Applied Random Matrix Theory

Joel A. Tropp
Steele Family Professor of
Applied & Computational Mathematics

Computing + Mathematical Sciences
California Institute of Technology
jtropp@cms.caltech.edu

Research supported by ONR, AFOSR, NSF, DARPA, Sloan, and Moore.
What is a Random Matrix?

Definition. A random matrix is a matrix whose entries are random variables, not necessarily independent.

A random matrix in captivity:

\[
\begin{pmatrix}
0.0000 & -1.3077 & -1.3499 & 0.2050 & 0.0000 \\
1.8339 & 0.0000 & -1.3077 & 0.0000 & 0.2050 \\
-2.2588 & 1.8339 & 0.0000 & -1.3077 & -1.3499 \\
2.7694 & 0.0000 & 1.8339 & 0.0000 & -1.3077 \\
0.0000 & 2.7694 & -2.2588 & 1.8339 & 0.0000
\end{pmatrix}
\]

What do we want to understand?

- Eigenvalues
- Eigenvectors
- Singular values
- Singular vectors
- Operator norms
- ...

Sources: Muirhead 1982; Mehta 2004; Nica & Speicher 2006; Bai & Silverstein 2010; Vershynin 2010; Tao 2011; Kemp 2013; Tropp 2015; ...
Random Matrices in Statistics


A comparison of equation (8) with the corresponding results (1) and (2) for uni-variate and bi-variate sampling, respectively, indicates the form the general result may be expected to take. In fact, we have for the simultaneous distribution in random samples of the $n$ variances (squared standard deviations) and the $\frac{n(n-1)}{2}$ product moment coefficients the following expression:

$$d_p = \left(\frac{n}{\pi}\right)^{\frac{n(n-1)}{2}} \prod_{i=1}^{n} \Gamma\left(\frac{N-1}{2}\right) \Gamma\left(\frac{N-2}{2}\right) \cdots \Gamma\left(\frac{N-n}{2}\right)$$

$$e^{-\frac{1}{2} \sum_{m=1}^{n} c_m^2} \cdot \prod_{m=1}^{n} \Delta_{m,m}$$

$$\left| \begin{array}{cccc}
    A_{11} & A_{12} & \cdots & A_{1n} \\
    A_{21} & A_{22} & \cdots & A_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    A_{n1} & A_{n2} & \cdots & A_{nn} \\
\end{array} \right|$$

$$\left| \begin{array}{cccc}
    a_{11} & a_{12} & \cdots & a_{1n} \\
    a_{21} & a_{22} & \cdots & a_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{n1} & a_{n2} & \cdots & a_{nn} \\
\end{array} \right|$$

where $c_{pq} = s_p s_q r_{pq}$, and $A_{pq} = \frac{\Delta_{pq}}{\Delta} \cdot \Delta_{pq}$, $\Delta$ being the determinant

$$|\rho_{pq}|, p, q = 1, 2, 3, \ldots n,$$

and $\Delta_{pq}$ the minor of $\rho_{pq}$ in $\Delta$.

---

John Wishart

Sample covariance matrix for the multivariate normal distribution

Sources: Wishart, Biometrika 1928. Photo from apprendre-math.info.

Joel A. Tropp (Caltech), Applied RMT, Foundations of Computational Mathematics (FoCM), Barcelona, 13 July 2017
now combining (8.6) and (8.7) we obtain our desired result:

\[
\text{Prob } (\lambda > 2\sigma^2 r n) < \frac{(r n)^{n-1/2} e^{-r n \pi^{1/2} e^{n}} \cdot 2^{n-2}}{\pi n^{n-1} (r - 1)^n}
\]

(8.8)

\[
= \left( \frac{2r}{e^{r-1}} \right)^n \times \frac{1}{4(r - 1)(r\pi n)^{1/2}}. 
\]

We sum up in the following theorem:

(8.9) The probability that the upper bound $|A|$ of the matrix $A$ of (8.1) exceeds $2.72\sigma n^{1/2}$ is less than $0.027 \times 2^{-n-1/2}$, that is, with probability greater than $99\%$ the upper bound of $A$ is less than $2.72\sigma n^{1/2}$ for $n = 2, 3, \cdots$

This follows at once by taking $r = 3.70$.
Random Matrices in Nuclear Physics

Random sign symmetric matrix

The matrices to be considered are $2N + 1$ dimensional real symmetric matrices; $N$ is a very large number. The diagonal elements of these matrices are zero, the non diagonal elements $v_{ik} = v_{ki} = \pm v$ have all the same absolute value but random signs. There are $R = 2^{N(2N+1)}$ such matrices. We shall calculate, after an introductory remark, the averages of $(H')_{00}$ and hence the strength function $S'(x) = \sigma(x)$. This has, in the present case, a second interpretation: it also gives the density of the characteristic values of these matrices. This will be shown first.

Eugene Wigner

Model for the Hamiltonian of a heavy atom in a slow nuclear reaction

Classical RMT

\[
\begin{bmatrix}
0 & + & - & + & + & - & + \\
0 & + & - & - & - & - & + \\
0 & + & - & + & + & & \\
0 & - & - & - & & & \\
& 0 & + & - & & & \\
& & 0 & + & & & \\
& & & & & & 0 \\
\end{bmatrix}
\]

Wigner \((n = 7)\)

Distribution of eigenvalues \((n = 10^3)\)

- Highly symmetric models
- Very precise results
- Strong resonances with other fields of mathematics
Contemporary Applications of RMT

- Numerical linear algebra
- Numerical analysis
- Uncertainty quantification
- High-dimensional statistics
- Econometrics
- Approximation theory
- Sampling theory
- Machine learning
- Learning theory
- Mathematical signal processing
- Optimization
- Computer graphics and vision
- Quantum information theory
- Theory of algorithms
- Combinatorics
- ...


Per Google Scholar, at least 26,100 papers on RMT since 2000! Equivalent to search for donald trump junior fredo corleone.
Contemporary RMT

\[
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
\end{bmatrix}
\]

(sample random columns)

\[
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\
0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
\end{bmatrix}
\]

- Wide range of examples, many data-driven
- Results may sacrifice precision for applicability
- Theory is still developing
Modern applications demand new random matrix models and new analytical tools
Goal: For a random matrix $Z$, find probabilistic bounds for

$$\| Z - \mathbb{E} Z \|$$

An upper bound on this quantity ensures that

- Singular values of $Z$ and $\mathbb{E} Z$ are close
- Singular vectors of $Z$ and $\mathbb{E} Z$ are close (for isolated singular values)
- Linear functionals of $Z$ and $\mathbb{E} Z$ are close
- Spectral norm of $Z$ is controlled: $\| Z \| = \| \mathbb{E} Z \| \pm \| Z - \mathbb{E} Z \|$

$\| \cdot \| = $ spectral norm = largest singular value = $\ell_2$ operator norm
The Independent Sum Model

\[ Z = \sum_k S_k \]

with \( S_k \) independent

**Useful observation:** \( \mathbb{E} Z = \sum_k \mathbb{E} S_k \)

**Exercise:** Express the sample covariance matrix in this model

**Exercise:** Express column sampling (with replacement) from a fixed matrix
The Bernstein Inequality

Fact 1 (Bernstein 1920s). Suppose

- $S_1, S_2, S_3, \ldots$ are independent real random variables
- Each one is centered: $\mathbb{E} S_k = 0$
- Each one is bounded: $|S_k| \leq L$

Then, for $t > 0$,

$$
\mathbb{P}\left\{ \left| \sum_k S_k \right| \geq t \right\} \leq 2 \cdot \exp\left( \frac{-t^2/2}{v + Lt/3} \right)
$$

where the variance proxy is

$$
v = \text{Var} \left( \sum_k S_k \right) = \sum_k \mathbb{E} S_k^2
$$

Sources: Bernstein 1927; Boucheron et al. 2013.
The Matrix Bernstein Inequality I

Theorem 2 (T 2011). Suppose

- \( S_1, S_2, S_3, \ldots \) are independent random matrices with dimension \( d_1 \times d_2 \)
- Each one is centered: \( \mathbb{E} S_k = 0 \)
- Each one is bounded: \( \| S_k \| \leq L \)

Then, for \( t > 0 \),

\[
\mathbb{P} \left\{ \left\| \sum_k S_k \right\| \geq t \right\} \leq (d_1 + d_2) \cdot \exp \left( \frac{-t^2/2}{\nu + Lt/3} \right)
\]

where the matrix variance proxy is

\[
\nu = \max \left\{ \left\| \sum_k \mathbb{E}(S_k^* S_k) \right\|, \left\| \sum_k \mathbb{E}(S_k^* S_k^*) \right\| \right\}
\]

The Matrix Bernstein Inequality II

Theorem 3 (T 2011). Suppose

- S₁, S₂, S₃, … are independent random matrices with dimension d₁ × d₂
- Each one is centered: E Sₖ = 0
- Each one is bounded: ∥Sₖ∥ ≤ L

Then

\[ E \left\| \sum_k S_k \right\| \leq \sqrt{2v \cdot \log(d_1 + d_2)} + \frac{1}{3} L \cdot \log(d_1 + d_2) \]

where the matrix variance proxy is

\[ v = \max \left\{ \left\| \sum_k E(S_k S_k^*) \right\| , \left\| \sum_k E(S_k^* S_k) \right\| \right\} \]

Example: Matrix Sparsification

\[
A = \begin{bmatrix}
1 & 2 & 3 & 4 \\
2 & 4 & 6 & 8 \\
3 & 6 & 9 & 12 \\
4 & 8 & 12 & 16
\end{bmatrix}
\quad \rightarrow \quad
\hat{A} = \begin{bmatrix}
2 \\
4 \\
3 & 6 & 9 & 12 \\
12 & 16
\end{bmatrix}
\]

**Goal:** Find a sparse matrix \( \hat{A} \) for which \( \| A - \hat{A} \| \) is small

**Approach:** Non-uniform randomized sampling

Let $A$ be a fixed $d_1 \times d_2$ matrix

Construct a probability mass $\{p_{ij}\}$ on the matrix indices

Define a 1-sparse random matrix $S$ where

$$S = \frac{a_{ij}}{p_{ij}} E_{ij} \quad \text{with probability } p_{ij}$$

The random matrix $S$ is an unbiased estimator for $A$

$$\mathbb{E} S = \sum_{ij} \frac{a_{ij}}{p_{ij}} E_{ij} \cdot p_{ij} = \sum_{ij} a_{ij} E_{ij} = A$$

To reduce the variance, average $r$ independent copies of $S$

$$\hat{A}_r = \frac{1}{r} \sum_{k=1}^r S_k \quad \text{where } S_k \sim S$$

By construction, $\hat{A}_r$ has at most $r$ nonzero entries and approximates $A$
Sparsification: Analysis

Recall: $S = (a_{ij} / p_{ij})E_{ij}$ with probability $p_{ij}$

Bound for spectral norm:

$$\|S - E S\| \leq 2 \cdot \max_{ij} \frac{|a_{ij}|}{p_{ij}}$$

Bound for variance:

$$\|E (S - E S)(S - E S)^*\| \leq \|E SS^*\| = \left\| \sum_i \left( \sum_j \frac{|a_{ij}|^2}{p_{ij}} \right) E_{ii} \right\| = \max_i \sum_j \frac{|a_{ij}|^2}{p_{ij}}$$

$$\|E (S - E S)^*(S - E S)\| \leq \|E S S^*\| = \left\| \sum_j \left( \sum_i \frac{|a_{ij}|^2}{p_{ij}} \right) E_{jj} \right\| = \max_j \sum_i \frac{|a_{ij}|^2}{p_{ij}}$$

Construct probability mass $p_{ij} \propto |a_{ij}| + |a_{ij}|^2$ to control all terms
Proposition 4 (Kundu & Drineas 2014; T 2015). Suppose

\[ r \geq \varepsilon^{-2} \cdot \text{srank}(A) \cdot \max\{d_1, d_2\} \log(d_1 + d_2) \]  

(0 < \varepsilon \leq 1)

Then the relative error in the \( r \)-sparse approximation \( \hat{A}_r \) satisfies

\[ \frac{E \| A - \hat{A}_r \|}{\| A \|} \leq 4\varepsilon \]

The stable rank

\[ \text{srank}(A) := \frac{\| A \|_F^2}{\| A \|_2^2} \leq \text{rank}(A) \]

The proof is an immediate consequence of matrix Bernstein
Theorem 5 (Kyng & Sachdeva 2016). Suppose

- $G$ is a weighted, undirected graph with $n$ vertices and $m$ edges
- $L$ is the combinatorial Laplacian of the graph $G$

Then, with high probability, the SPARSECHOLESKY algorithm produces

- A lower-triangular matrix $C$ with $O(m \log^3 n)$ nonzero entries that satisfies

$$\frac{1}{2}L \preceq CC^* \preceq \frac{3}{2}L$$

- The running time is $O(m \log^3 n)$

In particular, we can solve $Lx = b$ to relative error $\epsilon$ in time $O(m \log^3 n \log(1/\epsilon))$
**SparseCholesky** (Caricature)

\[
L = \begin{bmatrix} a & u^* \\ u & L_2 \end{bmatrix}_{n \times n} \rightarrow L_{2} - a^{-1} \begin{bmatrix} uu^* \end{bmatrix}_{(n-1) \times (n-1)}
\]

Subtract rank-1

\[
\rightarrow L_{2} - a^{-1} \begin{bmatrix} \times & \times \\ \times & \times \end{bmatrix}_{(n-1) \times (n-1)}
\]

Sparsify rank-1

- Direct computation of Cholesky factorization requires \(O(n^2)\) operations per step
- Randomized approximation in \(O((m/n) \log^3 n)\) operations per step (amortized)
- Sampling probabilities are computed using graph theory
- Proof depends on Bernstein inequality for matrix martingales!

A Virtuous Cycle

Models $\rightarrow$ Theory

Applications
Workshop B5: Random Matrices

Organizers: Michel Ledoux, Sheehan Olver, Joel A. Tropp

Semi-plenaries:

- Ioana Dumitriu: "Spectra of Random Regular and Quasi-Regular Graphs"
- Amit Singer: "Variations on PCA"

Talks:

- Folkmar Bornemann
- Djalil Chafai
- Alan Edelman
- Noureddine El Karoui
- Elizabeth Meckes
- Mark Meckes
- Ramis Movassagh
- Raj Rao Nadakuditi
- Jelani Nelson
- Konstantin Tikhomirov
- Thomas Trogdon
- Ke Wang

Poster: Plamen Koev
email: jtropp@cms.caltech.edu
web: http://users.cms.caltech.edu/~jtropp

Monograph:

Papers:
- “Freedman's inequality for matrix martingales.” *ECP*, 2011
- “From the joint convexity of relative entropy to a concavity theorem of Lieb.” *PAMS*, 2012
- “Improved analysis of the subsampled randomized Hadamard transform.” *AADA*, 2011
- “The masked sample covariance estimator” with R. Chen & A. Gittens. *I&I*, 2012
- “Tail bounds for all eigenvalues of a sum of random matrices” with A. Gittens. Caltech ACM Report 2014-02
- “Subadditivity of matrix $\phi$-entropy and concentration of random matrices” with R. Chen. *EJP*, 2014
- “Second-order matrix concentration inequalities.” *ACHA*, 2016