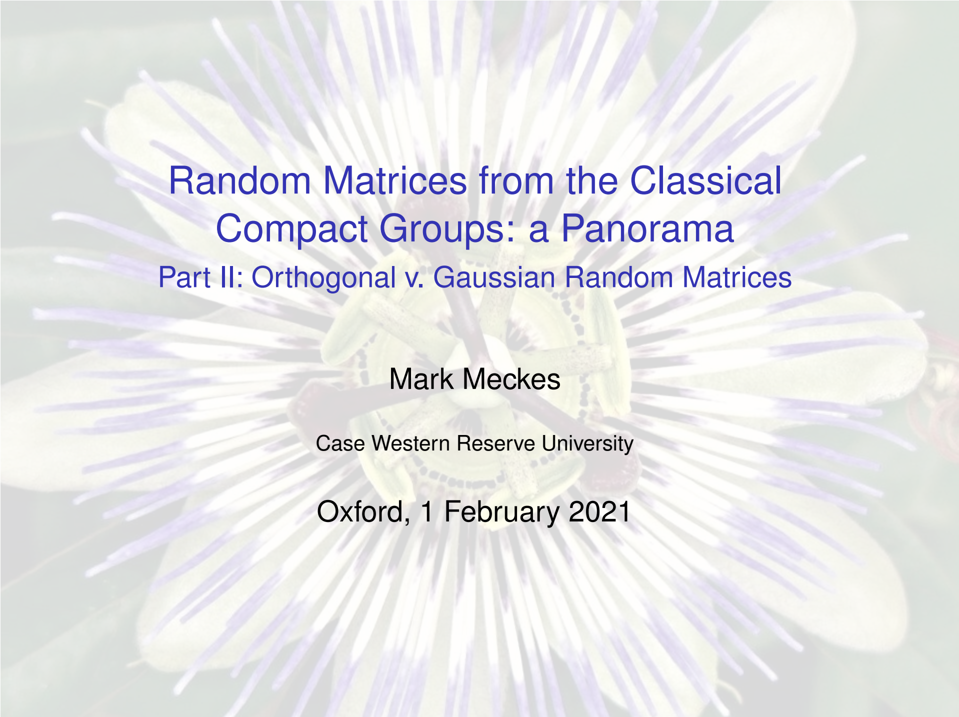

Random Matrices from the Classical Compact Groups: a Panorama

Part II: Orthogonal v. Gaussian Random Matrices

Mark Meckes

Case Western Reserve University

Oxford, 1 February 2021

Gaussian and orthogonal random matrices

Until further notice: U is random in $\mathbb{O}(n)$ (mainly for simplicity).

$G \in M_n(\mathbb{R})$ has independent $N(0,1)$ entries.

Today's slogan:

U and $\frac{1}{\sqrt{n}}G$ are a lot alike.

Caveats

- G has independent entries, U doesn't.
- U has bounded entries ($|u_{ij}| \leq 1$), G doesn't.
- U has orthogonal columns/rows, G doesn't.
- U acts as an isometry ($\|Ux\|_2 = \|x\|_2$), G doesn't.

So why is a raven like a writing-desk?

- U has almost independent entries.
- $\frac{1}{\sqrt{n}}G$ has almost bounded entries.
- G has almost orthogonal columns/rows.
- $\frac{1}{\sqrt{n}}G$ acts almost almost as an isometry.

Action on a single vector

Fix $x \in \mathbb{R}^n$. Then $Ux \sim \mathrm{unif}(\|x\|_2 \, S^{n-1})$.

$Gx \sim N(0, \|x\|_2^2 \, I_n)$, but more usefully:

- $Gx = \|Gx\|_2 \, \frac{Gx}{\|Gx\|_2}$,
- $\|Gx\|_2$ and $\frac{Gx}{\|Gx\|_2}$ are independent.
- $\frac{Gx}{\|Gx\|_2} \sim \mathrm{unif}(S^{n-1})$,
- $\mathbb{E} \, \|Gx\|_2 \approx \sqrt{n} \, \|x\|_2$ and $\|Gx_2\| \approx \sqrt{n} \, \|x\|_2$ with overwhelming probability (for large n).

More on this last point next time!

The caveats aren't all bad!

Ux and Gx are a lot alike, but Gx has independent entries.

Independence $\Rightarrow$ Gx is easy to calculate with.
"A lot alike" $\Rightarrow$ we can transfer results from Gx to Ux.

Simple example:

Write $\theta = Ue_1 \sim \mathrm{unif}(S^{n-1})$, $g = Ge_1 \sim N(0, I_n)$.
Then $g \sim R\theta$, with $R = \|g\|_2$ independent of θ.

$P(x) = x_1^{k_1} \cdots x_n^{k_n}$, $k = k_1 + \cdots + k_n$.

Then

$$\mathbb{E}P(g) = \mathbb{E}P(R\theta) = (\mathbb{E}R^k)(\mathbb{E}P(\theta))$$

and so

$$\mathbb{E}P(\theta) = \frac{\mathbb{E}P(g)}{\mathbb{E}R^k} = \frac{(\mathbb{E}g_1^{k_1}) \cdots (\mathbb{E}g_n^{k_n})}{\mathbb{E}R^k},$$

and the latter expectations are elementary.

Action on subspaces

The Grassmann manifold $G_{n,k}$ is the set of all k-dimensional linear subspaces of $\mathbb{R}^n$.

$G_{n,k}$ has a Haar measure: a unique notion of a random k-dimensional subspace $E \subseteq \mathbb{R}^n$ such that $V(E) \sim E$ for each fixed $V \in \mathbb{O}(n)$.

If $F \in G_{n,k}$ is fixed, then

$$U(F) \sim G(F) \sim E.$$

Norms

Theorem (Folklore?)

There are constants $c, C > 0$ such that if $\|\cdot\|$ is any seminorm on $M_n(\mathbb{R})$, then

$$c\mathbb{E}\,\|U\| \le \frac{1}{\sqrt{n}}\mathbb{E}\,\|G\| \le C\mathbb{E}\,\|U\|.$$

Again, this is mostly used in situations where

- we want to work with $\mathbb{E}\,\|U\|$, but
- $\mathbb{E}\,\|G\|$ is easier to estimate.

The form of this result is typical of nonasymptotic random matrix theory.

Asymptotic distribution of single entries

Let $\theta = Ue_1 \sim \mathrm{unif}(S^{n-1})$, $g = Ge_1 \sim N(0, I_n)$.

Recall:

- $g \sim R\theta$,
- $R = \|g\|_2$ and θ are independent.
- $R \approx \sqrt{n}$ with high probability (for large n).

This suggests:

For large n, entries of $\sqrt{n}\theta$ are distributed similarly to entries of g, i.e., independent standard Gaussians.

This was apparently first observed by Maxwell, and a rigorous version (for one entry) was first proved by Borel.

In their honor it is sometimes referred to as the "Poincaré limit".

Asymptotic distribution of entries in one column

The total variation distance between random vectors $X \sim f_X$ and $Y \sim f_Y$ is

$$\begin{aligned} d_{TV}(X, Y) &= 2 \sup_A |\mathbb{P}[X \in A] - \mathbb{P}[Y \in A]| \\ &= \sup_{\|\psi\|_\infty \leq 1} |\mathbb{E}\psi(X) - \mathbb{E}\psi(Y)| \\ &= \int_{\mathbb{R}^n} |f_X - f_Y| . \end{aligned}$$

Theorem (Diaconis–Freedman)

$$d_{TV}\left(\sqrt{n}(\theta_1, \ldots, \theta_k), (g_1, \ldots, g_k)\right) \leq C\frac{k}{n}.$$

In this sense, as many as $o(n)$ entries of $\theta = Ue_1$ are asymptotically independent Gaussians.

Asymptotic distribution of a submatrix

Recall: the upper-left $p \times q$ submatrix M of U has a density proportional to

$$\det(I_q - M^T M)^{\frac{n-p-q-1}{2}}.$$

For fixed p, q and $n \to \infty$, the density of $\sqrt{n}M$ looks like

$$\begin{aligned}
\det\left(I_q - \frac{1}{n}M^T M\right)^{\frac{n-p-q-1}{2}} &= \prod_{i=1}^{q}\left(1 - \frac{1}{n}\lambda_i(M^T M)\right)^{\frac{n-p-q-1}{2}} \\
&\approx \prod_{i=1}^{q} e^{-\lambda_i(M^T M)/2} \\
&= e^{-\frac{1}{2}\operatorname{Tr} M^T M} = e^{-\frac{1}{2}\|M\|_{HS}^2}.
\end{aligned}$$

This suggests there is a matrix version of the Diaconis–Freedman result.

Asymptotic distribution of a submatrix: total variation

Theorem (Jiang–Ma, Stewart)

Let p, q depend on n with $pq \xrightarrow{n\to\infty} \infty$ and $pq = o(n)$.
Let M_n be the upper-left $p \times q$ submatrix of $U \in \mathbb{O}(n)$, and let G_n be a $p \times q$ Gaussian random matrix.
Then $d_{TV}(\sqrt{n}M_n, G_n) \xrightarrow{n\to\infty} 0$.

This result is sharp: if $\limsup_{n\to\infty} \frac{pq}{n} > 0$, then $d_{TV}(\sqrt{n}M_n, G_n) \not\to 0$.

Basic idea of proof:

$$d_{TV}(\sqrt{n}M_n, G_n) = \int |f_M - f_G| = \int \left| \frac{f_M}{f_G} - 1 \right| f_G = \mathbb{E} \left| \frac{f_M(G)}{f_G(G)} - 1 \right|$$

followed by a much more precise version of the asymptotic analysis on the last slide.

Asymptotic distribution of a submatrix: in probability

Theorem (Jiang)

Let G be an $n \times n$ Gaussian random matrix, and let $U \in \mathbb{O}(n)$ be obtained from G by the Gram–Schmidt process.
Then

$$\mathbb{P}\left[\max_{1\le i\le n,\ 1\le j\le m} \left|\sqrt{n}u_{ij} - g_{ij}\right| > \varepsilon\right] \xrightarrow{n\to\infty} 0$$

for every $\varepsilon > 0$ if and only if $m = o\left(\frac{n}{\log n}\right)$.

The proof consists of a careful probabilistic analysis of the Gram–Schmidt process.

Arbitrary linear functions of the entries

If $A \in M_n(\mathbb{R})$ is fixed then

$$\operatorname{Tr} AU = \left\langle U, A^T \right\rangle = \sum_{i,j=1}^{n} a_{ji} u_{ij}.$$

Theorem (E. Meckes)

If $\|A\|_{HS} = \sqrt{n}$ and $g \sim N(0,1)$, then

$$d_{TV}(\operatorname{Tr} AU, g) \leq \frac{C}{n}.$$

The proof is by Stein's method.

Arbitrary linear projections

The L^1-Wasserstein distance between random vectors X and Y is

$$W_1(X, Y) = \sup_{\|\psi\|_L \leq 1} |\mathbb{E}\psi(X) - \mathbb{E}\psi(Y)|,$$

where $\|\psi\|_L = \sup_{x \neq y} \frac{|\psi(x) - \psi(y)|}{\|x - y\|_2}$.

Theorem (Chatterjee–E. Meckes)

Suppose $A_1, \ldots, A_k \in M_n(\mathbb{R})$ are orthogonal (w.r.t. the Hilbert–Schmidt inner product) with $\|A_i\|_{HS} = \sqrt{n}$, and $g \sim N(0, I_k)$. Then

$$W_1\left((\operatorname{Tr} A_1 U, \ldots, \operatorname{Tr} A_k U), g\right) \leq C\frac{k}{n}.$$

In this sense, arbitrary projections of dimension $k = o(n)$ are asymptotically Gaussian.

Traces

When $A = I_n$, the Gaussian approximation is amazingly good:

Theorem (Johansson)

$d_{TV}(\operatorname{Tr} U, g) \leq Ce^{-cn}$ *if* $U \in \mathbb{O}(n)$.

$$d_{TV}\left(\operatorname{Tr} U, \frac{1}{\sqrt{2}}(g_1 + ig_2)\right) \leq Cn^{-cn} \text{ if } U \in \mathbb{U}(n).$$

The proof is by Fourier analysis, and we'll return to the result itself later.

For $\mathbb{U}(n)$, Keating–Mezzadri–Singphu prove a rate of convergence for $\operatorname{Re}\operatorname{Tr} AU$ to Gaussian which

- depends on the singular values of A,
- is $O(n^{-1})$ always, and
- is $O(n^{-2})$ when the singular values of A are bounded (as for $A = I_n$).

Additional references

- Stanisław Szarek, "Condition numbers of random matrices", *Journal of Complexity* 7 no. 2, pp. 131–149, 1991.

- Joel Tropp, "A comparison principle for functions of a uniformly random subspace", *Probab. Theory Related Fields* 153 no. 3–4, pp. 759–769, 2012.

- Zakhar Kabluchko, Joscha Prochno, and Christoph Thäle, "A new look at random projections of the cube and general product measures", to appear in *Bernoulli*.