeley, Calif., 1972. Univ. California Press.

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

[80] C. Stein. The accuracy of the normal approximation to the distribution of the traces of powers of random orthogonal matrices. Technical Report No. 470, Stanford University Department of Statistics, 1995.

[81] N. Vilenkin. *Special Functions and the Theory of Group Representations.* Translated from the Russian by V. N. Singh. Translations of Mathematical Monographs, Vol. 22. American Mathematical Society, Providence, R. I., 1968.

[82] H. Weyl. The classical groups. *Princeton University Press*, 1946.

[83] S. Zelditch. Uniform distribution of eigenfunctions on compact hyperbolic surfaces. *Duke Math. J.*, 55(4):919–941, 1987.

[84] L. Zhao, Z. Bai, C.-C. Chao, and W.-Q. Liang. Error bound in a central limit theorem of double-indexed permutation statistics. *Ann. Statist.*, 25(5):2210–2227, 1997.